|| Volume 5 || Issue 12 || April 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH & ENGINEERING TRENDS
Multidisciplinary Journal
Double-Blind Peer Reviewed Refereed Open Access International Journal

SECURE STORAGE AT CLOUD WITH DUPLICATION CHECKING

Dhamane Yogesh Arun[1], Ruchika Pundlikrao Tekade[2], Rohit Kumar Agalave[3], Salunke Shubham Sanjay[4],

Prof. M. R. Dhage[5]

Dept of Computer Engineering, SKN SITS, Lonavala,

Savitribai Phule Pune University, Pune, Maharashtra, India.

----------***----------

Abstract: Cloud computing plays a major role in the business domain today as computing resources are delivered as a utility on demand to customers over the Internet. Cloud storage is one of the services provided in cloud computing which has been increasing in popularity. The main advantage of using cloud storage from the customers' point of view is that customers can reduce their expenditure in purchasing and maintaining storage infrastructure while only paying for the amount of storage requested, which can be scaled-up and down upon demand. With the growing data size of cloud computing, a reduction in data volumes could help providers reducing the costs of running large storage system and saving energy consumption. So data deduplication techniques have been brought to improve storage efficiency in cloud storages. With the dynamic nature of data in cloud storage, data usage in cloud changes overtime, some data chunks may be read frequently in period of time, but may not be used in another time period. Some datasets may be frequently accessed or updated by multiple users at the same time, while others may need the high level of redundancy for reliability requirement. Therefore, it is crucial to support this dynamic feature in cloud storage. However current approaches are mostly focused on static scheme, which limits their full applicability in dynamic characteristic of data in cloud storage. In this project, we propose a dynamic deduplication scheme for cloud storage, which aiming to improve storage efficiency and maintaining redundancy for fault tolerance.

Keywords: *Data deduplication, cloud, AES, MD5, Java, JSP & Servlet, etc*

---------- ***----------

I INTRODUCTION

The standard ABE system fails to achieve secure deduplication which is a technique to save storage space and network bandwidth by eliminating redundant copies of the encrypted data stored in the cloud. On the other hand, to the best of our knowledge, existing constructions, for secure deduplication are not built on attribute-based encryption. Nevertheless, since ABE and secure deduplication have been widely applied in cloud computing, it would be desirable to design a cloud storage system possessing both properties. In this system consider the following scenario in the design of an attribute-based storage system supporting secure deduplication of encrypted data in the cloud, in which the cloud will not store a file more than once even though it may receive multiple copies of the same file encrypted under different access policies. However, endowing such a tag checking ability to the private cloud is not sufficient to achieve deduplication in the attribute-based storage system which employs CP-ABE for data encryption. In the proposed attributed-based system, the same file could encrypted to different cipher texts associated with different access policies, storing only one cipher text of the file means that users w hose attributes satisfy the access policy of a discarded cipher text (but not that of the stored cipher text) will be denied to access the data that they are entitled to. To overcome this problem, we equip the private cloud with another capability named ciphertext regeneration.

Concerning the adversarial model of our storage system, we assume that the private cloud is curious-but-honest such that it will attempt to obtain the encrypted messages but it will honestly follow the protocols, whereas the public cloud is distrusted such that it might tamper with the label and ciphertext pairs outsourced from the private cloud (note that such a misbehavior will be detected by either the private cloud or the user via the accompanied label). Another difference between the private cloud and the public cloud is that the former cannot collude with users, but the latter could collude with users. This assumption is in line with the real world practice where the private cloud is trusted more than the public cloud. We assume that data users may try to access data beyond their authorized privileges. In addition to trying to obtain plaintext data from the cloud, malicious outsiders may also commit duplicate faking attacks as described before. The system can deduce that both security and performance are critical for the next generation large-scale systems, such as clouds. There-fore, in this project, collectively approach the issue of security and performance as a secure data replication problem. In the system present Division and Replication of Data in the Cloud for Optimal Performance and Security that judicially fragments user files into pieces and replicates them at strategic locations within the cloud. The division of a file into fragments is performed based on a given user criteria such that the individual fragments do not contain any meaningful information. Each of the cloud nodes (In this system use the term node to represent computing, storage, physical, and virtual machines) contains a distinct fragment to increase the data security.

II RELATED WORK:

In 2013, Chun-Ho Ng et.al proposed RevDedup algorithm to find and remove duplicates from virtual machine images. Whenever new VM image comes, the RevDedup find the similarity with old data and removes it from old data [2]. In the same year, Mihir Bellare et.al proposed a cryptographic approach called Message-Locked Encryption (MLE). In MLE the keys used for encryption and decryption are derived from message itself. It was the secured way to carry out deduplication [3]. In 2014, Zhou Lei et.al proposed a mechanism using fixed size block method to store images. This method calculates compact digest called fingerprint for each image file and hence make the directory of fingerprints. For new image input it calculates fingerprint and compares with available fingerprint library [4]. In the same year, Waraporn Leesakul et.al proposed a new scheme to improve efficiency of cloud storage space using dynamic data deduplication. This scheme improved storage space along with maintaining the redundancy [5]. In the same year, Issa M. Khalil et.al identified 28 cloud security issues through his survey on security issues in clouds and security solutions [6]. In 2015, N. Jayapandian et.al proposed the scheme based on authorization. This system has the feature to protect user data confidentiality using differential privileges based on duplicate check [7]. In the same year, Mi Wen et.al developed a scheme using convergent encryption technique for secure deduplication scheme [8]. In the same year, Lakshmi Pritha et.al developed a system using RSS key to provide secure access to cloud resources and demonstrated ALG technique for data deduplication [9]. In the same year, Chun-I Fan et.al proposed check block mechanism for encrypted data deduplication [10]. In the same year, Mr. Dame Tirumala Babu et.al presented a method for data deduplication based on authorization to secure data [11]. In 2016, Shuai Wang et.al proposed a RRMFS file system to support data deduplication. [12].In the same year, Zheng Yan et.al presented a scheme for ownership and re-encryption to deduplicate encrypted data stored in cloud [13]. In the same year, Naresh Kumar et.al performed a comparative analysis of various deduplication techniques is done using destor tool. Data deduplication technique uses several chunking algorithms fixed length and variable length chunking [14]. In the same year, Jun Ren et.al proposed a method based on differential privacy for secure data deduplication [15]. In the same year, Saurabh Singh etc. All provided cloud security survey with discussion about security issues and challenges [16]. In the same year, Feilong Tang et.al introduced an approach called Load Balanced Flow Scheduling approach for dynamic load balancing and to maximize network throughput [17]. In 2017, Danoing Li et.al proposed a method called CSPD using modified DCT-

|| Volume 5 || Issue 12 || April 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH & ENGINEERING TRENDS
Multidisciplinary Journal
Double-Blind Peer Reviewed Refereed Open Access International Journal

based Perceptual Image Hash (D-phash) to improve the accuracy of the duplicate check [18]. In the same year, Hui Cui et.al implemented an ABE encryption system for cloud storage based on attributes [19]. In the same year, Rayan Dasoriya et.al presented a dynamic load balancing algorithm to distribute the load across multiple connected network links [20]. In the same year, Shunrong Jiang et.al proposed data confidentiality and ownership management system for data deduplication based on Proof of Ownership (PoW) technique [21]. In the same year, Himshai Kambo et.al implemented a secure deduplication mechanism based on CDC and MD5 algorithm. CDC used to break the data streams using randomization and MD5 algorithm creates the hash values for the segments or chunks created by CDC. It was used to improve network bandwidth [22].

III PROPOSED SYSTEM:

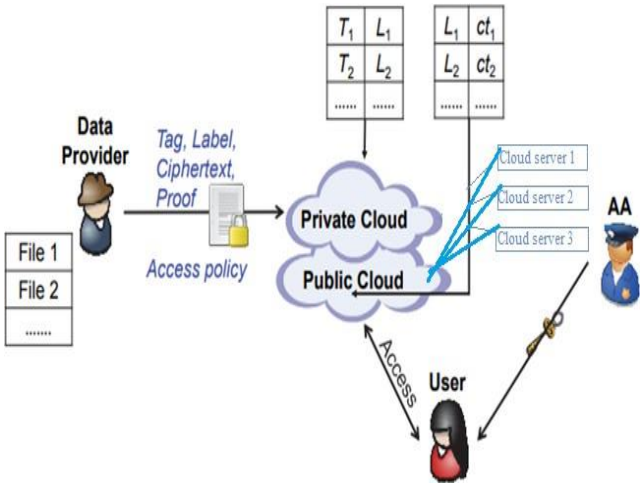

Fig 1: Proposed System

An attribute-based storage system supporting secure de-duplication. Our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage. Attribute based storage system supporting secure de duplication of encrypted data in the cloud, in which the cloud will not store a file more than once even though it may receive multiple copies of the same file encrypted under different access policies. The Attribute Authority issues every user a decryption key associated with the set of attributes. The attribute based storage system check the duplication of the file. The duplication is not occur, the file is stored. If the duplication is occurring, the attribute authority changes the ownership permission. In this system utilizing client accreditations to check the confirmation of the client. In that cases cloud is available two sort of cloud such private cloud and open cloud. In private cloud store the client accreditation and in the open cloud client information present out. The system have utilized a half and half cloud construction modeling as a part of proposed. In this system have to need to mind the file name in record information duplication and information DE duplication is checked at the square level. On the other hand, client needs to recover his information or download the information record he have to download both of the document from the cloud server this will prompts perform the operation on the same record this abuses the security of the distributed storage. Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that collectively approaches the security and performance issues. In this project, DROPS methodology, divide a file into fragments, and replicate the fragmented data over the cloud nodes. Each of the nodes stores only a single fragment of a particular data file that ensures that even in case of a successful attack, no meaning-ful information is revealed to the attacker.

Algorithm:

Step 1.Append Padding Bits. The message is ”padded” (extended) so that its length (in bits) is congruent to 448, modulo 512. ...

Step 2. Append Length. ...

Step 3. Initialize MD Buffer. ...

Step 4. Process Message in 16-Word Blocks. ...

Step 5. Output.

In cryptography, MD5 (Message-Digest algorithm 5) is a widely used cryptographic hash function with a 128-bit hash value. As an Internet standard (RFC 1321), MD5 has been employed in a wide variety of security applications, and is also commonly used to check the integrity of files.

|| Volume 5 || Issue 12 || April 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH & ENGINEERING TRENDS
Multidisciplinary Journal
Double-Blind Peer Reviewed Refereed Open Access International Journal

An MD5 hash is typically expressed as a 32 digit hexadecimal number.

IV RESULT:

|| Volume 5 || Issue 12 || April 2021 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH & ENGINEERING TRENDS
Multidisciplinary Journal
Double-Blind Peer Reviewed Refereed Open Access International Journal

V CONCLUSION:

Thus we are going to develop a system for secure deduplication in cloud computing. Here the files will be first checked either they have been already uploaded or not , and if any file is already uploaded then it will not be uploaded again. This system will help to improve the efficiency of the cloud storage system. It will solve the problem of availability of storage space to great extent.

REFERENCES:

[1] Shyam Patidar, Dheeraj Rane, Pradesh Jain, "A Survey Paper on Cloud Computing", Proceeding ACCT '12 Proceedings of the 2012 Second International Conference on Advanced Computing & Communication Technologies, pp 394-398, January 07 - 08, 2012.

[2] Chun-Ho Ng, Patrick P. C. Lee, "RevDedup: A Reverse Deduplication Storage System Optimized for Readsto Latest Backups", Proceeding APSys '13 Proceedings of the 4th Asia-Pacific Workshop on Systems, Article No. 15, Singapore, July 29 - 30, 2013.

[3] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart, "Message-Locked Encryption and Secure Deduplication", Annual International Conference on the Theory and Applications of Cryptographic Techniques, EUROCRYPT 2013: Advances in Cryptology – EUROCRYPT, Lecture Notes in Computer Science, vol 7881, Springer, Berlin, Heidelberg, pp 296-312, 2013.

[4] Zhou Lei, ZhaoXin Li, Yu Lei, YanLing Bi, Luokai Hu, Wenfeng Shen, "An Improved Image File Storage Method Using Data Deduplication", TrustCom 2014, The 13th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, Beijing, China, pp 638-643, 24-26 September 2014.

[5] Waraporn Leesakul, Paul Townend and Jie Xu, "Dynamic Data Deduplication in Cloud Storage", SOSE 2014, IEEE Eighth International Symposium On Service-Oriented System Engineering Oxford, United Kingdom, pp. 7-11 April 2014.

[6] Issa M. Khalil, Abdallah Khreishah and Muhammad Azeem, "Cloud Computing Security: A Survey", Article in 'Computers', Open Access Journal, Vol and Issue 3(1), pp. 1-35, 3 February 2014.

[7] N.Jayapandian, Dr.A.M.J.Md.Zubair Rahman and I.Nandhini, "A Novel Approach for Handling Sensitive Data with Deduplication Method in Hybrid Cloud", Online International Conference on Green Engineering and Technologies, November 2015.

[8] Mi Wen, Kejie Lu, Jingsheng Lei, Fengyong Li, Jing Li, "BDO-SD: An Efficient Scheme for Big Data Outsourcing

with Secure Deduplication", the Third International Workshop on Security and Privacy in Big Data, IEEE 2015.

[9] N. Lakshmi Pritha and N.Velmurugan, "Deduplication Based Storage and Retrieval of Data from Cloud Environment" in International Conference on Innovation Information in Computing Technologies, Chennai, pp. 1-6, IEEE 2015.

[10] Chun-I Fan and Shi-Yuan Huang, "Encrypted Data Deduplication in Cloud Storage", Article in 'ASIAJCIS' 15 Proceedings of the 2015 10th Asia Joint Conference on Information Security, pp.18-25, May 24-26, 2015, IEEE Computer Society, Washington, ISBN: 978-1-4799-1989-5.

[11] Dama Tirumala Babu and Yaddala Srinivasulu, "A Survey on Secure Authorized Deduplication Systems", International Research Journal of Engineering and Technology. Volume: 02 Issue: 05. Aug-2015.

[12] Shuai Wang and Jianhai Du "A Storage Solution for Multimedia Files to Support Data Deduplication", 2016 2nd International Conference on Cloud Computing and Internet of Things, Dalian, China, pp-78-8, 2016.

[13] Zheng Yan and Wenxiu Ding," Deduplication on Encrypted Big Data in Cloud", IEEE Transactions on Big Data, Vol. 2, No. 2, April-June, 2016.

[14] Naresh Kumar, Preeti Malik, Sonam Bhardwaj, Sushil Chandra Jain, "Comparative Analysis of Deduplication Techniques for Enhancing Storage Space", 4th International Conference on Parallel, Distributed and Grid Computing. IEEE, 2016.

[15] Jun Ren and Zhiqiang Yao, "A Secure data deduplication scheme based on differential privacy", IEEE 22nd International Conference on Parallel and Distributed System, pp-1241-1246, 2016.

[16] Saurabh Singh and Young-Sik Jeong, "A Survey on Cloud Computing Security: Issues, Threats, and Solutions", in Journal of Network and Computer Applications, pp-1-30, 2016.

[17] Feilong Tang and Laurence T. Yang, "A Dynamical and Load-Balanced Flow Scheduling Approach for Big Data Centers in Clouds", IEEE Transactions On Cloud Computing 2016.

[18] Danping Li, Chao Yang, Chengzhou Li, Qi Jiang, Xiaofeng Chen, Jianfeng Ma, and Jian Ren, "A Client-based Secure Deduplication of Multimedia Data", Communication and Information Systems Security Symposium. IEEE, 2017.

[19] Hui Cui, Robert H. Deng, Yingjiu Li, and Guowei Wu, "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud", IEEE Transactions on Cloud computing, year: 2017.

[20] Mr. Rayan Dasoriya, Ms. Purvi Kotadiya, Ms. Garima Arya, Mr. Priyanshu Nayak, "Dynamic Load Balancing in Cloud: A Data-Centric Approach", International Conference on Networks & Advances in Computational Technologies. IEEE, 2017.Shunrong Jiang, Tao Jiang and Liangmin Wang, "Secure and Efficient Cloud Data Deduplication with Ownership Management", IEEE Transactions on Services Computing. IEEE, 2017.

[21] Himshai Kambo, Bharati Sinha, "Secure Data Deduplication Mechanism based on Rabin CDC and MD5 in Cloud Computing Environment", 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT).Bangalore, pp 400-404, May 19-20, 2017, India.