ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCAIT-2017

JSS Academy of Technical Education, Bengaluru

Vol. 6, Special Issue 5, May 2017

Analysis of Encryption Algorithms and Deduplication Technique in Cloud

Punashri M S[1], Nirmala H[2]

PG Student, Department of Computer Science, SJBIT, Bengaluru, India[1]

Associate Professor, Department of Computer Science, SJBIT, Bengaluru, India[2]

Abstract: The most well known and vital popular service of Cloud is Data Storage. Information are regularly put away in an Encrypted shape to secure and protect information security of holders. Traditional deduplication schemes cannot work on encrypted data. Existing solutions of encrypted data deduplication suffer from security weakness. They cannot flexibly support data access control and revocation.New difficulties for cloud information deduplication is presented from Encrypted information. In Cloud this ends up noticeably significant for information preparing and capacity. Existing arrangements of information deduplication have less security in others words security shortcoming. Repudiation and information get to controls. In this an answer for deduplicate encoded information in cloud in light of intermediary re-encryption and possession challenges. With get to control it coordinates cloud information deduplication. The execution assessment depends on Computer reproductions and broad investigation. Results show the better performance of encryption algorithms, and data deduplication in Cloud storage.

Index Terms: Cloud Computing, Encryption Algorithms, Digital signature, Deduplication.

I. INTRODUCTION

Cloud Computing offers a new way technology by arranging resources of various types like computing and storage, based on the demands cloud provide information to the users or clients. It has desirable features like elasticity, fault-tolerance, security, encryption, pay as you use model. Promising platform for clients to store and manage a data in remote server where user can access from anywhere. The better approach for data innovation benefit offered by distributed computing is modifying different assets and the information is given to clients on their requests. This is turned into a promising administration stage due to a few properties. For example, adaptation to non-critical failure, pay per utilize versatility, and flexibility are the alluring properties of Cloud Computing.

The clients of cloud transfer secret or individual information to the datacentre of CSP (Cloud Service Provider such as Amazon, Google and so on.) and enable it to keep up these information. Because of a few assaults and interruption towards touchy information at CSP are not avoidable. Cloud clients can't completely believe the CSP. The security issue turns out to be more genuine because of alternate investigation innovations and the quick improvement of Data Mining. Some of the time the deduplicated information in encoded frame to CSP might be transferred by same or distinctive cloud clients. Putting away similar information in scrambled frame or ordinary information or Data deduplication squanders assets of system, entangles the administration, part of vitality devours. For the information holders it is hard to keep up the deduplication because of many reasons. For example, 1) Storage deferral is brought about Data holders may not be in online dependably or accessible for such administration, 2) Deduplication turn out to be excessively confused as far as Computation and correspondence to include information proprietors into deduplication prepare. 3) the way toward finding the deduplication may barge in the security of information holders. Hence Cloud benefit furnishes can't coordinate with information holders on information stockpiling deduplication as a rule. High cost saving is achieved and proved by Deduplication. Reducing upto 65% in file systems and 90-95 % storage needs backup applications. Existing Systems are not able to deduplicate the encrypted data and cannot ensure security privacy authentication, reliability. When data holders are not online it's hard to manage the deduplication due to many reasons, and it causes storage delay. This paper works on encryption algorithms, to find which performs better. Using 4 algorithm such as ECC(Elliptic Curve Cryptography), DES(Data Encryption Standard), AES(Advanced Encryption Standard) and RSA. Data ownership challenges, digital signature, to manage data which is encrypted use PRE. Our goal is to solve data duplication problem and to save storage space in other way saving money. Already user saved the document in the cloud and when the other user try to save the same content with the different name, it should tell the second user that the content is already existing.

In this work try to avoid Duplication to save Storage Space as the user is are paying for cloud its necessary to think about the storing space, user should not store same

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCAIT-2017

JSS Academy of Technical Education, Bengaluru

Vol. 6, Special Issue 5, May 2017

data more than one time, if user do that its waste of storage space in other words users are wasting our own money.
Encrypted data introduce new challenges for cloud data de duplication and Traditional de duplication schemes cannot work on encrypted data.
The deduplicated data in encrypted form to CSP may be uploaded by same or different cloud users. Storing the same data in encrypted form or normal data cause Data deduplication wastes resources of network, complicates the management, lot of energy consumes. For the data holders it is difficult to maintain the deduplication due to many reasons. Objective of this work is:

- To design and implement a solution to deduplicate the encrypted big data in cloud.
- To increase the efficiency, effectiveness and applicability.
- To save the storage space in cloud and protect the privacy of data holders or cloud users.
- The solution can flexibly supports sharing the data even when the data owner is not in online.

The rest of paper is organised as follows, Part 2 Literature survey, gives brief overview of related work, Part 3 System Design and implementation, Part 4 Result, Part 5 Results, Part 6 finally, Conclusion for performed work.

II. LITERATURE SURVEY

Service Providers of Cloud Storage such as Dropbox[1], Amazon, Google and other performs deduplication to save storage space by saving the document only one time, in other words uploading a copy single time in the cloud. Clients encrypt their data, saving storage by deduplication is completely lost. Encrypted information's are saved as various contents by applying various encryption keys. Existing techniques are failed in encrypted data deduplication.

DeDu[2] is an efficient deduplication system, but it not able to handle encrypted information or cipher text. This deduplication consists of two important components one is Hadoop Distributed File System [HDFS] and Front-end deduplication application. Back end distribution file system is HDFS in Hadoop. To use Hadoop Database to build a mass storage system and to build up a fast indexing system. There are two problems addressed, 1) how the duplications are identified by the system? 2) how to manage the data in the system? This work is not only to back up their data of organization it can be used To store their private data of common users.

DupLESS[3] CSP Providers such as Dropbox, Google, Amzon, Mozy etc. Should perform deduplication to save storage space by uploading a single copy of file or a document. Obtained by key server message based keys to encrypt the data of client PRF Protocol. In flexible way Data access of other users Cannot control.

Table I Comparison of Deduplication Technique

Approach	Encryption Scheme	Deduplication Strategy used
Message-locked encryption and secure deduplication	Message locked encryption	File level
BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication	Block Level Message locked encryption	Dual level: File level and Block level
HEDup: Secure Deduplication with Homomorphic Encryption	Homomorphic encryption	File level
DupLESS: Server-Aided Encryption for Deduplicated Storage	Enhanced Message level encryption to support security against Brute force attack	File level
ClouDedup:Secure Deduplication with Encrypted Data for Cloud Storage	Convergent encryption with added access control mechanisms	File level
Secure Deduplication with Efficient and Reliable Convergent Key Management	Convergent encryption	Block level
Twin clouds: An architecture for secure cloud computing	Convergent encryption	File level
A hybrid cloud approach for secure authorized deduplication	Convergent encryption	File level

Message-Locked Encryption[4] [MLE] Encryption and Decryption are performed derived from the message. Secure deduplication is achieved by MLE. ROM security is provided analyses of a family of MLE. Key K= H(M) M is message encrypted message C= E(K,M) =E(H(M),M) tag T=H(C) H is cryptographic hash function and E is cipher block this is derived by Alice. 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. provide definitions both for privacy and for a form of integrity that call tag consistency. Based on this foundation, make both practical and theoretical contributions.
Bleumer, and Strauss (BBS) proposed an application called atomic PRE[5]. In which converted semi – trusted proxy converts to cipher text Bob from cipher text Alice. Secure and fast re-encryption predicts will become most vital method for encrypted file systems Management

III. SYSTEM DESIGN

In the architecture data is stored in remote server called Cloud, First the user as to register for the cloud and he as to sign in to upload the data to the cloud, While uploading the data user can upload in both encrypted form and plain

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCAIT-2017

JSS Academy of Technical Education, Bengaluru

Vol. 6, Special Issue 5, May 2017

text form. For encryption using the ECC Algorithm. And during upload time the data it should check for duplication, if the content exists it should pop a window telling that the content is already existing. While uploading the data user can upload in both encrypted form and plain text form. For encryption using ECC Algorithm. and as our experiment ECC perform better compare to other algorithms. There are 5 modules in this work such as: Cloud Server, User, Register, Sign in, Auditor.

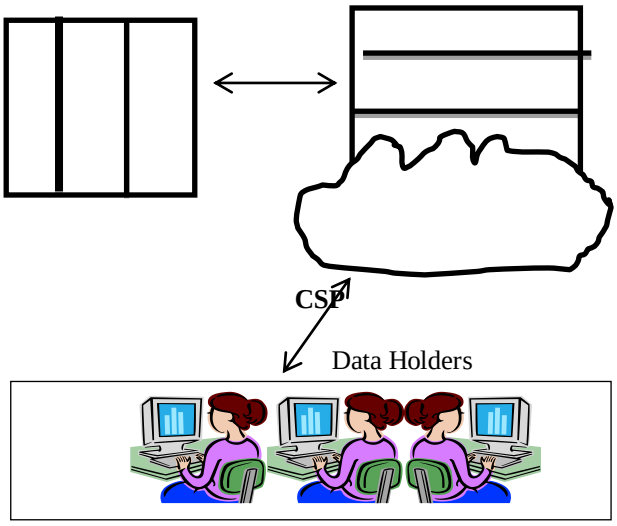

Fig.1. System Model

The Cloud Server to takes the request from the user and send back the response to the client or data user and check for the authorization to store the data to the cloud server. The Register module to hold the information or details of the data owner and data user. The Sign in module is to check the authorization of the person trying upload or download the file to or from Cloud. The Data User Module is the module for user with the authorization use digital signature for the privacy. The Auditor checks for the file modification.
In Register module, Name, password, user name, email id is necessary, for sign in only user name and password is enough.

Algorithm for Encryption,
In ECC
Step 1: Key Generation Q=d*P
The key is generated for encryption and decryption purpose.
Step 2: Encryption C_1=k*P, C_2=M+k*Q
Encryption is done using the above equation. Converting Plain text into cipher text.
Step 3: Decryption: M=C_2-d*C_1
Decryption is done using the above equation. Converting the cipher text into original form or plain text.

Proof:
M= C_2-d*C_1 M can be represented as C_2-d*C_1
C_2-d*C_1=(M+K*Q)-d*(K*Q) (C_2=M+K*Q and C_1=K*P)
=M+K*d*P-d*K*P

AES Analysis
Flexibility of key length for a level of future-fixing
The segments of AES are according to the accompanying
- Symmetric key symmetric piece figure
- 128-piece data, 128/192/256-piece keys
- More grounded and speedier than Triple-DES
- Give full specific and design purposes of intrigue
- Programming implementable in C and Java

To begin with the AES
- Include round key
- Blend areas
- Move lines
- Byte substitution

TABLE II SYSTEM NOTATION

Key	Description
Q	Public Key
P	Private Key
D	Random Number
M	Original Message
C	Cipher Text
H()	Hash Function
E	Encryption Algorithm
D	Decryption Algorithm

IV. SOME COMMON MISTAKES

Signature verification
- Input: system parameters and signature σ = (E0,E1,E2,ACOM,BCOM,c,taux,taus,tauePrime,taut,tauE)
- Output: True or False.

This work contains following aspects:
A. Encrypted Data Uploaded: If the duplicate check is negative, the data user can upload the data to the cloud by encrypting it using the encryption key.
B. Data Deduplication: this occurs at the time when the users try to upload the same data which is already existing in the cloud. Comparison is done in back end using Secure Hash Algorithm (SHA-256).
C. Data Owner Management: It should be managed because to save the data from the unauthorized user, this is done by using the digital signature algorithm.
D. Updating Encrypted Data: if the DEK is Updated by the owner of the data with DEK` and the new cipher text replace the old data, with the support of AP the CSP provide the re encrypted DEK` to all the data owners.

IV. IMPLEMENTATION

The Implementation steps are shown in below flow chart:

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

NCAIT-2017

JSS Academy of Technical Education, Bengaluru

Vol. 6, Special Issue 5, May 2017

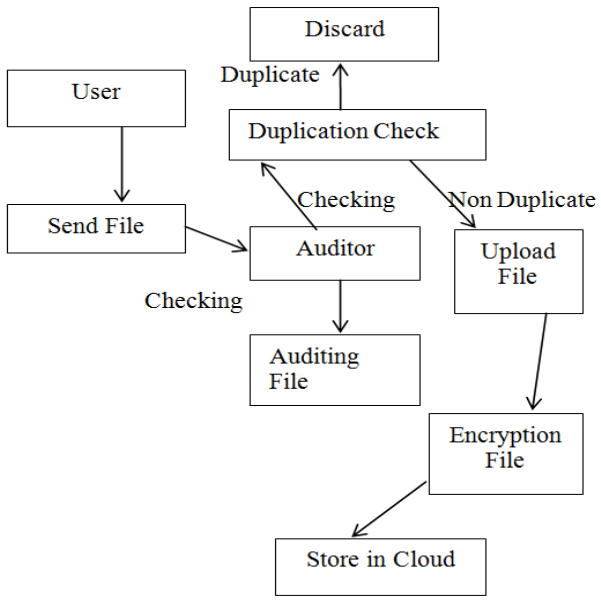

Fig.2. Data Flow Diagram

First the user try to upload the file, the file sent to the auditor, he checks for the duplication, if any duplication is existing it discards the file and send a message to the user telling that the data is existing, If there is no duplication it encrypt the data using the algorithm and store to the cloud, the auditor also checks for data modification. For implementing this to the system using the Java, eclipse and Json, Dropbox for the cloud. user have to set path to the cloud to store the data. As users are paying for the cloud he should not store the single data many times into the storage space.

It reduces the efficiency and then it cost more. Waste of money and performance decrease. The below figure explains how the data travel from one part to another and what and all the things are active during the upload session, and the user can also download his encrypted file. The downloaded file will be in original form, the decrypted file is downloaded for the user if he requests

V. RESULTS

Analysis of the Encryption Algorithms gives the below output, which shows the ECC algorithm perform better compare to other algorithm, and using ECC for encryption technology in our work. Below figure shows the graph. Performance evaluated based the time taken by the algorithm to encrypt the data. Deduplication saves the cloud space.

In this experiment, selected 192-bit field of elliptic curve (160-bit ECC has a security level comparable to 1024- bit RSA), 256-bit AES, 1024-bit PRE and 10M uploaded data. Tested the operation time of each step of data ownership Verification as presented in Fig. 3.

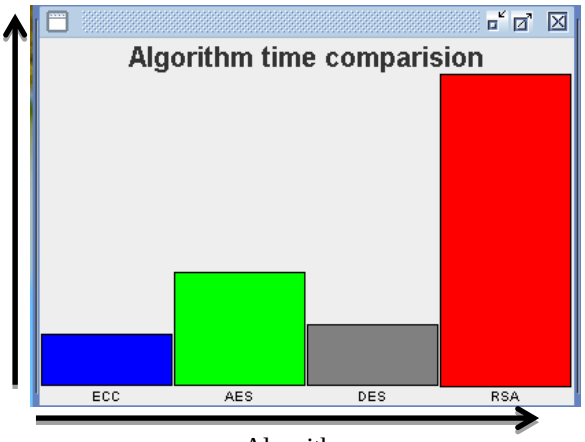

Fig. 3. Analysis of Encryption Technique

VI. CONCLUSION

Performance and Analysis of Encryption Technique helps us to find the better encryption algorithm and deduplication method helps us to save the storage space in cloud which minimizes the cost. Only authorization users can encrypt and decrypt the data using the keys generated during encryption. In this experiment it works better for encryption data and performs in a better way.

REFERENCES

[1] Dropbox, A file-storage and sharing service. (2016). [Online]. Available: http://www.dropbox.com

[2]Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355, doi:10.1109/CSCWD.2011.5960097.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. Cryptology—EUROCRYPT, 2013, pp. 296–312, doi:10.1007/978-3-642-38348-9_18.

[5] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inform. Syst. Secur., vol. 9, no. 1, pp. 1– 30, 2006, doi:10.1145/1127345.1127346.

[6] Y. Z. Zhou, Y. X. Zhang, H. Liu, N. X. Xiong, and A. V. Vasilakos, "A bare-metal and asymmetric partitioning approach to client virtualization," IEEE Trans. Serv. Comput., vol. 7, no. 1, pp. 40–53, Jan.-Mar. 2014, doi:10.1109/TSC.2012.32.

[7] N. X. Xiong, et al., "Comparative analysis of quality of service and memory usage for adaptive failure detectors in healthcare systems," IEEE J. Select. Areas Commun, vol. 27, no. 4, pp. 495–509, 2009, doi:10.1109/JSAC.2009.090512.

[8] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc 27th Annu. ACM Symp. Appl. Comput., 2012, pp. 441–446.