ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 10, October 2016

A Novel Approach for Secure Data Deduplication using Shared Authority in Cloud Storage

Prof.Ravi Rathod
M.E In Computer Science
PDEA's COEM

Ms.Akshada Chorage
Dept. of Computer Engineering
PDEA's COEM

Ms.Bhagyashri Landge
Dept. Of Computer Engineering
PDEA's COEM

Ms.Kadambari Londhe
Dept. of Computer Engineering
PDEA's COEM

Ms. Shilpa Waghmare
Dept. Of Computer Engineering
PDEA's COEM

***Abstract*- From the past few years, there has been a rapid progress in cloud, with the continuous and exponential increase of the number of users and the size of data .Data deduplication becomes more and more a necessry for cloud storage providers. For elimaniting duplicate copies of data we use data deduplication process. It is one of the important data compression technique for eliminating duplicate copies of repeating data. During the data accessing, different users may be in a collaborative relationship, and thus data sharing becomes significant to achieve productive benefits. And there is a necessity for protecting the data of various users using centralized resource The proposed system provides a solution for preserving the data in cloud with the deduplication. Also available of encryption protocol. The modules are presented here namely data owner, Deduplication, Encryption/Decryption, Third Party Administrator, and CSP. Advanced Encryption Standard algorithm is implemented in the current work for encrypting the data which has to be stored in the cloud. This data can be retrieved by retailer on providing the valid key to decrypt the data. It indicates that the proposed protocol realizing privacy-preserving data access authority sharing, is attractive for multi-user collaborative cloud applications.**

Keywords- Deduplication, Shared Authority, Encryption/decryption, CSP, TPA, Reliability, Tag Consistency, Authentication protocol, Privacy preservation.

INTRODUCTION

Now a day there is growth in information. With infinite storage space provide by cloud service provider users tend to use as much space as they can and vendors constantly look for techniques aimed to minimize redundant data and maximize space savings. Users will access information according to their needs and most users access same information again and again, the cost of computation, application hosting, content storage and delivery is reduced significantly. The cloud makes it possible for you to access your information from anywhere at any time. Cloud provides benefits such as, flexibility, disastracter, recovery, software updates automatically, pay- per-use model and cost reduction.[3] While a traditional computer setup requires you to be in the same location as your data storage device, the cloud takes away that step. The cloud removes the need for you to be in the same physical location as the hardware that stores your data. Each provider serves a specific function, giving users more or less control over their cloud depending on the type. Your cloud needs will vary depending on how you intend to use the space and resources associated with the cloud. Cloud computing refers to the use of computers which access Internet locations for computing power, storage and applications, with no need for the individual access points to maintain any of the infrastructure.

Data deduplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. Along with low ownership costs and flexibility, users require the protection of their data and confidentiality guarantees through encryption. To make data management scalable deduplication we are use Encryption for secure deduplication services.[1] Unfortunately, deduplication and encryption are two conflicting technologies. While the aim of deduplication is to detect identical data segments and store them only once, the result of encryption is to make two identical data segments in distinguishable after being encrypted. This means that if data are encrypted by users in a standard way as like shared authority, the cloud storage provider cannot apply deduplication since two identical data segments will be

different after encryption. On the other hand, if data are not encrypted by users, confidentiality by cannot be guaranteed and data are not protected against curious cloud storage providers.

There are two types of deduplication in terms of the size: (i) file-level deduplication, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and (ii) blocklevel deduplication, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixedsize blocks simplifies the computations of block boundaries, while using variable-size blocks. A technique which has been proposed to meet these two conflicting requirements is Tag generation and AES Scheme whereby the encryption key is usually the result of the hash of the data segment. Although encryption seems to be a good candidate to achieve confidentiality and deduplication at the same time, it unfortunately suffers from various well-known weaknesses. The confidentiality issue can be handled by encrypting sensitive data before outsourcing to remote servers.[13]

Along with low ownership costs and flexibility, users require the protection of their data and confidentiality guarantees through encryption.[9] In this paper, we address the a for mentioned privacy issue to propose a shared authority to the files which Deduplicted based privacy preserving authentication for the cloud data storage, which realizes authentication and authorization without compromising a user's private information. The main contributions are as follows.

1) Identify a new privacy challenge in cloud storage, and address a subtle privacy issue , in which the challenged request itself cannot reveal the user's privacy no matter whether or not it can obtain the access authority.

2) Propose an authentication protocol to enhance a user's access request related privacy, and the shared access authority is achieved by anonymous access request matching mechanism.

3) Apply cipher text-policy attribute based access control to realize that a user can reliably access its own data fields, and adopt the proxy re-encryption to provide temp authorized data sharing among multiple users.[10]

EXISTING SYSTEM

In a cloud management system to guarantee high data reliability in deduplication system is a critical problem. Most of the previous de duplication systems have only been considered in a single server setting. However, as lots of de duplication systems and cloud storage systems are intended by users and applications for higher reliability, especially in archival storage systems where data are critical and should be preserved over long time periods. The deduplication storage systems provide reliability comparable to other high available systems. Furthermore, the challenge for data privacy also arises as more and more sensitive data are being outsourced by users to cloud. Encryption mechanisms have usually been utilized to protect the confidentiality before outsourcing data into cloud. Commercial storage service providers are hesitant to apply encryption over the data because it makes de duplication impossible. The reason is that the traditional encryption mechanisms, including public key encryption and symmetric key encryption, require different users to encrypt their data with their own keys. As a result, identical data copies of different users will lead to different cipher texts.

Disadvantages:

- The traditional de duplication methods cannot be directly extended and applied in distributed and multi server systems.
- Maintain high data reliability in de duplication system is a critical one.
- This system only deduplicate data ,not provide shared authority concept.

PROPOSED SYSTEM

System Architecture:

1]The term deduplication defines that the process of discarding the duplicate copies or same copies in the cloud storage.And the shared authority defines to provide the security to the file which are been uploaded in the cloud .

2]The very first process the user has to login in the cloud and if he has already registered his account then directly

login in the sytem.And after the registretation of the user all the information is been saved in detail in the database.
3] Data owner and user is the one who will upload the file on cloud ,as the file will be uploaded on cloud the main thing it will do is to check the duplicate copies of the file which has been uploaded by the user.
4]The deduplication of the file will be done block wise deduplication each word in the file will be crossed checked and after that the next process if any duplicate copy is found that file will be discared and if not the file will be ready to be upload on cloud.
5]The file which is to be uploaded will be in encrypted format so that the other user will not be able to access the data from the file.

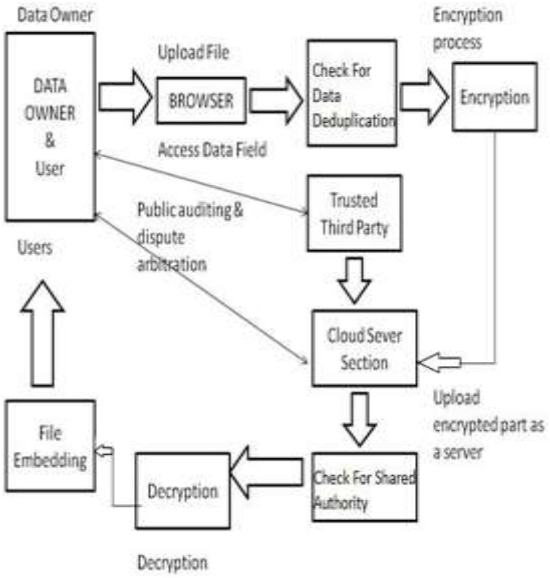

Fig: System Architecture

6]The second most important part of the project is shared authority as the file will be uploaded on cloud if the other user wants to access the users file he has to send a request to the cloud sever section and then the third party will ask the data owner to accept the request if he wants to allow the other user to access his data if not then the permission will be denied.
7]And if the data owner wants to allow the other user to access the data if will accept the request from the third party and the next the will will be converted in decrypted format so that the other user will be able to access the data from the file.
8]And as the file is in decrypted format the user can download the file from cloud storage.

SYSTEM MODULE

User Registration:
In this module an owner has to upload its files in a cloud server, he/she should register first. Then only he/she can be capable of doing it. For that he requires to fill the details in the registration form. These details are maintained in a database.

User Authenticate:
In this module,any of the above mentioned person have to authenticate,they should authenticate by giving their email-id and password
.**Utilizer Registration:**
In this module if a utilizer wants to access the data which is stored in a cloud, he/she should register their details first. These details are maintained in a Database.
Utilizer Authenticate:
If the utilizer is a sanctioned utilizer, he/she candownload the file by utilizing file id which has been stored by data owner when it was uploading.
Data Deduplication:
Check or Deduplicate file/data using Block level Deduplication & tag Generation algorithm for block level deduplication technique. We are going to discard the Duplicate file .Unique file will upload on cloud server.
S-CSP:
A CSP who handles cloud servers (CSs) and offers paid storage space on its infrastructure to store the owner's files.
Encryption & Decryption:
Here we are utilizing this aes_encrypt & aes_decrypt for encryption and decryption. The file we have uploaded which has to be in encrypted form and decrypt it.
File Upload:
In this module Owner uploads the file(along with Meta data) into database, with the avail of this metadata and its contents, the cessation utilizer has to download the file. The uploaded file was in encrypted form, only registered utilizer can decrypt it.
File Download:
The Sanctioned users can download the file from cloud database.
TTP (trusted third party) authenticate:
In this module TTP has monitors the data owners file by verifying the data owner's file and stored the file in a database .Also ttp checks the CSP(CLOUDACCOMMODATION PROVIDER), and ascertain whether the csp is sanctioned one or not.

Advantages:
- Improve the reliability of data
- Achieving the confidentiality of the users' outsourced data.
- Unique feature of the proposal is that data integrity, as well as tag consistency, can be achieved.

- Here we proposed the secured system and data owner can decide whether the user can access the system or not.

Mathematical model

Let S be the system object.
It consist of following
S={U,F,TPA,CSP}
U= no of users
U={u1,u2,u3,…..un}
F= no of files
F={f1,f2,f3,…..fn}
B=no of blocks.
B{B1,B2,…,Bn}
TPA= Third Party Auditor
TPA={C,PF,V,POW}
C=challenge
PF =proof by CSP
V= verification by TPA
POW= proof of ownership
CSP= Cloud Service provider
CSP={PF,F}
PF=proof
F=files

Used Algorithms

KeyGen(F) :
The key generation algorithm gives a input as file content F and generate outputs
as the convergent key ckF for F.

Encrypt(ckF;F) :
The encryption algorithm gives input as the convergent key ckF with file content F and generate the ciphertext ctF as output.

Decrypt(ckF; ctF) :
The decryption algorithm gives input, the convergent key ckF with ciphertext ctF and generate the plain file F as output.

TagGen(F) :
The tag generation algorithm gives input, a file content F and generate the tag tagF of F.

Advanced Encryption Standard :
The following AES steps of encryption for a 128-bit block are given below:
1. Derive the set of round keys from the cipher key
2. Initialize the state array with the block data (plaintext).
3. Add the initial round key to the starting state array.
4. Perform nine rounds of state manipulation.
5. Perform the tenth and final round of state manipulation.
6. Copy the final state array out as the encrypted data (ciphertext).

CONCLUSION AND FUTURE SCOPE

We proposed the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the users and also shared authority outsourced data with an encryption mechanism. Four constructions were proposed to support file-level and block-level data deduplication. The security of tag consistency and integrity were achieved.We implemented our deduplication systems using the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.In this work, we have identified a new privacy challenge during data accessing in the cloud computing to achieve privacy-preserving access authority sharing for deduplicated files. Authentication is established to guarantee data confidentiality and data integrity. Data anonymity is achieved since the wrapped values are exchanged during transmission. User privacy is enhanced by access requests to privately inform the cloud server about the users access desires .Forward security I srealized by the session identifiers to prevent the session correlation. It indicates that the proposed scheme is possibly applied for enhanced privacy preservation in cloud applications.

REFERENCES

1.Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang Senior Member, IEEE and Mohammad Mehedi Hassan Member, IEEE and Abdulhameed Alelaiwi Member, IEEE "Secure Distributed Deduplication Systems with Improved Reliability"IEEE Transactions on Computers Volume: PP Year: 2015
2. Miss Prachi D. Thakar, Dr. D.G.Harkut "Cloud based Hybrid Model for Authorized Deduplication" International Journal of Application or Innovation in Engineering & Management (IJAIEM). May 2015
3. T. Sivashakthi1, Dr. N Prabakaran "A Survey on Storage Techniques in Cloud Computing" International Journal of Emerging Technology and Advanced Engineering 2013
4. Mr. Dama Tirumala Babu, Prof.Yaddala Srinivasulu "A Survey on Secure Authorized Deduplication Systems" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056Volume: 02 Issue: 05 | Aug-2015
5. Sunita S. Velapure1, S. S. Barde "A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal of Science and Research (IJSR 2014)
6. Boyang Wang, Baochun Li, Member, IEEE, and Hui Li, Member, IEEE "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud" ieee 5th international conference on cloud computing year 2014.
7. Ayad F. Barsoum and M. Anwar Hasan "Provable Multicopy Dynamic Data Possessionin Cloud Computing Systems" ieee transactions on information forensics and security, 2015
8.N. Hemalatha M. Phil Scholar "A Comparative Analysis of Encryption Techniques and Data Security Issues in Cloud Computing" International Journal of Computer Applications 16, 2014.
9. Pasquale Puzio SecludIT and EURECOM "ClouDedup: Secure Deduplication with EncryptedData for Cloud Storage" International Journal of Computer Applications June 2014.

10. Hong Liu, Student Member, ieee, Huansheng Ning, Senior Member, ieee, Qingxu Xiong, Member, IEEE,and Laurence T. Yang, Member, IEEE "Shared Authority Based Privacy-preserving Authentication Protocol in Cloud Computing" IEEE transactions on parallel and distributed systems year 2014
11. Gore Swapnali, Gore Supriya, Tengale Kanchan, Tengale Varsha, Asst.Prof. S. B. Bandgar
" Modern Secure Distributed Deduplication Systems with Improved Reliability" International Journal of Advanced Research in Computer Science and Software Engineering October-2015
12. Mr.Kulakarni Harish, Mr.Ravi kumar chandu " Shared Authority Based Privacy-preserving Authentication Protocol in Cloud Computing" International Journal of Engineering Research and Applications (ijera national conference on Developments, Advances & Trends in Engineering Sciences (January 2015)
13. R.Bindu, U.Veeresh, CH. Shashikala "Provable Multicopy Dynamic Data Possession in Cloud Computing Systems" International Journal of Computer Engineering In Research Trends January-2016
14. Waghamare Amol Arjun "A Secure Hybrid Cloud Approach to Avoid Deduplication" International Journal of Computer Science and Mobile Computing April 2015
15. Zodge Kalyani1, Amruta Amune "review on secure distributed deduplication systems with improve reliability" Global Journal of Advanced Engineering Technologies 2016