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 6, June 2015

Encrypted Data Storage with Deduplication Approach on Twin Cloud

Shweta D. Pochhi[1], Prof. Pradnya V. Kasture[2]

M.E. Student, Dept. of Computer Engineering, RMD Sinhgad School of Engineering, Pune University, India[1]

Assistant Professor, Dept. of Computer Engineering, RMD Sinhgad School of Engineering, Pune University, India[2]

ABSTRACT: Deduplication of Data is a technique for reducingthe amount of storage space an organization needs to save itsdata. It is a specialized data compression technique for removingduplicate copies of repeating data. To protect the discretion ofsensitive data while supporting deduplication, the CE convergentencryption technique has been proposed to encrypt the databefore outsourcing. The differential privileges of users are consideredin duplicate check besides the data itself which has notdone in traditional deduplication systems. The proposed systemsupports authorized duplicate check in a hybrid cloud architecture.Proposed authorized duplicate check scheme acquiresminimal overhead compared to normal operations. To reduce theweaknesses of convergent encryption, we are proposing LFSR(Linear Feedback Shift Register) encryption technique. Securityanalysis determines that this system is secure in terms of thedefinitions specified in the proposed security model.

KEYWORDS: Deduplication, Convergent Encryption, Proof of ownership, Authorized Duplicate Check,Differential Authorization

I. INTRODUCTION

Cloud computing provides a scalable, low cost, location independent infrastructure for data management and storage. Cloud computing provides infinite virtualized resources to users as services across the entire Internet, while hiding platform and implementation details. CSP (cloud service providers) deal at relatively low costs with both highly available storage and especially parallel computing resources. One major challenge of cloud storage services is the management of the ever-growing volume of data as cloud computing becomes universal, an amount of data is being stored and common by users with specified privilege, which define the access rights of the stored data The rapid implementation of Cloud services is convoyed by increasing volumes of data stored at remote servers hence techniques for saving disk space and network bandwidth are needed. A central up and coming concept is deduplication where the server stores a single copy of each file, in spite of how many clients asked to store that file. All clients that store the file simply use links to the single copy of the file stored at the server. Furthermore, if the server already has a copy of the file then clients do not even need to upload it again to the server, thus saving bandwidth as well as storage. In a usual storage system with deduplication, a client first sends to the server only a hash of the file and the server checks if that hash value already exists in its database. If the hash is not in the database then the server scans for the entire file. Otherwise, since the file already exists at the server, it tells the client that there is no need to send the file itself. Either way the server marks the client as an owner of that file, and from that point on there is no difference between the clientand the original party who has uploaded the file. The client can therefore ask to restore the file, regardless of whether he was asked to upload the file or not.

Data deduplication has certain benefits to Eliminating redundant data can extensively shrink storage requirements and increase bandwidth efficiency. Since primary storage has gotten cheaper over time, typically store many versions of the same information so that new workers can reuse earlier work done. Some operations like backup store extremely redundant information. Data deduplication is data compression technique for eliminating duplicate copies of repeating data in storage. This technique is used to improve storage utilization and can also be applied to network data transfers to decrease the number of bytes that must be sent. Deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy instead of keeping multiple data copies with the same content. Deduplication can take place at the file level and the block level. It eliminates duplicate copies of the same file at file level. And eliminates duplicate blocks of data that occur in non-identical files at the block level.

II. RELATED WORK

MihirBellare [2] the encryption key depends on the valueof the plaintext, an attacker who has access to the storagecan perform the dictionary attacks by comparing the ciphertexts resulting from the encryption of well-known plaintextvalues from a dictionary with the stored ciphertexts. Indeed,even if encryption keys are encrypted with users private keysand stored somewhere else, the potentially malicious cloudprovider who has no access to the encryption key but hasaccess to the encrypted blocks can easily perform offlinedictionary attacks and discover expected files.

Yuan et al.[4] proposed Secure Deduplication permits deduplication ofboth files and their equivalent authentication tags. Storagededuplication and data integrity auditing are achieved simultaneously.Here propose the public and constant cost storageintegrity auditing scheme with secure deduplication whichcan also efficiently handle multiple auditing requests withbatch operations. Data integrity and storage efficiency areimportant requirements for cloud storage. POR (Proof ofRetrievability) and PDP (Proof of Data Possession) techniquesassure data integrity for cloud storage. Proof of Ownership(POW) increases storage efficiency by securely removing unnecessarilyduplicated data on the storage server. To reach bothdata integrity auditing and storage deduplication, insignificantsolution is to directly combine an existing POR/PDP schemewith a POW scheme. This solution will result in a O(W)storage overhead for each file where W is the number ofowners of this file because the data owners who are lackingmutual trust, need to separately store their own authenticationtags in cloud for file integrity auditing.

Bellare et al[5] showedhow to protect the data confidentiality by transforming thepredictable message into unpredictable message to enhance thesecurity of deduplication and protect the data confidentiality.Here, another third party called key server is introduced togenerate the file tag for duplicate check. This system also formalizedprimitive as messagelocked encryption, and exploredits application in space-efficient secure outsourced storage.

Stanek et al[6] explained encryption scheme that providesdifferential security for popular data and unpopular data.Popular data are not particularly sensitive hence the traditionalconventional encryption is performed. Convergent encryptionenables duplicate encrypted files to be recognized as identical,but there remains the problem of performing this identificationenables duplicate encrypted files to be recognized as identical, but there remains the problem of performing this identification across a large number of machines in a robust and decen-tralized manner. Convergent encryption purposely discloses information. Some Other research has considered unintentional leaks through side channels such as computational timing, measured power consumption, or response to injected faults.

Xu J. [7] proposed a formulation for client-side deduplication of encrypted files. The system defends confidentiality of users sensitive files against both malicious outside adversaries and honest but curious inside adversaries. Also emphasis on cross-user client-side deduplication over users sensitive data files and protect data privacy from both outside adversaries and the honest-but-curious cloud storage server. It confirms data privacy in deduplication. Also it addressed the problem and showed a secure convergent encryption without considering issues of the key-management and block-level deduplication.

Halevi et al[8] proposed to solve the problem of using a small hash value as a proxy for the entire file, we want to design a solution where a client shows to the server that it indeed has the file. The aim of this paper is to confirm that the file does not have a small representation that when leaked to an attacker allows the attacker to obtain the file from the server. Ideally, we would like the smallest representation of the file to be as long as the amount of entropy in the file itself. The proof of ownership (PoW) for deduplication systems so that a client can efficiently prove to the cloud storage server that he/she owns a file without uploading the file itself.

Bugiel et al[10] recommend an architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. In this, the user converses with a trusted cloud either a private cloud or built from multiple secure hardware components which encrypts and validates the data stored and operations performed in the untrusted commodity cloud. Here the system divides the computations such that the trusted cloud is frequently used for security

DOI: 10.15680/ijircce.2015.0306115

serious operations in the less time critical setup phase whereas queries to the outsourced data are processed in parallel by the fast commodity cloud on encrypted data. This system also provided an architecture consisting of twin clouds for secure outsourcing of data and random computations to an untrusted commodity cloud. It also presented the hybrid cloud techniques to support privacy aware data rigorous computing. To address the authorized deduplication problem over data in public cloud. The security model of systems is similar to those related work where the private cloud is assume to be sincere but curious. Z. Wilcox-OHearn and B. Warner[12] proposed a Tahoe which is a storage grid designed to provide secure long- term storage such as for backup applications. It consists of user space processes running on commodity PC hardware and communicating with one another over TCP/IP. Tahoe was designed the Principle of Least Authority each user or process that needs to accomplish a task should be able to perform that task without having or wielding more authority than is necessary.

III. PROPOSED ALGORITHM

To cope with the convergent encryption weaknesses by using LFSR(Linear Feedback Shift Register) encryption technique including block level deduplication which preserves confidentiality and privacy even against potentially malicious cloud storage providers.

A. *Goals and objectives:*

The below objectives listed to solve the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system

- Covergent Key Issue

To overcome convergent encryption issue the LFSR technique for Encryption which preserves the combined advantages of block level deduplication and convergent encryption. It assures block level deduplication and data confidentiality while coping with weaknesses raised by convergent encryption. Block-level deduplication renders the system more flexible and efficient and preserves confidentiality and privacy even against potentially malicious cloud storage providers.

- User Authorization

To perform duplicate check based on his privileges, each authorized user is able to get his/her individual token of his file. Other user will not allow to generate a token for duplicate check out of his privileges or without the aid from the private cloud server.

- Authorized Duplicate Check

To create query for certain file and the privileges he/she owned with the help of private cloud , authorized user is able to use his/her individual private keys, while the public cloud performs duplicate check directly and tells the user if there is any duplicate.

B. *System Architecture*:

The proposed system is divided into three sections: Client, Public cloud where the client likes to outsource the data and the Private cloud where the token generation will be performed for each file. Before uploading the data or file to public cloud, the client will send the file to private cloud for token generation which is unique for each file. Private cloud then generate a hash and a token and send the token to client. Token and hash keep in the private cloud itself so that whenever next file comes for token generation, the private clod can refer the same token. Once Client gets token foe a particular file, public cloud search for the similar token if it exists or not. If the token exist public cloud will return a pointer of the already existing file otherwise it will send a message to upload a file. Public cloud breaks the fie into blocks and generate the key with Linear feedback Shift Register (LFSR) technique. The token and tag generate on public cloud which will then send to private cloud to update that the token has been generated for the particular file.

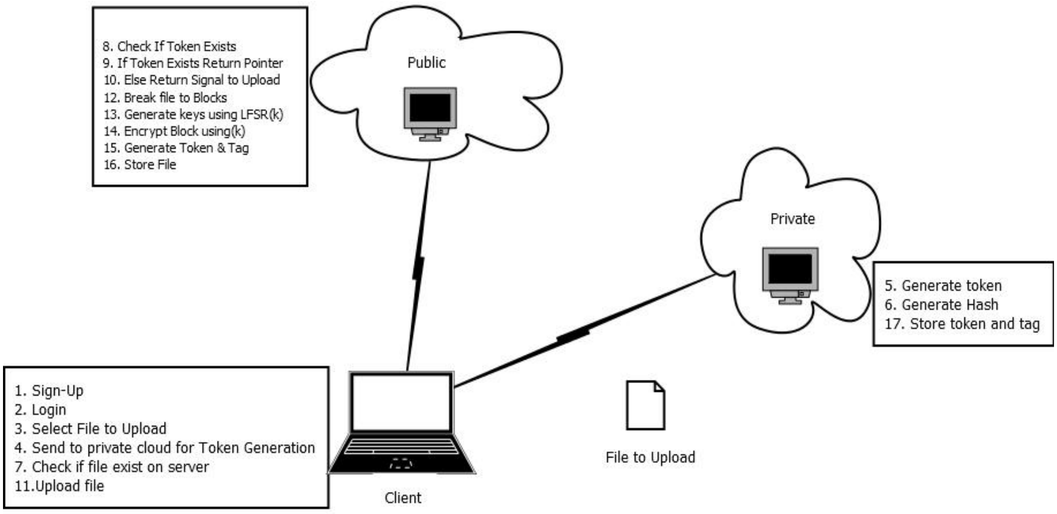

Fig 1: Architecture

IV. Algorithms

Algorithm 1: Advance Encryption Standard with LFSR technique.

```
Cipher(byte in[16], byte out[16], key array round key[Nr+1])
begin

byte state[16]; state = in;
LFSR(round key, tap); AddRoundKey(state, round key[0]); for i = 1 to Nr-1 stepsize 1 do
SubBytes(state);

ShiftRows(state);

MixColumns(state); AddRoundKey(state, round key[i]); end for

SubBytes(state);

ShiftRows(state); AddRoundKey(state, round key[Nr]); end
```

LFSR function :

```
uint16 t start state = 0xACE1u; uint16 t lfsr = start state; unsigned bit;

unsigned period = 0; do

{

bit = ((lfsr 0)^(lfsr 2)^(lfsr 3)^(lfsr 5) lfsr = (lfsr 1) j (bit 15 );
++period;

} while (lfsr 6= start state); return 0;
```

Algorithm 2: MD5:

Step 1.Appending Padding Bits. The original message is ”padded” (extended) so that its length (in bits) is congruent to 448, modulo 512.

Step 2.Appending Length. 64 bits are appended to the end of the padded message to indicate the length of the original message in bytes.

Step 3.Initializing MD Buffer. MD5 algorithm requires a 128-bit buffer with a specific initial value.

Step 4.Processing Message in 512-bit Blocks.which loops through the padded and appended message in blocks of 512 bits each. For each input block, 4 rounds of operations are performed with 16 operations in each round.

Step 5.Output. The contents in buffer words A, B, C, D are returned in sequence with low-order byte first.

V. RESULTS

The system will improve the memory usage while storing the file in cloud storage since block level deduplication points each duplicated blocks to existing data in cloud. Instead of checking throughout a file, block level deduplication will divide the file into certain blocks and identical blocks will be pointed to existing data stored on the cloud. This system will achieve the protection against COF and LRI attacks.

Fig. 2Storage space required without deduplication. The file stored as it is on the server. However, it does not check the duplicate contents while storing a file. It shows a directed graph with file size and storage area.

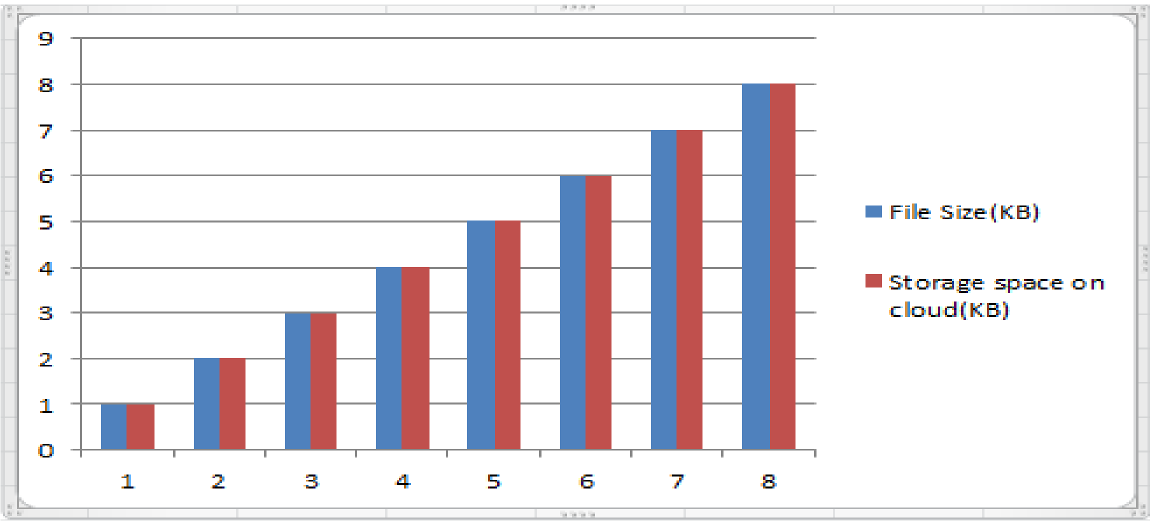

Fig 2: Storage space without deduplication

Fig. 3 describes the storage space required with deduplication. System will improve the memory usage while storing the file in cloud storage. (withdeduplication) .

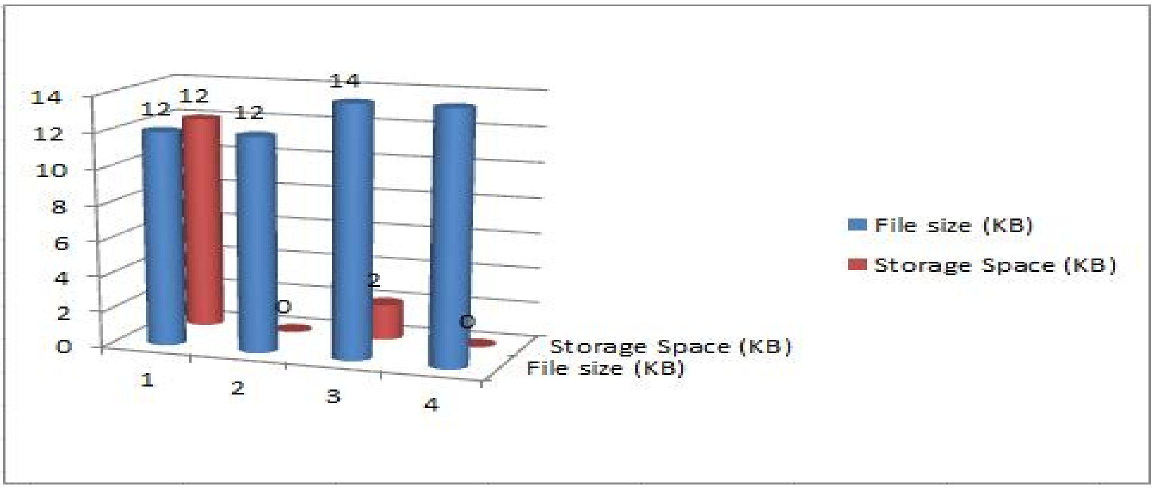

Fig 3: Storage space without deduplication

VI. Conclusion and Future Work

We are implementing a system which achieves confidentialityand enables block-level deduplication at the same time.This system is built on top of convergent encryption withLFSR (Linear Feedback Shift Register) encryption technique.A shift registers identifying function is shifting its contentsinto adjacent positions within the register or, in the case ofthe position on the end, out of the register. The positionon the other end is left empty unless some new contentis shifted into the register. The contents of a shift registerare usually thought of as being binary, that is, ones andzeros. We are performing block-level deduplication instead offile level deduplication since the gains in terms of storagespace. Block-level data deduplication operates on the subfilelevel. As its name implies, the file is typically brokendown into segments chunks or blocks that are examined forredundancy vs. previously stored information. Authorized datadeduplication was proposed to protect the data security byincluding differential privileges of users in the duplicate check.The duplicate-check tokens of files are generated by the privatecloud server with private keys.

References

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

[2] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart," Message-locked encryption and secure deduplication".Springer, In Advances in Cryptology EUROCRYPT 2013, pages 296312, 2013.

[3] Perttula. Attacks on convergent encryption. http://bit.ly/yQxyvl.

[4] J. Yuan and S. Yu., "Secure and constant cost public cloud Storage auditing with deduplication". IACR Cryptology Print Archive, 2013:149, 2013.

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart."Dupless: Serveraided encryption for deduplicated storage". In USENIX Security Symposium, 2013.

[6] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. "A secure data deduplication scheme for cloud storage". In Technical Report, 2013.

[7] J. Xu, E. -C. Chang, and J. Zhou. "Weak leakage-resilient client-side deduplication of encrypted data in cloud storage". In ASIACCS, pages 195206, 2013.

[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 Communica-tions Security, pages 491500. ACM, 2011.

[9] W. K. Ng, Y. Wen, and H. Zhu. Pr,"ivate data deduplication protocols in cloud storage". In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441446. ACM, 2012.

[10] 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.

[11] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: privacyaware data intensive computing on hybrid clouds". In Proceedings of the 18th ACM conference on Computer and communications security, CCS11, pages 515526, New York, NY, USA, 2011. ACM.

[12] Z. Wilcox-OHearn and B. Warner." Tahoe: the least authority filesystem". In Proc. of ACM StorageSS, 2008.

BIOGRAPHY

ShwetaPochhi is ME Computer Engineering student at RMDSinhgad School of Engineering, Warje, Pune-58

Prof. Pradnya V. Kasture is Assistant Professor (Computer Department)with RMD Sinhgad School of Engineering, Warje,Pune-58