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. 3, Issue 8, August 2015

A Dynamic Authentication for Client Side Deduplication in Cloud Storage Environment

E. Seetha[1], D. Ponniselvi [2]

M.Phil Research Scholar, Dept. of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women, Namakkal, TamilNadu, India [1]

Assistant Professor, Dept. of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women, Namakkal, TamilNadu, India [2]

ABSTRACT: Cloud computing is the most important emerging computing in which resources are shared over the internet. However, cloud storage environment faces one serious problem such as management of large volume of data. To manage vast amount of data, deduplication technique is used. Data deduplication is one of the important data compression techniques for eliminating duplicate copies of repeating data. To keep the confidentiality of sensitive data while supporting the deduplication, to encrypt the data before outsourcing convergent encryption technique has been proposed .To better protect data security, this project makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. In this paper we will examine about all methods, processes that are used in data deduplication. The proposed security models contain the demonstration of security analysis scheme. As a proof of concept contains the implementation framework of proposed authorized duplicate check scheme and conduct test bed experiments using these prototypes. In proposed system contain authorized duplicate check scheme incurs minimal overhead compared to normal operations.

KEYWORDS: cloud computing; deduplication; duplicate check; security; convergent encryption; proof of ownership

I. INTRODUCTION

Cloud computing is a model for delivering information technology services in which resources are retrieved from the internet through web-based interface and application, instead of direct connection to a server. Cloud storage provides a service for the evergreen management of vast amount of data in order to reduce the space and bandwidth. To make reliable and scalable management of data in the cloud computing, deduplication plays a vital role as a conventional technique. Deduplication is a data compression technique which is most commonly used for eliminating repeated copies of data/files in cloud storage to reduce space and bandwidth. This technique is used for reliable storage utilization and to provide scalable network data transfers to reduce number of bytes that must be sent. Data deduplication may occur as file level as well as block level data deduplication. Keeping multiple duplicate copies of file/data with similar content, deduplication detects and eliminates the redundant data by keeping original physical copy. This paper introduces a new cryptographic method for secure Proof of Ownership (PoW), based on the joint use of convergent encryption, for providing dynamic sharing between users and ensuring efficient data deduplication.

The remainder of this work is organized as follows. First section II describes the security analysis and related work in secure deduplication concepts. Then, section III introduces the proposed system and (PoW) system model. Finally, Section IV presents performance evaluation in our secure client side deduplication (PoW) scheme before concluding in section V.

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. 3, Issue 8, August 2015

II. RELATED WORK AND SECURITY ANALYSIS

A. Security Analysis

The security will be analyzed in terms of two aspects,

- Confidentiality of data and
- Authorization of duplicate check.

We suppose that all the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under this assumption, two kinds of adversaries are considered, that is, adversaries who aim to extract secret information as much as possible from both public cloud and private cloud, and internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. The data will be encrypted in our deduplication system before outsourcing to the storage cloud to maintain the confidentiality of data. The data is encrypted with the traditional encryption scheme and the data encrypted with such encryption method which guarantees the security of data. System address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for Differential Authorization and Authorized Duplicate Check.

Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any unauthorized user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server. Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs the duplicate check directly and tells the user if there is any duplicate.

The security requirements considered in two folds, including the security of data files and security of file token. For the security of file token unauthorized users without appropriate privileges or file prevented from getting or generating the file tokens for duplicate check of any file stored at the Storage cloud. The users are not allowed to collude with the public cloud server. It requires that any user without querying the private cloud server for some file token, he cannot able to get any useful information from the token, which includes the privilege or the file information and to maintain the data confidentiality unauthorized users without appropriate privileges or files, prevented from access to the underlying plaintext stored at Storage cloud.[1]

B. Related Work

In 2002 J. R. Douceur et al. [15] studied the problem of deduplication in multi-tenant environment. The authors proposed the use of the convergent encryption, i.e., deriving keys from the hash of plaintext. Then M.W.Storer et al. [16] pointed out some security problems, and presented a security model for secure data deduplication. However, these two protocols focus on server-side deduplication and do not consider data leakage settings, against malicious users.

In this paper [17] M. Bellare et al. Provides either security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly in existing literature. Underlying these is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work. They also analyze a generic folklore construction that in particular yields identity-based identification and signature schemes without random oracles.

In this paper J. Xu et al. [18] proposed growing need for secure cloud storage services and the attractive properties of the convergent cryptography lead us to combine them, thus, defining an innovative solution to the data outsourcing security and efficiency issues. Our solution is based on a cryptographic usage of symmetric encryption used for enciphering the data file and asymmetric encryption for Meta data files, due to the highest sensibility of this information towards several intrusions.

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. 3, Issue 8, August 2015

In addition the Merkle tree properties, this proposal is shown to support data deduplication, as it employs a pre-verification of data existence, in cloud servers, which is useful for saving bandwidth. Besides, our solution is also shown to be resistant to unauthorized access to data and to any data disclosure during sharing process, providing two levels of access control verification. Finally, we believe that cloud data storage security is still full of challenges and of paramount importance, and many research problems remain to be identified. [7],[13]

In this paper P. Anderson et al. 2010 [1] proposed a solution here the data which is common between users to increase the speed of backup and reduce the storage requirement namely backup algorithm. Supports client-end per user encryption is necessary for confidential personal data. This provides the potential to significantly decrease backup times and storage requirement. Storing huge amount of data in personal computer or laptops causes poor connectivity also may be theft due to hardware failure. However Network bandwidth can be a bottle-neck and Backing up directly to a cloud can be very costly are not addressed. Conventional backup solutions are not well suited to this environment. So client side deduplication necessary for confidential personal data.

In this paper J.R.Douceur et al. The Farsite distributed file system provides availability by replicating each file onto multiple desktop computers. In the view of the fact that this replication consumes considerable storage space, it is essential to reclaim used space where possible. Measurement of over 500 desktop file systems shows that nearly half of all consumed space is occupied by duplicate files. So there is need to present a mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Our mechanism includes convergent encryption, which enables duplicate files to combine into the space of a single file, even if the files are encrypted with different users. [3]

In this paper M. Mulazzani et al. [7] throughout the past few years, an enormous number of online file storage services have been introduced. At the same time as several of these services provide basic functionality such as uploading and retrieving files by a specific user, more advanced services offer features such as shared folders, real-time association, and minimization of data transfers or unrestricted storage space. Overviews of existing file storage services and examine Dropbox, an advanced file storage solution, in depth. Based on the results they show that Dropbox is used to store copyright-protected files from a popular file sharing network.

In this paper M. Bellare et al. Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication, a goal currently targeted by numerous cloud-storage providers. MLE is a primitive of both practical and theoretical concern. [2]

III. PROPOSED SYSTEM

In the proposed system (fig 1)we are achieving the data deduplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud. The secure primitives used in our secure deduplication. [10], [11]

A. Symmetric Encryption

It uses a common secret key k to encrypt and decrypt information. A symmetric encryption scheme consists of three primitive functions:

- **KeyGenSE(** 1^{λ} **)** $\rightarrow \kappa$ is the key generation algorithm that generates κ using security parameter 1^{λ};
- **EncSE(**κ**,***M***)** $\rightarrow C$ is the symmetric encryption algorithm that takes the secret κ and message M and then outputs the cipher text C; and

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. 3, Issue 8, August 2015

➢ **DecSE(κ,C)** $\rightarrow M$ is the symmetric decryption algorithm that takes the secret κ and cipher text C and then outputs the original message M.

B. ***Convergent encryption***

With this convergent encryption [4], [8] we get secure confidentiality of de-duplication. Data owner gets convergent key from each original data copy and encrypts data copy with the convergent key. A tag is also providing to the user with the data copy, tag will be used to detect duplicates. If two data copies are same, then their tags are same. To identify and check the duplicates, the user first sends a tag to the server side to check. Server will reply, if the identical copy has been already stored or not. Both (confidentiality check and tag) are independently derived. Tag cannot used to reduce the convergent key and compromise data confidentiality. Tag and its encrypted data copy will be stored in the server side. With the four primitive functions we can define the convergent encryption scheme.

➢ **KeyGenCE(M)** $\rightarrow K$ is the key generation algorithm that maps a data copy M to a convergent key K;

➢ **EncCE(K,M)** $\rightarrow C$ is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs cipher text C;

➢ **DecCE(K,C)** $\rightarrow M$ is the decryption algorithm that takes both the cipher text C and the convergent key K as inputs and then outputs the original data copy M;

➢ **TagGen (M)** $\rightarrow T(M)$ is the tag generation algorithm that maps the original data copy M and outputs a tag $T(M)$.

C. ***Proof of Ownership***

Enable the users to provide their ownership of data copies to the storage server we choose proof of ownership. Proof of ownership is implemented as an interactive algorithm run by a power and verifier. From a data copy of M, the verifier derives a short value Φ(M). To prove the ownership of the data copy M, the user needs to send Φ to the verifier such that $\Phi' = \Phi(M)$. The formal security definition for PoW roughly follows the threat model in a content distribution network, where an attacker does not know the entire file, but has accomplices who have the file. The accomplices follow the "bounded retrieval model", such that they can help the attacker obtain the file, subject to the constraint that they must send fewer bits than the initial min-entropy of the file to the attacker [11].

D. ***Identification protocol***

With two phases we can describe the identification protocol, **Proof** and **Verify.** In the stage of proof, a user U demonstrates his identity to a verifier by performing the some identification proof related to his identity. Private $sk_{,,}$ is the input of the user i.e. sensitive information such as private key of a public key in his certificate, credit card number, etc. These types of numbers cannot share with others. With the help of input of public information $pk_{,,}$ related to $sk_{,,}$, the verifier performs the verification. At the end of protocol, the verifier output either accept or not to denote whether the proof is passed or not. Different types of identification protocols are there like, certificate based and identification based identification [5], [6].

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. 3, Issue 8, August 2015

ACRONYM	DESCRIPTION
S-CSP	Storage Cloud Service Provider
PoW	Proof Of Ownership
(pk_u , sk_u)	User' public key and secret key
k_f	Convergent encryption key for file F
P_u	Privilege set of a user U
P_F	Specified Privilege set of a user U
$Ø'_{F,P}$	Token of the file F with privilege

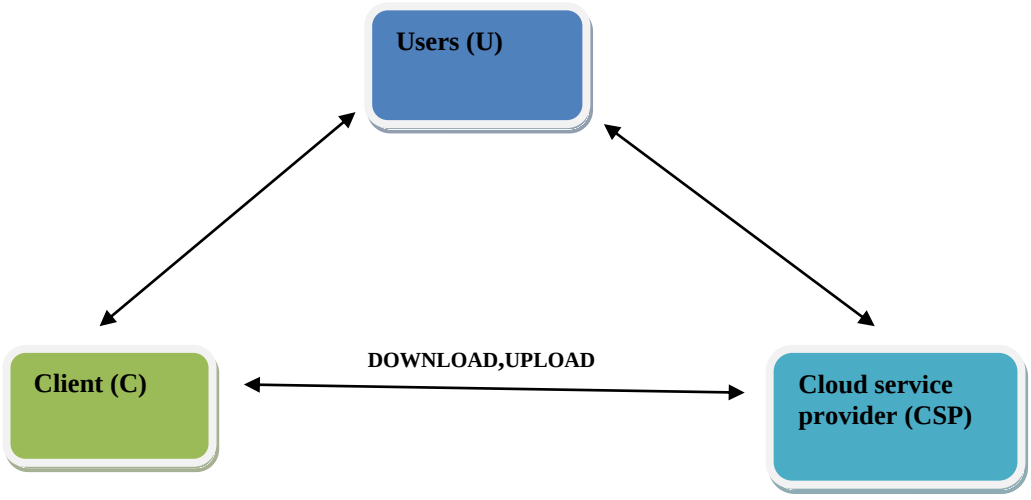


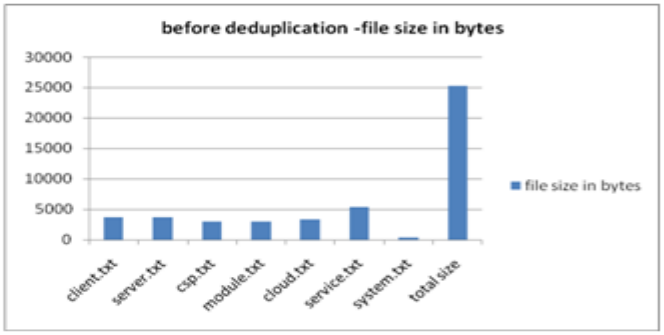
Figure 1: ARCHITECTURE OF CLOUD DATA STORAGE

IV. PERFORMANCE EVALUATION

A. ***Performance Evaluation:***

Results shown in the below given table and chart is the file size of Seven files of different size of .txt are chosen for deduplication. By applying deduplication approach on these files we are able to save the Cloud storage space up to 50%.

File Name	File Size In Bytes
client.txt	3734
server.txt	3736
csp.txt	3005
module.txt	3007
cloud.txt	3291
service.txt	5398
system.txt	309
total size	25266



a. File Size in bytes before deduplication

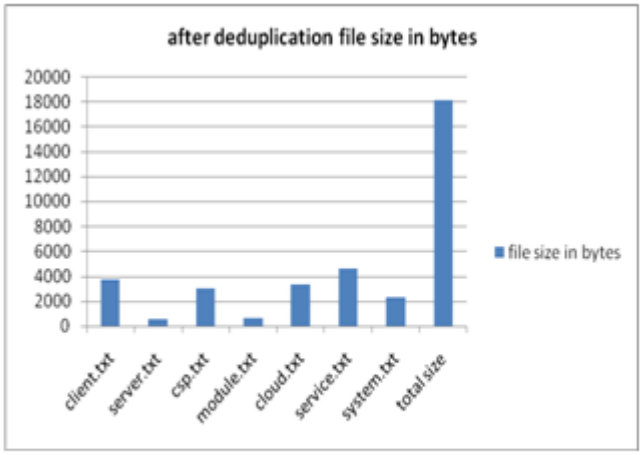
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

File name	File size in bytes
client.txt	3742
server.txt	536
csp.txt	3011
module.txt	607
cloud.txt	3301
service.txt	4610
system.txt	2301
total size	18108



b. File Size in bytes after deduplication

B. ***Result Analysis :***

The proposed system provides a safe cloud storage methodology which supports client side deduplication better than existing system. This thesis suggests that the security can be increased if the client side redundant files are avoided. Security mechanisms involved during client auditing of outsourced data is discussed. The methods are studied to achieve the data deduplication by providing the proof of data by the data owner. Three schemes are presented that can be applied in cloud storage environment to increase the security aspects. The common secret key is used to encrypt as well as decrypt data. Using three basic functions is achieved if first method is applied. The data confidentiality is achieved using the second method in which to derive convergent key for confidential convergent encryption The third method provides the security such that to prove that the data which user wants to upload or download is its own data. Finally convergent key and verifying data are provided to prove his ownership at server. It is proved that the proof of ownership scheme saves more amounts of space, bandwidth and cost compared to existing approach. The future study should focus on security proof and enhancements in data retrieval of the proposed framework.

V. CONCLUSION AND SCOPE FOR FUTURE ENHANCEMENT

The growing need for secure cloud storage services and the attractive properties of the convergent cryptography lead us to combine them, thus defining an innovative solution to the data outsourcing security and efficiency issues. The proposed work provides space saving as well as security to the data, but still there are various aspects which need to be address in future. Proposed work addresses only three types of file txt file, doc file and pdf file, we can extend this work for different types of file such as image and sound files and video files in future. We can also apply some searching technique that speed-up the operation by separating the hash key according to the file type.

VI. ACKNOWLEDGMENT

My abundant thanks to **Dr.B.T.SURESHKUMAR M.Sc., M.Phil., Ph.D.**, **Principal, Vivekananda College of Arts and Sciences for women, Namakkal** who gave this opportunity to do this research work. I am deeply indebted to **Dr.S.DHANALAKSHMI MCA., M.Phil., Ph.D., M.E., Head Department of Computer Science and Applications** at **Vivekananda College of Arts and Sciences for women, Namakkal** for this timely help during the paper. I express my deep gratitude and sincere thanks to my guide **Mrs. D. PONNISELVI, M.Sc., M.Phil., Assistant Professor, Department of Computer Science and Applications** at **Vivekananda College of Arts and Sciences for women, Namakkal** for her valuable, suggestion, innovative ideas, constructive, criticisms and inspiring guidance had enabled me to complete the work successfully.

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. 3, Issue 8, August 2015

REFERENCES

1. M. Bellare, S. Keelveedhi, and T. Ristenpart. **"Dupless: Server aided encryption for deduplicated storage"**. In USENIX Security Symposium, 2013.
2. P. Anderson and L. Zhang**." Fast and secure laptop backups with encrypted de-duplication"**. In *Proc. of USENIX LISA*, 2010.
3. 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.
4. 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.
5. J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. **"Secure deduplication with efficient and reliable convergent key management".** In *IEEE Transactions on Parallel and Distributed Systems*, 2013.
6. Ng and P. Lee. **"Revdedup: A reverse deduplication storage system optimized for reads to latest backups"**. In *Proc. of APSYS*, Apr 2013.
7. C.-K Huang, L.-F Chien, and Y.-J Oyang, **"Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs"** J. Am. Soc. for Information science and Technology, vol. 54, no. 7, pp. 638-649, 2003.
8. 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.
9. W. K. Ng, Y. Wen, and H. Zhu. **"Private data deduplication protocols in cloud storage"**. In S. Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012.
10. 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*, pages 81–82. ACM.
11. S. Quinlan and S. Dorward. Venti: **"A new approach to archival storage"**. In *Proc. USENIX FAST*, Jan 2002.
12. 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.
13. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. **"Role-based access control models"**. *IEEE Computer*, 29:38–47, Feb 1996.
14. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. **"A secure data deduplication scheme for cloud storage"**. In *Technical Report*, 2013.
15. 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 Proceedings of 22nd International Conference on Distributed Computing Systems (ICDCS, 2002).
16. M. W. Storer, K. Greenan, D. D. Long and E. L. Miller. **"Secure data deduplication".** In Proceedings of the 4th ACM International Workshop on Storage Security and Survivability, StorageSS 08, pages 1–10, 2008.
17. M. Bellare, C. Namprempre, and G. Neven, **"Security proofs for identity-based identification and signature scheme"**. 2009
18. J. Xu, E.-C.Chang, and J. Zhou, **"Weak leakage-resilient client side deduplication of encrypted data in cloud storage"**, 2013