Secure Storage at Cloud with Duplication Checking

Sujata Madhukar Pawar, Prof. S.V.Bodake

Dept of Computer Engineering,

PVPIT, Savitribai Phule Pune University, Pune, Maharashtra, India.

Abstract:

Cloud computing provides users with computer resources as a utility on demand over the Internet, and it is increasingly widely used in the commercial sector. Cloud storage has become one of the most popular cloud computing services. Customers benefit the most from cloud storage because they may save money on storage equipment purchases and maintenance by just paying for the storage capacity they need, which can be scaled up and down on demand. With the expanding amount of data in cloud computing, a reduction in data quantities could help providers save money by lowering operating costs and conserving energy. As a result, data deduplication techniques have been used in cloud storage to improve storage efficiency. Due to the dynamic nature of data in cloud storage, data utilisation in the cloud fluctuates over time. For example, some data items may be downloaded frequently one time but not the next. Some datasets are frequently viewed or edited by numerous persons at once, while others require a high level of redundancy to be stable. As a result, cloud storage with this dynamic functionality is critical. Current solutions, on the other hand, are primarily focused on a static system, which limits their application to the dynamic nature of data in cloud storage. . We describe a dynamic deduplication strategy for cloud storage in this research, with the goal of improving storage economy while maintaining fault tolerance redundancy.

Keywords: *hash generation, message digest, encryption, storage space, duplication, etc*

I. Introduction:

The basic ABE system does not support secure deduplication, a technique for reducing storage space and network bandwidth by deleting redundant copies of encrypted data stored in the cloud. Existing architectures for safe deduplication, on the other hand, do not use attribute-based encryption, as far as we know. Given the ubiquitous use of ABE and safe deduplication in cloud computing, a cloud storage solution that possesses both properties would be perfect. Consider the following scenario when creating an attribute-based storage system that allows for secure cloud deduplication of encrypted data: The cloud will not store the same material more than once, even if it gets several versions encrypted with different access permissions. However, in an attribute-based storage system that uses CP-ABE for data encryption, providing the private cloud with such a tag checking capability is insufficient to conduct deduplication. In the proposed attributed-based system, the same file could be encrypted to different cypher texts with different access policies; however, storing only one cypher text of the file means that users with attributes that match the access policy of a discarded cypher text will be denied access to the data they are entitled to. To address this issue, we've added ciphertext regeneration as a new feature to the private cloud. In terms of the adversarial model of our storage system, we assume that the private cloud is curious but honest, in that it will try to obtain the encrypted messages but will adhere to the protocols honestly, whereas the public cloud is distrusted, in that it may tamper with the label and ciphertext pairs outsourced from the private cloud (note that such a misbehaviour will be detected by either the private cloud or the user via the accompanied label). Another difference between public and private clouds is that the former cannot collaborate with users, whilst the latter can. This assumption is supported by real-world experience, in which the private cloud is seen as more reliable than the public cloud. We assume that data users will try to gain information beyond the scope of their permitted access. As

previously indicated, malicious outsiders may utilise duplicate phoney attacks in addition to attempting to steal plaintext data from the cloud. The system may determine that security and performance are critical for next-generation large-scale systems, such as clouds. As a result, we will use this project to address the issues of security and performance as a secure data replication challenge. User files are judicially split into bits and copied at significant cloud sites in the current approach, Division and Replication of Data in the Cloud for Optimal Performance and Security. A file is partitioned into fragments depending on a set of user criteria, with each fragment carrying no valuable information. Each cloud node (in this system, the term node refers to compute, storage, physical, and virtual machines) contains a unique fragment to improve data security.

II. Related Work:

Chun-Ho Ng et al. presented the RevDedup technique in 2013 to find and remove duplicates from virtual machine images. The RevDedup recognises a similarity with existing data when a new VM image is received and eliminates it from the old data [2]. In the same year, Mihir Bellare et al. devised a cryptographic approach called Message-Locked Encryption (MLE). MLE generates its encryption and decryption keys from the communication itself. It was the most secure deduplication method [3]. Zhou Lei et al. proposed a method for storing photos using the fixed size block method in 2014. This method generates a fingerprint directory by producing a succinct digest called a fingerprint for each image file. It generates fingerprints for new image input and compares them to a database of fingerprints [4]. In the same year, Waraporn Leesakul et al. proposed a new method that use dynamic data deduplication to boost cloud storage capacity efficiency. This strategy preserved redundancy while increasing storage space [5]. In the same year, Issa M. Khalil et al. discovered 28 cloud security vulnerabilities in their survey on cloud security issues and solutions [6]. In 2015, N. Jayapandian et al. proposed the authorization-based scheme. To protect user data confidentiality, this system uses differential privileges based on a duplicate check [7]. In the same year, Mi Wen et al. created a secure deduplication mechanism based on convergent encryption [8]. In the same year, Lakshmi Pritha et al. developed a system that employs RSS keys to provide secure access to cloud resources and showed the ALG data deduplication technique [9]. In the same year, Chun-I Fan et al. [10] presented a check block strategy for encrypted data deduplication. Mr. Dame Tirumala Babu and colleagues presented a data deduplication method based on authorization to secure data from the same year [11]. In 2016, Shuai Wang and colleagues proposed the RRMFS file system to help in data deduplication. [12]. The following year, Zheng Yan et al. proposed a mechanism for ownership and reencryption to deduplicate encrypted data stored in the cloud [13]. The destor tool was used by Naresh Kumar et al. in the same year to compare numerous deduplication algorithms. In the data deduplication strategy [14], fixed length and variable length chunking techniques are used. Jun Ren et al. published a safe data deduplication strategy based on differential privacy in the same year [15]. . The Load Balanced Flow Scheduling approach [17] was proposed by Feilong Tang et al. in the same year for dynamic load balancing and network performance maximization. The CSPD technique, proposed by Danoing Li et al. in 2017, uses a modified DCT-based Perceptual Image Hash (D-phash) to increase duplicate check accuracy [18]. In the same year, Hui Cui et al. developed an ABE encryption method for cloud storage based on characteristics [19]. In the same year, Rayan Dasoriya et al. [20] introduced a dynamic load balancing system that balanced the load across multiple connected network links. In the same year, Shunrong Jiang et al proposed a data secrecy and ownership management system based on Proof of Ownership (PoW) for data deduplication [21].

III. Proposed System:

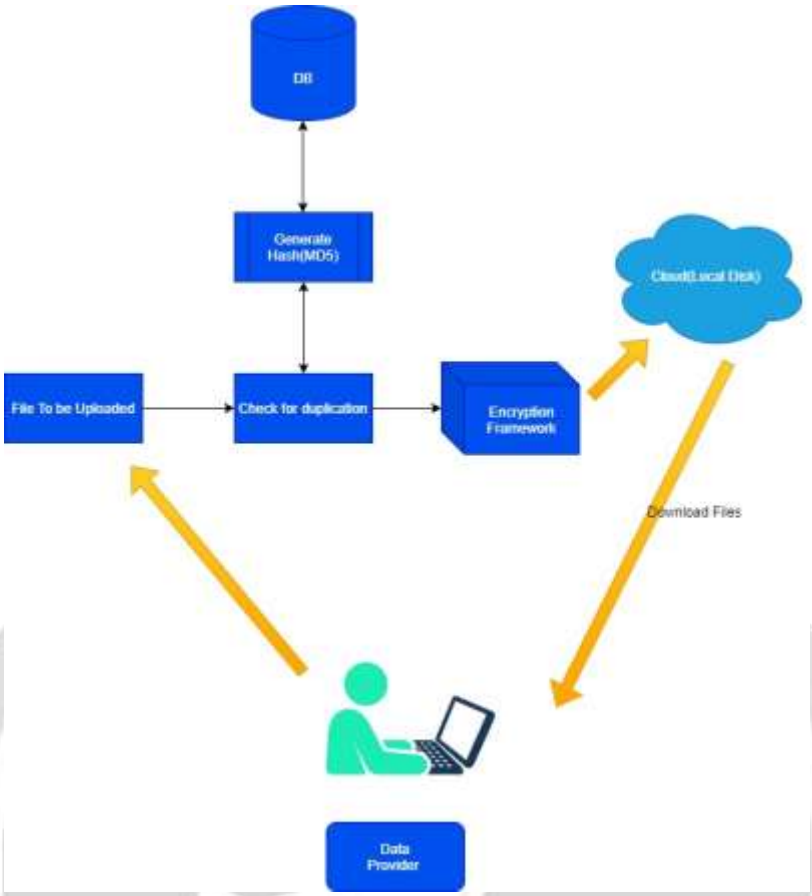

Fig: Proposed System

Secure de-duplication is supported by an attribute-based storage system. Our storage system is based on a hybrid cloud architecture, in which a private cloud controls compute and a public cloud controls storage. The cloud will not store a file more than once even if it receives several copies of the same file encrypted under various access permissions, thanks to an attribute-based storage system that supports secure de duplicating of encrypted data in the cloud. Every user is given a decryption key linked with the collection of characteristics by the Attribute Authority. The attribute-based storage system checks for file duplication. The file is not duplicated, and it is saved. The attribute authority modifies the ownership permission if duplication occurs.

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

IV. Results:

Fig: Registration Page

Fig: Login Page

Fig: File Uploading

V. Conclusion:

Thus we are going to develop a system for secure de duplication 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.