Reliability Analysis of Redundant Arrays of Inexpensive Disks*

MANISH MALHOTRA
Department of Computer Science, Duke University, Durham, North Carolina 27706

AND

KISHOR S. TRIVEDI
Department of Electrical Engineering, Duke University, Durham, North Carolina 27706

A reliability analysis of various disk array architectures (different levels of RAID) is performed. The dependence of reliability and mean time to data loss on various parameters of a disk array is characterized. A study of these characteristics reveals the impact of several design choices for a disk array on its reliability. Issues such as scalability of disk arrays, imperfect coverage of disk failures, cold versus hot spares, effect of predictive disk failures, and dependence of disk array reliability on data reconstruction time are studied. © 1993 Academic Press, Inc.

1. INTRODUCTION

To achieve high I/O performance, disk array architectures have been proposed. Unfortunately, an array of disks is more fault-prone than a single large drive [9] and may possess poor reliability. Failure of a disk may result in loss or corruption of data which may have been accumulated through years of research efforts and extensive experimentation. Thus, the demand is to design cost-effective disk systems which provide not only high performance but also high reliability.

Several fault-tolerant disk array architectures have been introduced by different researchers using varying degrees of hardware redundancy [1, 5, 7, 9, 11]. Patterson et al. [9] coined the term RAID (Redundant Array of Inexpensive Disks) for such disk array systems with redundancy. They unified the existing disk system architectures as different levels of RAID (levels 1, 2, 3, 4) and proposed a new high-performance architecture (RAID level 5). Gibson [2] and Patterson et al. [9] have analyzed reliability of RAID in terms of mean time to data loss (MTDL). Bitton and Gray [1] have analyzed MTTF for mirrored disks (RAID-1). Gibson and Patterson [4] provide a comprehensive analysis of RAID-1 and RAID-5 reliability. They compute MTDL for disk array models based on a simple approximation. Assuming that the time to failure of a group of disks is exponentially distributed, they compute the reliability of disk arrays using the approximate value of MTDL. The usefulness of these approximations is that MTDL and reliability can be expressed in closed form. However, it is known that the time to failure of a group of disks is not exponentially distributed even when individual disk failure times are.

In this paper, we quantify the reliability and mean time to data loss of disk array architectures without making the approximation used in [4]. We develop hierarchical reliability models for RAID-1, 2, 3, 4, 5, and use these models to answer important questions arising in the design of these architectures. To make the models realistic, we take into account several factors which have hitherto been not considered; mainly imperfect coverage of disk failures, predictive disk failures, and the type of spares (cold or hot). In Section 2, we describe a hierarchical reliability model for a general disk array architecture. In Section 3, we develop reliability models for different RAID architectures with finite number of cold and hot spares. In Section 4, numerical results are presented and discussed. Conclusions are presented in Section 5.

2. FAULT-TOLERANT DISK ARRAY ARCHITECTURES

To begin with, we introduce the terminology and state the assumptions. The disk array consists of N groups of D data disks and C check disks each. Time to failure of each disk is exponentially distributed with mean 1/λ (MTTF). Failure of a single disk in a group is tolerable since lost data can be reconstructed. Data reconstruction consists of replacement of failed disk by a spare and data construction on the spare disk using the data from the working disks and parity information of the group. However, if another disk in the same group fails while the reconstruction is underway, then data are lost (cannot be reconstructed) and that group of disks is considered failed. Data loss in any group implies the failure of the

* This work was supported in part by the National Science Foundation under Grant CCR-9108114 and by the Naval Surface Warfare Center under Grant N00014-91-J-4162.
entire disk array. Initially, we assume that each group has its own reconstruction mechanism independent of other disks and an unlimited supply of disk spares. Later we consider finite number of disk spares. All the disks are identical (come from the same manufacturer).

The data reconstruction time is assumed to be exponentially distributed with mean $1/\mu$. In reality, data reconstruction time is closer to deterministic. If a constant data reconstruction time is assumed, then our models become semi-Markov. Solution of semi-Markov models yields results that closely match the results obtained from Markov models in this case. Gibson and Patterson [4] also show that the assumption of exponential distribution yields results which closely match those obtained by simulation where data reconstruction time is taken to be constant.

Faults in a disk are covered with probability $p$ and uncovered with probability $1 - p$. An uncovered fault in a group causes data loss. Bit errors that are not detected, not corrected, or miscorrected are manifestations of what we call uncovered faults. These faults may also occur due to some fault within the error-correcting code circuitry (ECC) or if ECC is not properly invoked. Failure of support hardware components (such as power supply), catastrophic failures due to extreme environmental conditions, and occurrence of multiple bit errors (since ECC can correct only single bit errors) can also be accounted for by imperfect coverage.

2.1. RAID-1 (Mirrored Disks)

Mirroring is the traditional approach to improving the reliability of disk systems [1]. If the storage capacity of the system requires $N$ disks, then $2N$ disks are used in a mirrored disk system. This is the most expensive of the different RAID architectures. Each pair of mirrored disks forms a group. If one of the disks in a group fails, then a spare is switched in. The data reconstruction consists of copying the data onto the spare from the other working disk.

2.2. RAID-2

In this scheme, each group has $D$ data disks and $C$ (where $C \geq \log_2(D + C + 1)$) check disks. The check disks can correct single bit errors and detect double bit errors. A single failure of any disk in a group is tolerable.

2.3. RAID-3,4,5

The $C$ check disks for $D$ data disks in RAID-2 are basically needed to detect the incorrect bit position. Once the incorrect bit has been identified, then a single parity bit suffices for correction (reconstruction) of data which would otherwise be permanently lost. In RAID levels 3,4,5, the ability of the disk controller to detect a failed disk is utilized. Thus, we need only one check disk per group since the disk controller identifies the failed bit position. A RAID-3 architecture was proposed by Park and Balasubramaniam [8] and a RAID-4 architecture was proposed by Salem and Garcia-Molina [11]. RAID-3 allows only one I/O transfer per unit time per group. RAID-4 allows parallel transfers from a group; however, only the reads are parallelized. To parallelize writes, RAID-5 architecture was proposed by Patterson et al. [9]. In this scheme, parity information is spread across all the disks within a group (rotated parity). This scheme results in improved performance for reads as well as writes. However, the reliability models for RAID-3, RAID-4, and RAID-5 are identical since they all require equal number of disks per group.

RAID-3,4,5, rely on the disk controller’s ability to correctly predict the disk failures before they occur. Some implementations utilize this property to prevent data loss and reduce the reconstruction time. Assume that no loss of data occurs if the disk controller correctly predicts an impending disk failure. Further assume that the spare is electronically switched in and data is copied onto the spare before the failing disk is powered down. This sequence of operations does not result in a change of state of the system. However, the disk controller may not always be able to predict a disk failure. Failures resulting from uncovered faults are not predictable. With probability $(1 - \alpha)$, an impending failure due to a covered fault is not predicted.

There is also the possibility of false alarms when the disk controller erroneously predicts a disk failure. The time to next false alarm is assumed to be exponentially distributed with rate $\gamma$. However, false alarms are treated as correctly predicted failures and do not result in a change in system state. This is because we assume an unlimited supply of disk spares. However, they do result in monetary loss since a false alarm results in undue consumption of disk spares. In the next section, when we consider finite numbers of spares, the effect of false alarms becomes clear.

2.4. A General Reliability Model

Based on the above assumptions, we construct a two-level hierarchical reliability model which fits different RAID architectures. The overall reliability of the disk array is modeled by the reliability block diagram (RBD) shown in Fig. 1. Each block represents a group of disks. Assume that groups behave independently of each other. If $R_i(t)$ is the reliability of group $i$, then reliability of the

![FIG. 1. Reliability block diagram for RAID.](image-url)
disk array is given by
\[ R_{da}(t) = \prod_{i=1}^{N} R_i(t). \] (1)

To compute the reliability of a group, we use the simple Markov model shown in Fig. 2. In state 2, all the disks in a group are operational. After one of the disks fails, system state changes from 2 to 1 and data reconstruction is initiated. However, the disk array keeps functioning since data is available. If any other disk in the group fails before the reconstruction is completed, then data is lost and the disk array is considered failed. State 0 is the group failed state.

Closed form expressions for reliability \( R_i(t) \) of any group \( G_i \) can be obtained by using the Laplace transform approach [13]. It is given by
\[ R_i(t) = A_1 e^{\beta_1 t} + A_2 e^{\beta_2 t}, \] (2)

where
\[
\begin{align*}
\beta_1 &= \frac{-((D + C)(2 - p\alpha) - 1)\lambda + \mu + \sqrt{X}}{2}, \\
\beta_2 &= \frac{-((D + C)(2 - p\alpha) - 1)\lambda + \mu - \sqrt{X}}{2}, \\
A_1 &= \frac{\beta_1 + \mu + (D + C - 1)\lambda}{\beta_1 - \beta_2}, \\
A_2 &= \frac{\beta_2 + \mu + (D + C - 1)\lambda}{\beta_2 - \beta_1}, \\
X &= ((D + C)((D + C)p^2\alpha^2 - 2p\alpha + 1)\lambda^2 + \mu^2 + 2((D + C)p(2 - \alpha) - 1)\lambda\mu.
\end{align*}
\]

Mean time to data loss for a group of disks is given by
\[ MTDL_i = \int_0^\infty R_i(t) \, dt \]
\[ = \frac{\mu + ((D + C)(1 + p(1 - \alpha)) - 1)\lambda}{(D + C)\lambda(\mu(1 - p) + (D + C - 1)(1 - \alpha))}. \] (3)

The mean time to data loss for the disk array is given by
\[ MTDL_{da} = \int_0^\infty R_{da}(t) \, dt = \sum_{j=0}^{N} \frac{C_j}{\beta_1 j + \beta_2(N - j)}, \] (4)

The reliability and MTDL for different RAID architectures can be obtained by substituting the following values. For RAID-1, \( \alpha = 0 \) and \( D = C = 1 \). For RAID-2, \( \alpha = 0 \) and \( C \equiv \log_2(C + D + 1) \). For RAID-3,4,5, \( C = 1 \).

3. COLD VERSUS HOT DISK SPARES

In the previous section, we assumed unlimited supply of disk spares that did not fail. In reality, however, a fixed number of spares is maintained. The disk spares could be maintained hot or cold. A hot spare can fail even though it is not in active use. A cold spare does not fail unless it is switched in as a replacement for a failed disk. A hot spare can be switched in electronically after a disk fails and the time to perform the switch-in is negligible. The disadvantages of hot spares are: (1) The automated switch-in mechanism adds to the cost overhead, (2) Spares can fail while not in active use, and (3) The hardware used to carry out spare switch-in may fail (these failures can be accounted for by the coverage probability).

If cold spares are maintained, then a repairperson is called to install a spare after a disk fails. After installation, the disk reconfiguration and data reconstruction begin. Thus the total repair time increases. A better solution perhaps is a combination of two approaches. Few hot spares could be maintained, while the rest are kept cold. Each time a disk fails, a hot spare is used up and a cold spare is made hot. If a disk fails after all the spares are exhausted, then a new disk is ordered from the manufacturer. This increases the data reconstruction time. In practice, this could be avoided by always maintaining a minimum number of spares. Each time the number of available spares falls below this minimum, new disks can be ordered from the manufacturer. This scenario well approximates the case of unlimited spares.

In this section, we develop reliability models for different RAID architectures based upon different assumptions. Assume that each group has \( M \) spare disks and time to switch in a hot spare is negligible. Let us first consider the reliability model of a group of disks with \( M \) hot spares for RAID-3,4,5, shown in Fig. 3.

A state is a two-tuple \((i, j)\) where \( i \) is the number of active disks (operational data and check disks) and \( j \) is the number of spares left. State \( A \) is the group failed state. In this model, \( G = D + C \), the number of active disks, \( \lambda_1 = (D + C)\lambda_4 p(1 - \alpha) \), where \( \lambda_4 \) is the failure rate of a disk in active use, \( \lambda_4 = (D + C - 1)\lambda_4 \), and \( \lambda_1 = (D + C)\lambda_4 (1 - p) \). The transitions with rate \( \lambda_4 \) are transi-
tions representing failure of a disk during data reconstruction which results in data loss. Transitions with rate \( \lambda_1 \) and \( \lambda_1 \) respectively represent uncovered and covered failures of a disk. Transitions from states \((G-1, M-i+1)\) to states \((G-1, M-i)\) \((i = 1, ..., M)\) signify failure of hot spares. This also changes the transition rates from states \((G, M-i+1)\) to \((G, M-i)\) \((i = 1, ..., M)\). Rate \( \lambda_{si} = \lambda_{sp} + (D + C)\lambda_{4} p\alpha + \gamma \) where \( \lambda_{sp} \) is the failure rate of a hot spare such that \( \lambda_{sp} < \lambda_{4} \). \( \alpha \) and \( \gamma \) are the same as defined in Section 2. We assume that data-reconstruction time for correctly predicted failures and false alarms is negligible since a spare is installed and data-reconstruction completed before the failing disk is powered down. Thus, these events result only in a state change reflecting the decrease in the number of available spares by one. Data-reconstruction rate while at least one spare is available is \( \mu_1 \). In state \((G, 0)\), all the spares are exhausted. If a disk fails in this state, the data-reconstruction rate is \( \mu_2 \) where \( \mu_2 < \mu_1 \) because mean data-reconstruction time increases.

This model is specialized to all the other cases by slight modifications. In the case of RAID-1,2, with hot spares, the reliability model remains the same but the transition rates change since \( \alpha = \gamma = 0 \). Particularly, \( \lambda_1 = (D + C)\lambda_{4} p\alpha \) and \( \lambda_{si} = \lambda_{sp} \). In case of RAID-3,4,5, with cold spares, \( \lambda_{sp} = 0 \); i.e., there are no transitions from states \((G-1, M-i+1)\) to states \((G-1, M-i)\) \((i = 1, ..., M)\). The transition rate from states \((G, M-i+1)\) to \((G, M-i)\) \((i = 1, ..., M)\) becomes \( \lambda_2 = (D + C)\lambda_{4} p\alpha + \gamma \). Data-reconstruction rate while at least one spare is available is \( \mu_3 \) instead of \( \mu_1 \) and \( \mu_2 < \mu_3 < \mu_1 \). In case of RAID-1,2, with cold spares, besides \( \lambda_{sp} = 0 \), we also have \( \alpha = \gamma = 0 \); i.e., there are no transitions from states \((G, M-i+1)\) to \((G, M-i)\) \((i = 1, ..., M)\).

4. NUMERICAL RESULTS

We conducted several experiments with the models described in previous sections. We assume that each level of RAID uses identical disks (same capacity and same mean time to failure) which presumably come from the same manufacturer. The base model we use for RAID-1 has 32 data disks and 32 mirrored disks. For RAID-2, the model has 8 groups each consisting of 4 data disks and 3 check disks. For RAID-3, 4, and 5, the base model has 8 groups of 4 data disks and 1 check disk. The numerical values of some of the model parameters are chosen based upon the data given in [9]. The mean time to failure (MTTF) of an active disk (data and check disks) is 40,000 hr \((\lambda = 1/40,000 \text{ per hr})\). If hot spares are maintained, mean data-reconstruction time is 2 hr \((\mu_1 = 1/2 \text{ per hr})\). If cold spares are maintained, then mean data-reconstruction time is 50 hr. If no disk spares are maintained, then mean data-reconstruction time is 74 hr. Failure rate of a hot spare disk is \( \lambda_{sp} = 1/50,000 \text{ per hr} \). The coverage probability in each case is assumed to be 0.9. In models with predictive disk failures, rate of false alarm is chosen to be \( \gamma = 1/100,000 \text{ per hr} \) and probability that an impending failure is correctly predicted is chosen to be \( \alpha = 0.9 \). All these two-level hierarchical Markov models were solved using the software package SHARPE [10]. In the following experiments, reliability was evaluated at \( t = 1000 \text{ hr} \) unless otherwise mentioned.

4.1. How Reliable Should Each Disk Be?

The reliability gain of various disk array architectures as MTTF of a single disk increases is shown in Fig. 4. The reliability gain is significant as the MTTF of each disk is increased from 10,000 to 40,000 hr. However, the gain in reliability is not much as MTTF of a disk is increased beyond 40,000 hr.

4.2. How Much Does Improved Fault Coverage Help?

In Fig. 5. MTDL of disk array is plotted as a function of coverage probability. Tremendous improvement in MTDL is achieved with improved coverage of disk failures. Whereas it is impossible to get rid of catastrophic failures, it is possible to improve the reliability of support hardware, ECC, and disk controllers by introducing redundancy. It should be noted that coverage probability may well depend upon the RAID architecture. It can be
argued that RAID-1 will have higher coverage probability than RAID-3,4,5, because error detection and correction strategy of RAID-3,4,5, is more prone to uncovered failures. Experimental and analytic evaluation of the coverage parameters needs to be investigated further.

4.3. How Low Should the Data-Reconstruction Time Be?

We now consider the effect of mean time to data reconstruction (MTDR) on MTDL of disk array. The plots in Fig. 6 show that data-reconstruction time does not significantly affect the MTDL. The reason for this is that the MTTF of a disk is much larger than the MTDR (mean time to data reconstruction) of disk. Thus, taking expensive measures to reduce the data-reconstruction time would not yield significant gains as far as MTDL (or reliability) is considered. However, composite performance and reliability analysis may show gains in performance measures as MTDR is reduced.

4.4. How Many Disk Spares Are Needed?

In the previous models, we assumed an unlimited number of hot spares. Here, we analyze the dependence of array reliability on (limited) number of spares and on the kind of spares (hot or cold). For RAID-3,4,5, MTDL of disk array as a function of number of spares (cold and hot) is increased from zero to three is plotted in Fig. 7. The gain in MTDL is not much as the number of spares is increased beyond two per group. Similar trends are observed for RAID-1,2.

4.5. Is RAID Reliability Scalable?

A natural step towards more parallelism in I/O transfers and increased storage capacity demand would be to scale the disk arrays in two dimensions: the number of disks in a group and the number of groups (or both). We wish to find out if the reliability of the disk array scales appropriately. RAID-1 architecture is obviously not scalable. Increasing the number of disks reduces array reliability and MTDL. Moreover, it is not cost effective to duplicate all the disks in the system if we intend to use over 100 small disks. Thus, for RAID-1 architecture, the number of disks should be kept small and the size of each disk should be increased. The important thing to remember is that the MTTF of a large disk is not significantly lower than the MTTF of a small disk. Thus, the above solution yields a more reliable design of RAID-1.

Given the storage capacity for RAID-2,3,4,5, the choice of number of data disks in a group and number of groups is dictated mainly by performance considerations (e.g., the amount of parallelism in I/O transfers) and support hardware available (e.g., the number of I/O channels, array controllers, etc.). We illustrate how these choices affect array reliability.

Given a fixed number of disks in a group \( D = 8 \), the number of groups is varied. Figure 8 illustrates how reliability decreases as \( N \) increases. A similar trend is observed if the number of groups \( (N = 8) \) is fixed while the number of disks in a group is varied. These plots reveal that the reliability of all the RAID architectures falls below acceptable levels as the disk arrays are scaled up in dimensions. A simple solution is to scale the redundancy in hardware as the dimensions of a disk array are scaled. Another solution would be to come up with newer designs of RAID with more fault-tolerance like the one suggested in [3].
FIG. 8. Reliability vs storage capacity ($D = 8$ for RAID-2.3.4.5).

5. CONCLUSIONS

We have carried out a reliability analysis of different fault-tolerant disk array architectures classified as different levels of RAID. The reliability models formally capture the operational dependency of disk array system on array organization. Solution of these models provides useful insight into the dependence of disk array reliability on parameters such as MTTF of disks, mean data-reconstruction time, coverage of faults, and dimensions of disk arrays. Our results show that if the MTTF of a single disk is increased beyond a certain value, the gains in disk array reliability are not significant. Thus, unreliable individual disks are not the solution for ultrareliable disk arrays. Reducing data-reconstruction time does not yield significant array reliability gain either. However, a tremendous improvement in disk array reliability can be obtained with improved coverage of faults (i.e., improved fault-detection and more reliable support hardware).

Schulze et al. [12] showed that support hardware components (such as power supply and cooling equipment) reduce the reliability of a disk array and proposed a scheme to place support hardware components orthogonal to the parity groups. In a related paper [6], we model that effect and show that orthogonal placement of hardware components improves the overall reliability of the disk array. Thus, the key to improving disk array reliability is superior fault coverage and judicious placement of support hardware. Dimensional scaling of disk arrays results in reliability degradation. Therefore, hardware redundancy (both in number of disks and in support hardware) must be scaled as the dimensions of disk arrays are scaled to maintain high reliability.

REFERENCES


MANISH MALHOTRA received the B.Tech. degree in computer science from the Indian Institute of Technology, Delhi, India, in 1989, and the M.S. degree in computer science from Duke University, Durham, North Carolina, in 1990, where he is currently a Ph.D. candidate. He was a summer intern at AT&T Bell Labs, Holmdel, New Jersey, in 1990 and 1991. His research interests include dependability and performance modeling, numerical transient solution of Markov chains, and disk array systems.

KISHOR S. TRIVEDI holds a Ph.D. in computer science from the University of Illinois, Urbana–Champaign. He is the author of a text entitled Probability and Statistics with Reliability, Queuing, and Computer Science Applications, published by Prentice–Hall. His research interests are in reliability and performance evaluation. He has lectured and published extensively on these topics. He is a Fellow of the Institute of Electrical and Electronics Engineers. He is a professor of electrical engineering and computer science at Duke University, Durham, NC. He was an editor of the IEEE Transactions on Computers from 1983 to 1987. He is a codesigner of the HARP, SAVE, SHARPE, and SPNP modeling packages.