DEDUPLICATION IN CLOUD COMPUTING AND SECURE DATA ACCESS CONTROL

Shruthi P
Department of Computer Science And Engineering Malabar Institute of Technology ,Anjarakandy Kannur , Kerala -India
Email: spanand.nambiar@gmail.com

Akhil K K
Department of Computer Science And Engineering Malabar Institute of Technology ,Anjarakandy Kannur , Kerala - India
Email: akhilkk1@gmail.com

ABSTRACT

Deduplication is the process that eliminates the duplicated copies of data stored at cloud. The advantages of deduplication is high space and cost savings. In this paper, Data owner is only active at the time of deduplication. Here, block level deduplication can be performed at CSP.

Data owner wants to control not only data access but also its storage and usage. After deduplication data can be access securely from the cloud. The same data, although in an encrypted form, is only saved once at the cloud but can be accessed by different users based on the data owners' policies. Here, Elliptic Curve Digital Signature Algorithm (ECDS) can be used to maintain the data integrity of the data at the cloud. This algorithm can be executed by the third party who is trusted, it is an external entity related to the cloud. Access control is a security mechanism is to provide outsource able data security..

Index Terms- *Deduplication, Data Integrity, Cloud Computing*

1 INTRODUCTION

Cloud is the distributed computing systems that has several advantages are computing as utility, virtualization of resources, on demand access to computing resources, and outsourcing computing services. Cloud has an important platform in IT industry. Various cloud service providers (CSPs) has huge volumes of storage to maintain and manage the IoT data. These CSPs has service properties, such as scalability, elasticity, fault tolerance, and pay per use.

Deduplication is the process that eliminates the duplicated copies of data stored at cloud. The advantages of deduplication is high space and cost savings. In this paper, Data owner is only active at the time of deduplication. Here, block level deduplication can be performed at CSP. Data owner wants to control not only data access but also its storage and usage. After deduplication data can be access securely from the cloud. The same data, in an encrypted form, is saved once at the cloud. It can be accessed by different users based on the data owners' policies. Here, Elliptic Curve Digital Signature Algorithm (ECDS) can be used to maintain the data integrity of the data at the cloud. This algorithm can be executed by the third party who is trusted, it is an external entity related to the cloud. Access control is a security mechanism is to provide outsource able data security.

2 EXISTING SYSTEM

The attribute based encryption (ABE) scheme is used to deduplicate encrypted data stored in the cloud while at the same time supporting secure data access control. Data's hash code is used to check the deduplication. This system containing three types of entities: CSP, Data Owner, and Data holder. Data holders that are data users and could save the same data as the data owner at the CSP. The data owner has the highest priority for data storage management. In the existing sytsem, there is no method is employed to ensure the data integrity. Data Integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle. It refers to the overall completeness, accuracy and consistency of data. This can be indicated by the absence of alteration between two instances or between two updates of a data record, meaning data is intact and unchanged. Cryptographic hash functions can be used to
ensure the data integrity. The existing system is not fully secured. The data can be obtained by the attackers.

3 PROPOSED WORK

The proposed work uses the following methods and algo- rithms:

3.1 METHOD USED

(a)ABE-Algorithm

1: Input all the personal details of the user for registration.
2: Then goto login page.
3: If the user already exist, then directly goto login page.
4: else goto step 1
5: User can upload any type of files.
6: Read file as a byte array.
7: This byte array can be converted into base 64 string.
8: Apply attribute based encryption on the string.
9: Encrypted form stored as a byte array to the CSP.
10: Hash value of the data can also be calculated and stored to the CSP.
11: Deduplication can be performed at CSP based on the data owner'policies.

12: Encryption and decryption can be performed by using symmetric key.
13: Decrption can be performed as in same case of encryption.
14: User can download the original data from the CSP.

(b)ECDS-Algorithm

1: Elliptic Curve Digital Signature Algorithm (ECDS) can be used to provide
the data integrity at the cloud.
2: This algorithm can be executed by the trusted third party.
3: ECDS consists of two steps: Signature generation algorithm, Signature
Verification algorithm.
4:Signature generation Algorithm is as follows：

5: signature generation algorithm creates a key pair.
6: Private key and public key.
7:User digitally signed the data by using this private key.

8: This signature is the pair of integers.
9: The verification algorithm take place at TTP by using public key.
10: This public key can be sent by the user. TTP is an external entity who is trusted.
11 After verification, The original data was obtained and sent to the cloud.
12: If data is inconsistent, user will be notified via email.

3.2 SYSTEM ARCHITECTURE

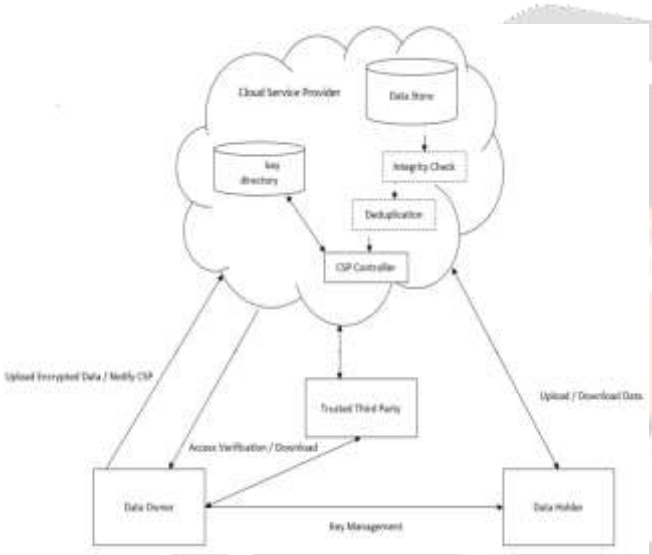

Fig:System-Architecture

The System architecture consists of 4 types of entities: Data Owner, Data holder, CSP and Trusted third party. In this system, Data owner has the highest priority of storage management. It should be active in the case of deduplication only. Data Holders are the eligible users who has upload/ download the data to and from the cloud. A CSP has a storage service and performs honestly on data storage and management to gain commercial profit but can't be fully trusted since it's curious about the contents of stored data. TTP is the trusted third party which is an external entity who is fully trusted. It used to provide the data integrity of the data in the cloud. ECDS (Elliptic Curve Digital Signature Algorithm) can be used to ensure the data integrity at the cloud. TTP executes this algorithm. Deduplication can be performed at CSP. Before uploading the data into CSP, it must be encrypted and hash code also be calculated. Hash code is used to check the deduplication.

4 CONCLUSION

Deduplication can be successfully performed at CSP based on data owner's policies. Deduplication can achieve high space and cost savings. ECDS Algorithm can be used for ensuring data integrity. There a trusted third party who is fully trusted and executed this algorithm. After checking the data integrity, users can access secured data from the CSP.

REFERENCES

[1] Zheng Yan, Mingjun Wang, Yuxiang Li, and Athanasios V Vasilakos.
Encrypted data management with deduplication in cloud computing.
IEEE Cloud Computing, 3(2):28–35, 2016.

[2] Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang. Secure data deduplication with dynamic ownership management in cloud storage.
[3] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Messagelocked encryption and secure deduplication. In Annual International Conference on the Theory and Applications of Cryptographic Techniques, pages 296–312. Springer, 2013.
[4] Yan Zhu, Hongxin Hu, Gail-Joon Ahn, and Mengyang Yu. Cooperativeprovable data possession for integrity verification in multicloud storage. IEEE transactions on parallel and distributed systems, 23(12):2231–2244, 2012.
[5] Kruti Sharma and Kavita R Singh. Seed block algorithm: A remote smart data back-up technique for cloud computing. In Communication Systems and Network Technologies (CSNT), 2013 International Conference on, pages 376–380. IEEE, 2013.
[6] Shweta Lamba and Monika Sharma. An e_cient elliptic curve digital signature algorithm (ecdsa). In Machine Intelligence and Research Advancement (ICMIRA), 2013 International Conference on, pages 179–183. IEEE, 2013.


IJARIIE