SECURE DEDUPLICATION FOR CLOUD STORAGE BY USING AES ALGORITHM

Prof. RM. SUGANYA, Ms. ROHINI N, Ms. NIVETHA P, Ms. SHAMILI P

Department of Information Technology
KLN COLLEGE OF ENGINEERING, Sivagangai, Tamil Nadu, India

***Abstract*: Data play an important role in every part of our life. Because processed information is obtained only from the processed data. Those data are gathered from various users and stored in a cloud storage. Cloud is a new technology that is developed to reduce the storage area and cost of storage. Users having same data share a common storage area and that data can be fetched whenever needed. These data's are encrypted and stored as cipher text to avoid data threat. The number of users and the data stored in cloud is increasing exponentially day by day. The only way to reduce the cloud storage is data deduplication, elimination of repeated data in cloud. The data deduplication becomes more and more a necessity for cloud storage providers. Rendering efficient storage and security for all data is very important for cloud computing. Securing and privacy preserving of data is of high priority when it comes to cloud storage. Therefore to provide efficient storage for cloud data owners and render high security for data this System consists of three segments: 1) Identity Management 2) Data deduplication 3) Secure Cloud Storage. Intrusion detection and prevention are performed manually by network operators in the existing system. In our proposed system data deduplication technique allows the cloud users to manage their cloud storage space effectively by avoiding storage of repeated data's and save bandwidth. The data are finally stored in cloud server namely CloudMe. To ensure data *confidentiality* the data are stored in an encrypted format using Advanced Encryption Standard (AES) algorithm.**

***Keywords*: Advanced Encryption Standard (AES), Deduplication, Authorization and Re-encryption.**

INTRODUCTION:

Distributed computing is the method for changing how data innovation (IT) is expended and managed, promising improved cost efficiencies, quickened advancement, quicker time-to-showcase, and the capacity to scale applications on interest. Be that as it may, as the distributed computing is rising and growing quickly both adroitly and as a general rule, the legitimate/legally binding, financial, administration quality, interoperability, security and protection issues still posture huge difficulties. In this undertaking, we portray different administration and arrangement models of distributed computing and distinguish real difficulties. Specifically, we talk about three basic difficulties: security and protection issues in distributed computing. A few answers for decrease these difficulties are additionally proposed alongside a concise introduction on the future patterns in distributed computing sending. The utilization of distributed computing has expanded quickly in a considerable lot of the organizations. Cloud registering gives numerous advantages as far as ease and expanding openness of information. Guaranteeing the security of distributed computing is a main consideration in the field of distributed computing condition, as clients frequently store touchy data with distributed storage suppliers yet these suppliers may not be trusted.

MOTIVATION:

To eliminate duplication of stored data. To implement the encryption using AES encryption effectively. Each and every stage user will be validate by admin and CSP. So the data will be securely stored and retrieved by the user. The data can be stored securely and avoid the duplicate data this is the main objective of the project.

PROBLEM STATEMENT:

There has been much research done from this relevant process. This different result has been due to diversity in different aspects of methods used in the research. Due to all this factors, it is not easy to compare and choose the method which can be said as the best One. Hence, there is always room for the development of better method suitable for specific application.

LITERATURE SURVEY:

PAPER 1

Hybrid cloud approach for secure authorized deduplication Author :Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou IEEE transaction on Information forensics and security

Data deduplication is one of the important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

PAPER 2

Secure Deduplication for Cloud Storage Using Interactive Message-Locked Encryption with Convergent Encryption, To Reduce Storage Space Author: Jayapandian N, Md Zubair Rahman A M J IEEE transaction on Information forensics and security

The digital data stored in the cloud requires much space due to copy of the same data. It can be reduced by dedupilcation, eliminating the copy of the repeated data in the cloud provided services. Identifying common checkoff data both files storing them only once. Deduplication can yield cost savings by increasing the utility of a given amount of storage. Unfortunately, deduplication has many security problems so more than one encryption is required to authenticate data. We have developed a solution that provides both data security and space efficiency in server storage and distributed content checksum storage systems. Here we adopt a method called interactive Message-Locked Encryption with Convergent Encryption (iMLEwCE). In this iMLEwCE the data is encrypted firstly then the cipher text is again encrypted. Block-level deduplication is used to reduce the storage space. Encryption keys are generated in a consistent configuration of data dependency from the chunk data. The identical chunks will always encrypt to the same cipher text. The keys configuration cannot be deduced by the hacker from the encrypted chunk data. So the information is protected from cloud server. This paper focuses on reducing the storage space and providing security in online cloud deduplication.

PAPER 3

Fuzzy Identity-Based Encryption 2020 Author: A. Sahai , and B. Waters IEEE transaction on Information forensics and security

We introduce a new type of Identity-Based Encryption (IBE) scheme that we call Fuzzy Identity-Based Encryption. In Fuzzy IBE we view an identity as set of descriptive attributes. A Fuzzy IBE scheme allows for a private key for an identity, ω, to decrypt a ciphertext encrypted with an identity, ω ′, if and only if the identities ω and ω ′ are close to each other as measured by the "set overlap" distance metric. A Fuzzy IBE scheme can be applied to enable encryption using biometric inputs as identities; the error-tolerance property of a Fuzzy IBE scheme is precisely what allows for the use of biometric identities, which inherently will have some noise each time they are sampled. Additionally, we show that Fuzzy-IBE can be used for a type of application that we term "attribute-based encryption".

In this paper we present two constructions of Fuzzy IBE schemes. Our constructions can be viewed as an Identity-Based Encryption of a message under several attributes that compose a (fuzzy) identity. Our IBE schemes are both error-tolerant and secure against collusion attacks. Additionally, our basic construction does not use random oracles. We prove the security of our schemes under the Selective-ID security model.

PAPER 4

Efficient revocation in ciphertext-policy attribute based encryption based cryptographic cloud storage 2019

Author: Y. Cheng et al

IEEE transaction on Information forensics and security

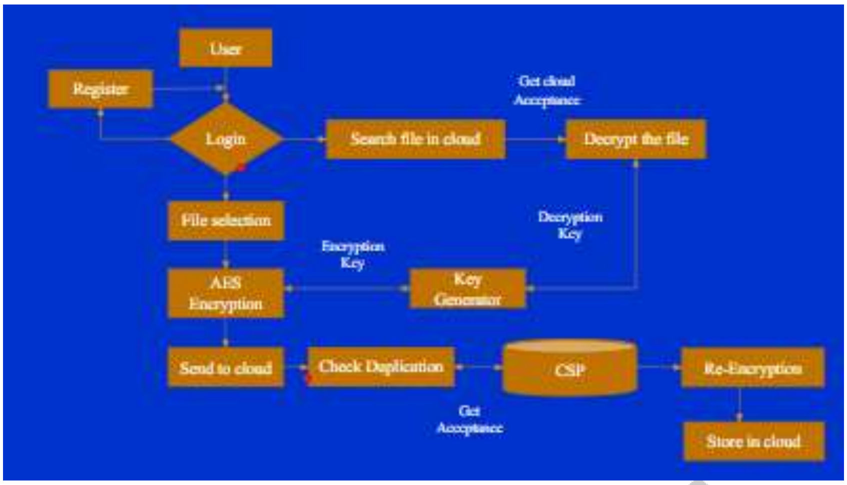
It is secure for customers to store and share their sensitive data in the cryptographic cloud storage. However, the revocation operation is a sure performance killer in the cryptographic access control system. To optimize the revocation procedure, we present a new efficient revocation scheme which is efficient, secure, and unassisted. In this scheme, the original data are first divided into a number of slices, and then published to the cloud storage. When a revocation occurs, the data owner needs only to retrieve one slice, and re-encrypt and re-publish it. Thus, the revocation process is accelerated by affecting only one slice instead of the whole data. We have applied the efficient revocation scheme to the ciphertext-policy attribute-based encryption (CP-ABE) based cryptographic cloud storage. The security analysis shows that our scheme is computationally secure. The theoretically evaluated and experimentally measured performance results show that the efficient revocation scheme can reduce the data owner's workload if the revocation occurs frequently.

PROPOSED SYSTEM:

In this system the security and low storage comes together at low cost, because we enhance the security by implementing high level security algorithm and a efficient storage service to reduce the storage. So, This will produce following features :

- Make the data more secured.
- Reduces storage space.
- Avoid Duplication.

SYSTEM ARCHITECTURE:



ADVANTAGES:

- Make the data more secured
- Reduce storage space
- Avoid duplication
- Backup also available because we used bi-cloud.

LIMITATION

- Internet Connection necessary

APPLICATIONS:

- O/S : Windows 7
- Language : Java
- IDE : Net Beans 8.2
- Data Base : MySql
- Server : WampServer

CONCLUSION:

In this framework, we scramble the information present in the cloud twice by utilizing intermediary reencryption technique.This is recommended that utilizes AES in half breed with symmetric intermediary reencryption conspire.So information can be re-encoded by cloud servers. The information proprietor just needs to create a lot of re-encryption keys and AES figure content scrambling the new keys then send both to the cloud for reencryption.

FUTURE WORK:

In the future, we will investigate how to achieve the same functionalities of AES with the same security guarantee without independent key servers. We notice that, our current encryption algorithm works fine with the text file. In future we intend to develop and implement an algorithm that could encrypt and decrypt other data formats like images, tables, graphs etc.

REFERENCES:

1. Secure Password-Protected Encryption Key for Deduplicated CloudStorage Systems Yuan Zhang; Chunxiang Xu; Nan Cheng; Xuemin Sherman Shen IEEE Transactions on Dependable and Secure Computing Year: 2021 | Early Access Article | Publisher: IEEE
2. A Hybrid Cloud Approach for Secure Authorized Deduplication" by Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Wenjing Lou in 2020.
3. "Secure Auditing and Deduplicating Data in Cloud" by Jingwei Li, Jin Li, Dongqing Xie, and Zhang Cai in 2106

4. A. Sahai, and B. Waters for "Fuzzy identity-based encryption,"Advances in Cryptology–EUROCRYPT 2005, LNCS, vol. 3494, pp.457-473.
5. “Efficient revocation in ciphertext-policy attribute-based encryption based cryptographic cloud storage” Y. Cheng etal IEEE transaction on Information forensics and security
6. “Conjunctive Broadcast and Attribute-Based Encryption“ F. Wang, J. Mickens, N. Zeldovich, V.Vaikuntanathan IEEE transaction on Information forensics and security
7. “Efficient Hybrid Proxy Re-Encryption for Practical Revocation and Key Rotation” S. Myers and A. Shull IEEE transaction on Information forensics and security.