

Application and Performance Analysis of EVENODD in RAID Architecture

Wei Ji Tao *

University of Electronic Science and Technology of China, Chengdu, 611731, China

* Corresponding Author Email: 2020190501010@std.uestc.edu.cn

Abstract. As storage systems are developed and used in different environments, people require more stable and reliable storage systems. As one of the main methods of storage system fault tolerance, EVENODD has received more and more attention. Compared with the traditional RS coding, EVENODD is completely based on the XOR operation, which is very important in the research of erasure correction. Firstly, the paper introduces the RAID architecture and the current typical and common erasure codes. Then, the possible prospects for using EVENODD and the coding implementation are described in detail. Besides, the paper compares the existing erasure codes from the perspective of encoding and decoding speed. Through the comparison and analysis, it is concluded that EVENODD has different degrees of defects in terms of fault tolerance and computational efficiency. How to improve the design of erasure codes with higher fault tolerance and computational efficiency based on EVENODD is an issue that deserves further research in the future.

Keywords: Storage system, RAID architecture, Erasure codes, EVENODD.

1. Introduction

Originally conceived in the late 1980s, the Redundant Array of Independent Disks (RAID) emerged as a response to the limitations of single-disk storage systems [1]. RAID configurations are often employed to enhance both the speed and dependability of data storage solutions. By merging several drives into one unified entity, RAID not only enhanced performance through parallelism but also introduced redundancy to safeguard against disk failures [1, 2]. Over the years, various RAID configurations, known as "levels", have been developed, each tailored to address specific needs regarding performance, redundancy, and cost [1, 2].

RAID is frequently utilized to boost both the efficiency and dependability of storage infrastructures. However, as data stores grow in size and complexity, so does the need for more robust fault-tolerance mechanisms [3]. Although a single disk failure in a RAID system can usually be managed using standard parity techniques, challenges arise when multiple disk failures occur simultaneously. Some basic reliability assessments have been conducted, illustrating why a single parity array might fall short of reliability requirements for certain applications, and highlighting the importance of contemplating the construction of an array capable of enduring two concurrent disk failures [3, 4]. Consequently, there's been a growing interest in large disk arrays and in the development of systems designed to preserve data, even in the face of simultaneous failures of multiple disks [3]. To address this, it's proposed to employ erasure-correcting codes [3, 4], which offer a superior error-correcting capacity compared to basic parity. In the context of coding theory, an erasure refers to an error with a known location [3]. In coding theory, to recover data from two erased disks, a minimum of two redundant disks are required, referred to as the Singleton bound [3, 5]. A logical approach to retrieve data lost from two disks would be to utilize the renowned Reed-Solomon codes [5]. According to [3], some very effective designs have been proposed. However, there is still a problem here. These approaches, while straightforward, still entail recursion during the encoding process and small write operations [3]. For instance, the above scheme requires an update of most redundant symbols every time the information symbols are updated [3]. But we hope that only a minimum number of redundant symbols need to be updated at this time.

EVENODD relies on straightforward exclusive-OR computations and independent parity checks, hence there's no recursion involved [3]. It has a simple encoding and decoding procedure. By using

EVENODD, the small write operation can be simplified. EVENODD is an efficient 2-erasure correcting code. Because it has only two redundancies to tolerate two failed disks, it is optimal [3, 4]. Therefore, it can be used in some applications requiring the correction of two erased symbols with minimal complexity [3].

The paper's structure is as follows: In the next section, some other algorithms that can be used for RAID structures, how these methods compare with EVENODD, and the advantages and limitations of EVENODD will be introduced. Then in the third part, the principle of EVENODD encoding will be explained. Later, in Part 4, applications of EVENODD used in RAID structures will be given. In the fifth part, the performance of EVENODD will be presented. Finally, in Section 6, we will make a concluding paragraph.

2. Related Work

EVENODD has excellent performance as an erasure code. But there are many other erasure codes out there so far.

Classic Reed-Solomon (RS) codes [6] are the sole maximum distance separable (MDS) encoding technique capable of supporting a flexible number of both data and redundant drives. RS codes are linear codes that are based on finite field arithmetic and can correct multiple symbol errors, not just single-bit errors [4, 6]. It is the encoding of polynomial operations (including addition and multiplication) on Galois domains for the corresponding domain elements. Due to its strong error correction capabilities, this code is popular in many applications, such as CDs, DVDs, QR codes, digital TV, etc. [6]. Reed-Solomon codes belong to a subset of non-binary cyclic codes and are based on polynomial interpolation. Through these techniques, Reed-Solomon codes generate check symbols for a given block of data. As shown in [4, 6], by using the generator matrix based on the Vandermonde matrix, the information can be encoded. However, EVENODD is more efficient than the Reed-Solomon (RS) scheme [3, 4]. Furthermore, the employment of the RS scheme entails computations over finite fields, leading to a more intricate implementation. The RS algorithm requires a larger number of operations during processing. In a disk array containing 15 disks, the implementation of EVENODD involves approximately half the number of exclusive-OR operations compared to what the RS scheme necessitates [3].

The complexity of multiplication and matrix inversion is an obstacle to the further development of standard RS coding. Are there alternative matrices to the Vandermonde matrix that can meet the criteria for RS coding's reconstruction algorithm? Is there a generator matrix that can streamline multiplication operations within Galois fields? The emergence of Cauchy Reed-Solomon (CRS) codes [7] gives an affirmative answer to these questions. CRS codes [7] are a variant of RS code, especially suitable for network coding and distributed storage systems, such as the application of erasure coding in distributed file systems. Like traditional RS codes, CRS codes are based on finite field arithmetic. However, it uses Cauchy matrices instead of traditional Vandermonde matrices, which makes encoding and decoding more efficient. It is also important to note that the CRS code operates on the entire strip [4, 7]. The differences between CRS codes and classic Reed-Solomon codes are shown in [4]. Compared with EVENODD, it has almost the same limitations as RS codes.

Row Diagonal Parity (RDP) [8] can be used in multiple device storage systems, such as disk array storage systems (RAID). It serves a crucial function in the realms of data preservation and information transfer, especially in environments where a high degree of reliability is required. RDP and EVENODD have some similarities. First, they are erasure codes designed for RAID6 storage systems, with two redundant disks to handle the situation where there are two erasures [4, 8]. Second, they both rely on XOR operations to generate parity, without requiring more complex mathematical operations [3, 4, and 8]. RDP has a lower computing cost than EVENODD. RDP excels in computational and I/O effectiveness, rendering it highly efficient both during regular functioning and in post-failure recovery [8]. It also requires fewer XOR operations compared to EVENODD for both parity construction and data reconstruction.

3. EVENODD Encoding Principle

The EVENODD encoding principle is based on the addition of two redundant disks to a RAID architecture. It employs simple exclusive-OR computations to generate the parity information for the data disks [3]. The parity information is computed using exclusive-OR operations between the data disks, and the parity disks are updated whenever a data disk is modified. This encoding process is highly efficient and solely relies on parity hardware, a component commonly found in standard RAID-5 controllers [3, 4]. Now, to briefly explain the encoding process of EVENODD, assume there are m data disks and two redundant disks used to store parity information, where m must be prime. It is crucial because, without this prerequisite, the scheme will fail [3]. However, the requirement for m to be prime isn't particularly stringent. If systems aim to store a variable number of disks, not strictly prime, the subsequent prime number after this variable number can be chosen while some disks are considered empty disks, whose information bits are set to be 0 [3]. Then, consider an array, m is a prime number. The last two disks are the redundant disks. The encoding method addresses the previously mentioned issue and exclusively uses XOR operations to calculate the redundancy. Here, some definitions in [3] must be considered. $\langle n \rangle_m = j$ if and only if $j \equiv n \pmod{m}$ and $0 \leq j \leq m - 1$. Let:

$$a_{l,m} = \bigoplus_{t=1}^{m-1} a_{l,t} \quad (1)$$

$$a_{l,m+1} = S \oplus \left(\bigoplus_{t=0, t \neq l+1}^{m-1} a_{\langle l-t \rangle_m, t} \right) \quad (2)$$

$$S = \bigoplus_{t=1}^{m+1} a_{m-1-t, t} \quad (3)$$

Here, the symbol $a_{i,j}$, where $0 \leq i \leq m-1$ and $0 \leq j \leq m+2$, is the i th data in the j th disk. For every l , $0 \leq l \leq m-2$. S is an important special sum. For EVENODD, redundancy is categorized into two types: horizontal redundancy and diagonal redundancy, as explained in [3]. Horizontal redundancy, represented by m th disk, is essentially the exclusive-OR sum of disks 0 through $m - 1$. Then, according to the formulas and Table 1, the calculation of diagonal redundancy can be explained clearly.

Table 1. A RAID6 example for EVENODD.

Disk	0	1	2	3	4	Horizontal redundancy	Diagonal redundancy
0	$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$	$a_{0,4}$	$a_{0,5}$	$a_{0,6}$
1	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{1,5}$	$a_{1,6}$
2	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,4}$	$a_{2,5}$	$a_{2,6}$
3	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{3,5}$	$a_{3,6}$

The data in the last column of Table 1 represents the diagonal redundancy, and the second to sixth columns represent the information stored in the information disk. According to the formulas, $a_{0,6}$ is an XOR sum of $a_{0,0}, a_{1,4}, a_{2,3}, a_{3,2}$ and S .? While $a_{1,6}$ is an XOR sum of $a_{0,1}, a_{1,0}, a_{3,3}, a_{2,4}$ and S . $a_{2,6}$ is an XOR sum of $a_{0,3}, a_{1,2}, a_{2,1}, a_{3,0}$ and S . As for $a_{3,6}$, it is an XOR sum of $a_{3,4}, a_{0,2}, a_{1,1}, a_{2,0}$ and S . Diagonal redundancy is all XOR sums of information in the data disks along different diagonals. It's essential to stress the assumption that the diagonals exclusively maintain either even or odd parity. It means that if S is ignored, it would not be possible to retrieve data. It is crucial for the MDS property of the EVENODD [3].

4. Applications of EVENODD Code

Various storage systems exhibit distinct requirements for error correction coding, and in this section, the possibilities of EVENODD applications in some different fields will be shown. Firstly, disk array systems, are distinguished by an extensive array of data disks paired with a relatively limited number of redundant disks. In such systems, computational complexity and speed often take

precedence. Error-correcting codes are typically implemented within disk controllers, making array codes better suited for this scenario. So EVENODD could be a big player in this field.

Next, for the P2P systems. These systems exhibit a multitude of data storage and redundancy options at a relatively low cost. However, due to the unique mobility of P2P networks, the likelihood of storage media errors or nodes leaving the network is notably high. Consequently, these systems demand error-correcting codes with robust fault-tolerance support. Considering these characteristics, EVENODD may not be the ideal choice for this type of application since its fault tolerance only supports the recovery of up to two disk errors. Nevertheless, exploring high fault-tolerant erasure correction codes based on EVENODD could become a significant avenue for future research in this domain.

Furthermore, there exist distributed storage systems. A distributed system refers to a network comprising numerous interconnected computers or servers dispersed across different geographical locations. These nodes collaborate to execute computational and data-processing tasks collectively. Typically, such systems are engineered to enhance performance, scalability, and reliability, facilitating the handling of extensive data and services. Distributed systems find applications in diverse domains such as cloud computing, large-scale data processing, distributed databases, distributed file systems, and various other sectors. In such systems, there is a substantial quantity of storage media, and high availability is imperative. Consequently, the fault tolerance and reconfiguration efficiency of codes play a pivotal role in the system's performance, with less emphasis on storage utilization. While EVENODD demonstrates commendable decoding efficiency, its relative lack of fault tolerance remains an aspect that requires enhancement within the system.

Lastly, in the context of archival storage, it is a storage approach employed for the preservation and administration of data across extended durations. This typically encompasses the storage of seldom-accessed data on relatively cost-effective and low-performance storage media, aimed at economizing high-performance storage resources and mitigating storage expenses. The paramount requirement for such systems is data reliability, necessitating the use of error-tolerant Error Correction Codes (ECCs) suitable for this category of system. Consequently, the current EVENODD, which provides support for a maximum of two disk errors, proves unsuitable for this application.

5. Performance evaluation and experiments

To evaluate the performance of various erasure codes, it is crucial to consider fault tolerance, storage utilization, and computational efficiency. Among these, the computing efficiency during encoding and reconstruction holds significant importance. This section will provide a comparison and analysis of prevalent erasure codes in this regard. First, fault tolerance forms the foundation of erasure coding algorithms. Simultaneously, numerous erasure codes impose specific criteria regarding the quantity of disks. The fault tolerance of many erasure codes is not very high, but there are also some codes that have good fault tolerance. But this encoding also has its own problems. For example, While RS coding can accommodate various fault tolerance requirements, it has significantly poor encoding efficiency. Moreover, RS has extremely strict requirements for generating matrices, making the construction of RS coding difficult. At present, the calculations of the constructed RS are known to be relatively complex. It is utilizing erasure coding results in redundant data, leading to added space overhead. Addressing how to maximize storage resources and enhance space efficiency has emerged as a critical concern in erasure coding. Besides, MDS property is also important.

The test experiments are comprehensively outlined in [4] and will not be reiterated here. That is, data words k equals 6 and coding words m equals 2. To conduct experiments, several open-source libraries were used, including Luby, Zfec, Jerasure and Cleversafe [4]. Open-source versions of EVENODD and RDP coding do not currently exist; however, they can be developed based on the research outlined in [4, 10]. The experimental results for encoding performance and decoding performance are shown in Figure 1 and Figure 2. Here, w , according to [3, 4, 6-8], in the RS code,

means that the stripe unit is a w -bit word, while in CRS, RDP and EVENODD, it implies that the stripe dimensions are w multiplied by the size of each packet.

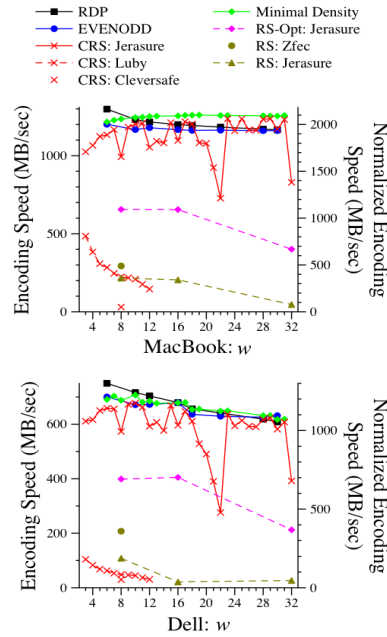


Figure 1. Encoding performance for $k=6$ and $m=2$ [4].

According to Figure 1, the data for MacBook and Dell are different. However, the performance of CRS coding implementations by Luby and Cleversafe is notably inferior to Jerasure for the encoding procedures in a two-machine setup. RDP, EVENODD and another RAID6 code, Minimal density RAID6 codes [9], exhibit significantly higher speed compared to the general-purpose codes. RDP encoding speed is the fastest at $w=6$. The encoding speed of EVENODD and RDP decreases with the increase of w . When $w=6$, the encoding speeds of EVENODD and Minimal Density RAID6 codes are equal in both two machines, and after that, it is always faster than EVENODD on MacBook. But on Dell, after w is greater than or equal to 16, there is no obvious difference in the encoding speed of these three RAID6 codes.

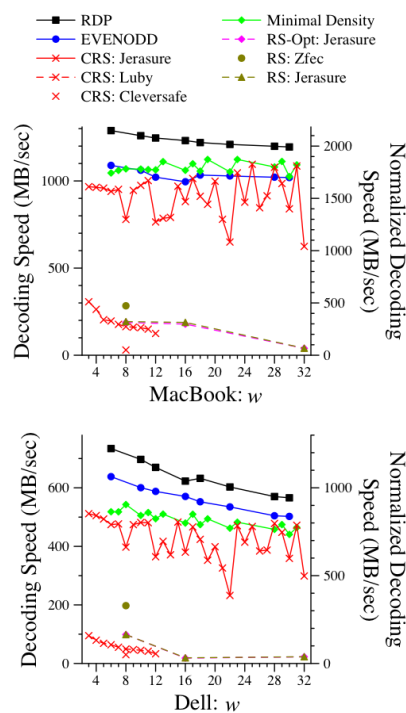


Figure 2. Decoding performance for $k=6$ and $m=2$ [4].

As for decoding performance, RDP was the fastest on both machines. At the same time, the three RAID6 codes are also much faster than the general-purpose codes. The decoding speed of EVENODD decreases as w increases on both machines. On a MacBook, it's always faster than minimal density and slower than RDP. On Dell, its speed is generally slower than minimal density. In some cases, the two are equally fast, such as when $w=10$.

6. Conclusion

As computer technology continues to advance, and storage systems experience significant growth, the focus on ensuring high reliability and availability has drawn increased attention from researchers in the field of storage systems. This article introduces EVENODD, a typical and widely used erasure code, and provides a comparative analysis of EVENODD against other existing erasure coding technologies. The analysis is conducted from the perspective of various critical performance indicators for evaluating the effectiveness of erasure codes. RS coding is one of the earliest erasure code types, it has undergone comprehensive research, leaving little room for further improvement. Conversely, array codes, exemplified by EVENODD, have garnered significant attention in recent years due to their reliance on XOR operations. Their simplicity and ease of implementation offer ample opportunities for continued research. Within the realm of array codes, various proposed RAID6 codes invariably exhibit some limitations, leaving no single erasure code universally recognized as the optimal solution for RAID6. Consequently, research into RAID6 encoding will likely remain a focal point in the foreseeable future. Simultaneously, as large-scale storage systems continue to evolve and expand, erasure codes with low fault tolerance rates are no longer adequate to meet the high-reliability requirements of these systems. Consequently, erasure codes with higher fault tolerance rates are destined to become a key focus of future development.

EVENODD, which is considered the precursor of array codes, possesses significant untapped development potential that should not be underestimated. It is particularly useful in RAID6 architectures. Its encoding process is highly efficient and relies only on parity hardware, a component commonly found in RAID-5 architecture. The EVENODD scheme finds applicability in systems that necessitate substantial symbols and relatively concise codes. The scheme is also optimal in redundancy, featuring highly efficient encoding and decoding algorithms, rendering it a valuable choice for applications requiring low-complexity correction of two erased symbols.

EVENODD may be used in data centers and cloud storage areas. In data centers and cloud storage environments, data reliability and durability are critical. EVENODD encoding can provide additional fault tolerance to help protect data stored in RAID systems, especially in large-scale storage environments where hard drive failures are common. When EVENODD encoding technology was proposed, it provided a relatively simple and efficient error correction solution for RAID6 systems, especially in scenarios where dual hard disk failures need to be dealt with. However, with the continuous development of technology and the emergence of new error correction coding schemes, EVENODD coding may face some challenges and opportunities. In the future, people may need erasure codes that can handle more disk errors. The future development of EVENODD should focus on the ability to handle multi-disk errors, while further reducing computational complexity and improving efficiency.

References

- [1] D. Patterson, G. Gibson, and R. Katz.: A case for redundant arrays of inexpensive disks (RAID). ACM SIGMOD 17(3), 109-116 (1998).
- [2] P. Chen et al.: RAID: high-performance, reliable secondary storage, ACM Computing Surveys 26 (2), 145-185, (1994).
- [3] M. Blaum, J. Brady, J. Bruck and Jai Menon.: EVENODD: an efficient scheme for tolerating double disk failures in RAID architectures. IEEE Transactions on Computers 44(2), 192-202 (1995).

- [4] J. S. Plank, J. Luo, C. D. Schuman, L. Xu, and Z. W. O'Hearn: A performance evaluation and examination of open-source erasure coding libraries for storage. In 7th USENIX Conference on File and Storage Technologies (FAST), pp. 253-265, 2009.
- [5] F. J. MacWilliams, N. J. A. Sloane.: The Theory of Error-Correcting Codes. North-Holland, Amsterdam, the Netherlands (1977).
- [6] I. S. Reed and G. Solomon.: Polynomial Codes over Certain Finite Fields. Journal of the Society for Industrial and Applied Mathematics 8(2), 300-304 (1960).
- [7] J. Blomer, M. Kalfane, M. Karpinski, R. Karp, M. Luby and D. Zuckerman. An XOR-Based Erasure-Resilient Coding Scheme. Technical Report TR-95-048, International Computer Science Institute, 1995.
- [8] P. Corbett, B. English, A. Goel, T. Grcanac, S. Kleiman, J. Leong, and S. Sankar.: Row Diagonal Parity for Double Disk Failure Correction. FAST-2004: 3rd Usenix Conference on File and Storage Technologies, San Francisco, CA (2004).
- [9] M. Blaum and R. M. Roth.: On Lowest Density MDS Codes. *IEEE Transactions on Information Theory*, 45(1), 46- 59, (1999).
- [10] C. Huang, J. Li and M. Chen.: On Optimizing XOR-Based Codes for Fault-Tolerant Storage Applications. ITW'07, Information Theory Workshop, IEEE, Tahoe City, CA, September, 2007, pp. 218-223.