IJCSN International Journal of Computer Science and Network, Volume 5, Issue 2, April 2016
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 1.02

Secured Authorized Block Level Deduplication on Hybrid Cloud

[1] Prerana T. Nitnaware, [2] Vikas B. Maral

[1] Department of Computer Engineering, KJ College of Engineering and Management Research
Pune, Maharashtra, India

[2] Department of Computer Engineering, KJ College of Engineering and Management Research
Pune, Maharashtra, India

Abstract - Data deduplication is a specialized data compression technique for eliminating duplicate copies of identical data, and it is mostly used in cloud storage to save bandwidth and reduce the amount of storage space. In many organizations, the storage systems contain duplicate copies of many pieces of data i.e. different users save the same file having same data in several different places. Deduplication eliminates these extra copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. It uses the convergent encryption technique to encrypt the data before outsourcing. For confidentiality we using the concept of Hybrid cloud. The protocol uses authorized duplicate check for hybrid cloud architecture. Different from traditional systems, the differential privileges of users are considered in duplicate check besides the data itself. Our proposed duplicate check scheme incurs minimal overhead compared to normal operations.

Keywords – *Block Level Deduplication, Authorized Duplicate Check, Confidentiality, Hybrid Cloud.*

1. Introduction

Cloud computing provides unlimited ―virtualized‖ resources to users as services across the whole Internet, while hiding platform and implementation details. Cloud storage service is management of ever increasing volume of data. To make data management scalable in cloud computing, deduplication has been standard technique. To reduce the memory management problem and to improve the storage space data de-duplication is an important technique that should be used by cloud computing.

Data de-duplication is done by two ways one is File level and another is Block level. In File level approach we can eliminate the identical files from the storage space and in block level approach we can delete some amount of data i.e. the block of data from the files which are not similar. In this paper we used Block level approach.

When a user uploads a file on the cloud, the file is split into a number of blocks, each block having a size of 2KB. Block is encrypted using a convergent key and subsequently a token is generated for it by using token generation algorithm. After encrypting the data using convergent key, users retain the key before sending the cipher text to the cloud. Due to the deterministic nature of encryption, if identical data copies are uploaded the same convergent keys and the same cipher text will be produced thus preventing the deduplication of data. Each block is then compared with the database of cloud. After comparing, if a match is found in the cloud database then only metadata of the block is stored in DB.

Although Data de-duplication decrease the storage space but the main problem is security of data and privacy of that data from hackers. To secure the data from attacker's users used 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 of data. After the data encryption and key generation process users retain the keys and send the ciphertext 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 ciphertext. A secure proof of ownership protocol is used to prevent the unauthorized access and also provide the proof to user regarding the duplicate is found of the same file.

2. Literature Review

Jin Li and Yan Kit Li presented hybrid cloud approach for secure authorized deduplication. It aims for solving the problem of the deduplication with different privileges in cloud computing [1].

Convergent Encryption: This guarantees information protection in deduplication. Bellare et al. [4] formalized a primary message-locked encryption, and analyzed its application in efficient space secure outsourced capacity storage. Xu et al. [10] additionally tended to the problem and demonstrated a protected convergent encryption for effective encryption, without considering problems of the block level deduplication and key-management. There are likewise different implementations of convergent encryption for secure deduplication. It is realized that some business cloud storage suppliers, for example, Bitcasa, likewise send convergent encryption.

Proof of Ownership: The thought of "Proof of ownership"(PoW) Halevi et al. [8] for deduplication frameworks, such that a customer can effectively prove to the cloud storage server that he owns a record without transferring the record itself. A few PoW developments established on [8] Merkle-Hash Tree is proposed to allow customer side deduplication, which include the delimited leakage setting. Pietro and Sorniotti [9] proposed an alternate PoW plan by selecting the projection of a record onto some randomly chosen bit-positions as the record verification. Note that all the above plans don't consider information security. Newly, Ng et al. [11] enhanced PoW for encryption documents, yet they don't show how to reduce the key management overhead.

Twin Clouds Architecture: Bugiel et al. [7] given a framework comprising of twin cloud for protected outsourcing of information and subjective processing to an untrusted service cloud. Zhang et al. [12] also introduced the hybrid cloud methods to support security conscious data intensive computing. The work considers pointing the authorized deduplication issue over information in public cloud.

3. Design Goal

In this paper, we propose a new deduplication system supporting for

1. Differential Authorization: Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges.
2. Authorized Duplicate Check: 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 duplicate check directly and tells the user if there is any duplicate. For the security of file token, two aspects are defined as unforgeability and indistinguishability of file token. The details are given below.
3. Unforgeability of file token/duplicate-check token: Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the public cloud.
4. Indistinguishability of file token/duplicate-check token: It requires that without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information or the privilege information.
5. Data Confidentiality: Unauthorized users without appropriate privileges or files, including the public cloud and the private cloud server, should be prevented from access to the underlying plaintext stored at public cloud.

4. System Requirement

4.1 Hardware Specification

1. Minimum 2 GB RAM
2. Minimum 30 GB Hard disk

4.2 Software Specification

1. Windows XP/ Windows 7/ Windows 8
2. MySQL
3. NetBeans
4. Java
5. Glassfish 3.1 and above

5. Proposed and Existing Work

5.1 Existing System

Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent; do not support the duplicate check with differential privileges.

In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique.

5.2 Proposed System

In previous de-duplication systems cannot support differential authorization of duplicate check, which is having importance in many of the applications. In such an authorized de-duplication system, each user is issued a set of privileges during system initialization. In this paper, aiming at efficiently solving the problem of deduplication with differential privileges in cloud computing, we consider hybrid cloud architecture.

A. Hybrid Cloud Architecture

Hybrid cloud architecture includes both public cloud and a private cloud. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.

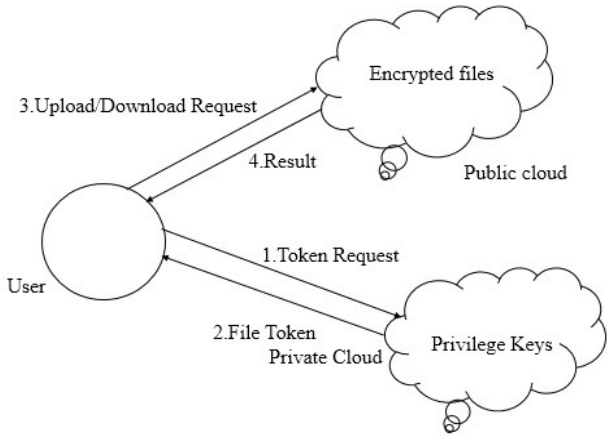

Fig. 1 Architecture for Authorized Deduplication.

B. Proposed Method:

Whenever the user wants to upload and download the file from cloud storage at that time first user request to the cloud server for uploading file. New user should register first, after registration user can login into the system means only authorized user can upload the file to the cloud for that purpose it use the proof of ownership protocol. When file is uploaded it divides into blocks that are block size 2KB by default. The file having all extension which open in notepad. According to file size the block occurs. Each block contain their own cipher text, token for the unique identification and private key. Given block compare with cloud storage if the block is already store in database it store only metadata of block.

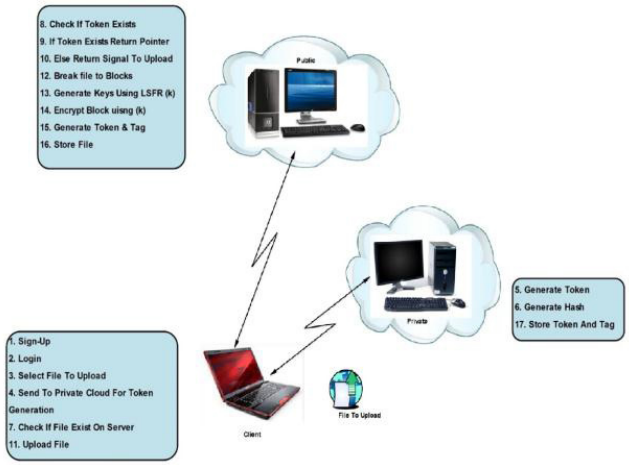

Fig. 2 Architecture of Proposed System.

The above Fig.2 shows the basic architecture of proposed system. Client, Private cloud and Public cloud are different sections of system.

Client (User): In this system user is an entity who wants to outsource data into public cloud and to access the data or files from public cloud. In a storage system supporting deduplication, the user only uploads unique data to save the upload bandwidth. Each file is protected by the convergent encryption key and privilege keys to recognize the authorized deduplication. Each user is assigned with a set of privileges.

Private Cloud: To facilitate user's secure use of cloud services this new entity is introduced. Private cloud manages the private keys for privilege, which supply the file token to users. 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 as the computing resources at data user/owner side are restricted and the public cloud is not fully trusted. This interface allows user to submit files and queries that are securely stored and computed respectively.

Public Cloud: This entity provides the data storage service. It reduces storage cost by using data deduplication technique.

C. Working Methodology:

This system is divided in to two sections one is upload file and another is download file.

Methodology for File Upload:

1. New User should register first.
2. User should Login to enter into the system.
3. User selects the file to upload.
4. File tag generation using Private cloud thus private cloud generate token.
5. After Token generation private cloud server generates Hash and Send it to public cloud server using SOAP connection.
6. After token generation, the next stage is to check whether file exist on server or not.
7. Public cloud server checks If token exist or not.
8. If token exists then it will return pointer.
9. If token does not exist then public cloud server return signal to upload the file on the server.
10. User Upload the file on the server.
11. When file is uploaded it divides into blocks.
12. According to file size the block occurs. Each block contain their own cipher text, token for the unique identification and private key.
13. According to the key information each block is encrypted.
14. After encryption of block Public cloud Generate token and tag.
15. Public cloud store the file.
16. Private cloud stores token and tag.

Methodology for File Download:

1. Public cloud return the identification token T to the user.
2. User Ask for file to download to Public cloud.
3. Public cloud checks the privileges of user.
4. If user has privileges it returns file info & decryption key to the user.
5. User sends file info and token to Private cloud.
6. Private cloud verifies the token and return file blocks to the user.
7. User decrypts the block and generates the original block.

6. Implementation

We implement system with data deduplication, in which we model three entities as separate programs. A Client program is used to model the data users to carry out the file upload/download process. A Private Server program is used to model the private cloud which manages the private keys and handles the Block token computation. A public cloud Server program is used to model the public cloud which manages deduplication. Followings are function calls used in the system.

- BlockTag(FileBlock) - It computes SHA-1 hash of the File block as file block Tag;
- DupCheckReq(Token) - It requests the Storage Server for Duplicate Check of the file block.
- FileUploadReq(FileBlockID, FileBlock, Token) – It uploads the File Data to the Storage Server if the file block is Unique and updates the file block Token stored.
- FileBlock Encrypt(Fileblock) - It encrypts the file block with Convergent Encryption, where the convergent key is from SHA Hashing of the file block;
- TokenGen(File Block, UserID) – the process loads the associated privilege keys of the user and generate token with HMAC-SHA-1.
- FileBlockStore(FileBlockID, FileBlock, Token) - It stores the FileBlock on Disk and updates the Mapping.

7. Result

Proposed system implemented by using block level deduplication which compare the given blocks with database, suppose the file is already stored in the database and that same file uploaded by another user at that time only metadata of file will be store not actually file so it reduce the storage space of data and proper utilization of space. The data will be store in encrypted format so it also maintains security because each block contains their own token, cipher text and private key. The database size will be reduced by using this technique. The proposed system has been compared with the existing system on the basis of database usage, and security using proof of ownership.

8. Future Scope

It saves the memory by dedupicating the data and thus provides us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data. In future we can use HADOOP.

9. Advantages and Limitations

Block level deduplication deliver the benefit of optimizing storage capacity. Lower block size provide more accurate de-duplication check results. However it has some limitations like, storing unique IDs in an index can slow the inspection process as it grows larger. For text files, we have checked de-duplication for different block sizes like 2kb, 4kb, and 8kb. As we reduced the block size number of block increases and hence mapping time increases.

10. Conclusion

In this paper secure deduplication occurs with the help of token generation and secure upload/download of file. It assures the user about high data security and also avoids data duplication in cloud storage. The concept of authorized data deduplication to protect the data security by providing differential privileges of users in the duplicate check. To support authorized duplicate check in hybrid cloud architecture new de-duplication constructions is provided, in which private cloud server is used for the duplicate check, 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. The proposed system aims to have minimal overhead in entire upload and download process and is negligible for moderate file size.

References

[1] Jin Li and Yan Kit Li ˙A Hybrid cloud approach for secure authorized deduplication, IEEE Transaction on parallel and distributed system, 2014.

[2] Anderson, Paul, and Le Zhang. "Fast and Secure Laptop Backups with Encrypted De-duplication." LISA. 2010.

[3] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "DupLESS: server-aided encryption for deduplicated storage." Proceedings of the 22nd USENIX conference on Security. USENIX Association, 2013.

[4] Bellare, Mihir, Sriram Keelveedhi, and Thomas Ristenpart. "Message-locked encryption and secure deduplication." Advances in Cryptology–EUROCRYPT 2013. Springer Berlin Heidelberg, 2013. 296-312.

[5] Bellare, Mihir, Chanathip Namprempre, and Gregory Neven. "Security proofs for identity-based identification and signature schemes." Journal of Cryptology 22.1 (2009): 1-61.

[6] Li, Jin, et al. "A Hybrid Cloud Approach for Secure Authorized Deduplication."

[7] Bugiel, Sven, et al. "Twin clouds: An architecture for secure cloud computing." Proceedings of the Workshop on Cryptography and Security in Clouds Zurich. 2011.

[8] Halevi, Shai, et al. "Proofs of ownership in remote storage systems." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.

[9] Di Pietro, Roberto, and Alessandro Sorniotti. "Boosting efficiency and security in proof of ownership for deduplication." Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. ACM, 2012.

[10] 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.

[11] Ng, Wee Keong, Yonggang Wen, and Huafei Zhu. "Private data deduplication protocols in cloud storage." Proceedings of the 27th Annual ACM Symposium on Applied Computing. ACM, 2012.

[12] Zhang, Kehuan, et al. "Sedic: privacy-aware data intensive computing on hybrid clouds." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.

[13] Wang, Cong, et al. "Privacy-preserving public auditing for data storage security in cloud computing." INFOCOM, 2010 Proceedings IEEE, 2010.

Author's Profile:

Prerana T. Nitnaware is pursuing ME in Computer Science and Engineering at the KJ College of Engineering & Management Research, Pune, affiliated with SPPU University, Pune, Maharashtra. She had worked with Centre for Development of Advanced Computing (C-DAC), the premier R&D organization of the Department of Electronics and Information Technology, Pune, India from 2008-2012. She received the BE degree in Computer Science and Engineering from Amravati University, Amravati, Maharashtra in 2007.

Prof. Vikas B. Maral received the BE & MTech degree in Computer Science and Engineering from SPPU University, Pune, Maharashtra in 2004 & 2011 respectively. He is now an Assistant Professor at the KJ College of Engineering & Management Research Pune affilited with SPPU University, Pune, Maharashtra. His current research interests include parallel and distributed computing, operating systems, cloud computing and systems security. He published more than 15 papers in national and International journals and conferences, such as International Conference on Computing, Communication Control and Automation, IEEE 2015, ICAC3, International Conference on Contemporary Computing and Informatics, IEEE 2014, International Conference on Computational Intelligence and Computing Research, IEEE 2013, IJARCSMS, IJCSIT, IJECS, IJSR.