ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Secured Deduplication of Encrypted Data in Cloud

Ch. Navya
Department of Computer Science & Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana 500075, India.

Dr. K. Jaya Shankar
Department of Computer Science & Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana 500075, India.

Dr. D. Vijaya Lakshmi
Department of Computer Science & Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana 500075, India.

Abstract

Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext.

INTRODUCTION

Cloud computing greatly facilitates data providers who want to outsource their data to the cloud without disclosing their sensitive data to external parties and would like users with certain credentials to be able to access the data. This requires data to be stored in encrypted forms with access control policies such that no one except users with attributes (or credentials) of specific forms can decrypt the encrypted data. An encryption technique that meets this requirement is called attribute-based encryption (ABE), where a user's private key is associated with an attribute set, a message is encrypted under an access policy (or access structure) over a set of attributes, and a user can decrypt a cipher text with his/her private key if his/her set of attributes satisfies the access policy associated with this cipher text. However, 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 paper, we present an attribute-based storage system which employs cipher text-policy attribute-based encryption (CP-ABE) and supports secure deduplication. Our main contributions can be summarized as follows. Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture.

Secondly, we put forth a methodology to modify a cipher text over one access policy into cipher texts of the

Cite this article as: Ch. Navya, Dr. K. Jaya Shankar & Dr. D. Vijaya Lakshmi, " Secured Deduplication of Encrypted Data in Cloud", International Journal & Magazine of Engineering, Technology, Management and Research, Volume 6 Issue 5, 2019, Page 35-41.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

same plaintext but under any other access policies without revealing the underlying plaintext. This technique might be of independent interest in addition to the application in the proposed storage system.

Thirdly, we propose an approach based on two cryptographic primitives, including a zero-knowledge proof of knowledge and a commitment scheme, to achieve data consistency in the system

Motivation

Recently, with the rapid development of network storage technology, cloud storage has become an important storage scheme. Owing to the rental cost lowness, outsourcing files of an enterprise to cloud storage can reduce its enterprise management costs and improve its competitiveness. To prevent files from information leakage, an enterprise user usually stores its files to cloud storage in an encrypted form. Encrypted file deduplication scheme can save its storage space and network bandwidth of cloud storage and improve its performance. However, in the enterprise application environment, different department employees have different permissions. Each employee can only access the files according to its permission. If an encrypted file deduplication scheme does not support permission checking, it will destroy the file permissions and bring some security problems. Li et al. proposed a secure authorized deduplication scheme based on a hybrid cloud (SADS) . They introduce a private cloud in SADS to preserve the user permissions and generate a permission tag for a user when it uploads a file. When the cloud storage performs the deduplication checking for a user, it needs to check the deduplication permission for the user, and if the user does not have the deduplication permission, the user needs to upload the file even though there exists the same file in the cloud storage. Only when the user has the deduplication permission and there exists the same file in the cloud storage can the cloud storage perform file deduplication.

The use of SADS can achieve the encrypted file deduplication, but there exist three shortcomings in SADS:

- Firstly, each permission is represented by a private key. If a user has multiple permissions, it needs to store multiple private keys secretly which can cause a great deal of trouble in the user key management.
- Secondly, when uploads a file or queries the duplication file of , the scheme needs to use permission keys to generate encrypted file tags for (If has been assigned permissions). So the scheme causes large network traffic.
- Thirdly, there exists a security weakness in SADS. Assuming Mike is an enterprise manager who manages department and department. Mike has the permissions of department and department.

At the same time, Mike is responsible for the financial department, so he also has the finance department permission. If cloud storage uses SADS to reduplicate the files in the cloud storage, SADS uses the private keys of department, department, and the finance department to generate three encrypted file tags. As a result, the staffs in department and department have the permission to deduplicate their files with the pay slip file. Suppose Mike has uploaded Alice's pay slip file to the cloud storage, if both Bob and Alice are employees of department. Bob wants to get the salary information of Alice. He can use the following steps (called online deduplication oracle attack) to attack SADS to obtain the salary information of Alice : (a) Bob first forges Alice's pay slip file . is a kind of small entropy file and it has a fixed format. Bob knows the file format or he even has the kind of file, i.e., he has his own pay slip. At the same time, he also knows that Alice's salary should be between and he just does not know the concrete salary value of Alice. So Bob can set the salary value to, respectively, and generate files.
(b) Bob uploads to the cloud storage, respectively. If the cloud storage deduplicates the file when he uploads a file to the cloud storage, Bob knows that the salary of Alice is the data in the uploaded file.

Obviously, the success reason for the attack is the authorization precision of SADS which is rough. When

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Mike generates an encrypted file tag, it has assigned the file deduplication permission to Bob and causes the file permission checking bypass. At the same time, when the cloud storage checks the file deduplication, it only checks whether the encrypted file query tag of the upload file matches the encrypted file tags stored in the cloud storage owner and does not check the user's permission. Therefore, we want to design a securely encrypted file deduplication scheme with permission to improve the file deduplication permission check of the user and avoid the security issues of SADS.

LITERATURE SURVEY

A Survey on Deduplication Strategies and Storage Systems

Now a day there is raising demands for systems which provide the data storage by keeping in mind security and cost factor. A duplicate file not only occupies a high volume space but also increases access time so that removing duplicates records become very necessary. But to achieve this is not that simple since neither duplicate files don't have a common key nor they contain error which makes duplicate matching a tedious task. There are different ways to eliminate duplicate data first at file level & then at chunk levels that reduces duplicate lookup overhead. In this survey paper we have discussed and describes about some of the deduplication strategies and some of storage systems like MAD2, Venti, HYDRAstor, Extreme Binning, Duplicate Data Elimination (DDE)

Message-Locked Encryption and Secure Deduplication:

We formalize a new cryptographic primitive, Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud-storage providers. We provide definitions both for privacy and for a form of integrity that we call tag consistency. Based on this foundation, we make both practical and theoretical contributions. On the practical side, we provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and we make connections with deterministic encryption, hash functions secure on correlated inputs and the sample-then-extract paradigm to deliver schemes under different assumptions and for different classes of message sources. Our work shows that MLE is a primitive of both practical and theoretical interest.

SYSTEM ANALYSIS:

Existing System:

When a user uploads data that already exist in the cloud storage, the user should be deterred from accessing the data that were stored before he obtained the ownership by uploading it (backward secrecy)2. These dynamic ownership changes may occur very frequently in a practical cloud system, and thus, it should be properly managed in order to avoid the security degradation of the cloud service. In the former approach, most of the existing schemes have been proposed in order to perform a PoW process in an efficient and robust manner, since the hash of the file, which is treated as a "proof" for the entire file, is vulnerable to being leaked to outside adversaries because of its relatively small size. a data owner uploads data that do not already exist in the cloud storage, he is called an initial uploaded; if the data already exist, called a subsequent uploaded since this implies that other owners may have uploaded the same data previously, he is called a subsequent uploaded.

Disadvantages:

- User deduplication on the client-side, cannot generate a new tag when they update the file. In this situation, the dynamic Ownerships would fail. As a summary, existing dynamic Ownerships cannot be extended to the multi-user environment. Whenever data is transformed, concerns arise about potential loss of data.
- By definition, data deduplication systems store data differently from how it was written. As a result, users are concerned with the integrity of their data.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

- One method for deduplicating data relies on the use of cryptographic hash functions to identify duplicate segments of data. If two different pieces of information generate the same hash value, this is known as a collision.
- The probability of a collision depends upon the hash function used, and although the probabilities are small, they are always non zero.

Proposed System:

This Project the goal of saving storage space for cloud storage services also is used for secure deduplication .but several process have been this same concept for deduplication. However this project flow some different modules in there. In this case, if two users upload the same file, the cloud server can discern the equal cipher texts and store. Only one copy of them. This process some authentication available in some issue for security purpose. Through this process for ensure secured deduplication. An owner wants to outsource data to the cloud and share it with users possessing certain credentials.

The Attribute Authority issues every user a decryption key associated with users set of attributes. Which is considered to be the most important challenge for efficient and secure cloud storage services in the environment where ownership changes dynamically? Every time data provider uploads file checking from cloud for save storage purpose. Most of the schemes have been proposed to provide data encryption, while still benefiting from a deduplication technique. Every user get secured key form admin for security purpose .user cannot take any key he cannot download chipper text file .they can download only encrypted data. Every detail manages and maintain by Attribute authority.

In this way, any user who downloads the file, after decryption, can check the correctness of the decrypted plaintext by matching it to the corresponding tag. To keep the notation succinct, we use c to denote the combination of the encrypted data and the corresponding access structure.

Advantages:

System has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a cipher text over one access policy into cipher texts of the same plaintext but under other access policies without revealing the underlying plaintext.

MODULES:

Data Provider:

Data provider uploading file to cloud with tag , label and security key , the proposed scheme guarantees data integrity against any tag inconsistency attack. Thus, security is enhanced in the proposed scheme.

Cloud Storage:

Secure Deduplication with the goal of saving storage space or cloud storage services, Douceur et al the first solution for balancing confidentiality and efficiency in performing deduplication called convergent encryption, where a message is encrypted under a message-derived key so that identical plaintexts are encrypted to the same ciphertexts. In this case, if two users upload the same file, the cloud server can discern the equal ciphertexts and store only one copy of them. which may violate the privacy of the data if the cloud server cannot be fully trusted . This is a client who owns data, and wishes to upload it into the cloud storage to save costs. A data owner encrypts the data and outsources it to the cloud storage with its index information, that is, a tag.

Deduplication:

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. This technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. Deduplication techniques take advantage of data similarity to identify the same data and reduce the storage space.

In contrast, encryption algorithms randomize the encrypted files in order to make ciphertext indistinguishable from theoretically random data.

Attribute Authority:

The AA issues every user a decryption key associated with user set of attributes At the user side, each user can download an item, and decrypt the ciphertext with the attribute-based private key generated by the AA if this user's attribute set satisfies the access structure.

SYSTEM DESIGN:

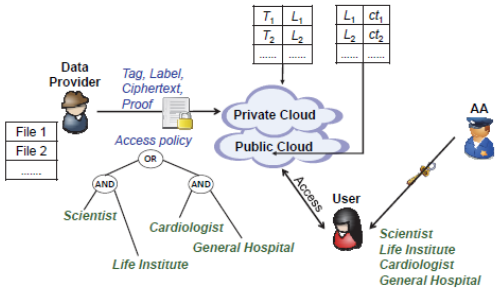

Fig 4.1: Architecture Diagram

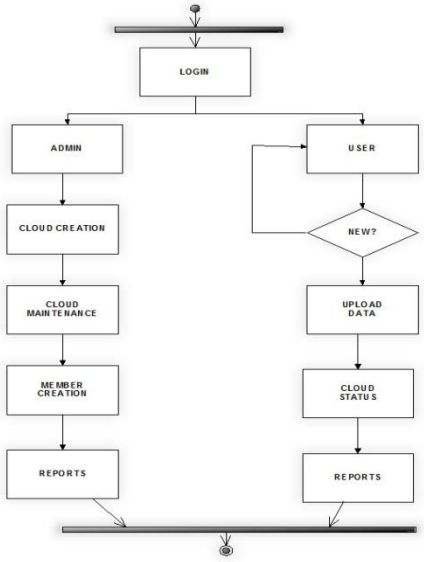

Fig.4.2. Activity Diagram

OUTPUT RESULTS:

Fig 5.1: Home Page

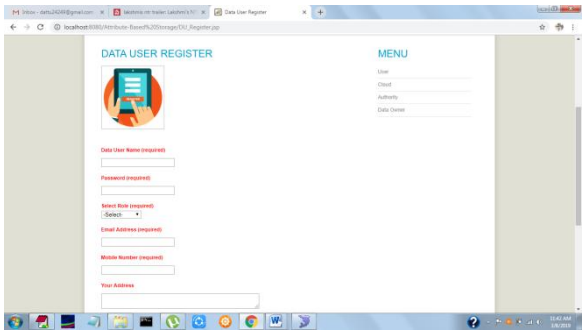

Fig 5.2: User registration Page

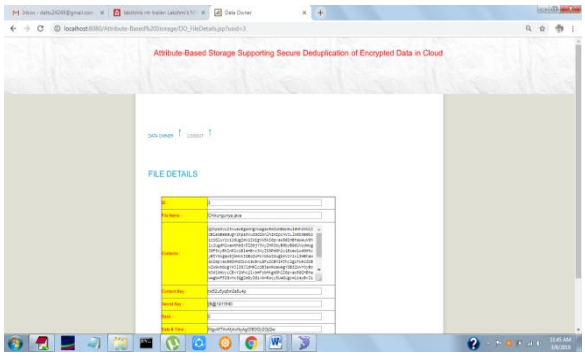

Fig 5.3: **File Details page**

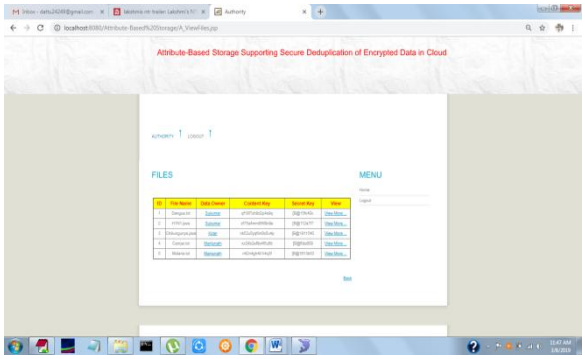

Fig 5.4: **File Upload details Page**

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

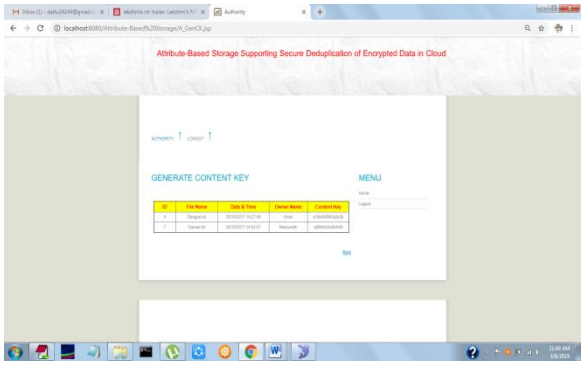

Fig 5.5: **Generate content key Page**

CONCLUSION

Attribute-based encryption (ABE) has been widely used in cloud computing where data providers outsource their encrypted data to the cloud and can share the data with users possessing specified c credentials. On the o there hand, deduplication is an important technique to save the storage space and network bandwidth, which eliminates duplicate copies of identical data. However, the standard ABE systems do not support secure deduplication, which makes them costly to be applied in some commercial storage services. In this paper, we presented a novel approach to realize 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. The private cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it can transfer the ciphertext over one access policy into ciphertexts of the same plaintext under any other access policies without being aware of the underlying plaintext. After receiving a storage request, the private cloud first checks the validity of the uploaded item through the attached proof. If the proof is valid, the private cloud runs a tag matching algorithm to see whether the same data underlying the ciphertext has been stored. If so, whenever it is necessary, it regenerates the ciphertext into a ciphertext of the same plaintext over an access policy which is the union set of both access policies. The proposed storage system enjoys two major advantages. Firstly, it can be used to confidentially share data with other users by specifying an access policy rather than sharing the decryption key. Secondly, it achieves the standard notion of semantic security while existing deduplication schemes only achieve it under a weaker security notion.

FUTURE ENHANCEMENT

We intent to propose time period key not to be based on operations instead strongly recommend generating time period key based on logging.

We propose fully homomorphic encryption auditing system for the data storage security in the cloud computing. We have integrated the fully homomorphic encryption in TPA auditing system. Proposed scheme of fully homomorphic provides auditing technique which not only preserves privacy but also provide authenticator that allow an unbounded number of verification. In future we intend to verify its approach practically and to add experimental analysis to derive some conclusion from it

As this complete paper narrates the different methodologies on enabling cloud storage auditing with key exposure resilience, but none of the methodologies seems to be perfect. So, this survey paper as a bit proposes a method of an effective key exposure resistance where we adopt the deduplication strategy of data. Moreover, it will check the duplicity of data and eliminate the redundant one using MD5 hashing algorithm. After the public and private keys are generated, it uses tile bitmap method wherein it will check the previous and the current versions of the data to ease the auditor's workload and to make the system more efficient.

The auditing performed by public verifier not only audits the data but also verifies the integrity of the data in cloud. The concept of user revocation allows revoking the invalid key registered. We formalize the definition and the security model of auditing protocol without key-exposure resilience, and then propose and verify the first practical solution. Further the duplicated files are prohibited but do not address the issues due to creation of such files. In future we need to identify the solution

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

for providing privacy to data that is not verified in public cloud.

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-of-the-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] A. Sahai and B. Waters, "Fuzzy identity-based encryption," in Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, ser. Lecture Notes in Computer Science, vol. 3494. Springer, 2005, pp. 457–473.

[7] B. Zhu, K. Li, and R. H. Patterson, "Avoiding the disk bottleneck in the data domain deduplication file system," in 6th USENIX Conference on File and Storage Technologies, FAST 2008, February 26- 29, 2008, San Jose, CA, USA. USENIX, 2008, pp. 269–282.

[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings, ser. Lecture Notes in Computer Science, vol. 7881. Springer, 2013, pp. 296–312.

[9] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lock-dependent messages," in Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I, ser. Lecture Notes in Computer Science, vol. 8042. Springer, 2013, pp. 374–391.

[10] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013. USENIX Association, 2013, pp. 179–194.