Scholar Publishing Group
Distributed Processing System
https://doi.org/10.38007/DPS.2020.010406
ISSN 2790-0916 Vol. 1, Issue 4: 41-48

Data Transmission Security Module of Distributed System Based on Network Coding

Hamdollah Ravand[*]

Univ Zagreb, Fac Elect Engn & Comp, Zagreb, Croatia

[*]*corresponding author*

Keywords: Network Coding, Distributed System, Data Transmission, Security Module

Abstract: Distributed storage systems often use replica mechanism to achieve redundancy, which often leads to large storage overhead and repair bandwidth overhead. In order to improve the overall performance of the system, network coding is introduced into the distributed storage system. This paper mainly studies the design of data transmission security module of distributed system based on network coding. This paper first analyzes the principle of network coding, and compares the transmission modes between butterfly network model and routing. Aiming at the problem of large amount of data encryption in distributed file system, this paper studies the encryption mechanism of HDFS system based on network coding, proposes a lightweight encryption mechanism, and carries out simulation experiments.

1. Introduction

Distributed storage systems (DSS) is to store data in multiple devices dispersedly. Because of its advantages such as high reliability, high availability and high access efficiency, it has become a mainstream solution for large-scale data storage in the industry [1]. At present, many cloud storage providers such as Amazon network service (AWS cloud service), Dropbox, Google drive and Microsoft onedrive have provided users with attractive storage services. Although DSS can enhance the reliability of data by storing data in different geographical locations, it will lead to more attack targets and increase the risk of personal sensitive data being eavesdropped [2]. The distributed storage mode brings the advantage of data reliability and also increases the risk of privacy infringement, which may lead to the leakage of personal sensitive data [3]. Qualitative analysis is one of the key problems of distributed storage. Such system applications include large data centers and P2P file storage systems such as OceanStore, total recall and dhash + +. A DSS must be able to repair faulty nodes to maintain system stability. At present, there are three main data repair technologies. Compared with the other two data repair technologies, i.e. replication and erasure correction code technology, network coding is applied to DSS, which is called regenerative code

(RC) technology. It has attracted wide attention because of its higher stability and smaller repair bandwidth. Research shows that there is a tradeoff between the node storage capacity and the repair bandwidth, and RC can obtain the points on this tradeoff curve [4-5].

In recent years, the anti eavesdropping research of RC technology used in DSS has received extensive attention at home and abroad. This is related to the security network coding in the recent literature, which is relatively difficult [6]. For example, a scholar's research shows that finding the security capacity of the network with eavesdropping nodes is an n-hard problem [7]. However, compared with the general network, the network representing DSS has structural symmetry, and the security problem becomes easy to handle in this network. For different node and link eavesdropping problems, scholars have derived the corresponding upper bound of system security capacity, and the RCS design that can reach the upper bound of security capacity and make the system obtain perfect security has been widely studied [82]. If there is eavesdropping, the system does not leak any information about the stored files to the eavesdroppers, saying that such a system has perfect security. In practice, perfect safety is usually not necessary, and different applications may have different safety level requirements [9]. For different node and link eavesdropping problems, scholars have derived the corresponding upper bound of system security capacity, and the corresponding security coding design has been widely studied [10]. However, there are some limitations in the research of data transmission security based on RC.

This paper mainly studies the research of distributed data transmission security protocol based on network coding, which extends from perfect security requirements to general scenarios with different security levels, and extends from the traditional model with strict data reconstruction and data repair constraints to the generalized model with huge potential performance gain, thus expanding its practical value.

2. Security of Distributed Data Transmission Based on Network Coding

2.1. Network Coding Principle

Network coding has broken through the boundaries of traditional information transmission and fundamentally changed the way of signal processing and transmission. It is a new hotspot of information theory research and has attracted the high attention and extensive research of scholars in the corresponding fields. In the past decade, the research field of network coding theory and its application has become more and more extensive, which has prompted researchers to use new mathematical tools (mainly algebra, matroid theory, graph theory, geometry, combinatorial mathematics and optimization theory) to study current network coding: extensive and complex mathematical models [11-12]. In addition, network coding is closely related to information theory and is dominated by information theory.

The traditional communication network generally adopts the store and forward mechanism. In the network, except for the source node and the sink node, the other nodes are only responsible for data routing and do not process the data [13]. While researchers integrate the network coding strategy into the routing process, the nodes perform coding operations on the information received in multiple channels and forward it to the downstream nodes. The information flow encoded by the network can make the network transmission reach the maximum flow limit of the network and improve the throughput and robustness of the network transmission [14].

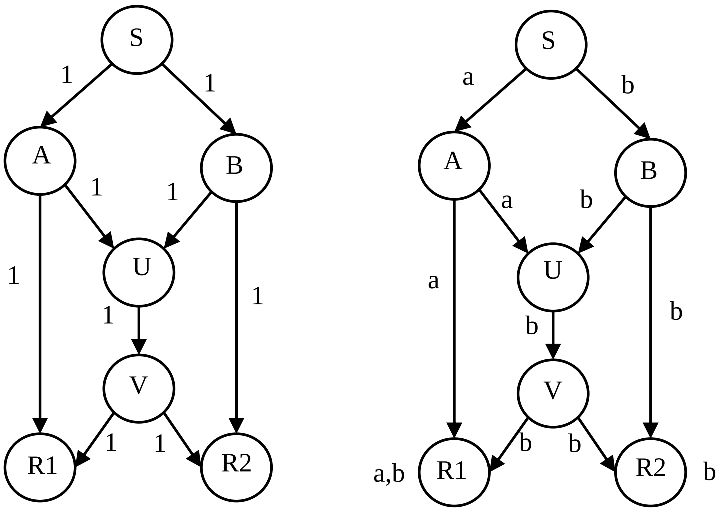

Figure 1. Comparison of network coding and routing in butterfly network

For a directed acyclic network, as shown in Fig. 1 (a), assuming that the link has a unit capacity and no delay, considering that the new sink nodes R1 and R2 can also receive the messages a and B (2 bits in total) sent by the source node s at the same time.

If the routing method is adopted, since the link UV is the transmission bottleneck of the network, only Lbit information (a or b) can be transmitted. As shown in Fig. 1 (b), the link UV only transmits message B, then the aggregation node can just receive message B, and I can only receive message B. the average throughput of each aggregation node is (2 + 1) / 2 = 1.5bit. Similarly, in Fig. 2 (a), the link UV can only transmit message a, then the sink node R1 can only receive message a, and I can also receive messages a and B. The average throughput of each aggregation node is still 1.5 bits.

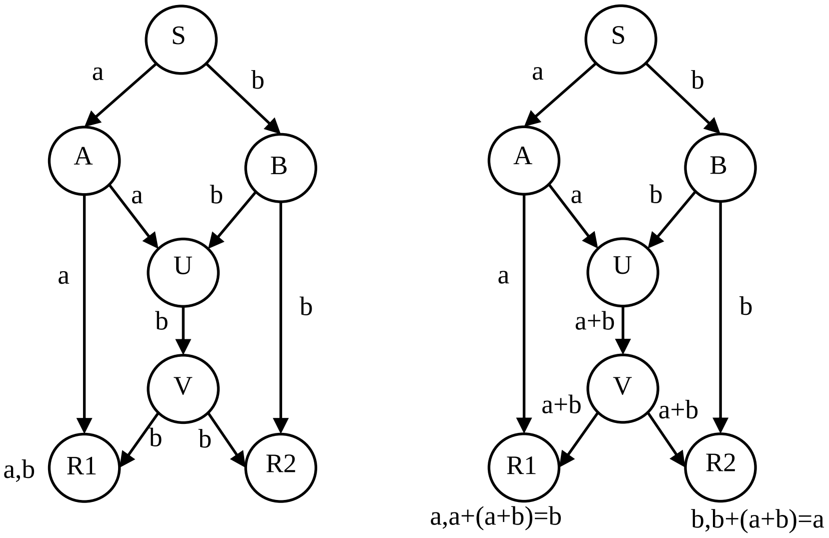

Figure 2. Comparison of network coding and routing

If the network coding method is adopted, as shown in Fig. 2 (b), the messages a and B are encoded on the node u, and then a + B are calculated and transmitted to the gang and R2. Finally, the messages a and a + B are received at the new node RL, and then the network coding is performed on the RL, where a + (a + b) = B, B is decoded, and the message hole B is finally received at the same time. For the aggregation node L, the same operation is performed, B + (a + b) = a, decoding a, and finally receiving messages such as B. thus, the average throughput of each sink node is (2 + 2) / 2 = 2bit, which is greater than 1.5bit of the routing mode, thus improving the network throughput [15].

It can be seen that the throughput achieved by network coding can be greater than that achieved by routing.

According to the above examples, the essence of network coding is that the intermediate node or the source node can encode the received information or the information transmitted by the upper layer application and then forward it [16]. If the operation is regarded as a function, the information received by the intermediate node is an independent variable, the result of the operation is a function value, and the coding rule is a function operation. The sink node decodes the received information and then recovers the information generated by the source point. Therefore, it must ensure the closeness of the operation, that is, select a set and operate the elements in the set. The budget result is still the elements in the set [17-18].

2.2. Data Transmission Encryption of Distributed Storage System

The system model proposed in this paper is that the original data will be stored in the distributed storage system after network coding. In distributed storage system, both coding matrix and coding data need to be stored. In different systems, the locations of matrix and data storage are different. In some distributed storage systems, the coding matrix and the coding data are stored in the same data node, that is, they are all distributed and stored in different nodes in the system; There are also some distributed storage systems, where the coding matrix and the coding data are stored separately, that is, the coding matrix is stored in a single location and can be called at any time during use. The storage system used in this paper is the latter, and the coding matrix will be stored in a location different from the coded data.

The attack model in this paper is: suppose there is a symmetric encryption algorithm enc with relatively secure semantics, in which encryption and decryption are respectively set as Enck () and dnck (), K represents the secret key, and the number of secret key bits is r. The content discussed in this paper is that it is assumed that the data transmission links in the network are secure, that is to say, the probability of an attacker obtaining information through the transmission links is basically zero. In this paper, it is set that the attacker's attack ability to the distributed system is limited, and he can only steal limited data information and read data. At the same time, it is assumed that the attacker will not tamper with or delete the data in the distributed storage system.

The traditional distributed storage system based on network coding has some defects. Taking HDFS as the target system, this paper proposes a lightweight encryption scheme. Network coding is applied in the HDFS system. The system needs to store both coding data and coding matrix.

This method is mainly used to protect the original data by encrypting the coding matrix. The two methods of encrypting the encoding matrix and encrypting the encoded data are different in the position where the encryption operation is performed. In the HDFS system based on network coding, the encryption operation is performed in the namenode node. After the namenode returns the encoding matrix to the encoding queue for encoding or waits for encoding, the encoding matrix is encrypted and then stored. Compared with directly encrypting coded data, one of the advantages of the encryption coding matrix is to reduce the amount of encrypted data.

Compared with the traditional encryption method, the encryption method of encryption coding matrix not only reduces the amount of encrypted data, but also has better support for the amount of calculation during data operation. As mentioned above, the way of encrypting the original data is easily affected by the diffusion of the encryption algorithm. Any slight change in the plaintext will lead to a large number of changes in the ciphertext. From this point of view, the method of encrypting the encoding vector in this paper will not have this defect, because this method does not directly encrypt the plaintext, so it can avoid large-scale changes in the ciphertext due to changes in the plaintext. In the way of encrypting encoded data, modifying the data of each data node requires the operation of encoding and decoding the encoded data, which leads to a large number of calculations and reduces the performance of the system. The method of encrypting the encoding vector proposed in this paper does not encrypt the encoded data, so the change of the data of the data node will not lead to any change of the encoded data.

The whole process of applying the encryption coding matrix in the HDFS system based on network coding is composed as follows:

$$\vec{K} = (K_1, K_2, \ldots K_N) \tag{1}$$

(1) Where is the secret key vector, and n secret keys are selected in the secret key space to form the secret key vector.

$$(Y, G) = Encode(M, k, n) \tag{2}$$

(2) Where is the encoding process. The original data M is divided into k pieces of data blocks, and the N pieces of encoded data generated by the encoding operation are expressed as y. The coding vector is expressed as G, and the relationship with y is: y = m * G.

$$g_i^{'} = ENC_{Ki}(gi) \tag{3}$$

(3) Where is the encryption algorithm. Select the corresponding secret key and encoding vector for operation.

3. Algorithm Simulation Experiment

In this paper, the optimization effect of data transmission of encryption transmission algorithm is experimentally analyzed. Through repeated experiments in different scale networks and different input symbol numbers, the data collection efficiency and data transmission performance of the protocol are evaluated and compared.

In this paper, several experiments are carried out under the network configuration of n = 128, n = 256, f = 15 and F = 25 respectively to evaluate the transmission efficiency of the codeword transmission process under the encryption transmission algorithm.

4. Analysis of Experimental Results

4.1.N = 128 Data Collection Results

5. As shown in Table 1 and figure 3, the average data collection efficiency of the algorithm in this paper under different network configurations is better than that of the traditional algorithm,

both at the initial stage and at the end of collection. With the expansion of network scale and data volume, there will be greater performance improvement.

Table 1. Data symbol collection process under F = 15

	0	100	200	400
Traditional algorithm	0	0.42	0.83	0.99
Algorithm in this paper	0	0.56	1	

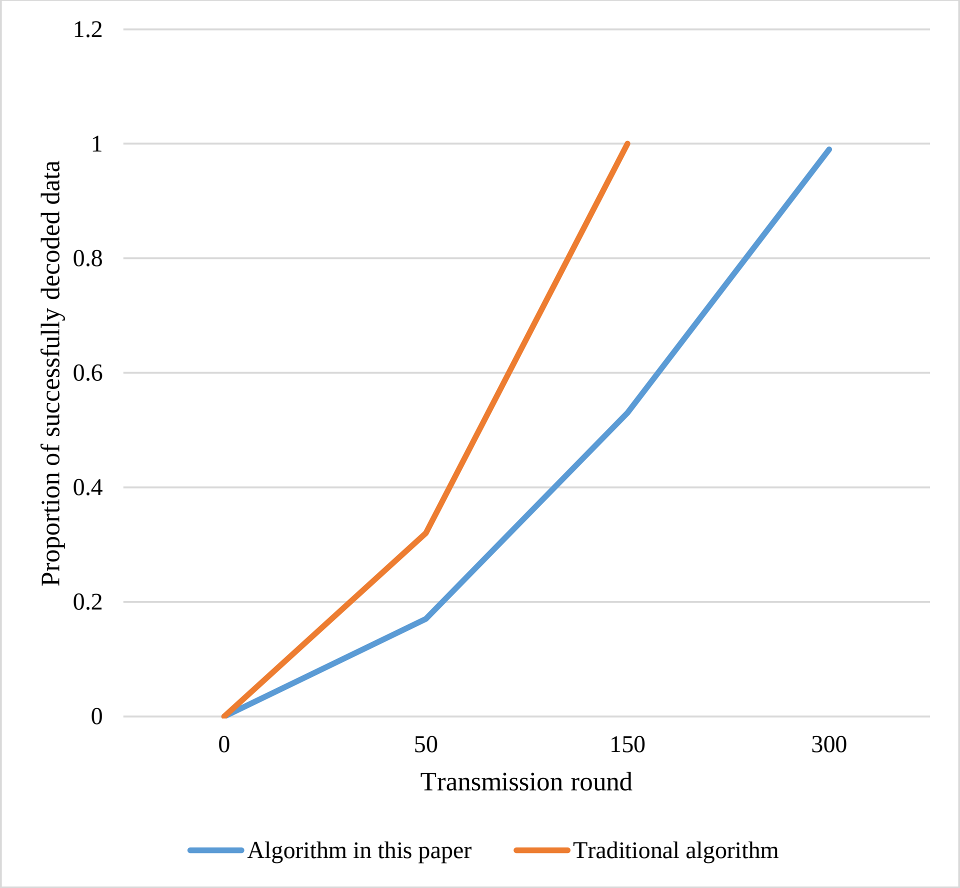

Figure 3. Data symbol collection process under f = 25

4.2. N = 256 Data Collection Results

As shown in Fig. 4, with the expansion of the network scale, the transmission efficiency gap between the algorithm in this paper and the traditional algorithm is also gradually widening. Especially in the slightly larger scale of N = 256 and F = 25, the gap between the two is more than 50%, which indicates that the algorithm in this paper has more advantages in large-scale networks. In large-scale networks, the symbol degree of freedom sequence of nodes plays a role in adjusting the data distribution, and the feedback information brings strong data orientation.

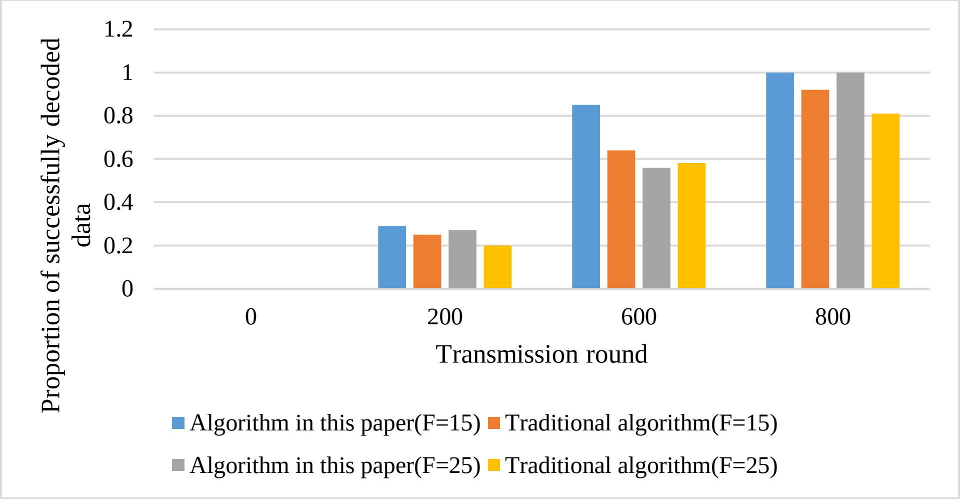

Figure 4. Data symbol collection process when the node is n = 256

5.Conclusion

Based on the open source distributed file storage system HDFS, this paper designs a Hadoop distributed file system based on network coding. A local repair code based on the system MSR code is introduced into the HDFS system. Based on the HDFS system based on network coding, a lightweight encryption method is proposed, and the HDFS system based on network coding is introduced. There are many places that can be further studied in the paper. The main research contents in the future include the following contents: the encryption method of the encryption coding matrix proposed in this paper is lightweight only when the parameters meet the conditions. In future research, more attention should be paid to the use of parameters to reduce the influence factors of parameters.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Amita N A, Kaur A, Kumar M. Reversible data hiding in absolute moment block truncation coding compressed images using adaptive multilevel histogram shifting technique.

International Journal of Information and Computer Security, 2018, 10(2/3):261.

[2] Rawal B S, Liang S, Gautam S, et al. Nth Order Binary Encoding with Split-protocol. International Journal of Rough Sets & Data Analysis, 2018, 5(2):95-118.

[3] Brahimi M A, Merazka F, Kurt G K. Secure network coding for data encoded using subspace codes. Physical Communication, 2020, 48(2):101408.

[4] Al-Aazzeh J, Ayyoub B, Faure E, et al. Telecommunication Systems with Multiple Access Based on Data Factorial Coding. International Journal on Communications Antenna and Propagation, 2020, 10(20):12.

[5] Chervyakov N I, Deryabin M A, Nazarov A S, et al. Secure and Reliable Data Transmission Over MANET Based On Principles of Computationally Secure Secret Sharing. Proceedings of the Institute for System Programming of RAS, 2019, 31(2):153-170.

[6] Bwalya M, Chembe C. A Security Framework for Mobile Application Systems: Case of Android Applications. Zambia ICT Journal, 2020, 3(2):31-43.

[7] Usman M A, Usman M R, Shin S Y. Quality assessment for wireless capsule endoscopy videos compressed via HEVC: From diagnostic quality to visual perception. Computers in Biology and Medicine, 2017, 91(4):112-134.

[8] Aljohani A J, Ng S X. Distributed Joint Source-Channel Coding-Based Adaptive Dynamic Network Coding. IEEE Access, 2020, PP(99):1-1.

[9] [1]Jonnahtan, Saltarin, Eirina, et al. Adaptive Video Streaming With Network Coding Enabled Named Data Networking. IEEE Transactions on Multimedia, 2017, 19(10):2182-2196.

[10] Marques B, Coelho I M, Sena A, et al. A network coding protocol for wireless sensor fog computing. International Journal of Grid and Utility Computing, 2019, 10(3):224.

[11] Naeem A, Rehmani M H, Saleem Y, et al. Network Coding in Cognitive Radio Networks: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 2017, 19(3):1945-1973.

[12] Vahid A, Lin S C, Wang I H. Erasure Broadcast Channels with Intermittent Feedback. IEEE Transactions on Communications, 2020, PP(99):1-1.

[13] Alhayani B, Abdallah A A. Manufacturing intelligent Corvus corone module for a secured two way image transmission under WSN. Engineering Computations, 2020, ahead-of-print(ahead-of-print).

[14] L Farzinvash∗. Online network coding-based multicast routing in multichannel multiradio wireless mesh networks. Turkish Journal of Electrical Engineering and Computer Sciences, 2019, 27(2):1387-1405.

[15] Pham H, Dang N T. Analog network coding aided multiuser visible light communication networks using optical CDMA. OSA Continuum, 2019, 2(9):2569.

[16] Spini G, Zemor G. Efficient Protocols for Perfectly Secure Message Transmission with Applications to Secure Network Coding. IEEE Transactions on Information Theory, 2020, PP(99):1-1.

[17] Nam W, Lee J, Shroff N B, et al. An Inter-Data Encoding Technique that Exploits Synchronized Data for Network Applications. IEEE Transactions on Mobile Computing, 2019, PP(99):1-1.

[18] Giambene G, Luong D, Cola T D, et al. Analysis of a Packet-Level Block Coding Approach for Terrestrial-Satellite Mobile Systems. IEEE Transactions on Vehicular Technology, 2019, 68(8):8117-8132.