Data Loss in RAID-5 and RAID-6 Storage Systems with Latent Errors

Ilias Iliadis
IBM Research – Zurich
8803 Rüschlikon, Switzerland
Email: ili@zurich.ibm.com

***Abstract*—Storage systems employ redundancy and recovering schemes to protect against device failures and latent sector errors, as well as to enhance reliability. The effectiveness of these schemes has been evaluated based on the Mean Time to Data Loss** (MTTDL) **and the Expected Annual Fraction of Data Loss** (EAFDL) **metrics. The reliability degradation due to device failures has been assessed in terms of both these metrics, but the adverse effect of latent errors has been assessed only in terms of the** MTTDL **metric. This article addresses the issue of evaluating the amount of data losses caused by latent errors. It presents a methodology for obtaining** MTTDL **and** EAFDL **of RAID-5 and RAID-6 systems analytically in the presence of unrecoverable or latent errors. A theoretical model capturing the effect of independent latent errors and device failures is developed, and closed-form expressions are derived for the metrics of interest.**

Keywords–Storage; Unrecoverable or latent sector errors; Reliability analysis; MTTDL; EAFDL; RAID; MDS codes; stochastic modeling.

I. INTRODUCTION

Today's large-scale data storage systems use data redundancy schemes to recover data lost due to device and component failures as well as to enhance reliability [1]. Erasure coding schemes are deployed that provide high data reliability as well as high storage efficiency. Special cases of erasure codes are the replication schemes and the Redundant Arrays of Inexpensive Disks (RAID) schemes, such as RAID-5 and RAID-6, which have been deployed extensively in the past thirty years [2-5]. The effectiveness of these schemes has been evaluated based on the Mean Time to Data Loss (MTTDL) [2-11] and, more recently, the Expected Annual Fraction of Data Loss (EAFDL) reliability metrics [12][13][14]. The latter metric was introduced because Amazon S3 considers the durability of data over a given year [15], and, similarly, Facebook [16], LinkedIn [17] and Yahoo! [18] consider the amount of lost data in given periods.

The reliability of storage systems is also degraded by the occurrence of unrecoverable sector errors, that is, errors that cannot be corrected by the standard sector-associated error-correcting code (ECC) nor by the re-read mechanism of hard-disk drives (HDDs). These sector errors are latent because their existence is only discovered when there is an attempt to access them. Once an unrecoverable or latent sector error is detected, it can usually be corrected by the RAID capability. However, if this is not feasible, these sectors are permanently lost, leading to an unrecoverable failure. Consequently, unrecoverable errors do not necessarily lead to unrecoverable failures. The effect of latent errors is quite pronounced in higher-capacity HDDs and storage nodes because of the high frequency of these errors [19-23]. The risk of permanent loss of data rises in the presence of latent errors.

Analytical reliability expressions for MTTDL that take into account the effect of latent errors have been obtained predominately using Markovian models, which assume that component failure and rebuild times are independent and exponentially distributed [8][21][22][24]. The effect of latent errors on MTTDL of erasure-coded storage systems for the practical case of non-exponential failure and rebuild time distributions was assessed in [23].

In this article, we consider the effect of latent errors not only on MTTDL, but also on the amount of lost data for the case of non-exponential failure and rebuild time distributions. Clearly, when a data loss occurs, the amount of data lost due to a device failure is much larger than the amount of sectors lost due to latent errors. We present a non-Markovian methodology for deriving the MTTDL and EAFDL metrics analytically for the case of RAID-5 and RAID-6 systems. We extend the methodology developed in prior work [12][13] to assess MTTDL and EAFDL of storage systems in the absence of latent errors. The validity of this methodology for accurately assessing the reliability of storage systems was confirmed by simulations in several contexts [4][9][12][25]. It has been demonstrated that theoretical predictions of the reliability of systems comprising highly reliable storage devices are in good agreement with simulation results. Consequently, the emphasis of the present work is on theoretically assessing the effect of latent errors on system reliability. This is the first work to study the effect of latent errors on EAFDL.

The key contributions of this article are the following. We consider the reliability of RAID storage systems that was assessed in our earlier work [1] for RAID-5 systems. We now extend our previous work by considering RAID-6 systems, which tolerate two device failures. We derive analytically the MTTDL and EAFDL reliability metrics. We subsequently establish that, for typical frequencies of sector errors, the probability of encountering an unrecoverable failure is much greater than that of encountering a device failure, which degrades MTTDL, but the EAFDL is practically unaffected in this range.

The remainder of the article is organized as follows. Section II describes the storage system model and the corresponding parameters considered. Section III considers the unrecoverable or latent errors and the frequency of their occurrence. Section IV presents the general framework and methodology

for deriving the MTTDL and EAFDL metrics analytically for the case of RAID systems and in the presence of latent errors. Closed-form expressions for relevant reliability metrics, such as the probability of data loss and the amount of data loss, are derived in Sections V and VI for RAID-5 and RAID-6 systems, respectively. Section VII presents numerical results demonstrating the effectiveness of the RAID-5 and RAID-6 schemes for improving system reliability and the adverse effect of unrecoverable or latent errors on the probability of data loss and on the MTTDL and EAFDL reliability metrics. Section VIII provides a discussion concerning the results obtained. Finally, we conclude in Section IX.

II. Storage System Model

The storage system considered here comprises n storage devices (nodes or disks), where each device stores an amount c of data such that the total storage capacity of the system is $n\,c$. User data is divided into blocks (or symbols) of a fixed size s (e.g., sector size of 512 bytes) and complemented with parity symbols to form codewords.

A. Redundancy

We consider an $(m, l) = (N, N-1)$ maximum distance separable (MDS) erasure code, which is a mapping from $N-1$ user-data symbols to a set of N symbols, called a codeword, having the property that any subset containing $N-1$ of the N symbols of the codeword can be used to decode (reconstruct, recover) the codeword. A single parity symbol is computed by using the XOR operation on $l = N-1$ user-data symbols to form a codeword with $m = N$ symbols in total. Such a scheme can tolerate a single erasure anywhere in the codeword. The N symbols of each codeword are stored on N distinct devices. More specifically, this scheme is used by the popular RAID-5 system, in which the n devices are arranged in groups (or arrays) of N devices, one of which is redundant [2][3]. The storage system therefore comprises n/N RAID-5 arrays, where each array has the ability to tolerate one device failure.

We also consider an $(m, l) = (N, N-2)$ MDS erasure code, which is a mapping from $N-2$ user-data symbols to a set of N codeword symbols having the property that any subset containing $N-2$ of the N symbols of the codeword can be used to decode (reconstruct, recover) the codeword. A codeword contains $l = N-2$ user-data symbols and two parity symbols for a total of $m = N$ symbols. Such a scheme can tolerate two erasures anywhere in the codeword. The N symbols of each codeword are stored on N distinct devices. More specifically, this scheme is used by the popular RAID-6 system, in which the n devices are arranged in groups (or arrays) of N devices, two of which are redundant [3]. The storage system therefore comprises n/N RAID-6 arrays, where each array has the ability to tolerate two device failures.

The storage efficiency $se^{(\mathrm{RAID})}$ of the system is [11][13]

$$se^{(\mathrm{RAID})} = \frac{l}{m} = \begin{cases} \frac{N-1}{N}, & \text{for RAID 5} \\ \frac{N-2}{N}, & \text{for RAID 6} \end{cases} \tag{1}$$

and the amount of user data U stored in the system is [13]

$$U = se^{(\mathrm{RAID})}\, n\, c = \frac{l\, n\, c}{m}\,. \tag{2}$$

TABLE I. Notation of system parameters

Parameter	Definition
n	number of storage devices
c	amount of data stored on each device
l	number of user-data symbols per codeword ($l \geq 1$)
m	total number of symbols per codeword ($m > l$)
(m, l)	MDS-code structure
s	symbol size
N	number of devices in a RAID array ($N = m$)
b	average reserved rebuild bandwidth per device
R	time required to read (or write) an amount c of data at an average rate b from (or to) a device
$F_R(.)$	cumulative distribution function of R
$F_\lambda(.)$	cumulative distribution function of device lifetimes
P_{bit}	Probability of an unrecoverable bit error
$se^{(\mathrm{RAID})}$	storage efficiency of RAID redundancy scheme ($se^{(\mathrm{RAID})} = l/m$)
U	amount of user data stored in the system ($U = se^{(\mathrm{RAID})}\, n\, c$)
C	number of codewords stored in a RAID array ($C = c/s$)
μ^{-1}	mean time to read (or write) an amount c of data at an average rate b from (or to) a device ($\mu^{-1} = E(R) = c/b$)
λ^{-1}	mean time to failure of a storage device ($\lambda^{-1} = \int_0^\infty [1 - F_\lambda(t)]dt$)
P_s	Probability of an unrecoverable sector (symbol) error

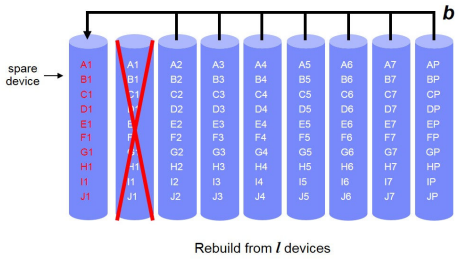

Figure 1. Rebuild for a RAID-5 array with $N = m = 8$ and $l = 7$.

Also, the number C of symbols in a device or, equivalently, the number of codewords in a RAID array is

$$C = \frac{c}{s}\,. \tag{3}$$

Our notation is summarized in Table I. The parameters are divided according to whether they are independent or derived, and are listed in the upper and lower part of the table, respectively.

B. Codeword Reconstruction

When a storage device of an array fails, the C codewords stored in the array lose one of their symbols. Subsequently, the system starts to reconstruct the lost codeword symbols using the surviving symbols of the affected codewords. We assume that device failures are detected instantaneously, which immediately triggers the rebuild process. A certain proportion of the device bandwidth is reserved for data recovery during the rebuild process, where b denotes the actual average reserved rebuild bandwidth per device. This bandwidth is usually only a fraction of the total bandwidth available at each device, the remaining bandwidth being used to serve user requests.

The rebuild process attempts to restore the codewords of the affected array sequentially. The lost symbols are reconstructed directly in a spare device as shown in Figure 1. Decoding and re-encoding of data are assumed to be done on the fly, so the reconstruction time is equal to the time taken to read and write the required data to the spare device. Consequently, the time required to recover the amount c of

lost data is equal to the time R required to read (or write) an amount c of data from (or to) a device. In particular, $1/\mu$ denotes the average time required to read (or write) an amount c of data from (or to) a device, which is defined by

$$\frac{1}{\mu} \triangleq E(R) = \frac{c}{b} . \quad (4)$$

C. Failure and Rebuild Time Distributions

We adopt the model and notation considered in [13]. The lifetimes of the n devices are assumed to be independent and identically distributed, with a cumulative distribution function $F_\lambda(.)$ and a mean of $1/\lambda$. We consider real-world distributions, such as Weibull and gamma, as well as exponential distributions that belong to the large class defined in [25]. The storage devices are characterized as *highly reliable* in that the ratio of the mean time $1/\mu$ to read all contents of a device (which typically is on the order of tens of hours), to the mean time to failure of a device $1/\lambda$ (which is typically on the order of thousands of hours) is very small, that is,

$$\frac{\lambda}{\mu} = \frac{\lambda\, c}{b} \ll 1 . \quad (5)$$

We consider storage devices whose cumulative distribution function F_λ satisfies the condition

$$\mu \int_0^\infty F_\lambda(t)[1 - F_R(t)]dt \ll 1, \quad \text{with } \frac{\lambda}{\mu} \ll 1 , \quad (6)$$

where $F_R(.)$ is the cumulative distribution function of the rebuild time R. Then the MTTDL and EAFDL reliability metrics tend to be insensitive to the device failure distribution, that is, they depend only on its mean $1/\lambda$, but not on its density $F_\lambda(.)$ [13].

III. Data Loss due to Unrecoverable Errors

The reliability of RAID systems is affected by the occurrence of unrecoverable or latent errors. Let P_{bit} denote the unrecoverable bit-error probability. According to the specifications, P_{bit} is equal to $1{\times}10^{-15}$ for SCSI drives and $1{\times}10^{-14}$ for SATA drives [8]. Assuming that bit errors occur independently over successive bits, the unrecoverable sector (symbol) error probability P_s is

$$P_s = 1 - (1 - P_{\text{bit}})^s , \quad (7)$$

with s expressed in bits. Assuming a sector size of 512 bytes, the equivalent unrecoverable sector error probability is $P_s \approx P_{\text{bit}} \times 4096$, which is $4.096{\times}10^{-12}$ in the case of SCSI and $4.096{\times}10^{-11}$ in the case of SATA drives. However, empirical field results suggest that the actual values can be orders of magnitude higher, reaching $P_s \approx 5{\times}10^{-9}$ [26].

IV. Derivation of MTTDL and EAFDL

The MTTDL metric assesses the expected time until some data can no longer be recovered and therefore is lost forever, whereas the EAFDL assesses the fraction of stored data that is expected to be lost by the system annually. We briefly review the general methodology for deriving the MTTDL and EAFDL metrics presented in [12]. This methodology does not involve Markovian analysis and holds for general failure time distributions, which can be exponential or non-exponential, such as the Weibull and gamma distributions that satisfy condition (6).

At any point in time, the system can be thought to be in one of two modes: normal or rebuild mode. During normal mode, all devices are operational and all data in the system has the original amount of redundancy. Any symbols encountered with unrecoverable or latent errors are usually corrected by the RAID capability. However, it may no be possible to recover multiple unrecoverable errors in a codeword, which therefore leads to data loss. A transition from normal to rebuild mode occurs when a device fails; we refer to the device failure that causes this transition as a *first device* failure. During rebuild mode, an active rebuild process attempts to restore the lost data in a spare device, which eventually leads the system either to a data loss (DL) with probability P_{DL} or back to the original normal mode by restoring initial redundancy, which occurs with probability $1 - P_{\text{DL}}$.

Let T be a typical interval of a fully operational period, that is, the interval from the time t that the system is brought to its original state until a subsequent first device failure occurs. For a system comprising n devices with a mean time to failure of a device $1/\lambda$, the expected duration of T is [12]

$$E(T) = 1/(n\,\lambda) , \quad (8)$$

and MTTDL is

$$\text{MTTDL} \approx \frac{E(T)}{P_{\text{DL}}} = \frac{1}{n\,\lambda\,P_{\text{DL}}} . \quad (9)$$

The EAFDL is obtained as the ratio of the expected amount $E(Q)$ of lost user data, normalized to the amount U of user data, to the expected duration of T [12, Equation (9)]:

$$\text{EAFDL} \approx \frac{E(Q)}{E(T)\cdot U} \overset{(8)}{=} \frac{n\,\lambda\,E(Q)}{U} \overset{(2)}{=} \frac{m\,\lambda\,E(Q)}{l\,c} , \quad (10)$$

with $E(T)$ and $1/\lambda$ expressed in years.

The expected amount $E(H)$ of lost user data, given that data loss has occurred, is determined by [12, Equation (8)]:

$$E(H) = \frac{E(Q)}{P_{\text{DL}}} . \quad (11)$$

It follows from (9) and (10) that the derivation of the MTTDL and EAFDL metrics requires the evaluation of P_{DL} and $E(Q)$, respectively. These quantities are derived using the direct path approximation [4][25][27], which, under conditions (5) and (6), accurately assesses the reliability metrics of interest [11][12][25][28].

V. RAID-5 Systems

Here we derive the reliability metrics for a RAID-5 system. When a storage device of a RAID-5 array fails, the C codewords stored in the array lose one of their symbols. Using the direct-path-approximation methodology, we proceed by considering only the subsequent potential data losses and device failures related to the affected array.

TABLE II. NOTATION OF VARIABLES

Parameter	Definition
I	number of codeword symbols with unrecoverable errors
L	number of codeword symbols lost
q	probability that a codeword can be restored
P_{DL}	probability of data loss
P_{UF}	probability of unrecoverable failures
S	number of lost symbols
Q	amount of lost user data
H	amount of lost user data, given that data loss has occurred

A. One Device Failure

The rebuild process attempts to restore the C codewords of the affected array sequentially. Let us consider such a codeword and let L_1 be the number of symbols permanently lost and I_1 be the number of symbols in the codeword with unrecoverable errors, as listed in Table II, where the subscript denotes the number of device failures. Owing to the independence of symbol errors, I_1 follows a binomial distribution with parameter P_s, the probability that a symbol has a latent (unrecoverable) error. Thus,

$$P(I_1 = i) = \binom{m-1}{i} P_s^i (1-P_s)^{m-1-i}, \text{ for } i = 0, \ldots, m-1 , \tag{12}$$

such that

$$E(I_1) = \sum_{i=1}^{m-1} i\, P(I_1 = i) \;=\; (m-1)\, P_s \; . \tag{13}$$

Clearly, the symbol lost due to the device failure can be corrected by the RAID-5 capability only if the remaining $m-1$ symbols can be read. Thus, $L_1 = 0$ if and only if $I_1 = 0$. Using (12), the probability q_1 that a codeword can be restored is

$$q_1 = P(L_1 = 0) = P(I_1 = 0) = (1 - P_s)^{m-1} \; , \tag{14}$$

which for very small values of P_s implies that

$$q_1 \approx \begin{cases} 1 - (m-1)\, P_s \, , & \text{for } P_s \ll \frac{1}{m-1} \\ 0 \, , & \text{for } P_s \gg \frac{1}{m-1} \; . \end{cases} \tag{15}$$

Note that if a codeword cannot be restored, then at least one of its l user-data symbols is lost. We now deduce that the conditional probability $P_{\text{UF}|1}$ of encountering an unrecoverable failure during the rebuild process of the C codewords given one device failure is

$$P_{\text{UF}|1} \;=\; 1 - q_1^C \tag{16}$$

$$\overset{(14)}{=} 1 - (1 - P_s)^{(m-1)\, C} \; . \tag{17}$$

Furthermore, such an unrecoverable failure entails the loss of user data. Let us denote by $N_{\text{UF}|1}$ the number of codewords that cannot be recovered owing to unrecoverable failures, referred to hereafter as *corrupted codewords*. Owing to the independence of symbol errors, codewords are independently corrupted. Therefore, $N_{\text{UF}|1}$ is binomially distributed with parameter $1 - q_1$, such that

$$E(N_{\text{UF}|1}) = C\,(1 - q_1) \; . \tag{18}$$

Remark 1: For very small values of P_s, it follows from (17) that

$$P_{\text{UF}|1} \approx \begin{cases} (m-1)\, C\, P_s \, , & \text{for } P_s \ll P_s^{(2)} \\ 1 \, , & \text{for } P_s \gg P_s^{(2)} \; . \end{cases} \tag{19}$$

where $P_s^{(2)}$ is obtained from the approximation (19)

$$P_{\text{UF}|1} \approx (m-1)\, C P_s^{(2)} = 1 \tag{20}$$

as follows:

$$\boxed{P_s^{(2)} \triangleq \frac{1}{C} \cdot \frac{1}{m-1}} \; . \tag{21}$$

Note also that from (15) and (18), it follows that

$$E(N_{\text{UF}|1}) \approx C\,(m-1)\, P_s \, , \quad \text{for } P_s \ll \frac{1}{m-1} \, . \tag{22}$$

In particular, for $P_s = P_s^{(2)}$, it holds that $E(N_{\text{UF}|1}) \approx 1$ and this, combined with the fact that $P_{\text{UF}|1} \approx 1$, implies that one of the C codewords is almost surely corrupted.

The expected number $E(N_{\text{UF}|1} | N_{\text{UF}|1} \geq 1)$ of the number of corrupted codewords, given that such codewords exist, is

$$E(N_{\text{UF}|1} | N_{\text{UF}|1} \geq 1) = \frac{E(N_{\text{UF}|1})}{P(N_{\text{UF}|1} \geq 1)} = \frac{E(N_{\text{UF}|1})}{P_{\text{UF}|1}} \tag{23}$$

$$\overset{(19)(22)}{\approx} \begin{cases} 1 \, , & \text{for } P_s \ll P_s^{(2)} \\ C\,(m-1)\, P_s \, , & \text{for } P_s^{(2)} \ll P_s \ll \frac{1}{m-1} \; . \end{cases} \tag{24}$$

When $I_1 > 0$, the number L_1 of lost symbols is $I_1 + 1$. Consequently, the expected number $E(L_1)$ of lost symbols is

$$E(L_1) = \sum_{i=1}^{m-1} (i+1)\, P(I_1 = i) = E(I_1) + 1 - P(I_1 = 0) \; , \tag{25}$$

which using (12), (13), and (14) yields

$$E(L_1) = 1 - q_1 + (m-1) P_s = 1 - (1 - P_s)^{m-1} + (m-1) P_s \; . \tag{26}$$

Remark 2: For small values of P_s, it follows from (15) that

$$E(L_1) \approx 2\,(m-1)\, P_s \, , \quad \text{for } P_s \ll \frac{1}{m-1} \, . \tag{27}$$

In particular, the expected number $E(L_1 | L_1 > 0)$ of lost symbols, given that the codeword is corrupted, is

$$E(L_1 | L_1 > 0) = \frac{E(L_1)}{P(L_1 > 0)} = \frac{E(L_1)}{1 - P(L_1 = 0)} \overset{(14)}{=} \frac{E(L_1)}{1 - q_1}$$

$$\overset{(15)(27)}{\approx} 2 \, , \qquad \text{for } P_s \ll \frac{1}{m-1} \, . \tag{28}$$

A subsequent (second) device failure may occur during the rebuild process, which is triggered by the initial device failure. The probability $P_{\text{DF}_2|R}$ of data loss due to two device failures, that is, the probability that one of the $m-1$ remaining devices in the array fails during the rebuild process, depends on the duration of the corresponding rebuild time R and the aggregate failure rate of these $m-1$ highly reliable devices, and is determined as follows [25]:

$$P_{\text{DF}_2|R} \approx (m-1)\, \lambda\, R \, . \tag{29}$$

In particular, it was shown in [29, Lemma 2] that, for highly reliable devices satisfying conditions (5) and (6), the fraction of the rebuild time R still remaining when another device fails is approximately uniformly distributed between 0 and 1. This implies that the probability $P(j | \text{DF}_2, R)$ that the second device

failure occurs during reconstruction of the jth $(1 \leq j \leq C)$ codeword does not depend on the rebuild duration R and is

$$P(j|\mathrm{DF}_2, R) = \frac{1}{C}\,, \quad \text{for } j = 1, 2, \ldots, C\,, \tag{30}$$

such that for large C, we have

$$E(J) = \sum_{j=1}^{C} j\, P(j|\mathrm{DF}_2, R) = \sum_{j=1}^{C} j\, \frac{1}{C} = \frac{C+1}{2} \approx \frac{C}{2}\,, \tag{31}$$

which implies that when the second failure occurs, on average half the codewords have already been considered for reconstruction.

The probability $P_{\mathrm{DF}_2}(j|R)$ that a device failure occurs during reconstruction of the jth $(1 \leq j \leq C)$ codeword, and given a rebuild time of R, is equal to the product of $P(j|\mathrm{DF}_2, R)$ and $P_{\mathrm{DF}_2|R}$, which, using (29) and (30), yields

$$P_{\mathrm{DF}_2}(j|R) \approx \frac{(m-1)\,\lambda\, R}{C}\,, \quad \text{for } j = 1, 2, \ldots, C\,. \tag{32}$$

The probability P_{DF_2} of data loss due to two device failures, that is, the probability of a device failure during the rebuild process, is obtained by unconditioning (29) on R, that is,

$$P_{\mathrm{DF}_2} = E(P_{\mathrm{DF}_2|R}) \approx (m-1)\,\lambda\, E(R) \overset{(4)}{=} (m-1)\,\frac{\lambda}{\mu}\,. \tag{33}$$

Consequently, the probability $P_{\mathrm{DF},1}$ of one device failure, that is, the probability of no subsequent device failure during the rebuild, is

$$P_{\mathrm{DF},1} = 1 - P_{\mathrm{DF}_2} \overset{(33)}{\approx} 1 - (m-1)\,\frac{\lambda}{\mu}\,. \tag{34}$$

Similarly, the probability $P_{\mathrm{DF}_2}(j)$ of a subsequent device failure during reconstruction of the jth $(1 \leq j \leq C)$ codeword is obtained by unconditioning (32) on R, that is,

$$P_{\mathrm{DF}_2}(j) = E(P_{\mathrm{DF}_2}(j|R)) \approx \frac{(m-1)\,\lambda\, E(R)}{C} \overset{(4)}{=} \frac{(m-1)}{C}\,\frac{\lambda}{\mu}\,. \tag{35}$$

The probability $P_{\mathrm{UF},1}$ of data loss due to a unrecoverable failures during rebuild in the case of one device failure is obtained by unconditioning the probability $P_{\mathrm{UF}|1}$ of unrecoverable failure during rebuild given one device failure, that is,

$$P_{\mathrm{UF},1} \;=\; P_{\mathrm{UF}|1}\, P_{\mathrm{DF},1} \overset{(34)}{=} (1 - P_{\mathrm{DF}_2})\, P_{\mathrm{UF}|1} \tag{36}$$

$$\overset{(16)(33)}{\approx} \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,(1 - q_1^C\,)\,, \tag{37}$$

where q_1 is determined by (14).

Corollary 1: It holds that

$$P_{\mathrm{UF},1} \;\approx\; P_{\mathrm{UF}|1} \overset{(16))}{=} 1 - q_1^C\,. \tag{38}$$

Proof: Immediate from (36) given that, according to (5), it holds that $P_{\mathrm{DF},1} = 1 - (m-1)\,\lambda/\mu \approx 1$ for small values of λ/μ . ■

Corollary 2: It holds that

$$P_{\mathrm{UF},1} \approx \begin{cases} P_{\mathrm{DF},1}\,(m-1)\,C\,P_s\,, & \text{for } P_s \ll P_s^{(2)} \\ P_{\mathrm{DF},1}\,, & \text{for } P_s \gg P_s^{(2)}\,, \end{cases} \tag{39}$$

where $P_{\mathrm{DF},1}$ and $P_s^{(2)}$ are given by (34) and (21), respectively.

Proof: Immediate by substituting (19) into (36). ■

Substituting (17) into (37) yields

$$P_{\mathrm{UF},1} \approx \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,\left[1 - (1-P_s)^{(m-1)\,C}\right]\,. \tag{40}$$

We now proceed to assess the amount of data loss. The expected number $E(S_{\mathrm{U}|1})$ of symbols lost due to unrecoverable failures during rebuild, given one device failure, is

$$E(S_{\mathrm{U}|1}) = C\,E(L_1)\,. \tag{41}$$

Substituting (26) into (41) yields

$$E(S_{\mathrm{U}|1}) = C\,[1 - (1-P_s)^{m-1} + (m-1)P_s]\,. \tag{42}$$

Remark 3: For small values of P_s, $E(L_1)$ is approximated by (27). Consequently, it follows from (41) that

$$E(S_{\mathrm{U}|1}) \approx 2\,C\,(m-1)\,P_s\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{43}$$

The expected number $E(S_{\mathrm{U},1})$ of symbols lost due to unrecoverable failures during rebuild in conjunction with one device failure is obtained by unconditioning the conditional expected number $E(S_{\mathrm{U}|1})$ of symbols lost due to unrecoverable failures during rebuild given one device failure, that is,

$$E(S_{\mathrm{U},1}) = E(S_{\mathrm{U}|1})\,P_{\mathrm{DF},1} \overset{(34)}{=} (1 - P_{\mathrm{DF}_2})\,E(S_{\mathrm{U}|1})\,. \tag{44}$$

Corollary 3: It holds that

$$E(S_{\mathrm{U},1}) \;\approx E(S_{\mathrm{U}|1}) = C\,E(L_1) \tag{45}$$

$$\overset{(27)}{\approx} 2\,C\,(m-1)\,P_s\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{46}$$

Proof: Immediate from (44) given that, according to (5), it holds that $P_{\mathrm{DF},1} = 1 - (m-1)\,\lambda/\mu \approx 1$ for small values of λ/μ, and also using (41). ■

Substituting (34) and (41) into (44) yields

$$E(S_{\mathrm{U},1}) = C\,E(L_1)\,\left[1 - (m-1)\,\frac{\lambda}{\mu}\right] \tag{47}$$

$$\overset{(26)}{\approx} C\,[1 - (1-P_s)^{m-1} + (m-1)P_s]\,\left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,. \tag{48}$$

B. Two Device Failures

As discussed in Section V-A, a subsequent (second) device failure may occur during the rebuild process, which is triggered by the first device failure. In particular, such a failure may occur during reconstruction of the jth $(1 \leq j \leq C)$ codeword with a probability of $P_{\mathrm{DF}_2}(j)$ determined by (35). This device failure divides the C codewords into two sets:

- $S_{1,j}$: the set of $j-1$ codewords already considered for reconstruction, and
- $S_{2,j}$: the set of remaining $C-j+1$ codewords, none of which can be reconstructed.

Note that the occurrence of the second device failure does not exclude the possibility of unrecoverable failures being

encountered prior to its occurrence. More specifically, the probability that no unrecoverable failures occur in $S_{1,j}$ is q_1^{j-1}, which implies that the probability $P_{\text{UF in } S_{1,j}|2}$ that one or more unrecoverable failures occur in $S_{1,j}$ is

$$P_{\text{UF in } S_{1,j}|2} = 1 - q_1^{j-1} \,. \tag{49}$$

Furthermore, such unrecoverable failures entail loss of user data. Let us denote by $N_{\text{UF in } S_{1,j}|2}$ the number of such corrupted codewords. Then it holds that

$$E(N_{\text{UF in } S_{1,j}|2}) = (j-1)\,(1-q_1) \,. \tag{50}$$

Also, each of the $C-j+1$ codewords in $S_{2,j}$ can no longer be reconstructed. In particular, the two symbols of the jth codeword that are stored on the two failed devices can no longer be recovered and are lost. Furthermore, any of its remaining $m-2$ symbols encountered with unrecoverable errors will also be lost, leading to unrecoverable failure. Thus, the probability p_2 of not encountering an unrecoverable failure due to unrecoverable errors in the remaining $m-2$ symbols is

$$p_2 = (1-P_s)^{m-2} \,, \tag{51}$$

which for very small values of P_s implies that

$$p_2 \approx \begin{cases} 1-(m-2)\,P_s \,, & \text{for } P_s \ll \frac{1}{m-2} \\ 0 \,, & \text{for } P_s \gg \frac{1}{m-2} \,. \end{cases} \tag{52}$$

The same applies for the remaining $C-j$ codewords. Therefore the probability that no unrecoverable failures occur in $S_{2,j}$ is p_2^{C-j+1}, which implies that the probability $P_{\text{UF in } S_{2,j}|2}$ that an unrecoverable failure occurs in $S_{2,j}$ is

$$P_{\text{UF in } S_{2,j}|2} = 1 - p_2^{C-j+1} \,. \tag{53}$$

Furthermore, such unrecoverable failures entail the loss of user data. Let us denote by $N_{\text{UF in } S_{2,j}|2}$ the number of such corrupted codewords. Then it holds that

$$E(N_{\text{UF in } S_{2,j}|2}) = (C-j+1)\,(1-p_2) \,. \tag{54}$$

Also, the probability $P_{\text{UF}|2}(j)$ that an unrecoverable failure occurs, given two device failures, is

$$P_{\text{UF}|2}(j) = 1 - q_1^{j-1}\, p_2^{C-j+1} \,. \tag{55}$$

The second device failure divides the C codewords into two sets:

S_1 : the set of the codewords already considered for reconstruction when the second device failure occurs, and

S_2 : the set of the remaining codewords, none of which can be reconstructed.

The probability $P_{\text{UF in } S_1|2}$ that an unrecoverable failure occurs in S_1 is obtained by unconditioning (49) on j as follows:

$$\begin{aligned} P_{\text{UF in } S_1|2} &= \sum_{j=1}^{C} P_{\text{UF in } S_{1,j}|2} P(j|\text{DF}_2, R) \\ &\overset{(30)}{\approx} \sum_{j=1}^{C} (1-q_1^{j-1})\,\frac{1}{C} = 1 - \frac{1-q_1^C}{C\,(1-q_1)} \,, \end{aligned} \tag{56}$$

which does not depend on the rebuild duration.

Corollary 4: It holds that

$$P_{\text{UF in } S_1|2} \approx \begin{cases} \dfrac{(C-1)(m-1)}{2}\,P_s \,, & \text{for } P_s \ll 2\,P_s^{(2)} \\ 1 \,, & \text{for } P_s \gg 2\,P_s^{(2)} \,, \end{cases} \tag{57}$$

where $P_s^{(2)}$ is determined by (21).

Proof: See Appendix A. ■

Remark 4: For $P_s \ll P_s^{(2)}$, and considering that C is large, it follows from (19) and (57) that

$$P_{\text{UF in } S_1|2} \approx \frac{1}{2}\,P_{\text{UF}|1} \,, \quad \text{for } P_s \ll P_s^{(2)} \,. \tag{58}$$

This is intuitively obvious because, according to (31), when the second failure occurs, on average half the codewords have already been considered for reconstruction, that is, $E(|S_1|) \approx C/2$.

The expected number $E(N_{\text{UF in } S_1|2})$ of corrupted codewords in S_1 is obtained by unconditioning (54) on j as follows:

$$\begin{aligned} E(N_{\text{UF in } S_1|2}) &= \sum_{j=1}^{C} E(N_{\text{UF in } S_{1,j}|2}) P(j|\text{DF}_2, R) \\ &\overset{(30)}{\approx} \sum_{j=1}^{C} (j-1)(1-q_1)\,\frac{1}{C} = \frac{C-1}{2}\,(1-q_1) \end{aligned} \tag{59}$$

$$\overset{(15)}{\approx} \frac{C}{2}\,(m-1)\,P_s \,, \quad \text{for } P_s \ll \frac{1}{m-1} \,. \tag{60}$$

The probability $P_{\text{UF in } S_2|2}$ that an unrecoverable failure occurs in S_2 is obtained by unconditioning (53) on j as follows:

$$\begin{aligned} P_{\text{UF in } S_2|2} &= \sum_{j=1}^{C} P_{\text{UF in } S_{2,j}|2} P(j|\text{DF}_2, R) \\ &\approx \sum_{j=1}^{C} (1-p_2^{C-j+1})\,\frac{1}{C} = 1 - \frac{p_2}{C}\,\frac{1-p_2^C}{1-p_2} \,, \end{aligned} \tag{61}$$

which does not depend on the rebuild duration.

Corollary 5: It holds that

$$P_{\text{UF in } S_2|2} \approx \begin{cases} \dfrac{(C+1)(m-2)}{2}\,P_s \,, & \text{for } P_s \ll P_{s^*}^{(2)} \\ 1 \,, & \text{for } P_s \gg P_{s^*}^{(2)} \,, \end{cases} \tag{62}$$

where $P_{s^*}^{(2)} = 2/[C\,(m-2)]$. Note that by virtue of (21), it holds that $P_{s^*}^{(2)} = [(m-1)(m-2)]\,2\,P_s^{(2)} > 2\,P_s^{(2)}$.

Proof: See Appendix B. ■

The probability $P_{\text{UF}|2}$ of data loss due to unrecoverable failures, given two device failures and a rebuild duration of R, is obtained by unconditioning (55) on j and using (30) as

follows:

$$P_{\text{UF}|2} = \sum_{j=1}^{C} P_{\text{UF}|2}(j)\, P(j|\text{DF}_2, R) \approx \sum_{j=1}^{C} \left(1 - q_1^{j-1}\, p_2^{C-j+1}\right) \frac{1}{C} = 1 - \frac{p_2}{C}\, \frac{p_2^C - q_1^C}{p_2 - q_1}\,, \tag{63}$$

which does not depend on the rebuild duration.

Corollary 6: It holds that

$$P_{\text{UF}|2} \approx \begin{cases} \left[\dfrac{C-1}{2} + (m-2)\, C\right] P_s\,, & \text{for } P_s \ll P_s^{(2)} \\ 1\,, & \text{for } P_s \gg P_s^{(2)}\,, \end{cases} \tag{64}$$

where $P_s^{(2)}$ is determined by (21).

Proof: See Appendix C. ■

The probability $P_{\text{UF},2}$ of data loss due to unrecoverable failures in conjunction with two device failures is obtained by unconditioning (63) via (33) as follows:

$$P_{\text{UF},2} = P_{\text{UF}|2}\, P_{\text{DF}_2} \approx \left(1 - \frac{p_2}{C}\, \frac{p_2^C - q_1^C}{p_2 - q_1}\right) (m-1)\, \frac{\lambda}{\mu}\,, \tag{65}$$

where q_1 and p_2 are determined by (14) and (51), respectively.

Corollary 7: It holds that

$$P_{\text{UF},2} \approx \begin{cases} \left[\frac{C-1}{2} + (m-2)\, C\right] (m-1)\, \frac{\lambda}{\mu}\, P_s\,, & \text{for } P_s \ll P_s^{(2)} \\ P_{\text{DF}_2} \approx (m-1)\, \frac{\lambda}{\mu}\,, & \text{for } P_s \gg P_s^{(2)}, \end{cases} \tag{66}$$

where $P_s^{(2)}$ is determined by (21).

Proof: Immediate by substituting (64) and (33) into (65). ■

Remark 5: From (39) and (66), it follows that

$$P_{\text{UF},1} \gg P_{\text{UF},2}\,, \tag{67}$$

because $P_{\text{UF},2}$ is of the order $O(\lambda/\mu)$, which by virtue of (5) is very small, whereas $P_{\text{UF},1}$ is not.

We now proceed to assess the amount of data loss. As discussed above, the two symbols of each of the $C - j + 1$ codewords in $S_{2,j}$ that are stored on the two failed devices can no longer be recovered and are lost. Thus, the total number $S_{\text{D}}(j)$ of symbols in these $C-j+1$ codewords that are stored on the two failed devices and are therefore lost is

$$S_{\text{D}}(j) = 2\,(C + 1 - j)\,. \tag{68}$$

Also, the expected total number $E(S_{\text{U},2}^{+}\,|\,\text{DF}_2 \text{ at } j)$ of symbols stored in these $C - j + 1$ codewords and lost due to unrecoverable failures is

$$E(S_{\text{U},2}^{+}\,|\,\text{DF}_2 \text{ at } j) = (C + 1 - j)\,(m-2)\, P_s\,. \tag{69}$$

Furthermore, each of the $j - 1$ codewords in $S_{1,j}$ loses an expected number of $E(L_1)$ symbols. Consequently, the expected total number $E(S_{\text{U},2}^{-}\,|\,\text{DF}_2 \text{ at } j)$ of symbols stored in these $j - 1$ codewords and lost due to unrecoverable failures is

$$E(S_{\text{U},2}^{-}\,|\,\text{DF}_2 \text{ at } j) = (j-1)\, E(L_1)\,. \tag{70}$$

Unconditioning (68), (69), and (70) on the event of a device failure during reconstruction of the jth codeword, and using (35), yields

$$E(S_{\text{D}}) \approx \sum_{j=1}^{C} 2\,(C + 1 - j)\, \frac{(m-1)}{C}\, \frac{\lambda}{\mu} \tag{71}$$

$$= (C+1)\,(m-1)\, \frac{\lambda}{\mu}\,, \tag{72}$$

$$E(S_{\text{U},2}^{+}) \approx \sum_{j=1}^{C} (C + 1 - j)\,(m-2)\, P_s\, \frac{(m-1)}{C}\, \frac{\lambda}{\mu} \tag{73}$$

$$= \frac{C+1}{2}\, (m-1)\,(m-2)\, \frac{\lambda}{\mu}\, P_s\,, \tag{74}$$

and

$$E(S_{\text{U},2}^{-}) \approx \sum_{j=1}^{C} (j-1)\, E(L_1)\, \frac{(m-1)}{C}\, \frac{\lambda}{\mu} \tag{75}$$

$$\overset{(26)}{=} \frac{C-1}{2}\, [1 - (1-P_s)^{m-1} + (m-1)P_s](m-1)\, \frac{\lambda}{\mu} \tag{76}$$

$$\overset{(27)}{\approx} (C-1)\,(m-1)^2\, \frac{\lambda}{\mu}\, P_s\,, \text{ for } P_s \ll \frac{1}{m-1}\,. \tag{77}$$

In particular, from (72) and considering that, for large values of C, $C + 1 \approx C$, it holds that

$$E(S_{\text{D}}) \approx C\,(m-1)\, \frac{\lambda}{\mu}\,. \tag{78}$$

The expected total number $E(S_{\text{U},2})$ of symbols lost due to unrecoverable errors in conjunction with two device failures is

$$E(S_{\text{U},2}) = E(S_{\text{U},2}^{+}) + E(S_{\text{U},2}^{-})\,, \tag{79}$$

where $E(S_{\text{U},2}^{+})$ and $E(S_{\text{U},2}^{-})$ are determined by (74) and (76), respectively.

Remark 6: From (48), (74), and (76), it follows that $E(S_{\text{U},1}) \gg E(S_{\text{U},2}^{-}) > E(S_{\text{U},2}^{+})$. The first inequality follows from the fact that $E(S_{\text{U},2}^{-})$ is of the order $O(\lambda/\mu)$, which is very small, whereas $E(S_{\text{U},1})$ is not. The second inequality follows from the fact that, for large values of C, we have $E(S_{\text{U},2}^{-})/E(S_{\text{U},2}^{+}) \approx [\,1 - (1-P_s)^{m-1} + (m-1)P_s\,]/[(m-2)P_s] > 1$. As discussed above, it follows that $E(S_{\text{U},1}) \gg E(S_{\text{U},2})$. Consequently, the symbols lost due to unrecoverable errors are predominately encountered during a rebuild that is completed without experiencing an additional device failure.

From (43), (45), and (78), it follows that

$$E(S_{\text{U},1}) \ll E(S_{\text{D}}) \;\Leftrightarrow\; 2\,C\,(m-1)\, P_s \ll C\,(m-1)\, \frac{\lambda}{\mu} \;\Leftrightarrow\; P_s \ll P_s^{(3)}\,, \tag{80}$$

where

$$\boxed{P_s^{(3)} \triangleq \frac{1}{2} \cdot \frac{\lambda}{\mu}}\,. \tag{81}$$

Remark 7: From (72), it follows that the expected number $E(S_{\text{D}})$ of symbols stored on the two failed devices and lost

is of the order $O(\lambda/\mu)$. Note that the above analysis does not exclude the possibility that additional device failures occur during rebuild. However, the corresponding expected number of the additional lost symbols can be ignored because it is of the order $O((\lambda/\mu)^2)$, which is much smaller than $O(\lambda/\mu)$. This is confirmed by (218), which is derived in Section VI-C and obtains the expected number of lost symbols in conjunction with three device failures.

C. Data Loss

Data loss during rebuild may occur because of another (second) device failure or an unrecoverable failure of one or more codewords, or a combination thereof.

Let P_{DL} denote the probability of data loss. Then, the probability $1-P_{\text{DL}}$ of the rebuild being completed successfully is equal to the product of $1 - P_{\text{DF}_2}$, the probability of not encountering a device failure during rebuild, and $1 - P_{\text{UF}|1}$, the probability of not encountering an unrecoverable failure during rebuild, namely, $1 - P_{\text{DL}} = (1 - P_{\text{DF}_2})(1 - P_{\text{UF}|1})$. Consequently,

$$P_{\text{DL}} = P_{\text{DF}_2} + (1 - P_{\text{DF}_2})\, P_{\text{UF}|1} \overset{(36)}{=} P_{\text{DF}_2} + P_{\text{UF},1}\,. \quad (82)$$

This expresses the fact that a data loss during rebuild may occur either because of two device failures or unrecoverable failures in the case of one device failure. These are two mutually exclusive events. Substituting (17) and (34) into (82) yields

$$P_{\text{DL}} \approx (m-1)\,\frac{\lambda}{\mu} + \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,\left[1 - (1-P_s)^{(m-1)\,C}\right]. \quad (83)$$

Corollary 8: For small values of λ/μ, it holds that

$$P_{\text{DL}} \approx P_{\text{DF}_2} + P_{\text{UF}|1} \quad (84)$$

$$\overset{(33)(17)}{\approx} (m-1)\,\frac{\lambda}{\mu} + 1 - (1-P_s)^{(m-1)\,C} \quad (85)$$

$$\overset{(33)(19)}{\approx} \begin{cases} (m-1)\,\left(\frac{\lambda}{\mu} + C\,P_s\right), & \text{for } P_s \ll P_s^{(2)} \\ 1\,, & \text{for } P_s \gg P_s^{(2)}\,, \end{cases} \quad (86)$$

where $P_s^{(2)}$ is determined by (21).

Proof: Immediate from (82) because $P_{\text{DF}_2} \ll 1$ due to (5). ■

Remark 8: When P_s increases and approaches 1, the P_{DL} obtained by (84) and (85) approaches $1 + (m-1)\lambda/\mu$ and therefore exceeds 1.

Remark 9: It follows from (19), (33), and (38) that the range $[0, P_s^{(1)})$ of P_s in which the probabilities $P_{\text{UF},1}$ and $P_{\text{UF}|1}$ are much smaller than the probability P_{DF_2} of encountering a device failure during rebuild is obtained by

$$P_{\text{UF}|1} \overset{(38)}{\approx} P_{\text{UF},1} \ll P_{\text{DF}_2}$$

$$\Leftrightarrow (m-1)\,C\,P_s \ll (m-1)\,\frac{\lambda}{\mu} \Leftrightarrow P_s \ll P_s^{(1)}, \quad (87)$$

where

$$\boxed{P_s^{(1)} \triangleq \frac{1}{C}\cdot\frac{\lambda}{\mu}}\,. \quad (88)$$

Also, it follows from (86) and (87) that

$$P_{\text{DL}} \approx \begin{cases} (m-1)\,\frac{\lambda}{\mu}\,, & \text{for } P_s \ll P_s^{(1)} \\ (m-1)\,C\,P_s\,, & \text{for } P_s^{(1)} \ll P_s \ll P_s^{(2)} \\ 1\,, & \text{for } P_s \gg P_s^{(2)}\,, \end{cases} \quad (89)$$

where $P_s^{(1)}$ and $P_s^{(2)}$ are determined by (88) and (21), respectively. Note that P_{DL}, as a function of P_s, exhibits two plateaus in the intervals $[0, P_s^{(1)})$ and $(P_s^{(2)}, 1]$, respectively.

Unrecoverable failures may occur in the cases of one device failure and two device failures. Consequently, the probability P_{UF} of encountering one or more unrecoverable failures during rebuild is

$$P_{\text{UF}} = P_{\text{UF},1} + P_{\text{UF},2} \quad (90)$$

$$\overset{(67)}{\approx} P_{\text{UF},1} \overset{(40)}{\approx} \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\left[1 - (1-P_s)^{(m-1)\,C}\right] \quad (91)$$

$$\overset{(38)}{\approx} P_{\text{UF}|1} \overset{(19)}{\approx} \begin{cases} (m-1)\,C\,P_s\,, & \text{for } P_s \ll P_s^{(2)} \\ 1\,, & \text{for } P_s \gg P_s^{(2)}\,, \end{cases} \quad (92)$$

where $P_{\text{UF},1}$ and $P_{\text{UF},2}$ are given by (40) and (65), respectively.

D. Amount of Data Loss

As discussed in Section V-C, data loss during rebuild may occur because of another (second) device failure or an unrecoverable failure of one or more codewords, or a combination thereof. Note that in all cases, data loss cannot involve only parity data, but also loss of user data.

Data loss during rebuild may occur because of unrecoverable failures in the cases of one device failure or two device failures. Consequently, the expected number $E(S_{\text{U}})$ of symbols lost due to unrecoverable errors is obtained as follows:

$$E(S_{\text{U}}) = E(S_{\text{U},1}) + E(S_{\text{U},2})\,, \quad (93)$$

where $E(S_{\text{U},1})$ and $E(S_{\text{U},2})$ are determined by (48) and (79), respectively. Moreover, according to Remark 6, it holds that

$$E(S_{\text{U}}) \approx E(S_{\text{U},1}) \quad (94)$$

$$\overset{(46)}{\approx} 2\,C\,(m-1)\,P_s\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \quad (95)$$

The expected total number $E(S)$ of lost symbols is

$$E(S) = E(S_{\text{D}}) + E(S_{\text{U}})\,, \quad (96)$$

where $E(S_{\text{D}})$ and $E(S_{\text{U}})$ are determined by (72) and (93), respectively.

Remark 10: It follows from (45), (78), (94), and (96) that

$$E(S) \approx C\left[(m-1)\,\frac{\lambda}{\mu} + E(L_1)\right] \quad (97)$$

$$\overset{(27)}{\approx} C\,(m-1)\left(\frac{\lambda}{\mu} + 2\,P_s\right), \text{ for } P_s \ll \frac{1}{m-1}\,, \quad (98)$$

where $E(L_1)$ is determined by (26). In particular, for $P_s = 0$, it holds that $E(S) = E(S_{\text{D}}) \approx C\,(m-1)\,\lambda/\mu$.

Remark 11: When P_s increases and approaches 1, it follows from (48), (72), (74), (76), (93), and (96) that $E(S)$ approaches $C\,m$. This is intuitively obvious because when

$P_s = 1$, all $C\,m$ symbols stored in the the RAID-5 array are lost owing to unrecoverable errors.

We now proceed to derive $E(Q)$, the expected amount of lost user data. First, we note that the expected number of lost user symbols is equal to the product of the storage efficiency and the expected number of lost symbols. Consequently, it follows from (1) that

$$E(Q) = \frac{l}{m}\, E(S)\, s \overset{(3)}{=} \frac{l}{m}\, \frac{E(S)}{C}\, c\,, \tag{99}$$

where $E(S)$ is given by (96) and s denotes the symbol size.

Similar expressions for the expected amounts $E(Q_{\text{DF}_2})$ and $E(Q_{\text{UF}})$ of user data lost due to device and unrecoverable failures are obtained from $E(S_{\text{D}})$ and $E(S_{\text{U}})$, respectively, as follows:

$$E(Q_{\text{DF}_2}) = \frac{l}{m}\, E(S_{\text{D}})\, s \overset{(3)}{=} \frac{l}{m}\, \frac{E(S_{\text{D}})}{C}\, c\,, \tag{100}$$

$$\overset{(78)}{\approx} \frac{l}{m}\,(m-1)\,\frac{\lambda}{\mu}\,c \tag{101}$$

and

$$E(Q_{\text{UF}}) = \frac{l}{m}\, E(S_{\text{U}})\, s \overset{(3)}{=} \frac{l}{m}\, \frac{E(S_{\text{U}})}{C}\, c \tag{102}$$

$$\overset{(95)}{\approx} 2\,\frac{l}{m}\,(m-1)\,c\,P_s\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,, \tag{103}$$

where $E(S_{\text{D}})$ and $E(S_{\text{U}})$ are determined by (72) and (93), respectively.

Substituting (97) and (98) into (99) yields

$$E(Q) \approx \frac{l}{m}\,\left[(m-1)\,\frac{\lambda}{\mu} + E(L_1)\right] c \tag{104}$$

$$\approx \frac{l}{m}\,(m-1)\left(\frac{\lambda}{\mu} + 2\,P_s\right) c, \text{ for } P_s \ll \frac{1}{m-1}\,, \tag{105}$$

where $E(L_1)$ is given by (26). In particular, for $P_s = 0$, it holds that $E(Q) = E(Q_{\text{DF}_2})$, which is determined by (101).

From (96), (99), (100), and (102), it holds that

$$E(Q) = E(Q_{\text{DF}_2}) + E(Q_{\text{UF}})\,. \tag{106}$$

Also, the expected amounts $E(Q_{\text{UF},1})$ and $E(Q_{\text{UF},2})$ of user data lost due to unrecoverable failures in the cases of one device failure and two device failures are

$$E(Q_{\text{UF},1}) = \frac{l}{m}\, \frac{E(S_{\text{U},1})}{C}\, c \tag{107}$$

and

$$E(Q_{\text{UF},2}) = \frac{l}{m}\, \frac{E(S_{\text{U},2})}{C}\, c\,, \tag{108}$$

where $E(S_{\text{U},1})$ and $E(S_{\text{U},2})$ are determined by (48) and (79), respectively.

Remark 12: From (94), (102), and (107), it follows that

$$E(Q_{\text{UF}}) \approx E(Q_{\text{UF},1})\,. \tag{109}$$

Remark 13: From (80), (94), (100), and (102), it follows that

$$E(Q_{\text{UF}}) \ll E(Q_{\text{DF}_2}) \Leftrightarrow P_s \ll P_s^{(3)}\,, \tag{110}$$

where $P_s^{(3)}$ is determined by (81).

Also, from (101), (103), (106), and (110), it follows that

$$E(Q) \approx \begin{cases} E(Q_{\text{DF}_2})\,, & \text{for } P_s \ll P_s^{(3)} \\ E(Q_{\text{UF}})\,, & \text{for } P_s \gg P_s^{(3)} \end{cases} \tag{111}$$

$$\approx \begin{cases} \frac{l}{m}\,(m-1)\,\frac{\lambda}{\mu}\,c\,, & \text{for } P_s \ll P_s^{(3)} \\ 2\,\frac{l}{m}\,(m-1)\,c\,P_s\,, & \text{for } P_s^{(3)} \ll P_s \ll \frac{1}{m-1}\,. \end{cases} \tag{112}$$

Remark 14: When P_s increases and approaches 1, from (99) and according to Remark 11, it follows that $E(Q)$ approaches $c\,l$. This is intuitively obvious because when $P_s = 1$, upon the first device failure, the entire amount $c\,l$ of user data stored in the RAID-5 array is lost owing to unrecoverable errors.

E. Reliability Metrics

The MTTDL normalized to $1/\lambda$ is obtained by substituting (83) into (9) as follows:

$$\lambda\,\text{MTTDL} \approx \frac{1}{n\left\{(m-1)\,\frac{\lambda}{\mu} + \left[1-(m-1)\,\frac{\lambda}{\mu}\right]\left[1-(1-P_s)^{(m-1)\,C}\right]\right\}}\,, \tag{113}$$

where C and λ/μ are determined by (3) and (5), respectively. In particular, substituting (86) into (9) yields

$$\lambda\,\text{MTTDL} \approx \begin{cases} \dfrac{1}{n\,(m-1)\,\left(\frac{\lambda}{\mu} + C\,P_s\right)}, & \text{for } P_s \ll P_s^{(2)} \\ \dfrac{1}{n}\,, & \text{for } P_s \gg P_s^{(2)}\,. \end{cases} \tag{114}$$

Note that MTTDL is insensitive to device failure and rebuild time distributions; it depends only on their means $1/\lambda$ and $1/\mu$, respectively. In particular, the normalized MTTDL depends only on the ratio λ/μ of their means. Also, for $\lambda/\mu \ll 1$ and $n = m = N$, the MTTDL derived in (113) is approximately equal to

$$\text{MTTDL} \approx \frac{1 + (2N-1)\,\frac{\lambda}{\mu}}{N\,\lambda\left\{(N-1)\,\frac{\lambda}{\mu} + \left[1-(1-P_s)^{(N-1)\,C}\right]\right\}}\,, \tag{115}$$

which is Equation (43) derived in [21]. Furthermore, for $P_s = 0$ and $n = m = N$, (113) yields

$$\text{MTTDL} \approx \frac{\mu}{N\,(N-1)\,\lambda^2}\,, \tag{116}$$

which is the same result as the one derived in [2, 3] (for a single array, namely, $n_G = 1$).

The EAFDL is obtained by substituting (99) into (10). In particular, the EAFDL normalized to λ is obtained by substituting (105) into (10) as follows:

$$\text{EAFDL}/\lambda \approx (m-1)\left(\frac{\lambda}{\mu} + 2\,P_s\right), \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{117}$$

where λ/μ is determined by (5). Note that EAFDL is insensitive to device failure and rebuild time distributions; it depends only on their means $1/\lambda$ and $1/\mu$, respectively. In particular,

the normalized EAFDL only depends on the ratio λ/μ of their means. Also, for $P_s = 0$, (117) is in agreement with Equation (74) of [14] (with $c/b = 1/\mu$ and $\phi = 1$).

The value of $E(H)$ is obtained by substituting (83) and (99) into (11). In particular, the $E(H)$ normalized to c is obtained by substituting (83) and (105) into (11) as follows:

$$E(H)/c \approx \frac{\frac{l}{m}(m-1)\left(\frac{\lambda}{\mu}+2\,P_s\right)}{(m-1)\,\frac{\lambda}{\mu}\,+\,\left[1-(m-1)\,\frac{\lambda}{\mu}\right]\,\left[1-(1-P_s)^{(m-1)\,C}\right]}\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,, \quad (118)$$

where C and λ/μ are determined by (3) and (5), respectively.

Note that $E(H)$ does not depend on the device failure nor the rebuild time distributions; it only depends on the ratio of their means λ/μ. Also, for $P_s = 0$, (118) yields $E(H)/c = l/m$, which is in agreement with Equation (75) of [14].

Similar to (11), the expected amounts $E(H_{\text{DF}_2})$ and $E(H_{\text{UF}})$ of user data lost due to device and unrecoverable failures, given that such failures have occurred, are

$$E(H_{\text{DF}_2}) = \frac{E(Q_{\text{DF}_2})}{P_{\text{DF}_2}}\,, \quad \text{and} \quad E(H_{\text{UF}}) = \frac{E(Q_{\text{UF}})}{P_{\text{UF}}}\,, \quad (119)$$

respectively.

From (11), (106), and (119), we deduce that the following relation holds

$$E(H) \;=\; \frac{P_{\text{DF}_2}}{P_{\text{DL}}}\,E(H_{\text{DF}_2}) + \frac{P_{\text{UF}}}{P_{\text{DL}}}\,E(H_{\text{UF}})\,. \quad (120)$$

Note that this is not a weighted average of $E(H_{\text{DF}_2})$ and $E(H_{\text{UF}})$ because a subsequent device failure and unrecoverable failures during rebuild are not mutually exclusive, and therefore, and according to (82) and (90), the sum of weights is not equal to but close to 1.

Remark 15: According to (21), (81), and (88), it holds that $P_s^{(1)} \ll \min(P_s^{(2)}, P_s^{(3)})$ and $P_s^{(2)} \lesseqgtr P_s^{(3)} \Leftrightarrow \lambda/\mu \gtrless 2/[C\,(m-1)]$. From (89) and (112), it follows that

$$E(H)/c \approx \begin{cases} \frac{l}{m}\,, & \text{for } P_s \ll P_s^{(1)} \\ \frac{l}{m}\,\frac{\lambda}{\mu}\,\frac{1}{C\,P_s}\,, & \text{for } P_s^{(1)} \ll P_s \ll P_{\min}^{(2,3)} \\ 2\,\frac{l}{m}\,\max(\frac{m-1}{2}\,\frac{\lambda}{\mu}, \frac{1}{C})\,, & \text{for } P_{\min}^{(2,3)} \ll P_s \ll P_{\max}^{(2,3)} \\ 2\,\frac{l}{m}\,(m-1)P_s\,, & \text{for } P_{\max}^{(2,3)} \ll P_s \ll \frac{1}{m-1}\,, \end{cases} \quad (121)$$

where

$$P_{\min}^{(2,3)} \triangleq \min(P_s^{(2)}, P_s^{(3)}) \;\text{ and }\; P_{\max}^{(2,3)} \triangleq \max(P_s^{(2)}, P_s^{(3)})\,. \quad (122)$$

Note that $E(H)$, as a function of P_s, exhibits two plateaus in the intervals $[0, P_s^{(1)})$ and $(P_{\min}^{(2,3)}, P_{\max}^{(2,3)})$, respectively.

Substituting (33), (91), (101), and (103) into (119) yields

$$E(H_{\text{DF}_2})/c \;\approx\; \frac{l}{m}\,, \quad (123)$$

and

$$\frac{E(H_{\text{UF}})}{c} \;\approx\; \frac{2\,\frac{l}{m}\,(m-1)\,P_s}{\left[1-(m-1)\,\frac{\lambda}{\mu}\right]\,\left[1-(1-P_s)^{(m-1)\,C}\right]}\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \quad (124)$$

where C and λ/μ are determined by (3) and (5), respectively.

Remark 16: Substituting (92) and (103) into (119) yields

$$\frac{E(H_{\text{UF}})}{c} \;\approx\; \begin{cases} 2\,\frac{l}{m}\,\frac{1}{C}\,, & \text{for } P_s \ll P_s^{(2)} \\ 2\,\frac{l}{m}\,(m-1)P_s\,, & \text{for } P_s^{(2)} \ll P_s \ll \frac{1}{m-1}\,. \end{cases} \quad (125)$$

Remark 17: When P_s increases and approaches 1, it follows from (11), (83), and Remark 14 that $E(H)$ approaches $c\,l$. This is intuitively obvious because when $P_s = 1$, the entire amount $c\,l$ of user data stored in the system is lost owing to unrecoverable errors.

VI. RAID-6 Systems

Here we derive the reliability metrics for a RAID-6 system. When a storage device of a RAID-6 array fails, the C codewords stored in the array lose one of their symbols. Using the direct-path-approximation methodology, we proceed by considering only the subsequent potential data losses and device failures related to the affected array.

A. One Device Failure

The rebuild process attempts to restore the C codewords of the affected array sequentially. Let us consider such a codeword and let L_1 be the number of symbols permanently lost and I_1 be the number of symbols in the codeword with unrecoverable errors. The probability distribution of I_1 is determined by (12). Clearly, the symbol lost due to the device failure can be corrected by the RAID-6 capability only if at least $m-2$ of the remaining $m-1$ symbols can be read. Thus, $L_1 = 0$ if and only if $I_1 \leq 1$. Using (12), the probability q_1 that a codeword can be restored is

$$q_1 = P(L_1 = 0) = P(I_1 \leq 1) \overset{(12)}{=} [1+(m-2)P_s](1-P_s)^{m-2}, \quad (126)$$

which for very small values of P_s implies that

$$q_1 \approx \begin{cases} 1 - \frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_s \ll \sqrt{\frac{2}{(m-1)(m-2)}} \\ 0\,, & \text{for } P_s \gg \sqrt{\frac{2}{(m-1)(m-2)}}\,. \end{cases} \quad (127)$$

Note that, if a codeword is corrupted, then at least one of its l user-data symbols is lost. We now deduce that the conditional probability $P_{\text{UF}|1}$ of encountering an unrecoverable failure during the rebuild process of the C codewords in the case of one device failure is

$$P_{\text{UF}|1} = 1 - q_1^C \overset{(126)}{=} 1 - [1+(m-2)P_s]^C(1-P_s)^{(m-2)\,C}\,, \quad (128)$$

Remark 18: For very small values of P_s, q_1 is approximated by (127). Consequently, it follows from (128) that

$$P_{\text{UF}|1} \approx \begin{cases} C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,. \end{cases} \quad (129)$$

where $P_{s^*}^{(4)}$ is obtained from the approximation (129)

$$P_{\text{UF}|1} \approx \frac{(m-1)(m-2)}{2}\, C\, P_s^2 = 1 \tag{130}$$

as follows:

$$\boxed{P_{s^*}^{(4)} \triangleq \sqrt{\frac{2}{C(m-1)(m-2)}}}\,. \tag{131}$$

Note also that, for $P_s \ll \sqrt{\frac{2}{(m-1)(m-2)}}$ and from (18) and (127), the expected number $E(N_{\text{UF}|1})$ of corrupted codewords is

$$E(N_{\text{UF}|1}) \approx C\, \frac{(m-1)(m-2)}{2}\, P_s^2\,. \tag{132}$$

In particular, for $P_s = P_{s^*}^{(4)}$, it holds that $E(N_{\text{UF}|1}) \approx 1$ and this, combined with the fact that $P_{\text{UF}|1} \approx 1$, implies that one of the C codewords is almost surely corrupted.

The expected number $E(N_{\text{UF}|1}|N_{\text{UF}|1} \geq 1)$ of corrupted codewords, given that such codewords exist, is derived using (23), (129), and (132) as follows:

$$\begin{aligned} E(N_{\text{UF}|1}|N_{\text{UF}|1} \geq 1) &= \frac{E(N_{\text{UF}|1})}{P(N_{\text{UF}|1} \geq 1)} = \frac{E(N_{\text{UF}|1})}{P_{\text{UF}|1}} \\ &\approx \begin{cases} 1\,, & \text{for } P_s \ll P_{s^*}^{(4)} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{s^*}^{(4)} \ll P_s \ll \sqrt{\frac{2}{(m-1)(m-2)}}\,. \end{cases} \end{aligned} \tag{133}$$

When $I_1 > 1$, the number L_1 of lost symbols is $I_1 + 1$. Consequently, the expected number $E(L_1)$ of lost symbols is

$$\begin{aligned} E(L_1) &= \sum_{i=2}^{m-1} (i+1)\, P(I_1 = i) \\ &= E(I_1) + 1 - P(I_1 = 0) - 2\, P(I_1 = 1) \\ &\overset{(12)(13)}{=} 1 - (1-P_s)^{m-1} + (m-1)\, P_s\, [1 - 2(1-P_s)^{m-2}]\,. \end{aligned} \tag{134}$$

Remark 19: From (134), it follows that

$$E(L_1) \approx \frac{3}{2}\,(m-1)\,(m-2)\,P_s^2\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{135}$$

In particular, the expected number $E(L_1|L_1 > 0)$ of lost symbols, given that the codeword cannot be restored, is

$$\begin{aligned} E(L_1|L_1 > 0) = \frac{E(L_1)}{P(L_1 > 0)} = \frac{E(L_1)}{1 - P(L_1 = 0)} &\overset{(126)}{=} \frac{E(L_1)}{1 - q_1} \\ &\overset{(127)(135)}{\approx} 3\,, \qquad \text{for } P_s \ll \frac{1}{m-1}\,. \end{aligned} \tag{136}$$

The probability $P_{\text{UF},1}$ of data loss due to unrecoverable failures during rebuild in the case of one device failure is obtained from (37), that is,

$$P_{\text{UF},1} \approx \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,(1 - q_1^C)\,, \tag{137}$$

where q_1 in the case of RAID-6 is determined by (126).

Corollary 9: It holds that

$$P_{\text{UF},1} \approx \begin{cases} P_{\text{DF},1}\,\dfrac{(m-1)(m-2)}{2}\,C\,P_s^2\,, & \text{for } P_s \ll P_{s^*}^{(4)} \\ P_{\text{DF},1}\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,, \end{cases} \tag{138}$$

where $P_{\text{DF},1}$ and $P_{s^*}^{(4)}$ are given by (34) and (131), respectively.

Proof: Immediate by substituting (129) into (36). ■

We now proceed to assess the amount of data loss. The expected number $E(S_{\text{U}|1})$ of symbols lost due to unrecoverable failures during rebuild, given one device failure, is determined by (41). Substituting (134) into (41) yields

$$E(S_{\text{U}|1}) = C\,\{1 - (1-P_s)^{m-1} + (m-1)\,P_s\,[1 - 2(1-P_s)^{m-2}]\}\,. \tag{139}$$

Remark 20: For small values of P_s, $E(L_1)$ is approximated by (135). Consequently, it follows from (41) that

$$E(S_{\text{U}|1}) \approx \frac{3}{2}\,C\,(m-1)\,(m-2)\,P_s^2\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{140}$$

The expected number $E(S_{\text{U},1})$ of symbols lost due to unrecoverable failures during rebuild, in the case of one device failure during rebuild, is determined by (44), and consequently by (47), that is,

$$E(S_{\text{U},1}) \approx C\,E(L_1)\left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,, \tag{141}$$

where $E(L_1)$ is determined by (134).

Remark 21: The relations given in (38) and (45), that is,

$$P_{\text{UF},1} \approx P_{\text{UF}|1} = 1 - q_1^C \tag{142}$$

and

$$E(S_{\text{U},1}) \approx E(S_{\text{U}|1}) = C\,E(L_1)\,, \tag{143}$$

hold for both RAID-5 and RAID-6 redundancy schemes, where q_1 is given by (14) and (126), respectively, and $E(L_1)$ by (26) and (134), respectively. In fact, these two relations are general because they apply to any redundancy scheme with the corresponding q_1 and $E(L_1)$ measures evaluated accordingly.

Remark 22: From (135) and (143), it follows that

$$E(S_{\text{U},1}) \approx \frac{3}{2}\,C\,(m-1)\,(m-2)\,P_s^2\,, \quad \text{for } P_s \ll \frac{1}{m-1}\,. \tag{144}$$

B. Two Device Failures

As discussed in Section V-B, a subsequent (second) device failure may occur during the rebuild process, which is triggered by the first device failure. In particular, such a failure may occur during reconstruction of the jth $(1 \leq j \leq C)$ codeword with a probability of $P_{\text{DF}_2}(j)$ determined by (35).

The rebuild process attempts to restore the jth codeword as well as the remaining $C-j$ codewords of the affected array sequentially. Let us consider such a codeword and let L_2 be the number of symbols permanently lost and I_2 be the number

of symbols in the codeword with unrecoverable errors. Clearly, I_2 is binomially distributed with parameter P_s, that is

$$P(I_2 = i) = \binom{m-2}{i} P_s^i (1-P_s)^{m-2-i}, \text{ for } i = 0, \ldots, m-2 , \tag{145}$$

such that

$$E(I_2) = \sum_{i=1}^{m-2} i\, P(I_2 = i) \; = \; (m-2)\, P_s . \tag{146}$$

The symbols lost due to the two device failures can be corrected by the RAID-6 capability only if the remaining $m-2$ symbols can be read. Thus, $L_2 = 0$ if and only if $I_2 = 0$. Using (145), the probability q_2 that a codeword can be restored is

$$q_2 = P(L_2 = 0) = P(I_2 = 0) = (1-P_s)^{m-2} \overset{(51)}{=} p_2 , \tag{147}$$

which for very small values of P_s implies that

$$q_2 \approx \begin{cases} 1-(m-2)\, P_s , & \text{for } P_s \ll \frac{1}{m-2} \\ 0 , & \text{for } P_s \gg \frac{1}{m-2} . \end{cases} \tag{148}$$

When $I_2 > 0$, the number L_2 of lost symbols is $I_1 + 2$. Consequently, the expected number $E(L_2)$ of lost symbols is

$$\begin{aligned} E(L_2) &= \sum_{i=1}^{m-2} (i+2)\, P(I_2 = i) \\ &= E(I_2) + 2 \sum_{i=1}^{m-2} P(I_2 = i) = E(I_2) + 2\,(1-q_2) \\ &\overset{(146)(147)}{=} (m-2)\, P_s + 2\,[1-(1-P_s)^{m-2}] . \end{aligned} \tag{149}$$

Remark 23: From (149), it follows that

$$E(L_2) \approx \; 3\,(m-2)\, P_s , \quad \text{for } P_s \ll \frac{1}{m-2} . \tag{150}$$

In particular, the expected number $E(L_2 | L_2 > 0)$ of lost symbols, given that the codeword cannot be restored, is

$$\begin{aligned} E(L_2 | L_2 > 0) = \frac{E(L_2)}{P(L_2 > 0)} = \frac{E(L_2)}{1 - P(L_2 = 0)} &\overset{(147)}{=} \frac{E(L_2)}{1-q_2} \\ &\overset{(127)(135)}{\approx} \; 3 , \qquad \text{for } P_s \ll \frac{1}{m-1} . \end{aligned} \tag{151}$$

As discussed in Section V-B, the C codewords are divided into two sets: $S_{1,j}$, the set of $j-1$ codewords already considered for reconstruction, and $S_{2,j}$, the set of the remaining $C-j+1$ codewords. In contrast to RAID-5, the codewords in $S_{2,j}$ could be reconstructed by the RAID-6 capability. Therefore, the probability that no unrecoverable failures occur in $S_{2,j}$ is q_2^{C-j+1}, which implies that the probability $P_{\text{UF in } S_{2,j}|2}$ that an unrecoverable failure occurs in $S_{2,j}$ is

$$P_{\text{UF in } S_{2,j}|2} = 1 - q_2^{C-j+1} . \tag{152}$$

Furthermore, such unrecoverable failures entail the loss of user data. Let us denote by $N_{\text{UF in } S_{2,j}|2}$ the number of such corrupted codewords. Then it holds that

$$E(N_{\text{UF in } S_{2,j}|2}) = (C-j+1)\,(1-q_2) . \tag{153}$$

Also, the probability $P_{\text{UF}|2}(j)$ that an unrecoverable failure occurs, given two device failures, is

$$P_{\text{UF}|2}(j) = 1 - q_1^{j-1}\, q_2^{C-j+1} . \tag{154}$$

As discussed in Section V-B, the second device failure divides the C codewords into two sets: S_1, the set of codewords already considered for reconstruction when the second device failure occurs, and S_2, the set of the remaining codewords. The probability $P_{\text{UF in } S_1|2}$ that an unrecoverable failure occurs in S_1 is given by (56), where q_1 is determined by (126).

Corollary 10: It holds that

$$P_{\text{UF in } S_1|2} \approx \begin{cases} \dfrac{(C-1)(m-1)(m-2)}{4}\, P_s^2 , & \text{for } P_s \ll P_{s^*} \\ 1 , & \text{for } P_s \gg P_{s^*} , \end{cases} \tag{155}$$

where P_{s^*} is obtained from the approximation (155)

$$P_{\text{UF in } S_1|2} \approx \frac{(C-1)(m-1)(m-2)}{4}\, P_s^2 = 1 \tag{156}$$

and considering that C is large, as follows:

$$P_{s^*} = \frac{2}{\sqrt{C(m-1)(m-2)}} \approx P_{s^*}^{(4)} . \tag{157}$$

Proof: See Appendix D. ■

Remark 24: For $P_s \ll P_{s^*}^{(4)}$, and considering that C is large, it follows from (129) and (155) that

$$P_{\text{UF in } S_1|2} \approx \frac{1}{2}\, P_{\text{UF}|1} , \quad \text{for } P_s \ll P_{s^*}^{(4)} . \tag{158}$$

This is intuitively obvious because, according to (31), when the second failure occurs, half the codewords on the average have already been considered for reconstruction, that is, $E(|S_1|) \approx C/2$.

The probability $P_{\text{UF in } S_2|2}$ that an unrecoverable failure occurs in S_2 is obtained by unconditioning (152) on j as follows:

$$\begin{aligned} P_{\text{UF in } S_2|2} &= \sum_{j=1}^{C} P_{\text{UF in } S_{2,j}|2} P(j|\text{DF}_2, R) \\ &\overset{(30)}{\approx} \sum_{j=1}^{C} (1 - q_2^{C-j+1})\, \frac{1}{C} = 1 - \frac{q_2}{C}\, \frac{1-q_2^C}{1-q_2} , \end{aligned} \tag{159}$$

which does not depend on the rebuild duration.

Corollary 11: It holds that

$$P_{\text{UF in } S_2|2} \approx \begin{cases} \dfrac{(C+1)(m-2)}{2}\, P_s , & \text{for } P_s \ll P_{s^*}^{(2)} \\ 1 , & \text{for } P_s \gg P_{s^*}^{(2)} , \end{cases} \tag{160}$$

where $P_{s^*}^{(2)}$ is obtained from the approximation (160)

$$P_{\text{UF in } S_2|2} \approx \frac{(C+1)(m-2)}{2}\, P_s = 1 \tag{161}$$

and is

$$\boxed{P_{s^*}^{(2)} \triangleq \frac{1}{C} \cdot \frac{2}{m-2}} . \tag{162}$$

Proof: Immediate from Corollary 5 and by recognizing that $P_{\text{UF in } S_2|2}$ derived by (159) is equal to the corresponding measure obtained in the case of RAID-5 as expressed by (61) because, according to (147), $q_2 = p_2$. ■

Remark 25: From (148) and for $P_s \ll 1/(m-2)$, it follows that $q_2 \approx 1$. Furthermore, $\log(q_2) = -(1-q_2) + O((1-q_2)^2) \approx -(1-q_2) \approx -(1-q_2)/q_2$. Consequently, substituting the term $q_2/(1-q_2)$ on the right-hand side of (159) with $-1/\log(q_2)$ yields

$$P_{\text{UF in } S_2|2} \approx 1 + \frac{1-q_2^C}{C\log(q_2)} = 1 + \frac{1-q_2^C}{\log(q_2^C)}, \quad (163)$$

where q_2 is determined by (147).

The expected number $E(N_{\text{UF in } S_1|2})$ of corrupted codewords in S_1 is determined by (59). Also, the expected number $E(N_{\text{UF in } S_2|2})$ of corrupted codewords in S_2 is obtained by unconditioning (153) on j as follows:

$$\begin{aligned} E(N_{\text{UF in } S_2|2}) &= \sum_{j=1}^{C} E(N_{\text{UF in } S_{2,j}|2}) P(j|\text{DF}_2, R) \\ &\overset{(30)}{\approx} \sum_{j=1}^{C} (C-j+1)(1-q_2)\frac{1}{C} \\ &= \frac{C-1}{2}(1-q_2) \approx \frac{C}{2}(1-q_2) \quad (164) \\ &\overset{(148)}{\approx} \frac{C}{2}(m-2)P_s, \text{ for } P_s \ll \frac{1}{m-2}. \quad (165) \end{aligned}$$

In particular, for $P_s = P_{s^*}^{(2)}$ as determined by (162), it holds that $E(N_{\text{UF in } S_2|2}) \approx 1$ and this, combined with the fact that $P_{\text{UF in } S_2|2} \approx 1$ as derived by (160), implies that one of the codewords in S_2 is almost surely corrupted. Also, (127) implies that $1-q_1 = 2(m-1)/[(m-2)C^2]$. Thus, from (59), it follows that $E(N_{\text{UF in } S_1|2}) \approx (m-1)/[(m-2)C]$, which is negligible compared with $E(N_{\text{UF in } S_2|2})$.

The expected number $E(N_{\text{UF in } S_2|2}|N_{\text{UF in } S_2|2} \geq 1)$ of corrupted codewords in S_2, given that such codewords exist, is

$$E(N_{\text{UF in } S_2|2}|N_{\text{UF in } S_2|2} \geq 1) = \frac{E(N_{\text{UF in } S_2|2})}{P_{\text{UF in } S_2|2}} \quad (166)$$

$$\overset{(160)(165)}{\approx} \begin{cases} 1, & \text{for } P_s \ll P_{s^*}^{(2)} \\ \frac{C}{2}(m-2)P_s, & \text{for } P_{s^*}^{(2)} \ll P_s \ll \frac{1}{m-1}. \end{cases} \quad (167)$$

The probability $P_{\text{UF}|2}$ of a data loss due to unrecoverable failures given two device failures and a rebuild duration of R is obtained by unconditioning (154) on j and using (30) as follows:

$$\begin{aligned} P_{\text{UF}|2} &= \sum_{j=1}^{C} P_{\text{UF}|2}(j)\, P(j|\text{DF}_2, R) \\ &\approx \sum_{j=1}^{C} \left(1 - q_1^{j-1} q_2^{C-j+1}\right)\frac{1}{C} = 1 - \frac{q_2}{C}\frac{q_1^C - q_2^C}{q_1 - q_2}, \end{aligned} \quad (168)$$

which does not depend on the rebuild duration.

Corollary 12: It holds that

$$P_{\text{UF}|2} \approx \begin{cases} \frac{(C+1)(m-2)}{2} P_s, & \text{for } P_s \ll P_{s^*}^{(2)} \\ 1, & \text{for } P_s \gg P_{s^*}^{(2)}, \end{cases} \quad (169)$$

where $P_{s^*}^{(2)}$ is determined by (162).

Proof: See Appendix E. ■

Remark 26: It follows from (155) and (160) that $P_{\text{UF in } S_1|2} \ll P_{\text{UF in } S_2|2}$ because the former is of the order $O(P_s^2)$, whereas the latter is of the order $O(P_s)$. Consequently, in conjunction with two device failures, an unrecoverable failure is more likely to be encountered after the second device failure and not before. This in turn implies that

$$P_{\text{UF}|2} \approx P_{\text{UF in } S_2|2}, \quad (170)$$

which is confirmed by comparing (160) with (169).

A subsequent (third) device failure may occur during the rebuild process of the remaining $C-j$ codewords, the duration of which is equal to $[(C-j)/C]R$. The probability $P_{\text{DF}_3}(j|R)$ that one of the $m-2$ remaining devices in the array fails during this rebuild depends on its duration and the aggregate failure rate of the $m-2$ highly reliable devices, and is

$$P_{\text{DF}_3}(j|R) \approx \frac{C-j}{C}(m-2)\lambda R. \quad (171)$$

Consequently, for a rebuild time duration of R, the probability $P_{\text{DF},2}(j|R)$ of encountering a second device failure during reconstruction of the jth codeword and not encountering a third device failure during the remaining rebuild time is determined by the product of $P_{\text{DF}_2}(j|R)$ and $1 - P_{\text{DF}_3}(j|R)$, that is,

$$P_{\text{DF},2}(j|R) \approx \frac{(m-1)\lambda R}{C}\left[1 - \frac{C-j}{C}(m-2)\lambda R\right], \quad \text{for } j = 1, 2, \ldots, C. \quad (172)$$

Therefore, the probability $P_{\text{DF},2}(j)$ of encountering a second device failure during reconstruction of the jth codeword and not encountering a third device failure during the remaining rebuild time is obtained by unconditioning (172) on R using (4) as follows:

$$P_{\text{DF},2}(j) \approx \frac{(m-1)}{C}\frac{\lambda}{\mu}\left[1 - \frac{C-j}{C}(m-2)\frac{\lambda}{\mu}f_R\right], \quad \text{for } j = 1, 2, \ldots, C, \quad (173)$$

where

$$f_R \triangleq \frac{E(R^2)}{[E(R)]^2} \geq 1. \quad (174)$$

Thus, the probability $P_{\text{DF},2}$ of two device failures during rebuild is obtained by (173) as follows:

$$\begin{aligned} P_{\text{DF},2} &\approx \sum_{j=1}^{C} P_{\text{DF},2}(j) \quad (175) \\ &\approx (m-1)\frac{\lambda}{\mu}\left[1 - \frac{C-1}{C}(m-2)\frac{\lambda}{\mu}f_R\right] \quad (176) \\ &\overset{(5)}{\approx} (m-1)\frac{\lambda}{\mu} \overset{(33)}{\approx} P_{\text{DF}_2}. \quad (177) \end{aligned}$$

The probability $P_{\text{UF},2}$ of data loss due to unrecoverable failures in conjunction with two device failures is obtained by unconditioning (154) on j and using (173) as follows:

$$\begin{aligned} P_{\text{UF},2} &= \sum_{j=1}^{C} P_{\text{UF}|2}(j)\, P_{\text{DF},2}(j) \\ &\approx \sum_{j=1}^{C} (1 - q_1^{j-1}\, q_2^{C-j+1})\, \frac{(m-1)}{C}\, \frac{\lambda}{\mu} \\ &\qquad \cdot \left[1 - \frac{C-j}{C}\,(m-2)\,\frac{\lambda}{\mu}\, f_R\right] \\ &= \left(1 - \frac{q_2}{C}\,\frac{q_1^C - q_2^C}{q_1 - q_2}\right)(m-1)\,\frac{\lambda}{\mu} \\ &\quad - \left[\frac{C\,(C-1)}{2} - q_2^2\,\frac{q_1^C - C\,q_1\,q_2^{C-1} + (C-1)\,q_2^C}{(q_1-q_2)^2}\right] \cdot \\ &\qquad \cdot \frac{(m-1)(m-2)}{C^2}\left(\frac{\lambda}{\mu}\right)^2 f_R\,, \end{aligned} \tag{178}$$

where f_R is determined by (174).

Corollary 13: It holds that

$$P_{\text{UF},2} \approx \begin{cases} A\,P_s\,, & \text{for } P_s \ll P_{s^*}^{(2)} \\ P_{\text{DF},2}\,, & \text{for } P_s \gg P_{s^*}^{(2)}\,, \end{cases} \tag{179}$$

where

$$A = \frac{C+1}{2}\left[1 - \frac{2}{3}\,\frac{C-1}{C}\,(m-2)\,\frac{\lambda}{\mu}\,f_R\right](m-1)(m-2)\,\frac{\lambda}{\mu} \tag{180}$$

$$\approx \frac{C}{2}\left[1 - \frac{2}{3}\,(m-2)\,\frac{\lambda}{\mu}\,f_R\right](m-1)(m-2)\,\frac{\lambda}{\mu} \tag{181}$$

$$\overset{(5)}{\approx} \frac{C}{2}\,(m-1)(m-2)\,\frac{\lambda}{\mu}\,, \tag{182}$$

and $P_{\text{DF},2}$ and $P_{s^*}^{(2)}$ are given by (176) and (162), respectively.

Proof: See Appendix E. ■

Remark 27: Note that the minuend of the difference shown on the right-hand side of (178) is much greater than the subtrahend because the former is of the order $O(\lambda/\mu)$, whereas the latter is of the order $O((\lambda/\mu)^2)$ and is also further reduced owing to division by the large number C^2. Consequently, from (178), it follows that

$$P_{\text{UF},2} \approx \left(1 - \frac{q_2}{C}\,\frac{q_1^C - q_2^C}{q_1 - q_2}\right)(m-1)\,\frac{\lambda}{\mu} \tag{183}$$

$$\overset{(33)(168)}{\approx} P_{\text{UF}|2}\, P_{\text{DF}_2} \tag{184}$$

$$\overset{(170)}{\approx} P_{\text{UF in } S_2|2}\, P_{\text{DF}_2} \tag{185}$$

$$\overset{(33)(163)}{\approx} \left[1 + \frac{1 - q_2^C}{\log(q_2^C)}\right](m-1)\,\frac{\lambda}{\mu}\,, \tag{186}$$

where q_1 and q_2 are given by (126) and (147), respectively.

Remark 28: According to (131) and (162) and for large values of C, it holds that $P_{s^*}^{(2)} \ll P_{s^*}^{(4)}$. Consequently, from (142), (129), (179), and (182), and given that $P_{s^*}^{(2)} \sim 1/C \ll \lambda/\mu$, it holds that

$$P_{\text{UF},1} \ll P_{\text{UF},2}\,, \qquad \text{for } P_s \ll P_{s^*}^{(2)}\,. \tag{187}$$

We now proceed to assess the amount of data loss. As discussed above, each of the $C-j+1$ codewords in $S_{2,j}$ loses an expected number of $E(L_2)$ symbols. Thus, the expected total number $E(S_{\text{U},2}^{+}\,|\,\text{DF}_2 \text{ at } j)$ of symbols stored in these $C-j+1$ codewords and lost due to unrecoverable failures is

$$E(S_{\text{U},2}^{+}\,|\,\text{DF}_2 \text{ at } j) = (C-j+1)\,E(L_2)\,, \tag{188}$$

where $E(L_2)$ is determined by (149). Furthermore, each of the $j-1$ codewords in $S_{1,j}$ loses an expected number of $E(L_1)$ symbols. Consequently, the expected total number $E(S_{\text{U},2}^{-}\,|\,\text{DF}_2 \text{ at } j)$ of symbols stored in these $j-1$ codewords and lost due to unrecoverable failures is

$$E(S_{\text{U},2}^{-}\,|\,\text{DF}_2 \text{ at } j) = (j-1)\,E(L_1)\,, \tag{189}$$

where $E(L_1)$ is determined by (134). Unconditioning (188) and (189) on the event of a device failure during the reconstruction of the jth codeword, and using (173), yields after some manipulations

$$\begin{aligned} E(S_{\text{U},2}^{+}) &= \sum_{j=1}^{C} E(S_{\text{U},2}^{+}\,|\,\text{DF}_2 \text{ at } j) P_{\text{DF},2}(j) \\ &= \sum_{j=1}^{C} (C-j+1)\,E(L_2)\,P_{\text{DF},2}(j) \\ &\approx \frac{1}{2}\,(C+1)\,E(L_2)\,(m-1)\,\frac{\lambda}{\mu} \\ &\qquad - \frac{(C+1)\,E(L_2)}{3}\cdot\frac{C-1}{C}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\,, \end{aligned} \tag{190}$$

and

$$\begin{aligned} E(S_{\text{U},2}^{-}) &= \sum_{j=1}^{C} E(S_{\text{U},2}^{-}\,|\,\text{DF}_2 \text{ at } j) P_{\text{DF},2}(j) \\ &= \sum_{j=1}^{C} (j-1)\,E(L_1)\,P_{\text{DF},2}(j) \\ &\approx \frac{1}{2}\,(C-1)\,E(L_1)\,(m-1)\,\frac{\lambda}{\mu} \\ &\qquad - \frac{(C-2)\,E(L_1)}{6}\cdot\frac{C-1}{C}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\,, \end{aligned} \tag{191}$$

where $E(L_1)$, $E(L_2)$, and f_R are determined by (134), (149), and (174), respectively.

The expected total number $E(S_{\text{U},2})$ of symbols lost due to unrecoverable errors in conjunction with two device failures is

$$E(S_{\text{U},2}) = E(S_{\text{U},2}^{+}) + E(S_{\text{U},2}^{-})\,, \tag{192}$$

where $E(S_{\text{U},2}^{+})$ and $E(S_{\text{U},2}^{-})$ are determined by (190) and (191), respectively.

Remark 29: From (135) and (150), it follows that $E(L_1) \ll E(L_2)$ because the former is of the order $O(P_s^2)$, whereas the latter is of the order $O(P_s)$. From (190) and (191), it follows that $E(S_{\text{U},2}^{+}) \sim E(L_2)$ and $E(S_{\text{U},2}^{-}) \sim E(L_1)$, respectively. Thus,

$$E(S_{\text{U},2}^{+}) \gg E(S_{\text{U},2}^{-})\,. \tag{193}$$

Corollary 14: It holds that

$$E(S_{\mathrm{U},2}) \approx E(S^{+}_{\mathrm{U},2}) \,. \tag{194}$$

Proof: Immediate from (192) and (193). ■

Remark 30: For small values of P_s and large values of C, from (190) and (194), and using (5) and (150), it follows that

$$E(S_{\mathrm{U},2}) \approx \frac{1}{2}\, C\, E(L_2)\, (m-1)\, \frac{\lambda}{\mu} \tag{195}$$

$$\approx \frac{3}{2}\, C\, (m-1)(m-2)\, \frac{\lambda}{\mu}\, P_s \,, \quad \text{for } P_s \ll \frac{1}{m-2} \,. \tag{196}$$

Remark 31: From (143) and (191), it follows that $E(S_{\mathrm{U},1}) \gg E(S^{-}_{\mathrm{U},2})$ because $E(S^{-}_{\mathrm{U},2})$ is of the order $O(\lambda/\mu)$, which is very small, whereas $E(S_{\mathrm{U},1})$ is not. Furthermore, from (144) and (196), it follows that

$$\begin{aligned} &E(S_{\mathrm{U},1}) \ll E(S_{\mathrm{U},2}) \\ &\Leftrightarrow\ \frac{3}{2}\, C\, (m-1)\, (m-2)\, P_s^2 \ll \frac{3}{2}\, C\, (m-1)(m-2)\, \frac{\lambda}{\mu}\, P_s \\ &\Leftrightarrow\ P_s \ll P^{(5)}_{s^*} \,, \end{aligned} \tag{197}$$

where

$$\boxed{P^{(5)}_{s^*} \triangleq \frac{\lambda}{\mu}} \,. \tag{198}$$

Consequently, for very small values of P_s, the symbols lost due to unrecoverable failures are predominately encountered during the rebuild phase after a second device failure.

C. Three Device Failures

As discussed in Section VI-B, after a second device failure during reconstruction of the jth $(1 \le j \le C)$ codeword, a subsequent (third) device failure may occur during the rebuild process of the remaining $C-j$ codewords, which form the $S_{2,j}$ set. In particular, such a failure may occur during reconstruction of the ith $(j+1 \le i \le C)$ codeword, which divides the codewords of the $S_{2,j}$ set into two subsets:

$S^{-}_{2,j,i}$: the set of $i-j$ codewords in $S_{2,j}$ already considered for reconstruction prior to the third device failure, and

$S^{+}_{2,j,i}$: the set of the remaining $C-i+1$ codewords in $S_{2,j}$, none of which can be reconstructed.

Analogous to (32), the probability $P_{\mathrm{DF}_3}(i|R)$ that a third device failure occurs during reconstruction of the ith $(j+1 \le i \le C)$ codeword during the rebuild process in conjunction with two device failures, and given a rebuild time of R, is

$$P_{\mathrm{DF}_3}(i|j,R) \approx \frac{(m-2)\,\lambda\, R}{C} \,, \quad \text{for } i = j+1, j+2, \ldots, C \,. \tag{199}$$

Consequently, the probability $P_{j,i}(R)$ of a second device failure during reconstruction of the jth codeword and a third device failure during reconstruction of the ith codeword, and given a rebuild time of R, is the product of $P_{\mathrm{DF}_2}(j|R)$ and $P_{\mathrm{DF}_3}(i|R)$ obtained from (32) and (199) as follows:

$$P_{j,i}(R) \triangleq P_{\mathrm{DF}_2}(j|R)\, P_{\mathrm{DF}_3}(i|j,R) \approx \frac{(m-1)(m-2)}{C^2}\, \lambda^2\, R^2 \,, \quad \text{for } j = 1, 2, \ldots, C \text{ and } i = j+1, j+2, \ldots, C \,. \tag{200}$$

Therefore, the probability $P_{j,i}$ of a second device failure during reconstruction of the jth codeword and a third device failure during reconstruction of the ith codeword is obtained by unconditioning (200) on R and using (4) and (174) as follows:

$$P_{j,i} = E(P_{j,i}(R)) \approx \frac{(m-1)(m-2)}{C^2}\, \left(\frac{\lambda}{\mu}\right)^2 f_R \,, \quad \text{for } j = 1, 2, \ldots, C \text{ and } i = j+1, j+2, \ldots, C \,. \tag{201}$$

Also, the probability $P_{\mathrm{DF}_3}(j|R)$ that, during reconstruction of the codewords in $S_{2,j}$, one of the $m-2$ remaining devices in the array fails is obtained from (199) as follows:

$$\begin{aligned} P_{\mathrm{DF}_3}(j|R) &= \sum_{i=j+1}^{C} P_{\mathrm{DF}_3}(i|j,R) \approx \sum_{i=j+1}^{C} \frac{(m-2)\,\lambda\, R}{C} \\ &= (C-j)\, \frac{(m-2)\,\lambda\, R}{C} \,, \end{aligned} \tag{202}$$

which is in agreement with (171).

The probability $P_{\mathrm{DF}_3}(R)$ of encountering three device failures given a rebuild duration of R is obtained by unconditioning (202) on j and using (32) as follows:

$$\begin{aligned} P_{\mathrm{DF}_3}(R) &\approx \sum_{j=1}^{C} P_{\mathrm{DF}_3}(j|R)\, P_{\mathrm{DF}_2}(j|R) \\ &\approx \sum_{j=1}^{C} \frac{C-j}{C^2}\, (m-1)(m-2)\, \lambda^2\, R^2 \\ &\approx \frac{C-1}{2\,C}\, (m-1)(m-2)\, \lambda^2\, R^2 \,. \end{aligned} \tag{203}$$

The probability P_{DF_3} of data loss due to three device failures, that is, the probability of two device failures during the rebuild process is obtained by unconditioning (203) on R and using (4) and (174) as follows:

$$P_{\mathrm{DF}_3} \approx \frac{C-1}{2\,C}\, (m-1)(m-2)\, \left(\frac{\lambda}{\mu}\right)^2 f_R \,. \tag{204}$$

In particular, for large values of C, it holds that

$$P_{\mathrm{DF}_3} \approx \frac{(m-1)(m-2)}{2}\, \left(\frac{\lambda}{\mu}\right)^2 f_R \,, \tag{205}$$

Note that the occurrence of the third device failure does not exclude the possibility of unrecoverable failures being encountered prior to its occurrence. Also, each of the $C-i+1$ codewords in $S^{+}_{2,j,i}$ can no longer be reconstructed. In particular, the three symbols of the ith codeword stored on the three failed devices can no longer be recovered and are lost. Furthermore, any of its remaining $m-3$ symbols with unrecoverable errors will also be lost, leading to unrecoverable failure. Thus, the probability p_3 of not encountering an unrecoverable failure due to unrecoverable errors in the remaining $m-3$ symbols is

$$p_3 = (1-P_s)^{m-3} \,, \tag{206}$$

which for very small values of P_s implies that

$$p_3 \approx \begin{cases} 1-(m-3)\, P_s \,, & \text{for } P_s \ll \frac{1}{m-3} \\ 0 \,, & \text{for } P_s \gg \frac{1}{m-3} \,. \end{cases} \tag{207}$$

The same applies for the remaining $C-i$ codewords. Therefore, the probability that no unrecoverable failures occur in $S_{1,j}$ is q_1^{j-1}, the probability that no unrecoverable failures occur in $S^-_{2,j,i}$ is q_2^{i-j}, and the probability that no unrecoverable failures occur in $S^+_{2,j,i}$ is p_3^{C-i+1}. Consequently, the probability $P_{\text{UF}|3}(j,i)$ that an unrecoverable failure occurs, given three device failures, is

$$P_{\text{UF}|3}(j,i) = 1 - q_1^{j-1}\, q_2^{i-j}\, p_3^{C-i+1}\ . \tag{208}$$

The probability $P_{\text{UF},3}$ of data loss due to unrecoverable failures in conjunction with three device failures is determined by the following proposition.

Proposition 1: It holds that

$$P_{\text{UF},3} = \left[\frac{C\,(C-1)}{2} - \frac{q_2\,p_3}{p_3-q_2}\left(\frac{q_1^C-p_3^C}{q_1-p_3} - \frac{q_1^C-q_2^C}{q_1-q_2}\right)\right] \cdot \\ \cdot\, \frac{(m-1)(m-2)}{C^2}\left(\frac{\lambda}{\mu}\right)^2 f_R\ . \tag{209}$$

Proof: See Appendix F. ■

Corollary 15: It holds that

$$P_{\text{UF},3} \approx \begin{cases} B\,P_s\ , & \text{for } P_s \ll P_{s^*}^{(2)} \\ P_{\text{DF}_3}\ , & \text{for } P_s \gg P_{s^*}^{(2)}\ , \end{cases} \tag{210}$$

where

$$B = \frac{(C-1)(C+1)}{6\,C}\,(2m-5)(m-1)(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R \tag{211}$$

$$\approx \frac{C}{6}\,(2m-5)(m-1)(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\ , \tag{212}$$

and P_{DF_3} and $P_{s^*}^{(2)}$ are given by (205) and (162), respectively.

Proof: See Appendix G. ■

Remark 32: From (210), it follows that for $P_s \gg P_{s^*}^{(2)}$, $P_{\text{UF}|3} = P_{\text{UF},3}/P_{\text{DF}_3} \approx 1$, which implies that a third device failure leads to data loss owing to unrecoverable failures. In this case, the probability of data loss is equal to that of a RAID-6 system in the absence of latent errors, that is, $P_{\text{UF},3} = P_{\text{DF}_3}$.

Remark 33: From (179) and (210), it follows that

$$P_{\text{UF},2} \gg P_{\text{UF},3}\ . \tag{213}$$

According to (182) and (212), it holds that $A \gg B$ because the former is of the order $O(\lambda/\mu)$, whereas the latter is of the order $O((\lambda/\mu)^2)$, and similarly, according to (177) and (205), $P_{\text{DF},2} \approx P_{\text{DF}_2} \gg P_{\text{DF}_3}$.

We now proceed to assess the amount of data loss. As discussed above, the three symbols of each of the $C-i+1$ codewords in $S^+_{2,j,i}$ stored on the three failed devices can no longer be recovered and are lost. Thus, the total number $S_{\text{D}}(j,i)$ of symbols stored on the three failed devices and lost is

$$S_{\text{D}}(j,i) = 3\,(C+1-i)\ . \tag{214}$$

Also, the expected number $E(S^+_{\text{U},3}\,|\,\text{DF}_2$ at j and DF_3 at $i)$ of symbols stored in these $C-i+1$ codewords and lost owing to unrecoverable errors is

$$E(S^+_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i) = (C+1-i)\,(m-3)\,P_s\ . \tag{215}$$

Moreover, each of the $i-j$ codewords in $S^-_{2,j,i}$ loses an expected number of $E(L_2)$ symbols. Consequently, the expected total number $E(S^-_{\text{U},3}\,|\,\text{DF}_2$ at j and DF_3 at $i)$ of symbols stored in these $j-i$ codewords and lost due to unrecoverable errors is

$$E(S^-_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i) = (i-j)\,E(L_2)\ . \tag{216}$$

Furthermore, each of the $j-1$ codewords in $S_{1,j}$ loses an expected number of $E(L_1)$ symbols. Consequently, the expected total number $E(S^{\odot}_{\text{U},3}\,|\,\text{DF}_2$ at j and DF_3 at $i)$ of symbols stored in these $j-1$ codewords and lost due to unrecoverable errors is

$$E(S^{\odot}_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i) = (j-1)\,E(L_1)\ . \tag{217}$$

Unconditioning (214), (215), (216), and (217) on the event of two device failures during the reconstruction of the jth and ith codewords, and using (201), yields

$$E(S_{\text{D}}) = \sum_{j=1}^{C}\sum_{i=j+1}^{C} S_{\text{D}}(j,i)\,P_{j,i} \\ \approx \frac{(C-1)\,(C+1)}{2\,C}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\ , \tag{218}$$

$$E(S^+_{\text{U},3}) = \sum_{j=1}^{C}\sum_{i=j+1}^{C} E(S^+_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i)\,P_{j,i} \\ \approx \frac{(C-1)\,(C+1)}{6\,C}\,(m-1)\,(m-2)\,(m-3)\left(\frac{\lambda}{\mu}\right)^2 f_R\,P_s\ , \tag{219}$$

$$E(S^-_{\text{U},3}) = \sum_{j=1}^{C}\sum_{i=j+1}^{C} E(S^-_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i)\,P_{j,i} \\ \approx \frac{(C-1)\,(C+1)}{6\,C}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\,E(L_2)\ , \tag{220}$$

and

$$E(S^{\odot}_{\text{U},3}) = \sum_{j=1}^{C}\sum_{i=j+1}^{C} E(S^{\odot}_{\text{U},3}\,|\,\text{DF}_2 \text{ at } j \text{ and } \text{DF}_3 \text{ at } i)\,P_{j,i} \\ \approx \frac{(C-1)\,(C-2)}{6\,C}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\,E(L_1)\ , \tag{221}$$

where $E(L_1)$, $E(L_2)$, and f_R are determined by (134), (149), and (174), respectively.

In particular, from (218) and considering that, for large values of C, $C+1 \approx C$, it holds that

$$E(S_{\text{D}}) \approx \frac{C}{2}\,(m-1)\,(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\ . \tag{222}$$

The expected total number $E(S_{\mathrm{U},3})$ of symbols lost due to unrecoverable errors in conjunction with two device failures is

$$E(S_{\mathrm{U},3}) = E(S^{+}_{\mathrm{U},3}) + E(S^{-}_{\mathrm{U},3}) + E(S^{\odot}_{\mathrm{U},3}) \,, \tag{223}$$

where $E(S^{+}_{\mathrm{U},3})$, $E(S^{-}_{\mathrm{U},3})$, and $E(S^{\odot}_{\mathrm{U},3})$ are determined by (219), (220), and (221), respectively.

Remark 34: From (219), (220), and (221), and by virtue of (135) and (150), it follows that $E(S^{-}_{\mathrm{U},3}) > E(S^{+}_{\mathrm{U},3})$ and $E(S^{-}_{\mathrm{U},3}) > E(S^{\odot}_{\mathrm{U},3})$. Also, from (195) and (220), it follows that $E(S_{\mathrm{U},2}) \gg E(S^{-}_{\mathrm{U},3})$ because $E(S_{\mathrm{U},2})$ is of the order $O(\lambda/\mu)$, whereas $E(S^{-}_{\mathrm{U},3})$ is of the order $O((\lambda/\mu)^2)$. From the discussion above, it follows that $E(S_{\mathrm{U},2}) \gg E(S_{\mathrm{U},3})$.

From (140), (143), (196), and (222), it follows that

$$E(S_{\mathrm{U},1}) \ll E(S_{\mathrm{D}}) \;\Leftrightarrow\; P_s \ll \sqrt{\frac{f_R}{3}}\, P^{(5)}_{s^*} \,, \tag{224}$$

$$E(S_{\mathrm{U},2}) \ll E(S_{\mathrm{D}}) \;\Leftrightarrow\; P_s \ll \frac{f_R}{3}\, P^{(5)}_{s^*} \,, \tag{225}$$

where $P^{(5)}_{s^*}$ is determined by (198).

Remark 35: From (81) and (198), it follows that $P^{(3)}_{s}$ and $P^{(5)}_{s^*}$ are of the same order.

Remark 36: From (218), it follows that the expected number $E(S_{\mathrm{D}})$ of symbols stored on the three failed devices and lost is of the order $O((\lambda/\mu)^2)$. Note that the above analysis does not exclude the possibility that additional device failures occur during rebuild. However, the corresponding expected number of additional lost symbols can be ignored because it is of the order $O((\lambda/\mu)^3)$, which is much smaller than $O((\lambda/\mu)^2)$.

D. Data Loss

Data loss during rebuild may occur because of an unrecoverable failure of one or more codewords in case of one, two or three device failures. These three mutually exclusive events imply that

$$P_{\mathrm{DL}} = P_{\mathrm{UF},1} + P_{\mathrm{UF},2} + P_{\mathrm{DF}_3} \,. \tag{226}$$

Substituting (137), (178), and (204) into (226) yields

$$\begin{aligned} P_{\mathrm{DL}} \approx{}& \left[1 - (m-1)\,\frac{\lambda}{\mu}\right]\,(1 - q_1^C\,) \\ &+ \left(1 - \frac{q_2}{C}\,\frac{q_1^C - q_2^C}{q_1 - q_2}\right)\,(m-1)\,\frac{\lambda}{\mu} \\ &- \left[\frac{(C-1)\,C}{2} - q_2^2\,\frac{q_1^C - C\,q_1\,q_2^{C-1} + (C-1)\,q_2^C}{(q_1-q_2)^2}\right]\cdot \\ &\qquad \cdot \frac{(m-1)(m-2)}{C^2}\left(\frac{\lambda}{\mu}\right)^2 f_R \,, \\ &+ \frac{C-1}{2\,C}\,(m-1)(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R \,. \end{aligned} \tag{227}$$

Thus, after some manipulations, (227) yields

$$\begin{aligned} P_{\mathrm{DL}} \approx{}& 1 - q_1^C\left[1 - (m-1)\,\frac{\lambda}{\mu}\right] - \frac{q_2}{C}\,\frac{q_1^C - q_2^C}{q_1 - q_2}\,(m-1)\,\frac{\lambda}{\mu} \\ &+ q_2^2\,\frac{q_1^C - C\,q_1\,q_2^{C-1} + (C-1)q_2^C}{(q_1-q_2)^2}\,\frac{(m-1)(m-2)}{C^2}\left(\frac{\lambda}{\mu}\right)^2 f_R \,, \end{aligned} \tag{228}$$

where q_1 and q_2 are given by (126) and (147), respectively.

Substituting (126) and (147) into (228) yields

$$\begin{aligned} P_{\mathrm{DL}} \approx{}& 1 - (1-P_s)^{C(m-2)} \\ &\cdot \Bigg\{ [1 + (m-2)P_s]^C \\ &+ \frac{m-1}{C\,(m-2)\,P_s}\,\frac{\lambda}{\mu}\,\Bigg[[1 + (m-2)P_s]^C[1 - C(m-2)P_s] \\ &\qquad - \frac{[1 + (m-2)P_s]^C - C(m-2)P_s - 1}{C\,P_s}\,\frac{\lambda}{\mu}\,f_R\Bigg]\Bigg\} \,. \end{aligned} \tag{229}$$

Corollary 16: For $P_s \ll P^{(2)}_{s^*}$, it holds that

$$P_{\mathrm{DL}} \approx P_{\mathrm{DF}_3} + P_{\mathrm{UF},2} \overset{(179)}{\approx} P_{\mathrm{DF}_3} + A\,P_s \tag{230}$$

$$\overset{(205)(182)}{\approx} \frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\left(\frac{\lambda}{\mu}\,f_R + C\,P_s\right), \tag{231}$$

where P_{DF_3} and A are given by (204) and (180), respectively.

Proof: Immediate from (226) by considering (187). ■

Remark 37: It follows from (179), (182), and (205) that the range $[0, P^{(1)}_{s^*})$ of P_s in which the probability $P_{\mathrm{UF},2}$ is much smaller than the probability P_{DF_3} of two device failures occurring during rebuild is obtained by

$$\begin{aligned} P_{\mathrm{UF},2} \ll P_{\mathrm{DF}_3} \;&\Leftrightarrow\; A\,P_s \ll \frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R \\ &\Leftrightarrow\; P_s \ll P^{(1)}_{s^*} \,, \end{aligned} \tag{232}$$

where

$$\boxed{P^{(1)}_{s^*} \triangleq \frac{1}{C}\cdot\frac{\lambda}{\mu}\cdot f_R} \,. \tag{233}$$

From the above, it follows that $P_{\mathrm{UF},2} \gg P_{\mathrm{DF}_3}$ for $P_s \gg P^{(1)}_{s^*}$ and, therefore, P_{DL} is dominated by $P_{\mathrm{UF},2}$.

Remark 38: From (169), it follows that $P_{\mathrm{UF}|2} \approx 1$ for $P_s \gg P^{(2)}_{s^*}$, which implies that a second device failure leads to data loss owing to unrecoverable failures. In this case, the probability of data loss is equal to that of a RAID-5 system in the absence of latent errors, that is, $P_{\mathrm{UF},2} = P_{\mathrm{DF}_2}$, as derived in (179) and (177).

Remark 39: From (131) and (162), it follows that $P^{(2)}_{s^*} \ll P^{(4)}_{s^*}$. Consequently, the range $[0, P^{(3)}_{s^*})$ of P_s in which probability $P_{\mathrm{UF},2}$ is much greater than probability $P_{\mathrm{UF},1}$ is obtained from (129), (142), and (179) as follows:

$$\begin{aligned} &P_{\mathrm{UF},1} \ll P_{\mathrm{UF},2} \\ &\Leftrightarrow \begin{cases} C\,\frac{(m-1)(m-2)}{2}\,P_s^2 \ll A\,P_s \,, & \text{for } P_s \ll P^{(2)}_{s^*} \ll P^{(4)}_{s^*} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2 \ll P_{\mathrm{DF},2} \,, & \text{for } P^{(2)}_{s^*} \ll P_s \ll P^{(4)}_{s^*} \end{cases} \end{aligned} \tag{234}$$

$$\overset{(177)(182)}{\Leftrightarrow} \begin{cases} C\,\frac{(m-1)(m-2)}{2}\,P_s^2 \ll \frac{C}{2}\,(m-1)(m-2)\,\frac{\lambda}{\mu}\,P_s \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2 \ll (m-1)\,\frac{\lambda}{\mu} \end{cases} \tag{235}$$

$$\Leftrightarrow\; P_s \ll P^{(3)}_{s^*} \,, \tag{236}$$

where, by virtue of (162),

$$P_{s^*}^{(3)} \triangleq \begin{cases} \dfrac{\lambda}{\mu}\,, & \text{for } \dfrac{\lambda}{\mu} \leq \dfrac{1}{C}\cdot\dfrac{2}{m-2} = P_{s^*}^{(2)} \\ \sqrt{\dfrac{2}{C(m-2)}\,\dfrac{\lambda}{\mu}}\,, & \text{for } \dfrac{\lambda}{\mu} \geq \dfrac{1}{C}\cdot\dfrac{2}{m-2} = P_{s^*}^{(2)} \end{cases}. \tag{237}$$

From the above discussion, it follows that for $P_s \gg P_{s^*}^{(3)}$, P_{DL} is dominated by $P_{\text{UF,1}}$.

Corollary 17: For $P_s \gg P_{s^*}^{(2)}$, it holds that

$$P_{\text{DL}} \approx P_{\text{UF,1}} + P_{\text{UF,2}} \tag{238}$$

$$\overset{(138)(179)}{\approx} \begin{cases} P_{\text{DF,2}} + P_{\text{DF,1}}\,\frac{(m-1)(m-2)}{2}\,C\,P_s^2\,, & \text{for } P_s \ll P_{s^*}^{(4)} \\ P_{\text{DF,2}} + P_{\text{DF,1}}\,, & \text{for } P_s \gg P_{s^*}^{(4)} \end{cases} \tag{239}$$

$$\overset{(34)(177)}{\approx} \begin{cases} (m-1)\left[\dfrac{\lambda}{\mu} + \dfrac{C\,(m-2)}{2}P_s^2\right], & \text{for } P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}, \end{cases} \tag{240}$$

where $P_{\text{DF,1}}$ and $P_{\text{DF,2}}$ are determined by (34) and (176), respectively.

Proof: Immediate from (226) by considering (232) and $P_s \gg P_{s^*}^{(2)} \gg P_{s^*}^{(1)}$ and $P_{\text{DF,1}} \approx 1$. ■

Corollary 18: For small values of λ/μ, it holds that

$$P_{\text{DL}} \approx P_{\text{UF|1}} + P_{\text{UF,2}} + P_{\text{DF}_3} \tag{241}$$

$$\overset{(184)}{\approx} P_{\text{UF|1}} + P_{\text{UF|2}}\,P_{\text{DF}_2} + P_{\text{DF}_3} \tag{242}$$

$$\overset{(128)(186)(205)}{\approx} 1 - q_1^C + \left[1 + \frac{1-q_2^C}{\log(q_2^C)}\right](m-1)\,\frac{\lambda}{\mu} + \frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R\,. \tag{243}$$

where q_1, q_2, and f_R are determined by (126), (147), and (174) respectively.

Proof: Immediate by substituting (142) into (226). ■

According to (162) and (237), it holds that $P_{s^*}^{(2)} \lessgtr P_{s^*}^{(3)} \Leftrightarrow \lambda/\mu \gtrless 2/[C\,(m-2)] = P_{s^*}^{(2)}$. Depending on the values of λ/μ, m and C, we consider the following two cases:

Case 1: $\frac{\lambda}{\mu} \geq \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. From (233), (162), (237), and (131), it holds that $P_{s^*}^{(1)} \ll P_{s^*}^{(2)} \leq P_{s^*}^{(3)} < P_{s^*}^{(4)}$. Also, from (241), and considering (129), (177), (179), (182), and (205), it follows that

$$P_{\text{DL}} \approx \begin{cases} \frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R\,, & \text{for } P_s \ll P_{s^*}^{(1)} \\ C\,\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,P_s\,, & \text{for } P_{s^*}^{(1)} \ll P_s \ll P_{s^*}^{(2)} \\ (m-1)\,\frac{\lambda}{\mu}\,, & \text{for } P_{s^*}^{(2)} \ll P_s \ll P_{s^*}^{(3)} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{s^*}^{(3)} \ll P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,, \end{cases} \tag{244}$$

where $P_{s^*}^{(1)}$, $P_{s^*}^{(2)}$, $P_{s^*}^{(3)}$, and $P_{s^*}^{(4)}$ are determined by (233), (162), (237), and (131), respectively.

Case 2: $\frac{\lambda}{\mu} < \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. From (233), (162), (237), (131), and (198), it holds that $P_{s^*}^{(1)} \ll P_{s^*}^{(3)} = P_{s^*}^{(5)} = \lambda/\mu < P_{s^*}^{(2)} < P_{s^*}^{(4)}$. Also, from (241), and considering (129), (179), (182), and (205), it follows that

$$P_{\text{DL}} \approx \begin{cases} \frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R\,, & \text{for } P_s \ll P_{s^*}^{(1)} \\ C\,\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,P_s\,, & \text{for } P_{s^*}^{(1)} \ll P_s \ll P_{s^*}^{(5)} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,, \end{cases} \tag{245}$$

where $P_{s^*}^{(1)}$, $P_{s^*}^{(2)}$, $P_{s^*}^{(4)}$, and $P_{s^*}^{(5)}$ are determined by (233), (162), (131), and (198), respectively.

Remark 40: In the first case, that is, when $\lambda/\mu \geq 2/[C\,(m-2)]$, (244) implies that P_{DL}, as a function of P_s, exhibits three plateaus in the intervals $[0, P_{s^*}^{(1)})$, $(P_{s^*}^{(2)}, P_{s^*}^{(3)})$, and $(P_{s^*}^{(4)}, 1]$, respectively. However, in the second case, that is, when $\lambda/\mu < 2/[C\,(m-2)]$, it holds that $P_{s^*}^{(3)} < P_{s^*}^{(2)}$ and therefore the second plateau vanishes. In this case, (244) degenerates to (245).

Unrecoverable failures may occur in conjunction with one, two or three device failures. Consequently, the probability P_{UF} of one or more unrecoverable failures during rebuild is

$$P_{\text{UF}} = P_{\text{UF,1}} + P_{\text{UF,2}} + P_{\text{UF,3}} \tag{246}$$

$$\overset{(213)}{\approx} P_{\text{UF,1}} + P_{\text{UF,2}} \tag{247}$$

$$\overset{(142)}{\approx} P_{\text{UF|1}} + P_{\text{UF,2}} \tag{248}$$

$$\overset{(128)(186)}{\approx} 1 - q_1^C + \left[1 + \frac{1-q_2^C}{\log(q_2^C)}\right](m-1)\,\frac{\lambda}{\mu}\,, \tag{249}$$

where $P_{\text{UF,1}}$, $P_{\text{UF,2}}$, and $P_{\text{UF,3}}$ are obtained by (137), (178), and (209), respectively. Also, q_1 and q_2 are determined by (126) and (147), respectively. We proceed by considering the previous two cases:

Case 1: $\frac{\lambda}{\mu} \geq \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. From (248) and considering (129), (177), (179), and (182), it follows that

$$P_{\text{UF}} \approx \begin{cases} C\,\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,P_s\,, & \text{for } P_s \ll P_{s^*}^{(2)} \\ (m-1)\,\frac{\lambda}{\mu}\,, & \text{for } P_{s^*}^{(2)} \ll P_s \ll P_{s^*}^{(3)} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{s^*}^{(3)} \ll P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,, \end{cases} \tag{250}$$

where $P_{s^*}^{(1)}$, $P_{s^*}^{(2)}$, $P_{s^*}^{(3)}$, and $P_{s^*}^{(4)}$ are determined by (233), (162), (237), and (131), respectively.

Case 2: $\frac{\lambda}{\mu} < \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. From (248) and considering (129), (179), and (182), it follows that

$$P_{\text{UF}} \approx \begin{cases} C\,\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,P_s\,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ C\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll P_{s^*}^{(4)} \\ 1\,, & \text{for } P_s \gg P_{s^*}^{(4)}\,, \end{cases} \tag{251}$$

where $P_{s^*}^{(4)}$ and $P_{s^*}^{(5)}$ are given by (131) and (198), respectively.

E. Amount of Data Loss

As discussed in Section VI-D, data loss during rebuild may occur because of two additional (second and third) device failures or an unrecoverable failure of one or more codewords, or a combination thereof. Note that in all cases, data loss cannot involve only parity data, but also loss of user data.

Data loss during rebuild may occur because of unrecoverable failures in conjunction with one, two or three device failures. Consequently, the expected number $E(S_{\mathrm{U}})$ of symbols lost due to unrecoverable failures is obtained as follows:

$$E(S_{\mathrm{U}}) = E(S_{\mathrm{U},1}) + E(S_{\mathrm{U},2}) + E(S_{\mathrm{U},3}) \,, \tag{252}$$

where $E(S_{\mathrm{U},1})$, $E(S_{\mathrm{U},2})$, and $E(S_{\mathrm{U},3})$ are determined by (141), (192), and (223), respectively. Moreover, according to (197) and Remark 34, it holds that

$$E(S_{\mathrm{U}}) \approx \begin{cases} E(S_{\mathrm{U},2}) \,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ E(S_{\mathrm{U},1}) \,, & \text{for } P_s \gg P_{s^*}^{(5)} \end{cases} \tag{253}$$

$$\overset{(144)(196)}{\approx} \begin{cases} \frac{3}{2}\, C\,(m-1)(m-2)\,\frac{\lambda}{\mu}\,P_s & \text{for } P_s \ll P_{s^*}^{(5)} \\ \frac{3}{2}\, C\,(m-1)\,(m-2)\,P_s^2 \,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll \frac{1}{m-1} \end{cases}, \tag{254}$$

where $P_{s^*}^{(5)}$ is determined by (198).

The expected total number $E(S)$ of symbols lost is

$$E(S) = E(S_{\mathrm{D}}) + E(S_{\mathrm{U}}) \,, \tag{255}$$

where $E(S_{\mathrm{D}})$ and $E(S_{\mathrm{U}})$ are determined by (218) and (252), respectively.

Remark 41: It follows from (143), (195), (222), and (252) that

$$E(S) \approx C \left[E(L_1) + \frac{1}{2}\, E(L_2)\,(m-1)\,\frac{\lambda}{\mu} + \frac{(m-1)(m-2)}{2}\,\left(\frac{\lambda}{\mu}\right)^2 f_R \right] \tag{256}$$

$$\overset{(135)(150)}{\approx} C\,\frac{(m-1)(m-2)}{2}\left[\left(\frac{\lambda}{\mu}\right)^2 f_R + 3\,\frac{\lambda}{\mu}\,P_s + 3\,P_s^2\right], \quad \text{for } P_s \ll \frac{1}{m-1} \,, \tag{257}$$

where $E(L_1)$, $E(L_2)$, and f_R are determined by (134), (149), and (174), respectively. In particular, for $P_s = 0$, it holds that $E(S) = E(S_{\mathrm{D}}) \approx (1/2)\, C\,(m-1)\,(m-2)\,(\lambda/\mu)^2\, f_R$.

Remark 42: When P_s increases and approaches 1, from (134), (141), (149), (190), (191), (192), (218), (219), (220), (221), (252), and (255), it follows that $E(S)$ approaches $C\,m$. This is intuitively obvious because when $P_s = 1$, all $C\,m$ symbols stored in the system are lost owing to unrecoverable errors.

We now proceed to derive $E(Q)$, the expected amount of lost user data. First, we note that the expected number of lost user symbols is equal to the product of the storage efficiency to the expected number of lost symbols. Consequently, it follows from (1) that

$$E(Q) = \frac{l}{m}\, E(S)\, s \overset{(3)}{=} \frac{l}{m}\,\frac{E(S)}{C}\, c \,, \tag{258}$$

where $E(S)$ is given by (255) and s denotes the symbol size.

Similar expressions for the expected amounts $E(Q_{\mathrm{DF}_3})$ and $E(Q_{\mathrm{UF}})$ of lost user data due to device and unrecoverable failures are obtained from $E(S_{\mathrm{D}})$ and $E(S_{\mathrm{U}})$, respectively, as follows:

$$E(Q_{\mathrm{DF}_3}) = \frac{l}{m}\, E(S_{\mathrm{D}})\, s \overset{(3)}{=} \frac{l}{m}\,\frac{E(S_{\mathrm{D}})}{C}\, c \tag{259}$$

$$\overset{(222)}{\approx} \frac{l}{m}\,\frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R\, c \tag{260}$$

and

$$E(Q_{\mathrm{UF}}) = \frac{l}{m}\, E(S_{\mathrm{U}})\, s \overset{(3)}{=} \frac{l}{m}\,\frac{E(S_{\mathrm{U}})}{C}\, c \tag{261}$$

$$\overset{(254)}{\approx} \begin{cases} 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,c\,P_s \,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,c\,P_s^2 \,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll \frac{1}{m-1} \end{cases}, \tag{262}$$

where $E(S_{\mathrm{D}})$, $E(S_{\mathrm{U}})$, and $P_{s^*}^{(5)}$ are determined by (218), (252), and (198), respectively.

Substituting (256) and (257) into (258) yields

$$E(Q) \approx \frac{l}{m}\,\left[E(L_1) + \frac{1}{2}\, E(L_2)\,(m-1)\,\frac{\lambda}{\mu} + \frac{(m-1)(m-2)}{2}\,\left(\frac{\lambda}{\mu}\right)^2 f_R\right] c \tag{263}$$

$$\approx \frac{l}{m}\,\frac{(m-1)(m-2)}{2}\left[\left(\frac{\lambda}{\mu}\right)^2 f_R + 3\,\frac{\lambda}{\mu}\,P_s + 3\,P_s^2\right] c \,, \quad \text{for } P_s \ll \frac{1}{m-1} \,, \tag{264}$$

where $E(L_1)$, $E(L_2)$, and f_R are obtained by (134), (149), and (174), respectively. In particular, for $P_s = 0$, it holds that $E(Q) = E(Q_{\mathrm{DF}_3})$, which is determined by (260).

From (255), (258), (259), and (261), it holds that

$$E(Q) = E(Q_{\mathrm{DF}_3}) + E(Q_{\mathrm{UF}}) \,. \tag{265}$$

Also, the expected amounts $E(Q_{\mathrm{UF},1})$, $E(Q_{\mathrm{UF},2})$, and $E(Q_{\mathrm{UF},3})$ of lost user data due to unrecoverable failures in conjunction with one, two, and three device failures are as follows:

$$E(Q_{\mathrm{UF},1}) = \frac{l}{m}\,\frac{E(S_{\mathrm{U},1})}{C}\, c \,, \tag{266}$$

$$E(Q_{\mathrm{UF},2}) = \frac{l}{m}\,\frac{E(S_{\mathrm{U},2})}{C}\, c \,, \tag{267}$$

and

$$E(Q_{\mathrm{UF},3}) = \frac{l}{m}\,\frac{E(S_{\mathrm{U},3})}{C}\, c \,, \tag{268}$$

where $E(S_{\mathrm{U},1})$, $E(S_{\mathrm{U},2})$, and $E(S_{\mathrm{U},3})$ are determined by (141), (192), and (223), respectively.

Remark 43: From (253), (261), (266), and (267), it follows that

$$E(Q_{\mathrm{UF}}) \approx \begin{cases} E(Q_{\mathrm{UF},2}) \,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ E(Q_{\mathrm{UF},1}) \,, & \text{for } P_s \gg P_{s^*}^{(5)} \,. \end{cases} \tag{269}$$

Remark 44: Note that the inequalities derived in Remark 34 together with (224) and (225), and by virtue of (259) and (261), imply that

$$E(Q_{\rm UF}) \ll E(Q_{\rm DF_3}) \;\Leftrightarrow\; P_s \ll P_{s^*}^{(5)}\,, \tag{270}$$

where $P_{s^*}^{(5)}$ is determined by (198).

Also, from (260), (262), (269), and (270), it follows that

$$E(Q) \approx \begin{cases} E(Q_{\rm DF_3})\,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ E(Q_{\rm UF})\,, & \text{for } P_s \gg P_{s^*}^{(5)} \end{cases} \tag{271}$$

$$\approx \begin{cases} \frac{l}{m}\,\frac{(m-1)(m-2)}{2}\left(\frac{\lambda}{\mu}\right)^2 f_R\, c\,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\, c\, P_s^2\,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll \frac{1}{m-1}\,, \end{cases} \tag{272}$$

where $P_{s^*}^{(5)}$ is determined by (198).

Remark 45: When P_s increases and approaches 1, from (258) and according to Remark 42, it follows that $E(Q)$ approaches cl. This is intuitively obvious because when $P_s = 1$, upon the first device failure, the entire amount cl of user data stored in the RAID-6 array is lost owing to unrecoverable errors.

F. Reliability Metrics

The MTTDL is obtained by substituting (229) into (9). From (229), it follows that MTTDL is insensitive to device failure distribution, but it depends on the rebuild time distribution through f_R and on their means $1/\lambda$ and $1/\mu$, respectively. In particular, the normalized MTTDL depends only on f_R and the ratio λ/μ of their means. Note that for $P_s = 0$, $n = m = N$, and for an exponential rebuild time distribution, for which it holds that $f_R = E(R^2)/[E(R)]^2 = 2$, (231) implies that

$$\text{MTTDL} \;\approx\; \frac{\mu^2}{N(N-1)(N-2)\,\lambda^3}\,, \tag{273}$$

which is the same result as that reported in [3]. Also, for small values of P_s, (231) yields

$$\text{MTTDL} \;\approx\; \frac{\mu^2}{N(N-1)(N-2)\,\lambda^2\left(\,\lambda + \frac{1}{2}\,\mu\, C\, P_s\right)}\,, \tag{274}$$

whereas Equation (110) of [11] yields

$$\text{MTTDL} \;\approx\; \frac{\mu^2}{N(N-1)(N-2)\,\lambda^2\,(\,\lambda + \mu\, C\, P_s)}\,. \tag{275}$$

Their difference is the factor 1/2, which is attributed to the fact that the probability $P_{\rm uf}^{(2)}$ of data loss due to an unrecoverable failure given two device failures is obtained in [11] by assuming that all C codewords are to be recovered. It is subsequently obtained in [11] by expression (94) and is equal to $(N-2)\,C\,P_s$. This measure corresponds to $P_{\rm UF|2}$ whose value, according to (169), is roughly equal to $1/2\,(N-2)\,C\,P_s$, which is half that of $P_{\rm uf}^{(2)}$. This is because, after the second device failure and according to (31), the expected number of codewords to be recovered in the critical mode is only half the total of codewords C.

The EAFDL is obtained by substituting (258) into (10). In particular, the EAFDL normalized to λ is obtained by substituting (272) into (10) as follows:

$$\text{EAFDL}/\lambda \approx \begin{cases} \frac{1}{2}\,(m-1)(m-2)\left(\frac{\lambda}{\mu}\right)^2 f_R\,, & \text{for } P_s \ll P_{s^*}^{(5)} \\ \frac{3}{2}\,(m-1)(m-2)\,P_s^2\,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll \frac{1}{m-1}\,, \end{cases} \tag{276}$$

where λ/μ and $P_{s^*}^{(5)}$ are determined by (5) and (198), respectively. Note that EAFDL is insensitive to the device failure distribution, but it depends on the rebuild time distribution through f_R and on their means $1/\lambda$ and $1/\mu$, respectively. In particular, the normalized EAFDL depends only on f_R and the ratio λ/μ of their means. Also, for $P_s = 0$, and according to (276), we obtain $\text{EAFDL}/\lambda \approx (1/2)\,(m-1)\,(m-2)\,(\lambda/\mu)^2\,f_R$, which is in agreement with Equation (74) of [14] (with $c/b = 1/\mu$ and $\phi = 1$).

The value of $E(H)$ is obtained by substituting (229) and (258) into (11). In particular, depending on the values of λ/μ, m and C and for the two cases considered in Section VI-D, $E(H)$ normalized to c is obtained by substituting (244) or (245) and (272) into (11) as follows:

Case 1: $\frac{\lambda}{\mu} \geq \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. Depending on the value of λ/μ, we consider two subcases:

(a) $\frac{\lambda}{\mu} \geq \sqrt{\frac{2}{C(m-1)(m-2)}}$. In this case it holds that

$$P_{s^*}^{(1)} < P_{s^*}^{(2)} < P_{s^*}^{(3)} < P_{s^*}^{(4)} \leq P_{s^*}^{(5)}\,. \tag{277}$$

(b) $\frac{1}{C}\cdot\frac{2}{m-2} \leq \frac{\lambda}{\mu} < \sqrt{\frac{2}{C(m-1)(m-2)}}$. In this case it holds that

$$P_{s^*}^{(1)} < P_{s^*}^{(2)} < P_{s^*}^{(3)} \leq P_{s^*}^{(5)} < P_{s^*}^{(4)}\,. \tag{278}$$

Then, the $E(H)$ normalized to c is obtained by substituting (244) and (272) into (11) as follows:

$$E(H)/c \approx \begin{cases} \frac{l}{m}\,, & \text{for } P_s \ll P_{s^*}^{(1)} \\ \frac{l}{m}\,\frac{\lambda}{\mu}\, f_R\,\frac{1}{C\,P_s}\,, & \text{for } P_{s^*}^{(1)} \ll P_s \ll P_{s^*}^{(2)} \\ \frac{l}{m}\,\frac{m-2}{2}\,\frac{\lambda}{\mu}\, f_R\,, & \text{for } P_{s^*}^{(2)} \ll P_s \ll P_{s^*}^{(3)} \\ \frac{l}{m}\,(\frac{\lambda}{\mu})^2\, f_R\,\frac{1}{C\,P_s^2}\,, & \text{for } P_{s^*}^{(3)} \ll P_s \ll P_{\rm min}^{(4,5)} \\ \frac{l}{m}\,\max\left(\frac{(m-1)(m-2)}{2}\,(\frac{\lambda}{\mu})^2\, f_R\,,\,\frac{3}{C}\right), & \text{for } P_{\rm min}^{(4,5)} \ll P_s \ll P_{\rm max}^{(4,5)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,P_s^2\,, & \text{for } P_{\rm max}^{(4,5)} \ll P_s \ll \frac{1}{m-1}\,, \end{cases} \tag{279}$$

where

$$P_{\rm min}^{(4,5)} \triangleq \min(P_{s^*}^{(4)}, P_{s^*}^{(5)}) \quad\text{and}\quad P_{\rm max}^{(4,5)} \triangleq \max(P_{s^*}^{(4)}, P_{s^*}^{(5)})\,, \tag{280}$$

and $P_{s^*}^{(1)}$, $P_{s^*}^{(2)}$, $P_{s^*}^{(3)}$, $P_{s^*}^{(4)}$, and $P_{s^*}^{(5)}$ are determined by (233), (162), (237), (131), (198), respectively. Note that $E(H)$, as a function of P_s, exhibits three plateaus in the intervals $[0, P_{s^*}^{(1)})$, $(P_{s^*}^{(2)}, P_{s^*}^{(3)})$, and $(P_{\rm min}^{(4,5)}, P_{\rm max}^{(4,5)})$, respectively.

Case 2: $\frac{\lambda}{\mu} < \frac{1}{C}\cdot\frac{2}{m-2} = P_{s^*}^{(2)}$. In this case it holds that

$$P_{s^*}^{(1)} < P_{s^*}^{(3)} = P_{s^*}^{(5)} < P_{s^*}^{(2)} < P_{s^*}^{(4)}\,. \tag{281}$$

The value of $E(H)$ normalized to c is obtained by substituting (245) and (272) into (11) as follows:

$$\frac{E(H)}{c} \approx \begin{cases} \frac{l}{m} \,, & \text{for } P_s \ll P_{s^*}^{(1)} \\ \frac{l}{m}\,\frac{\lambda}{\mu}\, f_R\, \frac{1}{C\,P_s} \,, & \text{for } P_{s^*}^{(1)} \ll P_s \ll P_{s^*}^{(5)} \\ 3\,\frac{l}{m}\,\frac{1}{C} \,, & \text{for } P_{s^*}^{(5)} \ll P_s \ll P_{s^*}^{(4)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,P_s^2 \,, & \text{for } P_{s^*}^{(4)} \ll P_s \ll \frac{1}{m-1} \,, \end{cases} \tag{282}$$

where $P_{s^*}^{(1)}$, $P_{s^*}^{(4)}$, and $P_{s^*}^{(5)}$ are determined by (233), (131), and (198), respectively. Note that $E(H)$, as a function of P_s, exhibits two plateaus in the intervals $[0, P_{s^*}^{(1)})$ and $(P_{s^*}^{(5)}, P_{s^*}^{(4)})$, respectively.

Analogous to (119), the expected amounts $E(H_{\text{DF}_3})$ and $E(H_{\text{UF}})$ of user data lost due to device and unrecoverable failures, given that such failures have occurred, are

$$E(H_{\text{DF}_3}) = \frac{E(Q_{\text{DF}_3})}{P_{\text{DF}_3}} \,, \quad \text{and} \quad E(H_{\text{UF}}) = \frac{E(Q_{\text{UF}})}{P_{\text{UF}}} \,, \tag{283}$$

respectively. Also, analogous to (120), the relation between $E(H)$, $E(H_{\text{DF}_3})$, and $E(H_{\text{UF}})$ is

$$E(H) \;=\; \frac{P_{\text{DF}_3}}{P_{\text{DL}}}\, E(H_{\text{DF}_3}) + \frac{P_{\text{UF}}}{P_{\text{DL}}}\, E(H_{\text{UF}}) \,. \tag{284}$$

Substituting (205) and (260) into (283) yields

$$E(H_{\text{DF}_3})/c \;\approx\; \frac{l}{m} \,. \tag{285}$$

Also, depending on the values of λ/μ, m and C and for the two cases considered in Section VI-D, the $E(H_{\text{UF}})$ normalized to c is obtained by substituting (250) or (251), and (262) into (283) as follows:

Case 1: $\frac{\lambda}{\mu} \geq \frac{1}{C} \cdot \frac{2}{m-2} = P_{s^*}^{(2)}$.

$$E(H_{\text{UF}})/c \approx \begin{cases} 3\,\frac{l}{m}\,\frac{1}{C} \,, & \text{for } P_s \ll P_{s^*}^{(2)} \\ 3\,\frac{l}{m}\,\frac{m-2}{2}\,P_s \,, & \text{for } P_{s^*}^{(2)} \ll P_s \ll P_{s^*}^{(3)} \\ 3\,\frac{l}{m}\,\frac{\lambda}{\mu}\,\frac{1}{C\,P_s} \,, & \text{for } P_{s^*}^{(3)} \ll P_s \ll P_{\text{min}}^{(4,5)} \\ 3\,\frac{l}{m}\,\max\left(\frac{(m-1)(m-2)}{2}\,\frac{\lambda}{\mu}\,P_s \,,\, \frac{1}{C}\right) , & \text{for } P_{\text{min}}^{(4,5)} \ll P_s \ll P_{\text{max}}^{(4,5)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,P_s^2 \,, & \text{for } P_{\text{max}}^{(4,5)} \ll P_s \ll \frac{1}{m-1} \,, \end{cases} \tag{286}$$

where $P_{s^*}^{(1)}$, $P_{s^*}^{(2)}$, $P_{s^*}^{(3)}$, $P_{\text{min}}^{(4,5)}$ and $P_{\text{max}}^{(4,5)}$ are determined by (233), (162), (237), and (280), respectively. Note that $E(H_{\text{UF}})$ generally increases with P_s, but in the interval $(P_{s^*}^{(3)}, P_{\text{min}}^{(4,5)})$ decreases with P_s.

Case 2: $\frac{\lambda}{\mu} < \frac{1}{C} \cdot \frac{2}{m-2} = P_{s^*}^{(2)}$.

$$\frac{E(H_{\text{UF}})}{c} \approx \begin{cases} 3\,\frac{l}{m}\,\frac{1}{C} \,, & \text{for } P_s \ll P_{s^*}^{(4)} \\ 3\,\frac{l}{m}\,\frac{(m-1)(m-2)}{2}\,P_s^2 \,, & \text{for } P_{s^*}^{(4)} \ll P_s \ll \frac{1}{m-1} \,, \end{cases} \tag{287}$$

where $P_{s^*}^{(4)}$ is determined by (131).

VII. Numerical Results

A. A RAID-5 System

We consider a RAID-5 array comprised of $n = 8$ devices with $N = m = 8$, $l = 7$, $\lambda/\mu = 0.001$, capacity $c = 1$TB, and symbol size s equal to a sector size of 512 bytes, such that the number of codewords stored in a device is $C = c/s = 1.9{\times}10^9$.

The probability of data loss P_{DL} is determined by (83) as a function of the unrecoverable error probability P_s of a symbol (sector), and shown in Figure 2. According to (33), the probability P_{DF_2} of a device failure occurring during rebuild is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 2. It follows from (82) and (87) that, for $P_s \ll P_s^{(1)}$, an unrecoverable failure most likely occurs in the case of one device failure with the corresponding probability $P_{\text{UF,1}}$ being much smaller than the probability P_{DF_2} of encountering a device failure during rebuild, as shown in Figure 2. However, when P_s is in the range $(P_s^{(1)}, P_s^{(2)})$, $P_{\text{UF,1}}$ becomes greater than P_{DF_2}, which implies that a data loss is most likely caused by an unrecoverable failure that occurs in the case of one device failure. In particular, for $P_s \ll P_s^{(2)}$, and according to (39), $P_{\text{UF,1}}$ increases linearly with P_s, as indicated by the dotted green line in Figure 2. It follows from (88) and (21), and for the parameters considered here, that $P_s^{(1)} = 5{\times}10^{-13}$ and $P_s^{(2)} = 7{\times}10^{-11}$, as shown in Figure 2. For $P_s \gg P_s^{(2)}$, and according to (89), (91), and (92), it follows that $P_{\text{UF,1}}$, P_{UF} and, in turn, P_{DL} approach 1 and are essentially independent of P_s. In this range and according to (19), a device failure leads to data loss because one of the codewords is almost surely corrupted. Note that this also holds in the case when a subsequent (second) device failure occurs during rebuild. Consequently, and according to (66), $P_{\text{UF,2}}$ approaches P_{DF_2}, as indicated by the dotted magenta line in Figure 2. As expected, and according to (86), the total probability of data loss P_{DL} increases monotonically with P_s and exhibits two plateaus in the intervals $[0, P_s^{(1)})$ and $(P_s^{(2)}, 1]$, respectively.

The normalized λMTTDL measure is obtained from (113) and is shown in Figure 3 as a function of the unrecoverable symbol error probability. The various regions and plateaus are also depicted and correspond to the ranges discussed above regarding the probability of data loss.

The normalized expected amount $E(Q)/c$ of lost user data relative to the amount of data stored in a device is obtained from (99) as a function of the unrecoverable symbol error probability P_s, and shown in Figure 4. According to (101), the normalized expected amount $E(Q_{\text{DF}_2})/c$ of user data lost due to a subsequent device failure during rebuild is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 4. The normalized expected amount $E(Q_{\text{UF}})/c$ of user data lost due to unrecoverable failures, and according to (109), is roughly equal to $E(Q_{\text{UF,1}})/c$, which is determined by (107) and shown in Figure 4. In particular, for small values of P_s, and according to (103), it increases linearly with P_s, as indicated by the dotted green line in Figure 4. Also, the expected amount $E(Q_{\text{UF,2}})$ of user data lost due to unrecoverable failures in conjunction with two device failures, and according to Remark

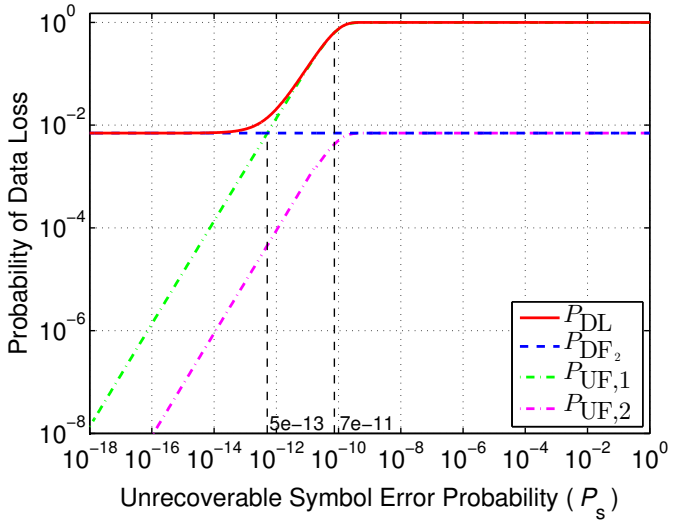

Figure 2. Probability of data loss P_{DL} for a RAID-5 array with latent errors ($\lambda/\mu = 0.001$, $m = N = 8$, $l = 7$, $c = 1$TB, and $s = 512$ B).

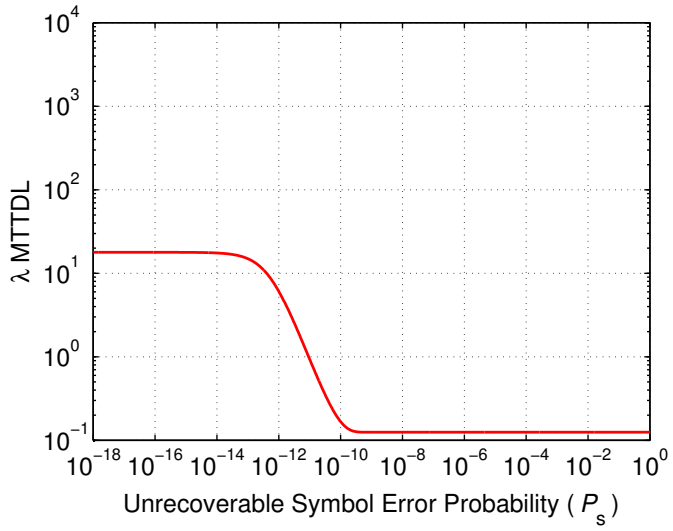

Figure 3. Normalized MTTDL for a RAID-5 array with latent errors ($\lambda/\mu = 0.001$, $m = N = 8$, $l = 7$, $c = 1$TB, and $s = 512$ B).

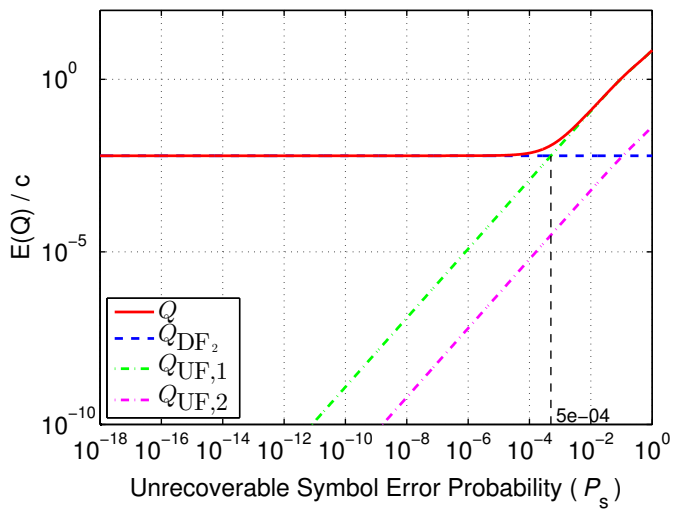

Figure 4. Normalized amount of data loss $E(Q)$ for a RAID-5 array with latent errors ($\lambda/\mu = 0.001$, $m = N = 8$, $l = 7$, $c = 1$TB, and $s = 512$ B).

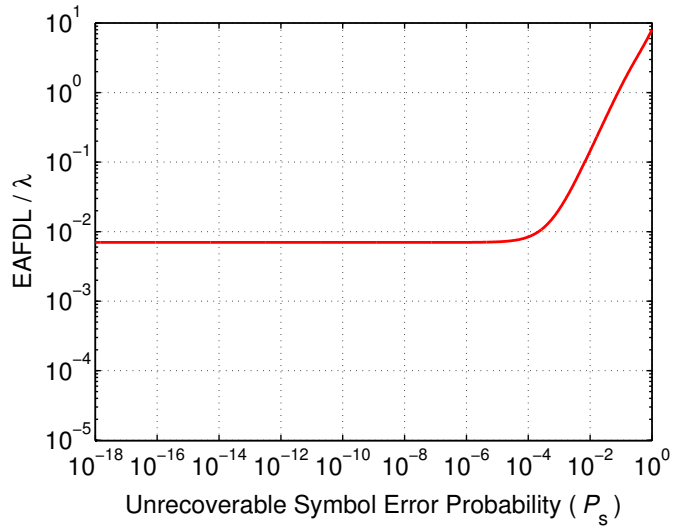

Figure 5. Normalized EAFDL for a RAID-5 array with latent errors ($\lambda/\mu = 0.001$, $m = N = 8$, $l = 7$, $c = 1$TB, and $s = 512$ B).

6, is much smaller than $E(Q_{\mathrm{UF},1})$, as indicated by the dotted magenta line in Figure 4. According to (110), $E(Q_{\mathrm{UF}})$ exceeds $E(Q_{\mathrm{DF}_2})$ when $P_s \gg P_s^{(3)} = 5{\times}10^{-4}$. As expected, and according to (105), the total expected amount $E(Q)$ of lost user data increases monotonically with P_s. In particular, when P_s approaches 1 and according to Remark 14, the normalized expected amount $E(Q)/c$ of lost user data approaches $l = 7$, as all user data in the array is lost.

The normalized EAFDL/λ measure is obtained by substituting (99) into (10) and is shown in Figure 5 as a function of the unrecoverable symbol error probability. Equation (10) suggests that this measure is proportional to $E(Q)$, which implies that the above discussion regarding the behavior of $E(Q)$ also holds here and therefore EAFDL increases monotonically with P_s.

The normalized expected amount $E(H)/c$ of lost user data, given that a data loss has occurred, relative to the amount of data stored in a device is obtained from (118) as a function of the unrecoverable symbol error probability P_s and shown in Figure 6. In contrast to the P_{DL}, EAFDL, and $E(Q)$ measures that increase monotonically with P_s, we observe that $E(H)$ does not do so. Data losses occur because of a subsequent device failure or unrecoverable failures of codewords, or a combination thereof. According to (123), the expected amount $E(H_{\mathrm{DF}_2})$ of lost user data associated with a subsequent device failure, given that such a device failure has occurred during rebuild, is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 6. Such a device failure causes the loss of a large number of symbols as opposed to a small number of additional symbols that may be lost owing to unrecoverable failures. The expected amount $E(H_{\mathrm{UF}})$ of user data lost due to

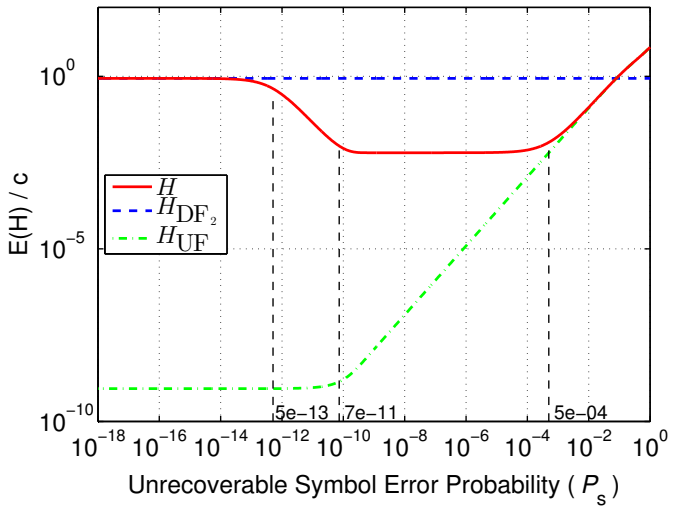

Figure 6. Normalized $E(H)$ for a RAID-5 array with latent errors ($\lambda/\mu = 0.001$, $m = N = 8$, $l = 7$, $c = 1$TB, and $s = 512$ B).

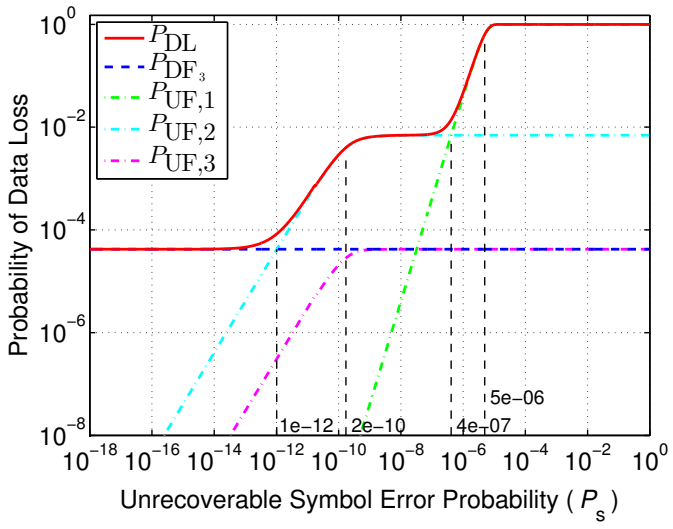

Figure 7. Probability of data loss P_{DL} for a RAID-6 array with latent errors ($\lambda/\mu = 0.001$, $f_R = 2$, $m = N = 8$, $l = 6$, $c = 1$TB, and $s = 512$ B).

unrecoverable failures, given that such failures have occurred, is obtained from (124) and shown in Figure 6. According to Remark 1, (24), and (28), when $P_s \ll P_s^{(2)} = 7{\times}10^{-11}$, an unrecoverable failure is most likely caused by a single corrupted codeword that loses two symbols. Consequently, and according to (125), for $P_s \ll P_s^{(2)}$, the expected amount $E(H_{\mathrm{UF}})$ of user data lost due to unrecoverable failures, given that such unrecoverable failures have occurred, is independent of P_s, as indicated by the horizontal part of the dotted green line in Figure 6. Also, the amount of lost data, which corresponds to the two lost symbols, is negligible compared with the amount of data lost due to a subsequent device failure, that is, $E(H_{\mathrm{UF}}) \ll E(H_{\mathrm{DF}_2})$. According to (24) and (28), when $P_s \gg P_s^{(2)}$, unrecoverable failures are most likely caused by multiple corrupted codewords that each loses two symbols. Moreover, (24) and (125) imply that the number of the corrupted codewords and the corresponding amount of lost data increase linearly with P_s, as indicated by the dotted green line shown in Figure 6.

The combined expected amount $E(H)$ of lost user data, given that data loss has occurred, is an average of $E(H_{\mathrm{DF}_2})$ and $E(H_{\mathrm{UF}})$ with the weights determined in (120). For $P_s \ll P_s^{(1)} = 5{\times}10^{-13}$, a data loss is most likely attributed to two device failures, which results in the first plateau obtained in (123). However, for values of P_s in the range ($5{\times}10^{-13}$, $7{\times}10^{-11}$), this is reversed, meaning that an unrecoverable failure is more likely to occur than a device failure, and this causes P_{DL} to increase as shown in Figure 2. Consequently, as the weight of the $E(H_{\mathrm{DF}_2})$ component decreases, so does $E(H)$. Subsequently, as P_s increases further, this weight along with $E(H)$ can no longer decrease because P_{DL} has reached its maximum value of 1. But, $E(H)$ cannot increase either because, although $E(H_{\mathrm{UF}})$ increases, it still remains negligible compared with $E(H_{\mathrm{DF}_2})$. As a result, $E(H)$ stabilizes at the second plateau level at $(l/m)\,(m-1)\,(\lambda/\mu)\,c$, as obtained by (121). As P_s increases further and exceeds $P_s^{(3)} = 5{\times}10^{-4}$, according to (110), the increasing amount of data lost due to unrecoverable failures $E(Q_{\mathrm{UF}})$ far exceeds $E(Q_{\mathrm{DF}_2})$, which in turn leads $E(H)$ to be essentially equal to $E(H_{\mathrm{UF}})$ and therefore to increase with P_s. In particular, when P_s approaches 1, and according to Remark 17, the amount $l\,c$ of user data stored in the array is lost owing to unrecoverable errors, which in turn implies that the normalized expected amount $E(H)/c$ of lost user data approaches $l = 7$.

B. A RAID-6 System

We consider a RAID-6 array with the same characteristics as the RAID-5 array considered in the previous section, except that the parameter l is now equal to 6. Also, in contrast to a RAID-5 system, some of the reliability metrics for a RAID-6 system depend on the rebuild time distribution. We consider a rebuild time distribution, such as the exponential one, for which it holds that $E(R^2) = 2\,[E(R)]^2$, which implies that $f_R = 2$.

The probability of data loss P_{DL} is determined by (229) as a function of the unrecoverable error probability P_s of a symbol (sector), and shown in Figure 7. According to (204), the probability P_{DF_3} of two device failures occurring during rebuild is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 7. It follows from (230) and (232) that, for $P_s \ll P_{s^*}^{(1)}$, an unrecoverable failure most likely occurs in conjunction with two device failures with the corresponding probability $P_{\mathrm{UF},2}$ being much smaller than the probability P_{DF_3} of three device failures, as shown in Figure 7. However, when P_s is in the range $(P_{s^*}^{(1)}, P_{s^*}^{(2)})$, $P_{\mathrm{UF},2}$ becomes greater than P_{DF_3}, which implies that data loss is most likely caused by an unrecoverable failure that occurs in conjunction with two device failures. In particular, for $P_s \ll P_{s^*}^{(2)}$, and according to (179), $P_{\mathrm{UF},2}$ increases linearly with P_s, as indicated by the dotted cyan line in Figure 7. It follows from (233) and (162), and for the parameters considered here, that $P_{s^*}^{(1)} = 1{\times}10^{-12}$ and $P_{s^*}^{(2)} = 2{\times}10^{-10}$, as shown in Figure 7. It follows from (230) and (232) that, for $P_s \gg P_{s^*}^{(1)}$, the probability P_{UF} of encountering an unrecoverable failure is much greater than that

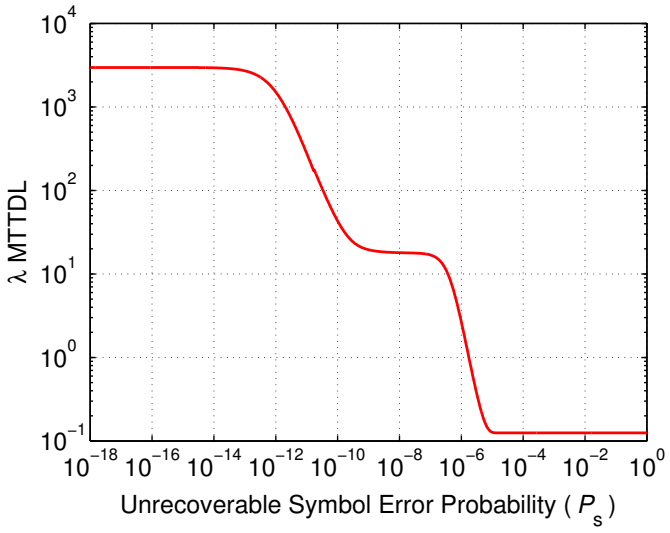

Figure 8. Normalized MTTDL for a RAID-6 array with latent errors ($\lambda/\mu = 0.001$, $f_R = 2$, $m = N = 8$, $l = 6$, $c = 1$TB, and $s = 512$ B).

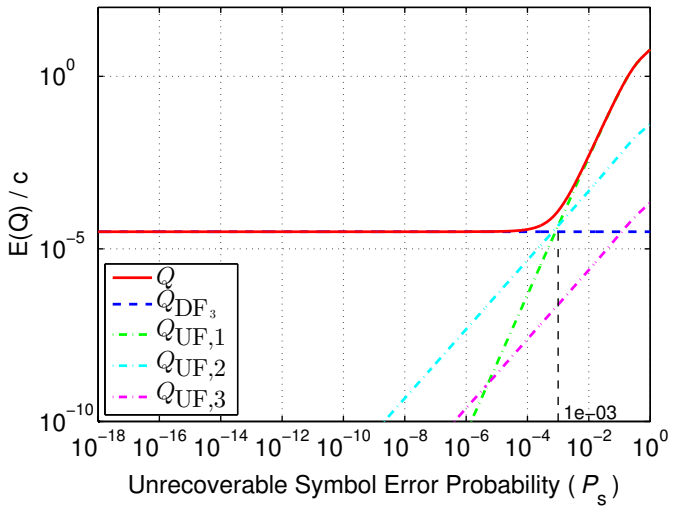

Figure 9. Normalized amount of data loss $E(Q)$ for a RAID-6 array with latent errors ($\lambda/\mu = 0.001$, $f_R = 2$, $m = N = 8$, $l = 6$, $c = 1$TB, and $s = 512$ B).

of encountering a data loss due to three device failures. In particular, from (169), (177), and (179), it follows that, when $P_s \gg P_{s^*}^{(2)}$, $P_{\text{UF}|2}$ approaches 1 and, in turn, $P_{\text{UF},2}$ approaches P_{DF_2} and they are essentially independent of P_s. In this range and according to Remark 38, a second device failure leads to data loss because one of the remaining codewords is almost surely corrupted, which implies that the probability of data loss is equal to that of a RAID-5 system in the absence of latent errors. Note that this also holds in the case when a subsequent (third) device failure occurs during rebuild. Consequently, and according to (210), $P_{\text{UF},3}$ approaches P_{DF_3}, as indicated by the dotted magenta line in Figure 7.

Subsequently, according to Remark 39, when $P_s \gg P_{s^*}^{(3)}$, the probability $P_{\text{UF},1}$ of data loss due to unrecoverable failures in the case of one device failure becomes greater than $P_{\text{UF},2}$, which implies that a data loss is most likely caused by an unrecoverable failure in conjunction with one device failure. In particular, for $P_s \ll P_{s^*}^{(4)}$, and according to (138) and (129), $P_{\text{UF},1}$ increases quadratically with P_s, as indicated by the dotted green line in Figure 7. It follows from (81) and (131), and for the parameters considered here, that $P_{s^*}^{(3)} = 4{\times}10^{-7}$ and $P_{s^*}^{(4)} = 5{\times}10^{-6}$, as shown in Figure 7. For $P_s \gg P_{s^*}^{(4)}$, and according to (187), (244), (247), and (250), it follows that $P_{\text{UF},1}$, P_{UF} and, in turn, P_{DL} approach 1 and are essentially independent of P_s. In this range and according to (129), a device failure leads to data loss because one of the codewords is almost surely corrupted. As expected, the total probability of data loss P_{DL} increases monotonically with P_s and exhibits three plateaus in the intervals $[0, P_{s^*}^{(1)})$, $(P_{s^*}^{(2)}, P_{s^*}^{(3)})$, and $(P_{s^*}^{(4)}, 1]$, respectively.

The normalized λMTTDL measure is obtained by substituting (229) into (9) and is shown in Figure 8 as a function of the unrecoverable symbol error probability. The various regions and plateaus are also depicted and correspond to the ranges discussed above regarding the probability of data loss.

The normalized expected amount $E(Q)/c$ of lost user data relative to the amount of data stored in a device is obtained from (258) as a function of the unrecoverable symbol error probability P_s, and shown in Figure 9. According to (260), the normalized expected amount $E(Q_{\text{DF}_3})/c$ of user data lost due to two subsequent device failures during rebuild is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 9. The normalized expected amount $E(Q_{\text{UF}})/c$ of user data lost due to unrecoverable failures, when $P_s \ll P_{s^*}^{(5)} = 10^{-3}$ and according to (262) and (269), is roughly equal to $E(Q_{\text{UF},2})/c$ and increases linearly with P_s, as indicated by the dotted cyan line in Figure 9. For $P_s \gg P_{s^*}^{(5)} = 10^{-3}$, $E(Q_{\text{UF}})/c$ is roughly equal to $E(Q_{\text{UF},1})/c$ and increases quadratically with P_s, as indicated by the dotted green line in Figure 9. Also, the expected amount $E(Q_{\text{UF},3})$ of user data lost due to unrecoverable failures in conjunction with three device failures, and according to Remark 34, is much smaller than $E(Q_{\text{UF},2})$, as indicated by the dotted magenta line in Figure 9. According to (270), $E(Q_{\text{UF}})$ exceeds $E(Q_{\text{DF}_3})$ when $P_s \gg P_{s^*}^{(5)} = 10^{-3}$. As expected, the total expected amount $E(Q)$ of lost user data increases monotonically with P_s. In particular, when P_s approaches 1 and according to Remark 12, the normalized expected amount $E(Q)/c$ of lost user data approaches $l = 6$, as all user data in the array is lost.

The normalized EAFDL/λ measure is obtained by substituting (258) into (10) and is shown in Figure 10 as a function of the unrecoverable symbol error probability. Equation (10) suggests that this measure is proportional to $E(Q)$, which implies that the preceding discussion regarding the behavior of $E(Q)$ also holds here and therefore EAFDL increases monotonically with P_s.

The normalized expected amount $E(H)/c$ of lost user data, given that a data loss has occurred, relative to the amount of data stored in a device is obtained from (279) as a function of the unrecoverable symbol error probability P_s and shown in Figure 11. In contrast to the P_{DL}, EAFDL, and $E(Q)$ measures that increase monotonically with P_s, we observe that $E(H)$

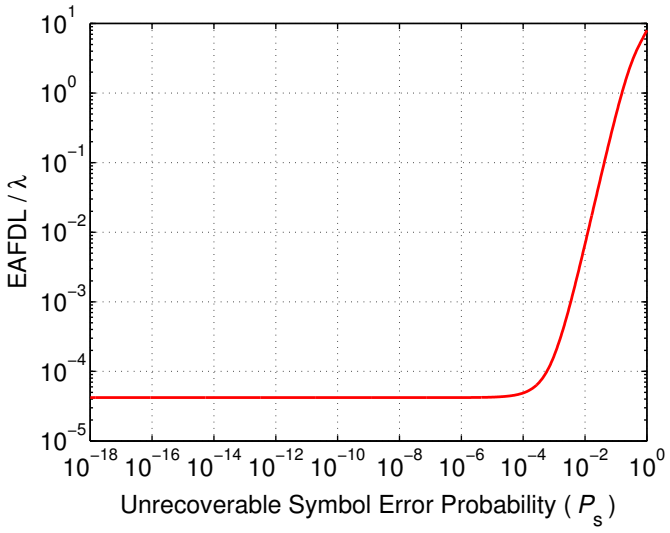

Figure 10. Normalized EAFDL for a RAID-6 array with latent errors ($\lambda/\mu = 0.001$, $f_R = 2$, $m = N = 8$, $l = 6$, $c = 1$TB, and $s = 512$ B).

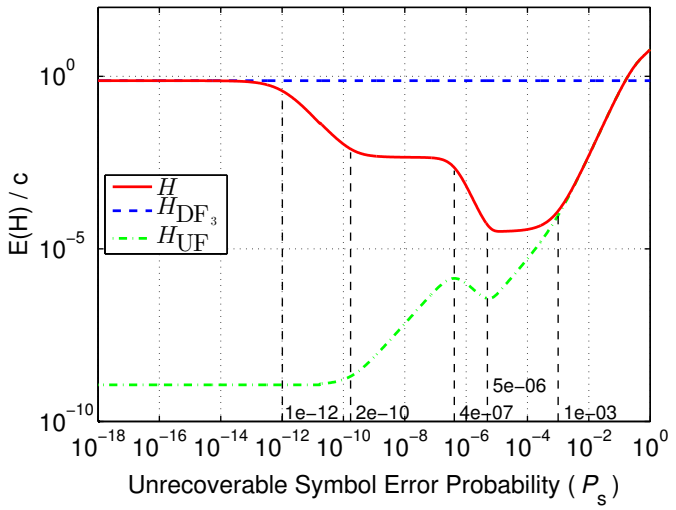

Figure 11. Normalized $E(H)$ for a RAID-6 array with latent errors ($\lambda/\mu = 0.001$, $f_R = 2$, $m = N = 8$, $l = 6$, $c = 1$TB, and $s = 512$ B).

does not do so. Data losses occur because of two subsequent device failures or unrecoverable failures of codewords, or a combination thereof. According to (285), the expected amount $E(H_{\mathrm{DF}_3})$ of lost user data associated with two subsequent device failures, given that such device failures have occurred during rebuild, is independent of the unrecoverable symbol error probability, as indicated by the horizontal dotted blue line in Figure 11. Such device failures cause the loss of a large number of symbols as opposed to a small number of additional symbols that may be lost owing to unrecoverable failures. The expected amount $E(H_{\mathrm{UF}})$ of user data lost due to unrecoverable failures, given that such failures have occurred, is determined by (286) and shown in Figure 11. According to (236), when $P_s \ll P_{s^*}^{(3)}$, $P_{\mathrm{UF},2}$ is much greater than $P_{\mathrm{UF},1}$, which implies that an unrecoverable failure most likely occurs in conjunction with two device failures. According to Corollary 11, (151), and (167), when $P_s \ll P_{s^*}^{(2)} = 2{\times}10^{-10}$, an unrecoverable failure is most likely caused by a single corrupted codeword that loses three symbols and is encountered after the second device failure. Consequently, and according to (286), for $P_s \ll P_{s^*}^{(2)}$, the expected amount $E(H_{\mathrm{UF}})$ of user data lost due to unrecoverable failures, given that such unrecoverable failures have occurred, is independent of P_s, as indicated by the horizontal part of the dotted green line in Figure 11. Also, the amount of lost data, which corresponds to the three lost symbols, is negligible compared with the amount of data lost due to a subsequent device failure, that is, $E(H_{\mathrm{UF}}) \ll E(H_{\mathrm{DF}_2})$. According to (151) and (167), when $P_s \gg P_s^{(2)}$, unrecoverable failures are most likely caused by multiple corrupted codewords that each loses three symbols and are encountered after the second device failure. Moreover, (167) and (286) imply that the number of corrupted codewords and the corresponding amount of lost data increase linearly with P_s in $(P_{s^*}^{(2)}, P_{s^*}^{(3)})$, as indicated by the dotted green line in Figure 11. Subsequently, when $P_s \gg P_{s^*}^{(3)}$, and according to (133) and (136), unrecoverable failures may also be caused by a single corrupted codeword that is encountered in conjunction with one device failure and loses three symbols. This in turn results in a reduction of the amount of lost data, as expressed by (286). When $P_s \gg P_{s^*}^{(4)}$, and according to (129), (133), and (136), the expected number of corrupted codewords in the case of one device failure also increases. In particular, unrecoverable failures are more likely encountered in conjunction with one device failure, but although $E(hufo)$ increases, it remains negligible compared with $E(huft)$, which also increases. This in turn results in an increase of the amount of lost data, as expressed by (286).

The combined expected amount $E(H)$ of lost user data, given that data loss has occurred, is an average of $E(H_{\mathrm{DF}_3})$ and $E(H_{\mathrm{UF}})$ with the weights determined in (284). For $P_s \ll P_{s^*}^{(1)} = 1{\times}10^{-12}$, a data loss is most likely attributed to three device failures, which results in the first plateau obtained in (285). However, for higher values of P_s, this is reversed, in that it becomes more likely to encounter an unrecoverable failure than a device failure. As in the case of RAID-5, $E(H)$ initially decreases, but in the case of RAID-6 it exhibits three plateaus, as obtained in (279). As P_s increases further and exceeds $P_{s^*}^{(5)} = 10^{-3}$, according to (270), the increasing amount of data lost due to unrecoverable failures $E(Q_{\mathrm{UF}})$ far exceeds $E(Q_{\mathrm{DF}_3})$, which in turn leads $E(H)$ to be essentially equal to $E(H_{\mathrm{UF}})$ and therefore to increase with P_s. In particular, when P_s approaches 1, and according to Remark 45, the amount $l\,c$ of user data stored in the array is lost owing to unrecoverable errors, which in turn implies that the normalized expected amount $E(H)/c$ of lost user data approaches $l = 7$.

VIII. Discussion

As discussed in Section III, field results suggest that the probability of unrecoverable sector errors lies in the range $(4.096{\times}10^{-11}, 5{\times}10^{-9})$. Figures 3 and 8 show that MTTDL is significantly degraded by the presence of latent errors, whereas Figures 5 and 10 reveal that EAFDL is practically unaffected in this range. When the probability of unrecoverable sector errors lies in the range of practical interest, the probability of an unrecoverable failure is much greater than that of a data loss due to device failures, which degrades MTTDL. However,

the amount of data corresponding to sectors lost due to latent errors is negligible compared with the amount of data lost due to device failures, which in turn implies that EAFDL remains practically unaffected. In contrast, Figures 6 and 11 reveal that the expected amount $E(H)$ of lost user data, given that data loss has occurred, decreases in the range of practical interest. This is because when a data loss occurs, it is more likely caused by unrecoverable failures that involve the loss of a small number of sectors rather than by multiple device failures that result in a significantly greater amount of lost data.

It follows from (88) and (81) that

$$P_s^{(1)} = \frac{1}{C} \cdot \frac{\lambda}{\mu} \;\ll\; \frac{1}{2} \cdot \frac{C+1}{C} \cdot \frac{\lambda}{\mu} = P_s^{(3)} \; . \tag{288}$$

Similarly, from (233) and (198), it holds that

$$P_{s^*}^{(1)} = \frac{1}{C} \cdot \frac{\lambda}{\mu} \cdot f_R \;\ll\; \frac{\lambda}{\mu} = P_{s^*}^{(5)} \; . \tag{289}$$

Consequently, increasing P_s first affects $P_{\rm DL}$, MTTDL, and $E(H)$ and then $E(Q)$ and EAFDL.

IX. Conclusions

The effect of latent sector errors on the reliability of RAID-5 and RAID-6 data storage systems was investigated. A methodology was developed for deriving the Mean Time to Data Loss (MTTDL) and the Expected Annual Fraction of Data Loss (EAFDL) reliability metrics analytically. Closed-form expressions capturing the effect of unrecoverable latent errors were obtained. Our results demonstrate that RAID-6 storage systems achieve a higher reliability than that of RAID-5 storage systems. We established that the reliability of storage systems is adversely affected by the presence of latent errors. The results demonstrated that the effect of latent errors depends on the relative magnitudes of the probability of a latent error versus the probability of a device failure. It was found that, for actual values of the unrecoverable sector error probability, MTTDL is adversely affected by the presence of latent errors, whereas EAFDL is not.

Extending the methodology developed to derive the MTTDL and EAFDL reliability metrics of erasure-coded systems in the presence of unrecoverable latent errors is a subject of further investigation.

Appendix A

Proof of Corollary 4.

For small values of x, it holds that

$$(1+x)^n \approx 1 + nx + \frac{n(n-1)}{2}\,x^2 + \frac{n(n-1)(n-2)}{6}\,x^3 + \frac{n(n-1)(n-2)(n-3)}{24}\,x^4 \; . \tag{290}$$

Consequently, for $x = -P_s$ and $n = (m-1)\,C$, (290) yields

$$q_1^C \overset{(14)}{=} (1-P_s)^{(m-1)C} \approx 1 - (m-1)C\,P_s + \frac{(m-1)C[(m-1)C-1]}{2}\,P_s^2 \; . \tag{291}$$

By setting $C = 1$, (291) yields

$$q_1 \approx 1 - (m-1)\,P_s + \frac{(m-1)(m-2)}{2}\,P_s^2 \; . \tag{292}$$

Substituting (291) and (292) into (56) yields

$$P_{\text{UF in } S_1|2} \approx 1 - \frac{(m-1)C\,P_s - \frac{(m-1)C[(m-1)C-1]}{2}\,P_s^2}{C\,[(m-1)\,P_s - \frac{(m-1)(m-2)}{2}\,P_s^2]} = 1 - \frac{1 - \frac{[(m-1)C-1]}{2}\,P_s}{1 - \frac{m-2}{2}\,P_s} \approx \frac{(C-1)(m-1)}{2}\,P_s \; . \tag{293}$$

This approximation holds when $(1/2)(C-1)(m-1)\,P_s \ll 1$ or, equivalently, $P_s \ll 2/[(C-1)(m-1)]$, which is roughly twice the value of $P_s^{(2)}$ as given by (21). ■

Appendix B

Proof of Corollary 5.

For $x = -P_s$ and $n = (m-2)\,C$, (290) yields

$$p_2^C \overset{(51)}{=} (1-P_s)^{(m-2)C} \approx 1 - (m-2)C\,P_s + \frac{(m-2)C[(m-2)C-1]}{2}\,P_s^2 \; . \tag{294}$$

By setting $C = 1$, (294) yields

$$p_2 \approx 1 - (m-2)\,P_s + \frac{(m-2)(m-3)}{2}\,P_s^2 \; . \tag{295}$$

Substituting (294) and (295) into (61) yields

$$P_{\text{UF in } S_2|2} \approx 1 - [1-(m-2)P_s]\,\frac{1 - \frac{C(m-2)-1}{2}\,P_s}{1 - \frac{m-3}{2}\,P_s} \approx \frac{\frac{(C+1)(m-2)}{2}\,P_s}{1 - \frac{m-3}{2}\,P_s} \approx \frac{(C+1)(m-2)}{2}\,P_s \; . \tag{296}$$

This approximation holds when $(1/2)(C+1)(m-2)\,P_s \ll 1$ or, equivalently, $P_s \ll 2/[(C+1)(m-2)]$, which for large C is roughly equal to $2/[C(m-2)]$. ■

Appendix C

Proof of Corollary 6.

From (14) and (51) it follows that

$$q_1 = (1-P_s)\,p_2 \; . \tag{297}$$

Then it holds that

$$\frac{p_2}{C}\,\frac{p_2^C - q_1^C}{p_2 - q_1} \overset{(297)}{=} \frac{p_2}{C}\,\frac{[1-(1-P_s)^C]p_2^C}{P_s\,p_2} = \frac{[1-(1-P_s)^C]}{C\,P_s}\,p_2^C \overset{(52)}{\approx} \frac{[1-(1-CP_s+\frac{C(C-1)}{2}\,P_s^2)]}{C\,P_s}\,[1-(m-2)\,P_s]^C \approx \left[1 - \frac{(C-1)}{2}\,P_s\right][1-C\,(m-2)\,P_s] = 1 - \left[\frac{C-1}{2} + (m-2)\,C\right]P_s + O(P_s^2) \tag{298}$$

Substituting (298) into (63) yields (64). ■

APPENDIX D

Proof of Corollary 10.

For $x = (m-2)\,P_s$ and $n = C$, (290) yields

$$\begin{aligned}[1+(m-2)\,P_s]^C &\approx 1 + C(m-2)\,P_s + \frac{C(C-1)}{2}\,[(m-2)\,P_s]^2 \\ &\quad + \frac{C(C-1)(C-2)}{6}\,[(m-2)\,P_s]^3 \\ &\quad + \frac{C(C-1)(C-2)(C-3)}{24}\,[(m-2)\,P_s]^4\,. \quad (299)\end{aligned}$$

Also, for $x = -P_s$ and $n = (m-2)\,C$, (290) yields

$$\begin{aligned}(1-P_s)^{(m-2)\,C} &\approx 1 - (m-2)\,CP_s + \frac{(m-2)\,C[(m-2)\,C-1]}{2}\,P_s^2 \\ &\quad - \frac{(m-2)\,C[(m-2)\,C-1][(m-2)\,C-2]}{6}\,P_s^3 \\ &\quad + \frac{(m-2)\,C[(m-2)\,C-1][(m-2)\,C-2][(m-2)\,C-3]}{24}\,P_s^4\,. \quad (300)\end{aligned}$$

From (126), (299), and (300), it follows that

$$\begin{aligned}q_1^C &\approx 1 - \frac{C(m-1)(m-2)}{2}P_s^2 + \frac{C(m-1)(m-2)(m-3)}{3}P_s^3 \\ &\quad + \frac{C(m-1)(m-2)}{8}[C(m-1)(m-2) - 2(m^2-5m+7)]P_s^4\,. \quad (301)\end{aligned}$$

By setting $C = 1$, (301) yields

$$\begin{aligned}q_1 &\approx 1 - \frac{(m-1)(m-2)}{2}\,P_s^2 + \frac{(m-1)(m-2)(m-3)}{3}\,P_s^3 \\ &\quad - \frac{(m-1)(m-2)(m-3)(m-4)}{8}\,P_s^4\,. \quad (302)\end{aligned}$$

Substituting (301) and (302) into (56) yields

$$\begin{aligned}P_{\text{UF in } S_1|2} &\approx 1 \\ &\quad - \frac{12 - 8(m-3)P_s - 3[C(m-1)(m-2) - 2(m^2-5m+7)]P_s^2}{12 - 8(m-3)P_s + 3(m-3)(m-4)P_s^2} \\ &= \frac{3(C-1)(m-1)(m-2)P_s^2}{12 - 8(m-3)P_s + 3(m-3)(m-4)P_s^2} \\ &\approx \frac{(C-1)(m-1)(m-2)}{4}\,P_s^2\,. \quad (303)\end{aligned}$$

■

APPENDIX E

Proof of Corollaries 12 and 13.

From (126) and (147) it follows that

$$q_1 = (1+x)\,q_2\,, \quad (304)$$

where

$$x \triangleq (m-2)P_s\,. \quad (305)$$

Then it holds that

$$\begin{aligned}\frac{q_2}{C}\,\frac{q_1^C - q_2^C}{q_1 - q_2} &\overset{(304)}{=} \frac{q_2}{C}\,\frac{[(1+x)^C - 1]q_2^C}{x\,q_2} = \frac{[(1+x)^C - 1]}{C\,x}\,q_2^C \\ &\overset{(148)(305)}{\approx} \frac{[(1 + Cx + \frac{C(C-1)}{2}\,x^2 - 1]}{C\,x}\,(1-x)^C \\ &\approx \left[1 + \frac{C-1}{2}\,x\right](1 - C\,x) \\ &= 1 - \frac{C+1}{2}\,x + O(x^2) \\ &= 1 - \frac{C+1}{2}\,(m-2)\,P_s + O(P_s^2) \quad (306)\end{aligned}$$

and

$$\begin{aligned}q_2^2\,\frac{q_1^C - C\,q_1\,q_2^{C-1} + (C-1)q_2^C}{(q_1-q_2)^2} &\overset{(304)}{=} q_2^2\,\frac{[(1+x)^C - C\,(1+x)\,+(C-1)]q_2^C}{(x\,q_2)^2} \\ &\approx \frac{[(1 + Cx + \frac{C(C-1)}{2}\,x^2 + \frac{C(C-1)(C-2)}{6}\,x^3 - Cx - 1)]}{x^2}\,q_2^C \\ &\overset{(148)(305)}{\approx} \left[\frac{C(C-1)}{2} + \frac{C(C-1)(C-2)}{6}\,x\right](1-x)^C \\ &\approx \frac{C(C-1)}{2}\left[1 + \frac{C-2}{3}\,x\right](1 - C\,x) \\ &= \frac{C(C-1)}{2} - \frac{C\,(C-1)\,(C+1)}{3}\,x + O(x^2) \\ &\overset{(305)}{=} \frac{C(C-1)}{2} - \frac{C\,(C-1)\,(C+1)}{3}\,(m-2)\,P_s + O(P_s^2) \quad (307)\end{aligned}$$

Substituting (306) into (168) yields the first part of (179). Given that $P_{\text{UF},2} = P_{\text{UF}|2}\,P_{\text{DF},2} \le P_{\text{DF},2}$, this part is valid when $P_{\text{UF},2} = A\,P_s \le P_{\text{DF},2}$ or, equivalently, $P_s \le P_{\text{DF},2}/A$, which by virtue of (162), (177), and (182) implies that $P_s \le P_{s^*}^{(2)}$. For $P_s \gg P_{s^*}^{(2)}$, $P_{\text{UF},2} \approx P_{\text{DF},2}$, which is the second part of (179). ■

APPENDIX F

Proof of Proposition 1.

The probability $P_{\text{UF},3}$ of data loss due to unrecoverable failures, given that three device failures have occurred, is obtained by unconditioning (208) on i and j, and using (201) as follows:

$$\begin{aligned}P_{\text{UF},3} &= \sum_{j=1}^{C}\sum_{i=j+1}^{C} P_{\text{UF}|3}(j,i)\,P_{j,i} \\ &\approx \sum_{j=1}^{C}\sum_{i=j+1}^{C}\left(1 - q_1^{j-1}\,q_2^{i-j}\,p_3^{C-i+1}\right)\frac{(m-1)(m-2)}{C^2}\,\left(\frac{\lambda}{\mu}\right)^2 f_R\,. \quad (308)\end{aligned}$$

It holds that

$$\sum_{j=1}^{C}\sum_{i=j+1}^{C} 1 = \sum_{j=1}^{C}(C-j) = \frac{C\,(C-1)}{2}\,, \quad (309)$$

and

$$\sum_{j=1}^{C}\sum_{i=j+1}^{C} q_1^{j-1}\, q_2^{i-j}\, p_3^{C-i+1} = \sum_{j=1}^{C} q_1^{j-1}\, q_2\, p_3^{C-j} \sum_{i=j+1}^{C} q_2^{i-j-1}\, p_3^{j-i+1} = \sum_{j=1}^{C} q_1^{j-1}\, q_2\, p_3^{C-j} \sum_{k=0}^{C-j} \left(\frac{q_2}{p_3}\right)^k = \sum_{j=1}^{C} q_1^{j-1}\, q_2\, p_3^{C-j}\, \frac{1-\left(\frac{q_2}{p_3}\right)^{C-j}}{1-\frac{q_2}{p_3}} = \frac{q_2\, p_3}{p_3-q_2} \sum_{j=1}^{C} q_1^{j-1} \left(p_3^{C-j} - q_2^{C-j}\right) = \frac{q_2\, p_3}{p_3-q_2} \left(\frac{q_1^C - p_3^C}{q_1-p_3} - \frac{q_1^C - q_2^C}{q_1-q_2}\right) . \quad (310)$$

Substituting (309) and (310) into (308) yields (209). ■

APPENDIX G

Proof of Corollary 15.

For small values of P_s, and from (127) , (148), and (207), it follows that

$$q_1^{j-1} \approx 1 - \frac{(j-1)(m-1)(m-2)}{2} P_s^2 , \quad (311)$$

$$q_2^{i-j} \approx 1 - (i-j)(m-2)\, P_s , \quad (312)$$

$$p_3^{C-i+1} \approx 1 - (C-i+1)(m-3)\, P_s . \quad (313)$$

Consequently,

$$\begin{aligned} & q_1^{j-1}\, q_2^{i-j}\, p_3^{C-i+1} \\ & \quad \approx 1 - [(i-j)(m-2) + (C-i+1)(m-3)]P_s + O(P_s^2) \\ & \quad \approx 1 - [(C+1)(m-3) - (m-2)j + i\,]\, P_s . \end{aligned} \quad (314)$$

Therefore,

$$\begin{aligned} & \sum_{j=1}^{C}\sum_{i=j+1}^{C} \left(1 - q_1^{j-1}\, q_2^{i-j}\, p_3^{C-i+1}\right) \\ & \quad \approx \sum_{j=1}^{C}\sum_{i=j+1}^{C} [(C+1)(m-3) - (m-2)j + i]P_s \\ & \quad \approx \sum_{j=1}^{C} (C-j)\,[(C+1)(m-3) - (m-2)j\,]P_s + \sum_{j=1}^{C}\sum_{i=j+1}^{C} i\, P_s \\ & \quad \approx [(C+1)(m-3)\,]P_s \sum_{j=1}^{C} (C-j) \\ & \qquad - (m-2) \sum_{j=1}^{C} [(C-j)\, j]\, P_s + \sum_{j=1}^{C}\sum_{k=0}^{C-j-1} (k+j+1)\, P_s \\ & \quad \approx [(C+1)(m-3)\,]P_s\, \frac{(C-1)\, C}{2} \\ & \qquad - (m-2) \sum_{j=1}^{C} [(C-j)\, j]\, P_s + \sum_{j=1}^{C}\sum_{k=0}^{C-j-1} (k+j+1)\, P_s \\ & \quad \approx \frac{(C-1)\, C\, (C+1)}{2} (m-3)\, P_s - (m-2) \sum_{j=1}^{C} [(C-j)\, j]\, P_s \\ & \qquad + \sum_{j=1}^{C} \left[(C-j)(j+1) + \frac{(C-j-1)(C-j)}{2}\right] P_s \\ & \quad \approx \frac{(C-1)\, C\, (C+1)}{2} (m-3)\, P_s - (m-3) \sum_{j=1}^{C} [(C-j)\, j]\, P_s \\ & \qquad + \frac{1}{2} \sum_{j=1}^{C} (C-j)\, P_s + \frac{1}{2} \sum_{j=1}^{C} (C-j)^2\, P_s \\ & \quad \approx \frac{(C-1)\, C\, (C+1)}{2} (m-3)\, P_s \\ & \qquad - (m-3)\, \frac{(C-1)\, C\, (C+1)}{6} P_s \\ & \qquad + \frac{1}{2}\, \frac{(C-1)\, C}{2} P_s + \frac{1}{2}\, \frac{(C-1)\, C\, (2\,C-1)}{6} P_s \\ & \quad \approx \frac{(C-1)\, C\, (C+1)}{3} (m-3)\, P_s \\ & \qquad + \frac{1}{2}\, \frac{(C-1)\, C}{2} P_s + \frac{1}{2}\, \frac{(C-1)\, C\, (2\,C-1)}{6} P_s \\ & \quad \approx \frac{(C-1)\, C}{6} \left[2\,(C+1)(m-3) + \frac{3}{2} + \frac{2\,C-1}{2}\right] P_s \\ & \quad \approx \frac{(C-1)\, C\, (C+1)}{6} (2\,m-5)\, P_s \end{aligned} \quad (315)$$

Substituting (315) into (308) yields the first part of (210). Given that $P_{\text{UF},3} = P_{\text{UF}|3}\, P_{\text{DF}_3} \leq P_{\text{DF}_3}$, this part is valid when $P_{\text{UF},3} = B\, P_s \leq P_{\text{DF}_3}$ or, equivalently, $P_s \leq P_{\text{DF}_3}/B$, which by virtue of (205) and (212) implies that $P_s \leq 3/(C\,(2m-5))$, which is of the same order as $P_{s^*}^{(2)}$. For $P_s \gg P_{s^*}^{(2)}$, $P_{\text{UF},3} \approx P_{\text{DF}_3}$, which is the second part of (210). ■

REFERENCES

[1] I. Iliadis, "Data loss in RAID-5 storage systems with latent errors," in Proceedings of the 12th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ), Mar. 2019, pp. 1–9.

[2] D. A. Patterson, G. Gibson, and R. H. Katz, "A case for redundant arrays of inexpensive disks (RAID)," in Proceedings of the ACM SIGMOD International Conference on Management of Data, Jun. 1988, pp. 109–116.

[3] P. M. Chen, E. K. Lee, G. A. Gibson, R. H. Katz, and D. A. Patterson, "RAID: High-performance, reliable secondary storage," ACM Comput. Surv., vol. 26, no. 2, Jun. 1994, pp. 145–185.

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

[5] I. Iliadis, D. Sotnikov, P. Ta-Shma, and V. Venkatesan, "Reliability of geo-replicated cloud storage systems," in Proceedings of the 2014 IEEE 20th Pacific Rim International Symposium on Dependable Computing (PRDC), Nov. 2014, pp. 169–179.

[6] M. Malhotra and K. S. Trivedi, "Reliability analysis of redundant arrays of inexpensive disks," J. Parallel Distrib. Comput., vol. 17, Jan. 1993, pp. 146–151.

[7] A. Thomasian and M. Blaum, "Higher reliability redundant disk arrays: Organization, operation, and coding," ACM Trans. Storage, vol. 5, no. 3, Nov. 2009, pp. 1–59.

[8] I. Iliadis, R. Haas, X.-Y. Hu, and E. Eleftheriou, "Disk scrubbing versus intradisk redundancy for RAID storage systems," ACM Trans. Storage, vol. 7, no. 2, Jul. 2011, pp. 1–42.

[9] V. Venkatesan, I. Iliadis, and R. Haas, "Reliability of data storage systems under network rebuild bandwidth constraints," in Proceedings of the 20th Annual IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Aug. 2012, pp. 189–197.

[10] J.-F. Pâris, T. J. E. Schwarz, A. Amer, and D. D. E. Long, "Highly reliable two-dimensional RAID arrays for archival storage," in Proceedings of the 31st IEEE International Performance Computing and Communications Conference (IPCCC), Dec. 2012, pp. 324–331.

[11] I. Iliadis and V. Venkatesan, "Most probable paths to data loss: An efficient method for reliability evaluation of data storage systems," Int'l J. Adv. Syst. Measur., vol. 8, no. 3&4, Dec. 2015, pp. 178–200.

[12] ——, "Expected annual fraction of data loss as a metric for data storage reliability," in Proceedings of the 22nd Annual IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Sep. 2014, pp. 375–384.

[13] ——, "Reliability evaluation of erasure coded systems," Int'l J. Adv. Telecommun., vol. 10, no. 3&4, Dec. 2017, pp. 118–144.

[14] I. Iliadis, "Reliability evaluation of erasure coded systems under rebuild bandwidth constraints," Int'l J. Adv. Networks and Services, vol. 11, no. 3&4, Dec. 2018, pp. 113–142.

[15] Amazon Simple Storage Service. [Online]. Available: http://aws.amazon.com/s3/ [retrieved: January 2019]

[16] D. Borthakur et al., "Apache Hadoop goes realtime at Facebook," in Proceedings of the ACM SIGMOD International Conference on Management of Data, Jun. 2011, pp. 1071–1080.

[17] R. J. Chansler, "Data availability and durability with the Hadoop Distributed File System," ;login: The USENIX Association Newsletter, vol. 37, no. 1, 2013, pp. 16–22.

[18] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in Proceedings of the 26th IEEE Symposium on Mass Storage Systems and Technologies (MSST), May 2010, pp. 1–10.

[19] Hitachi Global Storage Technologies, Hitachi Disk Drive Product Datasheets. [Online]. Available: http://www.hitachigst.com/ [retrieved: January 2019]

[20] E. Pinheiro, W.-D. Weber, and L. A. Barroso, "Failure trends in a large disk drive population," in Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST), Feb. 2007, pp. 17–28.

[21] A. Dholakia, E. Eleftheriou, X.-Y. Hu, I. Iliadis, J. Menon, and 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, no. 1, May 2008, pp. 1–42.

[22] I. Iliadis, "Reliability modeling of RAID storage systems with latent errors," in Proceedings of the 17th Annual IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Sep. 2009, pp. 111–122.

[23] V. Venkatesan and I. Iliadis, "Effect of latent errors on the reliability of data storage systems," in Proceedings of the 21th Annual IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), Aug. 2013, pp. 293–297.

[24] I. Iliadis and V. Venkatesan, "Rebuttal to 'Beyond MTTDL: A closed-form RAID-6 reliability equation'," ACM Trans. Storage, vol. 11, no. 2, Mar. 2015, pp. 1–10.

[25] V. Venkatesan and I. Iliadis, "A general reliability model for data storage systems," in Proceedings of the 9th International Conference on Quantitative Evaluation of Systems (QEST), Sep. 2012, pp. 209–219.

[26] I. Iliadis and X.-Y. Hu, "Reliability assurance of RAID storage systems for a wide range of latent sector errors," in Proceedings of the 2008 IEEE International Conference on Networking, Architecture, and Storage (NAS), Jun. 2008, pp. 10–19.

[27] V. Venkatesan and I. Iliadis, "Effect of codeword placement on the reliability of erasure coded data storage systems," in Proceedings of the 10th International Conference on Quantitative Evaluation of Systems (QEST), Sep. 2013, pp. 241–257.

[28] I. Iliadis and V. Venkatesan, "An efficient method for reliability evaluation of data storage systems," in Proceedings of the 8th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ), Apr. 2015, pp. 6–12.

[29] V. Venkatesan and I. Iliadis, "Effect of codeword placement on the reliability of erasure coded data storage systems," IBM Research Report, RZ 3827, Aug. 2012.