International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 12, Number 2 (2016), pp. 203-207
© Research India Publications
http://www.ripublication.com

Secure Deduplication in Cloud Computing

Shruthi P. and Indulekha K.

[1,2] Malabar Institue of Technology,
APJ Abdul Kalam Technological University, Kerala
E-mail: spanand.nambiar@gmail.com, indusreesan@gmail.com

Abstract

To ensure data security, existing research proposes to outsource only encrypted data to CSPs. The same or different users could save duplicated data by using different encryption schemes at the cloud. Although cloud storage space is big, duplication wastes networking resources, consumes excess power, and complicates data management. The advantage of Deduplication can achieve high space and cost savings. The data owners want CSPs to protect their personal data from an unauthorized access. CSPs should perform access control based on the data owners management. The data owners want to control not only data access but also its storage. Data deduplication should cooperate with data access control mechanisms. That is, the same data, in an encrypted form, is only saved once at the cloud but can be accessed by different users based on the data owners policies.To provide data integrity, ECDS Algorithm can be used with the help of Trusted Third Party. This system will provide more security to the data stored in the cloud.

Keywords: Deduplication, Data integrity, Cloud Computing, Data Access Control.

INTRODUCTION

Clouds are distributed computing systems built around core concepts such as computing as utility, virtualization of resources, on demand access to computing resources, and outsourcing computing services. Due to these concepts the clouds as

an attractive platform for businesses enabling them to outsource some of their IT operations. Cloud computing system has an ability to provide on demand access to always-on computing utilities..cloud computing offers a new way to deliver services by rearranging resources over the Internet and providing them to users on demand. It plays an important role in supporting data storage, processing, and management in the Internet of Things (IoT). Various cloud service providers (CSPs) offers huge volumes of storage to maintain and manage the IoT data, which can include videos, photos, and personal health records. These CSPs provide service properties, such as scalability, elasticity, fault tolerance, and pay per use. Thus, cloud computing has become a service paradigm to support IoT applications and IoT system deployment.

To ensure data privacy, existing research proposes to outsource only encrypted data to CSPs. However, the same or different users could save duplicated data by using different encryption schemes at the cloud. Although cloud storage space is big, duplication wastes networking resources, consumes excess power, and complicates data management. Thus, saving storage is becoming a crucial task for CSPs. Deduplication can achieve high space and cost savings. .At the same time, data owners want CSPs to protect their personal data from an unauthorized access. CSPs should therefore perform access control based on the data owners policies. the data owners want to control not only data access but also its storage and usage. The data deduplication should cooperate with data access control mechanisms. The same data, in an encrypted form, is only saved once at the cloud but can be accessed by different users based on the data owners policies. Data integrity can be provided to enhance the security of the data in the cloud. ECDSA, Elliptic Curve Digital Signature Algorithm can be used to check the integrity with the help of Trusted third party. After successful deduplication, data integrity can be checked to improve the security in the cloud.

RELATED WORK

The current industrial deduplication solutions cant handle encrypted data. Existing solutions for deduplication are vulnerable to brute-force attacks and cant flexibly support data access control and revocation. It raises issues relating to security and ownership. Many users are likely to encrypt their data before outsourcing them to the cloud storage to preserve privacy, but this hampers de duplication because of the randomization property of encryption.

Message-Locked Encryption provides to achieve secure deduplication, goal currently targeted by numerous cloud storage providers. Message-Locke Encryption where the key under which encryption and decryption are performed is itself derived from the message. It is used in a wide variety of commercial and research storage service systems. A client first computes a key K H(M) by applying a cryptographic hash function H() to M, and then computes the ciphertext C E(K, M) via a deterministic symmetric encryption scheme. A second client encrypting the same file M will produce the same C, enabling deduplication.

Cloud storage has become a faster profit growth by a low cost, scalable, position for clients data. Since cloud computing environment is constructed based on architectures and interfaces, it has the ability to incorporate multiple internal and/or external cloud services together to provide high interoperability. The availability and integrity of outsourced data in cloud storages can be checked. There is a basic approach called PDP (Provable Data Possession). It is a proof technique for a storage provider to prove the integrity and ownership of clients data without downloading data. The proof-checking without downloading makes it especially important for large-size files and folders to check whether these data have been tampered with or deleted without downloading the latest version of data. PDP is used to ensure and enhance the integrity of data in the proposed system.

In cloud computing, data generated in electronic form are large in amount. To maintain this data efficiently, there is a necessity of data recovery services. There is a smart remote data backup algorithm, Seed Block Algorithm (SBA).The objective of this algorithm is twofold; first it help the users to collect information from any remote location in the absence of network connectivity and second to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. The time related issues are also being solved by SBA such that it will take minimum time for the recovery process. SBA also focuses on the security concept for the back-up files stored at remote server, without using any of the existing encryption techniques. SBA can be used to efficiently recover lost data and also enhance data integrity.

The Elliptic Curve Digital Signature Algorithm is the Elliptic Curve analogue to the more widely used Digital Signature Algorithm (DSA).It is the application of ECC to digital signature generation and verification. Its security is based on the elliptic curve discrete logarithm problem. Elliptic Curve Digital signature represents one of the most widely used security technologies for ensuring un-forge-ability and non-repudiation of digital data. The steps involved in ECDSA are formation of key-pair, signature-generation and signature-verification . The digital signature is typically

created using the hash function. The transmitter sends the encrypted data along with signature to the receiver. The receiver in possession of senders public key and domain parameters can authenticate the signature. ECDSA Provides more security. This algorithm is more secure against intruders.
ECDSA is used to provide more security in the cloud environment.

EXISTING SYSTEM

In the existing system, the encrypted data can be deduplicated based on the Attribute based Encryption Scheme while at the same time supporting secure data acess control. Here, Data's hash code is used to check the deduplication. This system containing three types of entities: CSP, Data Owner, and Data holder. There is only one data owner for one data. Data holders that are eligible data users and could save the same data as the data owner at the CSP. Data holders that are eligible 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. If diiferent users can upload the same data, CSP performs deduplication.

PROBLEM DEFINITION

In the existing system, There is no method is employed to ensure the data integrity. Only a simple secured acess control mechanism is implemented. 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.

PROPOSED SYSTEM

The proposed system provides the auditing services for data storage security in cloud. Data integrity is an important factor to ensure in almost any data and computation related context. It serves not only as one of the qualities of service, but also an important part of data security and privacy. In the proposed system, successful deduplication of encrypted data can be performed and secured data acess control by trusted third party. After deduplication in the cloud , data integrity can be checked by using ECDS , Elliptic Curve Digital Signature Algorithm with the help of trusted third

party. ECDS Algorithm can be run by Trusted Third party. ECDS Algorithm can be used for checking data integrity in the cloud.

CONCLUSION

Proposed model uses ECDS Algorithm, to ensure the data integrity along with the acess control mechanism. There is a Trusted Third Party is used to execute this algorithm. ECDS Algorithm is a key pair generation algorithm. It is used to check if authorized person is used to modify the data or not. ECDS Algorithm can be performed by the Trusted third party. This proposed system gives the high security to the data in the cloud.

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

[3] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Message locked 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. Cooperative provable 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