Data Storage and Reliability Analysis on the Internet of Things Based on Pyramid Code

Xudong Zhao *

School of Engineering & Applied Science, George Washington University, Washington, DC, USA

* Corresponding author: xudong.zhao@gwmail.gwu.edu

Abstract. As the IoT (Internet of Things) develops, computers and devices generate a lot of data that needs to be processed and stored. Pyramid code, as an update of the Reed-Solomon code, is suitable for correcting data failure on various sides as an erasure code. Not only does Pyramid code reduce the cost of computation, but it can also meet the reliability and security requirements of IoT. The convenience of the Pyramid code is evaluated by comparing the (15, 12) RS code to the (16, 12) Pyramid code in this paper. Then, section 4 investigates how Pyramid code can match the needs of reliability, security, cost-effectiveness, etc. After that, the following section aims to give the applications of Pyramid code in IoT. There are many applications and this section concludes the several outstanding scenarios, such as healthcare, smart city, etc.

Keywords: IoT, Pyramid Code, Data storage.

1. Introduction

IoT is the term used to describe the connection of devices or subjects to the internet, which enables them to gather information without any human interference. IoT includes many parts, for example, connected devices, data collection and exchange, communications, cloud integration, and applications. The number of interconnected devices is still increasing and is expected to reach 41.6 billion by 2025 [1]. Data, which includes different types (textual data, binary data, relational data, image data. video data, etc.) and different formats (JSON, MessagePack, Protobuf, CBOR, etc.), plays a critical role in IoT. IoT is expected to generate large amounts of data from diverse locations that are aggregated at very high velocity. It demands better methods for indexing, storing, and processing such data which in turn requires developing a technique that converts this data into a knowledge base [2].

IoT devices, ranging from simple sensors to sophisticated cameras, generate data continuously. Therefore, both researchers and engineers are faced with the challenge of handling this massive heterogeneous data in highly distributed environments, especially in cloud platforms [3]. When storing vast amounts of data, ensuring its security, vulnerability, and velocity is essential. IoT devices, sometimes, are broken, fail, or attacked. Also, real-time data generation requires repaid storage and processing date solutions. As a result, recovery of failure data and express data processing in both encoding and decoding become a challenging thing.

As the IoT system continues its speedy development, ensuring the reliability of data becomes more and more urgent. Pyramid codes, a type of erasure coding, offer an efficient, flexible, and reliable method to protect the data from loss by adding redundancy, which refers to backup data. In conclusion, pyramid codes are an irreplaceable tool in the world of IoT.

The following is how this paper is structured. Section 2 will focus on the introduction of erasure coding and comparison of traditional Reed-Solomon codes and pyramid codes. Section 3 will explain the fundamentals of pyramid codes and their improvements for storing data. Section 4 will explore how pyramid codes can meet the needs of data storage and its application in the IoT system. Section 5 is going to use the experiment to show the performance of pyramid codes. Section 6 gives the numerical data to analyze the performance of Pyramid code. Section 7 concludes the advantages of pyramid codes and the future of their development.

2. Research on Erasure Coding

In settings such as that of a large data center, an important consideration is the efficient repair of a failed node. Efficient repair calls for erasure codes that in the face of node failure, are efficient in terms of minimizing the amount of repair data transferred over the network, the amount of data accessed at a helper node as well as the number of helper nodes contacted [5]. For example, Windows Azure Storage must erasure coding for data storage. Windows Azure Storage (WAS) is a scalable cloud storage system that has been in production since November 2008. [6]. Erasure coding can be classified into various types, such as Turbo Codes, Low-Density Parity-Check Codes, Regenerating Codes, and so on. Reed-Solomon code is one of the most dominant error-correcting codes used in IoT ecosystems. In the diverse IoT devices, Reed-Solomon code can always be an efficient method for repairing errors effectively. When the devices, with limited computational power, are resource-constrained, RS code offers a balance between error-fixing ability and computational complexity. WSNs (wireless sensor networks), IoT data storage, and IoT device firmware updates are all suitable for RS code, which can provide them with reliable data communication.

Reed-Solomon codes have been an integral part of the telecommunications revolution in the last half of the twentieth century. [7] The advantage of RS code is that it has a lower storage cost than a backup strategy with the same error-tolerant capability. Based on this, extending the RS code by splitting the last parity disk into two disks is possible to get a new coding scheme called Pyramid code, which can reduce reconstruction cost and increase the error-tolerant ability [8]. Lower cost means higher profit, especially in this huge IoT ecosystem.

3. Principle of Pyramid Codes

A classical system using Pyramid codes is shown in Fig. 1.

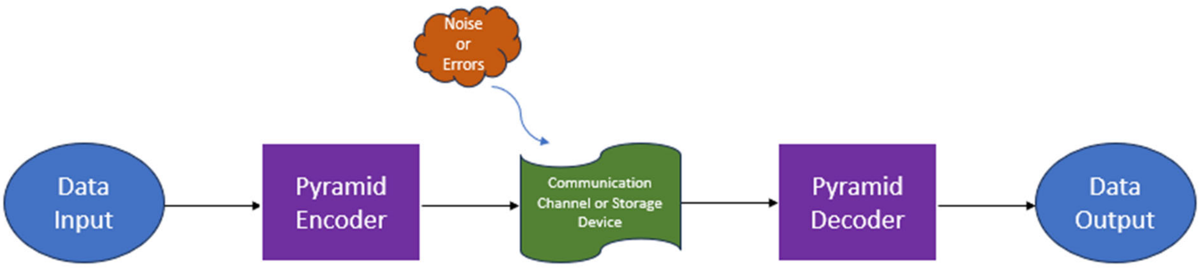

Figure 1. Data System

A block of digital data is taken by the Pyramid encoder and redundancy is added. Errors happen during transmission or storage due to several reasons including noise, device broken, etc. Every block is processed by the Pyramid decoder to retrieve the original data. The properties of Pyramid codes determine how many failures can be fixed. The definition of a Pyramid Code is that it contains n, k, and s-bits symbols, which are similar to Reed-Solomon codes. The pyramid code (16, 12) is used as an illustration. Fig. **2** shows it in detail.

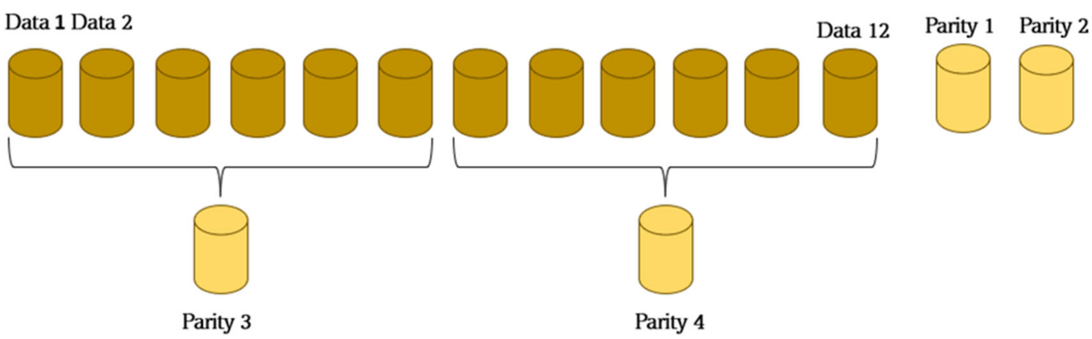

Figure 2. (16, 12) Pyramid Codes

As in Fig. 2, there are 12 data disks and 4 parties, which represent the additional data that is derived from the original data. Parity 1 and 2 data are both derived from data disk 1 to data disk 12. Parity 3 is derived from data disk 1 to data disk 6. Parity 4 is derived from data disk 7 to data disk 12. Four polynomials match the four parities accordingly and are shown below. p_{1k} to p_{4k} represent parity 1 to parity 4. d_{1k} to d_{12k} represent data disk 1 to 12. k is an index.

$$p_{1k} = d_{1k} + d_{2k} + \cdots + d_{12k} \tag{1}$$

$$p_{2k} = d_{1k} + 2d_{2k} + 2^2 d_{3k} + \cdots + 2^{11} d_{12k} \tag{2}$$

$$p_{3k} = d_{1k} + 4d_{2k} + 4^2 d_{3k} + \cdots + 4^5 d_{6k} \tag{3}$$

$$p_{4k} = 4^6 d_{7k} + 4^7 d_{8k} + \cdots + 4^{11} d_{12k} \tag{4}$$

4. Data storage requirement in the environment of IoT

IoT is an interconnected system of devices, which are capable of accumulating and trading data. As more IoT tools come to be extra prominent, their data production leads to considerable storage space concerns. Here are the key demands in an IoT setting:

Security: IoT applications are becoming part of personal lives and data collected is rather sensitive and private to an individual. Privacy and Security issues must be addressed in all of these environments [9]. The recent voluminous DDoS attacks (October 2016) on DYN's servers brought down many popular online services in the US, letting us know what can happen when attackers can leverage up to 150,000 unsecured IoT devices as malicious terminals [10]. Pyramid codes provide redundancy without replicating the exact data. By using pyramid codes to encode data, you can store it across multiple IoT devices or servers. The original data can still be reconstructed from the remaining pieces if a device fails or data is corrupted. Also, Distributed storage solutions can be advantageous in IoT environments. A network of devices can store data instead of relying on a central server. Pyramid codes can optimize this storage by guaranteeing data availability even when some devices are not operational.

Scalability: With billions of IoT devices connected, the amount of data generated is monumental. Storage solutions must scale quickly in response to this rapid data generation; Pyramid Codes were specifically created for large distributed systems - ideal candidates for accommodating IoT device-generated data.

Durability and Reliability: Data loss in IoT scenarios can have far-reaching ramifications ranging from minor inconvenience to catastrophic outcomes in critical applications like healthcare or transportation. Pyramid Codes have been specifically created to accommodate for failures; should any parts of data become unavailable they enable reconstruction to ensure optimal durability and reliability for high levels of durability and reliability.

Latency: Many IoT applications, particularly time-sensitive ones like real-time monitoring systems, require timely data access. By carefully placing encoded fragments to optimize for speed of retrieval and ensure low latency data access for certain IoT apps such as real-time monitoring systems or fleet management applications, Pyramid Codes provide support for fulfilling such needs.

Cost-Efficiency: Due to the massive volume of data that IoT devices produce, storage solutions must be cost-efficient. By applying the Pyramid code instead, costs will decrease while details will be covered during its performance analysis.

Data Integrity: Because data can become corrupt over time, mechanisms to ensure its long-term integrity are essential. Erasure codes like Pyramid Codes offer ways of keeping files intact even as individual storage nodes fail or become compromised within distributed systems.

5. Application of Pyramid Code in Data Storage of IoT

5.1. Smart Cities

In smart cities, various sensors and devices collect continuous data on everything from traffic management systems to utility monitoring services - which plays a pivotal role in everyday operations. Consider traffic cameras and sensors spread throughout a city that continuously feeds data to both central and decentralized storage systems, often unattended for extended periods. Due to its criticality and large amount of data being transmitted between storage systems, any storage fault could cause serious disruptions and problems that require urgent resolution.

Pyramid codes enable cities' IT infrastructures to store data efficiently across multiple nodes without needing all nodes' access, thus guaranteeing traffic management systems can gain continuous access to necessary data.

5.2. Industrial IoT (IoT):

Manufacturing and industrial environments employ machines equipped with sensors that continuously assess their health, performance, and environmental conditions. At a large manufacturing plant, every machine sends operational data directly into one central system for analysis by predictive maintenance software. This information serves an integral purpose in preventative maintenance strategies.

Pyramid codes offer several applications in industry and agriculture that could prove highly advantageous: maintenance data access can still be achieved in case certain storage nodes fail, thus minimizing costly downtimes.

5.3. Agriculture:

IoT devices have quickly become indispensable tools in precision agriculture, where sensors monitor soil conditions, weather patterns, and crop health. A large farm uses IoT devices to track soil moisture levels to optimize irrigation practices.

These sensors use Pyramid codes to store sensor data distributed; in case one or more storage nodes fail due to outdoor conditions, however, their irrigation systems still can access all required information to function optimally.

5.4. Healthcare

IoT applications in healthcare involve wearable devices that monitor patients' vital signs and transmit them for medical analysis. Imagine patients in remote locations wearing devices that send health metrics directly to a clinic nearby - potentially saving lives!

At our clinic, data storage uses Pyramid codes so in case of storage problems patient data remains accessible ensuring continuous monitoring and timely medical intervention.

5.5. Energy Sector

Smart grids and renewable energy sources often utilize IoT devices for monitoring and optimization purposes. A wind farm equipped with sensors on each turbine sends information regarding health status and output to its control center for analysis and management purposes.

Pyramid codes can be used to store this data efficiently. Should one of its nodes fail, however, its data can quickly be recreated from other nodes within a wind farm and remain operationally effective.

6. Analysis of Performance

Analysis of Repair Cost:

Access to six disks, including five data disks and one parity disk, is necessary for the (16, 12) Pyramid code if there is a single disk failure. The repair rate is 100%. Ten data disks and two parity disks are required for the Pyramid code to function properly. The repair rate is 100%. To achieve triple disk failure, the Pyramid code needs ten data disks and three parity disks. The repair rate is 100%. To prevent quadruple disk failure, the Pyramid code needs ten data disks and four parity disks. The repair rate is around 86.15%.

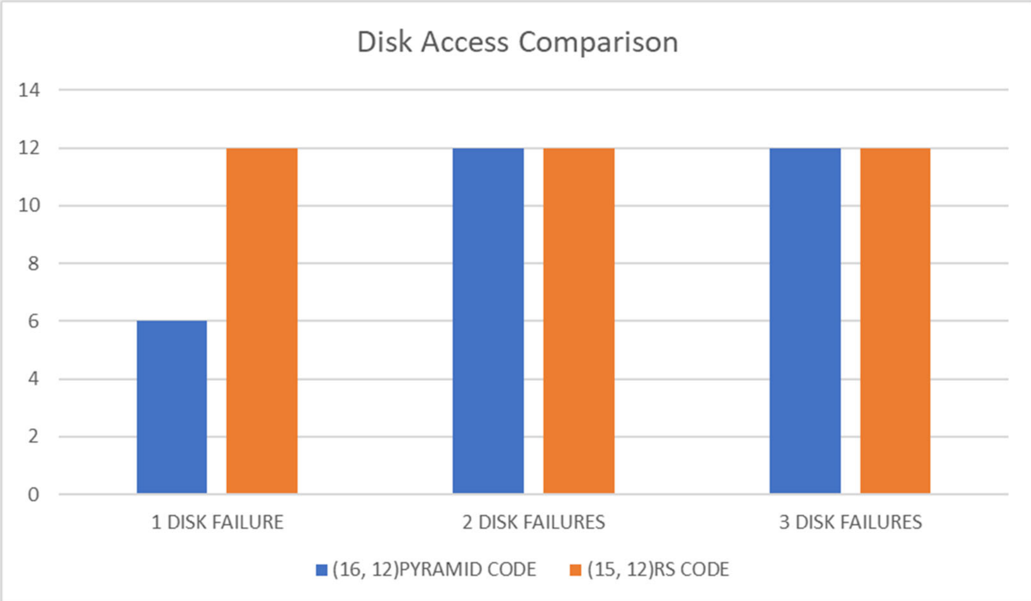

Figure 3. Disk Access Comparison between Pyramid code and RS code

In Figure 3, double and triple diss failure are costed same for both Pyramid code and RS code. However, for a single disk failure, the cost of Pyramid code is half of RS code and only access to 6 data disks can fix 1 disk failure. Also, the RS code cannot correct 4 disk failures happening at the same time, but the Pyramid code can correct around 86.15% of cases of 4 disk failures, which can be boosted to 99% by adding more parities or changing the different combination of parity 3 and parity 4.

7. Conclusion

In this paper, we discuss what are the Pyramid codes and why the IoT needs them. The Pyramid code model (16, 12) can correct all three error cases and almost all four error situations, as demonstrated in Section 3. The comprehensive Pyramid code in section 4 is capable of meeting the needs of the youth system, particularly when it comes to security and reliability. Section 5 illustrates how the Pyramid Code can be applied to various scenarios. Comparing the RS code and Pyramid codes is what we do in section 5. As a result, the Pyramid codes are more efficient in one disk failure case. Additionally, the pyramid can handle most of the four error cases with various combinations of parity, but the RS code in the same dimension cannot accomplish that. In the future, we expect more applications to be used in different systems, except for IoT systems.

References

[1] C. MacGillivray and D. Reinsel, "Worldwide global datasphere IoT device and data forecast 2019 – 2023", May 2019.

[2] J. A. Stankovic, "Research directions for the Internet of Things", IEEE Internet Things J., vol. 1, no. 1, pp. 3 - 9, Feb. 2014.

[3] H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, "IoT-Based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges," in IEEE Internet of Things Journal, vol. 4, no. 1, pp. 75 - 87, Feb. 2017, doi: 10.1109/JIOT. 2016. 2619369.

[4] F K Wang, "Application of big data and cloud computing in Internet of things [J]", Information & computer (theory edition), vol. 34, no. 08, pp. 183 - 185, 2022.

[5] Balaji S B, Krishnan M N, Vajha M, et al. Erasure coding for distributed storage: an overview. Sci China Inf Sci, 2018, 61 (10): 100301, https://doi.org/10.1007/s11432 - 018 - 9482 – 6.

[6] Cheng, H., Huseyin, S., Yikang, X., Aaron, O., Brad, C., Parikshit, G., Jin, L., & Sergey, Y. (2012). Erasure Coding in Windows Azure Storage. 2012 USENIX Federated Conferences Week. Microsoft Corporation.

[7] Stephen B. Wicker, Vijay K. Bhargava, Reed-Solomon Codes and Their Applications, Wiley-IEEE Press, 1994.

[8] Zhou Song and Wang Yijie, "EXPyramid: A Flexible Coding Scheme with High Fault Tolerance and Low Repair Cost Based on Array Structure [J]", Computer research and development, vol. 48, no. S1, pp. 30 - 36, 2011.

[9] H. Garg and M. Dave, "Securing IoT Devices and Securely Connecting the Dots Using REST API and Middleware," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-6, doi: 10.1109/IoT-SIU. 2019. 8777334.

[10] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and other botnets", Computer, vol. 50, no. 7, pp. 80 - 84, 2017.