A Survey On: De-Duplication With Security On Various Technologies

K.Uma , Research scholar,
PG &Research Department of computer science,
Government Arts College(Autonomous), Salem – 636007

Mr.E.Jayabalan, Assistant professor ,
PG &Research Department of computer science,
Government Arts College(Autonomous), Salem – 636007

The great significance technique De-duplication is removing duplicate copies from cloud, also to reduce the storage space, time consumption and save bandwidth. To keep the data securely at the same time as help of de-duplication, to encrypt the large data set. It is maintaining the data in cloud by encrypted form. Encrypted data de-duplication suffers from some security weakness. In cost effective manner product massive amount of data in cloud. In this paper we explain about enormous information of de-duplication on encrypted large data set, performance of primary storage system in cloud, common techniques of deduplication and also secure deduplication of encrypted cloud center.

Keyword: De-duplication, Secure de-duplication, Primary storage, Proxy re-encryption, Storage capacity, proof of ownership

1.INTRODUCTION

Cloud computing provide a storage space with the help of remote server and software networks over the internet. It has independent location infrastructure for data storage. Depend upon the user demand they rearranging various resources and providing by new technology. Many of the users store their data in cloud and also they are increasing their volume of data. To make the data safe and secure with the help of some technique in low of cost and time efficiency.

The importance of cloud service is data storage service. Most of the cloud users upload their personal details and business work. Cloud service provider managing data with security.

In This paper explorer to avoid deduplication in cloud computing. Hence, then the cloud computing Segmentized into private and public or hybrid cloud.

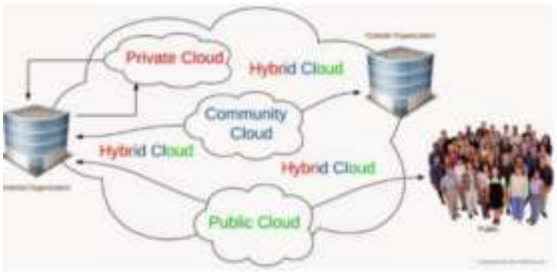

Figure 1.1 (deployment model)

Private cloud: It operated only within a single organization. Inside or by third party may manage it. It has some advantages that are called higher security and privacy control and cost and energy efficiency. It is internal cloud. It is safe guard by firewall. It only for authorized user for company purpose [17].

Public cloud: It is a render services and cloud hosting type, also it allows systems and services to be easily accessible to general public. It has some more advantage that is cost effective, reliability, flexibility, location independence, utility style costing and high scalability. It suite for business requirements [17].

Hybrid cloud: It is a concept of mishmash of public and private cloud. Noncritical activities are performed using public cloud and the critical activities are performed using private cloud. It has some more advantages that is scalability, flexibility, cost efficiencies, security [17].

Community cloud: It manually shared between many organization and business.

In 2016 the technique of data ownership challenge and proxy reencryption (PRE) to manage encrypted data storage with de-duplication is proven by zhengya et al [1].

In cloud highly increasing the data of organization and all. We need to share data with de-duplication even when data holder are offline. The encrypted data should be securely accessed because of only authorized data holders can obtain the symmetric keys used form the data decryption.

In cloud some of the attacker is there, for example Brute-Force attack, this one is in cryptography, a brute-force attack to consist of attacker trying to use hope of guessing passwords correctly. The password should be secure with encrypted data.

2.LITERATURE SURVEY

Following are the different which are used in data de-duplication in cloud storage.

Zhengyan et al [1], proposed ownership challenge and PRE to handle the encrypted big data in cloud. This is the method of flexible delay data sharing and update with deduplication even offline data holder also can do it. Encrypted data can be securely obtained because only authorized data holders can obtain the symmetric keys used for data decryption.

Pasqalepuzio et al [2], proposed cloud de-dup, a secure and systematic storage service which assures block-level de-duplication and data trustworthy at the same time.

Tin-yu et al [3], Proposed the INS to process not only file. Compressing, chunk matching, data de-duplication, real-time feedback control, IP information, and busy level index servile, but also file storage, optimized node selection, and sever load balancing.

Rongmaochan et al [4], Proposed scheme is proven secure in the random oracle modelm asked whether it is possible to design efficient BL-MLE schemes that are poven secure in th standard model.

Jibinwang et al [5], proposed Isieve, a high performance inline deduplication system for use in cloud storage, then design novel index tables to satisfy the I-sieve, architecture, since it is a bridge between frontend and backend system, and also implement a prototype of I-sieve based on iSCSI, and evaluate it with virtual machines and the IOmeter tool.

Yifengzheng et al [6], presented a secure system architecture design as our initial effort towards this direction, which bridges together the advancements of video coding techniques and secure encrypted deduplication.

Bo Mao, et al [7], proposed POD, a performance –oriented de-duplication scheme, to improve the performance of primary storage systems in the cloud by leveraging data deduplication on the I/O path to remove redundant write requests while also saving storage space.

Jin Li et al [8], presented several new de-don constructions supporting authorized duplicate check in hybrid cloud architecture, in which the deduplicate check tokens of files are generated by the private cloud server with private keys. Then security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model.

SchengmeiLuo et al [9], Presented Boafft, a cloud Storage system with distributed de-duplication when achieved scalable throughput and capacity using multiple storage nodes to de-duplicate in parallel, with a minimal loss of deduplication ratio. Finally they improved the data deduplication ratio in single node with the help of cache container of hot fingerprint based on access frequency.

Xunzhao et al [10], Proposed liquid ,which is a distributed file system particularly designed to simultaneously address the above problems faced in largescale VM deployment, while client side breaks VM images into small data blocks, references them by their fingerprints, and uses de-duplication techniques to avoid storing redundant data blocks.

Jin Li, et al [11], Proposed new distributed de-duplication systems with higher reliability in which the data chunk are distributed across multiple cloud servers, when the security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous de-duplication systems are secure in terms of the definitions specified in the proposed security model after that As a proof of concept, implemented the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.

Jin Li et al [12], proposed dekey , an efficient and reliable convergent key management scheme for secure deduplication . Dekey applies de-duplication among convergent keys and distributes convergent key shares across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data. We implement Dekey using the Ramp secret sharing scheme and demonstrate that is incurs small encoding/decoding overhead compared to the network transmission overhead in the regular upload/download operations.

Jingwei Li et al [13], proposed SecCloud and SecCloud+. SecCloud introduces an auditing entity with maintenance of a Map Reduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. Jingwei Li et al [14], presented REED, an encrypted deduplication storage system that aims for secure and lightweight rekeying. The core rekeying design of REED is to renew a key of a deterministic all-or-nothing-transform (AONT) package. Proposed two encryption schemes for REED: the basic scheme has higher encryption performance, while the enhanced scheme is resilient against key leakage.

3.OVERVIEW OF THE DE-DUPLICATION

De-duplication is a technique of reduce storage space from cloud as well as maintain the data in cloud by encrypted form. The main thing of de-duplication is also time consumption and save bandwidth.

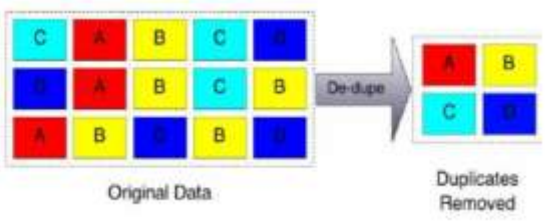

Figure 1.2(de-duplication)

3.1 De-duplication types [15]

3.1.1 Source De-duplication

Before transmission to backup target it reduce the duplicate form data in cloud. Comparing new data blocks on the primary device by it uses the client software.

3.1.2 Target De-duplication

In backup appliance it removes the duplicated copies from data set in cloud computing most often on virtual tape library or NAS device.

3.1.3 Inline De-duplication

Being return to a backup device it removing repeated data and to reduce storage space in cloud. It reduces the repeated data in an application and backup disk target.

3.1.4 Post-process De-duplication

It starts to remove duplicate copies from data set in cloud through after the backup data into the cache. It uses in cloud for backup application.

3.1.5 Global De-duplication

In multiple de-duplication device when data backing up then it is a method of preventing redundant data. Across multiple systems it removes all the possible repeated data from cloud.

3.2 Common methods for de-duplication technique [16]

3.2.1 File-based compared

This is the simple method for removing duplication. File system based algorithm just compare with in the file depend upon data name, size and date to remove the repeated data from cloud. If with the file stored same file then the duplicate file automatically removed.

3.2.2 File-based delta versioning and hashing

File-based delta versioning are update to file and it just stores the delta as other version. File-based delta hashing are create unique mathematical hash of file, and it compare original for new file. If there is match hashing then we can find it is already available in cloud. Hashing algorithm results depend upon some methods. There is available few more methods for hash algorithm.

3.2.3 Sub-block delta versioning and hashing

Comparing block level is the subblock delta versioning is more efficient for removing repeated data from cloud. For eg: UNIX format disks. But more hashes are required for smaller chunk usage. Hashing technique to providing deleting duplicates and reducing storage space for some particular data set from cloud. And there is not having huge disadvantage.

PROPOSED SYSTEM

We propose to improve the efficiency and high speed, preprocessing, flash memory, high performance of my research with the help of new Algorithm. There are more new techniques available for this. For example CIDS algorithm and ECO algorithm. We design and implement by CIDS algorithm for product the massive amount of data with security through WAN.

CONCLUSION

De-duplication is also very important to manage data with security while sharing information. Here we presented encrypted data service especially for big data, file based de-duplication etc. Here, We proposed scheme to improve the efficiency and high speed, preprocessing, flash memory, high performance of work with the help of new algorithm. We also present several new deduplication constructions which are supporting authorized duplicate check in hybrid cloud architecture.

REFERENCES:

[1] ZhengYan,Wenziu Ding, Xixun Yu, Haiqi Zhu and Robert H.Deng, “Deduplication on encrypted big data in cloud” 2016 IEEE Transactions on Vol.2, No.2.

[2] P. Puzio, R. Molva, M. Onen, and S. Loureiro, “ClouDedup: Secure deduplication with encrypted data for cloud storage,” in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370, doi:10.1109/CloudCom.2013.54. [3] Tin-Yu Wu, Jeng-Shyang Pan and Chia-Fan Lin, “Improving Accessing Efficiency of Cloud Storage Using De-Duplication and Feedback Schemes”, IEEE SYSTEMS JOURNAL, VOL. 8, NO. 1, MARCH 2014. [4] Rongmao Chen*, Yi Mu*, Guomin Yang, and FuchunGuo, BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication”: DOI 10.1109/TIFS.2015.2470221, 2015. [5] Jibin Wang, Zhigang Zhao, ZhaogangXu, Hu Zhang, Liang Li, and Ying guo, “I-Sieve: An Inline High Performance De-Duplication System Used in Cloud Storage”,IEEEvol 20, no 1, February 2015. [6] YifengZheng, Xingliang Yuan, Xinyu Wang, Jinghua Jiang, Cong Wang, and XiaolinGui, “Towards Encrypted Cloud Media Center with Secure Deduplication”, IEEE: DOI 10.1109/TMM.2016.2612760, [7] Bo Mao, Hong Jiang, Suzhen Wu, and Lei Tian, “Leveraging Data Deduplication to Improve the Performance of Primary Storage Systems in the Cloud” IEEE: DOI 10.1109/TC.2015.2455979, 2015.

[8] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, “A Hybrid Cloud Approach for Secure Authorized Deduplication”:DOI 10.1109/TPDS.2014.2318320, IEEE Vol:PP No:99 year 2014. [9] ShengmeiLuo, Guangyan Zhang, Chengwen Wu, Samee U. Khan, and Keqin Li, Boafft: Distributed Deduplication for Big Data Storage in the Cloud, IEEE Transactions On Cloud Computing, vol. 61, no. 11, January 2015 [10] Xun Zhao, Yang Zhang, Yongwei Wu, Kang Chen, Jinlei Jiang, and Keqin Li “Liquid: A Scalable Deduplication File System for Virtual Machine Images” IEEE, vol. 25, no. 5, May 2014

[11] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang, and Mohammad Mehedi Hassan and AbdulhameedAlelaiwi “Secure Distributed Deduplication Systems with Improved Reliability”, 10.1109/TC.2015.2401017, year 2015.

[12] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, Secure Deduplication with Efficient and Reliable Convergent Key Management, IEEE, vol. 25, no. 6, JUNE 2014

[13] Jingwei Li, Jin Li, DongqingXie and Zhang Cai, ”Secure Auditing and Deduplicating Data in Cloud”, IEEE: DOI 10.1109/TC.2015.2389960 Year 2015. [14] Jingwei Li1,Chuan Qin1, Patrick P. C. Lee1, and Jin Li2, “Rekeying for Encrypted Deduplication Storage” 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. [15] http://www.slideshare. net / And rewMaclen/data de-duplication-and-itsdifferent-types.

[16] http://www.computerworld.com/ article/2475106/cloudcomputing /datade-duplication-in-the-cloud explained—part-two—the-deepdive.html.

[17] https://www.ibm.com/developer works /community