Theoretical analysis of the data recovery time in the fault-tolerant disk arrays RAID-5 and RAID-6 with data striping

P A Rahman¹, M D Kviring¹ and V E Bolnokin²

¹ Ufa State Petroleum Technological University, 1, Kosmonavtov Street, Ufa, 450062, Bashkortostan, Russian Federation
² Mechanical Engineering Research Institute, Russian Academy of Sciences, 4, Maliy Haritonievsky str., Moscow, 101000, Russian Federation

E-mail: pavelar@yandex.ru

Abstract. This scientific paper deals with the fault-tolerant disk arrays RAID-5 and RAID-6 with data striping. The structure of the disk arrays RAID-5 and RAID-6 and mathematical model of the redundant data coding in these disk arrays, based on the arithmetic of the binary Galois field GF(2⁸), are discussed. The obtained by the authors theoretical formulas for estimation of the data recovery time in the case of failure and replacement of single disk in the RAID-5 arrays, and data recovery time in the cases of failure and replacement of single and two disks in the RAID-6 arrays are also presented. Finally, the examples of theoretical estimation of the data recovery time in the RAID-5 and RAID-6 arrays are also discussed.

1. Introduction

In present days the data storage systems based on the disk arrays [1, 2] are widely used as the hardware platform for the modern information systems. To provide high performance and fault-tolerance of data storage systems the different kind of redundant arrays of inexpensive disks (RAID) [3, 4] are used. In particular, the fault-tolerant disk arrays RAID-5 and RAID-6 with data striping are widely used. One of the most important technical parameters of the RAID-5 and RAID-6 arrays is the data recovery time in case of failure and replacement of disks, which directly affects the performance and reliability of these disk arrays. Accordingly, theoretical analysis of the data recovery time in the RAID-5 and RAID-6 arrays is a quite urgent scientific task.

Within the scope of scientific research in the field of reliability of the disk arrays [5, 6], the authors analyzed the structure of the RAID-5 and RAID-6 arrays and mathematical model of the redundant data coding in these disk arrays, based on the arithmetic of the Galois Field GF(2⁸) [7-10]. As result, the authors obtained formulas for estimation of the data recovery time in the case of failure and replacement of single disk in the RAID-5 arrays, and data recovery time in the cases of failure and replacement of single and two disks in the RAID-6 arrays.

2. The structure and data coding technology of the RAID-5 arrays

The disk array RAID-5 consist of \( n \geq 3 \) identical disks and it is tolerant to the failure of any single disk. In case of failure of two or more disks, all data of the disk array will be lost.

The actual capacity of the RAID-5 array, which can be used for user data, is the \( (n-1) / n \) part of the total capacity of all disks in the array. On each of the disks the \( 1 / n \) part of disk space is used for
the redundant (control) data, which is calculated on the basis of user data, located on the other disks. Thus, in case of failure of any single disk the unavailable data can be calculated by using the user and control data of the remaining \(n-1\) disks. Figure 1 shows an example of distribution of user and control data in the RAID-5 array with \(n = 5\) disks.

![Disk array RAID-5](image)

**Figure 1.** Distribution of user and control data in the RAID-5 array.

The data coding technology of the RAID-5 array divides user and control data into blocks and places them in the «horizontal rows». Each of the «horizontal rows» contains \(n-1\) blocks \(D_a \ldots D_b\) with user data and one block \(S_{a-b}\) with control data. In turn, blocks consist of \(m\) bytes and each of the bytes in the control blocks are calculated by using the appropriate bytes in the user data blocks.

The number of the «horizontal rows» is determined by the disk capacity \(V\) (in bytes) and block size \(m\) (in bytes) and it is equal to \(\lceil V/m \rceil \cdot n\). Accordingly, the number of the «horizontal rows» is a multiple of \(n\). Moreover, the size of disk space, used for user and control data on each of the disks, is equal to \(\lceil V/m \rceil \cdot n \cdot m\) and this value slightly differs from the disk capacity \(V\). However, for the modern large disks (hundreds of gigabytes and more) this difference is negligible.

For each of the «horizontal rows» the bytes of the control block \(S_{a-b}\) are calculated by using the bytes of the appropriate \(n-1\) user data blocks \(D_a \ldots D_b\) and following calculation formula:

\[
\begin{align*}
  k &= 0 \ldots \lceil V/m \rceil \cdot n - 1; \\
  a &= (n - 1)k + 1; \quad b = (n - 1)(k + 1); \\
  j &= 0 \ldots m - 1; \\
  s_{a-b,j} &= d_{a,j} + \ldots + d_{b,j} = \sum_{i=a}^{b} d_{i,j}.
\end{align*}
\]

The bytes of the control block are calculated as the bitwise sum modulo 2 of the bytes of the \(n-1\) user data blocks, and it can be interpreted as summation of elements of the Galois Field \(GF(2^8)\) using the addition operation of the \(GF(2^8)\).

3. **Estimation of the data recovery time in the RAID-5 array in case of failure of single disk**

In case of failure of single disk in the RAID-5 array, all blocks of this disk become unavailable. Moreover, in some «horizontal rows» the user data block \(D_a\) becomes unavailable and in other «rows» the control block \(S_{a-b}\) becomes unavailable.

In the \(k\)-th «horizontal row» with the unavailable control block \(S_{a-b}\) this block is calculated using the available \(n-1\) user data blocks \(D_i, a \leq i \leq b, a = (n - 1)k + 1, b = (n - 1)(k + 1)\):

\[
\begin{align*}
  k &= 0 \ldots \lceil V/m \rceil \cdot n - 1; \\
  a &= (n - 1)k + 1; \quad b = (n - 1)(k + 1); \\
  j &= 0 \ldots m - 1; \\
  s_{a-b,j} &= d_{a,j} + \ldots + d_{b,j} = \sum_{i=a}^{b} d_{i,j}.
\end{align*}
\]
In the $k$-th «horizontal row» with the unavailable user data block $D_i$, this block is calculated using the available control block $S_{a-b}$ and $n-2$ user data blocks $D_i, a \leq i \leq b; i \neq u, \ a = (n-1)k + 1, \ b = (n-1)(k + 1)$:

$$
\begin{align*}
    j &= 0 \ldots m - 1; \\
    s_{a-b,j} &= \sum_{i=a}^{b} d_{i,j}.
\end{align*}
$$

Thus, unavailable blocks in all of the «rows» are calculated using the formulas (2), (3) and addition operation of the Galois Field $GF(2^8)$.

For theoretical estimation of the data recovery time in case of failure and replacement of single disk in the RAID-5 array, the authors analyzed the computational complexity of the formulas (2) and (3), and amount of the «horizontal rows», in which these formulas are used for calculation of the unavailable blocks. It was also considered that in each of the «rows» the $n-1$ blocks, required for calculation of the unavailable block, are located on different $n-1$ disks and they can be read simultaneously. With taking into account all of the previously mentioned, the authors obtained the following formula for theoretical estimation of the data recovery time (in hours) for the case of failure and replacement of single disk in the RAID-5 array:

$$
T_{RSSF} = \left\lceil \frac{V}{m} \cdot \frac{1}{n} \cdot \frac{n \cdot m}{3600} \cdot \frac{1}{v_{RD}} + \frac{1}{v_{WR}} + (n-1) \cdot \tau_{ADD} \right\rceil,
$$

where $n$ – number of disks; $V$ – disk capacity (in bytes); $m$ – block size (in bytes); $v_{RD}$ – linear read speed of disk (in bytes/sec); $v_{WR}$ – linear write speed of disk (in bytes/sec); $\tau_{ADD}$ – time (in seconds) of the addition operation for the elements of the Galois Field $GF(2^8)$.

4. The structure and data coding technology of the RAID-6 arrays

The disk array RAID-6 consist of $n \geq 4$ identical disks and it is tolerant to failure of any single disk, as well as to the failures of any two disks. In case of failure of three or more disks, all data of the disk array will be lost (figure 2).

![Figure 2](image-url)
redundant (control) data, which is calculated on the basis of user data, located on the other disks. Thus, in case of failure of any two disks the unavailable data can be calculated by using the user and control data of the remaining \( n - 2 \) disks. In case of failure of single disk, the unavailable data also can be calculated by using user and control data of the remaining \( n - 1 \) disks. Figure 2 shows an example of distribution of user and control data in the RAID-6 array with \( n = 5 \) disks.

The data coding technology of the RAID-6 array divides user and control data into blocks and places them in the «horizontal rows». Each of the «horizontal rows» contains \( n - 2 \) blocks \( D_a \ldots D_b \) with user data and two blocks with control data \( S_{a-b} \) and \( R_{a-b} \). In turn, blocks consist of \( m \) bytes and each of the bytes in the control blocks \( S_{a-b} \) and \( R_{a-b} \) are calculated by using the appropriate bytes in the user data blocks.

The number of the «horizontal rows» is determined by the disk capacity \( V \) (in bytes) and block size \( m \) (in bytes) and it is equal to \([V / m] \cdot n \cdot n - 1\). Accordingly, the number of the «horizontal rows» is a multiple of \( n \). Moreover, the size of disk space, used for user and control data on each of the disks, is equal to \([V / m] \cdot n \cdot m\) and this value slightly differs from the disk capacity \( V \). However, for the modern large disks (hundreds of gigabytes and more) this difference is negligible.

For each of the «horizontal rows» the bytes of the control blocks \( S_{a-b} \) and \( R_{a-b} \) are calculated by using the bytes of the appropriate \( n - 2 \) user data blocks \( D_a \ldots D_b \) and following calculation formulas:

\[
\begin{align*}
  k &= 0 \ldots [V / m] / n \cdot n - 1; \\
  a &= (n - 2)k + 1; \quad b = (n - 2)(k + 1); \\
  j &= 0 \ldots m - 1; \\
  \alpha &= 2; \\
  s_{a-b,j} &= d_{a,j} + \ldots + d_{b,j} = \sum_{i=a}^{b} d_{i,j}; \\
  \alpha^n &= \sum_{i=a}^{b} d_{i,j} \alpha^{i-a}. 
\end{align*}
\]

The bytes of the control block \( S_{a-b} \) are calculated as the bitwise sum modulo 2 of the bytes of the \( n - 2 \) user data blocks, and it can be interpreted as summation of elements of the Galois Field \( GF(2^8) \) using the addition operation of the \( GF(2^8) \).

The bytes of the control block \( R_{a-b} \) are calculated as the weighted summation of elements of the Galois Field \( GF(2^8) \) using the addition and multiplication operations of the \( GF(2^8) \). The powers of the primitive of element \( \alpha \) of the Galois Field \( GF(2^8) \) are used as the weights.

It is very important to note, that all the powers \( \alpha^n, \alpha^{i}, \ldots, \alpha^{254} \) of the primitive element \( \alpha \) of the Galois Field \( GF(2^8) \) are stored as the pre-calculated constants in memory of the RAID-array controller. Such approach provides high performance of calculations and excludes necessity of the multiplication operations for calculation of the powers of the primitive element.

5. Estimation of the data recovery time in the RAID-6 array in case of failure of single disk

In case of failure of single disk in the RAID-6 array all blocks of this disk become unavailable. In some «horizontal rows» the user data block \( D_b \) becomes unavailable, in other «rows» the control block \( S_{a-b} \) becomes unavailable, and in third «rows» the control block \( R_{a-b} \) becomes unavailable.

In the \( k \)-th «horizontal row» with the unavailable control block \( S_{a-b} \) this block is calculated using the available \( n-2 \) user data blocks \( D_i, a \leq i \leq b, a = (n-2)k+1, b = (n-2)(k+1) \):

\[
\begin{align*}
  k &= 0 \ldots [V / m] / n \cdot n - 1; \\
  a &= (n - 2)k + 1; \quad b = (n - 2)(k + 1); \\
  j &= 0 \ldots m - 1; \\
  \alpha &= 2; \\
  s_{a-b,j} &= d_{a,j} + \ldots + d_{b,j} = \sum_{i=a}^{b} d_{i,j}; \\
  \alpha^n &= \sum_{i=a}^{b} d_{i,j} \alpha^{i-a}. 
\end{align*}
\]
In the $k$-th «horizontal row» with the unavailable control block $R_{a-b}$ this block is calculated using the available $n-2$ user data blocks $D_i, a \leq i \leq b$, $a = (n-2)k+1$, $b = (n-2)(k+1)$:

$$\left\{ \begin{array}{l}
j = 0 \ldots m-1; \\
s_{a-b,j} = \sum_{i=a}^{b} d_{i,j}. \\ \end{array} \right. \quad (6)$$

In the $k$-th «horizontal row» with the unavailable user data block $D_a$ this block is calculated using the available control block $S_{a-b}$ and $n-3$ user data blocks $D_i, a \leq i \leq b; i \neq u$, $a = (n-2)k+1$, $b = (n-2)(k+1)$:

$$\left\{ \begin{array}{l}
j = 0 \ldots m-1; \\
r_{a-b,j} = \sum_{i=a}^{b} d_{i,j} \alpha^{i-a}. \\ \end{array} \right. \quad (7)$$

Thus, unavailable blocks in all of the «rows» are calculated using the formulas (6)-(8), and addition and multiplication operations of the Galois Field $\text{GF}(2^8)$.

For theoretical estimation of the data recovery time in case of failure and replacement of single disk in the RAID-6 array, the authors analyzed the computational complexity of the formulas (6)-(8), and amount of the «horizontal rows», in which these formulas are used for calculation of the unavailable blocks. It was also considered that in each of the «rows» the $n-2$ blocks, required for calculation of the unavailable block, are located on different $n-2$ disks and they can be read simultaneously. With taking into account aforesaid, the authors obtained the following formula for theoretical estimation of the data recovery time (in hours) for the case of failure and replacement of single disk in the RAID-6 array:

$$T_{\text{R6SF}} = \left[ \frac{V}{m} \right] \cdot \frac{1}{n} \cdot \frac{n \cdot m}{3600} \cdot \left( \frac{1}{v_{\text{RD}}} + \frac{1}{v_{\text{WR}}} + (n-2) \cdot \left( \tau_{\text{ADD}} + \frac{\tau_{\text{MUL}}}{n} \right) \right),$$

where $\tau_{\text{MUL}}$ – time (in seconds) of the multiplication operations for elements of the Galois Field $\text{GF}(2^8)$.

6. Estimation of the data recovery time in the RAID-6 array in case of failure of two disks

In case of failure of two disks in the RAID-6 array all blocks of these disks become unavailable. In some «horizontal rows» the control blocks $S_{a-b}$ and $R_{a-b}$ become unavailable, in other «rows» the control block $R_{a-b}$ and user data block $D_a$ become unavailable, in third «rows» the control block $S_{a-b}$ and user data block $D_a$ become unavailable, and in fourth «rows» the two user data blocks $D_a$ and $D_i$ become unavailable.

In the $k$-th «horizontal row» with the unavailable control blocks $S_{a-b}$ and $R_{a-b}$ these blocks are calculated using the available $n-2$ user data blocks $D_i, a \leq i \leq b$, $a = (n-2)k+1$, $b = (n-2)(k+1)$:

$$\left\{ \begin{array}{l}
j = 0 \ldots m-1; \\
s_{a-b,j} = \sum_{i=a}^{b} d_{i,j}; \quad r_{a-b,j} = \sum_{i=a}^{b} d_{i,j} \alpha^{i-a}. \\ \end{array} \right. \quad (10)$$
In the $k$-th «horizontal row» with the unavailable control block $R_{a-b}$ and user data block $D_a$ firstly the block $D_u$ is calculated using the available control block $S_{a-b}$ and $n-3$ user data blocks $D_i, a \leq i \leq b, i \neq u, a = (n-2)k+1, b = (n-2)(k+1)$. After that, the control block $R_{a-b}$ is calculated using the $n-2$ user data blocks:

$$d_{u,j} = s_{a-b,j} + \sum_{i=a}^{b} d_{i,j}; \quad r_{a-b,j} = \sum_{i=a}^{b} d_{i,j} \alpha^{i-a}. \quad (11)$$

In the $k$-th «horizontal row» with the unavailable control block $S_{a-b}$ and user data block $D_a$ firstly the block $D_u$ is calculated using the available control block $R_{a-b}$ and $n-3$ user data blocks $D_i, a \leq i \leq b, i \neq u, a = (n-2)k+1, b = (n-2)(k+1)$. After that, the control block $S_{a-b}$ is calculated using the $n-2$ user data blocks:

$$d_{u,j} = \frac{1}{\alpha^{u-a}} \left( r_{a-b,j} + \sum_{i=a}^{b} d_{i,j} \alpha^{i-a} \right); \quad s_{a-b,j} = \sum_{i=a}^{b} d_{i,j}. \quad (12)$$

In the $k$-th «horizontal row» with the two unavailable user data blocks $D_u$ and $D_v$, these blocks are calculated using the available control blocks $S_{a-b}$ and $R_{a-b}$, and $n-4$ user data blocks $D_i, a \leq i \leq b, i \neq u; i \neq v, a = (n-2)k+1, b = (n-2)(k+1)$:

$$d_{u,j}^{(k)} = \frac{1}{\alpha^{u-a} + \alpha^{v-a}} \left\{ \alpha^{a-u} \left( s_{a-b,j} + \sum_{i=a}^{b} d_{i,j} \right) + \left( r_{a-b,j} + \sum_{i=a}^{b} d_{i,j} \alpha^{i-a} \right) \right\};$$

$$d_{v,j}^{(k)} = \frac{1}{\alpha^{u-a} + \alpha^{v-a}} \left\{ \alpha^{a-u} \left( s_{a-b,j} + \sum_{i=a}^{b} d_{i,j} \right) + \left( r_{a-b,j} + \sum_{i=a}^{b} d_{i,j} \alpha^{i-a} \right) \right\}. \quad (13)$$

Thus, unavailable blocks in all of the «rows» are calculated using the formulas (10)-(13), and addition, multiplication and division operations of the Galois Field $GF(2^5)$.

For theoretical estimation of the data recovery time in case of failure and replacement of two disks in the RAID-6 array, the authors analyzed the computational complexity of the formulas (10)-(13), and number of «horizontal rows», in which these formulas are used for calculation of the unavailable blocks. It was also considered that in each of the «rows» the $n-2$ blocks, required for calculation of the two unavailable blocks, are located on different $n-2$ disks and they can be read simultaneously. Finally, it was considered that the division operation for the elements of the Galois field $GF(2^5)$ takes similar time as the multiplication operation. With taking into account aforesaid, the authors obtained the following formula for theoretical estimation of the data recovery time (in hours) for the case of failure and replacement of two disks in the RAID-6 array:

$$T_{RDF} = \left[ \frac{V}{m} \right] \cdot \frac{1}{n} \cdot \frac{n \cdot m}{3600} \left( \frac{1}{v_{RD}} + \frac{1}{v_{WR}} + \left( \frac{n-6}{n} \right) \cdot (2 \tau_{ADD} + \tau_{MUL}) \right). \quad (14)$$

7. Examples of theoretical estimation of the data recovery times
For theoretical estimation and comparison of the data recovery times of the RAID-5 and RAID-6 disk arrays the authors carried out calculations by using the obtained formulas (1.4), (2.5) and (2.10) for the following number of disks $n = 4 \ldots 12$ and parameters of the disks and RAID-array controller:
$V = 10^{12}$ bytes; $v_{RD} = 125 \cdot 10^6$ bytes/sec; $v_{WR} = 80 \cdot 10^6$ bytes/sec; $m = 65536$ bytes; $\tau_{ADD} = 10^{-9}$ sec; $\tau_{MUL} = 4 \cdot 10^{-9}$ sec.

Table 1 shows the results of theoretical estimation of the data recovery time $T_{RSSF}$ in case of failure and replacement of single disk in the RAID-5 array and data recovery times $T_{RRSF}$ and $T_{RD5}$ in cases of failure of single and two disks in the RAID-6 array.

| $n$   | $T_{RSSF}$, hours | $T_{RRSF}$, hours | $T_{RD5}$, hours |
|------|-------------------|-------------------|------------------|
| 4    | 6.53              | 6.81              | 9.86             |
| 5    | 6.81              | 7.19              | 12.03            |
| 6    | 7.08              | 7.55              | 14.03            |
| 7    | 7.36              | 7.88              | 15.93            |
| 8    | 7.64              | 8.19              | 17.78            |
| 9    | 7.92              | 8.50              | 19.58            |
| 10   | 8.19              | 8.81              | 21.36            |
| 11   | 8.47              | 9.10              | 23.12            |
| 12   | 8.75              | 9.40              | 24.86            |

Obviously, the data recovery time in the RAID-6 arrays in case of failure and replacement of two disks significantly more depends on the number of disks in comparison with the data recovery times in case of failure and replacement of single disk in the RAID-5 and RAID-6 arrays.

8. Conclusion

Thus, within the scope of this scientific paper the structure of the fault-tolerant disk arrays RAID-5 and RAID-6 and redundant data coding technology of these disk arrays, based on the arithmetic of the Galois Field GF(2^8), are presented. The obtained by the authors theoretical formulas for estimation of the data recovery time in case of failure and replacement of single disk in the RAID-5 arrays, and data recovery time in cases of failure and replacement of single and two disks in the RAID-6 arrays are also discussed. Finally, the examples of theoretical estimation of the data recovery times in the RAID-5 and RAID-6 arrays are also presented.

The obtained results can be used for assessment of efficiency of implementation of the data coding and recovery procedures in the modern RAID-5 and RAID-6 controllers.

Acknowledgements

The authors are grateful to professor I. I. Ladygin from Moscow Power Engineering Institute for scientific support and theoretical base in the field of data protection technologies.

References

[1] Nelson S 2011 Pro data backup and recovery (Apress)
[2] Thomasian A and Blaum M 2009 ACM Transactions on Storage 5(3) 7
[3] Schwarz T J E and Burkhard W A 1995 IEEE Transactions on Magnetics 31(2) 1161-6
[4] Elerath J G and Pecht M 2009 IEEE Transactions on Computers 58(3) 289-99
[5] Rahman P A and Novikova F S G D 2018 IOP Conf. Ser.: Mater. Sci. Eng. 327 022086
[6] Rahman P A and Novikova F S G D 2018 IOP Conf. Ser.: Mater. Sci. Eng. 327 022087
[7] Lidl R and Niederreiter H 1983 Finite fields (Reading, Addison-Wesley)
[8] Mullen G L and Panario D 2013 Handbook of Finite Fields (CRC Press)
[9] Blahut R E 1983 Theory and Practice of Error Control Codes (Reading, Addison-Wesley)
[10] Moon T K 2005 Error correcting coding: mathematical methods and algorithms (New Jersey, John Wiley & Sons Inc.)