# Reliability Equations for Cloud Storage Systems with Proactive Fault Tolerance

Jing Li, Peng Li, Rebecca J. Stones, Gang Wang, *Member, IEEE*, Zhongwei Li, and Xiaoguang Liu

**Abstract**—As cloud storage systems increase in scale, hard drive failures are becoming more frequent, which raises reliability issues. In addition to traditional reactive fault tolerance, proactive fault tolerance is used to improve a system's reliability. However, there are few studies which analyze the reliability of proactive cloud storage systems, and they typically assume an exponential distribution for drive failures. This paper presents closed-form equations for estimating the number of data-loss events in proactive cloud storage systems using RAID-5, RAID-6, 2-way replication, and 3-way replication mechanisms, within a given time period. The equations model the impact of proactive fault tolerance, operational failures, failure restorations, latent block defects, and drive scrubbing on the systems reliability, and use time-based Weibull distributions to represent processes (instead of homogeneous Poisson processes). We also design a Monte-Carlo simulation method to simulate the running of proactive cloud storage systems. The proposed equations closely match time-consuming Monte-Carlo simulations, using parameters obtained from the analysis of field data. These equations allow designers to efficiently estimate system reliability under varying parameters, facilitating cloud storage system design.

**Index Terms**—proactive fault tolerance, cloud storage systems, reliability, time-variant failure rates, latent block defects



## 1 Introduction

Modern-day data centers usually host hundreds of thousands of servers, using hard drives as the primary data storage device. Many challenges are faced for such large data center management [1], [2]. The failure of an individual hard drive might be rare, but a system with thousands of drives will regularly experience failures [3], [4], [5], [6]. Drive failure can result in service unavailability, hurting the user experience, and even permanent data loss. Therefore, high reliability is one of the biggest concerns in such systems.

Traditional cloud storage systems adopt redundancy, e.g., erasure codes and replication, to reconstruct data when drive failure occurs, which is known as *reactive* fault tolerance. To provide satisfactory reliability in large-scale cloud storage systems with high failure frequency, multi-erasure codes (or multiple replicas) must be used, which brings high construction and maintenance cost and heavy read/write overhead. Thus, reactive fault tolerance alone cannot meet the demands of the high reliability and service quality in modern data centers. *Proactive* fault tolerance [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17] instead predicts drive failures and handles them in advance; with sufficient prediction accuracy and effective warning handling, it can significantly enhance the system reliability and reduce costs.

When designing proactive cloud storage systems and tweaking parameters to optimize performance, designers must consider factors such as coding redundancy, failure prediction, the amount of bandwidth used to reconstruct or migrate data after a drive fails or is predicted to fail, and drive scrubbing for eliminating block defects. As such, designers could benefit from an accurate and easy-to-use way to assess the effects of these factors on overall reliability.

So far there are only a few relevant studies which analyze the reliability of proactive cloud storage systems [11], [18], [19]. There are some drawbacks to the current research: (a) Inaccurate failure distribution models—the reliability estimates are based on the assumption that both hard drive failures and their repairs follow a homogeneous Poisson process with constant failure and restoration rates, which was contested in [20], [21], [22], [23]. (b) Incomplete consideration of failures—focusing on whole-drive failures, without incorporating latent block level failure mode, which, with increases in single-drive and whole-system capacity, can not be ignored [24]. (c) Unrealistic reliability metric—mean time to data loss (MTTDL) is excessive relative to the actual run time of cloud storage systems and does not adequately reflect reliability [25].

One could simulate the system to assess reliability more accurately, but at the expense of usability, which would require specialized code, extensive computation, and is time consuming. It is preferable to have reliability equations which are easy to use, applicable to different configurations and executed quickly, especially when frequently tweaking the parameters of a cloud storage system where the executed workload and rates of failures, warnings, etc., fluctuate over time.

- J. Li is with the College of Computer Science and Technology, Civil Aviation University of China, Tianjin, China. E-mail: lijing@nbjl.nankai.edu.cn.
- P. Li, R.J. Stones, G. Wang, Z. Li, and X. Liu are with Nankai-Baidu Joint Lab, College of Computer and Control Engineering, Nankai University, Tianjin, China.
E-mail: {lipeng, becky, wgzwp, lizhongwei, liuxg}@nbjl.nankai.edu.cn.

*Manuscript received 18 May 2017; revised 13 Aug. 2018; accepted 15 Nov. 2018. Date of publication 0 . 0000; date of current version 0 . 0000.*
*(Corresponding author: Xiaoguang Liu.)*
*Recommended for acceptance by S. Kundu.*
*For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.*
*Digital Object Identifier no. 10.1109/TDSC.2018.2882512*

TABLE 1
Mathematical Models for the Expected Number of Data Loss Events in RAID-5, RAID-6, 2-way Replication, and 3-way Replication Systems, Each With Proactive Fault Tolerance

**Inputs:**

| Weibull parameters | | | |
|---|---|---|---|
| Operational failure | Rebuild time | Latent block defect | Scrubbing time |
| $\alpha_f$, $\beta_f$ | $\alpha_r$, $\beta_r$ | $\alpha_l$, $\beta_l$ | $\alpha_s$, $\beta_s$ |
| | $\text{MTTR} = \alpha_r\Gamma(1+1/\beta_r)$ | $\text{MTTB} = \alpha_l\Gamma(1+1/\beta_l)$ | $\text{MTTS} = \alpha_s\Gamma(1+1/\beta_s)$ |

Failure detection rate: FDR

**Models:**

| Parameters | System | Expected no. data loss events | Eq. no. |
|---|---|---|---|
| $g$-drive group | RAID-5 | $(R_{\text{op}} + R_{\text{def}})\,(g-1)\,\hat{H}(t)$ | (7) |
| | RAID-6 | $(R_{\text{op-op}} + R_{\text{op-def}})\,(g-2)\,\hat{H}(t)$ | (8) |
| $r$ racks; $n$ nodes; $d$ drives; $b$ blocks | 2-way | $\left(P_{\text{2-way}}\,(r-1)nd\,D_{\text{op}} + rnd\,(1-A_{\text{def}})\right)\hat{H}(t)$ | (10) |
| | 3-way | $\left(P_{\text{3-way}}\left((r-1)nd\,D^{(1)}_{\text{op-op}} + 2(n-1)d\,D^{(2)}_{\text{op-op}}\right) + 2\,D_{\text{op}}(1-A_{\text{def}})\right)\hat{H}(t)$ | (13) |

**Auxiliary functions:**

adjusted drive availability (operational fail)

$$\hat{A}_{\text{op}} = \frac{\alpha_p}{\alpha_p + (1-\text{FDR})\,\text{MTTR}}; \quad \alpha_p = \frac{\alpha^\beta}{t^{\beta-1}}$$

drive availability (block defect)

$$A_{\text{def}} = \frac{\text{MTTB}}{\text{MTTB}+\text{MTTS}}$$

adjusted cumulative hazard rate

$$\hat{H}(t) = (1-\text{FDR})\,\frac{t^\beta}{\alpha^\beta}$$

**Auxiliary functions:**

$$R_{\text{op}} = 1 - \hat{A}_{\text{op}}^g$$
$$R_{\text{def}} = 1 - A_{\text{def}}^g$$
$$R_{\text{op-op}} = 1 - \hat{A}_{\text{op}}^g - g\,\hat{A}_{\text{op}}^{g-1}\,(1-\hat{A}_{\text{op}})$$
$$R_{\text{op-def}} \simeq 1 - \hat{A}_{\text{op}}^g - A_{\text{def}}^g + (\hat{A}_{\text{op}}\,A_{\text{def}})^g$$

$$D_{\text{op}} = 1 - \hat{A}_{\text{op}}^{rnd}$$
$$D^{(1)}_{\text{op-op}} = 1 - \left((\hat{A}_{\text{op}}^d)^n + n(\hat{A}_{\text{op}}^d)^{n-1}(1-\hat{A}_{\text{op}}^d)\right)^r$$
$$D^{(2)}_{\text{op-op}} = 1 - \hat{A}_{\text{op}}^{rnd} - r\,(\hat{A}_{\text{op}}^{nd})^{r-1}(1-\hat{A}_{\text{op}}^{nd})$$
$$P_{\text{2-way}} = 1 - \left(1 - \tfrac{1}{(r-1)\,nd}\right)^b$$
$$P_{\text{3-way}} = 1 - \left(1 - \tfrac{2}{3(r-1)\,n(n-1)d^2}\right)^b$$

Elerath et al. defined two reliability equations for reactive RAID-5 [22] and RAID-6 [23] groups without disk failure prediction. In this paper, we extend the work to proactive RAID systems and proactive replication systems with disk failure prediction; in these settings the reliability analysis is more intricate due to the need to factor in failure prediction and replica dispersal. We make two main contributions: (a) to incorporate drive failure prediction, we modify the calculation for the cumulative hazard rate and drive availability; and (b) to incorporate replica dispersal, we mathematically derive the probability of data loss.

Specifically, this work generalizes the equations by Elerath et al. [22], [23] for assessing the reliability of proactive RAID-5 and RAID-6 systems, and we further propose two new equations for assessing the reliability of proactive 2-way replication, and 3-way replication systems; these four systems are the most commonly used methods in current data centers. The equations allow for expressions of time-dependent failure and restoration rates, and incorporate block defects, media scrubbing processes, and proactive fault tolerance on reliability. We also design an event-driven Monte-Carlo-based simulation method to simulate cloud storage systems with proactive fault tolerance. The proposed equations and simulations are consistent with one another. Using these equations, system designers can easily assess trade-offs, compare schemes, and understand the effects of the parameters on the overall cloud storage system reliability, allowing them to better design and optimize infrastructures.

The rest of the paper is organized as follows: Section 2 describes the relevant background knowledge. The equations we present are listed in Table 1, and their derivation is given in Section 3. We evaluate their accuracy versus simulations in Section 4, and explore these systems' sensitivity to varying parameters. Section 5 describes some limitations and the efficacy of the models.

# 2 Background

## 2.1 Related Work

Reliability, the focus of this paper, is one of the most important aspects of storage systems and has been studied extensively, especially for RAID systems.

Gibson et al. [26] found that hard drive failure rates typically followed a "bathtub curve". However they still considered the exponential distribution a useful simplifying assumption for modeling drive failure events. Recently, researchers found that drive failure events were not adequately modeled by homogeneous Poisson processes [20], [21], [22], [23], and some other work focus on device failure [27].

In this paper, we use Weibull distributions for modeling drive failure events. This is motivated by Schroeder and Gibson [20], who found that hard drive failure rates were not constant with age, and recommended the Weibull distribution for modeling drive failure, which can account for both "infant mortality" and aging drives.

Elerath et al. defined two equations for assessing the reliability of RAID-5 [22] and RAID-6 [23] groups, and these papers form the basis of the equations proposed in this paper. Greenan et al. [28] argued that MTTDL was a bad reliability metric, and Elerath et al. went so far as to say it should be "put to rest". However, Iliadis and Venkatesan [29] offered a rebuttal. In this paper, we use the expected number of data loss events to measure reliability, consistent with Elerath et al. Elerath et al.'s equations use time-dependent failure and repair rates and included the contributions of both sector defects and data scrubbing. They considered RAID groups without disk failure prediction. In this work, we extend this work to include (a) proactive fault tolerance and (b) replication systems.

For proactive cloud storage systems, there are only a few studies focusing on their reliability. Eckart et al. [18] used Markov models to demonstrate the effect of failure prediction has on a system's MTTDL. They devised models for a single hard drive, RAID-1, and RAID-5 with proactive fault tolerance. Li et al. extended this study to RAID-6 groups [11] and replication systems [19]. However, there are some drawbacks in those studies on reliability of proactive systems (as mentioned in the introduction), which we overcome in this paper.

### 2.2 Drive Failure Modes

*Intermittent failure* is the most common mode of failure. In this case, a drive or some sectors cannot be accessed temporarily, and are often restored by retrying several times.

*Latent block defects* are commonly caused by latent sector errors, a permanent inability to access data from certain sectors (possibly due to physical defects e.g., a scratch), and data corruption, where data stored in a block are incorrect. Latent sector errors are not reported by the drive until the particular sector is accessed. Data corruption can not be reported by the drive even when a defective block is read; it is silent and could have greater impact than other errors.

To detect and protect data against the block defects, cloud storage systems usually perform *drive scrubbing* during idle periods. Drive scrubbing is a background process that proactively reads and checks data from all drive blocks. If a defective block is detected, the system reconstructs the corrupted content from the available data. The time required to scrub an entire drive varies with the drive capacity and the drive scrubbing rate.

A serious type of failure is an *operational failure*, where a whole drive is permanently no longer accessible. Such failure can be repaired only by replacing the drive. Proactive fault tolerance only protects against data loss caused by operational failures.

When an operational failure occurs, the cloud storage system initiates a rebuild process during which it restores the missing data using the accessible surviving data. The rebuild time depends on the amount of data that is transferred during the process, and the data transfer rate. Moreover, to maintain the quality of user service, usually only a fraction of the total bandwidth available is used for a rebuild process. As such, the rebuild time will also be influenced by the foreground activity.

Only the last two failure types—latent block defects and operational failures—impact a storage system's reliability, so we focus on them in this paper.

### 2.3 Proactive Fault Tolerance

Self-Monitoring, Analysis, and Reporting Technology (SMART) is implemented within modern hard drives [30]. SMART monitors and compares drive attributes with preset thresholds, and issues warnings when attributes exceed the thresholds. As a result, systems can act in advance of drive failures, such as by migrating data. This typifies proactive fault tolerance, which fundamentally improves system reliability.

Moreover, to improve prediction accuracy, statistical and machine learning methods have been proposed to build hard drive failure prediction models based on SMART attributes [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], some of which achieve good prediction performance, which is reasonable to use in practice.

Existing drive failure prediction models are mostly only able to predict operational failures in advance; block-defect prediction models are at an early stage in development and do not have suitable practical performance. For example, recent work by Mahdisoltani et al. [31] does not predict the location of block defects, but only if a disk will incur a block defect. Currently, drive scrubbing and reactive fault tolerance are used to cope with block defects. Thus, this paper only considers the prediction of operational failures (and not the prediction of block defects).

In a proactive cloud storage system, a drive failure prediction model runs in the background and monitors the drives in real time (with a minor resource cost [14], [16]), periodically outputting their health states (such as once an hour). When an alarm is raised by the model, the data on an at-risk drive is ordinarily migrated (or backed up) to healthy drives immediately.

There are two prediction deployment schemes: (a) intra-drive prediction, in which the backups are dealt with locally by the drive; and (b) intra-system prediction, in which backups are dealt with by the system. The latter has greater flexibility, so we study proactive cloud storage systems with intra-system prediction. Since failure prediction eliminates most drive failures, it reduces the rate of data loss events.

In this paper, we use a simple model for proactive fault tolerance, where a proportion of drives that are about to fail are predicted to fail in advance; the proportion is called the *failure detection rate* (FDR). Rebuild processes for drives that are predicted to fail are assumed to be completed before failure actually occurs.

### 2.4 Reactive Fault Tolerance

In a real-world setting, some drives will inevitably incur operational failure in proactive cloud storage systems since the FDR will not be 100 percent and it takes time to migrate data. It is therefore necessary for proactive cloud storage systems to also use reactive fault tolerance to ensure reliability. In this paper, we include four different reactive fault tolerance methods: RAID-5, RAID-6, 2-way replication, and 3-way replication.

In the presence of reactive fault tolerance, data loss requires simultaneous operational failures and/or block defects. Moreover, these failures will not result in data loss if rebuilding has finished and/or scrubbing has taken place.

RAID-5 and RAID-6 are popular RAID schemes, in which hard drives are arranged in *RAID groups* of $g$ drives. Data *stripes* are spread across multiple drives and are accessed in parallel. Each stripe of a RAID-5 group can tolerate a single failure—an operational failure or block defect—but data loss may occur as the result of simultaneous failures. Each stripe of a RAID-6 group can tolerate any two failures.

In a replication cloud storage system, each data block has a certain number of replicas, and the replicas are dispersed over different nodes to improve the probability of blocks available when multiple nodes fail concurrently. Replication systems have two storage properties: no two replicas of a data block are stored on the same node; and replicas of a data block must be found on at least two racks.

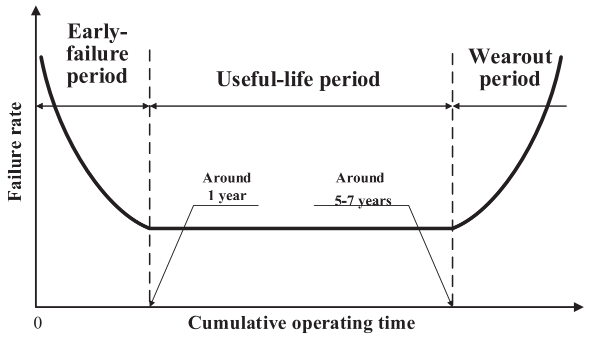



Fig. 1. Depicting the typical “bathtub curve” behavior of hard-drive failure rates, deduced in [34] from drives in the field.

### 2.5 Weibull Distribution

In a real-world setting, drive failure rates typically follow a “bathtub curve” with high failure rates at the beginning and the end of a drive’s life-cycle [32]. Fig. 1 depicts the failure rate for a hard drive’s life-cycle [32], [33], [34]: after the initial infant mortality, the failure rate enters in low-risk state and starts to wear out after 5 to 7 years. Schroeder and Gibson [20, Fig. 8] subsequently found the Weibull distribution is a suitable fit for the empirical cumulative distribution function of time between drive replacements observed in a field-gathered dataset, while the exponential distribution provides a poorer fit.

The time required for restoring data after an operational failure (or scrubbing an entire drive) depends on the drive’s capacity and the data transfer rate of the drive (or the rate of media scrubbing). To improve the reliability of cloud storage systems while maintaining a quality user service, the data transfer rates for repairing (and scrubbing rates) are adjusted according to the foreground activity, with high rates when idle and low rates when busy [23]. This was verified by Elerath and Schindler [23, Figs. 3, 4, 5, Tab. III], who found the Weibull distribution is also a suitable match for time-to-failure, time-to-repair, and scrubbing-time distributions derived from field data.

Therefore, we use the 2-parameter Weibull distribution to model occurrences of failures, rebuilds, and scrubbing in a cloud storage system. For parameters $\alpha$ and $\beta$ in the Weibull distribution, the probability density function $f$, cumulative density function $F$, hazard rate $h$, and cumulative hazard rate $H$ are given as follows:

$$f(t) = \beta \frac{t^{\beta-1}}{\alpha^{\beta}} \exp(-(t/\alpha)^{\beta}), \tag{1}$$

$$\begin{aligned} F(t) &= 1 - \exp(-(t/\alpha)^{\beta}), \\ h(t) &= \beta \frac{t^{\beta-1}}{\alpha^{\beta}}, \text{ and} \\ H(t) &= \frac{t^{\beta}}{\alpha^{\beta}}. \end{aligned} \tag{2}$$

(See e.g., [35, Sec. 3.1.1].) The parameter $\alpha$ is the *scale parameter* denoting the characteristic life, and $\beta$ is the *shape parameter* controlling the shape of the distribution.

Weibull distributions are used to express various time-dependent distributions with increasing, decreasing, or constant occurrence rates. If $\beta > 1$, the hazard rate $h$ increases over time, i.e., the probability of an operational failure increases, simulating an aging cloud storage system. With $0 < \beta < 1$, we model a system with “infant mortality”, and with $\beta = 1$ we have the traditional exponential distribution. The parameter $\alpha$ gives the characteristic life of drives. Younger drives may have a decreasing failure rate, while the older drives may have increasing failure rates.

We compare models using the cumulative hazard rate, i.e., the expected number of failure events from time 0. For a proactive cloud storage system, however, we assume that all drives that are classified as at risk of failure have sufficient *time in advance*, i.e., the time between warning and actual failure. Thus, for proactive cloud storage systems, we scale the cumulative hazard rate accordingly:

$$\hat{H}(t) := (1 - \mathrm{FDR}) \frac{t^{\beta}}{\alpha^{\beta}}, \tag{3}$$

where FDR is the *failure detection rate*. The function $\hat{H}$ instead gives the expected number of failure events from time 0 which are not predicted in advance.

A drive’s *availability* is the proportion of time it can provide service to its users. The availability of a drive in the presence of operational failures, can be estimated using [23], [36]:

$$A_{\mathrm{op}}(t) := \frac{\alpha_p(t)}{\alpha_p(t) + \mathrm{MTTR}},$$

where MTTR denotes the *mean time to repair* a drive after an operational failure, and $\alpha_p$ is the drive *pseudo-characteristic life*, modeled by

$$\alpha_p(t) = \frac{\alpha^{\beta}}{t^{\beta-1}},$$

where $\alpha$ and $\beta$ are the parameters of the Weibull distribution for operational failures, and $t$ is time.

For proactive cloud storage systems, we adjust this to account for operational failure prediction:

$$\hat{A}_{\mathrm{op}}(t) := \frac{\alpha_p(t)}{\alpha_p(t) + (1 - \mathrm{FDR})\,\mathrm{MTTR}}, \tag{4}$$

as only unpredicted operational failures need to be repaired. Since the system has intra-system prediction, proactive fault tolerance does not affect the pseudo-characteristic life of drives.

We also use a 2-parameter Weibull distribution to model occurrences of block defects, restorations, and scrubbing in a cloud storage systems (block defects are not predicted in advance). The steady-state availability of drives in the presence of block defects is modeled in [36] by

$$A_{\mathrm{def}} := \frac{\mathrm{MTTB}}{\mathrm{MTTB} + \mathrm{MTTS}}, \tag{5}$$

where MTTB denotes the *mean time to block defect*, and MTTS denotes the *mean time to scrubbing*, i.e., the mean time between drive scrubbings.

For the Weibull distribution, Eq. (1), we have

$$\mathrm{MTTR} = \alpha\,\Gamma(1 + 1/\beta), \tag{6}$$

TABLE 2
Parameters of the Weibull Distributions [23]

| | | Drive A SATA | Drive B SATA | Drive C FC/SCSI |
|---|---|---|---|---|
| Operational failure | $\alpha_f$ | 302,016 | 4,833,522 | 1,058,364 |
| | $\beta_f$ | 1.13 | 0.576 | 0.721 |
| Latent block defect | $\alpha_l$ | 12,325 | 42,857 | 50,254 |
| | $\beta_l$ | 1 | 1 | 1 |
| Rebuild time | $\alpha_r$ | 22.7 | 20.25 | 6.75 |
| | $\beta_r$ | 1.65 | 1.15 | 1.4 |
| Scrubbing time | $\alpha_s$ | 186 | 160 | 124 |
| | $\beta_s$ | 1 | 0.97 | 2.1 |

where $\Gamma$ is the gamma function, so

$$\Gamma(1 + 1/\beta) = \int_0^\infty x^{1/\beta} e^{-x}\, dx.$$

The values of MTTB and MTTS are also calculated as per Eq. (6), with appropriate parameters (listed in Table 2 for our experiments).

## 3 Mathematical Models

In this section, we derive equations for the expected number of data loss events within a period of time, for RAID-5, RAID-6, 2-way replication, and 3-way replication systems, each with proactive fault tolerance. The equations are listed in Table 1.

### 3.1 Proactive RAID Systems

If an operational failure or block defect occurs for a RAID group, the group is said to be in *degraded mode*. Degraded mode triggered by an operational failure initiates a rebuild process to repair the failure, whereas degraded mode as a result of a block defect will be resolved in the course of scrubbing.

Consistent with [22], [23], we consider a sequence of failures as negligible if (a) it requires two or more block defects, as it is improbable that two block defects simultaneously occur which affect the same data on two separate drives (with the number of blocks per drive typically in the tens of thousands), and (b) if the only surviving copy of a block is lost through a block defect, as the time between block defects (usually thousands or even tens of thousands of hours) is much larger than the rebuild time (typically tens of hours).

The expected number of data loss events in time period $t$ in a stripe is thus modeled by:

$$\Pr\left(\begin{matrix}\text{being at risk of data}\\ \text{loss by op. failure}\end{matrix}\right) \times \left(\begin{matrix}\text{no. ways in which an}\\ \text{op. failure could result}\\ \text{in data loss}\end{matrix}\right) \times \hat{H}(t),$$

where $\hat{H}(t)$ is the adjusted cumulative hazard rate given in Eq. (3).

Modifying [22, Eq. 5], we estimate the expected number of data loss events in a $g$-drive RAID-5 group within a time period $t$ as

$$N_{\mathrm{R5}}(t) := (R_{\mathrm{op}} + R_{\mathrm{def}})\,(g-1)\,\hat{H}(t), \qquad (7)$$

where $R_{\mathrm{op}} = 1 - \hat{A}^g_{\mathrm{op}}$ is the probability of the group having at least one operational failure and $R_{\mathrm{def}} = 1 - A^g_{\mathrm{def}}$ is the probability of the group having at least one block defect. Here, $\hat{A}_{\mathrm{op}}$ is the adjusted availability of a drive in the presence of operational failures given in Eq. (4), and $\hat{A}^g_{\mathrm{op}}$ is the probability of all the $g$ drives having no operational failure at a given time. After a drive fails, the remaining $g-1$ drives are subject to operational failure, thereby giving Eq. (7).

Again modifying [23, Eq. 5], we estimate the expected number of data loss events in a $g$-drive RAID-6 group within a time period $t$ as

$$N_{\mathrm{R6}}(t) := (R_{\text{op-op}} + R_{\text{op-def}})\,(g-2)\,\hat{H}(t), \qquad (8)$$

where $R_{\text{op-op}}$ is the probability of the group having at least two operational failures, so

$$\begin{aligned} R_{\text{op-op}} &= \Pr(\text{at least 2 op. fails})\\ &= 1 - \Pr(\text{no op. fail}) - \Pr(\text{exactly 1 op. fail})\\ &= 1 - \hat{A}^g_{\mathrm{op}} - g\,\hat{A}^{g-1}_{\mathrm{op}}\,(1 - \hat{A}_{\mathrm{op}}) \end{aligned}$$

and $R_{\text{op-def}}$ is the probability of the group having at least one operational failure and one block defect on two distinct drives, so

$$\begin{aligned} R_{\text{op-def}} &= \Pr\left(\begin{matrix}\text{at least 1 op. fail and 1 block}\\ \text{defect on two distinct drives}\end{matrix}\right)\\ &\simeq 1 - \Pr(\text{no op. fail}) - \Pr(\text{no block defect})\\ &\quad + \Pr(\text{no op. fail and no block defect})\\ &= 1 - \hat{A}^g_{\mathrm{op}} - A^g_{\mathrm{def}} + (\hat{A}_{\mathrm{op}}\,A_{\mathrm{def}})^g. \end{aligned}$$

For proactive RAID systems, we use the adjusted drive availability $\hat{A}_{\mathrm{op}}$ and cumulative hazard rate $\hat{H}(t)$ to account for operational failure prediction in Eqs. (7) and (8), rather than $A_{\mathrm{op}}$ and $H(t)$ in [22, Eq. 5] and [23, Eq. 5], respectively. When the failure detection rate $\mathrm{FDR} = 0$, Eqs. (7) and (8) become [22, Eq. 5] and [23, Eq. 5], respectively, as in reactive RAID systems.

### 3.2 Proactive Replication Systems

In a replication system, we use $r$ to denote the number of racks, $n$ to denote the number of nodes in each rack, $d$ to denote the number of drives in each node, and $b$ to denote the number of blocks on each drive. There are thus $rnd$ drives in total.

After an operational failure or block defect, we say the system enters *degraded mode* during the rebuild or scrubbing process. Note, this definition of "degraded mode" is distinct from that for RAID groups: for replication, we consider all of the $rnd$ drives in the system (whereas for RAID, we consider those within a group separately). As such, we no longer consider *if* the system has block defects, but *how many* block defects it has.

#### 3.2.1 Proactive 2-Way Replication

In a 2-way replication system, every data block has two copies, and the two copies must be stored on two separate racks. We call such an arrangement a *replica pair*. Given a drive, we can choose any of the $(r-1)nd$ drives on different racks to extend it to a replica pair.

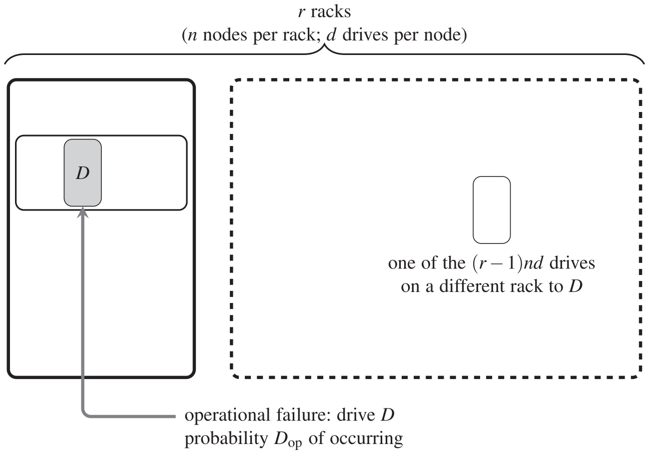



Fig. 2. 2-way replication. We have an operational failure for some drive $D$ (i.e., degraded mode due to an operational failure). We lose data with probability $P_{2\text{-way}}$ if an operational failure occurs on any drive on another rack to $D$.

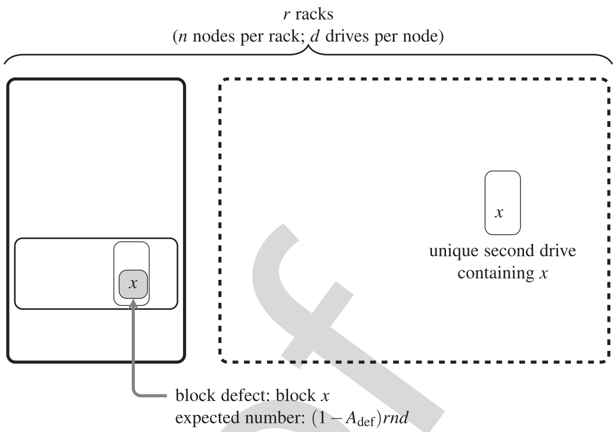



Fig. 3. 2-way replication. We have a block defect for block $x$. We will lose $x$ if an operational failure occurs to the unique second drive containing $x$.

Suppose both drives in a replica pair incur operational failures, and drive $D$ is one of them. The probability that a data block $x$ on $D$ is also stored on the other failed drive is

$$p = \frac{1}{(r-1)\,nd},$$

assuming duplicate blocks are stored in random replica pairs.

The probability that no block on $D$ is lost (i.e., not stored on the other failed drive) is $(1-p)^b$, so the probability of data loss caused by these two concurrent failures is

$$P_{2\text{-way}} = 1 - \left(1-p\right)^b, \\ = 1 - \left(1 - \frac{1}{(r-1)\,nd}\right)^b. \quad (9)$$

Data loss can only occur when at least two racks simultaneously have corrupted data due to operational failures or block defects. As with RAID systems, we consider the case of two block defects erasing shared data to be negligible.

*Case (a):* Two operational failures.

Assuming the system is in degraded mode due to an operational failure of some drive on one rack, data loss occurs with probability $P_{2\text{-way}}$ when there is a second operational failure on one of the $(r-1)nd$ drives on the other racks. This situation is illustrated in Fig. 2.

*Case (b):* One operational failure and one block defect.

'Given the system has a block defect for block $x$, data loss will occur when an operational failure occurs for the unique second drive containing $x$. By definition, a drive has a block defect with probability $1 - A_{\text{def}}$, so the expected number of drives with block defects is $(1 - A_{\text{def}})rnd$. This situation is illustrated in Fig. 3.

Combining cases (a) and (b), the expected number of data loss events within a time period $t$ is thus

$$N_{2\text{-way}}(t) = \left(P_{2\text{-way}}\,(r-1)nd\,D_{\text{op}} + rnd\,(1 - A_{\text{def}})\right)\hat{H}(t), \quad (10)$$

where

$$D_{\text{op}} = 1 - \hat{A}_{\text{op}}^{rnd} \quad (11)$$

is the probability of being in degraded mode due to an operational failure.

### 3.2.2 Proactive 3-Way Replication

In 3-way replication, each data block is stored 3 times, on 3 distinct nodes, out of which 2 nodes are on the same rack and 1 node is on another rack. We call such an arrangement a *replica set*.

**Lemma 1.** *The number of replica sets is*

$$r(r-1)\binom{n}{2}nd^3.$$

*The number of replica sets containing a given drive is*

$$\frac{3}{2}(r-1)\,n(n-1)d^2.$$

*The number of replica sets containing two given drives on different nodes in the same rack is* $(r-1)nd$.

*The number of replica sets containing two given drives on different racks is* $2(n-1)d$.

**Proof.** To count the number of replica sets, we pick one of the $r$ racks, from which we choose two nodes (in one of $n2$ ways), and one of the $r-1$ other racks, from which we choose one of the $n$ nodes, and from each of the three nodes, we choose one drive (in one of $d^3$ ways). This gives the first claim in the lemma statement.

Given a drive $D$, it can belong to two types of replica sets; the number of replica sets containing $D$ and another drive in the same rack as $D$ (but different node to $D$) is

$$\overbrace{(n-1)d}^{\text{choose drive on same rack as } D} \times \overbrace{(r-1)nd}^{\text{choose drive on different rack to } D}$$

and there are $(r-1)n2d^2$ replica sets containing $D$ with the other two drives in a different rack to $D$. Summing these simplifies to give the second claim.

For the third claim, given two drives on different nodes in the same rack, we can choose any of the $(r-1)nd$ drives on different racks to extend them to a replica set.

For the fourth claim, given two drives on different racks, we can choose any of the $2(n-1)d$ drives on different nodes in those two racks to extend them to a replica set. □

Suppose all the three drives in a replica set incur operational failures, and drive $D$ is one of them. By the second claim in Lemma 1, the probability of a block on $D$ is also stored on the other two failed drives is

$$p_{\text{dup}} = \frac{2}{3(r-1)\,n(n-1)d^2}.$$

Then the probability [37] of at least one data block being lost caused by the concurrent failures is

$$\begin{aligned} P_{3\text{-way}} &= 1 - \left(1 - p_{\text{dup}}\right)^b, \\ &= 1 - \left(1 - \frac{2}{3(r-1)\,n(n-1)d^2}\right)^b. \end{aligned} \tag{12}$$

There are four situations in which data loss occurs as the result of an additional operational failure, Cases I to IV in the following.

*Case I:* Two operational failures in one rack.

For a given $d$-drive node, the probability of no drive in that node incurring an operational failure is $\hat{A}^d_{\text{op}}$. Consequently, for a given $n$-node rack, the probability of operational failures occurring on at least two drives on different nodes in that rack is

$$\begin{aligned} F_{\text{rack}} &:= 1 - \Pr(\text{no op. fails}) - \Pr\begin{pmatrix}\text{op. fails only} \\ \text{on one node}\end{pmatrix} \\ &= 1 - (\hat{A}^d_{\text{op}})^n - n(\hat{A}^d_{\text{op}})^{n-1}(1 - \hat{A}^d_{\text{op}}) \end{aligned}$$

and the probability $D^{(1)}_{\text{op-op}}$ that at least one of the $r$ racks incurs operational failures on at least two drives on different nodes in that rack is given by

$$\begin{aligned} D^{(1)}_{\text{op-op}} &= 1 - (1 - F_{\text{rack}})^n \\ &= 1 - \left((\hat{A}^d_{\text{op}})^n + n\,(\hat{A}^d_{\text{op}})^{n-1}(1 - \hat{A}^d_{\text{op}})\right)^r. \end{aligned}$$

Given two operational failures occurring on two drives on different nodes in the same rack, an operational failure on any one of the $(r-1)nd$ drives in the other racks causes data loss with probability $P_{3\text{-way}}$.

*Case II:* Two operational failures on different racks.

For a given rack, the probability of at least one drive incurring an operational failure is $1 - \hat{A}^{nd}_{\text{op}}$. The probability of at least two racks incurring an operational failure is thus

$$\begin{aligned} D^{(2)}_{\text{op-op}} &= 1 - \Pr(\text{no op. fails}) - \Pr\begin{pmatrix}\text{exactly one rack} \\ \text{with op. fails}\end{pmatrix} \\ &= 1 - \hat{A}^{rnd}_{\text{op}} - r\,(\hat{A}^{nd}_{\text{op}})^{r-1}(1 - \hat{A}^{nd}_{\text{op}}). \end{aligned}$$

Given two operational failures occurring in two different racks, an operational failure on any one of the $2(n-1)d$ drives in those racks but in distinct nodes causes data loss with probability $P_{3\text{-way}}$.

*Case III:* An operational failure and a block defect on the same rack.

This case is illustrated in Fig. 4. Given an operational failure, the expected number of drives on different nodes in the same rack with block defects is

$$\Pr(\text{block def.}) \times (\text{no. such drives}) = (1 - A_{\text{def}})(n-1)d.$$

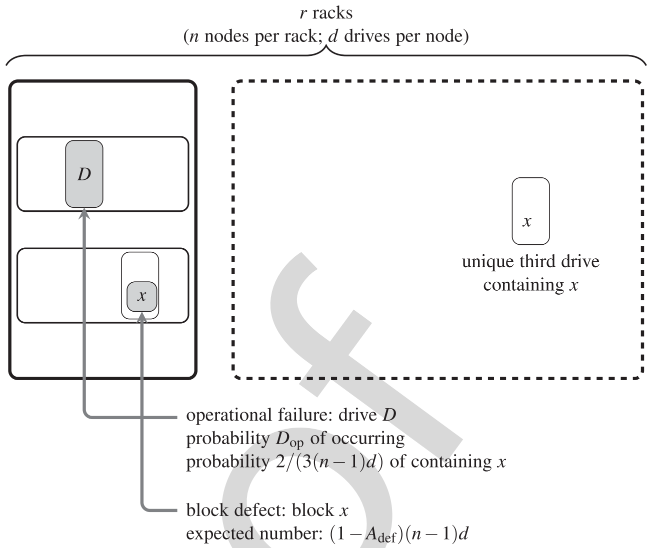



Fig. 4. Simultaneously, we have an operational failure for drive $D$ and a block defect for block $x$ (on a drive on a different node in the same rack to $D$). If $D$ happens to contain a copy of $x$, then $x$ is lost if an operational failure occurs to the unique third drive containing $x$.

Given an operational failure for some drive $D$ and a drive $X$ on a different node in the same rack with defective block $x$,

$$\begin{aligned} \Pr\begin{pmatrix}\text{a copy of } x \\ \text{occurs in } D\end{pmatrix} &= \frac{\text{no. replica sets containing } D \text{ and } X}{\text{no. replica sets containing } X} \\ &= \frac{2}{3(n-1)d} \end{aligned}$$

by Lemma 1. Thus, given an operational failure for some drive $D$, the expected number of drives on a different node in the same rack with defective block $x$ which also occurs on $D$ is

$$\frac{2}{3}(1 - A_{\text{def}}).$$

Finally, given an operational failure for some drive $D$ and a drive on a different node in the same rack with defective block $x$ which also occurs on $D$, there is a unique drive on another rack which must incur an operational failure to lose block $x$.

*Case IV:* An operational failure and a block defect on the different racks.

This case is illustrated in Fig. 5. Given an operational failure, the expected number of drives in different racks with block defects is

$$(1 - A_{\text{def}})(r-1)nd.$$

Given an operational failure for some drive $D$ and a drive $X$ on a different rack with defective block $x$,

$$\begin{aligned} \Pr\begin{pmatrix}\text{a copy of } x \\ \text{occurs in } D\end{pmatrix} &= \frac{\text{no. replica sets containing } D \text{ and } X}{\text{no. replica sets containing}} \\ &= \frac{4}{3(r-1)nd} \end{aligned}$$

by Lemma 1. Thus, given an operational failure for some drive $D$, the expected number of drives on a different rack with defective block $x$ which also occurs on $D$ is

$$\frac{4}{3}(1 - A_{\text{def}}).$$

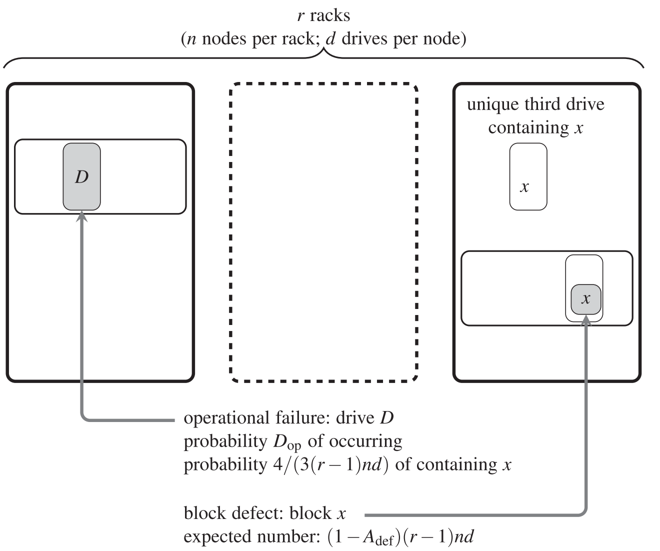



Fig. 5. Simultaneously, we have an operational failure for drive $D$ and a block defect for block $x$ (on a drive on a different rack as $D$). If $D$ happens to contain a copy of $x$, then $x$ is lost if an operational failure occurs to the unique third drive containing $x$.

Finally, given an operational failure for some drive $D$ and a drive on a different rack with defective block $x$ which also occurs on $D$, there is a unique drive on another rack which must be erased to lose block $x$.

*Combining Cases I–IV:* The expected number of data loss events within a time period $t$ is thus

$$N_{\text{3-way}}(t) \simeq \big(P_{\text{3-way}}\left((r-1)nd\, D^{(1)}_{\text{op-op}} + 2(n-1)d\, D^{(2)}_{\text{op-op}}\right) + 2\, D_{\text{op}}(1-A_{\text{def}})\big)\, \hat{H}(t), \tag{13}$$

where $D_{\text{op}}$ is given by Eq. (11).

## 4 Evaluation

In this section, we compare the predictions of the equations in Table 1 to simulations and give an analysis of the sensitivity of the equations.

### 4.1 Equation Verification

#### 4.1.1 Experimental Setup

To test the accuracy of the equations in Table 1, we compare them against Monte-Carlo simulations. We design event-driven Monte-Carlo simulations, in which there are six types of events that drive the virtual time forward:

(a) Weibull distributed *operational failure events*, with parameters $\alpha_f$ and $\beta_f$, potentially occurring on each drive;
(b) Weibull distributed *block defect events*, with parameters $\alpha_l$ and $\beta_l$, potentially occurring on each drive;
(c) Weibull distributed *failure-rebuild complete events*, with $\alpha_r$ and $\beta_r$ denoting recovery time, occurring after an operational failure;
(d) Weibull distributed *scrubbing complete events*, with $\alpha_s$ and $\beta_s$ denoting scrubbing time, periodically occurring on each drive to eliminate simulated block defects on it;
(e) *warning events*, occurring 300 hours before operational failure for a proportion of drives determined by the failure detection rate FDR; and
(f) Weibull distributed *warning-rebuild complete events*, with $\alpha_r$ and $\beta_r$ denoting pre-warning recovery time (chosen to be the same as for failure-rebuild complete events, for simplicity), occurring after a warning event.

Events (e) and (f) simulate proactive fault tolerance. Events (c) and (f) trigger the introduction of a new drive (with future failure and block defects). After (b) completes, a future block defect event is added to the drive. When events of type (a) and/or (b) simultaneously occur, under appropriate conditions (depending on the system) we incur a data loss event, and new data is added to maintain system scale.

We enumerate the number of data loss events over a 5-year time period, and compare the results to the equations' predictions. A 5-year simulation is repeated until the total number of data loss events is 10 or more, and we average the results.

Our choice of time in advance $\text{TIA} = 300$ hours is motivated by [11], where the classification tree prediction model predicted over 95 percent of failures with the TIA around 360 hours on a real-world dataset. For the four events (a)–(d), we use the parameters in [23], listed in Table 2, from three representative drive models, which had been in the field for several years.

For RAID systems, we set the number of drives in each group $g = 15$ for RAID-5 and $g = 16$ for RAID-6, so each RAID group has 14 data drives. We choose 400 RAID groups to model a deployment typical for a single RAID system (5,600 data drives). For 2-way replication, we set the number of racks $r = 200$, the number of nodes in each rack $n = 14$, the number of drives in each node $d = 4$, and the number of blocks in each drive $b = 10^7$, and for 3-way replication we set $(r, n, d, b) = (300, 14, 4, 10^7)$. In this way, all four systems store the same amount of user data.

#### 4.1.2 Accuracy with Failure Prediction

Fig. 6 plots the expected number of data loss events predicted by the equations in Table 1 and enumerated by simulation as the failure detection rate varies from 0 to 0.95. We include proactive RAID and replication systems for drives $A$, $B$, and $C$. For all the drive models, the equation-based results of every system closely match the simulation-based values; in the average case, they disagree by around 10 percent (Elerath and Schindler [23] found the difference between equation and simulation of around 20 percent).

A reliability analysis is often used to assess trade-offs, to compare schemes, and to estimate the effect of several parameters on storage system reliability. In this setting, a 10 percent error would not significantly influence the analysis, especially considering the Monte-Carlo simulations themselves are also approximations. In a situation where Monte-Carlo simulations are required, the given equations could be used e.g., to quickly reject inferior parameter combinations, after which we can use Monte-Carlo simulations on the remaining cases.

Ordinarily, the equations and simulations agree closely, but for high FDR, they begin to disagree in some cases, e.g., by a factor of 1.9 for the 3-way replication system with $\text{FDR} = 0.95$ on drive C. However, this occurs when the expected number of data loss events is low.

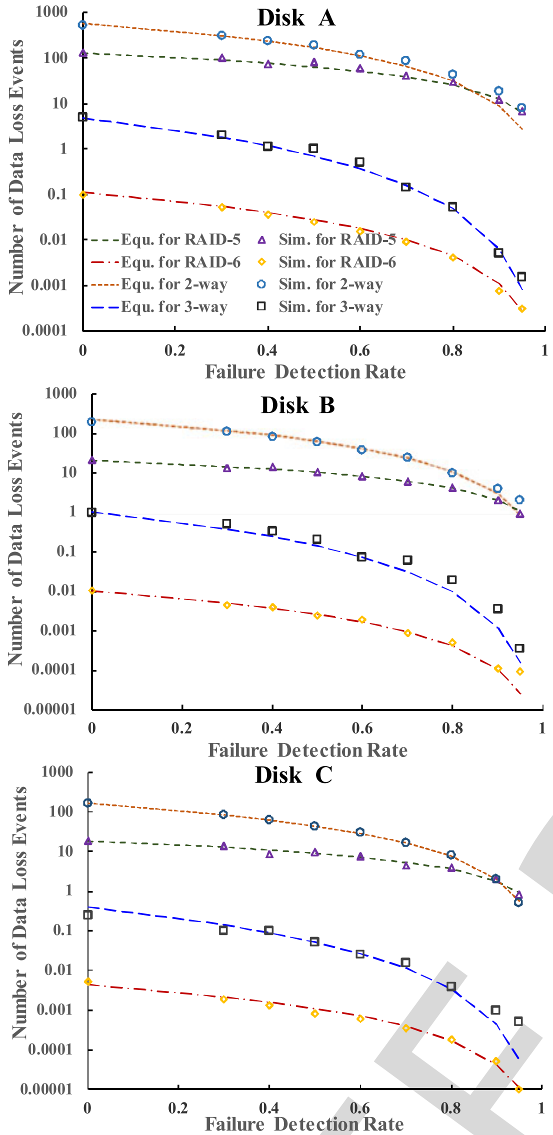



Fig. 6. Number of data loss events predicted by the mathematical model and enumerated during simulation as the failure detection rate (FDR) varies, for drives $A$, $B$, and $C$, respectively.

### 4.1.3 Accuracy as Time Varies

Fig. 7 plots the equational and simulated number of data loss events over a $t$ year period, as $t$ varies from 1 year to 10 years. Here we use drive A and set $\mathrm{FDR} = 0.8$. We find that the equational results closely match the simulated results. In the worst case, the equation and simulation disagree by around 30 percent. In the average case, they disagree by around 10 percent.

### 4.1.4 Accuracy as System Scale Varies

Fig. 8 plots the equational and simulated number of data loss events over a 5-year period, as the effective storage space (i.e., the amount of user data, excluding redundancy) varies. We vary the storage space via the number of groups in RAID systems and the number of racks in replication systems. Here we continue to use drive A and set $\mathrm{FDR} = 0.8$.

We again find that the equational results closely match the simulated results. In the worst case, the equation and simulation disagree by around 30 percent. In the average case, they disagree by around 10 percent. The experiment

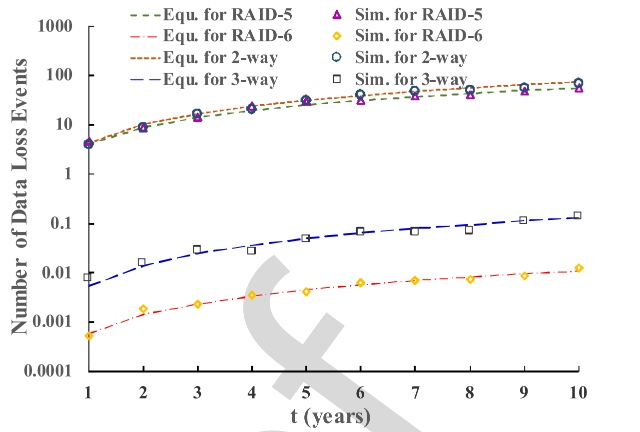



Fig. 7. The expected number of data loss events over $t$ years, for drive A and $\mathrm{FDR} = 0.8$.

results verify the effectiveness of the reliability equations on cloud storage systems with various scales.

Moreover, the equations yield comparable results much faster than the simulations. On a standard PC desktop, with e.g., MATLAB, we can quickly calculate the equational results (in approximately 1ms), while the simulations usually take between tens of seconds and tens of hours (even hundreds of hours for a system with high FDR, where the expected number of data loss events is very low) to produce the results for a single set of inputs.

If the failure statistics for solid-state storage systems could be obtained (i.e., the inputs in Table 1), the proposed equations could be used for solid-state storage systems.

## 4.2 Sensitivity Analysis

In this section, we illustrate how the equations can be used to analyze system sensitivity to varying system parameters. We compare RAID-6 and 3-way replication, which are the most common redundancy schemes. Unless otherwise stated, we use drive A's parameters and set $t = 5$ years.

### 4.2.1 Sensitivity to Drive Model

While both drives A and B are near-line SATA models, they have the different failure distributions. In particular, drive A has higher operational failure and block defect rates than drive B. Fig. 9 plots the expected number of data loss events for these two drives.

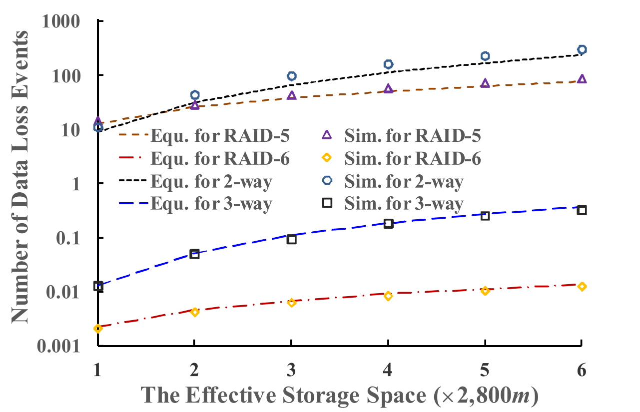



Fig. 8. The expected number of data loss events as the effective storage space varies, for drive A and $\mathrm{FDR} = 0.8$. The storage capacity of a drive is denoted $m$.

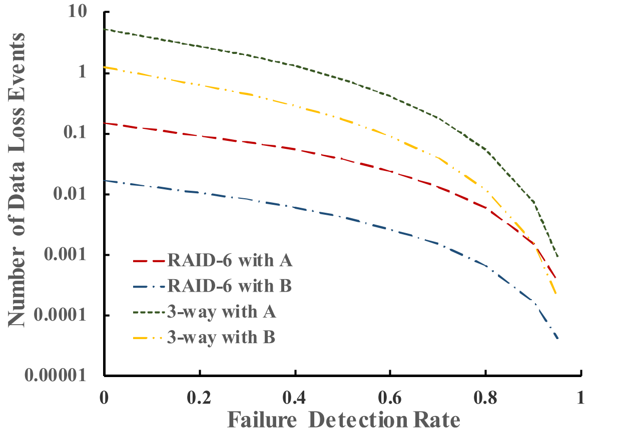



Fig. 9. The expected number of data loss events of drives A and B, with RAID-6 or 3-way replication (rebuild parameters: $\alpha_r = 24$, $\beta_r = 2$; scrubbing parameters: $\alpha_s = 240$, $\beta_s = 1$).

In terms of data loss events, a designer could learn from this which drive performs better (drive B, in this case) and which storage scheme performs better (RAID-6, in this case). Fig. 9 also shows the change in performance as FDR changes, and the significant difference between proactive and purely reactive fault tolerance. We see a stronger sensitivity to FDR in 3-way replication and RAID-6 systems, than in 2-way replication and RAID-5 systems.

#### 4.2.2 Sensitivity to Weibull Parameters

The reliability of a system is affected by the rebuild time, as concurrent operational failures are more likely to happen with a longer rebuilding process.

Fig. 10 plots the expected number of data loss events for proactive RAID-6 and 3-way replication systems as the Weibull parameters vary. We vary the rebuild time via $\alpha_r$, the time for latent block defects via $\alpha_l$, and the scrubbing time via $\alpha_s$. All other parameters are those for drive A in Table 2.

We see that rebuild time plays a significant role in the expected number of data loss events, and that 3-way replication is more sensitive to the rebuild time than RAID-6. This arises as RAID-6 is far more sensitive to block defects (and hence $A_{\text{def}}$) than 3-way replication, which is apparent from the model: for RAID-6, the expected number of data loss events scales linearly with $A^g_{\text{def}}$, whereas for 3-way replication, the expected number of data loss events scales linearly with $A_{\text{def}}$, and with MTTB in the order of years and MTTS in the order of hours, we have $A_{\text{def}}$ close to 1.

Fig. 10 shows MTTB and MTTS has a negligible role for 3-way replication, but not for RAID-6. This is consistent with the models: in 3-way replication, two operational failures can result in the loss of a defective block $x$ in only one way, but in RAID-6, two operational failures can result in the loss of block $x$ in $\binom{g-1}{2}$ ways.

#### 4.2.3 Sensitivity to System Scale Parameters

There are several other sensitivity properties that concern system designers and reliability engineers, such as the effects of system scale parameters on the overall cloud storage system reliability. In this subsection, we investigate the sensitivity of a system's reliability to some system scale parameters (including the number of data drives per RAID group, rack

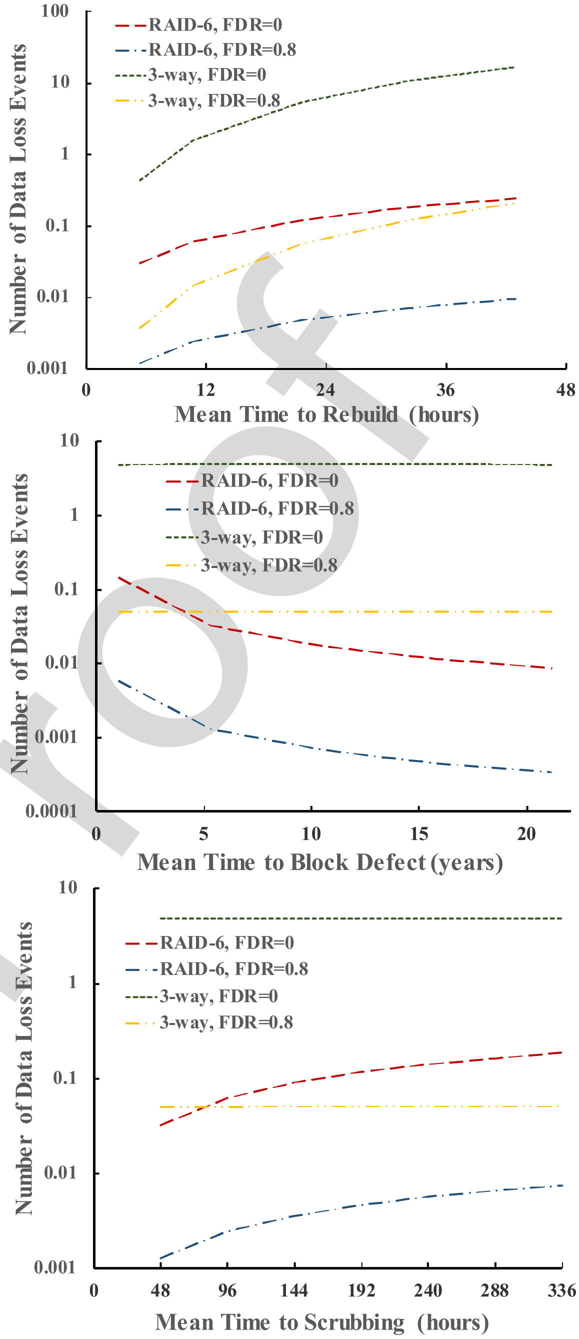



Fig. 10. The expected number of data loss events for drive A as the mean time to rebuild, mean time to block defect, and mean time to scrubbing changes. We include proactive fault tolerance with $\text{FDR} = 0.8$ and reactive fault tolerance with $\text{FDR} = 0$.

size, node size, and drive capacity). Unless otherwise stated, we use the system parameters described in Section 4.1.1.

Fig. 11 plots the expected number of data loss events for RAID-6 and 3-way replication systems, as the number of data drives per group (i.e., $g - 2$), and the number of nodes per rack (i.e., rack size $n$) varies. We see that the number of data drives in RAID-6 group and the rack size have a significant impact on RAID-6 and 3-way replication systems reliability, respectively.

Fig. 12 plots the expected number of data loss events for 3-way replication systems as the number of drives per node $d$ changes, and as the number of blocks per drive $b$ changes.

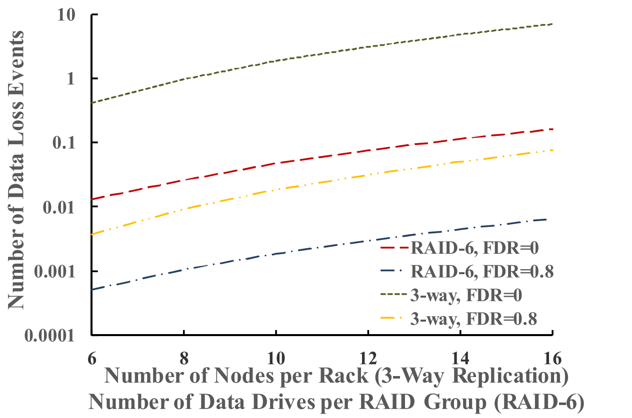



Fig. 11. The expected number of data loss events of drive A, as the number of data drives in RAID-6 group, and the rack sizes in 3-way replication system vary. We include proactive fault tolerance with $\mathrm{FDR} = 0.8$ and reactive fault tolerance with $\mathrm{FDR} = 0$.

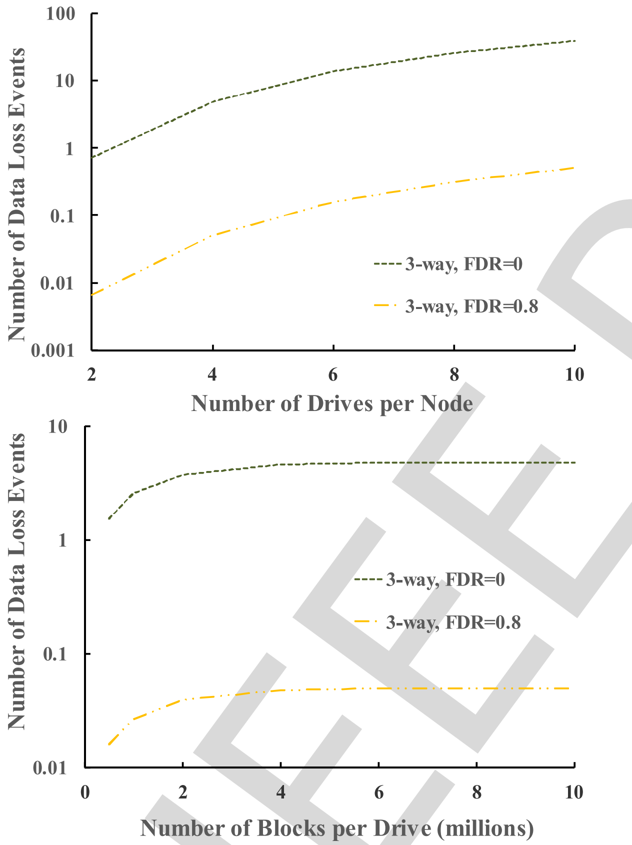



Fig. 12. The expected number of data loss events for 3-way replication systems with drive A as the node size (top) and drive size (bottom) change. We include proactive fault tolerance with $\mathrm{FDR} = 0.8$ and reactive fault tolerance with $\mathrm{FDR} = 0$.

We see that the number of drives per node (i.e., the node size $d$) has a significant impact on 3-way replication systems reliability. We see that the number of data loss events increases as $b$ increases.

As per the analysis in Section 3.2.2, provided each replica set shares at least one data block, data loss occurs when the three drives in a replica set simultaneously fail, and otherwise occurs with probability $P_{\text{3-way}}$ given in Eq. (12). Consequently, the larger the value of $b$, the higher the probability of a data loss event through simultaneous failures, but when the $b$ is greater than a certain value (depending on system size), the probability is approximately 1, consistent with Eq. (12).

## 5 Conclusion and Future Work

In this paper, we present four equations for proactive RAID and replication cloud storage systems, by which one can predict the overall reliability of systems in the presence of operational drive failures and latent block defects. The equations also incorporate time-variant failure rates and media scrubbing processes. We use Weibull distributions for failure rates, which is now considered more realistic than the simpler exponential distribution.

We indicate the usefulness of these equations by investigating the impact of proactive fault tolerance and the system parameters on reliability. While simulations need specialized code and take much longer, they give comparable results to the equations. As such, the equations can help designers to readily explore the design space for their system:

(a) To ensure availability, a designer desires to minimize rebuild and warning migration bandwidth, but this influences the rebuild time, which impacts successfully protecting at-risk data, which will negatively affect the reliability against data loss. The equations can aid the designer in optimizing this trade-off.
(b) In general, a high failure detection rate incurs a high false alarm rate (FAR), resulting in unnecessary processing costs. The equations can help designers choose a failure predictor with an FDR to achieve a specific level of reliability, while minimizing FAR.
(c) The reliability of systems is also significantly affected by drive model (see Fig. 9). The equations can thus help e.g., to decide whether to construct a system with less reliable, but cheaper drives.
(d) The intermediate results in the mathematical model are meaningful, which may assist an operator pinpoint how data loss occurs in a storage system.

There are some limitations to the mathematical models we present:

- The models have a static mean rebuild time, which may not accurately reflect fluctuation due to system usage (which might vary according to the time of the day) and system utilization (how much data is stored on each drive).
- The assumptions break down when FDR approaches 1. With $\mathrm{FDR} = 1$, there are no operational failures, and the models will predict no data loss events. Realistically, even with $\mathrm{FDR} = 1$, data loss events will still occur in cases where there is insufficient time in advance, and as a result of concurrent block defects. We see this behavior in Fig. 6.

In practice, cloud storage systems at petabyte or exabyte scales are both dynamic and heterogeneous, since new drives will continuously enter the system as old ones leave due to failure or age. Moreover, correlated failures (e.g., node failures, or simultaneous block defects as a result of a scratch) will also occur in system, which may influence the system's reliability. Therefore, in future work, we plan to: (a) modify the equations to suit other cloud storage systems, such as deduplication and heterogeneous systems, to model their reliability; and (b) extend the models to incorporate correlated failures.

## Acknowledgments

This work is partially supported by the NSF of China (grant number: 61702521), NSF of Tianjin (grant number: 4117JCYBJC15300), the Scientific Research Foundation of Civil Aviation University of China (grant number: 2017QD03S), and the Fundamental Research Funds for the Central Universities. Stones also supported by the Thousand Youth Talents Plan in Tianjin.

## References

[1] I. Corderí, T. Schwarz, A. Amer, and D. D. E. Long, "Self-adjusting two-failure tolerant disk arrays," in *Proc. Petascale Data Storage Workshop*, 2012, pp. 1–5.

[2] T. Schwarz, A. Amer, T. Kroeger, E. Miller, D. Long, and J. F. Paris, "RESAR: Reliable storage at exabyte scale," in *Proc. Model. Anal. Simul. Comput. Telecommun. Syst.*, 2016, pp. 211–220.

[3] Q. Xin, E. L. Miller, S. J. Schwarz, and J. E. Thomas, "Evaluation of distributed recovery in large-scale storage systems," in *Proc. High Perform. Distrib. Comput.*, 2004, pp. 172–181.

[4] S. Mitra, R. K. Panta, M.-R. Ra, and S. Bagchi, "Partial-parallel-repair (PPR): A distributed technique for repairing erasure coded storage," in *Proc. EuroSys*, 2016, pp. 1–14.

[5] V. Gramoli, G. Jourjon, and O. Mehani, "Disaster-tolerant storage with SDN," in *Proc. Int. Conf. Networked Syst.*, 2015, pp. 278–292.

[6] V. Estrada-Galiñanes, J. F. Pâris, and P. Felber, "Simple data entanglement layouts with high reliability," in *Proc. Perform. Comput. Commun. Conf.*, 2017, pp. 1–8.

[7] J. F. Murray, G. F. Hughes, and K. Kreutz-Delgado, "Machine learning methods for predicting failures in hard drives: A multiple-instance application," *J. Mach. Learn. Res.*, vol. 6, pp. 783–816, 2005.

[8] G. F. Hughes, J. F. Murray, K. Kreutz-Delgado, and C. Elkan, "Improved disk-drive failure warnings," *IEEE Trans. Rel.*, vol. 51, no. 3, pp. 350–357, Sep. 2002.

[9] J. F. Murray, G. F. Hughes, and K. KreutzDelgado, "Hard drive failure prediction using non-parametric statistical methods," in *Proc. Artif. Neural Netw.*, 2003, pp. 1–4.

[10] Y. Zhao, X. Liu, S. Gan, and W. Zheng, "Predicting disk failures with HMM- and HSMM-based approaches," in *Proc. Adv. Data Mining: Appl. Theoretical Aspects*, 2010, pp. 390–404.

[11] J. Li, X. Ji, Y. Jia, B. Zhu, G. Wang, Z. Li, and X. Liu, "Hard drive failure prediction using classification and regression trees," in *Proc. Dependable Syst. Netw.*, 2014, pp. 383–394.

[12] A. Ma, F. Douglis, G. Lu, D. Sawyer, S. Chandra, and W. Hsu, "RAIDShield: Characterizing, monitoring, and proactively protecting against disk failures," in *Proc. USENIX File Storage Technol.*, 2015, pp. 241–256.

[13] S. Wu, H. Jiang, and B. Mao, "Proactive data migration for improved storage availability in large-scale data centers," *IEEE Trans. Comput.*, vol. 64, no. 8, pp. 2637–2651, Sep. 2015.

[14] C. Xu, G. Wang, Z. Li, and X. Liu, "Health status and failure prediction for hard drives with recurrent neural networks," *IEEE Trans. Comput.*, vol. 65, no. 11, pp. 3502–3508, Nov. 2016

[15] S. Pang, Y. Jia, R. J. Stones, X. Liu, and G. Wang, "A combined Bayesian network method for predicting drive failure times from SMART attributes," in *Proc. Int. Joint Conf. Neural Netw.*, 2016, pp. 4850–4856.

[16] J. Li, R. J. Stones, G. Wang, Z. Li, X. Liu, and X. Kang, "Being accurate is not enough: New metrics for disk failure prediction," in *Proc. Symp. Reliable Distrib. Syst.*, 2016, pp. 71–80.

[17] J. Li, R. J. Stones, G. Wang, X. Liu, Z. Li, and X. Ming, "Hard drive failure prediction using decision trees," *Rel. Eng. Syst. Safety*, vol. 164, pp. 55–65, 2017.

[18] B. Eckart, X. Chen, X. He, and S. L. Scott, "Failure prediction models for proactive fault tolerance within storage systems," in *Proc. Model. Anal. Simul. Comput. Telecommun. Syst.*, 2008, pp. 1–8.

[19] J. Li, M. Li, G. Wang, X. Liu, Z. Li, and H. Tang, "Global reliability evaluation for cloud storage systems with proactive fault tolerance," in *Proc. Algorithms Archit. Parallel Process.*, 2015, pp. 189–203.

[20] B. Schroeder and G. A. Gibson, "Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?" in *Proc. USENIX File Storage Technol.*, 2007, pp. 1–16.

[21] E. Pinheiro, W.-D. Weber, and L. A. Barroso, "Failure trends in a large disk drive population," in *Proc. USENIX File Storage Technol.*, 2007, pp. 17–29.

[22] J. G. Elerath, "A simple equation for estimating reliability of an $N+1$ redundant array of independent disks (RAID)," in *Proc. Dependable Syst. Netw.*, 2009, pp. 484–493.

[23] J. G. Elerath and J. Schindler, "Beyond MTTDL: A closed-form RAID-6 reliability equation," *ACM Trans. Storage*, vol. 10, no. 2, 2014, Art. no. 7.

[24] A. Dholakia, E. Eleftheriou, X. Y. Hu, I. Iliadis, J. Menon, and K. K. Rao, "A new intra-disk redundancy scheme for high-reliability RAID storage systems in the presence of unrecoverable errors," *ACM Trans. Storage*, vol. 4, pp. 1–42, 2008.

[25] K. M. Greenan, E. L. Miller, and J. J. Wylie, "Reliability of flat XOR-based erasure codes on heterogeneous devices," in *Proc. Dependable Syst. Netw.*, 2008, pp. 147–156.

[26] G. A. Gibson, *Redundant Disk Arrays: Reliable, Parallel Secondary Storage*, vol. 368. Cambridge, MA, USA: MIT Press, 1992.

[27] W. Dweik, M. A. Majeed, and M. Annavaram, "Warped-Shield: Tolerating hard faults in GPGPUs," in *Proc. Dependable Syst. Netw.*, 2014, pp. 431–442.

[28] K. M. Greenan, J. S. Plank, and J. J. Wylie, "Mean time to meaningless: MTTDL, Markov models, and storage system reliability," in *Proc. USENIX Hot Topics Storage File Syst.*, 2010, pp. 1–5.

[29] I. Iliadis and V. Venkatesan, "Rebuttal to "Beyond MTTDL: A closed-form RAID-6 reliability equation"," *ACM Trans. Storage*, vol. 11, no. 2, 2015, Art. no. 10.

[30] B. Allen, "Monitoring hard disks with SMART," *Linux J.*, vol. 1, no. 117, pp. 74–77, 2004.

[31] F. Mahdisoltani, I. Stefanovici, and B. Schroeder, "Proactive error prediction to improve storage system reliability," in *Proc. USENIX Annu. Tech. Conf.*, 2017, pp. 391–402.

[32] J. G. Elerath, "Specifying reliability in the disk drive industry: No more mtbf's," in *Proc. Rel. Maintainability Symp.*, 2000, pp. 194–199.

[33] J. G. Elerath, "AFR: Problems of definition, calculation and measurement in a commercial environment," in *Proc. Rel. Maintainability Symp.*, 2000, pp. 71–76.

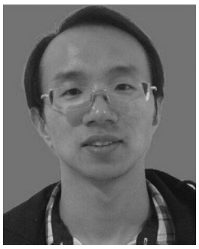
[34] J. Yang and F. B. Sun, "A comprehensive review of hard-disk drive reliability," in *Proc. Rel. Maintainability Symp.*, 1999, pp. 403–409.

[35] J. I. McCool, *Using the Weibull Distribution: Reliability, Modeling and Inference*. Hoboken, NJ, USA: Wiley, 2012.

[36] K. Trivedi, "Availability modeling," 2016. [Online]. Available: http://amod.ee.duke.edu/, Accessed on: Oct. 5, 2016.

[37] P. Li, J. Li, R. J. Stones, G. Wang, Z. Li, and X. Liu, "ProCode: A proactive erasure coding scheme for cloud storage systems," in *Proc. IEEE 35th Symp. Reliable Distrib. Syst.*, 2016, pp. 219–228.

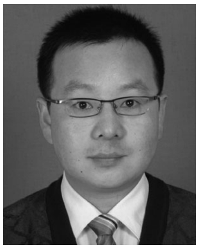
**Jing Li** received the BSc and MSc degrees in computer science and technology from Shandong University, Jinan, China, in 2004 and 2007, respectively and the PhD degree in computer science from Nankai University, Tianjin, China, in 2016. She is currently a teacher with the College of Computer Science and Technology, Civil Aviation University of China. Her research interests include mass data storage and machine learning.



**Peng Li** received the BSc degree in computer science and technology from Tianjin Normal University, Tianjin, China, in 2012. Now he is working toward the PhD degree in the College of Computer and Control Engineering, Nankai University, Tianjin, China. His research interests include cloud storage, distributed systems and erasure coding.

**Rebecca J. Stones** received the PhD degree in pure mathematics from Monash University, in 2010. She now has diverse research interests, including combinatorics and graph theory, codes, search engines and data storage, phylogenetics, and quantitative psychology.

**Gang Wang** received the BSc, MSc, and PhD degrees in computer science from Nankai University, Tianjin, China, in 1996, 1999, and 2002, respectively. He is currently a professor with the College of Computer and Control Engineering, Nankai University. His research interests include storage systems and parallel computing. He is a member of the IEEE.



**Zhongwei Li** received the PhD degree in computer science and technology from Harbin Engineering University, Harbin, China, in 2006. He is currently an associate professor with the College of Software, Nankai University, Tianjin, China. His research interests include machine learning and mass data storage.

**Xiaoguang Liu** received the BSc, MSc, and PhD degrees in computer science from Nankai University, Tianjin, China, in 1996, 1999, and 2002, respectively. He is currently a professor in computer science with Nankai University, Tianjin, China. His research interests include parallel computing and storage system.