An adaptive selection method for repair nodes in distributed storage systems

Rong Chang¹*, Chuanxu Yang¹, Yuankang Lei¹
¹Yunnan Power Grid Co., Ltd, Yuxi Power Supply Bureau

*Corresponding author e-mail: changrong@yunnanpg.org

Abstract. In distributed storage system, how to optimize the repair time of failure data to ensure the high reliability of the system has aroused wide attention. In recent years, studies have found that different node selection mechanisms in the repair process have a great impact on the data regeneration time, and previous work has proposed the node selection under the single node failure scenario. In the case that multiple nodes are repaired at the same time in the system, the huge space-time overhead of SPSN algorithm makes the data regeneration time no longer optimal. Make statistics on failure data and causes of existing real systems; in order to solve the replication strategy in the storage overhead and the erasure code strategy in the repair of bandwidth overhead, network coding technology, called regenerative code, is introduced into distributed storage system to balance the storage cost and bandwidth cost. In this paper, single-node repair algorithm and multi-node collaboration algorithm are mainly studied.

Keywords: Distributed Storage System, Node Repair, Adaptive Selection Method

1. Preface

With the rapid development of information technology and related industries, information processing and network services have gradually penetrated into all aspects of social life, generating a large amount of data all the time. According to reliable data, the amount of data generated in a network environment every 18 months is approximately equal to the total amount of data in the past. It is expected that by 2020, the accumulated data volume generated in the global network environment will be 50 times the current information volume, indicating that data sea quantification has become a trend. When node failure occurs in the system, to ensure continuous reliability, the system needs to restore the data on the failed node to the new node. The living node that provides data download at the time of repair is called the provisioning node, and the new node that stores regenerative data at the time of repair is called the newborn node. The traditional network storage system usually adopts the centralized storage mode, storing all the data in the central storage server, which has low security, reliability and scalability, and is not suitable for the storage of mass data [1].

In a situation where the storage system is under great pressure and challenge, distributed storage has attracted wide attention due to its advantages such as low cost and easy extension. At the same time,
based on the network technology, it separates the data processing and data storage in the server system to realize the mass storage of data.

This paper mainly studies the method of node repair of distributed storage system based on regenerative code, including single-node repair algorithm and multi-node cooperation algorithm.\[2\]

2. Information flow diagram
Along with the social informationization and the information mass emergence, as well as the people to the information request surge. The flow of information forms an intricate and ever-changing pattern. This movement can occur between people, between people and institutions, within institutions, and between institutions, including tangible flows and intangible flows. Before we get into the two fix nodes, let's take a look at the information flow diagram. The information flow graph is a directed acyclic graph, which can be used to analyze node repair problems in distributed storage systems. As shown in the figure, the information flow diagram contains three kinds of nodes, source node S, storage node X and data collection node D. In the information flow diagram, there is only one source node, representing the original file. Each storage node X is jointly represented by node Xin and node Xout, where node Xin represents the storage input node and node Xout represents the storage output node. There is a directed edge between node Xin and node Xout, and from node Xin points to node Xout. The capacity of this edge represents the storage space of the node. The data receive node D represents the user connected to a node in the system to restore the entire original file.\[3\]

![Diagram of Information flow diagram](image)

3. A single node repair algorithm for distributed storage system based on MSR code
In the traditional node repair algorithm of distributed storage system, based on the replicated node repair method, the new node only needs to connect any node with the same data stored by the failed node and download its data to complete the repair process. Therefore, the repair process only needs to download the same amount of data as the node storage, without additional bandwidth consumption, but in order to maintain the reliability of the system, the replication strategy requires a large number of backup nodes to store the backup data of the original files, which brings a lot of storage consumption to the system. In the case of node failure in distributed storage system, single node failure is more common, so it is very necessary to repair the single node problem in distributed storage system.\[4\]
4. Multi-node cooperative repair algorithm based on MBR code in distributed storage system

The storage mode of the file in the system is (n, k, d). The file is divided into K data blocks and placed into N. When a node in a distributed storage system fails, the system needs to download information from d surviving nodes to prepare to recover the data on the failed node. Assuming that the file is stored in (4, 2, 2) MDS code, that is, the file is encoded into blocks and placed on four nodes, and the network topology with the redundancy of the file is 2. The provisioning node only transfers data to the newborn node during data repair, the newborn node not only downloads data from the provisioning node, but also needs to transmit data to other newborn nodes. Therefore, the topology network does not show the transmission bandwidth between the newborn node and the provisioning node and between provisioning nodes. \[6\]

Aiming at the problem of multi-node repair in distributed storage system, this paper presents an improved coding scheme of multi-node cooperative repair based on MBR code, starting with the minimum bandwidth cost theory of multi-node cooperative repairers of pure electric vehicles. \[6\]

Figure 2. Single node repair diagram

Figure 3. Multi-node cooperative repair diagram
5. Conclusion
In order to optimize the repair time, relevant studies verify the influence of node selection mechanism on the repair time in single node failure scenarios. However, the existing algorithms are not suitable for multi-node failure scenarios, and their serializability makes the data repair performance no longer optimal. Research on node repair problem of distributed storage system based on Network coding, at present, most of them are still at the theoretical level, and there are still many practical problems to be considered. In view of the research content of this paper, there are some contents worthy of further research:

(1) Single node repair algorithm based on MSR code in distributed storage system
(2) Multi-node cooperative repair algorithm based on MBR code in distributed storage system

References
[1] Cadambe V R, Huang C, Li J. Permutation code: Optimal exact-repair of a single failed node in MDS code based distributed storage systems[C]// IEEE International Symposium on Information Theory Proceedings. IEEE, 2011:1225-1229.
[2] Zhang Z, Lian Q, Lin S, et al. BitVault: a highly reliable distributed data retention platform[J]. Acm Sigops Operating Systems Review, 2007, 41(2):27-36.
[3] Dimakis A G, Godfrey P G, Wainwright M J, et al. Network coding for peer - to - peer storage[C]// 2007.
[4] Dimakis A G, Godfrey P B, Wu Y, et al. Network Coding for Distributed Storage Systems[C]// IEEE Trans.inf.theory, 2010, 56(9):4539-4551.
[5] Xin Q, Miller E L, Schwarz T, et al. Reliability mechanisms for very large storage systems[C]// 20 Th Ieee/11 Th NASA Goddard Conference on MASS Storage Systems and Technologies. IEEE Computer Society, 2003:146.
[6] Weatherspoon H, Kubiatowicz J D. Erasure Coding Vs. Replication: A Quantitative Comparison[C]// Revised Papers from the First International Workshop on Peer-to-Peer Systems. Springer-Verlag, 2002:328-338.
[7] Wu Y, Dimakis A, Ramchandran K. Deterministic regenerating codes for distributed storage[C]// Allerton Conference on Control, Computing, and Communication. 2007.