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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Analysis of Empower Auditing and Secure Deduplication in Hybrid Cloud

Namrta Singh[1], Puneet Sharma[2]

Research Scholar, Department of Computer Science and Engineering, Amity University, Uttar Pradesh, India[1]

Assistant Professor, Department of Computer Science and Engineering, Amity University, Uttar Pradesh, India[2]

ABSTRACT: Cloud computing have brought great convenience for data sharing and data storage. The large no. of participants are sharing data in cloud. security issues like integrity, efficiency and privacy of the user should maintained while sharing the data in cloud. In cloud storage services, managing data is a major challenge as data is increasing continuously. For the management of data, deduplication is a well-known technique of data compression to remove duplicate copies of data in storage over a cloud. Data deduplication provides lot of benefits in cloud for security and privacy of data. Deduplication highly used in cloud storage to minimize storage space. In traditional deduplication used convergence encryption which provides confidentiality but do not check duplicates with some differential privileges. In this paper secured deduplication implemented to protect data security with differential privileges of user. The files are encrypted with differential privilege keys to provide robust security. Only the marked files are allowed to duplicate check by the user. The third party auditor can audit the data and verify the users presence of file deduplication in cloud. Auditor audits and verifies the uploaded file on a time .This paper useful for user and storage provider by auditing technique and deduplication technique respectively.

KEYWORDS: Auditing, Deduplication, cloud Security**.**

I. INTRODUCTION

1.1 CLOUD: It is a widespread innovative technology. This one is the growth of many computing (distributed, grid, parallel), and is the evolution and amalgamation of Virtualization, Utility computing, Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). Cloud is basically a representation to define web as a space where computing has been pre-installed which exists as an application; facility, storage, operating systems, data and processing power exists on the web ready to be shared. Cloud computing is a Pay-per-Use-On-Demand mode to users. Shared IT resources can be conveniently accessed by it through the Internet. IT resources include storage, application, network, server, service and so on and they can be deployed with much easy and quirk manner and it requires minimum management and also easy interactions with service providers. Cloud computing can improve the obtainability of IT resources and has many advantages over other computing techniques. IT infrastructure can be used by users with Pay-per-Use-On-Demand mode. It would bar the cost to purchase the somatic resources which might be vacant. Cloud computing is the result of evolution of IT services [3].

Internet has been a lashing force towards innumerable technologies that have been technologically advanced. Cloud computing paradigm has witnessed an enormous shift towards its adoption over the last few years and it has become a trend in the IT space as it assures new business potential and significant cost reductions to its users also providers. There are many leads of using cloud computing. Some of them include: [4]

i) Reduced hardware and conservation cost
ii) Flexibility and highly automated processes
iii) Accessibility around the globe

A. NECESSITY

Cloud storage provides extensively huge amount of data where it is been stored in virtualized pools of storage which are commonly hosted by 3rd parties. Cloud storage provides profits to the customers and helps in saving cost. These features connect more no. of customers and users to use the stored data their personal data to the cloud storage.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Cloud storage is one of the important service of the cloud computing. user can modify and share their outsourced data anywhere and anytime in the cloud[1].

In cloud storage data owner exchange data from their local computing system to the cloud .large no. Of data owners are ready to store data in the cloud [2].

Data Deduplication is an efficient process used in cloud storage to eliminate duplicate copies of data so that to improve storage space and optimize bandwidth.

Deduplication is a great mechanism for data compression in cloud storage and to remove redundant data from the cloud storage.

After secure deduplication process data auditing can be performing. The verification of the file can be done after Deduplication in cloud with the help of third party auditor. Auditor audits and verifies the uploaded file in a given time.

II. RELATED WORK

Cloud storage reduces the heavy data load for the storage management and maintenance with some low cost location independent scalable platform. To solve security problem, audit services are used for the maintenance of outsourced data. PDP is a cryptographic method to verify data integrity and prevent data from unauthorized access. Audit services are the good technique to minimize the heavy workload on the cloud storage services [8].

TPA is use to verify the dynamic data i.e stored in the cloud storage .it is an efficient approach to secure the data [9].

POR protocol is used to prevent a static achieved file in cloud storage .it provides flexible and cost effective environments. POR allows critical information stored storage-as-a service in encrypted format. Hash tree is used for the block tag authentication to secure cloud storage [10].

Current auditing specifies a cloud, which presents an auditing method with the use of a MAP Reduce cloud which is use to create data tags formerly uploading and audit the reliability of data warehoused in the cloud[11].

Data deduplication is a technique which is to use to remove duplicate copies of data. The main objective of data deduplication is to improve storage efficiency. In traditional, deduplication chunks identify and store only one replica of data, for other replica logical pointers are created instead of

sharing redundant data .Deduplication minimizes the both storage space and network bandwidth [7]. Deduplication is a process that makes data more scalable in cloud [6].

However, the current deduplication technique is a distributed deduplication to achieve reliability of data and the confidentiality of the user without using an encryption algorithm [5].

III. SYSTEM MODEL

Cloud Storage: In cloud computing, cloud user is one who is responsible to take outsource data on public cloud storage which is used as public cloud. In this system always try to provide authentication to enter in system provide in which data is uploaded with set of privileges for accessing the data to download.

Private cloud: Data owner and user are used for deduplication data with differential keys.
Auditor: PA is an auditor which used as expertise and capabilities, whenever cloud user do not have trust to assess the cloud storage reliability.
For each set of privileges and symmetric key is assigned and stored in private cloud. During the registration time, privileges are assigned to user according to identity given by the user, when user registers in to the system. Then, data owner with set of privileges upload and share file to users, the data owners is used for the identification and to send the file tag to the cloud server, responsible for the verification of the data owner and computes the file token and allows the token send back to the data owner.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Now, this file token is sent by the data owner and request to upload a file to the storage provider. If any duplicates found the user request to proof of ownership protocol with the storage provider to prove he/she has an ownership of respective file. PoW results: if PoW of file is passed then pointer is allotted to that file. Otherwise no duplicates are found for the file. For the proof, storage provider will return the signature for the particular file. User sends the set of privileges and gives confirmation to the private cloud server and requested towards uploading the file.
Now signature is verified by the private cloud server while receiving the call to upload file for the user. If signature verification result is passed, token file will be computed by private cloud, with each privilege from the privilege set given by the user which will return back to user.
Now, user performs the encryption, user encrypts the file with a key k and the key k is encrypted in to cipher text with each key in the file token given by the private cloud server. Then, encrypted file, tag and encrypted key are uploaded by the user. To download file F, user decrypt the encoded file with a key k and acquire the original file F.
The user is not sure about the presence of his/her file in the cloud. Now the auditing process supports to audit the files stored in the public storage. user assigns auditor from the cloud and sends the metadata of file which is going to upload in the cloud to the auditor. Auditor allocates an audit message to make assured that cloud server remembered the data file F in a given time, then a response message is generated from a function of the warehoused data file F and its metadata is verified by executing Gen Proof via the Verify Proof. TPA verifies the response for that particular data file.

IV. NOTATION AND PRELIMINARIES

Acronym	Description
(pkv,skv)	Users public and secret key pair
Kf	Convergent encryption for file F
Pv	Privilege set of a user
PF	Specified Privilege set of File F
Ø'F,P	Token of the File with privilege p
TPA	Third Party Auditor

Table 1.Notation

A.Symmetric encryption
A secret key k is used for encryption and decryption in symmetric encryption. There are three main function of symmetric encryption:

- k is key generation algorithm to generates k with the use of some security parameter.
- Then, symmetric encryption algorithm uses a secret key k to convert plaintext in to cipher text.
- Symmetric decryption algorithm uses a secret key k to convert cipher text in to plaintext.

B.Convergent encryption
During deduplication process convergent encryption supports data confidentiality[12],[13]. A convergent key is used to encrypt the original data by the data owner or user. A tag is generated for each data Files, which is used to check duplicates. If two data copies are comparable then their tags are also identical. For the identification of duplicates , user sends tag to the server and check whether the identical copies of the data present in the cloud storage or not. Convergent encryption and tags are derived independently; the encrypted data and its respective tag will be piled at the server. There are main functions of convergent encryption:

1. A key generation algorithm is derived to map data copy to a convergent key.
2. Symmetric key algorithm takes convergent key and data facsimile as an input and then converts in to cipher text as a output.
3. Then, decryption key algorithm takes convergent key and cipher text as a input and converts it in to original text as a output.
4. Tag is generated by using an set of rules that maps the imaginative data copy and tag as output.

C.Proof of ownership
PoW is an interactive algorithm which is hosted by prover and verifier. PoW allows user to access data copies storage

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

in server by proving their ownership for it[14].

D. Identification Protocol

Resilient and substantiate are the two phases of an identification protocol. In the first phase of Proof, A user or prover manifest his/her own identity to a verifier by identification proof which is related to his/her identify. A private key pkv is used by user or prover as a input , which is sensitive information as private key of public key in his/her certificate that he would not share with other user to maintain privacy. Now the verifiers allows for the verification with input of the public information pkv related to skv. The outcome of protocol is that verifier outputs either consent or discard for proof is passed or fail[15].

E. MAC- based Solution

There are two methods to use MAC for authentication of a data. In trifling, data blocks are uploaded with their respective MACs to the server. Now, secret key sk sends to TPA. Then data blocks with their MACs can be retrieved by the TPA and check the correctness by sk. For verification process, TPA requires more knowledge about data blocks rather than high communication and computation complexities[16]

V. CONCLUSION

Cloud have become more popular and are widely used for data sharing and data storage. The large no. of participants are added and use huge amount of data. As outsourced data are untreatable, security issues like integrity and confidentiality of data should maintained in cloud. Auditing technique is used to manage and audit the data and prevent data from unauthorized access in a hybrid cloud. As data in cloud is increasing continuously, some duplicates are present in cloud storage. Data deduplication is an effective method used to eliminate redundant data. Deduplication method maintains storage space and optimizes the bandwidth. In traditional, deduplication provides confidentiality with the use of convergent encryption but not allowed to check duplicates with differential privileges. To provide greater security files are encrypted with differential privilege key. Only the marked files are allowed for duplicate check by the user. The verification of the file can be done after deduplication in cloud with the help of third party auditor. Auditor audits and verifies the uploaded file in a certain time. This paper useful for user and storage provider by auditing technique and deduplication technique respectively.

REFERENCES

1. R.Buyya,C.S,Yeo,S.Venugopal,J.Broberg.and I.Brandic,"Cloud computing and emergingit platforms:Vision, hype,and reality for delivering computing as the fifth utility."Future generation computer system,vol,25, no.6,pp.599-616,2009.
2. CCachin,I,Keider,and A.Shrarer,"Trusting the cloud,"Acm Sigact News,Vol.40,no.2,pp.81-86 2009.
3. http://www.tutorialspoint.com/cloud_computing/index.html.
4. http://www.thecloudtutorial.com.
5. J.Yuan ands.Yu,"Secure and constant cost pubic cloud storage auditing with deduplication." In IEEE conference on Communications and Network Security(CNS),2013,pp.145-153.
6. S.Quinlan and S. Dorward.,"Venti:A new approach to archival storage ,storage,"USENIX FAST.Jan 2002.
7. SNIA,"Advanced Deduplication concepts,"2011.
8. Yan Zhuan,b,Hongxin Huc,Gail_Joon Ahnc,Stephen S.Yauc,"Efficient Audit Service Outsourcing For Data Integrity In Clouds",In The journal of systems and Software 82(2012).
9. Wang.Q,C.Wang,K.Ren,W.Lou,and J.Li,"Enabling Public Audit Ability And Data Dynamics For Storage Security In Cloud Computing",In IEEE Tras.Parllel Distributed System ,vol.22,no.5,pp.847-859,May 2011.
10. Juels.A and J.Burton,S.Kailiski,"Pors:Proof of Retrievability For LargemFiles".In Proc ACMConf.Computer and comm. Security(CCS'07),pp ,584-597,Oct 2007.
11. Jingwei Li,Jin Li,Dongqing Xie and Zhang Cai,"Secure Auditing and Deduplication Data in Cloud",IEEE Transactions on Computers 2015.
12. A.Adya,W.J.Bolosky,D.Simon , M.Theimer.J.R.Douceur,"Reclaimin Space from duplicate files in a serverless distributed,"ICDCS,pp.617-624,2002.
13. Sriram Keelveedhi,Thomas Ristenpart Mihir Bellare,"Message locked encryption and secure deduplication,"in springer Berlin Heidelberg,International Association for Crptologic Research Advances in Crptology-EUROCRYPT 2013, Athens,Greece,March 2013,pp.296-312.
14 . D.Harnik,B.Pinkas and A.Shulman-Peleg.S.Halevi,"Proof of ownership in remote storage system,"ACM Conference on zcomputer and Communications Security,pp 491-500,2011.
15. Chanathip Namprempre,Gregory Neven Mihir Bellare,"Security Proofs for Identity-Based Identification and Signature schemes,"Journal of Cryptology,Springer-Verlag,vol.22,no.1,pp.1-61,Janaury 2009.
16. Sherman S-M.Chow,Qian Wang,Kui Ren,and Wenjing Lou Cong Wang."Privacy-Preserving Public Auditing for Secure Cloud Storage."Computers ,IEEE Transactions,vol.62.no.2,pp.362-375,Feb2013.