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 9, September 2015

A Survey on: Secure Authorized Deduplication for Hybrid Cloud Storage

Sulakshana S.Patange, Ganesh V. Kadam

PG Scholar, Department of Computer Engineering, JSPM NTC, RajashriShahu School of Engineering and Research, Narhe, Pune, India

Assistant Professor, Department of Computer Engineering, JSPM NTC, RajarshiShahu School of Engineering and Research, Narhe, Pune, India

ABSTRACT: Systems providing secured data storage are now in greater demand. These systems provide data storage in a cost-effective manner. But a situation may arise, when the data storage consists of large amount of duplicate and redundant data. These duplicate records may occupy more space and access time. Hence, there is a need of banishing the duplicate records. Eliminating the duplicate records seems to be an easy task but requires a lot of work to do because the duplicate records don’t share any common key. Sometimes, errors occur as a result of transcription errors or incomplete information, lack of standard formats, or any combination of these errors. Many methods and approaches are introduced for dealing with the duplicate data. In this paper, we discuss some of the few methods along with its advantages and disadvantages for future directions.

KEYWORDS: Cloud Computing, Data storage, Deduplication.

I. INTRODUCTION

There are many services that are provided over the internet. Cloud Computing is one of the service that provides virtualized resources without limits to the users over the internet by hiding the platform and implementing details. Cloud data storage is the service by cloud computing that manages the huge mass of increasing data. This large data must be properly managed in cloud computing and the technique of deduplication is the conventional way of handling the data. Another technique known as data compression is used for eliminating the redundant copies of duplicate data for reducing the data deduplication. This technique can also be used in network data transfers as it reduces the total number of bytes to be sent along with improving the storage utilization. Keeping multiple data records with the similar contents is a task that increases the overhead. The technique of deduplication banishes the duplicate data and maintains one a single physical copy. If required, this single copy can be referred by the other redundant copies. The process of data deduplication can be carried on both, file level or block level. In the file level deduplication process, the redundant copies of identical files are eliminated. But in block level deduplication process, the duplicate blocks of data that occur in non-identical files are eliminated. There is a lot of well-being in the technique of data deduplication with comparison to the conventional techniques of encryption. As the traditional encryption process provides security and privacy to the users’ sensitive data against the attacks from both insiders and outsider attacks, this technique is contrary to the technique of data deduplication.

Providing data deduplication along with maintaining the data confidentiality, the technique known as convergent encryption can be used. This technique also provides feasible data deduplication. In convergent encryption, the data copy is encrypted and decrypted by using a convergent key. Once the data is encrypted and the process of key generation is done, the user keeps hold of the keys and sends the corresponding ciphertext towards the cloud storage. The encryption process of this technique is very determinative and derived from the data contents. Hence, in case if similar data copies of the original data are generated, then the convergent encryption of the data copies will produce the same ciphertext. Another protocol known as the secure proof of ownership protocol can be used to prevent the unauthorized access and also provide the proof to the user regarding the duplicates found of the same file.

During this review, we present the methods used in secure deduplication in section II. In section III, we briefly review the views of different researchers towards secure data deduplication in cloud storage and in section IV we describe the summary of the review paper.

II. METHODS USED IN SECURE DEDUPLICATION

1. Symmetric Encryption

Symmetric encryption uses a common secret key *k* to encrypt and decrypt information. A symmetric encryption scheme made up of three primary functions.

- KeyGen SE (1λ)→ k is the key generation algorithm that generates k using security parameter 1λ;
- Enc SE (k, M)→ C is the symmetric encryption algorithm that takes the secret k, and message M and then outputs the ciphertext C, and
- Dec SE (k, C) → *M* is the symmetric decryption algorithm that takes the secret *k* and ciphertext*C* and then outputs the original message *M*.
-

2. Convergent Encryption

Convergent encryption [1], are used to provide data confidentiality in deduplication. A user uses original data copy to derive a convergent key and encrypt the data by using convergent encryption. User derives tag for each data copy. such that to detect duplicates tag will be used. If the two data copies are similar then tags are also same. The user first sends the tag to the server side to check if the identical copy has been already stored for detect duplicates.[2].

3. Proof of Ownership

The notion of proof of ownership (PoW) [3] enables users to prove their ownership of data copies to the storage server. Specifically, Proof of ownership is implemented as an interactive algorithm run by a user and a storage server.

4. Identification Protocol

The identification of protocol having two phases as follows: 1. Proof: The user can demonstrate his identity to a verifier by performing some identification proof related to his identity. 2. Verify: The verifier occurs verification with input of public information.

III. LITERATURE SURVEY

Following are the different methods which are used in secure data deduplication in cloud storage are presented in Table 1.

TABLE 1: COMPARATIVE STUDY OF SECURE AUTHORIZED DEDUPLICATION FOR HYBRID CLOUD STORAGE

Sr. No.	Author Name	Paper Title and Year	Description	Advantages	Disadvantages
1.	P. Anderson and L. Zhang	Fast and secure laptop backups with encrypted deduplication, 2010 [4]	It provides the potential to decrease storage requirements and backup times.	• Reduces storage requirements • Provides Confidentiality by supporting client end per user encryption.	• Backing up directly to cloud is very cosly. • Network bandwidth can be bottleneck.
2.	Bellare, S. eedhi, and T.	Dupless: Server aided Encryption for	DupLESS work transparently on top of	• High Performance • Space Saving	Failed to secure brute force attacks

	ıpart	deduplicatedstorage, 2013 [5]	any Storage interface	• Provide strong security against External attacks. • Resolve the cross user deduplication	
3.	PasqualoPuzio, RefikMolva ,MelekOnen	CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage,2013 [6]	Deduplication unfortunately come with a high cost in terms of new security and privacy challenges	• Provides Confidentiality	Does not impact the overall storage and computational Cost
4.	Iuon –Chang Lin, Po-chingChien	Data Deduplication Scheme for Cloud Storage,2012 [7]	It improves the speed of data.	1.Improves the speed of data deduplication 2.The Signature is computed for uploaded file for verifying the integrity of files.	The problem is of the worst case in that cloud storage server will regard all blocks as a new blocks and store all of these blocks, resulting in storing duplicate blocks.
5.	Halevi, D. k, B. Pinkas, A. Shulman-	Proofs of ownership in remote storage systems,2011 [8]	We identify attacks that exploit client side deduplication and attempts to identify reduplication. We implemented a prototype of the new protocol and ran it to evaluate performance and asses the Pow scheme benefits.	• Time Saving • Identify attacks • Rigorous security • Saving bandwidth	It is impossible to verify experimentally the assumption about the input distribution

6.	M. Shyamala Devi, V.VimalKhanna,NaveenBalaji	Enhanced Dynamic whole file De-duplication(DWFD) for space optimization in private cloud storage backup, 2014 [9]	DWFD scheme is designed To optimize the private cloud storage backup in order to provide high throughput to the users of the organization by increasing the de-duplication efficiency	• Improve the private cloud backup • Storage efficient • Reduce the de-duplication time	It is not sufficient to development of chunk level deduplication and block level reduplication
7.	JiaXu, Ee-Chien Chang and Jianying Zhou	Weak leakage – resilient client side Deduplication of encrypted data in cloud storage,2013 [10]	A secure client side deduplication scheme.	• Addressed a security concern in cross-user • Client side deduplication	Custom encryption and Convergent encryption methods are semantically unsecure.

IV. CONCLUSION

The advanced research work in the fields of Cloud Computing has taken the field on a maturity level which is leading towards a very productive phase. This means the topics of cloud computing have become interesting to work on and also the issues of cloud computing are handled and addressed. But yet, cloud computing is still as much as a research topic to many researchers. In this paper we reviewed the deduplication techniques for better confidentiality and security in cloud computing. The detection of redundant data and removal of this redundant data is an important task for keeping the cloud storage clean and scalable. This duplicate data elimination has a great advantage for cloud storage. We have surveyed various techniques for deduplication.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart.Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296– 312, 2013.

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

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart. “Dupless: Server aided encryption for deduplicated storage”. In *USENIX SecuritySymposium*, 2013.

[6] PasqualoPuzio, RefikMolva ,MelekOnen ,”CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage”, SecludIT and EURECOM,France.

[7] Iuon –Chang Lin, Po-chingChien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)

[8] 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.

[9] M. Shyamala Devi, V.VimalKhanna,NaveenBalaji ”Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup”,IACSIT, August,2014.

[10] JiaXu, Ee-Chien Chang and Jianying Zhou ,Weak Leakage-Resilient Client –Side deduplication of Encrypted Data in Cloud Storage

BIOGRAPHY

Sulakshana S. Patangeis P. G. Scholar in the Computer Engineering Department, JSPM's RajashriShahu School of Engineering and Research, Narhe, Pune, India She has received Bachelor of Engineering (B.E.) in Computer Engineering from Pune Institute Of Computer Technology (Pune University),India. She is currently working as Lecturer in MIT Polytechnic, Kothrud, Pune. Her research interests are Securityand software engineering.

Prof. Ganesh V.Kadamis a full time Assistant Professor at Department of Computer, JSPM's RajashriShahu School of Engineering and Research, Narhe, Pune, India.

DOI: 10.15680/IJIRCCE.2015. 0309143