ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

Survey on ABS Supporting Secure Deduplication of Encrypted Data in Cloud

Priyanka Bhosale, Prof. S. H. Patil

PG Students at Dept. of Computer Engineering, JSCOE Hadapsar, Pune, Maharashtra, India

Asst. Professor at Dept. of Computer Engineering, JSCOE Hadapsar, Pune, Maharashtra, India

ABSTRACT: In public cloud storage system protecting the data and controlling the data access is a challenging issue. Cipher text Policy Attribute-Based Encryption (CP-ABE) has been adopted as a promising technique to provide flexible, fine-grained and secure data access control for cloud storage with honest-but-curious cloud servers. However numerous works have been proposed using CP-ABE scheme, in which the single attribute authority must execute the time-consuming user legitimacy verification and secret key distribution and hence it results in a single-point performance bottleneck when a CP-ABE scheme is adopted in a large-scale cloud storage system. Clients may be stuck in the waiting line for a long stretch to get their mystery keys, which results in low-efficiency of the framework. Even though the multi authority access control plans have been proposed, these plans still cannot conquer the disadvantages of single-point bottleneck and low efficiency; because of the way that each of the authority still autonomously deals with a disjoint characteristic set. In order to overcome this disadvantage, there has been proposed a novel heterogeneous framework to remove the problem of single point performance bottleneck and provide a more efficient access control scheme with an auditing mechanism. Meanwhile, in this scheme, a CA (Central Authority) is introduced to generate secret keys for legitimacy verified users and each of the authorities in this scheme manages the whole attribute set individually. This system makes performance improvement in key generation and also guarantees security requirement. Still there are some security loopholes in this system such as there is no protocol to verify owner and If the owner is compromised then he/she may put wrong data or information in the data server and users may get wrong data. There is no way to know who has used the data.CA who generates secret keys, is assumed to be fully trusted ,If CA gets compromised he can collude with any user or AA to provide secret keys to illegitimate users. In order to overcome these disadvantages, we are going to propose a new framework where there will be a protocol to verify owner and his/her data to be uploaded and a log will be maintained to know who will be accessing the data. This framework will also propose to choose one among the AAs to act as CA instead of separate CA and will have an observer to trace if CA is working properly or not. If observer finds any discrepancy it will be creating a report. This will make the system more secure and efficient.

KEYWORDS*:* ABE, De-dupllication, storage.

I. INTRODUCTION

The most important and popular cloud service is data storage service. Cloud users upload personal or confidential data to the data center of a Cloud Service Provider (CSP) and allow it to maintain these data. Cloud computing schemes presented focus on warehoused data, where the outsourced data is kept unchanged over remote servers In cloud data storage system, users store their data in the cloud and longer possess the data locally. Thus, the correctness and availability of the multiple data files being on the distributed cloud servers must be guaranteed. In existing system, brute-force attack used to avoid multiple copies of dynamic When verifying multiple data copies, the overall system integrity check fails if there are on corrupted copies. This project Rewriting (HAR) algorithm to avoid de encrypted data stored in cloud based on ownership challenge and proxy re-encryption. It integrates cloud data deduplication with access control. scheme based on data owner-ship challenge and Proxy Re-Encryption (PRE) to manage encrypted data storage with deduplication.aim to solve the issue of deduplication in the situation where the data holder is not available or difficult to get involv method using double encryption key for encrypted data stored in cloud. First the data owner

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

provide the secret key to data user then authorized party (AP) send the secret key to data owner. Both AP key and private key generate the encrypted key for encryption. AES algorithm using the encrypt content stored in cloud.

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. We 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. We can deduce that both security and performance are critical for the next generation large-scale systems, such as clouds. There-fore, in this project, we collectively approach the issue of security and performance as a secure data replication problem. We 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 (we use the term node to represent computing, storage, physical, and virtual machines) contains a distinct fragment to increase the data security.

II. LITERATURE SURVEY

In this Paper, Nowadays regularly use cloud services in our daily life. There are various services provided by cloud such as a service, Platform as a service, and Infrastructure asa service. The used to keep our data,documents, and files on cloud. The data that store may be Personal, Private, secret data. So must be very sure that whatever the cloud service we use that must be secure. Cloud computing Provides number of services to client over internet. Storage service isone ofthe important services that people used now days for storing data on network so that they can access their data from anywhere and anytime. With the benefit of storage service there is an issue of security. To overcome security problem the proposed system contain two levels of security and to reduce the unwanted storage space de-duplication[1,2] technique is involved. To increase the level of security one technique is a session password.Session passwords can be used only once and every time a new password is generated.To protect the confidentiality of sensitive data while supporting de-duplication[1,2]the convergent encryption technique has been proposed to encrypt the data before outsourcing, Symmetrickey algorithm uses same key for both encryption and decryption. In this paper [4], I will focus on session based authentication for both encryptions for files and duplication check for reduce space of storage on cloud Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. Big Data in most companies are processed by Hadoop by submitting the jobs to Master. The Master distributes the job to its cluster and process map and reduces tasks sequentially. But nowadays the growing data need and the competition between Service Providers leads to the increased submission of jobs to the Master. This Concurrent job submission on Hadoop forces us to do Scheduling on

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

Hadoop Cluster so that the response time will be acceptable for each job. In this Deduplication techniques are most widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. So which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth? We present an attribute-based storage system with secure deduplication in a hybrid cloud setting, using public cloud and private cloud. Where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Instead of keeping multiple data copies with the same content, the system eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Each such copy can be defined based on user access policies. In this user will upload the file with access policies and then file type question with answer [8]. Then same file with different access policies to set the particular file to replace the reference. Where a user's private key is associated with an attribute set, a message is encrypted under an access policy over a set of attributes, and a user can decrypt a ciphertext with his/her private key if his/her set of attributes satisfies the access policy associated with this ciphertex.

III. PROBLEM STATEMENT

An inherent drawback of the existing approaches to achieve secure de-duplication is that they cannot satisfy the standard security definition for confidentiality such as semantic security. To solve this problem, a weaker security notion called privacy under chosen-distribution attacks was put forward under the assumption that the input message is sufficiently unpredictable. We present Division and Replication of Data in the Cloud for Optimal Performance and Security (DROPS) that judicially fragments user files into pieces [10] and replicates them at strategic locations within the cloud.

IV. SYSTEM ARCHITECTURE

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. We are utilizing client accreditation's 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. We have utilized a half and half cloud construction modeling as a part of proposed. We 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

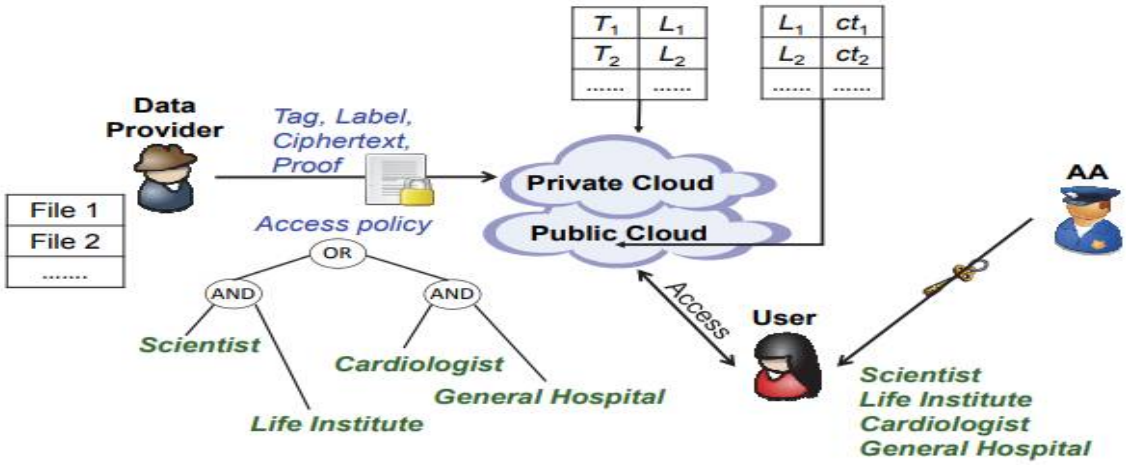

Figure 1. System architecture

V. ALGORITHM

Setup, Encrypt, Key Generation (KeyGen), and Decrypt.

1. Setup(1,λ) → (pars, msk). Taking the security parameter λ as the input, this setup algorithm outputs the public parameter pars and the master private key msk for the system.

2. Encrypt(pars, M, A)→ (skT , CT). Taking the public parameter pars, a message M and an access structure A over the universe of attributes as the input, this encryption algorithm outputs a trapdoor key skT and a tuple CT = (T, L, ct, pf), where T and L are the tag and the label associated with M respectively, ct is the ciphertext which includes the encryption of M as well as the access structure A, and pf is a proof on the relationship of tag T, label L and ciphertext ct. This algorithm is run by the data provider. Both skT and CT are forwarded to the private cloud. Note that skT can not be disclosed to any third party, so it must be sent to the private cloud in a secure manner.

KeyGen(pars, msk, A) → skA. Taking the public parameter pars, the master private key msk and an attribute set A as the input, this attribute-based private key generation algorithm generates an attribute based private key skA for the attribute set A.

4. Validity-Test(pars, CT)→ 1/0. Taking the public parameter pars and a tuple CT as the input, this validity testing algorithm parses CT as (T, L, ct, pf), and outputs 1 if pf is a valid proof for (T, L, ct) or 0 otherwise.

5. Equality-Test(pars, (T1, L1, ct1), (T2, L2, ct2)) → 1/0. Taking the public parameter pars and two tuples (T1, L1, ct1) and (T2, L2, ct2) as the input, this equality testing algorithm outputs 1 if both (T1, L1, ct1), (T2, L2, ct2) are generated from the same underlying message or 0 otherwise.

VI. CONCLUSION

In this system provided reason that our proposed framework information DE duplication of record is done approves way and safely. In this we have additionally proposed new duplication check system which produce the token for the private document. A proposed routine guarantees the information duplication safely. Future work includes efficient

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 12, December 2018

data ownership verification, scheme optimization with hardware acceleration at IoT devices for practical deployment, and development of a flexible solution to support deduplication and data access controlled by either the data owner or its representative agent.

REFERENCES

1. D. Quick, B. Martini, and K. R. Choo, Cloud Storage Forensics. Syngress Publishing / Elsevier, 2014. [Online]. Available: http://www.elsevier.com/books/cloud-storageforensics/ quick/978-0-12- 419970-5
2. K. R. Choo, J. Domingo-Ferrer, and L. Zhang, "Cloud cryptography: Theory, practice and future research directions," Future Generation Comp. Syst., vol. 62, pp. 51–53, 2016.
3. K. R. Choo, M. Herman, M. Iorga, and B. Martini, "Cloud forensics: State-ofthe-art and future directions," Digital Investigation, vol. 18, pp. 77–78, 2016.
4. Y. Yang, H. Zhu, H. Lu, J.Weng, Y. Zhang, and K. R. Choo, "Cloud based data sharing with fine-grained proxy re-encryption," Pervasive and Mobile Computing, vol. 28, pp. 122–134, 2016.
5. D. Quick and K. R. Choo, "Google drive: Forensic analysis of data remnants," J. Network and Computer Applications, vol. 40, pp. 179–193, 2014.
6. M. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," In Proc.of NDSS'12, 2012.
7. C. Wang, K. Ren, S. C. Yu, and K. M. R. Urs, "Achieving Usable and Privacy-assured Similarity Search over Outsourced Cloud Data," in Proc. of IEEE INFOCOM 2012, 2012, pp. 451-459.
8. N. Cao, C. Wang, M. Li, K. Ren. W. J. Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data," in Proc. of IEEE INFOCOM 2011, 2011, pp. 829-837
9. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. j. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.
10. C. Liu, L. H. Zhu, L. Li, and Y. Tan, "Fuzzy Keyword Search on Encrypted Cloud Storage Data with Small Index," in Proc. of IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), 2011, pp. 269-273.
11. C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," in Proc. of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), 2010, pp. 253-262.