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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

A Survey on Efficient Duplication Framework for Distributed Network Storage System

Tanmay Karande, Amol Gholve, Pankaj Taleng, Ankush Maratkar

Diploma Students, Dept. of Computer Engineering, Dr.D.Y.Patil. Polytechnic Akurdi, Maharashtra, India

ABSTRACT: Data deduplication is the most important Data abbreviate techniques used for to removing the duplicate copies of repeating data and it is mostly used in the cloud storage purpose of reduce the storage space and also save the bandwidth. To keep the initimate of hypersensitive data while supporting the deduplication, to encrypt the data before outsourcing convergent encoded technique has been proposed. To better protect the data security, this project makes the first attempt to formally address the problem of authorized redudent of data. Different from the traditional deduplication system, differential advantages of the user are further considered the duplicate check besides the data itself. The hybrid cloud architecture contains many new deduplication constructions supporting authorized duplicate check. The proposed security models contain the demonstration of security analysis system. As a proof of concept, contains the implementation framework of proposed authorized duplicate check scheme and handlling test bed experiments using these prototypes. In proposed system contain authorized duplicate check scheme obtain minimal overhead compared to normal operations.

KEYWORDS: Deduplication, authorized duplicate check, confidentiality, hybrid cloud, Proof of ownerships.

I. INTRODUCTION

Cloud computing provides limitless virtualized recourse to user as services across the whole internet while hiding the all platform and implementing details. Cloud storage service is an management of evergreen increasing mass of data. To make a data management scalable in cloud computing, deduplication has been a conventional technique. Data compression technique is mostly used for eliminating the duplicate copies of repeated data in cloud storage to reduce the data duplication. This technique is used for the improve storage utilization and also applied to network data transfers to reduce number of bytes that is sent. Keeping the multiple data copies with similar content, deduplication eliminates redundant data by keeping one physical copy and refer other redundant data to that copy. Data deduplication occurs file level as well as block level file. The duplicate copies of identical file eliminate by file level DEDUPLICATION. FOR the block level DUPLICATION, which is eliminates duplicates blocks of data that occur in non-identical files. Although data deduplication takes A MANY ADVANTAGE, security as well as privacy concerns arise as users' sensitive data are capable to the both insider as well as outsider attacks. In the traditional encryption providing data confidentiality, is contradictory with the data deduplication. Traditional encryption requires different users to encrypt their data with own keys.

For making feasible deduplication and maintain the data confidentiality used convergent encryption technique. It encrypts decrypts the data copy with a convergent key, the content of the data copy obtained by computing the cryptographic hash value of. After the data encryption and the key generation process users retain the keys and send the cipher text to the cloud. Since the encryption operation is a determinative and is derived from data content, similar data copies will generate the same convergent key and hence same cipher text. A secure proof of ownership protocol is used to prevent the unauthorized access and also provide the proof to the user regarding the duplicate is found of the same file.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

II. RELATED WORK

Following are the secure primitive used in secure deduplication.

A. SYMMETRIC ENCRYPTION

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

• KeyGen SE (1x)^ k is the key generation algorithm that generates k using security parameter 1x;

• Enc SE (k, M)^ C is the symmetric encryption algorithm that takes secret k, and message M and then outputs ciphertext C, and

• Dec SE (k, C) ^ M is a symmetric decryption algorithm that takes the secret k and ciphertext C and then outputs original message M.

B. CONVERGENT ENCRYPTION

Convergent encryption [3], provides the data confidentiality in the deduplication. A user derives a convergent key from the each original data copy and encrypts the data copy with convergent key. In addition, the user also derive tag for the data copy, such that to detect duplicates the tag will be used Here, we assume that the tag holds the property of correctness , i.e., if two data copies are the same, the tags of data also same. The user first sends the tag to the server side to check if identical copy has been already stored for detect duplicates. [2].

C. PROOF OF OWNERSHIP

The notion of proof of ownership (PoW) [7] enables users to prove the their ownership of data copies to the storage server. Specifically, The Proof of the ownership is implemented as a interactive algorithm run by user and storage server.

D. IDENTIFICATION PROTOCOL

The identification of the protocol having two phases as follows:

1. Proof: The user can demonstrate his identity to a verifier by performing some identification proof which is related to the his identity.

Verify: The verifier occurs verification with input of public information.

III. LITERATURE SURVEY

Following are the different types of methods which are used in the secure data deduplication in cloud storage.

A. DUPLESS SERVER-AIDED ENCRYPTION FOR DEDUPLICATED STORAGE

DupLess: Server aided encryption for the deduplicated storage for the cloud storage service provider like Mozy, Dropbox, and others perform deduplication to save the space by only storing one copy of each file uploaded .Message lock encryption is used to resolve the problem of clients to encrypt their file however the saving are lock. Dupless is used to provide secure deduplicated storage as well as the storage resisting brute- force attacks. Clients

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

encrypt under message-based keys obtained from a key-server via oblivious PRF protocol in dupless server. It allow clients to store encrypted data with an existing service, have service are occurs deduplication on their part , and yet achieves strong confidentiality guarantees. It show that encryption for deduplicated storage can be successfully reach the desired performance and space savings close to that of using the storage service with plaintext data [1].

Characteristic:

1. More Security .
2. Easily-deployed solution for encryption that supports the deduplication.
3. User Friendly: Use the command-line client that supports both Dropbox and the Google Drive.
4. Resolve the problem of the message lock Encryption.

B. PROOFS OF THE OWNERSHIP IN REMOTE STORAGE SYSTEMS

It stores only one copy of the duplicate data. Client-side deduplication tries to identify the deduplication chance already at the client and save bandwidth of uploading copies of the existing files to the server[7].To overcome the attacks Danny Harnik, Benny Pinkas, Alexandra Shulman-Peleg and Shai Halevi1 proposes the Proof of the ownership which lets a client efficiently prove to a server that the client keep a file, rather than the just some short information about it present solutions based on the Merkle trees, specific encodings, and analyse their security.[5]

Characteristic:

1. To identify attacks that is exploit the client-side deduplication.
2. Proofs of ownership provide rigorous security.
3. The Rigorous efficiency requirements Peta-byte scale storage systems.

C. A SECURE DEDUPLICATION WITH THE EFFICIENT AND THE RELIABLE CONVERGENT OF KEY MANAGEMENT

Data deduplication is a used for removing duplicate copies of the data, and has been mostly applied in cloud storage to reduce not only the storage space but also upload the banddwidth. Promising it is, an appearing challenge is to accomplish the secure deduplication in cloud storage. Although the convergent encryption has been extensively acquired secure deduplication, a uncertain issue of making the convergent encryption practical is efficiently and reliably manage a huge number of the convergent keys.

Techniques:

1. Key management
2. Convergent Encryption[2]

E. TWIN CLOUDS: AN ARCHITECTURE FOR AN SECURE CLOUD COMPUTING

S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider proposed architecture for the secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. In come towards, user communicates with an trusted cloud. Which is the encrypts as well as the verifies data stored and operations occurred in untrusted cloud .It divide computations such that the trusted cloud is used for the security- critical operations in less time-critical setup phase, whereas queries to the out sourced data which are processed in parallel by fast cloud on the encrypted data [6].

F. PRIVATE DATA DEDUPLICATION PROTOCOLS IN CLOUD STORAGE

The most important issue in cloud storage is utilization of the storage capacity. In this paper, there are two main categories or parts of data deduplication strategy, and extend the fault-tolerant digital signature scheme proposed by Zhang on examining redundancy of blocks to achieve data deduplication. The proposed scheme in this paper not only reduces cloud storage capacity, but also it improves speed of the techniques to reduce the deduplication in cloud

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

storage and maintain the security data deduplication. Furthermore, the signature is computed for the each and every uploaded file for verifying the integrity of files.[4]

Author	Method	Feature	Result
M.Bellare,S.Keelveedhi.and T.Ristenpart	DupLESS ServerAidded Encryption for the Deduplicated Storaged	• Space Saving • Resolve cross user deduplication • Security:Provide strong security against External attacks	Simple storage Interface
S.Halevi,D.Harnik,B.Pinkas,and A.ShulmanPeleg	Proofs of Ownership in Remote Storage Systems	• Time Saving • Rigorous security • Identify the attacks Saving bandwidth	Performance Measurments indicate that scheme incures only a small overhead compared to the nativ clientside deduplication.
J.Li,X.Chen,M.Li,J.Li,P.Lee,and W.Lou	A Secure deduplicationwith the efficient and reliable convergent key management	• Reduce • Storage space and bandwidth • Efficient • Reliable key	Convergent key share across the multiple server.
S.Bugiel,S.Nurnberger,A.Sadeghi,and T.Scheider	Twin clouds: An architecture for secure cloud computing	• Secure computation • Store large amount of the data • Low latency Secure expectation environment	Client user trusted Cloud as a proxy that provide a clearly defined interface to manage the outsourced data, programs, and the queries.
W.K.Ng,Y.Wen,and H.Zhu	Private data deduplication Protocol Is in cloud storage	• Improve the speed of data duplication • Fault tolerant Reduce the cloud storage capacity	Enhance efficiency of the data

Table 1: Comparison Of the Different Methods Of Data Deduplication In Cloud Storage

- DupLESS Server-Aided Encryption for Deduplicated Storage is used for simple storage interface and also provides strong security against the external attacks like brute force attack. It provides the high performance as well as the resolves cross user duplication.
- Proof of ownership presents Performance measurements indicate that scheme obtain only a small overhead compared to naive client-side deduplication. And It identifies attacks and saving bandwidth.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

- A Secure deduplication with the efficient and the reliable convergent key management for reduces the storage space and bandwidth. The Convergent key share across the multiple server.
- Twin Clouds: An architecture for secure cloud computing contain the Client which uses the trusted Cloud as a proxy that is provides a clearly defined interface to the manage outsourced data, programs, and also queries. It having low latency and also provide secure execution environment.
- The Private data deduplication Protocols in cloud storage Enhance efficiency of the data as well as Improve speed of data duplication.

IV. CONCLUSION AND FUTURE WORK

In this survey article proposed secure deduplication with the Help of token generation and Secure upload download.It can be assure the user about high data security and also avoid data deduplication. The security analysis determine that given schemes are very secure in terms of insider as well as the outsider attacks specified in proposed security model. As proof of concept, it executed a prototype of proposed authorized duplicate check scheme and conduct test bed experiments on given prototype.In future by using Cloud Service Provider (CSP) have significant resources to govern distributed cloud storage servers and to the manage its database servers. It also provides the virtual infrastructure to host application services. These services can be used by client to manage his data stored in the cloud servers. The CSP provides a web interface for the client to store the data into set of cloud servers, which are running in a cooperated and also distributed manner. In addition, the web interface is used by the users to retrieve, modify and restore the data from cloud, depending on their access rights. Moreover, the CSP relies on database servers to map client identities to their stored data identifiers and group identifiers.

V. ACKNOWLEDGEMENT

We sincerely thank reviewers for their comments and suggestions that helped us improve our work. Thisresearch has been supported by our Faculty members especially our project guide Prof. Asmita Mohite.

REFERENCES

1. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In USENIX Security Symposium, 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. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message- locked encryption and secure deduplication. In *EUROCRYPT*, pages 296- 312, 2013.
4. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.
5. 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 Security2012.
6. 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.
7. 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-
8. 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*, 2002, pp. 617–624.
9. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in *Proc. of StorageSS*, 2008
10. J. S. Plank, S. Simmerman, and C. D. Schuman, "Jerasure: A library in C/C++ facilitating erasure coding for storage applica- tions-Version1.2,"UniversityofTennessee,Tech.Rep.CS-08-627, August2008
11. M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," *Journal of the ACM*, vol. 36, no. 2, pp. 335–348, Apr. 1989 G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner,
12. Z. Peterson, and D. Song, "Provable data possession at untrusted stores," in *Proceedings of the 14th ACM conference on Computer and communications security*, ser. CCS '07. New York, NY, USA: ACM, 2007, pp. 598–609. [Online]. Available: http://doi.acm.org/10.1145/1315245.1315318
13. H. Shacham and B. Waters, "Compact proofs of retrievability," in*ASIACRYPT*, 2008, pp. 90–107.
14. Amazon, "Case Studies," https://aws.amazon.com/solutions/case- studies/#backup.