Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 02, PP. 103-114, 2025
DOI: https://doi.org/10.54216/JISIoT.140209

Research on Big Data efficient hybrid cloud storage model and algorithm based on 5G network

Lei Hu[1, 3,*], Yangxia Shu[2, 3]

[1]Operation and Maintenance Section of Assets Department, Jiangxi Institute of Fashion Technology, Nanchang 330201, Jiangxi, China

[2]College of Big Date Science, Jiangxi Institute of Fashion Technology, Nanchang 330201, Jiangxi, China

[3]Information Technology Integration Innovation Center, Intelligent Research and Innovation Team for Clothing (Jiangxi Institute of Fashion Technology), Nanchang 330201, Jiangxi, China

E-mail: leihulh@outlook.com; Yangxia. Shu@163.com

Abstract

Due to the large demand for big data storage capacity, the storage intensity index is not calculated in the current big data cloud storage process, resulting in a high storage space usage. This paper proposes a big data efficient hybrid cloud storage model and algorithm under 5G network. The model is based on the 5G network performance framework and consists of three parts: users, private cloud and public cloud storage service providers. The purpose of efficient hybrid cloud storage of big data is achieved by using consistent hash algorithm. The simulation results show that the above algorithm occupies less storage memory space, the device load variance is smaller, the overall system load is more stable and balanced, and the average response is fast, which provides a favorable basis for the efficient hybrid cloud storage algorithm of big data.

Received: March 14, 2024 Revised: June 19, 2024 Accepted: October 12, 2024

Key words: 5G network; Big data; Hybrid cloud storage; Cloud storage algorithm

1. Introduction

In the context of 5G, the amount of data surges, and cloud storage cannot meet its data storage requirements. This problem has also become a hot topic in related research fields. With the maturity of the Internet of Things technology, more and more Internet of Things data are available. In order to ensure the integrity of the database content, it can be stored on the basis of hybrid cloud [1]. Cloud computing is an emerging computing method based on Internet technology, which can ensure the storage efficiency of data while making efficient calculation and provide convenience for users [2]. A big data efficient hybrid cloud storage model under 5G network is proposed and researched, which consists of user entity, private cloud entity and public cloud storage service provider entity. It has good encryption effect and can effectively control storage space and management cost [3]. Achieve efficient hybrid cloud storage through the relationship between the elasticity, granularity, and utilization of the obtained big data. Experiments show that this method can effectively improve storage security and reduce space ratio, and is a reliable and scientific big data cloud storage method [4-5].

2. Working principles of the CSS

Compared with traditional storage technology, cloud storage technology is no longer directly stored on a specific device, and the access mode is more flexible. Cloud storage technology can integrate a large number of different types of devices through networks and specific system structures, and follow the rules of hardware virtualization, access networking, and interface standardization. The storage resources scattered on each physical node are integrated into a unified storage system at the logical level [6].

The underlying working mechanism of cloud Storage is still based on traditional physical storage technology, that is, the so-called physical storage layer of cloud storage system, which is the basis of cloud storage. The physical storage medium of traditional devices is mainly hard disk and similar Direct Attached Storage (DAS) [7]. Cloud storage devices are generally configured according to their designed capacity limits, which can achieve a higher cost performance, due to the improvement of today's hard disk technology, the capacity of a single device can also be large, can grow to several TB or even tens of TB magnitude, but due to the limitations of the network export of a single machine, if in a system with high throughput demand, Such bottleneck effect is revealed, but has an impact on the performance of the entire system. Therefore, in general, the capacity of a single device in a cloud storage system should be in a relatively high cost-effective range [8]. In addition, in the cloud storage system, due to the huge system and long construction cycle, the purchase of different batches of equipment, device performance will vary greatly, and heterogeneous system design is therefore necessary, which is a common situation in the physical storage layer of cloud storage. In some systems, a device management system is set up outside the physical storage layer to realize the monitoring and operation and maintenance of physical devices [9-10].

In addition to these traditional physical storage solutions, there are also SAN systems and NAS systems, among which NAS systems are the most similar to cloud storage technologies. However, in traditional physical storage technology, hardware structure and design are the core, and the software and architecture above are less involved. In the cloud storage system, hardware devices are virtualized by software due to the introduction of software. A storage system that isolates hardware particularities is an important difference between the traditional storage system and the cloud storage system. Generally, from the perspective of the technical working principle of cloud storage, it can be divided into storage layer, basic management layer, application interface layer, and access layer.

3. Features of the cloud storage system

Cloud storage is an Internet technology emerging after cloud computing. It connects a large number of physical storage devices that actually exist together through virtualization technology, and these devices seem to exist in a cloud. In order to realize storage services, they run cooperatively and finally achieve the purpose of providing storage services to users [11]. When the cloud computing system has a strong storage capacity due to the need to deal with massive data, it seems to have converted into a cloud storage system. Using the cloud storage system, users can access data on the cloud at any time and place, convenient and fast.

The cloud storage system can meet various requirements of network users. Based on the analysis of the existing cloud storage system, it can be concluded that the cloud storage system must have the following features:

a. Scalability and efficiency: In today's information age, information is growing explosively, and the amount of data on the network is increasing, which requires continuous addition of storage devices to the cloud. In addition, in the massive data, a petabyte file system access records may be as many as hundreds of millions of records, the consumption of storage space is very alarming. Therefore, the cloud storage system requires a highly scalable and efficient management mechanism for hardware and software devices to provide external services.

b. Data persistence: Whether it is data stored locally or in the cloud, due to hardware failure, user disoperation, hacker attacks and other unpredictable disasters, data may be damaged or even lost to varying degrees. Therefore, the cloud storage system must take some measures, such as redundant backup technology, to ensure that data can be stable and durable stored in the cloud.

c. Security model: Because information is always accompanied by security issues, especially cloud storage, which is a system for mass data storage and management, data security issues need to attract more attention. Therefore, it is very necessary to have a security model in the cloud storage platform to ensure that files can be legally accessed and operated [12-13].

d. Portability: Due to the wide variety of existing cloud storage systems and the lack of an authoritative standard, it is very difficult to efficiently transfer data from one CSP to another. Therefore, the cloud storage system should have the ability to realize data transfer between other storage systems at a low cost.

4. Big data efficient hybrid cloud storage architecture and model

4.1 Hybrid Cloud Storage System Architecture

A hybrid cloud storage system consists of private cloud storage, public cloud storage, and hybrid cloud storage systems. Private cloud storage mainly runs on enterprise networks. The hybrid cloud storage system also runs in the internal network of the enterprise, but it is responsible for data and information interaction with the public cloud storage service on the public network [14]. The hybrid cloud storage solution proposed in this paper provides users with access to the hybrid cloud storage system through the Web access interface. The hybrid cloud storage system initiates access requests to private cloud storage and public cloud storage on behalf of users, and users do not directly operate the cloud storage system. The specific architecture is shown in Figure 1 below.

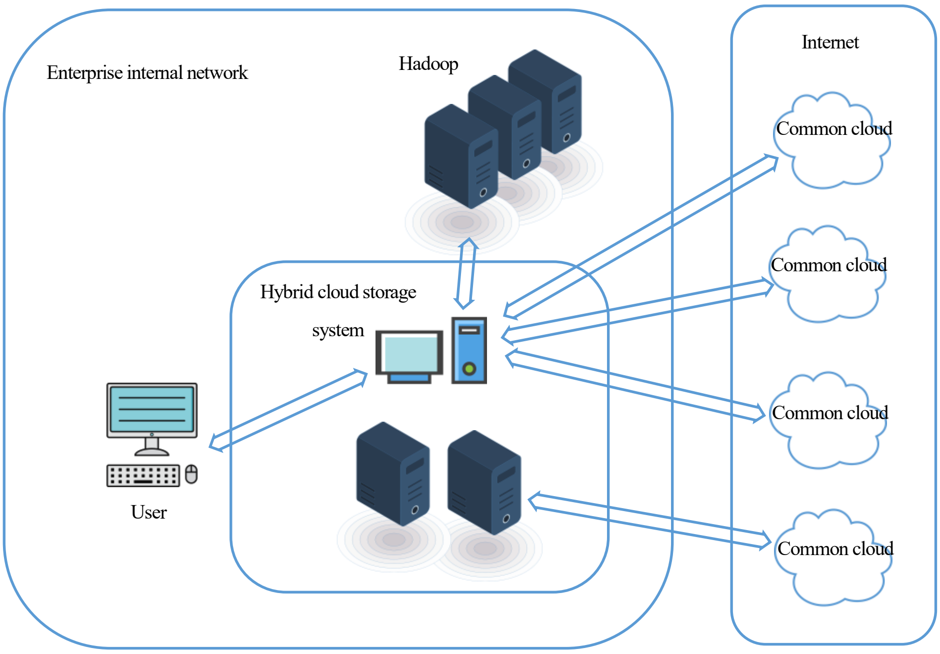

Figure 1. System architecture of hybrid cloud storage solution

The services of the hybrid cloud storage system and private cloud and public cloud storage constitute an organic whole. The hybrid cloud storage system is deployed within an enterprise and combines the enterprise's private cloud storage service with the public cloud storage service on the Internet.

Provides data storage services for enterprise users. The system itself provides the user interface through the Web service, which acts as the agent for uploading and downloading the user's data files. W server, database server and core service server jointly undertake the system services, and the system master service and other services will be deployed in the core service server. The failure of any server will cause a surge in the hybrid cloud storage system. Therefore, based on specific application scenarios and user scale, the system can be expanded through distributed, load balancing, and redundant backup technologies. Hadoop As a private cloud storage implementation, the distributed file system HDFS in Hadoop has high reliability and throughput, can ensure the performance and quality of service of the entire system, and can be deployed on a large number of inexpensive devices in the cluster, helping enterprises make full use of their storage and computing resources. The public cloud storage service of the system will choose the mainstream cloud service provider, which determines the elasticity and quality of the entire system service, and also plays a crucial role.

4.2 Hybrid cloud storage model

Traditional single-cloud storage means that user's hand over their data to a single cloud provider. In a single cloud storage system, user data is stored on multiple storage nodes. However, if the cloud storage service provider interrupts the service, user data may become unavailable for a short period of time or even for a long period of time, which may cause serious losses. In addition, due to provider lock-in issues, the cost of data migration is very expensive for users. The multi-cloud storage method can well avoid the problems of single-cloud storage, such as low data availability [15-16].

As Figure 2 shows, multi-cloud data hybrid storage is where data is distributed redundantly across multiple providers. The multi-cloud data hybrid storage service is aimed at a wide range of users. Individual users can store their data in the cloud, or they can use data and services stored in the cloud by other data owners through authorization, such as video on demand, audio on demand, etc. Enterprise users can use multi-cloud data storage to back up and archive enterprise data. Especially for some large websites, websites have to generate massive data every day, through the use of cloud storage service providers to provide space rental services, the massive data stored in the multi-cloud storage system, not only to ensure the security and stability of the data, but also to save their own storage equipment and data management costs. The source of the user's data can be a PC (personal computer) and portable computer connected to the network, or it can be a mobile phone. The biggest difference between data storage based on multi-cloud environment and traditional single cloud storage is that data redundancy in single cloud storage occurs within a single cloud storage service provider, that is, data is distributed redundantly on multiple storage nodes in the same cloud. The data storage based on multi-cloud environment distributes the data redundantly in the servers of two or more cloud storage service providers, which effectively avoids the Vendor lock-in problem and improves the availability of data. Even if a cloud storage service provider is interrupted or has other accidents, the data storage system will be more efficient. It still ensures that the user's data is available. In addition, the selection of appropriate data storage schemes, including appropriate data redundancy technology and appropriate cloud storage service providers, will also help improve data storage performance and efficiency and reduce data storage costs [17].

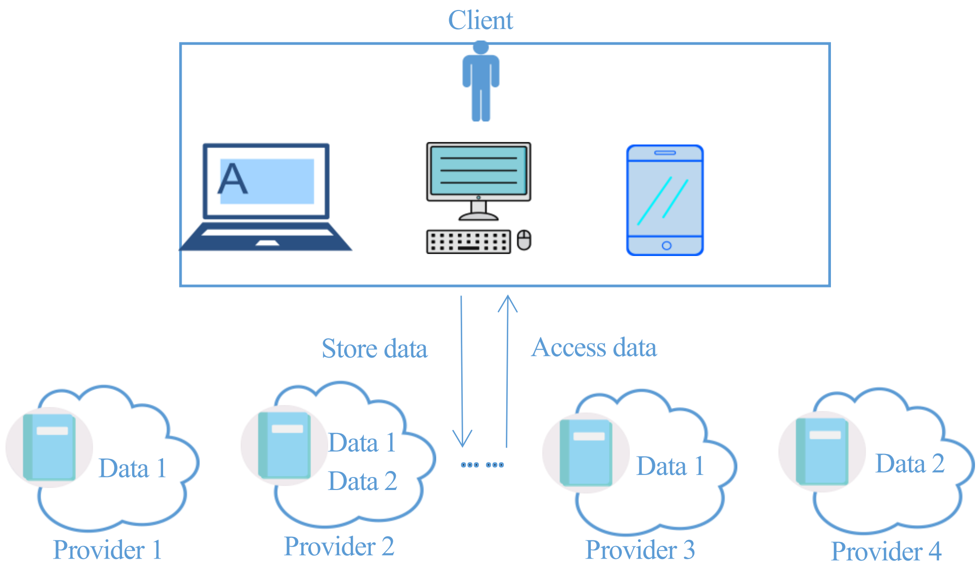

Figure 2. Multi-cloud data hybrid storage model

Hybrid cloud storage (including multi-cloud storage) is more complex and huger, it is composed of many components including network, cloud server, software, etc., and these parts logically constitute an organic whole, their respective division of labor, work together. The essence of cloud storage is actually to use virtual technology to achieve the goal of storage virtualization and even automation. Cloud providers provide configurable virtualization services and solutions for massive amounts of data through network technology.

In the evolution process of cloud storage, scholars have proposed a variety of structural models, as shown in Figure 3, which gives a typical hierarchical structure today. It consists of four layers from bottom to top: storage layer, basic management layer, application interface layer and access layer. The basic management layer is the most core part, and the lower layer provides services for the higher layer. Users can use the services provided by the cloud storage service provider directly through the access layer, without caring about the underlying composition and business logic.

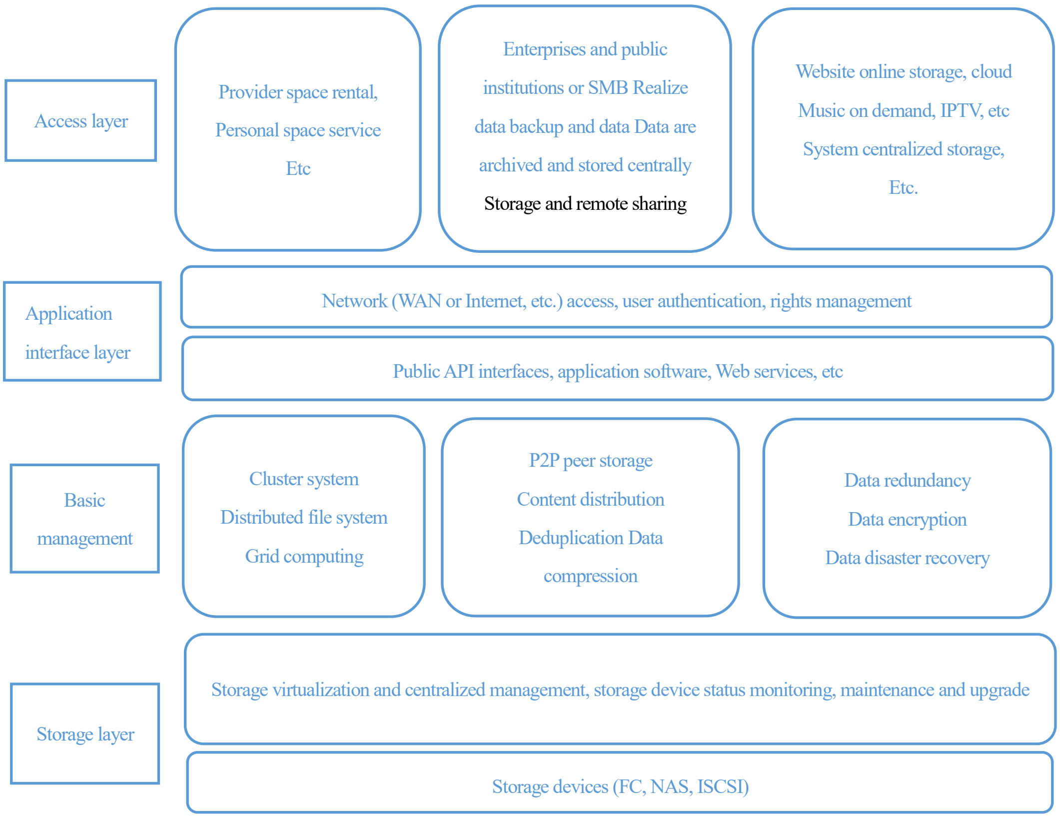

Figure 3. Hybrid cloud storage structure model

a. Storage layer the lowest storage layer is the most basic. There are various types of devices in the storage tier, such as Fiber channel, IP storage devices (such as NAS), and DAS storage devices (such as ISCSI). These devices are diverse and large in number, and they are often dispersed in different regions and connected through various networks, and then their status is monitored by management systems.

b. The basic management layer is the soul of the cloud storage system. It is responsible for the scheduling and decision-making of the entire cloud storage system. This layer determines the matching and scheduling of resources and data such as devices. It organizes devices in the entire system in a certain organizational form, slices or blocks data, or distributes it to corresponding devices in the form of objects according to certain strategies, and can realize data positioning, redundancy backup, load balancing, error recovery, etc., so as to achieve horizontal expansion of single machines. So that the system can be linearly expanded from the capacity, on the other hand, linear expansion is not only the capacity, in the corresponding read and write operation performance can also be linearly expanded on the whole, at the same time, through redundant coding or redundant copy technology, further improve the data failure protection ability, and can increase the load capacity of single file single data. In the actual system, this part can enhance the level, manage and schedule at the upper part of the implementation interface, and further improve the management and scheduling can be involved in the access side, and run in the application layer as part of the service, which can realize the distributed cloud storage system more quickly.

c. The application interface layer is an interface for the service application layer. It shields the internal implementation details of the cloud storage system and provides a friendly and unified operation interface externally. The interface layer provided by many cloud storage systems is their proprietary API or software. Through this interface software, you can perform file data operations on the cloud storage system.

d. The access layer refers to the specific business requirements of end users, such as audio and video playback, monitoring data access, personal file access, etc. This layer is only used to connect to the logic of service requirements. The other side performs data operations through the access interface layer of the cloud storage system. This section is responsible for implementing a wide variety of storage-related business requirements.

5. Efficient hybrid cloud storage algorithm implementations

5.1 multi-selection consistent hashing algorithm

As for the description of the MaCH (Multi-choices Consistent Hash) algorithm of the CMM system, as shown in Figure 4, the algorithm runs multiple hash function operations, but in fact, the algorithm in this paper only carries out one SHA-256 hash operation. And use the same hash has value for different segments [18-19]. When data is added to the system, the Distributed Hash Table (DHT) is mapped according to the segment corresponding to the hash value of its name, and the device with the lightest load is selected among the corresponding devices in each range for distribution. During the operation of the system, when the load of a certain device exceeds the alarm value and may be overloaded, the CMM system will use the M1CBF algorithm to select the hottest part of the data, and then carry out secondary allocation by MaCH algorithm to find the new corresponding device according to different hash values, and then find the most suitable device to reallocate the past. As shown in Figure 4, this series of processes are distributed by each client and storage device, without any central node deployment, eliminating the centralized single point problem, and improving the dynamic balancing problem of the general consistent hash algorithm.

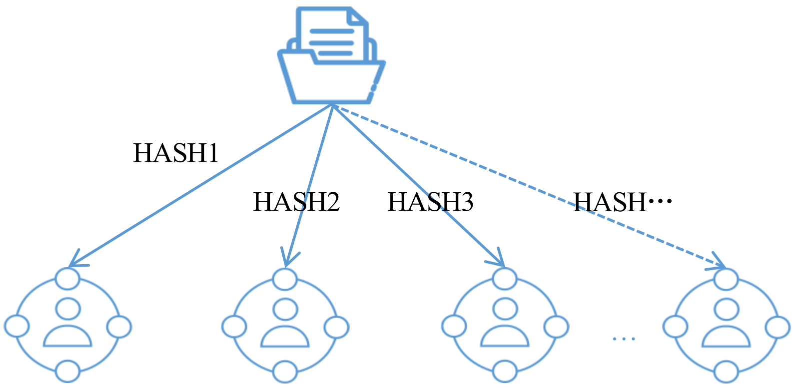

Figure 4. Schematic diagram of multi-choice consistent hashing algorithm

The CMM system scheme is also different from another class of non-central node design schemes, such as those that broadcast data relocation information. These broadcast schemes differ from the MaCH algorithm in that although they also use some Key-value database classes to improve performance, the way of broadcasting information still requires a lot of network and capacity resources. In the scheme of this paper, all operations are distributed, and there is no need to communicate in the way of broadcast. In the whole system, it is assumed that m' pieces of data in the system are relocated to the corresponding devices in the group, so the result is that the most heavily loaded devices have a high probability of having the load of $\frac{m}{n}+\frac{\log\log n}{\log|D|}+O(1)$, so you can see the imbalance of the load and the difference between the maximum load and the average load. In order to clearly describe the algorithm, a multi-dimensional mixed load is generally used to describe the comprehensive load situation, as described above, you can discuss the load of all resource combinations: CPU load, bandwidth, capacity, memory, etc. In fact, the most important thing to consider in the cloud storage system is the capacity and throughput efficiency of the device, so you can simplify the problem and only analyze these two main factors. So, the goal of the system should be to implement min ($\sum_{d\in D}(R_d - \bar{R}_D)$). Mixed resource utilization R_d can be expressed as:

$$R_d = \sum_{m'\in M'}(\frac{w_d^c c_d}{C_d} + \frac{w_d^b b_d p_{m'}}{B_d})V_d^{m'} \qquad (1)$$

Obviously, the overutilization of some resources limits the full utilization of other resources. Therefore, the weight of a resource is defined as being inversely proportional to the resource utilization of that resource dimension, thus:

$$w_d{}^c = \frac{\frac{1}{r_d^c}}{r}, \quad w_d^b = \frac{\frac{1}{r_d^b}}{r} \qquad (2)$$

Among $r = \frac{1}{r_d^b} + \frac{1}{r_d^c}$. r_d^c, r_d^b... It indicates the utilization rate of device d in terms of resources such as capacity and bandwidth, that is:

$$r_d^c = \frac{\sum_{m'\in M'} c_{m'} P_d^{m'} V_d^{m'}}{C_d} \qquad (3)$$

$$r_d^b = \frac{\sum_{m'\in M'} b_{m'} P_d^{m'} V_d^{m'}}{B_d} \qquad (4)$$

Even in the model of the already simplified design, the problem can be further simplified because of $2dm' \ll n$ and $m' \ll m$. Without further simplification, the computational complexity is still quite high, and most data optimization solutions are time-consuming, especially when large amounts of data need to be allocated and relocated to a large number of devices. Therefore, a simpler and more practical solution is needed to meet the real needs.

After experimental testing, the most practical scheme is the system to assign data to the device one by one, that is, m '=1. This allocation process includes both new data and relocated data. Therefore, when the target is transformed into the case of min $R_d^{m'=1}$, the problem faced by the algorithm becomes simple. In this case, the CMM system can use a simple heuristic algorithm to determine which specific device in the system should be selected, and because the execution efficiency of the heuristic algorithm is relatively high, a practical solution can be obtained. Clustering algorithms can be used for high-volume operations when a lot of data should be relocated or added, but the frequency of such operations is very low unless large-scale data migration is taking place. In the selection of data layout, both single data operation and multi-data group operation can use the distance algorithm to determine the choice of data layout location. Suppose A_d is the mixed weight of the capacity and bandwidth of the devices in each group, and B_m is the mixed weight of the data size and heat, and the distance measurement between them can be expressed as:

$$\mathrm{dist}_{dm'} = \sqrt{(A_d - B_{m'})^2} \qquad (5)$$

In fact, there are not many resource dimensions to consider and can be calculated separately to achieve more accurate results. Therefore, assuming that there are h resource types and the distance between the data and the device is directly calculated, the above indicators can be expressed as:

$$\mathrm{dist}_{dm'} = \sqrt{\sum_k (a_{dh} - b_{m'h})^2} \qquad (6)$$

The perfect situation of the system should be like this, in the relatively stable period of the system, the hot and small data is located on the equipment with small capacity and high throughput capacity, the large and low heat data is stored on the equipment with large capacity and weak throughput capacity, all the data is fully utilized in the order of the resource capacity of the equipment, and finally, the system can achieve the balance of resource utilization. There are no bottlenecks in the utilization of a single resource, and each resource dimension is fully utilized, while there are no bottlenecks in the central node and single points of failure.

6. Simulation experiment and analysis

Tracking validation and evaluation of algorithm performance in CMM system based on simulation results and real commercial data. In the experiment, the algorithm is simulated in each discrete time interval, each simulation interval is added with 2×10^5 data, and the average arrival rate of each period is w, following the segmented Poisson distribution, the peak average request arrival rate $\lambda_{\max} =$100visits/SEC, and the minimum $\lambda_{\min} =$1 visit/SEC. The size and heat of the data follow either a Zipf distribution or a Pareto distribution. The purpose of the clustering algorithm used in the system is to group the devices. In Table 1, the parameters of the devices in several groups separated by the clustering group are summarized. According to the previous goal evaluation model, it can be seen that the ultimate goal of the system is to reduce the difference in resource utilization, so as to improve the load balancing degree of each device, so as to prevent local overload of the system and use system resources more rationally.

Table 1: Simulation experiment parameters

Parameter description	Parameter value
Pareto distribution parameters for data size	α=1.318; k=241
Zipf distribution parameter for data size	α=1.1
Zipf distribution parameter for access popularity	α=1.09
Maximum average arrival rate ($\lambda_{\max}$)	100 requests per second
Minimum average arrival rate ($\lambda_{\min}$)	1 Request per second
Each round is used to estimate the number of arrivals	2000+
The total number of data additions and relocations	2×10^5

Table 2: Parameters of the device group

Group number	Number of devices in a group	Capacity	Bandwidth	CPU	Internal memory
1	10	128GB	100Mb/s	2.3GHz	1GB
2	50	256GB	1000Mb/s	2.3GHz	1GB
3	100	2TB	2Gb/s	2.3GHz	1GB
4	20	500GB	10Gb/s	2.3GHz	1GB
5	30	600GB	200Mb/s	2.3GHz	1GB
6	40	700GB	2000Mb/s	2.3GHz	1GB

A core goal of the CMM system is to close the gap between resource utilization of all devices, avoid overloads due to imbalances, and enable the entire system to meet as many requests as possible. The experiments in this section compare the consistent hashing algorithm, multi-choice hashing algorithm and the corresponding algorithm of CMM system. When the total number of devices n is determined, the resource usage variance of the consistent hashing algorithm is highly positively correlated with the number of data m, and its boundary $\frac{m}{n} + O(\sqrt{\frac{m \ln n}{n}})$ is discussed above. As the number of data increases, the effect of consistent hashing algorithm is much lower than the performance of other algorithms, as shown in Figure 5.

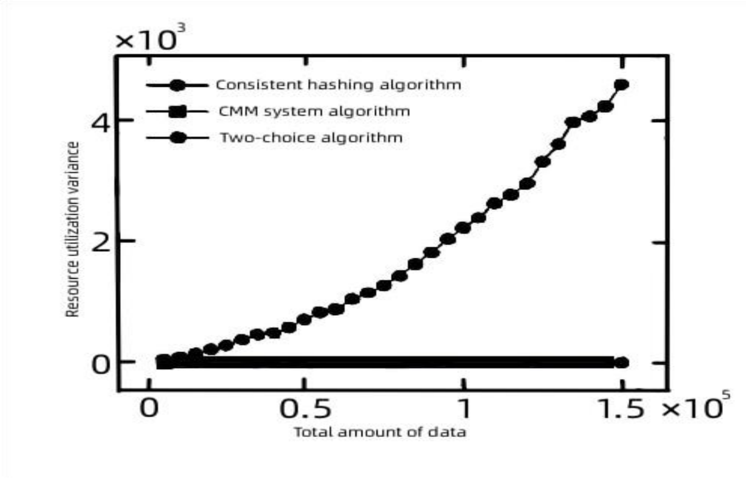

Figure 5. Comparison between generative hashing and multiple-choice algorithm

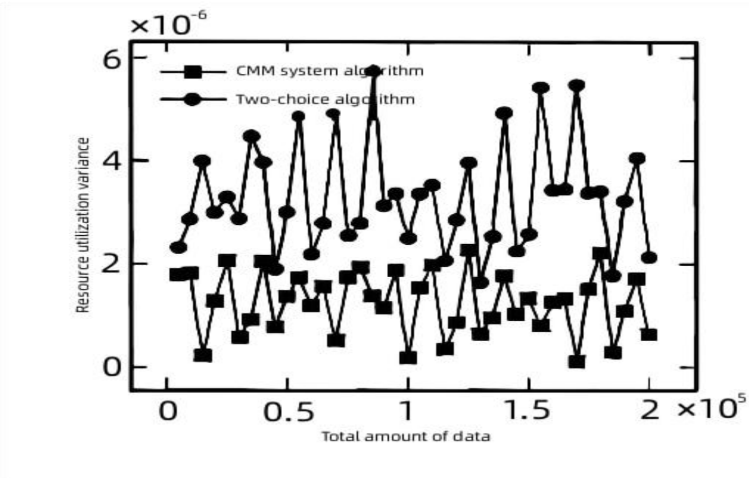

Figure 6. Comparison between CMM algorithm and two-choice algorithm

In the simulation, the consistent hashing algorithm even experienced severe device overload when the amount of data increased by 500 to around. Meanwhile, this class of algorithms (including the algorithm of CMM system) has no obvious positive correlation with the total number of data m, and its heaviest load upper bound is $\frac{m}{n}+\frac{\log\log n}{\log|D|}+O(1)$, which is only related to the number of devices n, so its upper bound will not change significantly in the real system for a period of time. As can be seen from FIG. 6, the variance of the two-choice hash algorithm (2-C) is approximately the variance of the algorithm in CMM, but it is obviously inferior to the algorithm in CMM. In addition, in the experiment, we also tested the case assuming that the resource consumption of each data item is uniform, which is the simplest case.

In this section, the CMM algorithm is used for further simulation experiments to make the analysis closer to the reality. As shown in Table 1, the parameters used in the simulation, these requests in the experiment are generated according to segmented Poisson distributions of J with different average rates, which have similar access patterns to those in the simulation experiment above. Data size and access heat follow the Pareto and Zipf distributions, respectively, and the improvement in the variance of resource usage is shown in Figures 7 and 8, indicating a more uniform load on device resources. It can also be clearly observed that the resource balance degree of each device in the CMM system is higher than that of the two-choice and multi-choice algorithms. Compared with these algorithms, according to the experimental results, the algorithm of the CMM system improves the balance degree by about 47% at most, and improves the balance degree by about 17% on average.

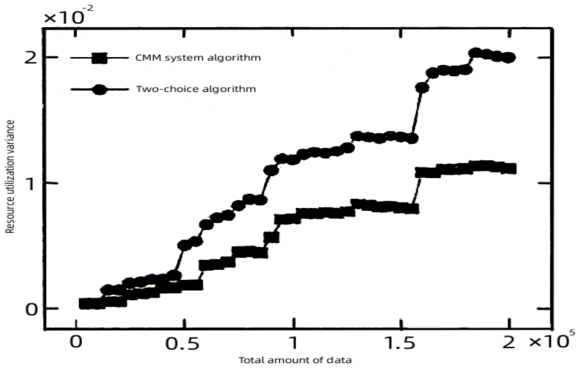

Figure 7. in the Zipf distribution

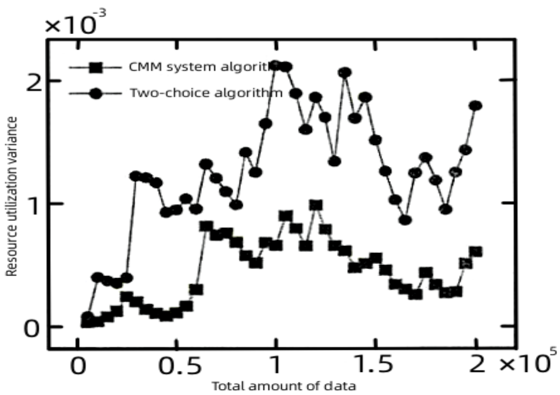

Figure 8. The situation under Pareto distribution

According to the Hadoop system manual, when using FUSE (the file system in user space), write operations can suffer efficiency losses of more than 30%. In fact, through actual testing, the corresponding loss in the actual HDFS client is about 50%. There is also a loss of about 10% performance on read operations, but because FUSE provides users with a convenient access interface, this is acceptable. As shown in Figure 9, the FUSE system and the CMM system have similar read speeds most of the time, because the experiments essentially hit the upper limits of the system's bandwidth and throughput capacity. However, as the amount of data increases, some devices in the FUSE system can begin to become overloaded, as shown in Figure 10, and the throughput and operational latency of the FUSE system begin to deteriorate later. In addition, because FUSE systems used in practice typically have many copies of data set up, the FUSE write speed is also reduced.

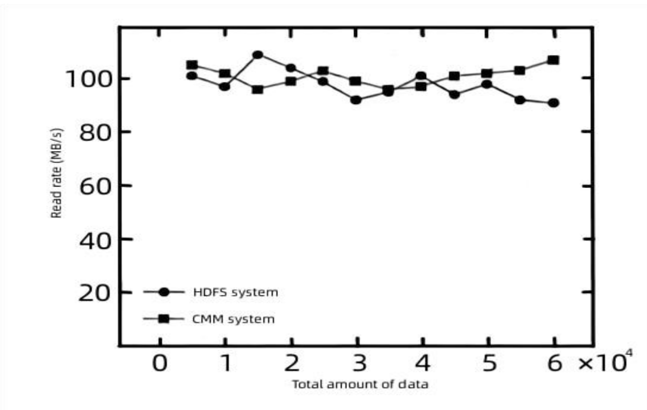

Figure 9. Comparison between the average read rate of the CMM system and the HDFS system

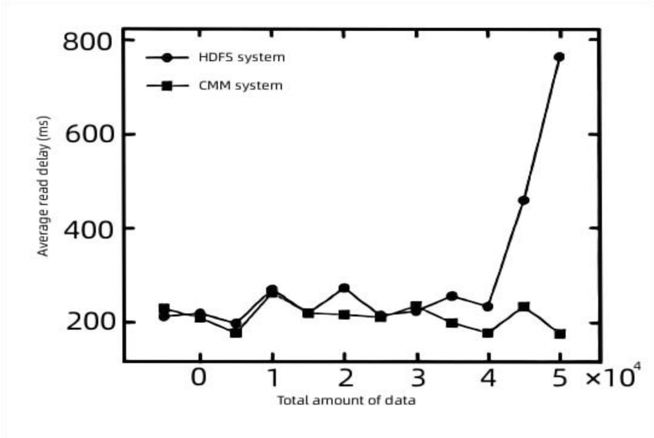

Figure 10. Comparison of average read operation latency between the CMM system and the HDFS system

7. Conclusion

With the clarity of the 5G concept, the industry is becoming more and more aware of the importance of 5G networks, and the cloud storage of big data is becoming more and more important. In order to solve the problems of poor storage security and high space occupancy in existing big data storage models and algorithms, an efficient hybrid cloud storage model and algorithm for big data under 5G network is proposed. The elasticity, utilization, and granularity of big data are determined as known values using 5G networks to achieve high access, buffering, and storage archiving of big data through three gradients. Simulation experiments show that the model and algorithm can realize hybrid cloud storage based on 5G network big data with low memory occupancy and high security, with smaller device load variance and more stable and balanced overall system load, which can provide reference for the development of this field.

References

[1] Li Jingyuan, Duan Juan. Cloud computing environment data security storage research. Journal of software Tribune, 2014, 13 (7): 3. DOI: CNKI: SUN: RJDK. 0.2014-07-054.

[2] Li Xiang, Chen Tianxiang, Liu Zhengyan. Data transmission and storage safety of cloud computing environment. Computer knowledge and technology, 2018 (2): 66-68. The DOI: CNKI: SUN: DNZS. 0.2018-23-027.

[3] Anonymous. User Privacy Hybrid encryption scheme and its performance analysis in cloud environment. Journal of chongqing university of posts and telecommunications: natural science edition, 2015, 27 (5): 8. DOI: 10.3979 / j.i SSN. 1673-825 - x. 2015.05.011.

[4] FanNing. Cloud storage platform design analysis. Journal of software Tribune, 2015, 14 (9): 2. DOI: 10.11907 / RJDK. 151654.

[5] Yang Ting, Wang Meng, Zhang Yajian, et al. Optimization Algorithm of HDFS Differential storage Energy Saving in Cloud Computing Data Center. Journal of Computer Science, 2019.

[6] Yang Xiaodan. Discussion on data storage under Computer Cloud Computing. Computer Programming Skills and Maintenance, 2021, 000(003):108-110.

[7] Liu Ruobing. Key technology research for big data cloud storage system. Journal of modern electronic technology, 2016, 39 (6): 5. DOI: 10.16652 / j.i SSN. 1004-373 - x. 2016.06.006.

[8] Ji Yilong, Yang Chuanbin. A hybrid cloud storage system design and implementation of campus. Journal of zhejiang normal university: natural science edition, 2019, 42 (2): 5. DOI: 10.16218 / j.i SSN. 1001-5051.2019.02.006.

[9] Li Yanhui, He Zuyuan, Tao Yingying. L1 fault detection for networked control systems with random measurement data loss and mixed time delay. Journal of northeast petroleum university, 2019, 43 (5): 10. DOI: CNKI: SUN: DQSY. 0.2019-05-012.

[10] She Wei, Chen Jiansen, Gu Zhihao, et al. Iot node location privacy protection based on the block chain model. Journal of applied science, 2020, 38 (1): 13. DOI: CNKI: SUN: YYKX. 0.2020-01-011.

[11] Xu Bin, Liang Xiaobing, Shen Bo. Big data privacy protection environment China Africa interactive query difference model. Computer engineering and application, 2020, 56 (7): 6. DOI: 10.3778 / j.i SSN. 1002-8331.1812-0081.

[12] Ma Xiuxiu. Discussion on Cloud Storage Data Security in Big Data Environment. China Equipment Engineering, 2022(6):2.

[13] Xue Chuanyu. Analysis on the problems and solutions of data security in cloud storage. Network Security Technology and Application, 2021(11):2.

[14] Qin Xiaobo, Huo Liwen, Chai Zhicheng. Key technology research for big data cloud storage system. Journal of geek, 2018 (6): 1. The DOI: 10.3969 / j.i SSN. 1672-528 - x. 2018.16.064.

[15] Guo Fangfang, Chao Luomeng, Zhu Jianwen. Multi-source data parallel preprocessing method based on similar connection. Journal of computer applications, 2019, 39 (1): 4. DOI: 10.11772 / j.i SSN. 1001-9081.2018071869.

[16] Zheng Baozhou, Wu Lili, Li Fuqiang, et al. Based on abnormal data preprocessing and WSN data fusion algorithm of adaptive estimation. Computer application research, 2019, 36 (9): 5. DOI: 10.19734 / j.i SSN. 1001-3695.2018.07.0345.

[17] Guan Zhaoxiong. The design of the hybrid storage algorithm under the cloud storage. Automation and instrumentation, 2018 (4): 4. DOI: 10.14016 / j.carol carroll nki. 1001-9227.2018.04.157.

[18] Li Yanmei. Gateway dump optimization algorithm based on hybrid cloud storage research. Journal of electronics, 2018 (7): 3. DOI: CNKI: SUN: DZZZ. 0.2018-07-020.

[19] Yang Yu-Jian, Lin Bo. Research on Consistency hashing algorithm in Distributed Storage Systems. Computer Knowledge and Technology: Academic Exchange, 2011.