RZ 3817 (# Z1203-001) 03/29/2012
Computer Science 12 pages

Research Report

A General Reliability Model for Data Storage Systems

V. Venkatesan, I. Iliadis

IBM Research – Zurich
8803 Rüschlikon
Switzerland

Research
Almaden • Austin • Brazil • Cambridge • China • Haifa • India • Tokyo • Watson • Zurich

A General Reliability Model for Data Storage Systems

Vinodh Venkatesan, Ilias Iliadis
IBM Research - Zurich
{ven, ili}@zurich.ibm.com

***Abstract*—Typical models for analysis of storage system reliability assume independent and exponentially distributed times to failure. Also the rebuild time periods are often assumed to be deterministic or to follow an exponential distribution, or alternatively a Weibull distribution. As a first step towards a generalization of these models, we consider more general non-exponential distributions for failure and rebuild times while still retaining the independence assumption. It is shown that the mean time to data loss (MTTDL) of storage systems is practically insensitive to the actual failure distribution when the storage nodes are generally reliable, that is, when their mean time to failure is much larger than their mean time to repair. This implies that MTTDL results previously obtained in the literature by assuming exponential node failure distributions may still be valid despite this unrealistic assumption. In contrast, it is shown that the MTTDL depends on the characteristics of the rebuild distribution.**

***Index Terms*—storage system, reliability, non-exponential, failure distribution, insensitivity**

I. Introduction

In today's information age, large-scale data storage systems store a significant proportion of all data in the world. Data centers are becoming larger not just because of the increase in the amount of new data generated each year but also because of agglomeration of many small data centers into massive data centers for reasons of cost-cutting, energy efficiency, and ease of management.

Although the individual storage nodes that make up the data centers are generally reliable, as the number of nodes in a storage system increases, there is an inevitable increase in the frequency of node failures in the system. This creates a need to model such node failures and to design schemes of data redundancy, placement, and repair, to reduce the chance of irrecoverable loss of data due to a catastrophic sequence of node failures in the system.

Models of data storage systems typically assume that the times to failure of individual storage nodes are independent and identically distributed. Furthermore, it is typically assumed that the distribution of the time to failure of a node is exponential as this allows the use of Markov models [1]. However, all these modeling assumptions have been contested by recent empirical studies of real world storage systems. It has been found that real world failures are neither independent nor exponentially distributed [2]. As a first step toward developing more realistic models, we extend the reliability modeling and analysis of storage systems by including more general non-exponential distributions. This is a non-trivial problem because non-exponential distributions preclude the use of continuous-time Markov chain based methods which are typically used for estimating the system reliability.

Using results from renewal theory and approximations such as the ones used in [3], we show in this paper that the mean time to data loss (MTTDL) of a system are essentially insensitive to the actual distribution of the times to failure of storage nodes as long as the storage nodes are *generally reliable*. By generally reliable nodes, we refer to nodes with mean time to failure being much larger than than their mean time to repair. The implication of this result is significant because it suggests that a large class of distributions for the times to failure of storage nodes essentially yields the same estimate of the MTTDL of the system. In this sense, this is an insensitivity result. Moreover, this result implies that the system MTTDL will not be affected if the failure distribution were changed to a corresponding exponential one with the same mean. This observation is also of great importance because it suggests that MTTDL results obtained previously in the literature by assuming exponential node failure distributions may still be valid despite this unrealistic assumption.

In addition, we also study the effect of the distribution of rebuild times on the reliability of storage nodes. It is shown that the rebuild distribution does affect the reliability. This is because of an effect also known as the *waiting time paradox*. In essence, failures occurring during rebuild tend to occur within rebuild intervals with a larger rather than a smaller duration. This means that the failures tend to occur during critical times when the rebuilds are slower than average and so the amount of data not rebuilt is higher. Thus, following a failure during rebuild, the amount of most-exposed data, that is, the data with the least number of replicas, is higher. This in turn implies that the duration required to rebuild the most-exposed data is also higher. This cascading effect is shown to have a significant effect on the reliability, especially if the rebuild distribution exhibits large variability.

II. Related Work

Several of the earliest works on the analysis of reliability of storage systems [4] have assumed independent and exponentially distributed times to failure. A vast majority of the publications have also assumed exponentially distributed times to rebuild as this allows the use of continuous-time Markov chain models to estimate the reliability of the system [5], [6], [7]. A few works have used other probabilistic methods [8], [9],

TABLE I
PARAMETERS OF A STORAGE SYSTEM

c	storage capacity of each node (bytes)
n	number of storage nodes
$1/\lambda$	mean time to failure of a storage node (s)
$1/\mu$	mean time to rebuild a storage node (s)
$c\mu$	average rebuild bandwidth at each storage node (bytes/s)

however, the probability of data loss in these works is obtained for the case when there are no rebuild operations performed. Publications based on real world failure data have shown that the distribution of failures is neither exponential nor independent [2]. Failure distributions other than exponential have been studied extensively through simulations [10], [11], [12]. In particular, it has been shown that the expected number of double disk failures in RAID-5 systems within a given time period can vary depending on the failure distribution [12]. In contrast, we consider the expected time to the first data loss event (which is equivalent to a double disk failure in case of a RAID-5 system) and we show that it is insensitive to the node failure distribution.

III. SYSTEM MODEL

The parameters of the storage system considered and the failure and rebuild distributions considered in the paper are described in this section. Table I lists the parameters used.

A. Storage System

Consider a data storage system with n nodes each with a storage capacity c. Some form of redundancy such as replication, RAID-5, RAID-6, or other erasure code, is assumed to be used to protect data from node failures in the system. Whenever a node failure occurs, a rebuild process is initiated to restore the redundancy lost due to that node failure. As more nodes fail during the rebuild process, some data tend to lose more of their redundancy until the point where irrecoverable data loss occurs. In other words, a data loss is said to have occurred when the rebuild process can no longer restore some of the data that was initially stored in the system due to a catastrophic sequence of node failures. Broadly, at any time, the system can be thought to be in one of two modes: fully-operational mode and rebuild mode. During the fully-operational mode, all data in the system has the original amount of redundancy and there is no active rebuild process. During the rebuild mode, some data in the system has less than the original amount of redundancy and there is an active rebuild process that is trying to restore the lost redundancy. A transition from fully-operational mode to rebuild mode occurs when a node fails; we refer to this node failure that causes this transition as a *first-node* failure. Following a first-node failure, a complex sequence of rebuilds and subsequent node failures may occur which may eventually lead the system either back to the original fully-operational mode or to irrecoverable data loss.

B. Failure and Rebuild Distributions

It is known that real world storage nodes are generally reliable, that is, the mean time to repair a node (which is of the order of tens of hours) is much smaller than the mean time to failure of a node (which is at least of the order of thousands of hours). Let us denote the mean time to rebuild a node by $1/\mu$ and the mean time to failure of node by $1/\lambda$. It now follows that generally reliable nodes satisfies the following condition:

$$1/\mu \ll 1/\lambda, \quad \text{or} \quad \lambda/\mu \ll 1. \tag{1}$$

In the subsequent analysis, this condition implies that terms involving powers of λ/μ greater than one are negligible compared to λ/μ and can be ignored.

Let the cumulative distribution functions of time to failure and rebuild time of each node be F_λ with mean $1/\lambda$ and G_μ with mean $1/\mu$, respectively, satisfying the following condition:

$$\mu \int_0^\infty F_\lambda(t)(1 - G_\mu(t))dt \ll 1, \quad \text{with } \frac{\lambda}{\mu} \ll 1. \tag{2}$$

The results of this paper are derived for the class of failure and rebuild distributions that satisfy the above condition. In particular, the MTTDL is shown to be insensitive to the failure distributions within this class. This result is of great importance because it turns out that this condition holds for a wide variety of failure and rebuild distributions including, most importantly, distributions that are seen in real world storage systems. As an illustration, let us consider the class of failure distributions that satisfy the above conditions, when the rebuild times are deterministic, that is,

$$G_\mu(t) = \begin{cases} 0, & \text{when } t < 1/\mu, \\ 1, & \text{when } t \geq 1/\mu. \end{cases} \tag{3}$$

Recognizing that F_λ is a monotonically non-decreasing function such that $F_\lambda(t) \leq F_\lambda(1/\mu)$ for $t \leq 1/\mu$, the left hand side of (2) reduces to

$$\mu \int_0^{1/\mu} F_\lambda(t)dt \leq F_\lambda(1/\mu). \tag{4}$$

When $\lambda/\mu \ll 1$, it can be seen that $F_\lambda(1/\mu) \ll 1$ for a wide variety of distributions including exponential, Weibull (with shape parameter greater than 1), and gamma (with shape parameter greater than 1). For instance, consider a Weibull distribution with shape parameter k and scale parameter β having the cumulative distribution function

$$F_\lambda^{\text{Weibull}}(t) = 1 - e^{-(t/\beta)^k}. \tag{5}$$

The mean of the Weibull distribution, $1/\lambda$, is equal to $\beta\Gamma(1 + 1/k)$, where $\Gamma(\cdot)$ denotes the gamma function. Therefore, the scale parameter β can be written in terms of the mean $1/\lambda$ as

$$\beta = 1/(\lambda\Gamma(1 + 1/k)). \tag{6}$$

Substituting (6) in (5) for $t = 1/\mu$ we get

$$F_\lambda^{\text{Weibull}}(1/\mu) = 1 - e^{-(\lambda\Gamma(1+1/k)/\mu)^k} \ll 1, \tag{7}$$

when $\lambda/\mu \ll 1$ and $k \geq 1$. However, if $k < 1$, the above inequality may not hold. Note that nodes that have Weibull lifetime distributions with $k < 1$ have high infant mortality

rate, whereas those with Weibull lifetime distributions with $k > 1$ gracefully age over time.

In general, it can be observed that failure distributions with high infant mortality rates do not satisfy condition (2). However, it has been observed that infant mortality is not present in real world storage nodes [2]. Furthermore, the effects of infant mortality can be eliminated from the system by stressing new nodes before adding them to the system. It can also be observed that (2) is satisfied by a wide variety of distributions for rebuild times, in particular, distributions with bounded support. Therefore, condition (2) is realistic as it is satisfied by practical storage systems.

Condition (2) can also be stated in the following alternate way: if F_λ and G_μ belong to a *family* of distributions characterized by λ and μ, respectively, then (2) is equivalent to

$$\lim_{\lambda/\mu \to 0} \mu \int_0^\infty F_\lambda(t)(1 - G_\mu(t))dt = 0. \tag{8}$$

For a fixed μ, this implies that

$$\lim_{1/\lambda \to \infty} \mu \int_0^\infty F_\lambda(t)(1 - G_\mu(t))dt = 0. \tag{9}$$

As $F_\lambda(t)(1-G_\mu(t)) \leq 1-G_\mu(t)$ and $1-G_\mu(t)$ is integrable, by the dominated convergence theorem the order of limit and integral can be exchanged. Therefore,

$$\lim_{1/\lambda \to \infty} \mu \int_0^\infty F_\lambda(t)(1 - G_\mu(t))dt = \mu \int_0^\infty \lim_{1/\lambda \to \infty} F_\lambda(t)(1 - G_\mu(t))dt. \tag{10}$$

Therefore, for a fixed μ, (9) holds only when

$$\lim_{1/\lambda \to \infty} F_\lambda(t) = 0, \, \forall \, t \text{ where } G_\mu(t) < 1, \tag{11}$$

and the convergence of F_λ is pointwise. Similarly, it can be shown that, for a fixed λ, (8) holds only when

$$\lim_{1/\mu \to 0} \mu(1 - G_\mu(t)) = 0, \, \forall \, t \text{ where } F_\lambda(t) > 0, \tag{12}$$

and the convergence is pointwise. Note that (11) and (12) can be equivalently written as

$$F_\lambda(t) \ll 1 \text{ when } G_\mu(t) < 1 \text{ and } \lambda \ll \mu, \tag{13}$$

$$\mu(1 - G_\mu(t)) \ll 1 \text{ when } F_\lambda(t) > 0 \text{ and } \mu \gg \lambda. \tag{14}$$

The next section provides known results from renewal theory that will be used later in our analysis.

IV. Preliminaries

A. Node Availability

A node i operates for a certain period of time with distribution F_λ before failing. Following the failure of a node, the node and all of its data is restored after a period of time with distribution G_μ. Therefore, the timeline of the node consists of successive periods of operation and repair. For $t \geq 0$, let us define

$$\nu_t^{(i)} := \begin{cases} 1, & \text{if node is operational at time } t, \\ 0, & \text{if node is under repair at time } t. \end{cases} \tag{15}$$

Then the node availability at time t is given by the probability

$$a_t^{(i)} := \Pr\{\nu_t^{(i)} = 1\}. \tag{16}$$

The following result is well known in renewal theory [13, Chap. 2, pp. 109–114]:

Lemma 1: The steady-state node availability a is given by

$$a := \lim_{t \to \infty} a_t^{(i)} = \frac{1/\lambda}{1/\lambda + 1/\mu}. \tag{17}$$

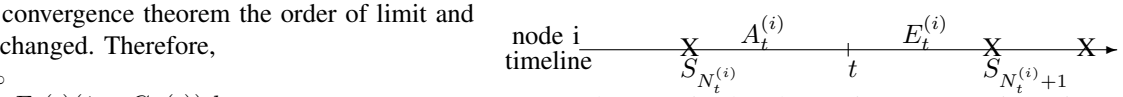
Note that the above result indicated that the steady-state node availability only depends on the means of the distributions F_λ and G_μ.

B. Age and Excess

Consider the timeline constructed by concatenating only the periods of operation of the node. In this timeline, let $N_t^{(i)}$ be the number of replacements of the node up to time t, and S_k be the time of the kth failure, for $k = 1, 2, \cdots$. Define the *age* $A_t^{(i)}$ and the *excess* $E_t^{(i)}$ of the node as

$$A_t^{(i)} := t - S_{N_t^{(i)}}, \tag{18}$$

$$E_t^{(i)} := S_{N_t^{(i)}+1} - t. \tag{19}$$



As can be seen in the above picture, at a given time t, the age $A_t^{(i)}$ is equal to the time that has passed since the last replacement of the node, and the excess $E_t^{(i)}$ is equal to the time until the next failure of the node. A well known result in renewal theory is the following [13, Chap. 2, pp. 109–114]:

Lemma 2:

$$\lim_{t \to \infty} \Pr\{A_t^{(i)} \leq \tau\} = \lim_{t \to \infty} \Pr\{E_t^{(i)} \leq \tau\} = \widetilde{F}_\lambda(\tau), \tag{20}$$

where

$$\widetilde{F}_\lambda(\tau) := \lambda \int_0^\tau (1 - F_\lambda(x))dx. \tag{21}$$

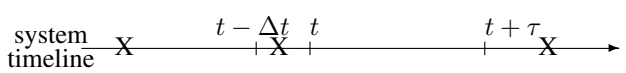
In other words, the cumulative distribution functions of $A_t^{(i)}$ and $E_t^{(i)}$ tend to $\widetilde{F}_\lambda$ as t tends to infinity. In fact, it can be shown that, if the density function corresponding to F_λ approaches zero exponentially fast, then the distributions of $A_t^{(i)}$ and $E_t^{(i)}$ also approach $\widetilde{F}_\lambda$ exponentially fast [13].

V. System Reliability

We model the system's behavior as a series of first-node-failure events each of which is followed by a potentially complex sequence of failure and rebuild events that either lead to data loss with probability P_{DL}, or back to the original fully-operational state of the system by restoring all replicas. As the rebuild times are much shorter than the times to failure, the time taken for these complex sequence of events is negligible compared to the time to first-node-failure, and therefore can be ignored.

Let $E(t, \Delta t)$ represent the event that the system was renewed to its original state in the interval $(t - \Delta t, t)$. Also, let

$E(t, \Delta t, \tau)$ represent the event that the system was renewed to its fully-operational state in the interval $(t-\Delta t, t)$ and continues to operate without any failures in the interval $(t, t+\tau)$.



We are interested in the reliability (or survival) function of the system, that is, the probability $p_t(\tau)$ defined as:

$$\begin{aligned} p_t(\tau) &:= \lim_{\Delta t \to 0} \Pr\{E(t, \Delta t, \tau) | E(t, \Delta t)\} \\ &= \lim_{\Delta t \to 0} \frac{\Pr\{E(t, \Delta t, \tau)\}}{\Pr\{E(t, \Delta t)\}}. \end{aligned} \tag{22}$$

Using this probability, the mean fully-operational period of the system, T_t, can be computed as

$$T_t = \int_0^\infty p_t(\tau) d\tau. \tag{23}$$

In other words, T_t is the mean time period between the first-node failure at time t and the subsequent first-node-failure event. As the system becomes stationary, $p_t(\tau)$ converges to $p(\tau)$ and T_t converges to T. By computing $p(\tau)$, it is shown in Appendix A that

$$T := \lim_{t \to \infty} T_t = \frac{1}{n\lambda}. \tag{24}$$

As each first-node-failure can result in data loss with probability P_{DL}, the expected number of first-node-failures until data loss occurs is $1/P_{DL}$. By neglecting the effect of the relatively short transient period of the system, the MTTDL is essentially the product of the expected time between two first-node-failure events, T, and the expected number of first-node-failure events, $1/P_{DL}$, that is,

$$\text{MTTDL} \approx \frac{T}{P_{DL}} = \frac{1}{n\lambda P_{DL}}. \tag{25}$$

VI. Replication Based Systems

As an application, we will showcase in detail the analysis of the reliability of storage systems that use replication to protect data from node failures. Such systems with different replica placement schemes have been analyzed with exponentially distributed times to node failures and deterministic rebuild times [3]. Here we extend that analysis to more general distributions of times to failure and rebuild completion.

A. System Model

Consider a block-based storage system comprising n storage nodes each with capacity c. Every user data block is replicated r times to protect the data from node failures. These r copies (replicas) are stored in the system such that no two replicas of a data block are in the same node. The way in which the r replicas of each data block are stored across the n nodes is determined by the placement scheme used. We will analyze the two schemes previously studied in [3], namely, declustered and clustered placement schemes, and we will compare the analyses and the respective results.

B. Replica Placement Schemes

We will now briefly describe the two placements schemes, namely, declustered and clustered.

Declustered Placement: In this placement scheme, all $\binom{n}{r}$ possible ways of placing r replicas across n nodes are equally used to store the data in the system. This way, the $r-1$ replicas of the data on each node are equally spread across the remaining $n-1$ nodes. It can be seen that, in this placement, any set of two nodes share replicas of exactly $\frac{r-1}{n-1}c$ amount of data. In general, any set of k nodes ($k \leq r$) share copies of exactly $c\prod_{i=1}^{k-1}\left(\frac{r-i}{n-i}\right)$ amount of data.

Clustered Placement: In this placement scheme, the n nodes are divided into disjoint sets of r nodes and all nodes in each set are mirrors of each other, that is, they store replicas of the same data.

One major way in which these two placement schemes affect the reliability of the system is through the rebuild process. In declustered placement, when a node fails, as the replicas of the data lost are spread across all $n-1$ nodes, a parallel rebuild process can be used to take advantage of the rebuild bandwidth available at all $n-1$ nodes. On the other hand, in clustered placement, when a node fails, the replicas of the data lost are available only on the remaining $r-1$ node of the cluster and therefore, the rebuild bandwidth available does not scale with n, the number of nodes in the system.

C. Rebuild Model

When nodes fail, data blocks lose one or more of their r replicas. The purpose of the rebuild process is to recover all replicas lost so that all data have r replicas. A good rebuild process needs to be both *intelligent* and *distributed*.

By an intelligent rebuild process, we mean that the system always attempts to first recover the copies (replicas) of the blocks that have the least number of replicas left. In contrast to the intelligent rebuild, one may consider an *unintelligent* rebuild, where lost replicas are being recovered in an order that is not specifically aimed at recovering the data blocks with the least number of replicas first. Clearly, an unintelligent rebuild is more vulnerable to data loss, but may have a lower implementation complexity than an intelligent rebuild. In the remainder of the paper we consider only intelligent rebuild.

In placement schemes such as the declustered scheme, the surviving replicas that the system needs to read to recover the lost replicas may be spread across several, or even all, surviving nodes. Broadly speaking, two approaches can be taken when recovering the lost replicas: the data blocks to be rebuilt can be read from all the nodes in which they are present, and either (i) copied directly to a new node, or (ii) copied to (reserved) spare space in all surviving nodes first and then to a new node. The latter method is referred to as distributed rebuild and has a clear advantage in terms of time to rebuild because it exploits parallelism when writing to many (surviving) nodes versus writing to only one (new) node.

During the rebuild process, an average read-write bandwidth of $c\mu$ bytes/s is assumed to be reserved at each node exclusively for the rebuild. This is usually only a fraction of the total

bandwidth available at each node; the remainder is being used to serve user requests. In clustered placement, it is assumed that there are spare nodes, and when a node fails, data is read from any *one* of the surviving nodes of the cluster to which the failed node belonged and written to a spare node at an average rate $c\mu$. Let $G_{\mu_\alpha^{\text{clus.}}}$ denote the cumulative distribution function of the time taken to rebuild a fraction α of the node which has a mean $1/\mu_\alpha^{\text{clus.}} = \alpha/\mu$. In declustered placement, it is assumed that sufficient spare space is reserved in each node for rebuild. During rebuild, the data to be rebuilt is read from *all* surviving nodes and copied to the spare space reserved in these nodes in such a way that no data block is copied to the spare space of a node in which a copy is already present. As data is being read from and written to each surviving node, the total average read-write rebuild bandwidth $c\mu$ of each node is equally split between the reads and the writes. So if there are $\tilde{n}$ surviving nodes, the total average speed of rebuild in the system is $(\tilde{n}c\mu)/2$. Therefore, the cumulative distribution function of the time taken to rebuild a fraction α of the node is $G_{\mu_\alpha^{\text{declus.}}}$ with $1/\mu_\alpha^{\text{declus.}} = \alpha/(\tilde{n}\mu/2)$. We assume that the distributions $G_{\mu_\alpha^{\text{clus.}}}$ and $G_{\mu_\alpha^{\text{declus.}}}$ satisfy (2). In addition, sufficient network bandwidth is assumed to be available to exploit parallelism when rebuilding from all nodes of the system.

D. Estimation of P_{DL}

We estimate P_{DL} by modeling the system using *exposure levels* [3].

1) Exposure Levels: At time t, let $D_l(t)$ be the number of distinct data blocks that have lost l replicas, with $0 \le l \le r$. The system is said to be in exposure level e at time t, $0 \le e \le r$, if

$$e = \max_{D_l(t)>0} l. \tag{26}$$

In other words, the system is in exposure level e if there exists at least one block with $r-e$ copies and no blocks with fewer than $r-e$ copies in the system, that is, $D_e(t) > 0$, and $D_l(t) = 0$ for all $l > e$. At $t = 0$, $D_l(0) = 0$ for all $l > 0$ and $D_0(0)$ is the total number of distinct data blocks stored in the system. Node failures and rebuild processes cause the values of $D_1(t), \cdots, D_r(t)$ to change over time, and when data loss occurs, $D_r(t) > 0$.

2) Direct Path to Data Loss: Consider the direct path of successive transitions from exposure level 1 to r. In [3] it was shown that P_{DL} can be approximated by the probability of the direct path to data loss, $P_{DL,\text{direct}}$, when nodes are generally reliable, that is,

$$P_{DL} \approx P_{DL,\text{direct}} = \prod_{e=1}^{r-1} P_{e\to e+1}, \tag{27}$$

where $P_{e\to e+1}$ denotes the probability of transition from exposure level e to $e+1$.

3) Rebuild Times at Each Exposure Level: Consider the direct path to data loss and let the rebuild times of the most-exposed data at each exposure level in this path be denoted by R_e, $e = 1, \cdots, r-1$ with means $1/\mu_e$, $e = 1, \cdots, r-1$. Next, we will derive the conditional distributions of these rebuild times given that the system goes through this direct path to data loss, and then we will compute probabilities $P_{e\to e+1}$. Let α_e be the fraction of the rebuild time R_e still left when a node failure occurs causing an exposure level transition. In Appendix B, it is shown that α_e is uniformly distributed, that is,

$$\alpha_e \sim U(0,1), \quad e = 1, \cdots, r-2. \tag{28}$$

a) Clustered Placement: Following the first-node-failure event, the system enters exposure level 1. As the amount of data to be rebuilt at this exposure level is equal to the capacity of the failed node, c, it holds that

$$\frac{1}{\mu_1} = E[R_1] = \frac{1}{\mu}, \quad R_1 \sim G_{\mu_1} = G_\mu. \tag{29}$$

Given α_1 and R_1, the remaining time to complete rebuild is $\alpha_1 R_1$. As the the average rebuild bandwidth is $c\mu$, it now follows that the expected amount of the most-exposed data not rebuilt when the exposure level transition occurred is $\alpha_1 R_1 c\mu$. At this instant, this data has lost 2 copies and is thus the most-exposed data in exposure level 2. As the average rebuild bandwidth remains unaffected, the expected rebuild time $1/\mu_2$ in the second exposure level is

$$\frac{1}{\mu_2} = E[R_2|R_1,\alpha_1] = \frac{\alpha_1 R_1 c\mu}{c\mu} = \alpha_1 R_1. \tag{30}$$

However the distribution of R_2 given R_1 and α_1 could be modeled in several ways. We propose two models, namely,

$$R_2|R_1,\alpha_1 \sim G_{\mu_2} \quad \text{(model A)} \tag{31}$$

$$R_2|R_1,\alpha_1 = 1/\mu_2 \quad \text{(model B)} \tag{32}$$

In model A, we assume that, following a node failure, the system has to reconfigure its rebuild process entirely to rebuild the most-exposed data blocks in the new exposure level. This model may be applicable for instance in the case of a clustered placement scheme, where the node from which data was being rebuilt failed and hence the system has to rebuild from another node in the cluster (if there is one). In this case, the rebuild time R_2 would be different from $\alpha_1 R_1$. In model B, we assume that, following a node failure, the system has to do little or no reconfiguration of the rebuild process to rebuild the most-exposed data in the new exposure level. This is the case where the newly failed node is different from the node from which data is being rebuilt. Similarly, for higher exposure levels, we have

$$\frac{1}{\mu_e} = E[R_e|R_{e-1},\alpha_{e-1}] = \alpha_{e-1}R_{e-1} \tag{33}$$

$$R_e|R_{e-1},\alpha_{e-1} \sim G_{\mu_e} \quad \text{(model A)} \tag{34}$$

$$R_e|R_{e-1},\alpha_{e-1} = 1/\mu_e \quad \text{(model B)} \tag{35}$$

b) Declustered Placement: In exposure level 1, the amount of data to be rebuilt is c, just like in clustered placement. However, the average rebuild bandwidth is $(n-1)c\mu/2$ because of the parallel rebuild process. Therefore

$$\frac{1}{\mu_1} = E[R_1] = \frac{c}{(n-1)c\mu/2} = \frac{1}{(n-1)\mu/2},\ R_1 \sim G_{\mu_1}. \quad (36)$$

Note however that, although $E[R_1]$ is lower for declustered than clustered placement, this does not necessarily mean that the probability of exposure level transition $P_{1\to 2}$ is also lower for declustered placement. This is because, in clustered placement, there are only $r-1$ nodes that can cause the system to go to exposure level 2, whereas in declustered placement, there are $n-1$ nodes that can cause the system to go to the next exposure level. Given α_1 and R_1, the remaining time to complete rebuild is $\alpha_1 R_1$. As the average rebuild bandwidth is $(n-1)c\mu/2$, it now follows that the expected amount of data not rebuilt in exposure level 1 is $\alpha_1 R_1 (n-1)c\mu/2$. However, copies of only a fraction of this data, $\frac{r-1}{n-1}\alpha_1 R_1 (n-1)c\mu/2$ were shared by the newly failed node due to the nature of the declustered placement scheme. This implies that, in exposure level 2, the amount of most-exposed data, that is, the data which have lost 2 copies, is $(r-1)\alpha_1 R_1 c\mu/2$. As the system performs intelligent rebuild, that is, rebuilding the most-exposed data first, and as the total average rebuild bandwidth is $(n-2)c\mu/2$, the expected rebuild time in exposure level 2 is

$$\frac{1}{\mu_2} = E[R_2|R_1,\alpha_1] = \frac{(r-1)\alpha_1 R_1 c\mu/2}{(n-2)c\mu/2} = \frac{r-1}{n-2}\alpha_1 R_1. \quad (37)$$

Following similar arguments as in clustered placement, the distribution of $R_2|R_1,\alpha_1$ could follow (31) or (32). Model B may be applicable here as the system has to adapt from rebuilding using, say $\tilde{n}$ nodes, to using $\tilde{n}-1$ nodes following a node failure and so there is not much change in the randomness associated with the rebuild process. Similarly, for higher exposure levels, we get

$$\frac{1}{\mu_e} = E[R_e|R_{e-1},\alpha_{e-1}] = \frac{r-e+1}{n-e}\alpha_{e-1}R_{e-1} \quad (38)$$

$$R_e|R_{e-1},\alpha_{e-1} \sim G_{\mu_e} \quad \text{(model A)} \quad (39)$$

$$R_e|R_{e-1},\alpha_{e-1} = 1/\mu_e \quad \text{(model B)} \quad (40)$$

4) Conditional Probability of Exposure Level Transition: Suppose there are $\tilde{n}(e)$ nodes in exposure level e whose failure before rebuild can cause the system to go to exposure level $e+1$. Denote the times to failure of these nodes by $E_t^{(i)}$, $i = 1,\cdots,\tilde{n}(e)$. According to Lemma 2, the distribution of these times is $\tilde{F}_\lambda$ given by (21). Denote by $P_{e\to e+1}(R_e)$ the conditional probability of transition to exposure level $e+1$ given that the rebuild time is R_e. Then,

$$\begin{aligned} P_{e\to e+1}(R_e) &= \Pr\{\min_i E_t^{(i)} \le R_e\} \\ &= 1-(1-\Pr\{E_t^{(i)} \le R_e\})^{\tilde{n}(e)}. \end{aligned} \quad (41)$$

Using (21) and (13), we have

$$\Pr\{E_t^{(i)} \le R_e\} = \lambda \int_0^{R_e} (1-F_\lambda(t))dt \approx \lambda R_e. \quad (42)$$

Substituting the above equation in (41), and ignoring higher powers of λR_e, we get

$$P_{e\to e+1}(R_e) \approx \tilde{n}(e)\lambda R_e. \quad (43)$$

For clustered placement, the number of nodes whose failure can cause an exposure level transition $e \to e+1$ is $\tilde{n}(e) = r-e$. Thus,

$$P^{\text{clus.}}_{e\to e+1}(R_e) \approx (r-e)\lambda R_e. \quad (44)$$

For declustered placement, $\tilde{n}(e) = n-e$. Thus,

$$P^{\text{declus.}}_{e\to e+1}(R_e) \approx (n-e)\lambda R_e. \quad (45)$$

5) Expression for P_{DL}: Consider a realization of the direct path to data loss with fractions α_e, $e = 1,\cdots,r-2$, and R_e, $e = 1,\cdots,r-1$, the rebuild times. Denote the vector $(\alpha_1,\cdots,\alpha_{r-2})$ by $\vec{\alpha}$ and $(R_1,\cdots,R_{r-1})$ by $\vec{R}$ for notational convenience, and denote the conditional probability of this direct path by $P_{DL,\text{direct}}(\vec{\alpha},\vec{R})$. Then, using (43),

$$P_{DL,\text{direct}}(\vec{\alpha},\vec{R}) = \prod_{e=1}^{r-1} P_{e\to e+1}(R_e) \approx \lambda^{r-1}\prod_{e=1}^{r-1}\tilde{n}(e)R_e. \quad (46)$$

By unconditioning on $\vec{\alpha}$ and $\vec{R}$, we now obtain

$$P_{DL,\text{direct}} = E\left[P_{DL,\text{direct}}(\vec{\alpha},\vec{R})\right] \quad (47)$$

$$\approx \lambda^{r-1}E[R_1R_2\cdots R_{r-1}]\prod_{e=1}^{r-1}\tilde{n}(e). \quad (48)$$

Then by the direct path approximation (27), the probability P_{DL} is given by

$$P_{DL} \approx P_{DL,\text{direct}} \approx \lambda^{r-1}E[R_1R_2\cdots R_{r-1}]\prod_{e=1}^{r-1}\tilde{n}(e), \quad (49)$$

where $\tilde{n}(e) = r-e$ for clustered placement and $\tilde{n}(e) = n-e$ for declustered placement.

E. Clustered vs. Declustered

As discussed in Section VI-D3, there are two models for the times to rebuild in higher exposure levels. As it turns out, closed form expressions of MTTDL under model A can be obtained for certain families of rebuild distributions G_μ including Weibull and exponential distributions. On the other hand, closed form expressions of MTTDL are available under model B for arbitrary rebuild distributions. Owing to space limitation, we now proceed to obtain the expressions for MTTDL under model B.

1) Clustered Placement: For clustered placement, substituting $\tilde{n}(e) = r-e$ in (49), we get

$$\begin{aligned} P^{\text{clus.}}_{DL} &\approx \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-1}R_e] \\ &= \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-2}R_e\cdot E[R_{r-1}|R_1,\cdots,R_{r-2}]]. \end{aligned}$$

Given R_{r-2}, R_{r-1} is independent of $R_1, \cdots, R_{r-3}$. Therefore, $E[R_{r-1}|R_1, \cdots, R_{r-2}] = E[R_{r-1}|R_{r-2}]$. Thus,

$$\begin{aligned} P_{DL}^{\text{clus.}} &\approx \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-2} R_e \cdot E[R_{r-1}|R_{r-2}]] \\ &= \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-2} R_e \cdot E[E[R_{r-1}|R_{r-2}, \alpha_{r-2}]]]. \end{aligned}$$

Substituting for $E[R_{r-1}|R_{r-2}, \alpha_{r-2}]$, using (33), and given that α_{r-2} is independent of $R_1, \cdots, R_{r-2}$ with $\alpha_{r-2} \sim U(0,1)$, we get

$$\begin{aligned} P_{DL}^{\text{clus.}} &\approx \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-2} R_e \cdot E[\alpha_{r-2}R_{r-2}|R_{r-2}]] \\ &= \lambda^{r-1}(r-1)!E[\frac{R_{r-2}^2}{2}\prod_{e=1}^{r-3} R_e]. \end{aligned} \quad (50)$$

Similarly, using the assumption of model B from (35), and the fact that $\alpha_e \sim U(0,1)$ such that $E[\alpha_e^k] = 1/(k+1)$, we get

$$E\left[R_e^k \prod_{i=1}^{e-1} R_i\right] = E\left[\frac{R_{e-1}^{k+1}}{k+1}\prod_{i=1}^{e-2} R_i\right] \quad \text{(for model B)} \quad (51)$$

Using the above recursion, we get

$$P_{DL}^{\text{clus.}} \approx \lambda^{r-1}(r-1)!E[R_1^{r-1}]\frac{1}{(r-1)!}. \quad (52)$$

Multiplying and dividing by $E[R_1]^{r-1}$ and noting that $E[R_1] = 1/\mu$ from (29), we get

$$P_{DL}^{\text{clus.}} \approx \left(\frac{\lambda}{\mu}\right)^{r-1} \frac{E[R_1^{r-1}]}{E[R_1]^{r-1}}. \quad (53)$$

According to (29), $R_1 \sim G_\mu$, and by denoting the kth moment of G_μ by $m_k(\mu)$,

$$P_{DL}^{\text{clus.}} \approx \left(\frac{\lambda}{\mu}\right)^{r-1} \frac{m_{r-1}(\mu)}{m_1(\mu)^{r-1}}. \quad (54)$$

An estimate for the MTTDL then follows from (25):

$$\text{MTTDL}^{\text{clus.}} \approx \frac{\mu^{r-1}}{n\lambda^r}\frac{m_1(\mu)^{r-1}}{m_{r-1}(\mu)}. \quad (55)$$

Note that the above expressions (54) and (55) are valid under rebuild model B as described in Section VI-D3.

Note that all calculations until (50) are also valid for model A. However, after that, expressions of the form $E[R_e^k|R_{e-1}, \alpha_{e-1}]$ need to be evaluated. From (33) and (34), it follows that $R_e|R_{e-1}, \alpha_{e-1} \sim G_{\mu_e}$, where $\mu_e = 1/(\alpha_{e-1}R_{e-1})$. Therefore, $E[R_e^k|R_{e-1}, \alpha_{e-1}] = m_k(\mu_e) = m_k(1/(\alpha_{e-1}R_{e-1}))$. From this point on, it is difficult to carry forward without a closed form expression of the function $m_k(\cdot)$ as we would then be requiring expectations of the form $E[R_{e-1}m(1/(\alpha_{e-1}R_{e-1}))]$. However, one point worth noting is that the above mentioned difficulties in model A do not arise for $r \leq 3$. In fact, model A and model B start to differ in the MTTDL estimates only for $r > 3$. Therefore, in fact the expression (55) is valid under both models for $r \leq 3$. As an example under model A, if G_μ is exponential, the expression for MTTDL is

$$\text{MTTDL}^{\text{clus.}} \approx \frac{\mu^{r-1}}{n\lambda^r}\frac{m_1(\mu)^{r-1}}{m_{r-1}(\mu)}\prod_{e=1}^{r-3}\frac{1}{(r-e-1)^e} \quad (56)$$

2) Declustered Placement: For declustered placement, substituting $\tilde{n}(e) = n - e$ into (49), we get

$$\begin{aligned} P_{DL}^{\text{declus.}} &\approx \lambda^{r-1}E[\prod_{e=1}^{r-1} R_e]\prod_{e=1}^{r-1}(n-e) \\ &= \lambda^{r-1}E[\prod_{e=1}^{r-2} R_e \cdot E[R_{r-1}|R_1, \cdots, R_{r-2}]] \\ &\quad \times \prod_{e=1}^{r-1}(n-e). \end{aligned} \quad (57)$$

Given R_{r-2}, R_{r-1} is independent of $R_1, \cdots, R_{r-3}$. Therefore, $E[R_{r-1}|R_1, \cdots, R_{r-2}] = E[R_{r-1}|R_{r-2}]$. Substituting this above,

$$\begin{aligned} P_{DL}^{\text{declus.}} &\approx \lambda^{r-1}(r-1)!E[\prod_{e=1}^{r-2} R_e \cdot E[R_{r-1}|R_{r-2}]] \\ &\quad \times \prod_{e=1}^{r-1}(n-e) \end{aligned} \quad (58)$$

$$\begin{aligned} &= \lambda^{r-1}E[\prod_{e=1}^{r-2} R_e \cdot E[E[R_{r-1}|R_{r-2}, \alpha_{r-2}]]] \\ &\quad \times \prod_{e=1}^{r-1}(n-e) \end{aligned} \quad (59)$$

Substituting for $E[R_{r-1}|R_{r-2}, \alpha_{r-2}]$, using (38), and recalling that $\alpha_{r-2} \sim U(0,1)$, we get

$$\begin{aligned} P_{DL}^{\text{declus.}} &\approx \lambda^{r-1}E[\prod_{e=1}^{r-2} R_e \cdot E[\alpha_{r-2}R_{r-2}|R_{r-2}]] \\ &\quad \times \frac{r-(r-1)+1}{n-(r-1)}\prod_{e=1}^{r-1}(n-e) \end{aligned} \quad (60)$$

$$\begin{aligned} &= \lambda^{r-1}E[R_{r-2}^2\prod_{e=1}^{r-3} R_e]\frac{1}{2}\frac{r-(r-1)+1}{n-(r-1)} \\ &\quad \times \prod_{e=1}^{r-1}(n-e). \end{aligned} \quad (61)$$

Similarly, using the assumption of model B from (40), and given that $\alpha_e \sim U(0,1]$ such that $E[\alpha_e^k] = 1/(k+1)$, we get

$$E\left[R_e^k \prod_{i=1}^{e-1} R_i\right] = E\left[\frac{R_{e-1}^{k+1}}{k+1}\prod_{i=1}^{e-2} R_i\right]\left(\frac{r-e+1}{n-e}\right)^k \quad \text{(for model B)} \quad (62)$$

Using the above recursion and rewriting the product, we get

$$P_{DL}^{\text{declus.}} \approx \lambda^{r-1}E[R_1^{r-1}]\frac{(n-1)^{r-1}}{(r-1)!}\prod_{e=1}^{r-2}\left(\frac{r-e}{n-e}\right)^{r-e-1} \quad (63)$$

Multiplying and dividing by $E[R_1]^{r-1}$ and noting that $E[R_1] = 2/((n-1)\mu)$ from (36), we get

$$P_{DL}^{\text{declus.}} \approx \left(\frac{\lambda}{\mu}\right)^{r-1} \frac{2^{r-1}}{(r-1)!} \frac{E[R_1^{r-1}]}{E[R_1]^{r-1}} \prod_{e=1}^{r-2} \left(\frac{r-e}{n-e}\right)^{r-e-1} \quad (64)$$

Recognizing from (36) that $R_1 \sim G_{(n-1)\mu/2}$,

$$P_{DL}^{\text{declus.}} \approx \left(\frac{\lambda}{\mu}\right)^{r-1} \frac{2^{r-1}}{(r-1)!} \prod_{e=1}^{r-2} \left(\frac{r-e}{n-e}\right)^{r-e-1} \times \frac{m_{r-1}(\frac{(n-1)\mu}{2})}{m_1(\frac{(n-1)\mu}{2})^{r-1}}. \quad (65)$$

An estimate for the MTTDL then follows from (25):

$$\text{MTTDL}^{\text{clus.}} \approx \left(\frac{\lambda}{\mu}\right)^{r-1} \frac{(r-1)!}{2^{r-1}} \prod_{e=1}^{r-2} \left(\frac{n-e}{r-e}\right)^{r-e-1} \times \frac{m_1(\frac{(n-1)\mu}{2})^{r-1}}{m_{r-1}(\frac{(n-1)\mu}{2})}. \quad (66)$$

Note that the above expressions (65) and (66) are valid under rebuild model B as described in Section VI-D3. It is also valid under model A for $r \leq 3$ as discussed in Section VI-E1.

F. Insensitivity to Failure Distributions

It can be observed by comparing the results of the analysis in (55) and (66) with that done in [3] using exponential distribution for times to node failure and deterministic rebuild times, that the results are indeed insensitive to the actual distributions of the times to failure as long as the nodes are generally reliable. It must however be noted that, we have an additional level of approximation in this paper when compared to [3] by neglecting the effect of the transient period of the system. In essence, the approximation lies in using the stationary node-excess-time distribution $\widetilde{F}_\lambda$ for the entire lifetime of the system. Note that, for an exponential failure distribution $\widetilde{F}_\lambda$ is the same as F_λ.

G. Sensitivity to Rebuild Distributions

Distribution of rebuilds affect the reliability of the system by a varying degree that depends on the replication factor. For replication factor two, the distribution of rebuild does not have any effect at all. For higher replication factors, randomness in rebuild times affects the reliability by a factor equal to the ratio of the $(r-1)$th moment to the $(r-1)$th power of the first moment. This is because of the effect commonly known as *waiting time paradox*. In this context, node failures tend to occur in longer rebuild intervals, which in turn translates to larger amounts of exposed data.

The modeling assumptions as discussed in Section VI-D3 also have an effect on the reliability. We considered two main modeling assumptions: 1) Under the so-called model A, we assume that the randomness in the rebuild process is completely refreshed every time an exposure level transition occurs. This has an effect of further increasing the randomness of the rebuild times as the system goes to higher exposure levels, and therefore results in lower reliability; 2) Under the so-called model B, we assume that the randomness in the rebuild process is not affected at all (or is affected negligibly) when an exposure level transition occurs. This model results in higher reliability than model A. It was also found that these two models agree on the reliability for $r \leq 3$ and start to differ only for $r > 3$.

VII. Simulation

The storage system is simulated using an event-driven simulator with three types of events that drive the simulation time forward: *(a) failure events*, *(b) rebuild-complete events*, and *(c) node-restore events*. The state of the system is maintained by the following variables: `time`, the simulated time, `failTimes`, a list of times to next failure of each node drawn according to the chosen failure distribution, `failedNodes`, the list of nodes that have failed in the system and are being rebuilt, `exposureLevel`, the exposure level, and a vector of length $(r+1)$ `dataExposure` $= (D_0, \cdots, D_r)$, where D_l is the number of distinct data blocks that have lost l replicas. The values of these variables are updated at each event, and when $D_r > 0$, data loss is said to have occurred and the simulation ends. For each set of parameters, the simulation is run 100 times, and the MTTDL and its bootstrap 95% confidence intervals are computed.

A. Theory vs. Simulation

Although some of the assumptions used in the theoretical analysis, such as independence of node failures, are also used in the simulation, the simulation results reflect a more realistic picture of the systems's reliability. This is because of the following key differences between the theoretical analysis and the simulations. The theoretical estimate of MTTDL in (25) takes into account only the time spent by the system in the failure-free state and ignores the rebuild times, whereas the simulations do not ignore the rebuild times when calculating the times to data loss. Furthermore, in (27), P_{DL} is approximated by the probability of the direct path to data loss, thereby implicitly assuming that this is the only path following a first-node-failure event that would lead to data loss. In simulations however, all the complex trajectories of the system through the different exposure levels are simulated by simulating random node failure events and updating the data exposure vector by taking partial rebuilds into account. In the theoretical analysis, the time required to restore new nodes in a declustered placement scheme (following successful rebuild of lost replicas in the spare space of surviving nodes) is ignored, whereas in the simulations, the time to restore new nodes is simulated as well. In addition, other approximations made in the analysis, such as neglecting the effect of the transient period of the system, are implicitly avoided in the simulations. Therefore, the simulations reflect a more complex picture of the system behavior than what is assumed in theory.

B. Simulation Results

Table II shows the range of parameters used for the simulations. Typical values for practical systems are used for all

TABLE II
RANGE OF VALUES OF DIFFERENT SIMULATION PARAMETERS

Parameter	Meaning	Range
c	storage capacity of each node	12 TB
n	number of storage nodes	4 to 50
r	replication factor	2, 3
$c\mu$	rebuild bandwidth available at each node	96 MB/s
$1/\lambda$	mean time to failure of a node	10^3 to 10^4 h

parameters, except for the mean times to failure of a node, which have been chosen artificially low (10000 h and 1000 h for replication factors 2 and 3, respectively) to shorten the simulation times. The running times of simulations with practical values of the mean times to node failure, which are of the order of 10000 h or higher, are prohibitively high; this is due to the fact that P_{DL} becomes extremely low thereby making the number of first-node-failure events that need to be simulated (along with the other complex set of events that restore all lost replicas following each first-node-failure event) extremely high for each run of the simulation. Although this approach scales down the MTTDL by making failure events more frequent, its use is justified because it preserves the ratios of MTTDLs of the various schemes [6].

Replication Factor 2: Fig. 1 shows the comparison of theoretically predicted and simulation-based MTTDL values for a system with replication factor 2 and mean time to failure of a node, $1/\lambda$, equal to 10000 h as the number of nodes n in the system is varied. From (55) and (66), for $r = 2$, the MTTDL of a system with deterministic rebuild times is given by

$$\text{MTTDL}^{\text{clus.}} \approx \mu/(n\lambda^2), \quad (67)$$

$$\text{MTTDL}^{\text{declus.}} \approx \mu/(2n\lambda^2), \quad (68)$$

irrespective of the underlying failure distribution. It is observed that the theoretically predicted values, although approximate, match well the simulation-based values as they typically lie within the 95% confidence intervals. This also holds when the failure distribution is varied from exponential to Weibull and when the shape parameter of the Weibull distribution is altered. We notice however that, there is a slight deviation from the theoretical value for Weibull distribution with shape parameter 0.7 as the number of nodes increases. This is however expected as Weibull shape parameters less than one represent nodes with high infant mortality and do not necessarily satisfy condition (2) as discussed in Section III. Note that, for replication factor 2, the MTTDL is also invariant with respect to the rebuild distribution. This is because the terms involving higher moments of the rebuild distribution do not appear in the expressions for MTTDL (55) and (66).

Although the mean of the times to data loss, that is the MTTDLs, are invariant with respect to the node failure distribution, the cumulative distribution functions of the times to data loss depends on the underlying node failure distribution. This is shown in Fig. 2 by the empirical distributions of the time to data loss for two cases of node failure distribution, namely, exponential and Weibull (with shape 1.2). Although the MTTDL is the same for both distributions (29314 days),

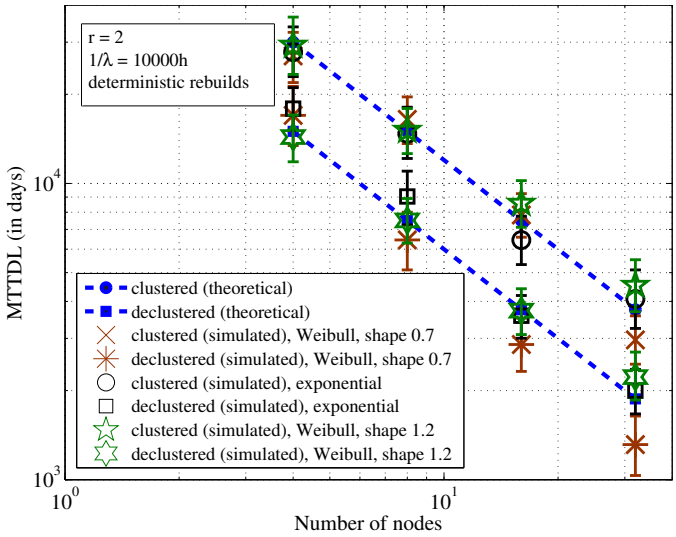


Fig. 1. Comparison of theoretically predicted and simulated values of MTTDL for a replication factor of two with mean time to failure of a node equal to 10000 h. Simulations are done with deterministic rebuild times and exponential or Weibull distributions for failure times. For the simulated results, 95% bootstrap confidence intervals are shown.

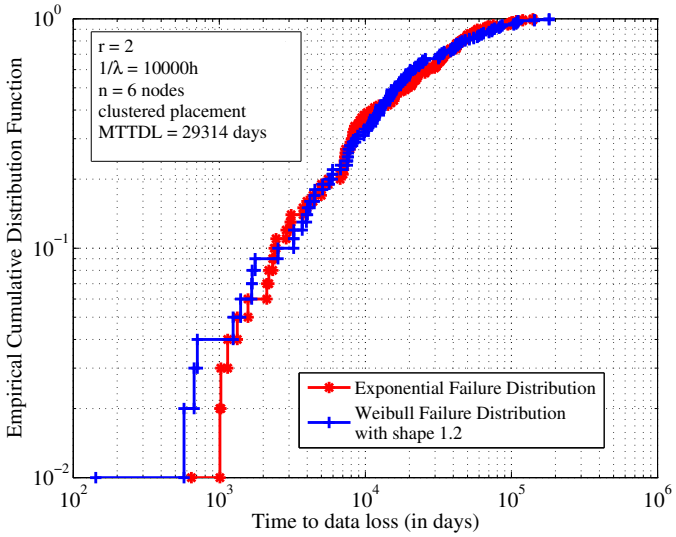


Fig. 2. Comparison of empirical cumulative distribution functions of the time to data loss for exponential and Weibull failure distributions. The above graph is for a system with a replication factor of two, six nodes, each with mean time to failure of 10000 h, clustered placement, and deterministic rebuilds.

the probability that data loss occurs within shorter durations (of the order of 1000 days) is much higher for Weibull distribution than for exponential distribution.

Replication Factor 3: From (55) and (66), for replication factor $r = 3$, the MTTDL of a system with deterministic rebuilds is given by

$$\text{MTTDL}^{\text{clus.}} \approx \mu^2/(n\lambda^3), \quad (69)$$

$$\text{MTTDL}^{\text{declus.}} \approx (n-1)\mu^2/(4n\lambda^3), \quad (70)$$

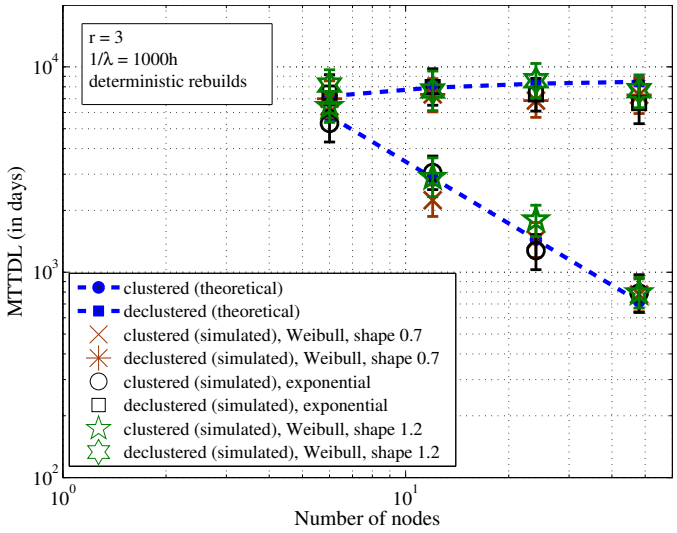


Fig. 3. Comparison of theoretically predicted and simulated values of MTTDL for a replication factor of three with mean time to failure of a node equal to 1000 h. Simulations are done with deterministic rebuild times and exponential or Weibull distributions for failure times. For the simulated results, 95% bootstrap confidence intervals are shown.

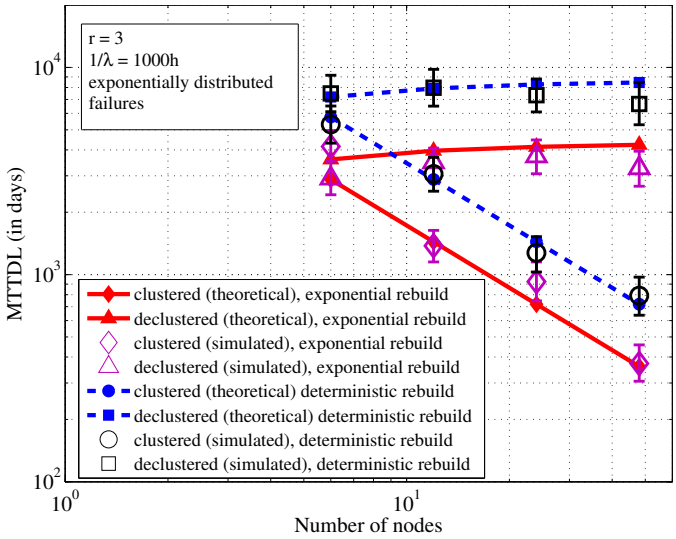


Fig. 4. Comparison of theoretically predicted and simulated values of MTTDL for a replication factor of three with mean time to failure of a node equal to 1000 h. Simulations are done with exponential distribution for failure times and deterministic or exponentially distributed rebuild times. For the simulated results, 95% bootstrap confidence intervals are shown.

irrespective of the underlying failure distribution. Theoretical estimates of MTTDL match well with the simulation-based values as seen in Fig. 3 as the underlying failure distribution is varied. In contrast, the MTTDL depends on the second moment of the rebuild distribution as seen in (55) and (66). For instance, for an exponential rebuild distribution, MTTDL expressions (55) and (66) reduce to

$$\text{MTTDL}^{\text{clus.}} \approx \mu^2/(2n\lambda^3), \tag{71}$$

$$\text{MTTDL}^{\text{declus.}} \approx (n-1)\mu^2/(8n\lambda^3). \tag{72}$$

irrespective of the underlying failure distribution. In essence, the MTTDLs are half of those corresponding to deterministic rebuilds. This is confirmed by simulations as shown in Fig. 4, where the failure distribution is kept the same (exponential) while the rebuild times are chosen to be either deterministic or exponentially distributed.

C. Summary of Findings

The following lists the findings of this paper:

- The MTTDL expressions (55) and (66) are invariant with respect to the distribution of times to node failures as long as the distribution satisfies (2). This implies that the results obtained in literature assuming exponential distributions may be applicable to other distributions as well.
- Condition (2) is satisfied by a wide variety of failure and rebuild distributions, including most importantly, real world distributions.
- Although, the MTTDL is invariant with respect to the failure time distribution, the actual distribution of the time to data loss may in fact depend on the failure time distribution.
- The MTTDL expressions (55) and (66) depend on the rebuild time distribution. More specifically, the MTTDL is inversely proportional to the $(r-1)$th moment of the rebuild time distribution, where r is the replication factor used. Therefore, rebuild distributions with higher variability lead to lower MTTDL.
- The dependence of MTTDL on rebuild distributions arises because of a phenomenon commonly referred to as the waiting time paradox. Essentially, given that a failure occurred during rebuild, it is more likely that the rebuild period was lengthier, and therefore a larger proportion of data not rebuilt becomes exposed. This leads to a cascading effect of a larger expected rebuild time for most-exposed data blocks.

VIII. Conclusions

In this paper, we presented a general reliability model for data storage systems that includes a wide-variety of failure and rebuild distributions, including, most importantly, those observed in real world systems. For systems with generally reliable nodes, our analysis demonstrated that the expressions for MTTDL are essentially invariant within a large class of failure distributions. This result is significant because this class includes both real world distributions, such as Weibull which are difficult to analyze theoretically, as well as the exponential distribution, which is amenable to theoretical analysis. This implies that many MTTDL results derived in the literature assuming exponential failure distributions may hold even under other more realistic distributions. We also showed that the MTTDL is affected by the variability of rebuild distributions; rebuild distributions having higher variability are shown to have lower MTTDL values. How the relative values of the

MTTDLs relate to the relative values of their corresponding reliability functions is a subject of further investigation.

IX. Acknowledgment

The authors would like to thank Prof. Rüdiger Urbanke and Prof. Christina Fragouli for their helpful comments through the preparation of this paper.

Appendix A
Mean Operational Period of the System

The following derivation of the mean operational period of a system has been adapted from [13, Chap. 2, pp. 139–140]. Let

$$\widetilde{F}^{(i)}_{\lambda,A_t}(\tau) := \Pr\{A^{(i)}_t \le \tau | \nu^{(i)}_t = 1\}, \quad (73)$$

$$\widetilde{F}^{(i)}_{\lambda,E_t}(\tau) := \Pr\{E^{(i)}_t \le \tau | \nu^{(i)}_t = 1\}. \quad (74)$$

According to Lemma 2, the above sequences, $\widetilde{F}^{(i)}_{\lambda,A_t}$ and $\widetilde{F}^{(i)}_{\lambda,E_t}$, converge pointwise to $\widetilde{F}_\lambda$:

$$\lim_{t\to\infty} \widetilde{F}^{(i)}_{\lambda,A_t}(\tau) = \lim_{t\to\infty} \widetilde{F}^{(i)}_{\lambda,E_t}(\tau) = \widetilde{F}_\lambda(\tau), \quad (75)$$

for $i = 1, \cdots, n$, where $\widetilde{F}_\lambda$ is given by (21). If $E^{(i)}(t, \Delta t, \tau)$ denotes the event that the node i was renewed in the interval $(t - \Delta t, t)$, that it operates without failure in $(t, t + \tau)$, and that the remaining nodes operate without failure in $(t, t + \tau)$, then the event $E(t, \Delta t, \tau)$ can be written as the disjoint union

$$E(t, \Delta t, \tau) = E^{(1)}(t, \Delta t, \tau) \cup \cdots \cup E^{(n)}(t, \Delta t, \tau), \quad (76)$$

by ignoring events that have probabilities of higher order in Δt such as more than one rebuild event within a Δt time period. Therefore,

$$\begin{aligned}\Pr\{E(t, \Delta t, \tau)\} &= \sum_{i=1}^{n} \Pr\{E^{(i)}(t, \Delta t, \tau)\} \\ &= \sum_{i=1}^{n} \left[\Pr\{A^{(i)}_t \le \Delta t, E^{(i)}_t > \tau, \nu^{(i)}_t = 1\} \right. \\ &\qquad \left. \times \prod_{\substack{j=1 \\ j\neq i}}^{n} \Pr\{E^{(j)}_t > \tau, \nu^{(j)}_t = 1\} \right], \quad (77)\end{aligned}$$

The first term in the summation above can be expanded as

$$\begin{aligned}&\Pr\{A^{(i)}_t \le \Delta t, E^{(i)}_t > \tau, \nu^{(i)}_t = 1\} \quad (78) \\ &\quad = \Pr\{\nu^{(i)}_t = 1\} \Pr\{A^{(i)}_t \le \Delta t | \nu^{(i)}_t = 1\} \\ &\qquad \cdot \Pr\{E^{(i)}_t > \tau | A^{(i)}_t \le \Delta t, \nu^{(i)}_t = 1\} \\ &\quad = a^{(i)}_t \widetilde{F}^{(i)}_{\lambda,A_t}(\Delta t)(1 - F_{\lambda,\Delta t}(\tau)) \quad (79)\end{aligned}$$

Here, (79) follows from (16), (73), and (74) and $F_{\lambda,\Delta t}(\tau)$ converges to $F_\lambda(\tau)$ as Δt tends to zero. Furthermore, as $\widetilde{F}^{(i)}_{\lambda,A_t}(\Delta t)$ converges to $\widetilde{F}_\lambda(\Delta t)$ by Lemma 2, using (21), we can write $\widetilde{F}^{(i)}_{\lambda,A_t}(\Delta t)$ as

$$\widetilde{F}^{(i)}_{\lambda,A_t}(\Delta t) = \lambda \Delta t + o(\Delta t), \quad (80)$$

where the small-'o' notation is used to denote that the term $o(\Delta t)$ tends to zero faster than Δt as Δt tends to zero. Therefore, (79) reduces to

$$\begin{aligned}&\Pr\{A^{(i)}_t \le \Delta t, E^{(i)}_t > \tau, \nu^{(i)}_t = 1\} \\ &\qquad = a^{(i)}_t \lambda \Delta t (1 - F_{\lambda,\Delta t}(\tau)) + o(\Delta t). \quad (81)\end{aligned}$$

Using (16) and (74), the product term in (77) can be expanded as follows:

$$\begin{aligned}&\prod_{\substack{j=1 \\ j\neq i}}^{n} \Pr\{E^{(j)}_t > \tau, \nu^{(j)}_t = 1\} \\ &\qquad = \prod_{\substack{j=1 \\ j\neq i}}^{n} \Pr\{\nu^{(j)}_t = 1\} \Pr\{E^{(j)}_t > \tau | \nu^{(j)}_t = 1\} \\ &\qquad = \prod_{\substack{j=1 \\ j\neq i}}^{n} a^{(j)}_t (1 - \widetilde{F}^{(j)}_{\lambda,E_t}(\tau)). \quad (82)\end{aligned}$$

Substituting (81) and (82) into (77), we get

$$\begin{aligned}\Pr\{E(t, \Delta t, \tau)\} &= \lambda \Delta t (1 - F_{\lambda,\Delta t}(\tau)) \\ &\quad \times \sum_{i=1}^{n} \left[a^{(i)}_t \prod_{\substack{j=1 \\ j\neq i}}^{n} a^{(j)}_t (1 - \widetilde{F}^{(j)}_{\lambda,E_t}(\tau)) \right] + o(\Delta t). \quad (83)\end{aligned}$$

Similar to the calculations above for $\Pr\{E(t, \Delta t, \tau\}$, the probability $\Pr\{E(t, \Delta t)\}$ can be computed by writing $E(t, \Delta t)$ as a disjoint union of events. The result is:

$$\Pr\{E(t, \Delta t)\} = \lambda \Delta t \sum_{i=1}^{n} \left[a^{(i)}_t \prod_{\substack{j=1 \\ j\neq i}}^{n} a^{(j)}_t \right] + o(\Delta t). \quad (84)$$

Substituting (83) and (84) into (22) and computing the limit as Δt tends to zero, we get

$$\begin{aligned}p_t(\tau) &= \lim_{\Delta t \to 0} \frac{\Pr\{E(t, \Delta t, \tau)\}}{\Pr\{E(t, \Delta t)\}} \\ &= (1 - F_\lambda(\tau)) \\ &\quad \times \frac{\sum_{i=1}^{n} \left[a^{(i)}_t \prod_{\substack{j=1 \\ j\neq i}}^{n} a^{(j)}_t (1 - \widetilde{F}^{(j)}_{\lambda,E_t}(\tau)) \right]}{\sum_{i=1}^{n} \left[a^{(i)}_t \prod_{\substack{j=1 \\ j\neq i}}^{n} a^{(j)}_t \right]}. \quad (85)\end{aligned}$$

Using (17), (75), and (21), (85) yields

$$\begin{aligned}\lim_{t\to\infty} p_t(\tau) &= (1 - F_\lambda(\tau))(1 - \widetilde{F}_\lambda(\tau))^{(n-1)} \\ &= -\frac{1}{n\lambda} \frac{d}{d\tau} (1 - \widetilde{F}_\lambda(\tau))^n, \quad (86)\end{aligned}$$

Thus,

$$T = \lim_{t\to\infty} T_t = \lim_{t\to\infty} \int_0^\infty p_t(\tau) d\tau. \quad (87)$$

From (85), it can be seen that $p_t(\tau) \le 1 - F(\tau)$. As $1 - F(\tau)$ is integrable, by the dominated convergence theorem, the

limit and the integral can be exchanged in the above equation. Therefore,

$$\begin{aligned} T &= \int_0^\infty \lim_{t\to\infty} p_t(\tau)d\tau \\ &= \int_0^\infty -\frac{1}{n\lambda}\frac{d}{d\tau}(1-\widetilde{F}_\lambda(\tau))^n = \frac{1}{n\lambda}. \end{aligned}$$

APPENDIX B
FRACTION OF DATA REBUILT

Suppose there are $\tilde{n}$ nodes whose failure during rebuild can cause an exposure level transition. Denote the times to failures of these nodes by $E_t^{(i)}$, $i = 1, \cdots, \tilde{n}$. According to Lemma 2 and the node-failure independence assumption, in the stationary period of the system, $E_t^{(i)}$ are independent and identically distributed according to $\widetilde{F}_\lambda$.

We are interested in the fraction α of rebuild time left when a node failure occurs, given that this failure occurred before rebuild completes, that is,

$$\alpha = (R - \min_i E_t^{(i)})/R, \text{ given that } \min_i E_t^{(i)} < R.$$

The distribution function of α for $x \in (0, 1]$ is

$$\begin{aligned} \Pr\{\alpha \le x\} &= \Pr\left\{ \frac{R - \min_i E_t^{(i)}}{R} \le x \middle| \min_i E_t^{(i)} < R \right\} \\ &= \frac{\Pr\{R(1-x) \le \min_i E_t^{(i)} < R\}}{\Pr\{\min_i E_t^{(i)} < R\}} \\ &= 1 - \frac{\Pr\{\min_i E_t^{(i)} < R(1-x)\}}{\Pr\{\min_i E_t^{(i)} < R\}} \\ &= 1 - \frac{1-(1-\Pr\{E_t < R(1-x)\})^{\tilde{n}}}{1-(1-\Pr\{E_t < R\})^{\tilde{n}}}. \quad (88) \end{aligned}$$

It is shown in Appendix C that

$$\begin{aligned} \Pr\{E_t < R\} &= \lambda/\mu + o(\lambda/\mu), \\ \Pr\{E_t < R(1-x)\} &= (1-x)\lambda/\mu + o((1-x)\lambda/\mu). \end{aligned}$$

Therefore,

$$\Pr\{\alpha \le x\} = 1 - \frac{\tilde{n}(1-x)\lambda/\mu + o((1-x)\lambda/\mu)}{\tilde{n}\lambda/\mu + o(\lambda/\mu)} \approx x. \quad (89)$$

This means that, for highly reliable systems, α is uniformly distributed between zero and one.

APPENDIX C
PROBABILITY OF FAILURE BEFORE REBUILD COMPLETION

According to (20) and (21), it holds that

$$\begin{aligned} \Pr\{E_t < R\} &= \int_{\tau=0}^\infty \widetilde{F}_\lambda(\tau) dG_\mu(\tau) \\ &= \int_{\tau=0}^\infty \lambda \int_{t=0}^\tau (1-F_\lambda(t)) dt dG_\mu(\tau). \end{aligned}$$

Changing the order of integrals, yields after some manipulations

$$\Pr\{E_t < R\} = \frac{\lambda}{\mu}\left(1 - \mu\int_{t=0}^\infty F_\lambda(t)(1-G_\mu(t))dt\right).$$

In the last step above, we used the fact that integrating the complementary cumulative distribution function $1 - G_\mu(t)$ gives the mean $1/\mu$. As the functions F_λ and G_μ satisfy (13) and (14) respectively, it can be seen that the second term inside the parantheses is $o(1)$. Therefore,

$$\Pr\{E_t < R\} = \lambda/\mu + o(\lambda/\mu).$$

Similarly the following can also be shown for any $x \in (0, 1)$:

$$\Pr\{E_t < Rx\} = x\lambda/\mu + o(x\lambda/\mu).$$

REFERENCES

[1] K. S. Trivedi, *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. Wiley, 2002.

[2] B. Schroeder and G. A. Gibson, "Understanding disk failure rates: What does an MTTF of 1,000,000 hours mean to you?" *ACM Transactions on Storage*, vol. 3, no. 3, pp. 1–31, October 2007.

[3] V. Venkatesan, I. Iliadis, C. Fragouli, and R. Urbanke, "Reliability of clustered vs. declustered replica placement in data storage systems," in *Proc. 19th Annual IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'11)*, 2011, pp. 307–317.

[4] D. A. Patterson, G. Gibson, and R. H. Katz, "A case for redundant arrays of inexpensive disks (raid)," in *SIGMOD '88: Proc. 1988 ACM SIGMOD international conference on Management of data*. ACM, 1988, pp. 109–116.

[5] Q. Xin, E. L. Miller, T. Schwarz, D. D. E. Long, S. A. Brandt, and W. Litwin, "Reliability mechanisms for very large storage systems," in *Proc. 20th IEEE / 11th NASA Goddard Conference on Mass Storage Systems and Technologies (MSS'03)*, 2003, pp. 146–156.

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

[7] S. Ramabhadran and J. Pasquale, "Analysis of long-running replicated systems," in *Proc. 25th IEEE International Conference on Computer Communications (INFOCOM'06)*, 2006, pp. 1–9.

[8] M. Leslie, J. Davies, and T. Huffman, "A comparison of replication strategies for reliable decentralised storage," *Journal of Networks*, vol. 1, no. 6, pp. 36–44, December 2006.

[9] A. Thomasian and M. Blaum, "Mirrored disk organization reliability analysis," *IEEE Transactions on Computers*, vol. 55, pp. 1640–1644, December 2006.

[10] Q. Xin, E. L. Miller, and T. J. E. Schwarz, "Evaluation of distributed recovery in large-scale storage systems," in *Proc. 13th IEEE International Symposium on High Performance Distributed Computing (HPDC'04)*, 2004, pp. 172–181.

[11] K. M. Greenan, E. L. Miller, and J. Wylie, "Reliability of flat XOR-based erasure codes on heterogeneous devices," in *Proc. 38th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'08)*, June 2008, pp. 147–156.

[12] J. Elerath and M. Pecht, "A highly accurate method for assessing reliability of redundant arrays of inexpensive disks (RAID)," *IEEE Transactions on Computers*, vol. 58, pp. 289–299, 2009.

[13] B. Gnedenko, I. Beliaev, A. Solovev, and R. Barlow, *Mathematical methods of reliability theory*, ser. Probability and mathematical statistics. Academic Press, 1969.