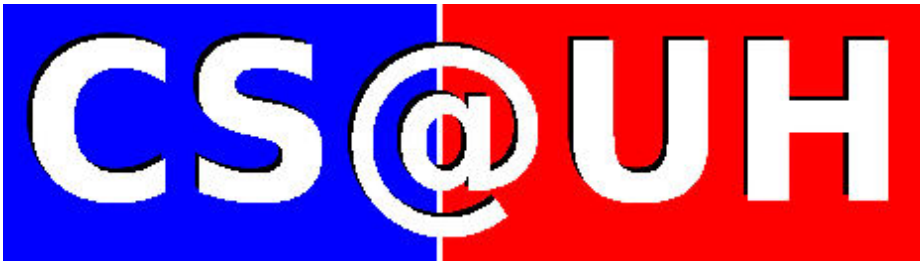

EVALUATING THE RELIABILITY OF STORAGE SYSTEMS

J.-F. Pâris[1], T. J. E. Schwarz[2], and Darrell D. E. Long[3]

Department of Computer Science
University of Houston
Houston, TX, 77204, USA
http://www.cs.uh.edu

Technical Report Number UH-CS-06-08

June 9, 2006

Keywords: fault-tolerant systems, storage systems, repairable systems, *k*-out-of-*n* systems

Abstract

Modern storage systems are often large complex distributed systems. Current techniques for evaluating their reliability function require the solution of a system of differential equations. We present a more elementary, intuitive approach that focuses on the steady-state behavior of each storage organization when it goes through repeated cycles of failures succeeded by repairs. As a result, our approach provides immediately a purely algebraic method for computing both the average failure rate and mean time to failure. We show how to apply our technique to model the high infant mortality of disk drives and the behavior of the so-called S.M.A.R.T. drives, which can warn users of impending disk failures.

[1] J.-F. Pâris is with the Department of Computer Science, University of Houston, Houston, TX 77204, USA.
[2] T. J. Schwarz is with the Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053, USA.
[3] D. D. E. Long is with the Department of Computer Science, University of California, Santa Cruz, CA 95064, USA.

EVALUATING THE RELIABILITY OF STORAGE SYSTEMS

J.-F. Pâris[1], T. J. E. Schwarz[2], and Darrell D. E. Long[3]

Abstract

Modern storage systems are often large complex distributed systems. Current techniques for evaluating their reliability function require the solution of a system of differential equations. We present a more elementary, intuitive approach that focuses on the steady-state behavior of each storage organization when it goes through repeated cycles of failures succeeded by repairs. As a result, our approach provides immediately a purely algebraic method for computing both the average failure rate and mean time to failure. We show how to apply our technique to model the high infant mortality of disk drives and the behavior of the so-called S.M.A.R.T. drives, which can warn users of impending disk failures.

Index Terms

Fault-tolerant systems, storage systems, repairable systems, *k*-out-of-*n* systems.

I. INTRODUCTION

All organizations maintain ever-increasing amounts of data on-line because it is now the most cost-effective way to preserve these data. In part this is due to the lower costs of storage that enable new uses of storage, in part due to regulatory pressure (*e.g*. HIPAA, and Sarbanes-Oxley), and in part simply to the cumulative effect of digital data production at increasing rates. Given the present state of the technology, this means storing these data on disk drives, devices that are known to be subject to unexpected failures well before the end of their useful lifetime.

As petabyte scale file systems become more common, disk failures will occur daily, if not more frequently [XM+03], while data loss at this rate can no longer be tolerated. Backups have been the traditional way of protecting data against equipment failures. Unfortunately, they suffer from several grave deficiencies. First, they do not scale well; indeed the amount of time required to make a copy of a large data set can exceed the interval between back ups. Second, the process is not as trustworthy as it should be due to both human error and the frailty of most recording media. Finally, backup technologies are subject to technical obsolescence, which means that saved data risk becoming unreadable after only a few years. Other traditional techniques such as RAID Level 5 no longer achieve the failure tolerance required of these massive storage systems [XM+03]. A much better solution is to introduce redundancy into our on-line storage systems, through the use of replication including techniques such as *m*-out-of-*n* codes. Assuming that we can achieve any level of data survivability by increasing the level of redundancy of our system, we must still decide what constitutes the appropriate level of redundancy for some specific data. These aspects are particularly important for archival storage systems, because these systems must guarantee the survival of huge amounts of data over very long periods of time. The owners of these systems must select a redundancy level that is sufficient to guarantee the long-term survival of their archived data while avoiding any unnecessary costs due to excessive redundancy. To solve this problem, the owners of any storage system must estimate in advance the survivability levels of the various solutions they require. This estimation becomes more difficult with increasing scale of systems. If systems are large enough, unlikely events become common and failure modes that can be neglected in smaller systems (such as disk infant mortality) suddenly have a measurable impact.

Estimating the survivability of data requires estimating the reliability of the storage system on which they reside, that is the probability $R(t)$ that the system will operate correctly over the time interval [0, t] given that it operated correctly at time $t = 0$. Computing that probability requires solving a system of linear differential equations, a task that becomes quickly unmanageable as the complexity of the system grows. As a result, discrete system simulation

[1] J.-F. Pâris is with the Department of Computer Science, University of Houston, Houston, TX 77204, USA.
[2] T. J. Schwarz is with the Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053, USA.
[3] D. D. E. Long is with the Department of Computer Science, University of California, Santa Cruz, CA 95064, USA.

usually constitutes the preferred tool for estimating the reliability of complex systems. Unfortunately, simulation has its own limitations. First, it only provides numerical values instead of closed form expressions. Second, it requires inordinate amounts of time to evaluate the probabilities of very infrequent events. This second observation is particularly true for archival storage systems: since we want these systems losing very few documents over long periods of time, we have to estimate the likelihood of very unlikely events.

A simpler option is to focus on the mean time to failure (MTTF) of the system that we want to analyze. There are methods that can obtain a symbolic solution for the MTTF directly from the transition rates, but to understand this solution we need to solve a linear differential equation (the Chapman-Kolmogorov equations). The main disadvantage of this approach is that it requires a solid background in Calculus and differential equations to fully understand it.

We propose a simpler, more elementary and more intuitive approach. It consists of focusing on the steady-state behavior of each storage organization over long periods of time as it goes through repeated cycles of failures and repairs. This allows us to describe the storage system directly by a system of linear equations without any recourse to the Chapman-Kolmogorov system of linear differential equations. While it provides the same closed-form expressions as existing techniques, our approach is much easier to learn, as it does not require any advanced mathematical training besides an introduction to Markov processes. Rather, it reduces the MTTF calculation to the manipulation of a simply derived linear system of equations. Even though all our examples come from the storage systems area, our new technique is not specific to that area and would apply equally well to all systems whose behavior can be described by first-order Markov models.

The remainder of this article is organized as follows. Section 2 surveys previous work on estimating the reliability of storage systems. Section 3 introduces our method and Section 4 shows how it can be applied to take into account the high infant mortality of disk drives and the behavior of the so called S.M.A.R.T. drives, which can predict when they are the most likely to fail. Finally, Section 5 has our conclusions.

II. Previous Work

Disk arrays need to combine high reliability with high performance and good storage utilization [GW+94]. Many authors have used Markov models to determine the reliability of disk arrays with redundancy (*e.g.* [BM93, BKJP01,GP93, Is93, LCZ05, Ng 94, RM05, SB95, WLK98, XSM05, ZJ+03, Z02]). Typically, the Markov models used are small and can be solved formally. Both formal and numerical methods for solving these models exist. If $\mathbf{p}(t)$ is the vector formed by the probabilities $p_i(t)$ of the system being in state i at time t, and $\mathbf{M}$ is the *transition* matrix (see below) with coefficient reflecting the then the fundamental *Chapman-Kolmogorov* system of differential equation describes the evolution of $\mathbf{p}$ over time:

$$\frac{d\mathbf{p}}{dt} = \mathbf{M} \cdot \mathbf{p} .$$

and once a researcher has obtained a formal solution for **p,** they can integrate in order to obtain a formal expression for the MTTF. For at least a generation, packages exist that do so automatically. For instance, *Acyclic Markov Chain Evaluator* (ACE) (1986) [MRT86] solves the Kolmogorov system for acyclic homogenous Markov models (*i.e.* exactly those that MTTF of storage systems calculations want to solve) formally as state probabilities

$$p_i(t) = \sum_{j=t}^{n} \left(\exp(\gamma_{ij} t) \sum_{k=1}^{n} a_{i,j,k} t^k \right)$$

where n is the number of states and the various coefficients are either directly transition rates (the γ_{ij}) or are calculated from them.

Most Markov models of interest in storage system reliability are *stiff*, that is, the transition rates vary by several orders of magnitude. An example is device failure rates (measured in months or years) and repair rates (measured in minutes, hours or possibly days). The resulting numerical instability forces a researcher to carefully select the numerical method [MMT94]. An additional problem that sometimes arises is the large state space of Markov models. For example, Markov models derived from Petri Net models. Additionally, the iterative methods that give accurate numerical results might be slow to converge and special methods need to be employed [HMT96]. Fortunately, there are packages such as SHARPE [ST87] that will select an appropriate numerical method and in some cases aggregate states to lower the complexity of the numerical task.

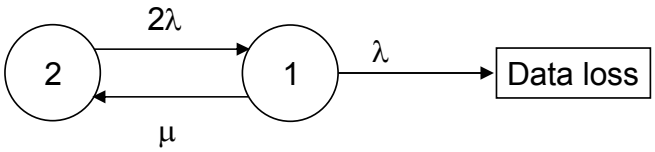

Fig. 1. State-transition diagram for data replicated on two drives.

In many circumstances, it is desirable and possible to derive MTTF directly and exactly without using these sophisticated tools. Examples for a theoretically rigorous yet more elementary and hence more understandable approach are Akhtar's calculation of *k*-out-of-*n*: G systems [Ak94] or Lu's and Liew's analysis of *r*-for-*N* protection systems [LL90].

III. Our Approach

Our system model consists of an array of drives with independent failure modes. When a drive fails, a repair process is immediately initiated for that drive. Should several drives fail, the repair process will be performed in parallel on those drives. We assume that drive failures are independent events and are exponentially distributed with mean λ, and repairs are assumed to be exponentially distributed with mean μ. We will first consider replicated organizations that replicate data on two, three or more drives then examine data organizations using *m*-out-of-*n* codes.

A. Replicated Organizations

The simplest redundant data organization consists of replicating data on two drives (disk mirroring or RAID-1). As Fig. 1 shows, that disk organization can be at any time in one out of three possible states, namely,

a) A state where both drives are operational, that is, state <2>;
b) A state where one disk drive has failed and waits to be repaired or replaced, that is, state <1>;
c) A state where both disk have failed and the data are lost.

Its three state transitions are

a) A transition from state <2> to state <1> that corresponds to the failure of either of the two drives; its rate is twice the failure rate λ of a single drive;
b) A transition from state <1> to state <2> that corresponds to the repair or replacement of the failed drive;
c) A transition from state <1> to the failed state that corresponds to the failure of the last operational drive.

Since the failed state is an absorbing state, the data eventually will be lost. If $p_i(t)$ represents the probability that the system is in state <*i*> at time *t*, the behavior of the system can be represented by the Kolmogorov system of differential equations

$$\frac{dp_2(t)}{dt} = -2\lambda p_2(t) + \mu p_1(t),$$
$$\frac{dp_1(t)}{dt} = 2\lambda p_2(t) - (\lambda + \mu) p_1(t)$$

with the initial conditions $p_2(t=0)=1$ and $p_1(t=0)=0$. The Laplace transforms of this system are

$$sp_2^*(s) - 1 = \mu p_1^*(s)$$
$$sp_1^*(s) = 2\lambda p_2^*(s)$$

where $p_2^*(s)$ and $p_1^*(s)$ respectively are the Laplace transforms of $p_2(t)$ and $p_1(t)$.

Solving the above system for $p_2^*(s)$ and $p_1^*(s)$, we obtain

$$p_2^*(s) = \frac{2\lambda}{s^2 + (3\lambda + \mu)s + 2\lambda^2}, \; p_1^*(s) = \frac{\lambda + \mu + s}{s^2 + (3\lambda + \mu)s + 2\lambda^2}$$

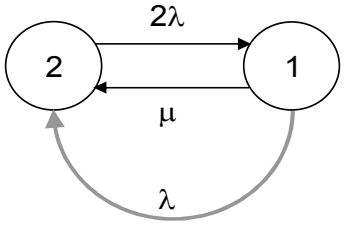

Fig. 2. State-transition diagram for data replicated on two drives assuming that the two drives are instantly repaired after a complete failure.

Since the reliability R(t) of the disk organization is given by

$$R(t) = p_2(t) + p_1(t)$$

the Laplace transform of R(t) is

$$R^*(t) = p_2^*(s) + p_1^*(s) = \frac{3\lambda + \mu + s}{s^2 + (3\lambda + \mu)s + 2\lambda^2}$$

and the MTTF of the system is given by

$$MTTF = R^*(0) = \frac{3\lambda + \mu}{2\lambda^2}.$$

This approach requires computing the Laplace transforms of the Kolmogorov system of differential equations and then the resolution of a system of linear equations. We now show a more elementary approach that obtains the same results by a purely algebraic method without ever using the properties of Laplace transforms.

Consider what happens if our system went through continuous cycles during which it would first operate correctly then lose its data and get instantly repaired and reloaded with new data. Fig. 2 shows the state-transition diagram corresponding to this cyclical behavior. It is identical to the state-transition diagram in Fig. 1 except for the transition from state <1> to the failed state, which we have replaced by a transition from state <1> to state <2>. We can evaluate the steady state by evaluating the flow between <1> and <2>, which gives us $2\lambda p_2 = (\lambda + \mu)p_1$. Together with the condition that the sum of the state probabilities equals one, *i.e.* that $p_1 + p_2 = 1$, we get a simple system of equations that yields:

$$2\lambda p_2 = (\lambda + \mu) p_1$$
$$p_2 + p_1 = 1$$

where p_i represents the steady-state probability of the system being in state <*i*>. The solution of this system is

$$p_2 = \frac{\mu + \lambda}{\mu + 3\lambda}, p_1 = \frac{2\lambda}{\mu + 3\lambda}$$

where $\rho = \lambda/\mu$ is the failure rate to repair rate ratio of the two disk drives.

The rate at which the system will fail is

$$L = \lambda p_1 = \frac{2\lambda^2}{\mu + 3\lambda}$$

and the MTTF of the system is

$$MTTF = \frac{1}{L} = \frac{\mu + 3\lambda}{2\lambda^2},$$

which also happens to be the mean time to data loss.

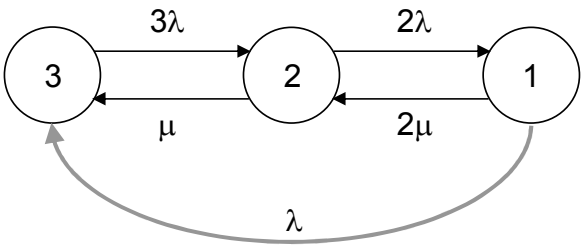

Fig. 3. State-transition diagram for data replicated on three drives assuming that the three drives are instantly repaired after a complete failure.

We can apply the same approach to data organizations that replicate data on three, four or more drives. Fig. 3 represents the state-transition diagram for data replicated on three drives assuming that the three drives are instantly repaired after a total failure. The corresponding system of equations is

$$3\lambda p_3 = \mu p_2 + \lambda p_1$$
$$(2\lambda + \mu) p_2 = 3\lambda p_3 + 2\mu p_1$$
$$(\lambda + 2\mu) p_1 = 2\lambda p_2$$

together with the condition that $p_1 + p_2 + p_3 = 1$, where p_i represents the steady-state probability of the system being in state <*i*>. The solution of this system is

$$p_3 = \frac{2\mu^2 + \lambda\mu + 2\lambda^2}{2\mu^2 + 7\lambda\mu + 11\lambda^2}, p_2 = \frac{6\lambda\mu + 3\lambda^2}{2\mu^2 + 7\lambda\mu + 11\lambda^2}, p_1 = \frac{6\lambda^2}{2\mu^2 + 7\lambda\mu + 11\lambda^2}$$

The rate at which the system will fail is

$$L = \lambda p_1 = \frac{6\lambda^3}{2\mu^2 + 7\lambda\mu + 11\lambda^2}$$

and the MTTF of the system is

$$MTTF = \frac{1}{L} = \frac{2\mu^2 + 7\lambda\mu + 11\lambda^2}{6\lambda^3}$$

The same technique can be applied to compute the MTTF of data organizations that replicate data on more than three drives. For instance, the MTTF of a data organization that replicates data on four drives is

$$MTTF = \frac{3\mu^3 + 13\lambda\mu^2 + 23\lambda^2\mu + 25\lambda^3}{12\lambda^4}$$

Let us now derive a lower bound of that MTTF for an arbitrary number of replicas. Consider Fig. 4. It represents the state transition diagram for a data organization that replicates data on *n* drives. The corresponding steady-state equations are

$$n\lambda p_n = \mu p_{n-1} + \lambda p_1$$
$$((n-1)\lambda + \mu) p_{n-1} = n\lambda p_n + 2\mu p_{n-2}$$
$$\dots$$
$$(2\lambda + (n-2)\mu) p2 = 3\lambda p_{n-3} + (n-1)\mu$$
$$(\lambda + (n-1)\mu) p_1 = 2\lambda p_2$$

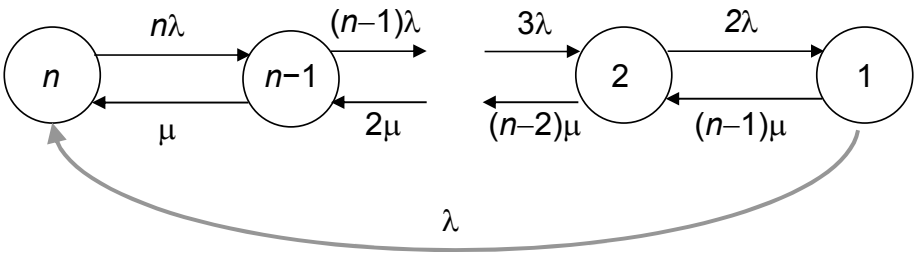

Fig. 4. State-transition diagram for data replicated on n drives assuming that all drives are instantly repaired after a complete failure.

together with the condition $p_1 + p_2 + \ldots p_n = 1$, from which we can derive the following inequalities

$$p_{n-1} = \frac{n\lambda}{\mu} p_n - \frac{\lambda}{\mu} p_1 < \frac{n\lambda}{\mu} p_n$$

$$p_{n-2} = \frac{(n-1)\lambda}{2\mu} p_{n-1} - \frac{\lambda}{2\mu} p_1 < \frac{(n-1)\lambda}{2\mu} p_{n-1}$$

$$\ldots$$

$$p_2 = \frac{3\lambda}{(n-2)\mu} p_3 - \frac{\lambda}{(n-2)\mu} p_1 < \frac{3\lambda}{(n-2)\mu} p_3$$

$$p_1 = \frac{2\lambda}{(n-1)\mu} p_2 - \frac{\lambda}{(n-1)\mu} p_1 < \frac{2\lambda}{(n-1)\mu} p_2$$

and an upper bound for p_1

$$p_1 < \frac{2.3.\cdots.(n-1).n.\lambda^{n-1}}{(n-1).(n-2).\cdots.2.1.\mu^{n-1}} p_n = \frac{n\lambda^{n-1}}{\mu^{n-1}} p_n < \frac{n\lambda^{n-1}}{\mu^{n-1}}$$

since $p_n < 1$. We thus have an upper bound for the rate at which the system will fail

$$L = \lambda p_1 < \frac{n\lambda^n}{\mu^{n-1}}$$

and a lower bound for the MTTF of the system

$$MTTF = \frac{1}{L} > \frac{\mu^{n-1}}{n\lambda^n}.$$

B. Data organizations using m-out-of-n codes

These data organizations store data on n distinct drives along with enough redundant information to allow access to the data in the event $n - m$ of these drives fail. The best-known organizations using these codes are RAID 5, which uses an $n - 1$ out of n code, and RAID 6, which uses an $n - 2$ out of n code.

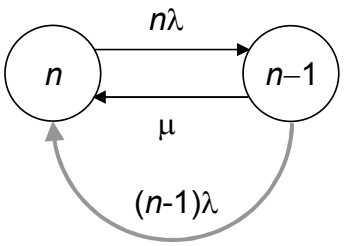

Fig. 5. State-transition diagram for data using an $n-1$ out of n code assuming that all drives are instantly repaired after a complete failure.

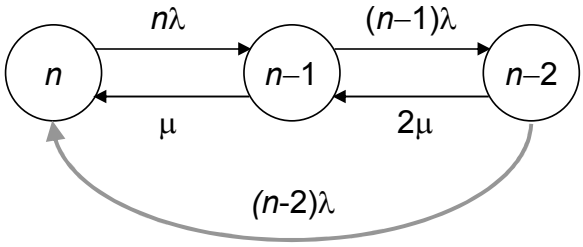

Fig. 6. State-transition diagram for data using an $n-2$ out of n code assuming that all drives are instantly repaired after a complete failure.

Consider first the case of a data organization using an $n-1$ out of n code. As shown on Fig. 5, this data organization will tolerate the failure of one of its n disk drives. Keeping the same notations as in the previous subsection, we can write the steady-state equations of the system as

$$n\lambda p_n = ((n-1)\lambda + \mu)p_{n-1}$$
$$((n-1)\lambda + \mu)p_{n-1} = n\lambda p_n$$
$$p_n + p_{n-1} = 1$$

where p_i represents the steady-state probability of the system being in state <i>. The solution of this system gives us

$$p_n = \frac{1+(n-1)\rho}{1+(2n-1)\rho}, p_{n-1} = \frac{n\rho}{1+(2n-1)\rho}$$

where $\rho = \lambda/\mu$ is the failure rate to repair rate ratio of the two disk drives. The MTTF of the system is

$$MTTF = \frac{1}{(n-1)\lambda p_{n-1}} = \frac{1+(2n-1)\rho}{n(n-1)\lambda\rho} = \frac{\mu+(2n-1)\lambda}{n(n-1)\lambda^2}$$

Let us now turn our attention to $n-2$ out of n codes. As shown on Fig. 5, this data organization will tolerate the failure of two of its n disk drives. The steady-state equations of the system are

$$n\lambda p_n = \mu p_{n-1} + (n-2)\lambda p_{n-2}$$
$$((n-1)\lambda + \mu)p_{n-1} = 3\lambda p_n + 2\mu p_{n-2}$$
$$((n-2)\lambda + 2\mu)p_{n-2} = 2\lambda p_{n-1}$$
$$p_n + p_{n-1} + p_{n-2} = 1$$

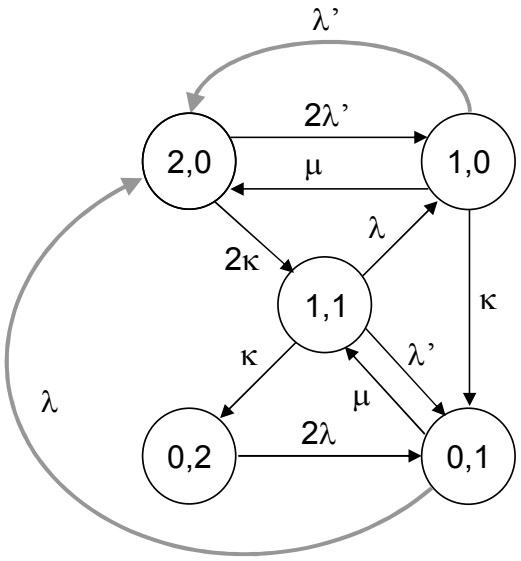

Fig. 7. State-transition diagram for data stored on one pair of drives when considering the higher infant mortality of the two drives.

With the result that

$$p_1 = \frac{n(n-1)\lambda^2}{2\mu^2 + (3n-2)\lambda\mu + (3n^2-6n+2)\lambda^2},$$

the MTTF of the system is

$$MTTF = \frac{1}{(n-2)\lambda p_{n-2}} = \frac{2\mu^2 + (3n-2)\lambda\mu + (3n^2-6n+2)\lambda^2}{n(n-1)(n-2)\lambda^3}.$$

IV. APPLICATIONS

The configurations that we have considered so far are all instances of *k*-out-of-*n* systems. Let us now apply our technique to models that take into account some of the specific characteristics of disk drives, in particular their higher infant mortality and the failure prediction capability of the new S.M.A.R.T. drives.

A. Taking into account infant mortality of disk drives

Disk drives are known to fail more frequently during the first year of deployment [XSM05]. Elerath and IDEMA [E00, I98] proposed a more detailed MTBF rating that incorporates four different values corresponding to drive ages of 0–3 months, 3–6 months, 6–12 months, and one year to *End of Design Life* (EODL). We propose a simpler two-stage model that assumes that recently deployed drives will have a higher failure rate λ' than the failure rate λ of the older drives. While we limit our discussion to the case of data replicated on two drives, nothing should prevent us to apply it to more complex data organizations.

As Fig. 7 shows, the state of the system will be presented by a pair of numbers <*j, k*>, where the first number represents the number of drives that have recently deployed and the second number represents the number of drives that have been deployed for more than a year. The system will start in state <2, 0> as it consists of two new drives. The aging of these drives will be represented by a transition of rate 2κ from state <2, 0> to state <1, 1>, and two transitions of rate κ with the first going from state <1,1> to state <0, 2> and the second going <from state <1, 0> to state <0, 1>. We assume that the repair process will always replace the failed drive with a new drive. Hence, our two repair transitions will be from state <1, 0> to state <2, 0> and from state <0, 1> to state <1, 1>. There is a transition from state <1,1> to state <0,2>, which corresponds to the aging of the replacement drive.

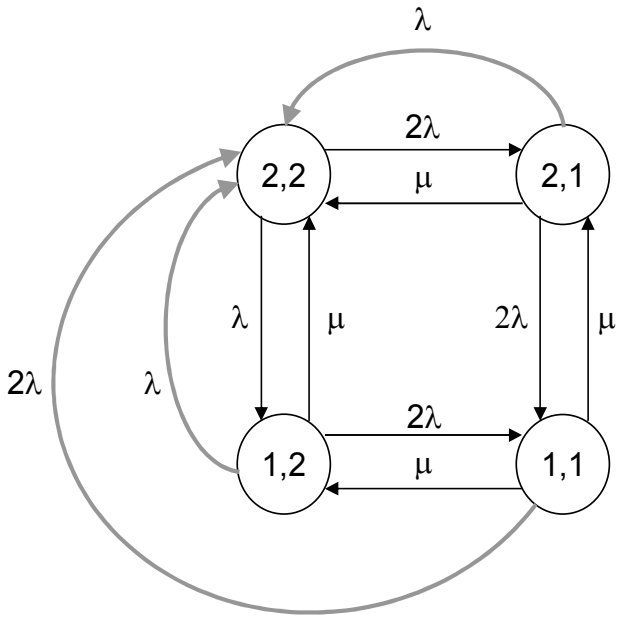

Fig. 8. State-transition diagram for two sets of data stored on two pairs of conventional drives.

Unlike the other systems we have examined, this system has two critical states, namely state <1, 0> and state <0, 1>. A failure from either of these two states will result in permanent data loss. The two corresponding transitions return the system to its original state <2, 0>.

The steady state equations of our system are

$$
\begin{aligned}
(2\lambda'+2\kappa)p_{20} &= (\mu+\lambda')p_{10} + \lambda p_{01} \\
(\lambda+\lambda'+\kappa)p_{11} &= 2\kappa p_{20} + \mu p_{01} \\
2\lambda p_{02} &= \kappa p_{11} \\
(\lambda'+\mu+\kappa)p_{10} &= 2\lambda' p_{20} + \lambda p_{11} \\
(\lambda+\mu)p_{01} &= 2\lambda p_{02} + \lambda' p_{11} + \kappa p_{10} \\
p_{20} + p_{02} + p_{11} + p_{10} + p_{01} &= 1
\end{aligned}
$$

where p_{ij} is the probability of the system being in state $<i, j>$. The solution of this system gives us the steady-state probabilities of the two critical states, namely, <1, 0> and <0, 1>. These are

$$
p_{10} = \frac{(4\lambda'^2+4\lambda'+4\kappa\lambda'+2\kappa)\lambda^2 + 4\lambda'\lambda^3}{D}
$$

$$
p_{01} = \frac{(2\kappa\lambda^2 + (2\kappa\lambda'+2\kappa^2)\lambda)\mu + (6\kappa\lambda'+2\kappa^2)\lambda^2 + (6\kappa\lambda'^2+8\kappa^2\lambda'+2\kappa^3)\lambda}{D}
$$

with

$$
\begin{aligned}
D = \; & (2\lambda^2+3\kappa\lambda+\kappa^2)\mu^3 + (2\lambda^3 + (8\lambda+7\kappa)\lambda^2 + (10\kappa\lambda'+5\kappa^2)\lambda + 3\kappa\lambda'+\kappa^3)\mu + \\
& (6\lambda'+2\kappa)\lambda^3 + (6\lambda'^2+15\kappa\lambda'+5\kappa^2)\lambda^2 + (7\kappa\lambda'^2+10\kappa^2\lambda'+3\kappa^3)\lambda
\end{aligned}
$$

The MTTF of the system is

$$
MTTF = \frac{1}{L} = \frac{1}{\lambda' p_{10} + \lambda p_{01}}
$$

Given the complexity of this expression, it is much easier to reason using a concrete example. Assume that the two drives have a MTTF of 10^5 hours each, that is, slightly more than eleven years, and a mean time to repair of 164 hours or exactly a week. Neglecting the higher infant mortality of disk drives, would give us an MTTF of

2.991×10^7 hours or 3,414 years. Assuming that the failure rate of the disk drives is three times their normal failure rate ($\lambda' = 3\lambda$) during the first year ($\kappa = 1/8760$), we find an MTTF of 2.227×10^7 hours, that is, 25 percent lower than the estimate that neglected infant mortality.

B. S.M.A.R.T. drives

Most major drive manufacturers now support to some extent the *Self-Monitoring, Analysis and Reporting Technology* (S.M.A.R.T.), whose purpose is to warn users of impending disk failures [W06]. The technique can only predict approximately 60 percent of hard drive failures since many failures are sudden and unpredictable. First, consider a system consisting of two pairs of drives with each pair containing two replicas of the stored data. We will consider that the system has failed whenever it has permanently lost some data because both replicas of stored data have been lost. As seen on Fig. 8, the system has four states, namely <2, 2>, <2, 1>, <1, 2> and <1, 1>, with each number indicating the number of operational drives in one of the two pairs. It has failure transitions from states <2, 1>, <1, 2> and <1, 1>, all of which bring the system to the original state <2, 2>. The steady-state equations of the system are

$$\begin{aligned}
4\lambda p_{22} &= (\mu+\lambda)(p_{21}+p_{12}) + 2\lambda p_{11}\\
(3\lambda+\mu)p_{21} &= 2\lambda p_{22} + \mu p_{11}\\
(3\lambda+\mu)p_{12} &= 2\lambda p_{22} + \mu p_{11}\\
(2\lambda+2\mu)p_{11} &= 2\lambda(p_{21}+p_{12})\\
p_{22}+p_{21}+p_{12}+p_{11} &= 1
\end{aligned}$$

The solution of this system gives us the steady-state probabilities of the three critical states, namely, <2, 1>, <1, 2> and <1, 1>. These are

$$p_{21} = p_{12} = \frac{2\lambda(\mu+\lambda)}{\mu^2+6\lambda\mu+11\lambda^2},\ p_{11} = \frac{4\lambda^2}{\mu^2+6\lambda\mu+11\lambda^2}$$

The MTTF of the system is

$$MTTF = \frac{1}{\lambda(p_{21}+\lambda p_{12})+2\lambda p_{11}} = \frac{\mu^2+6\lambda\mu+11\lambda^2}{4\lambda^2(\mu+3\lambda)}$$

Assume now that the four drives are S.M.A.R.T. drives and we get early warning of some disk failures. There is little we could do if a failure is predicted when the system is state <1, 1> as we have no spare capacity. This is not the case when the system is in either state <2, 1> or state <1, 2>. Whenever we get an early warning of the failure of the last operational drive of a pair, we could decide to start transferring the data from that drive to one of the two operational drives in the other pair, thus leaving the system in state <11> with a single copy of all the stored data. Define $\alpha \le 1$ as the probability that we get a warning of the impending failure of a drive and that this warning gives us enough time to transfer the data on the drive to another drive. As seen on Fig. 9, the state-transition diagram of the system remain the same as before but for the probabilities of the failure transitions from states <2, 1> and <1, 2>: a fraction α of the failures that occasioned data loss are now redirected to state <1, 1>. The steady-state equations of the system are

$$\begin{aligned}
4\lambda p_{22} &= (\mu+(1-\alpha)\lambda)(p_{21}+p_{12}) + 2\lambda p_{11}\\
(3\lambda+\mu)p_{21} &= 2\lambda p_{22} + \mu p_{11}\\
(3\lambda+\mu)p_{12} &= 2\lambda p_{22} + \mu p_{11}\\
(2\lambda+2\mu)p_{11} &= (2+\alpha)\lambda(p_{21}+p_{12})\\
p_{22}+p_{21}+p_{12}+p_{11} &= 1
\end{aligned}$$

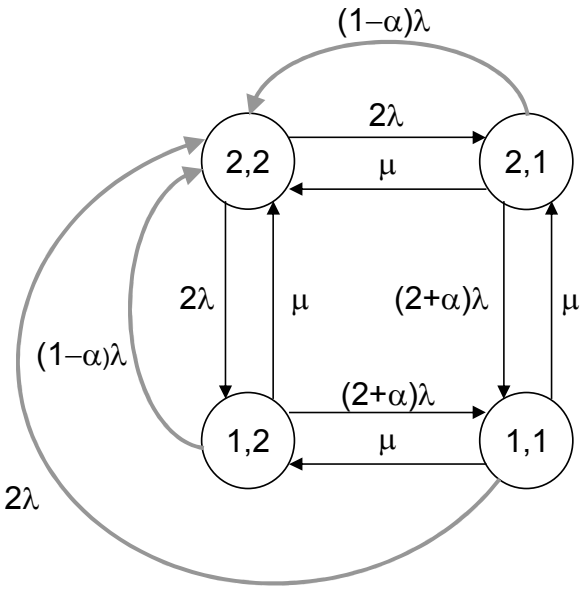

Fig. 9. State-transition diagram for two sets of data stored on two pairs of S.M.A.R.T. drives

and the steady-state probabilities of the three critical states are

$$p_{21} = p_{12} = \frac{2\lambda(\mu+\lambda)}{\mu^2 + (6-\alpha)\lambda\mu + (11-2\alpha)\lambda^2}, p_{11} = \frac{2(2+\alpha)\lambda^2}{\mu^2 + (6-\alpha)\lambda\mu + (11-2\alpha)\lambda^2}$$

The MTTF of the system is

$$MTTF = \frac{1}{(1-\alpha)\lambda(p_{21} + \lambda p_{12}) + 2\lambda p_{12}} = \frac{\mu^2 + (6-\alpha)\lambda\mu + (11-2\alpha)\lambda^2}{4\lambda^2(\mu - \alpha\mu + 3\lambda)}$$

Comparing this result with the MTTF of the same data organization using conventional drives, we find that the MTTF gained by using S.M.A.R.T. drives is

$$\Delta_{MTTF} = \frac{\mu^2 + (6-\alpha)\lambda\mu + (11-2\alpha)\lambda^2}{4\lambda^2(\mu - \alpha\mu + 3\lambda)} - \frac{\mu^2 + 6\lambda\mu + 11\lambda^2}{4\lambda^2(\mu + 3\lambda)} = \frac{\alpha(\mu+\lambda)(\mu^2 + 4\lambda\mu + 6\lambda^2)}{4\lambda^2(\mu+3\lambda)(\mu - \alpha\mu + 3\lambda)}$$

Since $\lambda << \mu$ and $\alpha \le 0.6$, Δ_{MTTF} is roughly equal to $\alpha\mu/(4\lambda^2(1-\alpha))$.

Fig. 10 displays the MTTF of data organizations storing two data sets on two sets of dives for selected values of α and selected Mean Time To Repair (MTTR) of replacing a failed disk drive. We assumed a drive MTTF of 10^5 hours, which corresponds to about one failure every eleven years. As we can see, a system that can predict 50 percent of failures sufficiently ahead of time to be able to save the data from the endangered drive would have an MTTF twice that of a system using conventional drives. This result is quite impressive as it achieved without adding any additional hardware, without assuming that the S.M.A.R.T. drives are more reliable than conventional drives, or assuming that we use a S.M.A.R.T warning to speed up repair.

V. CONCLUSIONS

We have presented a new technique for computing the mean time to failure (MTTF) of repairable systems. Unlike extant approaches, our technique is based on the steady-state analysis of the system when it goes through repeated cycles of failures and repairs. It is completely intuitive and MTTF calculation does not invoke the Chapman-Kolmogorov equations directly.

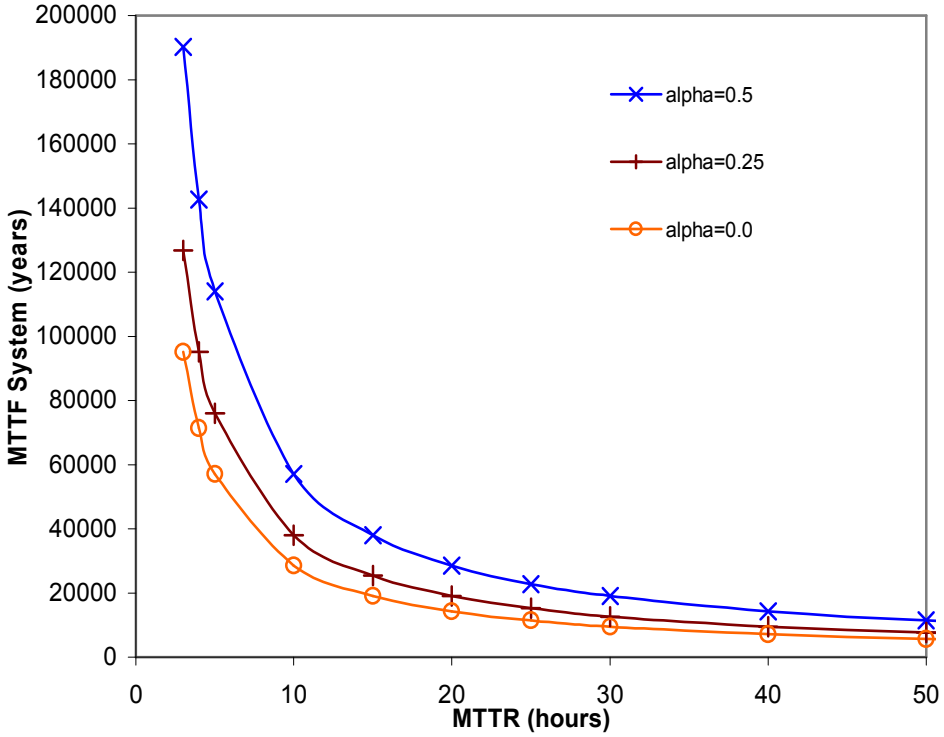

Fig. 10. MTTF for two sets of data stored on two pairs of drives for selected MTTR and α.

We have applied our technique to compute the MTTF's of various standard *k*-out-of-*n* systems and derive a general lower bound for the MTTF of 1-out-of-*n* systems. In addition, we have used our technique to analyze specific aspects of disk drive behaviors, namely, their higher infant mortality and the failure prediction capability of the new S.M.A.R.T. drives. We found out that the higher infant mortality of disk drives had a noticeable impact on the MTTF of replicated disk organizations: assuming that disk drives fail at three times their normal failure rate during their first year of operation reduced by 25 percent our estimate the MTTF of data replicated on a pair of drives. The effect on MTTF of the prediction capabilities of the new S.M.A.R.T. drives was even more impressive: using S.M.A.R.T. drives that can predict at least 50 percent of future failures could double the MTTF of two pairs of mirrored drives. Even though these examples all come from the storage system area, our new technique is not specific to that area and would apply equally well to all systems whose behavior can be described by first-order Markov models.

REFERENCES

[A00] Satoshi Asami, *Reducing the cost of system administration of a disk storage system built from commodity components.* Ph. D. Thesis, UC Berkeley, 2000. (www.eecs.berkeley.edu/Pubs/TechRpts/2000/CSD-00-1100.pdf)

[Ak94] S. Akhtar, "Reliability of *k* out of *n*:*G* systems with imperfect fault coverage." IEEE Transactions on Reliability, vol. 43(1), March 1994.

[AB+02] M. Ajtai, R. Burns, R. Fagin, D. D. E. Long, and L. Stockmeyer, "Compactly encoding unstructured inputs with differential compression," *Journal of the ACM*, Vol. 49, No. 3, pp. 318–367, May 2002.

[BK+01] S. H. Baek, B. W. Kim, E.J. Joung and C W. Park. "Reliability and performance of hierarchical RAID with multiple controllers." *Proc. 20th ACM Symposium on Principles of Distributed Computing*, pp. 246– 256, Aug. 2001.

[BM93] W. Burkhard and J. Menon, "Disk array storage system reliability." *Proc. 23rd International Symposium on Fault-Tolerant Computing* (FTCS-23), 432-441, 1993.

[CL+94] P. M. Chen, E. K. Lee, G. A. Gibson, R. Katz, and D. Patterson, "RAID, High-performance, reliable secondary storage." *ACM Computing Surveys*, Vol. 26, No. 2, pp. 145–185, 1994.

[E00] J. G. Elerath. "Specifying reliability in the disk drive industry: No more MTBF's." *Proc.2000 Annual Reliability and Maintainability Symposium*, pp. 194–199, 2000.

[GP93] G. A. Gibson and D. A. Patterson, "Designing disk arrays for high data reliability." *Journal of Parallel and Distributed Computing*, Vol. 17(1/2) p. 4, 1993

[GT90] R. Geist, K. S. Trivedi, "Reliability Estimation of Fault-Tolerant Systems: Tools and Techniques," *Computer*, Vol. 23, No. 7, pp. 52–61, July 1990.

[GW+94] G. Ganger, B. Worthington, R. Hou, Y. Patt, "Disk arrays: High-performance, high-reliability storage subsystems." *IEEE Computer* vol. 27(3), p. 30–36. 1994.

[HR94] A. Høyland, M. Rausand, System reliability theory: Models and statistical methods. Wiley & Sons, 1994.

[I98] The International Disk Drive Equipment & Materials Association (IDEMA). *R2-98: Specification of hard disk drive reliability*.

[Is93] S. Islam, "Performability analysis of disk arrays." *Proc. 36th Midwest Symposium on Circuits and Systems*, Vol.1, 158–160, 1993.

[LCZ05] Q. Lian, W. Chen, and Z. Zhang, "On the impact of replica placement to the reliability of distributed brick storage systems." *Proc. 25th IEEE International Conference on Distributed Computing Systems* (ICDCS'05), pp. 187–196, June 2005.

[LL90] K. Lu, C. Liew, "Analysis and applications of r-for N protection system. *Global Telecommunications Conference, 1990, and Exhibition.*

[M63] M. A. McGregor, "Approximation Formulas for Reliability with Repair," *IEEE Transactions on Reliability*, Vol. R-12, pp. 64–92, 1963.

[MMT94] M. Malhotra, J. Muppula, K. Trivedi, "Stiffness-tolerant methods for transient analysis of stiff Markov chains." *Microelectronics and Reliability,* Vol. 34(11), pp. 1825–1841, 1994.

[MRT86] R. Marie, A. Reibman, K. Trivedi, "Transient analysis of acyclic Markov chains." *Performance Evaluation,* Vol. 7, pp. 175–194, 1987.

[MT92] M. Malhotra and K. Trivedi, "Reliability modeling of disk array systems." Proc. 6th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation, Edinburgh, 1992.

[Ng 94] S. W. Ng, "Sparing for redundant disk arrays," *Distributed and Parallel Databases*, Volume 2, No 2, pp. 133–149, Apr 1994.

[PGK88] D. A. Patterson, G. A. Gibson, and R. H. Katz. "A case for redundant arrays of inexpensive disks (RAID)," *Proc. SIGMOD 1988 International Conference on Data Management,* pp. 109–116, June 1988.

[RM05] J. Risson, T. Moors, "Recovery of commodity multi-site email clusters." *Proc. IEEE International Conf. on Networks*, 2005.

[RT88] A. Reibman, K. Trivedi, "Numerical transient analysis of Markov models." *Computers and Operations Research*, Vol. 15, No. 1, pp. 19–36, 1988.

[SB92] T. J. E. Schwarz and W. A. Burkhard. "RAID Organization and Performance," *Proc. 12th International Conference on Distributed Computing Systems*, pp. 318–325, June 1992.

[SB95] T. J. E. Schwarz, and W. A. Burkhard, "Reliability and performance of RAIDs," *IEEE Transactions on Magnetics*, 1995, Vol 31(2), pp. 1161–1166. March 1995.

[SG+89] M. Schulze, G. Gibson, R. Katz and D. Patterson. "How reliable is a RAID?" *Proc. Spring COMPCON 89 Conference*, pp. 118–123, March 1989.

[ST83] A. Sahner, K. Trivedi, "Design of the hybrid automated reliability predictor." *Proc. 5th Digital Avionics Systems Conference*, November 1983.

[W06] *Self-Monitoring, Analysis and Reporting Technology – Wikipedia, the free encyclopedia*, http://en.wikipedia.org/wiki/Self-Monitoring,_Analysis_and_Reporting_Technology, accessed in April 2006.

[WLK98] X. Wu, J. Li and H. Kameda, "Reliability analysis of disk array organizations by considering uncorrectable bit errors," *IEICE Transactions on Information and Systems, E Series D*, 1998

[XM+03] Q. Xin, E. L. Miller, T. J. E. Schwarz, D. D. E. Long, S. A.Brandt, and W. Litwin, "Reliability mechanisms for very large storage systems," *Proc. 20th IEEE Conference on Mass Storage Systems and Technologies*, pages 146–156, Apr. 2003.

[XSM05] Q. Xin, T. J. E. Schwarz and E. L. Miller, "Disk infant mortality in large storage systems," Proc. 13th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS '05), August 2005.

[Z02] Y. Zhu, Design, implementation and performance evaluation of a cost-effective fault-tolerant parallel virtual file system (CEFT-PVFS), Thesis, University of Nebraska, 2002.

[ZJ+03] Y. Zhu, H. Jiang, X. Qin, D. Feng, and D. Swanson, "Design, implementation, and performance evaluation of a cost-effective fault-tolerant parallel virtual file system." *Proc. International Workshop on Storage Network Architecture and Parallel I/O* (SNAPI '03), 2003.