RAID+: Deterministic and Balanced Data Distribution for Large Disk Enclosures

Paper #36

1 More on RAID+ Reliability

1.1 Calculation Method of `MTTDL`

Here we follow the conventional way of calculating the `MTTDL` (mean time to data loss) of RAID-5, RAID-6, and RAID+ (i.e., RAID-x) using Continuous-time Markov Chains (CTMC) [1–3].

By assuming both disk failures and repairs follow the exponential distribution with rate μ and ν respectively, we first generate the state transition graph and the generator matrix $\boldsymbol{Q}$ of RAID-x's CTMC, where the state number is denoted as w. In the generator matrix, each element $q_{i,j}$ is defined as the transition speed from state i to state j. Then from the generator matrix, we need to get the embedded Markov chain of RAID-x's CTMC. Note that this embedded Markov chain is a discrete time Markov chain (DTMC). In this embedded Markov chain, the transition probability $p_{i,j}$ from state i to j can be calculated by Equation 1.

$$p_{i,j} = \frac{q_{i,j}}{\sum_{k \neq i} q_{i,k}} \tag{1}$$

As long as we get the state transition probability, the probability transition matrix $\boldsymbol{P}$ can be formed as $p_{i,j}$ is the value at row i and column j of the matrix. Denote that $\pi_i^e(x)$ is the probability that system will be at state i during step x in embedded Markov chain. These w $\pi_i^e(x)$, $i \in [0, w)$, form the vector $\boldsymbol{\pi}^e(x)$. In matrix $\boldsymbol{P}$, the state DL is an absorbing state, so each state in embedded Markov chain has stationary probabilities. Denote the stationary probability vector as $\boldsymbol{\pi}^e$ and stationary probability for state i as π_i^e.

Take n-disk RAID-5 as an example, we generate the state transition graph, matrix $\boldsymbol{Q}_{RAID-5}$, matrix $\boldsymbol{P}_{RAID-5}$ of RAID-5.

As depicted in Figure 1, the canonical RAID-5's CTMC contains three states, the state 0 is the normal state where n disks are working, the state 1 is the state for one disk failure, and the state DL is the state for data loss. As previous assumption, the disk failures follow the exponential distribution with rate μ. When one of n disk fails, the system will change to state 1. When another one of $n-1$ disks fails, the system will change to state DL. Thus the state transition from 0 to 1 follows the exponential distribution with rate $n\mu$, while the state transition from 1 to DL follows the exponential distribution with rate $(n-1)\mu$. As the disk repair makes the system change its state from 1 to 0, the state transition from 1 to 0 follow the exponential distribution with rate ν. Therefore, as shown in Equation 2, the variables in $\boldsymbol{Q}_{RAID-5}$ can be determined as $q_{0,1} = n\mu$, $q_{1,0} = \nu$, $q_{1,2} = (n-1)\mu$.

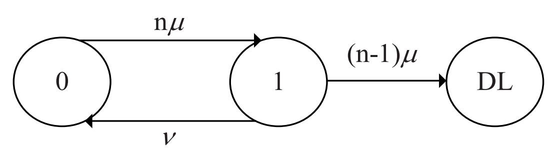

Figure 1: State transition graph of RAID5

$$\boldsymbol{Q}_{RAID-5} = \begin{bmatrix} -n\mu & n\mu & 0 \\ \nu & -\nu-(n-1)\mu & (n-1)\mu \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

As the transition probability of embedded Markov chain can be formulated as Equation 1, we can get the probability transition matrix of embedded Markov chain $\boldsymbol{P}_{RAID-5}$ in Equation 3, where $p_{i,j}$ is the value at row i and column j of the matrix. So, in Equation 3, we have $r_1 = \frac{\nu}{(n-1)\mu+\nu}$ and $s_1 = 1 - r_1$.

$$\boldsymbol{P}_{RAID-5} = \begin{bmatrix} 0 & 1 & 0 \\ r_1 & 0 & s_1 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

After the model is determined, we begin to calculate `MTTDL` of RAID-x. In embedded Markov chain, we denote $E(t_i^c)$ as the expected holding time of state i before the system goes into stationary. As described in equation 4, the `MTTDL` is the sum of $E(t_i^c)$, where $i \in [0, w-1)$. Note that the state DL doesn't have holding time as it's the dead state.

$$MTTDL_{RAID-x} = \sum_{i=0}^{w-2} E(t_i^c) \tag{4}$$

Now we need to get the $E(t_i^c)$ of each state except state DL. As the $E(t_i^c)$ is expected holding time before the system goes into stationary, it can be calculated using Equation 5. In Equation 5, $\pi_i^e(x)$ stands for the probability that the system is at state i in step x, and t_i is the average holding time of state i in one step. With distributive law, $E(t_i^c)$ can be regarded as the multiplication of t_i and $\sum_{x=0}^{\infty}\pi_i^e(x)$ and we denote $\sum_{x=0}^{\infty}\pi_i^e(x)$ as σ_i.

$$E(t_i^c) = \sum_{x=0}^{\infty} t_i \pi_i^e(x) = t_i \sum_{x=0}^{\infty} \pi_i^e(x) = t_i \sigma_i \tag{5}$$

Here, t_i for state i can be calculated by Equation 6.

$$t_i = E(H_i) = \frac{1}{\sum_{j\neq i} q_{i,j}} \tag{6}$$

So far, in order to calculate $MTTDL$, we only need to get the value of σ_i. First, we define the vector $\boldsymbol{\sigma}$ which have w elements, where the first $w-1$ elements are $\sigma_0, \sigma_1, \cdots, \sigma_{w-2}$, while the last element is set to 0. Then we get the $\boldsymbol{\pi}^{e0}(x)$, $\boldsymbol{\pi}^{e0}$ by setting the last element of $\boldsymbol{\pi}^{e}(x)$, $\boldsymbol{\pi}^{e}$ to 0. By setting all the elements in column $w-1$ of $\boldsymbol{P}$ to zero, we get a new matrix $\boldsymbol{P}^0$. As we have properties $\boldsymbol{\pi}^{e}(x+1) = \boldsymbol{\pi}^{e}(x)\boldsymbol{P}$ and $\boldsymbol{\pi}^{e} = \boldsymbol{\pi}^{e}\boldsymbol{P}$ in DTMC, $\boldsymbol{\pi}^{e0}$, $\boldsymbol{\pi}^{e0}(x)$, $\boldsymbol{P}^0$ also have the similar properties, we list these properties in equation 7. That's because state DL is the absorbing state.

$$\begin{aligned} \boldsymbol{\pi}^{e0}(x+1) &= \boldsymbol{\pi}^{e0}(x)\boldsymbol{P}^0, \\ \boldsymbol{\pi}^{e0} &= \boldsymbol{\pi}^{e0}\boldsymbol{P}^0 \end{aligned} \tag{7}$$

From the definition of σ_i, we have $\boldsymbol{\sigma} = \sum_{x=0}^{\infty}\boldsymbol{\pi}^{e0}(x)$. Thanks to absorbing state DL, all the elements in $\boldsymbol{\pi}^{e0}$ are 0, therefore, we have $\boldsymbol{\sigma} = \boldsymbol{\sigma} + \boldsymbol{\pi}^{e0}$. Using this and the properties in equation 7, we have equation 8 which gives us a solution to get the value of $\boldsymbol{\sigma}$.

$$\begin{aligned} \sum_{x=0}^{\infty}\boldsymbol{\pi}^{e0}(x) &= \sum_{x=0}^{\infty}\boldsymbol{\pi}^{e0}(x) + \boldsymbol{\pi}^{e0} \\ &= \boldsymbol{\pi}^{e0}(0) + (\sum_{x=1}^{\infty}\boldsymbol{\pi}^{e0}(x)) + \boldsymbol{\pi}^{e0} \\ &= \boldsymbol{\pi}^{e0}(0) + (\sum_{x=1}^{\infty}\boldsymbol{\pi}^{e0}(x-1))\boldsymbol{P}^0 \\ &= (\sum_{x=0}^{\infty}\boldsymbol{\pi}^{e0}(x))\boldsymbol{P}^0 + \boldsymbol{\pi}^{e0}(0) \\ &= (\sum_{x=0}^{\infty}\boldsymbol{\pi}^{e0}(x))\boldsymbol{P}^0 + \boldsymbol{\pi}^{e}(0), \\ \boldsymbol{\sigma} &= \boldsymbol{\sigma}\boldsymbol{P}^0 + \boldsymbol{\pi}^{e}(0) \end{aligned} \tag{8}$$

We rewrite equation 8 into element-wise simultaneous equation 9 which has $w-1$ equations.

$$\sigma_i = \sum_{j\neq i}\sigma_j p_{j,i} + \pi_i^e(0), i, j \in [0, w-1) \tag{9}$$

The initial probability vector of each state in RAID-x's embedded Markov chain $\boldsymbol{\pi}^{e}(0)$ can be formulated as $[1, 0, 0, \cdots]$, therefore, simultaneous equation 9 has unique solution.

Let's continue the example of RAID-5. First, we can get the average holding time t_i of each state except DL in RAID-5's Markov chain by equation 6. Therefore, the average holding time of state 0 is $t_0 = \frac{1}{n\mu}$ and the average holding time of state 1 is $t_1 = \frac{1}{(n-1)\mu+\nu}$. As the DL state is the dead state we assume its holding time is 0.

The σ_0 and σ_1 of RAID-5 can be calculated by simultaneous equation 10, which is derived from Equation 9.

$$\begin{cases} \sigma_0 &= r_1\sigma_1 + 1, \\ \sigma_1 &= \sigma_0 \end{cases} \tag{10}$$

The following equations give the rest inference of getting the `MTTDL` of RAID-5.

$$\sigma_0 = 1 + r_1\sigma_1 = 1 + \frac{\nu}{(n-1)\mu+\nu}\sigma$$

$$\sigma_1 = \sigma_0$$

$$\sigma_0 = \sigma_1 = \frac{\nu+(n-1)\mu}{(n-1)\mu}$$

$$\begin{aligned} MTTDL_{RAID-5} &= \sigma_0 t_0 + \sigma_1 t_1 \\ &= \frac{(n-1)\mu+\nu}{(n-1)\mu}(\frac{1}{n\mu} + \frac{1}{(n-1)\mu+\nu}) \\ &= \frac{(2n-1)\mu+\nu}{n(n-1)\mu^2} \\ &\approx \frac{\nu}{n(n-1)\mu^2} \end{aligned}$$

At the end of this subsection, we conclude above method as a recipe-like style. By following these steps, the `MTTDL` of RAID-x can be calculated.

1. Get the state transition graph and CTMC of RAID-x.
2. Get the generator matrix $\boldsymbol{Q}$.
3. Get the average holding time of each state $\boldsymbol{t}$ except state DL.
4. Get the transition probability matrix of embedded Markov chain $\boldsymbol{P}$.
5. Get the $\boldsymbol{\sigma}$ by simultaneous equation $\sigma_i = \sum_{j\neq i}\sigma_j p_{j,i} + \pi_i^e(0), i, j \in [0, w-1)$, where $\pi_i^e(0)$ is the initial probability of state i, and w is the number of states in Markov chain.
6. Finally, the `MTTDL` is calculated by $MTTDL = \sum_{i=0}^{w-2} t_i\sigma_i$.

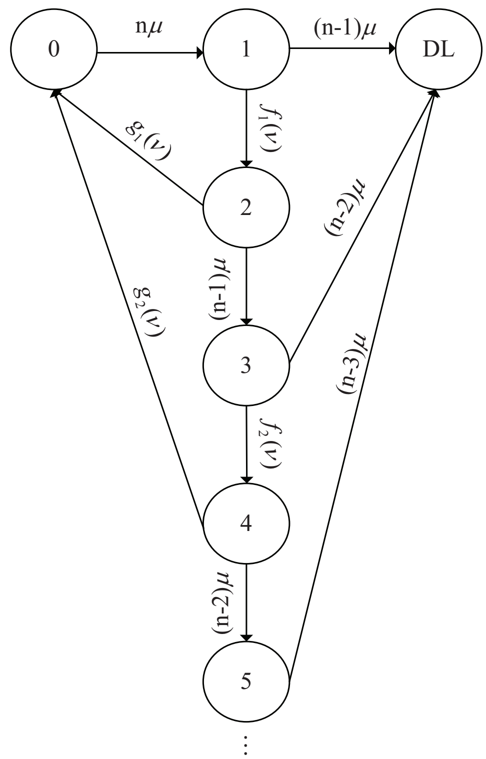

Figure 2: State transition graph of RAID+

1.2 RAID-5 organized RAID+'s Reliability and its comparison to RAID-50

From the lesson learned in subsection 1.1, we calculate the `MTTDL` of RAID-5 organized RAID+. Figure 2 gives us a part of state transition graph of RAID-5 organized RAID+. Suppose we have n disks and organized them into RAID+, in figure 2 there are 7 states where the state 0 stands for normal state that there are total n disks servicing for RAID+. State 1 is the state in which one disk fails and there are $(n-1)$ disks still online. As analyzed before, the $q_{0,1} = n\mu$. State 2 is a recovered state where parities are moved into the rest $n-1$ disks, therefore, there are $n-1$ disks online while RAID+ is still tolerant for one disk failure. The average repair time in RAID+ a function of conventional RAID-5's, therefore we use $f_1(\nu)$ to represent the disk repair rate with $n-1$ disks working online. Then disk adjunction is needed to maintain the disk numbers in RAID+ and we denote the adjunction rate for i disks as $g_i(\nu)$. In figure 2, we can see $q_{2,0} = g_1(\nu)$ as there is only one disk needed to maintain n-disk-RAID+. If another disk fails, State 2 will turn into state 3, then if recovered then the system would be in state 4 otherwise if another disk fails the data loss happens. The graph is only drawn with 5 abnormal state but in real application the abnormal state can be extended to tolerate more disk failures if the parities are stored in online disks.

State	t_i	r_i
0	$\frac{1}{n\mu}$	$\frac{1}{n\mu}$
1	$\frac{1}{f_1(\nu)+(n-1)\mu}$	$\frac{f_1(\nu)}{f_1(\nu)+(n-1)\mu}$
2	$\frac{1}{g_1(\nu)+(n-1)\mu}$	$\frac{g_1(\nu)}{g_1(\nu)+(n-1)\mu}$
3	$\frac{1}{f_2(\nu)+(n-2)\mu}$	$\frac{f_2(\nu)}{f_2(\nu)+(n-2)\mu}$
4	$\frac{1}{g_2(\nu)+(n-2)\mu}$	$\frac{g_2(\nu)}{g_2(\nu)+(n-2)\mu}$
5	$\frac{1}{f_3(\nu)+(n-3)\mu}$	$\frac{f_3(\nu)}{f_3(\nu)+(n-3)\mu}$
$\vdots$	$\vdots$	$\vdots$
$2i-1$	$\frac{1}{f_i(\nu)+(n-i)\mu}$	$\frac{f_i(\nu)}{f_i(\nu)+(n-i)\mu}$
$2i$	$\frac{1}{g_i(\nu)+(n-i)\mu}$	$\frac{g_i(\nu)}{g_i(\nu)+(n-i)\mu}$
$\vdots$	$\vdots$	$\vdots$

Table 1: The average holding time $\boldsymbol{t}$ and the recover transition probability $\boldsymbol{r}$ of the states in RAID+ (except state *DL*)

After forming the model of RAID+, we need to calculate the $\boldsymbol{t}$, $\boldsymbol{P}$, and $\boldsymbol{\sigma}$ to get the `MTTDL`. The average holding time $\boldsymbol{t}$ of the states in RAID+ (except state *DL*) and the probability transition matrix of embedded Markov chain $\boldsymbol{P}$ are given in table 1 and equation 3. Note that the $r_i s$ and the $s_i s$ are the recover transition probability and failure transition probability which indicates the transition probability of a safer state and a more dangerous state for each state. The values of $r_i s$ are given in table 1 and the values of $s_i s$ can be inferred with equation $r_i + s_i = 1$.

$$\boldsymbol{P}_{RAID-5+} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & r_1 & 0 & 0 & 0 & 0 & \cdots & s_1 \\ r_2 & 0 & 0 & s_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & r_3 & 0 & 0 & \cdots & s_3 \\ r_4 & 0 & 0 & 0 & 0 & s_4 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & s_5 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{11}$$

Then we get the value of $\boldsymbol{\sigma}$ according to the simultaneous equation $\sigma_i = \sum_{j \neq i} \sigma_j p_{i,j} + \pi_i^e(0)$. As we put the calculated figures into simultaneous equation, we get simultaneous equation 12.

$$
\begin{aligned}
\sigma_0 &= 1+\sigma_2 r_2+\sigma_4 r_4+\cdots \\
\sigma_1 &= \sigma_0 \\
\sigma_2 &= \sigma_1 r_1 \\
\sigma_3 &= \sigma_2 s_2 \\
\sigma_4 &= \sigma_3 r_3 \\
\sigma_5 &= \sigma_4 s_4 \\
&\vdots
\end{aligned} \tag{12}
$$

From the simultaneous equation 12, we can infer that $\sigma_0 = \frac{1}{1-\sum_{i=1}^{n}\prod_{j=1}^{2i-1}(I_{even}(j)r_i+(1-I_{even}(j))s_i)}$, in which $I_{even}(x)$ stands for the indicator function for even number, if the x is even, $I_{even}(x)=1$ and $I_{even}(x)=0$ if not. The rest elements in $\boldsymbol{\sigma}$ can be inferred by σ_0.

Since the result of $\boldsymbol{\sigma}$ is nearly unwritable in the paper, a simplified model of RAID+ with some approximation loss of `MTTDL` will be adopted. This simplified model is based on the simplification of the equation $MTTDL=\sum_{i=0}^{w-1}t_i\sigma_i$ and equation $\sigma_0 = \frac{1}{1-\sum_{i=1}^{n}\prod_{j=1}^{i}(I_{even}(j)r_i+(1-I_{even}(j))s_i)}$. In the first equation, we just remain the first three entries while in the second equation, we remain the first entry of $\sum_{i=1}^{n}\prod_{j=1}^{i}(I_{even}(j)r_i+(1-I_{even}(j))s_i)$ in denominator. Note that all the two simplifications will lead to a smaller $\boldsymbol{\sigma}$ and `MTTDL` but the model can't express the property that RAID+ could tolerate more failures after repairing.

In this paragraph, we will finish all the preparation for comparing the `MTTDL` of RAID-50 and RAID-5 organized RAID+, including the configurations, the analysis of the relationship between $f_i(\nu), g_i(\nu), \nu$. Assuming that there are n disks, RAID-50 organized k disks into RAID-5 and puts n/k RAID-5 into one RAID-0. In RAID+, the stripe size is set to k. Suppose we have S bytes contents in each disk, then if one disk fails, RAID-5 will repair with the average time of S/w_{disk} in which w_{disk} stands for the write bandwidth of one disk. In subsection 1.1, we assume that the disk repairs follow the exponential distribution with rate ν, therefore, the average repair time can be formulated by $1/\nu$. From above two analyses, we can get the equation $\nu = w_{disk}/S$. For RAID+, as the parity per disk is $S\frac{k}{n-i}$ when there are $n-i$ disks online, we can get $f_i(\nu)=\frac{(n-i)}{nk}\nu$. When adding the disks on RAID+, each disk is written S bytes with the speed w_{disk} thus $g_i(\nu)=\nu$.

Based on the preparation in the last paragraph, we can get the simplified `MTTDL` of RAID+ is equation 13. As released in subsection 1.1, the `MTTDL` of RAID-5 is $\frac{\nu}{k(k-1)\mu^2}$. When n/k RAID-5s organized into RAID0, the `MTTDL` of RAID-50 becomes $\frac{\nu}{n(k-1)\mu^2}$. Therefore, the ratio of `MTTDL` for two RAIDs $\frac{MTTDL_{RAID-5+}}{MTTDL_{RAID-50}}=\frac{k-1}{k}$. Considering the approximation when counting the estimated `MTTDL` of RAID+, the ratio is slightly larger than $\frac{k-1}{k}$.

$$
\begin{aligned}
MTTDL_{RAID-5+} &> \frac{1}{1-r_1}\frac{1}{n\mu}+\frac{1}{1-r_1}\frac{1}{(n-1)\mu+f_1(\nu)} \\
&= \frac{\nu}{nk\mu^2}
\end{aligned} \tag{13}
$$

1.3 RAID6 organized RAID+'s Reliability and its comparison to RAID-60

Following the conclusion in subsection 1.1, we calculated the `MTTDL` of RAID6 step by step:

1. Get the state transition graph and CTMC of RAID6, the state transition graph is shown in figure 3.
2. Get the generator matrix $\boldsymbol{Q}$, the generator matrix $\boldsymbol{Q}_{RAID6}$ is given in equation 14.
3. Get the average holding time of each state $\boldsymbol{t}$, the result is given in equation 15.
4. Get the transition probability matrix of embedded Markov chain $\boldsymbol{P}$, the result is given in equation 16.
5. Get the $\boldsymbol{\sigma}$ by simultaneous equation $\sigma_i = \sum_{j\neq i}\sigma_j p_{i,j}+\pi_i^e(0)$, where $\pi_i^e(0)$ is the initial probability of state i, the result is given in equation 17. The $p_{i,j}$ refers to the i,j-th element of $\boldsymbol{P}_{RAID6}$.
6. The `MTTDL` can be calculated by $MTTDL=\sum_{i=0}^{w-1}t_i\sigma_i$, where w is the number of states in Markov chain. The inference is given in equation 18.

$$
\begin{aligned}
&\boldsymbol{Q}_{RAID6} \\
&= \begin{bmatrix}
-n\mu & n\mu & 0 & 0 \\
\nu & -\nu-(n-1)\mu & (n-1)\mu & 0 \\
0 & \nu & -\nu-(n-2)\mu & (n-2)\mu \\
0 & 0 & 0 & 0
\end{bmatrix}
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
&\boldsymbol{t}_{RAID6} \\
&= \begin{bmatrix} \frac{1}{n\mu} & \frac{1}{\nu+(n-1)\mu} & \frac{1}{\nu+(n-2)\mu} \end{bmatrix}
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
&\boldsymbol{P}_{RAID6} \\
&= \begin{bmatrix}
0 & 1 & 0 & 0 \\
\frac{\nu}{\nu+(n-1)\mu} & 0 & \frac{(n-1)\mu}{\nu+(n-1)\mu} & 0 \\
0 & \frac{\nu}{\nu+(n-2)\mu} & 0 & \frac{(n-2)\mu}{\nu+(n-2)\mu} \\
0 & 0 & 0 & 1
\end{bmatrix}
\end{aligned} \tag{16}
$$

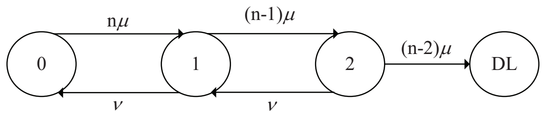

Figure 3: State transition graph of RAID6

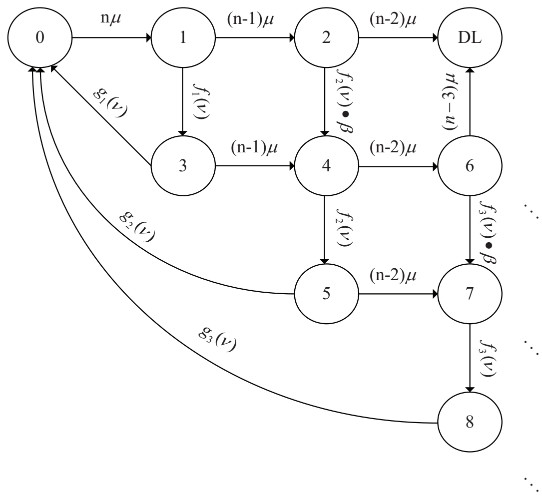

Figure 4: State transition graph of RAID6 based RAID+

$$\boldsymbol{\sigma}_{RAID6} = \left[1+\frac{p_{1,0}}{(1-p_{1,0})(1-p_{2,1})} \quad \frac{1}{(1-p_{1,0})(1-p_{2,1})} \quad \frac{1-p_{1,0}}{(1-p_{1,0})(1-p_{2,1})}\right] \tag{17}$$

$$\begin{aligned} &MTTDL_{RAID6} \\ &= \sum_{i=0}^{2} \sigma_i t_i \\ &= (1+\frac{(\mu(n-1)+\nu)(\mu(n-2)+\nu)}{\mu^2(n-1)(n-2)})\frac{1}{n\mu} \\ &+ (\frac{(\mu(n-1)+\nu)(\mu(n-2)+\nu)}{\mu^2(n-1)(n-2)})\frac{1}{(n-1)\mu+\nu} \\ &+ \frac{\mu(n-2)+\nu}{\mu(n-2)}\frac{1}{(n-2)\mu+\nu} \\ &= \frac{2(n-1)(n-2)\mu^2}{\mu^3 n(n-1)(n-2)} \\ &+ \frac{\mu^2 n(2n-3)+\mu n\nu+\nu^2+\mu(2n-3)\nu}{\mu^3 n(n-1)(n-2)} \\ &\approx \frac{\nu^2}{\mu^3 n(n-1)(n-2)} \end{aligned} \tag{18}$$

The `MTTDL` of RAID6 organized RAID+ is even more complicated than RAID-5 organized RAID+. The complexities which RAID+ brings to us are follows. First, the RAID+ will recover its fault-tolerant ability before the disk added on, which means the states of the same-level-tolerant are not the same state. This property of model is similar to RAID-5 organized RAID+. Second, in RAID6 organized RAID+, when two disks fail, the first level recovery will have less data to transfer, we denote that the ratio of original data transfer and optimized data transfer as β. with the property 6 in the main paper, we can get that $\beta = \frac{2(n-k)+k-2}{k-2}$. Figure 4 illustrates the state transition graph of RAID6 based RAID+, in this figure, we listed part of the states in RAID+ and here we'll explain them respectively. The state 0 is the normal state in which there are n disks in RAID+. In state 1, one disk fails and there are $n-1$ disks online. In state 2, two disks fails and there are $n-2$ disks online. In state 3, the disk failure in state 1 is recovered and there are $n-1$ disks online. In state 4, only one disk failure is recovered and there are $n-2$ disks online. In state 5, all the disk fails are recovered and there are $n-2$ disks online. In state 6, there are two disk failures and only $n-3$ disks online. In state 7, one disk failure is recovered in state 6 and only $n-3$ disks online. In state 8, disk failures are all recovered in state $6,7$ but only $n-3$ disks online. In state DL, more than two disks fail.

For the same simplification solution with RAID-5 organized RAID+ (with the `MTTDL` loss for RAID+), we only use those 10 states to estimate the `MTTDL` of RAID+. Follow the conventional way, we first get the $\boldsymbol{Q}_{RAID6+}$, for better fitting the size of paper, we use sparse matrix representation COO to present it. The $\boldsymbol{Q}_{RAID6+}$ is shown in table 2.

From the generator matrix $\boldsymbol{Q}_{RAID6+}$, we can get the $\boldsymbol{t}_{RAID6+}$ and $\boldsymbol{P}_{RAID6+}$. The results of them are listed in table 2 and table 3. Note that the r_i in $\boldsymbol{P}_{RAID6+}$ is given in 3 and the $s_i = 1 - r_i$.

After $\boldsymbol{P}_{RAID6+}$ is calculated, the simultaneous equation 20 can be listed. The simultaneous equation yields equation 19.

$$\begin{aligned} \sigma_1 = 1 \\ &/(1 - r_4 r_2 - r_5 r_6 s_4 r_2 \\ &- r_6 r_5 r_3 s_2 - r_7 s_5 r_3 r_2 \\ &- r_7 s_5 s_4 r_2 - s_6 r_5 s_4 r_2 - s_6 r_5 r_3 s_2) \end{aligned} \tag{19}$$

i	j	$q_{i,j}$	$p_{i,j}$
0	1	$n\mu$	1
0	0	$-n\mu$	0
1	2	$(n-1)\mu$	s_1
1	3	$f_1(\nu)$	r_1
1	1	$-(n-1)\mu - f_1(\nu)$	0
2	9	$(n-2)\mu$	s_2
2	4	$f_2(\nu)\beta$	r_2
2	2	$-(n-2)\mu - f_2(\nu)\beta$	0
3	0	$g_1(\nu)$	r_3
3	4	$(n-1)\mu$	s_3
3	3	$-(n-1)\mu - g_1(\nu)$	0
4	5	$f_2(\nu)$	r_4
4	6	$(n-2)\mu$	s_4
4	4	$-(n-2)\mu - f_2(\nu)$	0
5	0	$g_2(\nu)$	r_5
5	7	$(n-2)\mu$	s_5
5	5	$-(n-2)\mu - g_2(\nu)$	0
6	7	$f_3(\nu)\beta$	r_6
6	9	$(n-3)\mu$	s_6
6	6	$-(n-3)\mu - f_3(\nu)\beta$	0
7	8	$f_3(\nu)$	1
7	7	$-f_3(\nu)$	0
8	0	$g_1(\nu)$	1
8	8	$-g_1(\nu)$	0
9	9	0	1

Table 2: $\boldsymbol{Q}_{RAID6+}$ and $\boldsymbol{P}_{RAID6+}$

i	t_i	r_i
1	$\frac{1}{n\mu}$	0
2	$\frac{1}{(n-1)\mu+\frac{n-1}{k}\nu}$	$\frac{\frac{n-1}{k}\nu}{(n-1)\mu+\frac{n-1}{k}\nu}$
3	$\frac{1}{(n-2)\mu+\frac{n-2}{k}\frac{2(n-k)+k-2}{k-2}\nu}$	$\frac{\frac{n-2}{k}\frac{2(n-k)+k-2}{k-2}\nu}{(n-2)\mu+\frac{n-2}{k}\frac{2(n-k)+k-2}{k-2}\nu}$
4	$\frac{1}{(n-1)\mu+\nu}$	$\frac{\nu}{(n-1)\mu+\nu}$
5	$\frac{1}{(n-2)\mu+\frac{n-3}{k}\nu}$	$\frac{\frac{n-3}{k}\nu}{(n-2)\mu+\frac{n-3}{k}\nu}$
6	$\frac{1}{(n-2)\mu+\nu}$	$\frac{\nu}{(n-2)\mu+\nu}$
7	$\frac{1}{(n-3)\mu+\frac{n-3}{k}\frac{2(n-k)+k-2}{k-2}\nu}$	$\frac{\frac{n-3}{k}\frac{2(n-k)+k-2}{k-2}\nu}{(n-3)\mu+\frac{n-3}{k}\frac{2(n-k)+k-2}{k-2}\nu}$
8	$\frac{1}{\frac{n-3}{k}\nu}$	1
9	$\frac{1}{\nu}$	1

Table 3: $\boldsymbol{t}_{RAID6+}$ and $\boldsymbol{r}_{RAID6+}$

$$\begin{aligned}
\sigma_1 &= 1+\sigma_9+r_6\sigma_6+r_4\sigma_4 \\
\sigma_2 &= \sigma_1 \\
\sigma_3 &= s_2\sigma_2 \\
\sigma_4 &= r_2\sigma_2 \\
\sigma_5 &= r_3\sigma_3+s_4\sigma_4 \\
\sigma_6 &= r_5\sigma_5 \\
\sigma_7 &= s_5\sigma_5 \\
\sigma_8 &= r_7\sigma_7+s_6\sigma_6 \\
\sigma_9 &= \sigma_8
\end{aligned} \tag{20}$$

Different from RAID-5 organized RAID+, with simplification model the result of `MTTDL` of RAID6 is still too complicated to directly write here. Thus further approximation is adopted. We throw away all the states that have k failure tolerance and disks less than $n+k-2$ and just approximate the model as the disk adjunction is immediate. This approximation is not promise to have a bigger or smaller `MTTDL` thus it's only a reference method. We take the result of $f_i(x)$ and $g_i(x)$ in subsection 1.2 and set $n = 56$ and $k = 7$. The RAID-60's `MTTDL` is $\frac{\nu^2}{\mu^3 n(k-1)(k-2)}$ and the estimated `MTTDL` of RAID-6 organized RAID+ is in 21. Thus the ratio of two `MTTDL` is $\frac{MTTDL_{estimated.RAID-6+}}{MTTDL_{RAID-60}} = \frac{(2(n-k)+k-2)(k-1)}{k^2} \approx 12.61$.

$$\begin{aligned}
&MTTDL_{estimated.RAID-6+} \\
&= \frac{\nu_1\nu 2}{\mu^3 n(n-1)(n-2)} \\
&= \frac{\frac{2(n-k)+k-2}{k-2}\frac{n-2}{k}\nu\frac{n-1}{k}\nu}{\mu^3 n(n-1)(n-2)} \\
&= \frac{(2(n-k)+k-2)\nu^2}{\mu^3 nk^2(k-2)}
\end{aligned} \tag{21}$$

Last paragraph only gives an estimated ratio between two RAIDs. Here we use MATLAB to calculated the `MTTDL` for both RAIDs with different MTTF. The rest parameters are fixed as $n = 56$, $k = 7$, $\nu = \frac{1}{10000}$. We set MTTF to $3600 \times [100, 1000, 10000, 100000, 1000000, 10000000, 100000000]$ seconds and plot the `MTTDL` of both RAIDs. The result is shown in 5, in this figure, the smallest ratio of to RAIDs is $(\frac{MTTDL_{RAID-6+}}{MTTDL_{RAID-60}})_{min} = 12.6$. From this conclusion, we can infer that the approximation in the last paragraph is rather accurate and the reason comes that the probabilities of transmitting into most of the states are low.

References

[1] Kevin M Greenan. *Reliability and power-efficiency in erasure-coded storage systems*. University of Cal-

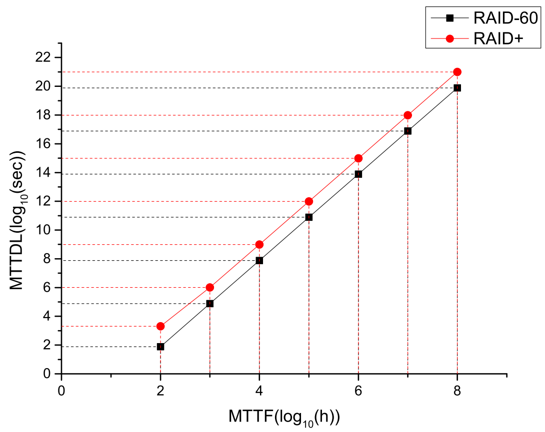

Figure 5: Comparison for RAID-60 and RAID+'s `MTTDL`

Config	252	1008	4032	16128
$RAID_H$	38.21	17.01	5.70	2.24
$RAID_R$	45.21	22.60	11.24	5.79

Table 4: Average COV of different window sizes. The first line of the table shows the window size (block number). The results are calculated through $\sum_{i=1}^{N} COV_i, N = \lfloor \frac{disk\_size}{window\_size} \rfloor$, where COV_i is the COV observed in window i.

ifornia, Santa Cruz, 2009.

[2] Kevin M Greenan, James S Plank, Jay J Wylie, et al. Mean time to meaningless: Mttdl, markov models, and storage system reliability. In *HotStorage*, 2010.

[3] Qin Xin, Ethan L Miller, Thomas Schwarz, Darrell DE Long, Scott A Brandt, and Witold Litwin. Reliability mechanisms for very large storage systems. In *Mass Storage Systems and Technologies, 2003.(MSST 2003). Proceedings. 20th IEEE/11th NASA Goddard Conference on*, pages 146–156. IEEE, 2003.