Secure Authorized Encryption for Deduplication Based Hybrid Cloud Approach

Ashwini V. Garole[1], Dr.Mrs.Sharmila K.Wagh[2]
[1, 2] Department of Computer Engineering
[1, 2] Modern Education Society's College of Engineering, Pune, India

Abstract- *Secure deduplication is a technique for eliminating duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. This paper makes an attempt to primarily address the problem of authorized data deduplication. To protect the confidentiality of important data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. Along with the data the privilege level of the user is also checked in order to assure whether he/she is an authorised user or not. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. We show that our proposed authorized duplicate check scheme has minimal overhead compared to normal operations.*

Keywords- Authorized Deduplication, Secured, duplicate check, confidentiality, Hybrid Cloud computing.

I. INTRODUCTION

In computing, data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. This technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

A Hybrid Cloud is a combined form of private clouds and public clouds in which some critical data resides in the enterprise's private cloud while other data is stored in and accessible from a public cloud. Hybrid clouds seek to deliver the advantages of scalability, reliability, rapid deployment and potential cost savings of public clouds with the security and increased control and management of private clouds. As cloud computing becomes famous, an increasing amount of data is being stored in the cloud and used by users with specified privileges, which define the access rights of the stored. Figure 1. Architecture of cloud computing

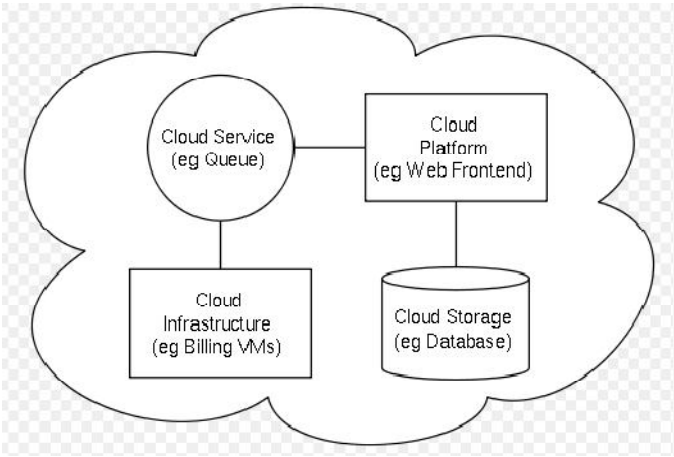

Figure 1: Architecture of cloud computing

The critical challenge of cloud storage or cloud computing is the management of the continuously increasing volume of data. Data deduplication or Single Instancing essentially refers to the elimination of redundant data. In the deduplication process, duplicate data is deleted, leaving only one copy (single instance) of the data to be stored. However, indexing of all data is still retained should that data ever be required. In general the data deduplication eliminates the duplicate copies of repeating data. The data is encrypted before outsourcing it on the cloud or network. This encryption requires more time and space requirements to encode data. In case of large data storage the encryption becomes even more complex and critical. By using the data deduplication inside a hybrid cloud, the encryption will become simpler.

II. LITERATURE SURVEY

J.R. Douceur et.al [1] implemented the feasible deduplication technique and maintain the data confidentiality

using convergent encryption technique. It encrypts decrypts a data copy with a convergent key, the content of the data copy obtained by computing the cryptographic hash value. After the data encryption and key generation process users retain the keys and send the cipher text to the cloud. Since the encryption operation is determinative and is derived from the data content, similar data copies will generate the same convergent key and hence the same cipher text.Authorized deduplication technique is proven by J.Li et.al [2] which avoid the duplicate content in cloud storage system and incurs minimal overhead as compared to the normal operation by using convergent key encryption. It also provides the security to the given data..S. Halevi, D. Harnik et.al [3] proposes Proofs of Ownership in Remote Storage Systems which contain Performance measurements indicate that the scheme incurs only a small overhead compared to naive client-side deduplication. A hybrid cloud is a combination of private cloud and public cloud in which the data which is most critical that resides on a private cloud and the data which is easily accessible is resides on a public cloud. M. Bellare et.al[4] Hybrid cloud is helpful for reliability, extensibility and fast deployment and cost saving of public cloud with more security with private cloud. P. Anderson et.al[5] Implemented The complex challenge of cloud storage or cloud computing is the arrangement of large volume of data duplication is a process of eliminating of duplicate data in de-duplication techniques.In the previous old system the data is encrypted back to outsourcing.W. K. Ng et.al [6]Proposes private data deduplication Protocols in cloud storage to Enhance the efficiency of data. R D.Pietro et.al [7]presented deduplication technique in hybrid cloud.The encryption technique become simpler. As we all of knows that the network has large amount of data which being shared by many users. Many large networks uses data cloud to store the data and share that data on the networkBoga Venkatesh et al.[8] proposed a deduplication system in the cloud storage to reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality, Bellare et al.[9]showed how to protect the data confidentiality by transforming the predicatable message into unpredicatable message. In their system, another third party called key server is introduced to generate the file tag for duplicate check. Token generation technique and identity based signature to provide security to the give data in cloud storage. N.B.Kadu et al[10] presented a novel encryption scheme that provides the essential security for popular data and unpopular data. For popular data that are not particularly sensitive, the traditional conventional encryption is performed. Another two-layered encryption scheme with stronger security while supporting deduplication is proposed for unpopular data. In this way, they achieved better trade between the efficiency and security of the out-sourced data. Jin Li et.al [11] developed the Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. This technique is used to improve storage utilization.Shai Halevi1, Danny Harnik,et.al[12] proposes the Proof of ownership which lets a client efficiently prove to a server that that the client keep a file, rather than just some short information about it present solutions based on Merkle trees and specific encodings, and analyse their securityJin Li, Xiaofeng Chen et.al [13] proposed Dekey, an efficient and reliable convergent key management scheme for secure deduplication. Dekey applies deduplication among convergent keys and distributes convergent key shares across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data. They implement Dekey using the Ramp secret sharing scheme and demonstrate that it incurs small encoding/decoding overhead compared to the network transmission overhead in the regular upload/download operations. A. Rahumed et.al [14]addressed Anonymous Authentication for data storing to clouds. Anonymous authentication is the process of validating the user without the details or attributes of the user. So the cloud server doesn't know the details or identity of the user, which provides privacy to the users to hide their details from other users of that cloud. S. Bugiel, S. Nurnberger,et.al[15] proposed architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. In come towards, the user communicates with a trusted cloud. Which encrypts as well as verifies the datastored and operations occurred in the untrusted cloud.It divide the computations such that the trusted cloud isused for security-critical operations in the less time-critical setup phase, whereas queries to the outsourced data are processed in parallel by the fast cloud on encrypted data.J.XU et.at.[16]showed how to protect the data confidentiality by transforming the predicatable message into unpredicatable message. In their system, another third party called key server is introduced to generate the file tag for duplicate check. J. Yuan et.al.[17] proposed a deduplication system in the cloud storage to reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality.Aparna patil et.al[18] proposed .To better protect data security, this project makes the first attempt to formally address the problem of authorized data deduplication .Different from the traditional deduplication system, differential benefits of the user are further considered the duplicate check besides the data itself. Hybrid cloud architecture contains several new deduplication constructions supporting authorized duplicate check. A.AlZa[19] in presented twin clouds: architecture for secure cloud computing for Client uses the trusted Cloud as a proxy that provides a clearly defined interface to manage the outsourced data, programs, and queries.Token generation technique and

identity based signature for provide security to the give data in cloud storage

III. PROBLEMS IN EXISTING SYSTEM

To make data management scalable in cloud computing, deduplication has been a well-known technique and has attracted more and more attention recently. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For filelevel deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.Data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges.Such architecture is practical and has attracted much attention from researchers.The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

These are the problems occur in the existing system as following:-

1. Users sensitive data are susceptible to both insider and outsider attacks.
2. Some times deduplication impossible.
3. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.
4. Identical data copies of different users will lead to different cipher texts, making deduplication impossible.

IV. PROPOSED SYSTEM

A) Significance of problem definition:

Todays cloud service providers offer both highly available storage and parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud
and shared by users with specified privileges, which define the access rights of the stored data. Our system handles critical challenge of cloud storage services is the management of the ever-increasing volume of data

B) Mathematical Model

Set representations:

Let ‘S’ be the system
S={I/P,O/P,SC,FC,Σ}
Where,
I/P=Input to the System
O/P=Output to the System
SC=Success Case
FC=Failure Case
Σ= Number of functions

I/P={LoginId, Password, Keyreq, Keyres}
O/P={Original data copy}
Success case={Valid I/P, Valid Key}
Failure case={Wrong Id, Wrong Password, Wrong Key}
Σ={Number of functions}
Σ={A1,A2,A3,A4}
A1={Key Generation Algorithm}
A2={Encryption Algorithm}
A3={Decryption Algorithm}
A4={Tag Generation Algorithm}

- The following four phases of mathematical model:-

A) Authentication

Suppose the data owner wants to upload a file, the owner must be privileged user.

Input= User, Data owner, Private cloud, key Authentication involves following process:

1) User must be a privileged one
2) He generates a key which he canuse that for Decryption, another kind of authentication
3) The generated key will be stored on the private cloud

B) Key Generation

A unique key is generated for each file which helps to identify the file duplication
Process= user, file data, key

1) Key generation involves following process. storage inside the public cloud, key generation generate a
2) User inputs a file. unique key which differ to each file. For retrieving data, user
3) A unique is generated for each file. can directly download data from cloud storage but only after
4) Key will differs from file to file. specifying the user authentication.

C) Anonymization

Once the file is checked in the cloud ,if the cloud does not the file content, the file will be encrypted before it got uploaded in the file.

Upload= File data, key It involves following procedure

1) File that user inputed
2) Key which is generated in key generation

D) De-Anonymization

If the user wants to download contents from the cloud user must specify the key and download the file contents of the data.

Download= file data, User Specified key It involves following procedures

1) Anonymized data
2) User Specified key

E) Hybrid Architecture for Secure Deduplication:-

In this paper, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model. In this paper, we will only consider the file level deduplication for simplicity. In another word, we refer a data copy to be a whole file and file-level deduplication which eliminates the storage of any redundant files. Actually, block-level deduplication can be easily deduced from file-level deduplication, Specifically, to upload a file, a user first performs the file-level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well; otherwise, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. Each data copy (i.e., a file or a block) is associated with a token for the duplicate check.

- *S-CSP:-* This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.
- *Data Users:-* A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized deduplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.
- *Private Cloud:-* Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

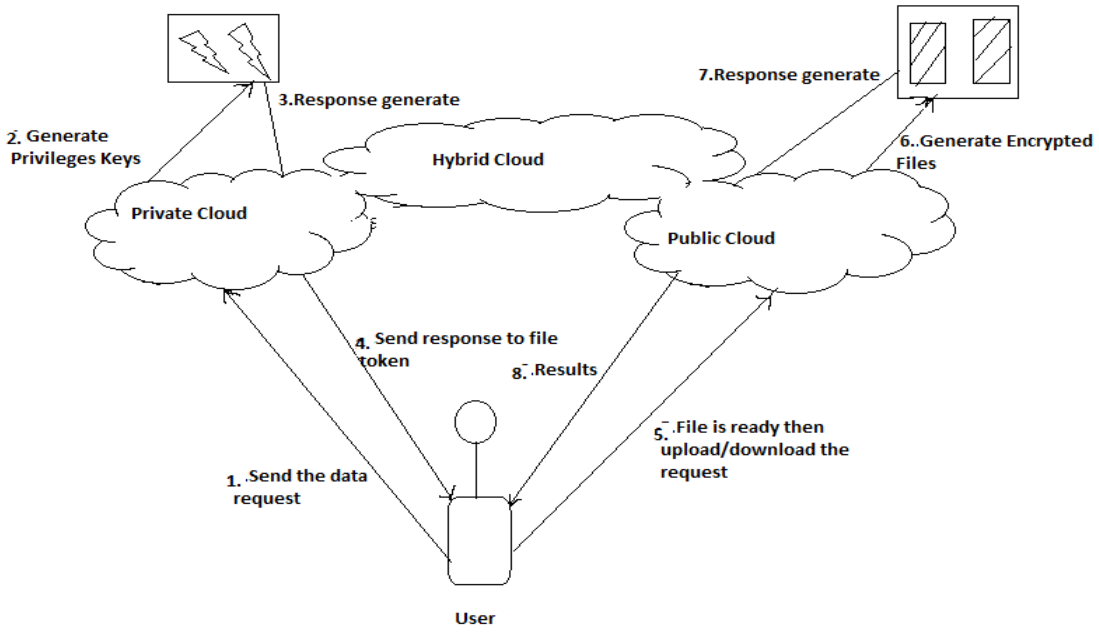

Figure 2. Architecture for Authorized Deduplication

- **Confidential Encryption:**

It provides data confidentiality in deduplication. A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the data copy, such that the tag will be used to detect duplicates.

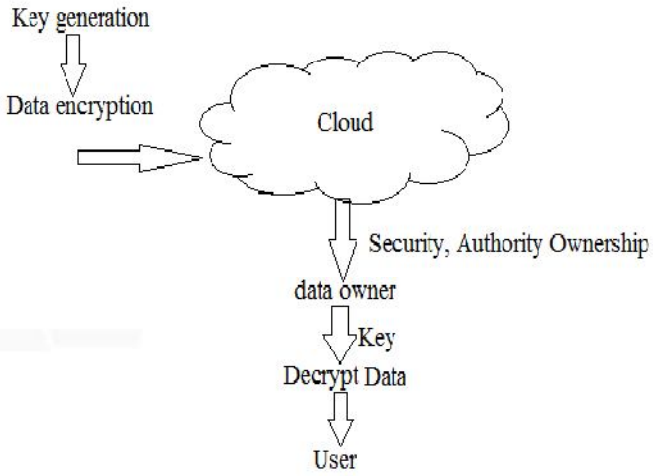

Figure 3: Confidential Data Encryption

- **Advantages of Proposed System**

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality.

V. ALGORITHM MEASURES AND METRICS

A) Performance Measures Used

- SHA 256 algorithm to check duplication of files.
- SHA 256 generates sha token for individual uploaded files in cloud based on file-name and file-contents .
- DES encryption algorithm used for confidentiality purpose, file stored in cloud as encrypted format because third person can not be read file contents.
- By OAuth scheme improve security of system.
- By using above all these methods in project performance of system is increased.

B) Comparison with Similar System

Table 1. Comparison with Similar System

Existing System[2]	Proposed System
1. Time Consuming process for uploading file.	1.Take few microseconds for uploading file.
2. File duplication check based on either filename and file contents.	2. File duplication check based on both filename and file contents.
3. Wastage of storage space.	3. Save storage space.

VI. CONCLUSION & FUTURE SCOPE

Authorized data de-duplication was proposed to protect the data security by including differential privileges of

users in the duplicate check. We also presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test-bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provide us with sufficient memory. It provides authorization to the private firms and protect the confidentiality of the important data.

REFERENCES

[1] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617–624, 2002.

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[4] M. Bellare, C. Namprempre , and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 2009.

[5] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

[6] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.

[7] R. D. Pietro and A. Sorniotti . Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security 2012.

[8] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST- NCSC National Computer Security Conf., 1992.

[9] Boga Venkatesh,Anamika Sharma,Gaurav Desai and Dadaram JadhavSecure Authorised Deduplication by Using Hybrid Cloud Approach.,In an International Journal of Innovative Research in Advanced Engineering (IJIRAE), Volume1 Issue 10 (November 2014).

[10] N.B. Kadu, Amit Tickoo, Saurabh I. Patil, Nilesh B. Bhagat, Ganesh B. DivteA Hybrid Cloud Approach for Secure Authorized Deduplication,International Journal of Scientific and Research Publications,Volume 5, Issue 4, April 2015 ISSN 2250-3153.

[11] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing LouA Hybrid Cloud Approach for Secure Authorized Deduplication,IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEM VOL:PP NO:99 YEAR 2014.

[12] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[13] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou "Secure Deduplication with Efficient and Reliable Convergent Key Management" IEEE Transactions On Parallel And Distributed Systems, VOL. 25, NO. 6, JUNE 2014.

[14] A. Rahumed , H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011.

[15] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[16] J. Xu, E.C. Chang and J. Zhou. Weak leakage resilient client side deduplication of encrypted data in cloud storage. In ASIACCS, pages 195–206, 2013

[17] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication .IACR Cryptology ePrint Archive, 2013.

[18] Aparna Ajit Patil.Dhanashree Kulkarni" A Survey on: Secure Data Deduplication on Hybrid Cloud Storage Architecture International Journal of Computer Applications (0975 –8887) Volume 110 –No. 3, January 2015

[19] A.AlZain "Cloud Computing Security: From Single to Multi-Clouds" 2012 45th Hawaii International Conference on System Sciences