ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Secure De Duplication over the Cloud Environment Using Multi Authority Attribute Based Encryption

Vina M Lomte[1], Ashlesha D Gaikwad [2]

Department of Computer Engineering, RMD Sinhgad School of Engineering, Warje, Pune, India[1, 2]

ABSTRACT: Attribute based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted due to a cloud service provider and can share the data with user possessing specific credentials or attributes. Standard ABE System does not support secure de duplication which crucial for eliminating duplicates copies of identical data in order to save storage space In this paper present an attributes based storage detection and public cloud manager the storage. Compared with prior data De duplication the main two advantages is it can be used to confidentially share data with user by specifying access policies rather than sharing decryption key. And secondly it achieves the standard notion of semantic security for data confidentiality. In Advances the Modify a cipher text over one access policy into cipher text of same plaintext but under other access policies

KEYWORDS: De- duplication, cloud, cloud security, Authorized check duplicates, confidentially, auditing, Encrypted Data.

1. INTRODUCTION

In cloud computing, authority accepts the user enrollment and creates some parameters. Cloud service provider (CSP) is the manager of cloud servers and provides multiple services for client. Data owner encrypts and uploads the generated ciphertext to CSP. User downloads and decrypts the interested ciphertext from CSP. The shared files usually have hierarchical structure. That is, departments of files are divided into a number of hierarchy sub departments located at different access levels. If the files in the same hierarchical structure could be encrypted by an integrated access structure, the storage cost of ciphertext and time cost of encryption could be saved. Presently a day's more number of plans utilized encryption for control the information in Cloud. It empowers clients with restricted computational assets to outsource their expansive calculation workloads to the cloud, and monetarily appreciate the monstrous computational power, data transfer capacity, stockpiling, and even proper programming that can be partaken in a compensation for each utilization way.

To realize scalable, flexible and fine-grained access control of outsourced data in cloud computing. The outsourced computation workloads contain sensitive information such as business financial records, proprietary research data or personally identifiable health records etc. Users may try to access the data files outside their privileges. Hence a hierarchy is proposed where a particular department of users trusts a domain authority. The domain authority in turn trusts the trusted authority.

Background: The cloud computing refers to both the application delivered as services over the Internet and hardware and system in data centers that provides those services. The services themselves have long been. Referred to as software as services (SaaS) and some other terms as the Iaas(Infrastructure as a services) and PaaS(Platform as services).In main cloud computing system does not work as the Data Integrity and de-duplication. In cloud de duplicated data are storage in cloud. Moreover the same name of file and same name of content are stored in cloud then the space are major disadvantage.

DOI: 10.15680/IJIRCCE.2018.0605052

Aim:

1. Security of data to store in cloud when Data Owner send the File to Cloud then as Security double Encryption is used.
2. Data Integrity is used in application
3. Reduction of Storage when system used the de duplication concept then some name of file and same content are not stored in cloud.
4. File Sharing is used for cloud when the Multiple file are used for Multiple user.
5. Accessibility.

Scope: Data integrity auditing and secure duplication enables the guarantee of file confidentiality. Double Encryption is used for Security purpose first encryption is used in admin side and Second encryption is used for the Cloud Server provider Side. The major goal of this web application is to help the users to store their data on the cloud with confidentiality and security. De-duplication of data is the main focus in the entire web application. Providing storage of data on a large scale with multiple file sharing. Auditing helps the user to check the integrity of the data.

II. RELATED WORK

1. A Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing

AuthorQian Wang

Cloud Computing system has been predicted as the next-generation architecture of IT Enterprise. It moves the application software and databases to the centralized with large data centers, where the management of the data and services may not be fully trustworthy. This unique ensample brings about many new security challenges, which have not been well understood. Our research work examine the problem of assuring the integrity of data storage in Cloud Computing. In particular, we consider the task of allowing a third party auditor (TPA), on concern of the cloud client, to verify the integrity of the dynamic data stored in the cloud. The introduction of TPA dismiss the involvement of client through the auditing of whether user's data stored in the cloud is truly intact, which can be important in achieving economies of scale for Cloud Computing. The support for data dynamics through the most general forms of data operation, such as block modification, insertion and deletion, is also more powerful step to - ward practicality, since services in Cloud Computing are not limited to archive or backup data only. While presiding work on ensure remote data integrity often lack the supports of either public verifiability or dynamic data operation.

2. B.Proofs of Ownership in Remote Storage Systems

AuthorShaiHalevi

Cloud storage systems are becoming more and more popular. A promising technology that keeps their cost down is de duplication, which stores only a single copy of duplicating data. Client-side deduplication attempts to identify deduplication opportunities already at the client side and save the bandwidth of uploading copies of existing files to the server. In this work we identify attacks that exploit client-side deduplication, granting an attacker to gain access to arbitrary-size files of other users based on a very small hash signature of these files. More specifically, an attacker who knows the hash signature of a file can assure the storage service that it owns that file, hence the server lets the attacker download the entire file.

3. DupLESS: Server-Aided Encryption for De duplicated Storage

Author: MihirBellare

Cloud storage service providers such as Dropbox, Mozy, and others perform deduplication to save space by only storing one copy of each file uploaded. Should clients frequently encrypt their files, however, savings are lost. Message-locked encryption (the most remarkable manifestation of which is convergent encryption) resolves this tension. However it is inherently vulnerable to brute -force attacks that can recover files falling into a known set. We propose an architecture that provides secure deduplicated storage opposing brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt the under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an current service, have the service perform deduplication on their behalf, and yet achieves strong confidentiality guarantees. We show that encryption for deduplicated storage can achieve performance and space savings near to that of using the storage service with plaintext data.

4. D.Provable Data Possession at Untrusted Stores

Authors: Giuseppe Ateniese

We introduce a model for provable data possession (PDP) that allows a client that has stored data at an untrusted server to verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a s mall, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking supports large data sets in widely - distributed storage systems.

5. E.Remote Data Checking Using Provable Data Possession

Authors: Giuseppe Ateniese

We suggest a model for provable data possession (PDP) that can be used for remote data checking: A client that has stored data at an untrusted server can verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking is lightweight and supports large data sets in distributed storage systems. The model is also robust in that it incorporates mechanisms for mitigating arbitrary amounts of data corruption.

6. E.Remote Data Checking Using Provable Data Possession

Authors: Giuseppe Ateniese

We suggest a model for provable data possession (PDP) that can be used for remote data checking: A client that has stored data at an untrusted server can verify that the server possesses the original data without retrieving it. The model generates probabilistic proofs of possession by sampling random sets of blocks from the server, which drastically reduces I/O costs. The client maintains a constant amount of metadata to verify the proof. The challenge/response protocol transmits a small, constant amount of data, which minimizes network communication. Thus, the PDP model for remote data checking is lightweight and supports large data sets in distributed storage systems. The model is also robust in that it incorporates mechanisms for mitigating arbitrary amounts of data corruption.

III. ALGORITHMS

International Data Encryption Algorithms (IDEA):-

- First of all 64 bits plan text is divided into 4 16 bit parts and they are taken as input in first round.
- At end of first encryption round four 16 bits values are produced which used input to second encryption are round.
- The process is repeated in each of the subsequent 8 encryption round.
- Note that in 9th round we have to use only 4 key (k49, K50, K51, K52) and have to perform different operation as guided in previous slide.

Message Digest Algorithms MD5:

Steps of Algorithms

1. Appends padding bits – the input message is “padded” (extended) so that its length (in bits) equal to 448 mod 512. Padding is always performed, even if length of message is already 448 mod 512.

2. Padding is performed as follows a single "1 " bit is appended to message and then "0" bits are appended so that length in bits of padded message become congruent to 448 mod 512 A least one bit and at most 512 bits are appended.
3. Append length- A 64 bit representation of length of message is appended to result of result of step 1.if length of message is greater than 2 pow 64 only the low order 64 bits will be used.
4. The Resulting message (After padding with bits and with b) has a length that is an exact multiple of 512 bits. The input message will have that is an exact multiple of 16(32 bits) word.
5. Initialize MD buffer – A four word buffer (A,B,C,D) is used to compute the message digest. Each of A,B,C,D is an 32 bits register Theses register are initialized to fallowing values in hexadecimal , low order bytes first.
6. Process message in 16 bits block – Four Function will be defined such that each function taken an input of three 32 bits words and produces a 32 bit word output.

MATHEMATICAL MODEL

Variable used in Mathematical model

1. $X=(x_1,x_2,x_3,x_4.....................x_B)$
2. Here B block of data
3. $H(x)=B \log_2 q$ bits.
4. Let *N* be the number of available CSPs for the user to store data. Before storing the file, the user encodes the *B* blocks of data into *n* blocks. We use
5. $f: F^B = F^n$
6. Which maps ***x*** into ***y***, to denote the *encoding function*
7. $y=(y_1,y_2,y_3,\ldots\ldots\ldots\ldots y_n)$
    a. for $i=1,2,3\ldots\ldots\ldots\ldots N$
    b. here N is sub-vectors
8. let $y_i=(y_{i,1},y_{i,2},\ldots\ldots\ldots y_{i,ni})$ € F^{ni}
    a. Be the data stored on CSP(Cloud Services provider)
    b. $n=\sum_{i=1}^{N} n_I$ (1)
    c. n= total number of encoded block
        i. $\sum_{i=1}^{N} n_I$ =sum of Number of Encoded block stored in each CSp.
    d. Let,
    e. V_i for i € N be the amount of blocks which can be downloaded from CSP *i* within a predefined time delay, it is required that
    f. $n_i \leq Vi$ (2)
    g. In this work we assume *Vi* 's are integers, $Vi \geq 1$ and distinct for $i \in N$. We call *Vi*the *budget* of the stored data on CSP *i*. Let *Ci*be the cost for storing one block of data on CSP *i* and *Ci*'s are all distinct. The total storage cost is given by
    h. $C=\sum_{i=1}^{N} C_i n_i$ (3)
    i. $H(x|y(S)) =0, \forall S \cap_{k\,N}$ (4)
9. The Above bound on H(x) holds for any $S \cap_k N$ and any T is a subset of S.
    a. Therefore (5) can be re-written as
    b. $H(x) \leq \log_2 q \min(\sum n_i - \sum n_i)$ (5)

IV. PROPOSED SYSTEM APPROACH

To solve this problem on exiting system we propose two secure systems Sec Cloud and Sec Cloud + while generated better and efficient system for accessing massive data on cloud. In this, firstly encrypted the plain data file and perform integrity auditing on that encrypted file. Sec Cloud system has achieved both integrity auditing and file de duplication in this process Server doesn't known the contain in file. In other word the functionalities of integrity auditing and secure de duplication are only imposed on plain text.SecCloud + managing de duplication on encrypted files. On other word operation perform on secure file.

Advantage

1. Sec Cloud and Sec Cloud + are used**.**
2. Sec Cloud is supported on both block level and sector level.

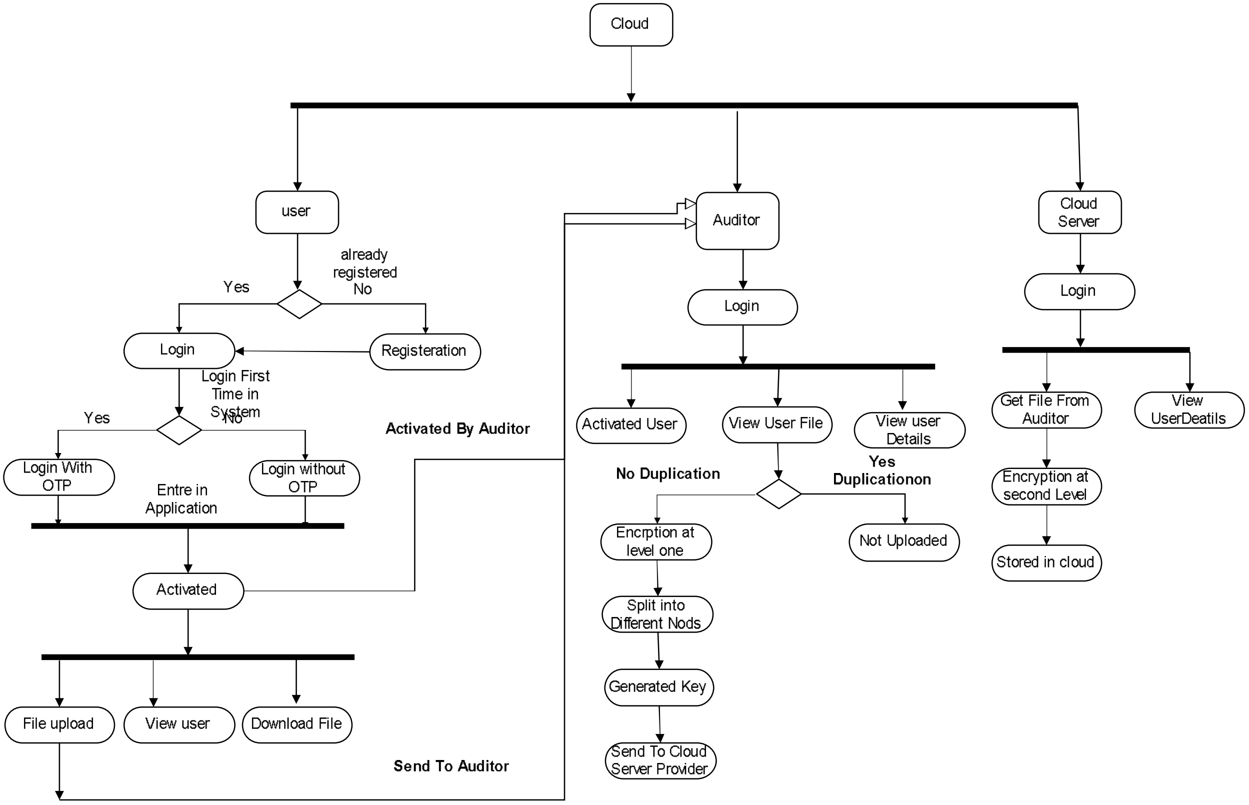

Figure 1: System Diagram

Module:

Module 1 : User / End User: - user responsible to upload the Data in cloud and that file send to Auditor.

Module 2: -Auditor: -Auditor check the file of user and check file copy is already present or not and Encryption at file level.

Module 3:- Cloud Server Provider: - CSP check the user details and Encrypted at second level and generated the key.

VI. RESULTS

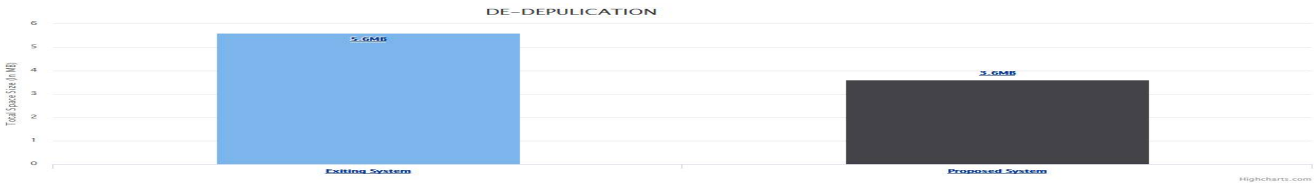

Figure 1 Exiting System and Proposed System

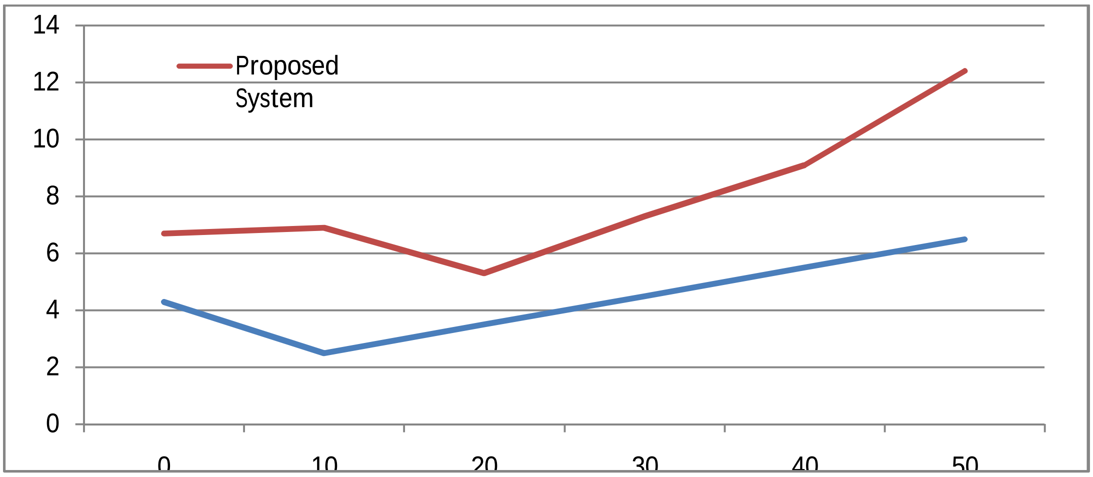

Figure 2 Proposed System for Content duplication and FileName Dedlication

File Size in Kb	Time for uploading in ms	Time For Downloading File
1	1.5	1
2	2	1.5
3	2.5	2
4	3	2.5

Table1: Comparison between Time for Uploading and Downloading

Metric	Existing system	Proposed System
Data Dynamic	No	Yes
Public Audit ability	Yes	Yes
Verifier Storage	No	Yes
Every time key Generation	No	Yes
KASE	No	Yes
Double Encryption	No	Yes

Table 2: Comparison between Exiting System and Proposed System

Our System Proposed attributes based storage system with secure de duplication over different algorithms IDEA and MD5 which is Rest of time in Second of attributes based storage is quick good. Shows the computation complexity of the proposed attribute-based storage system supportingsecure deduplication in terms of algorithms: key generation algorithm IDEA encryption algorithm Encrypt (Graph 1) and decryption algorithm Decrypt (Graph 2). As illustrated in Graph has the best performance, while Exiting System has the most expensive computational cost among all the curves.

VII. 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 credentials. On the other hand, de duplication is an important technique to save the storage space which eliminates duplicate copies of identical data. And Encryption data are also check the De duplication Process. However, the standard ABE systems do not support secure de duplication, 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 de duplication.Our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage. When the Encryption is done with are it regenerates the ciphertext into a ciphertext of the same plaintext over an access policy which is the union set of both access policies

REFERENCES

1. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia , ―A view of cloud computing,‖ Communication of the ACM, vol. 53, no. 4, pp.50–58, 2010.
2. J. Yuan and S. Yu, ―Secure and constant cost public cloud storage auditing with deduplication,‖ in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.
3. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, ―Proofs of ownership in remote storage systems,‖ in Proceedings of the 18th ACM Conference on Computer and Communications Security . ACM, 201, pp. 491– 500.
4. S. Keelveedhi, M. Bellare, and T. Ristenpart, ―Dupless: Serveraided encryption for deduplicated storage,‖ in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington, D.C.: USENIX Association, 2013, pp. 179–194. [Online]. Available:https://www.usenix.org/conference/usenixsecurity13/technicalsessions/presentation/bellare
5. G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, ―Provable data possession at untrusted stores,‖ in Proceedings of the 14th ACM Conference on Computer and Communications Security, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609.
6. G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, ―Remote data checking using provable data possession,‖ ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 12:1–12:34, 2011

DOI: 10.15680/IJIRCCE.2018.0605052