INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE V, MAY 22, WWW.IJCEA.COM, ISSN 2321-3469

APPLICATION OF MD5 FOR DE DUPLICATION IN CLOUD ENVIRONMENT

Prof. Smita Gumaste, Vipul Kargaonkar, Indranil oza, Shubham Kosaiker, Yash Khade

Dept of Computer Engineering, JSCOE, Savitribai Phule Pune University, Pune, Maharashtra, India.

ABSTRACT:

Computing resources are given as a utility on demand to consumers over the Internet, and cloud computing plays a significant role in the commercial domain today. Cloud storage is one of the services offered by cloud computing that has grown in popularity. Customers benefit the most from cloud storage since they can cut their expenditures on purchasing and maintaining storage equipment by simply paying for the amount of storage they need, which can be scaled up and down on demand. With cloud computing's expanding data size, a reduction in data quantities could assist providers in lowering the costs of running huge storage systems and conserving energy. As a result, data deduplication techniques have been implemented in cloud storage to improve storage efficiency. Because of the dynamic nature of data in cloud storage, data utilization in the cloud fluctuates over time. For example, some data chunks may be accessed often one time but not the next. Some datasets may be viewed or updated frequently by several users at the same time, while others may require a high amount of redundancy for stability. As a result, it's critical that cloud storage provide this dynamic functionality. Current techniques, on the other hand, are primarily focused on a static scheme, which limits their full applicability in the dynamic nature of data in cloud storage. We propose a dynamic deduplication strategy for cloud storage in this research, with the goal of increasing storage economy while maintaining redundancy for fault tolerance.

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

1. Introduction:

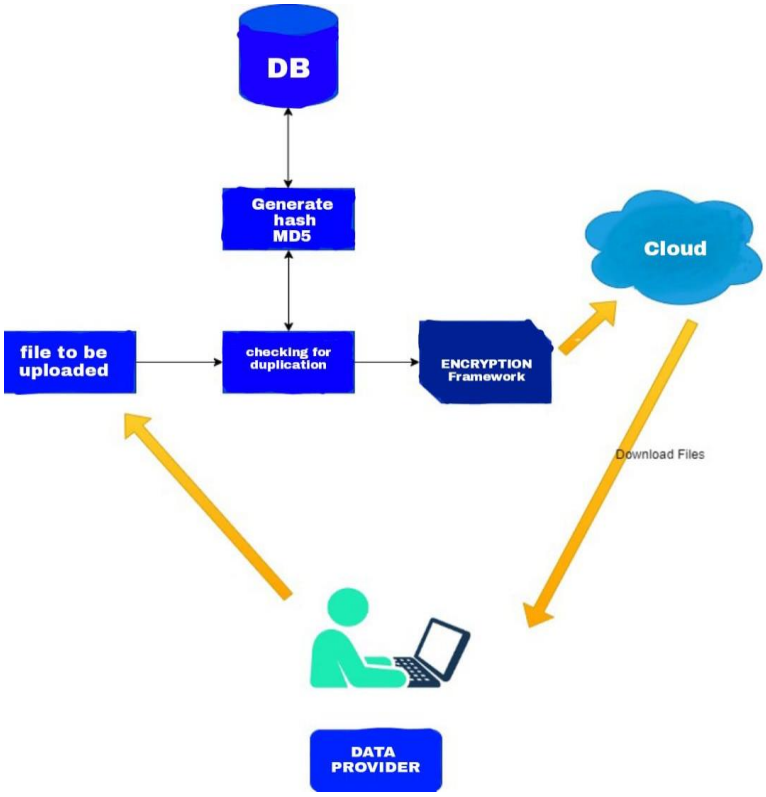
Secure deduplication, a technique for saving storage space and network bandwidth by removing redundant copies of encrypted data saved in the cloud, is not possible with the basic Attribute Based Encryption (ABE) system. Existing architectures for safe deduplication, on the other hand, are not, to our knowledge, based on attribute-based encryption. However, given the widespread use of ABE and safe deduplication in cloud computing, it would be ideal to build a cloud storage system that has both qualities. Consider the following scenario while designing an attribute-based storage system that supports secure deduplication of encrypted data in the cloud in this system, Even if it receives numerous copies of the same material encrypted under various access controls, the cloud will not store it more than once. However, providing the private cloud with such a tag checking capability is insufficient to perform deduplication in an attribute-based storage system that uses CP-ABE (Cipher Policy ABE) for data encryption. The same file could be encrypted to different cypher texts associated with different access policies in the proposed attributed-based system; however, storing only one cypher text of the file means that users whose attributes satisfy the access policy of a discarded cypher text will be denied access to the data they are entitled to. To solve this problem, we add a new feature to the private cloud called ciphertext regeneration. In terms of our storage system's adversarial model, we assume that the private cloud is curious but honest, in that it will try to obtain the encrypted messages but will follow 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 mis-behavior will be detected by either the private cloud or the user via the accompanied label). Another distinction between the private and public clouds is that the former cannot collaborate with users, whilst the latter may. This assumption corresponds to real-world reality, in which the private cloud is seen as more trustworthy than the public cloud. We presume that data users will attempt to obtain information beyond their allowed permissions. Malicious outsiders may use duplicate fake assaults, in addition to attempting to steal plaintext data from the cloud, as previously mentioned. The system may determine that for next-generation large-scale systems, such as clouds, both security and performance are crucial. As a result, we will address the problems of security and performance as a secure data replication challenge in this project. In the current approach, Division and Replication of Data in the Cloud for Optimal Performance and Security, user files are judicially fragmented into pieces and replicated at important cloud sites. The partition of a file into fragments is done based on a set of user criteria, with the individual fragments containing no useful information. To strengthen data security, each cloud node (in this system, the term node refers to computing, storage, physical, and virtual machines) carries a unique fragment.

2. Related Work:

The RevDedup technique was proposed by Chun-Ho Ng et al. in 2013 to locate and eliminate duplicates from virtual machine pictures. When a new VM image is received, the RevDedup

detects a similarity with existing data and removes it from the existing data [2]. Mihir Bellare et al. introduced a crypto graphic technique called Message-Locked Encryption in the same year (MLE). The encryption and decryption keys of MLE are generated from the message itself. It was the most secure method of deduplication [3]. In 2014, Zhou Lei et al. suggested a methodology for storing photos that used the fixed size block method. This approach creates a directory of fingerprints by calculating a compact digest termed fingerprint for each image file. It calculates fingerprints for new image input and compares them to a fingerprint library [4]. In the same year, Waraporn Leesakul et al. suggested a new system that uses dynamic data deduplication to increase the efficiency of cloud storage capacity. This method increased storage space while preserving redundancy [5]. Issa M. Khalil et al. found 28 cloud security vulnerabilities in their survey on security challenges in clouds and security solutions [6] in the same year. N. Jayapandian et al. introduced the authorization-based scheme in 2015. This system uses differential rights based on duplicate check to safeguard user data confidentiality [7]. Mi Wen et al. devised a secure deduplication strategy employing convergent encryption technique in the same year [8]. Lakshmi Pritha et al. built a system that uses RSS keys to offer safe access to cloud resources and demonstrated the ALG technique for data deduplication the same year [9]. Chun-I Fan et al. presented a check block approach for encrypted data deduplication the same year [10]. Mr. Dame Tirumala Babu et al. presented a solution for data deduplication based on authorization to protect data the same year [11]. Shuai Wang and colleagues suggested the RRMFS file system to facilitate data deduplication in 2016. [12]. Zheng Yan et al. introduced a strategy for ownership and re encryption to deduplicate encrypted data saved in the cloud the next year [13]. In the same year, Naresh Kumar et al. used the destor tool to do a comparative examination of several deduplication approaches. The fixed length and variable length chunking techniques are used in the data deduplication approach [14]. In the same year, Jun Ren et al. introduced a secure data deduplication approach based on differential privacy [15]. Saurabh Singh et al. published a cloud security survey in the same year, with a discussion of security issues and challenges [16]. In the same year, Feilong Tang et al. proposed the Load Balanced Flow Scheduling technique for dynamic load balancing and network performance maximization [17]. In 2017, Danoing Li et al. proposed the CSPD approach, which uses a modified DCT-based Perceptual Image Hash (D-phash) to improve duplicate check accuracy [18]. Hui Cui et al. built an ABE encryption system for cloud storage based on attributes in the same year [19]. Rayan Dasoriya et al. presented a dynamic load balancing technique in the same year [20], which distributed the load across various connected network links. In the same year, Shunrong Jiang et al suggested a Proof of Ownership (PoW)-based data secrecy and ownership management system for data deduplication [21].

3. Implementation System:



An attribute-based storage system supporting secure deduplication. 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 full information is revealed to the attacker.

4. Algorithm:

1. MD-5:
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. After padding, 64 bits are inserted at the end, which is used to record the original input length. The resulting message has a length multiple of 512 bits.
Step 3: Initialize MD Buffer. A four-word buffer is used to compute the values for the message digest.
Step 4: Process Message in 16-Word Blocks. MD5 uses the auxiliary functions, which take the input as three 32-bit numbers and produce 32-bit output. These functions use logical operators like OR, XOR, NOR.
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.

MD5 is primarily used to authenticate files.
MD4 was considered not secure because its hash calculation wasn't sufficiently complex. While MD4 hashes resemble MD5 hashes, there's a lot more going on behind the scenes with MD5, many more steps were added to the calculation to increase complexity.
MD5 calculates faster than SHA, making it a convenient solution for software vendors.

2. AES:
Step 1: Derive the set of round keys from the cipher key.
Step 2: Initialize the state array with the block data (plaintext).
Step 3:Add the initial round key to the starting state array.
Step 4: Perform ninth rounds of state manipulation.
Step 5: Perform the tenth and final round of state manipulation
Step 6: Copy the final state array out as the encrypted data (ciphertext).

	DES	AES
Developed	1977	2000
Key Length	56 bits	128, 192, or 256 bits
Cipher Type	Symmetric block cipher	Symmetric block cipher
Block Size	64 bits	128 bits
Security	Proven inadequate	Considered secure

5. Results:

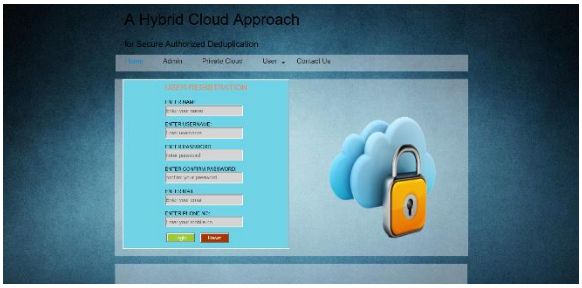


Fig: Registration

Fig: Login

Fig: Upload File

Fig: Download File

6. Future Scope:

Deduplication is increasingly popular in primary storage systems with the rapid growth of application data. Employing deduplication to improve I/O performance will play a more important role in high-performance storage systems than saving storage space. Read and delete operations frequently occur in primary storage systems. Addressing these problems with variable primary storage workloads in systems with different storage devices, such as, disks, DRAM memory, non volatile memory devices, will be an interesting future research direction.

• Emerging applications. Deduplication can benefit applications besides disk storage such as employing deduplication to extend the lifetime of SSDs and PCMs and eliminating visual redundancy for images and videos. We believe that there will be more applications for deduplication, such as storage systems for tapes or shingled disks, since it will help reduce the growing redundant data in large-scale storage systems.

7. 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.

8. 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.