www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 4 April 2015, Page No. 11212-11214

Improving Accessing Efficiency of Cloud Storage Using De-Duplication and Feedback Schemes

R.K.Saranya[1], R.Sanjana[2], Steffi Miriam Philip[3], Shahana M.S.A[4]

[1]Assistant Professor, Department of Computer Science and Engineering, [2, 3] B.E Final Year Students

Jeppiaar Engineering College, Chennai

saranya.rks@gmail.com[1], sanjanaramesh15@gmail.com[2], miriam.steffi@gmail.com[3], shahanamsa@gmail.com[4]

Abstract *File storage in cloud storage is handled by third parties. Files can be integrated, so that the users are able to access the files using the centralized management. Due to the great number of users and devices in the cloud network, the managers cannot effectively manage the efficiency of storage node. Therefore, hardware is wasted and the complexity for managing the files also increases. In order to reduce workloads due to duplicate files, we propose the index name servers (INS). It helps to reduce file storage, data de-duplication, optimized node selection, and server load balancing, file compression, chunk matching, real-time feedback control, IP information, and busy level index monitoring. Performance is also increased. By using INS the files can also be reasonably distributed and workload can be decreased*

Key Words *de-duplication,Load Balancing, Hash-Code Function.*

I.Introduction

Files stored in the cloud can be accessed at any time from any place so long as we will have Internet access. Another benefit is that cloud storage provides organizations with off-site backups of data which reduces costs associated with disaster recovery. Cloud storage can provide the benefits of greater accessibility and reliability; rapid deployment; strong protection for backup, archival and disaster recovery purposes; and lower overall storage costs as a result of not having to purchase, manage and maintain expensive hardware. However, cloud storage will have the potential for security and compliance concerns.

Data deduplication is one of the hottest technologies in storage right now because it enables companies to save a lot of money on storage costs to store the data and on the bandwidth costs to move the data when replicating it offsite for DR. This is great news for cloud providers, because if you store less, you need less hardware. If you can deduplicate what you store, you can better utilize your existing storage space, which can save money by using what you have more efficiently. If you store less, you also back up less, which again means less hardware and backup media. If you store less, you also send less data over the network in case of a disaster, which means you save money in hardware and network costs over time. The business benefits of data deduplication.

Load balancing distributes workloads across multiple computing resources, such as computers, a computer cluster, network links, central processing units or disk drives. Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid overload of any single resource. Using multiple components with load balancing instead of a single component may increase reliability through redundancy. Load balancing usually involves dedicated software or hardware, such as a multilayer switch or a Domain Name System server process.

II.Related Work

To decrease the workload caused by duplicated files, this paper proposes a new data management structure: Index Name Server (INS), which integrates data de-duplication with nodes optimization mechanisms for cloud storage performance enhancement. INS can manage and optimize the nodes according to the client-side transmission conditions. By INS, each node can be controlled to work in the best status and matched to

suitable clients as possible. IT is improved the that the performance of the cloud storage system efficiently distribute the files to reduce the load of each storage node. Techniques, such as run length encoding (RLE), dictionary coding, calculation for the digital fingerprinting of data chunks, distributed hash table (DHT), and bloom filter, there have been several investigations into load balancing in cloud computing systems. A digital fingerprint is the essential feature of a data chunk. Each data chunk has its unique fingerprint, and different chunks have different fingerprints.If it has same hash values, we can say that data with the same hash values must have the same original data, and that data with different hash values must have different original input data.

The bloom filter is composed of a long binary vector and a series of random mapping functions. The bloom filter is presented to test whether an element is included in the set. However, with the increase of the elements in the set, more storage space will be needed and the retrieval speed will be slowed down.

A DHT node does not maintain and possess all the information in the network, but stores only its own data and those of its neighboring nodes. This greatly reduces hardware and bandwidth consumption. Essentially, DHTs features include Decentralization, Scalability, Fault Tolerance.

keeps changing.

In existing system, the opportunistic load balancing (OLB) algorithm is used which keep the node busy. Thus, OLB does not consider the current workload of each node, but distributes the unprocessed tasks randomly to available nodes. Although OLB is easy and direct, this scheduling algorithm does not consider the expected task execution time and therefore cannot achieve good execution time in make span.

III. Present System & Framework

The INS(Index Name Server) uses a complex P2P-like structure to manage the cloud data. The INS principally handles the one-to-many matches between the storage nodes' IP addresses and hash codes.Three main functions of INS include:

1) Switching the fingerprints to their corresponding storage nodes;

2) Confirming and balancing the load of the storage nodes;

3) Fulfilling user requirements for transmission as possible.

In the present work, we are implementing the SHA-1 function to improve the efficiency. SHA-1 function is an advance technology that provides an enhanced functionality for cloud storage security and it protects the stored file. This novel technique will improve the service of cloud storage.This technique will notify the user if a duplicate file is present in cloud. Therfore it will automatically remove the duplicate files using the hash code functions. By doing this, the space in the cloud is increased, duplicate files are reduced, load balancing takes place and selection of optimised node.

IV. Algorithm

STEP 1: .1) R(k): The initial expected value

STEP 2: F(k): The output feedback;

STEP 3: M(k): The modified feedback;

STEP 4: Fs(k): The modified internal function of the storage node;

STEP 5: D(k): The external random variable;

STEP 6: X(k): The result within the storage node;

STEP 7: Y(k): The actual result;

STEP 8: KINS: The optimal node determined by the SHA-1 based on the feedback.

V. Conclusion

We proposed the SHA-1 to process not only file compression, chunk matching, data de-duplication, real-time feedback control, IP information, and busy level index monitoring, but also file storage, optimized node selection, and server load balancing.

Based on several SHA parameters that monitor IP information and the busy level index of each node, our proposed scheme can determine the location of maximum loading and trace back to the source of demands to determine the optimal backup node.

According to the transmission states of storage nodes and clients, the SHA-1 received the feedback of the previous transmissions and adjusted the transmission parameters to attain the optimal performance for the storage nodes. By compressing and partitioning the files according to the chunk size of the cloud file system.

REFERENCES

[1] Y.-M. Huo, H.-Y. Wang, L.-A. Hu, and H.-G. Yang, "A cloud storage architecture model for data-intensive applications," in Proc Int. Conf Comput. Manage., May 2011, pp. 1–4.

[2] L. B. Costa and M. Ripeanu, "Towards automating the configuration of a distributed storage system," in Proc. 11th IEEE/ACM Int. Conf. Grid Comput., Oct. 2010, pp. 201–208.

[3]C.-Y. Chen, K.-D. Chang, and H.-C. Chao, "Transaction pattern based anomaly detection algorithm for IP multimedia subsystem, IEEE Trans Inform. Forensics Security, vol. 6, no. 1, pp. 152–161, Mar. 2011.

[4]G. Urdaneta, G. Pierre, and M. Van Steen, "A survey of DHT security techniques," ACM Comput. Surveys (CSUR), vol. 43, no. 2, pp. 8:1–8:49, Jan. 2011.

[5]T.-Y. Wu, W.-T. Lee, and C. F. Lin, "Cloud storage performance enhancement by real-time feedback control and de-duplication," in Proc Wireless Telecommun. Symp., Apr. 2012, pp. 1–5.

[6]H. He and L. Wang, " P&P: A combined push-pull model for resource monitoring in cloud computing environment," in Proc. IEEE 3rd Int Conf. Cloud Comput., Jul. 2010, pp. 260–267

[7]R. Tong and X. Zhu,. "A load balancing strategy based on the combination of static and dynamic," in Proc. 2nd Int. Workshop Database Technol. Appl., Nov. 2010, pp. 1–4

[8]T.-Y. Wu, W.-T. Lee, Y.-S. Lin, Y.-S. Lin, H.-L. Chan, and J.-S. Huang, "Dynamic load balancing mechanism based on cloud storage," in Proc Comput. Com. Appl. Conf., Jan. 2012, pp. 102–106.

[9]Y. Zhang, C. Zhang, Y. Ji, and W. Mi, " A novel load balancing scheme for DHT-based server farm," in Proc. 3rd IEEE Int. Conf Comput. Broadband Netw. Multimedia Technol., Oct. 2010, pp. 980–984.