SECURE DEDUPLICATION SCHEME FOR CLOUD STORAGE

Swapnil Patil

M TECH (IT Dept.) SRM UNIVERSITY CHENNAI, INDIA

Mr.K.Venkatesh

ASSISTANT PROFESSOR (Sr.G) SRM UNIVERSITY CHENNAI, INDIA

Abstract- Nowadays, the explosive growth of digital contents continues to rise the demand for new storage and network capacities, along with an increasing need for more cost effective use of storage and network bandwidth for data transfer.

As such, the use of remote storage systems is gaining an expanding interest, namely the cloud storage based services, since it provides cost efficient architectures. These architectures support the transmission, storage in a multi-tenant environment, and intensive computation of outsourced data in a pay per use business model. Security and privacy are among top concerns for the cloud environments. Towards these security challenges,

This paper introduces a new cryptographic method for secure Proof of Ownership, based on the joint use of convergent encryption and the Merkle-based Tree, for improving data security in cloud storage systems, providing dynamic sharing between users and ensuring efficient data deduplication. Our idea consists in using the Merkle-based Tree over encrypted data, in order to derive a unique identifier of outsourced data. On one hand, this identifier serves to check the availability of the same data in remote cloud servers. On the other hand, it is used to ensure efficient access control in dynamic sharing scenarios.

Keywords- Cloud Storage, Data Security, Deduplication, Confidentiality, Proof of Ownership

I.INTRODUCTION

Cloud storage services commonly use de duplication, which eliminates redundant data by storing only a single copy of each file or block. De duplication reduces the space and bandwidth requirements of data storage services, and is most effective when applied across multiple users, a common practice by cloud storage offerings.

Data deduplication in the cloud is a new technology that caters to the rapidly increasing amount of digital data in data storage. Data deduplication is the process of identifying the redundancy in data and then removing it. The resulting unique single copy is stored and will then serve all of the authorized users.

Despite these significant advantages in saving resources, client data deduplication brings many security issues, considerably due to the multi-owner data possession challenges. For instance, several attacks target either the bandwidth consumption or the confidentiality and the privacy of legitimate cloud users. For example, a user may check whether another user has already uploaded a file, by trying to outsource the same file to the cloud

Although the existing schemes aim at providing integrity verification for different data storage systems, data dynamics has not been fully addressed. How to achieve a secure and efficient design to seamlessly integrate these two important components

ISSN: 2278 – 1323

for data storage service remains an open challenging task in Cloud storage.

Although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity.

Second, there do exist various motivations for CSP to behave unfaithfully toward the cloud users regarding their outsourced data status.

In particular, simply downloading all the data for its integrity verification is not a practical solution due to the expensiveness in I/O and transmission cost across the network. Besides, it is often insufficient to detect the data corruption only when accessing the data, as it does not give users correctness assurance for those unaccessed data and might be too late to recover the data loss or damage.

Encryption does not completely solve the problem of protecting data privacy against cloud service provider but just reduces it to the complex key management domain. Unauthorized data leakage still remains possible due to the potential exposure of decryption keys.

The proposed system is using the convergent encryption, i.e., deriving keys from the hash of plaintext. Then, Storer et al. 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.

II.RELATED WORKS

The Proof of Ownership (PoW) is introduced by Halevi. It is a challenge-response protocol enabling a storage server to check whether a requesting entity is the Data owner, based on a short value. That is, when a user wants to upload a data file to the cloud, he first computes and sends a hash value hash=H (f) to the storage server. This latter maintains a database of hash values of all received files, and looks up hash. If there is a match found, then fis Already outsourced to cloud servers. As such, the cloud tags the cloud user as an owner of data with no need to upload the file to remote storage servers. If there is no match, then the user has to send the file data fto the cloud.

This Secure deduplication, referred as hash-as-a proof, presents several security challenges, mainly due to the trust of cloud user's assumption. This Section presents a security analysis of existing PoW schemes.

III.PROPOSED SYSTEM

The Data Owner will create Remote users by specifying username. The Remote user can login with user name and auto generated password. The Data owner browses and upload the Files to Cloud Server while uploading Data Owner generates the Key to each file and it will be stored in cloud server. The file splits into parts it will be stored in encrypted form. The Remote user has to use correct key to access the files and file names. If anyone is wrong then he is detected as attacker.

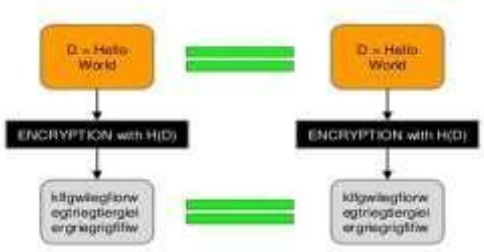

Fig 1.Convergent Encryption

ISSN: 2278 – 1323

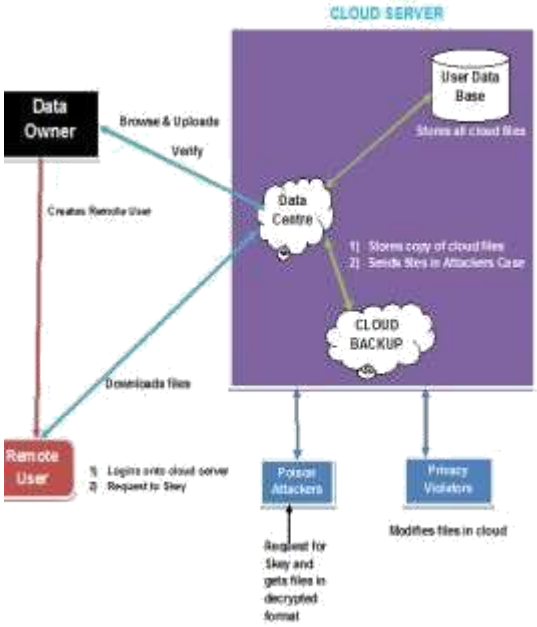

Fig 2.Achitecture

The Attributes are File Management, Cloud Storage, Data Security, Deduplication, Confidentiality, Proof of Ownership, Blocking, unblocking.

This proposed System for secure deduplication Consist of Five Major component Data Owner, Receiver, Attacker, Data Centre, Backup Server

Data Flow –First data owner will creates blocks of file and uploads to cloud server, Second Data centre checks file is safe or not? If it safe takes from user DB otherwise from Backup DB & sends back to user. Third the server uploads the file to both User DB and Backup DB.

Sequence of Operation-

Data Owner request the cloud servers for file operations. When upload start the file is divided into blocks and uploads the blocks into cloud DB. After completing the file operations cloud server sends back the confirmation to the data owner.

After file operation request data owner will check the file for intergrity.it will send a verification request to the cloud server, cloud server will reply back with verification details.

When intended receiver or client will request for file first it will request to the cloud server for file, then cloud server will check the filename or secret key for verification of receiver and send response message to the client, at last it will send back the file to the receiver.

Fig 3. Blocks of hash

IV.IMPLEMENTATION

Data provider-

In this module, the data provider is responsible for creating Remote user by specifying user name; Data provider will auto generate the password. The Data provider uploads their data to the cloud server. For the security purpose the data provider encrypts the data file and then stores in the cloud in parts (splited form). The Data owner can have capable of manipulating the encrypted data file.

ISSN: 2278 – 1323

Cloud Server

The cloud server is responsible for data storage and file authorization for an end user. The data file will be stored in user data base and Backup DB in blocks with their tags such as file name, secret key, hmac1, hmac2, hmac3, hmac4, and hmac5 and owner name. The data file will be sending based on the privileges. If the privilege is correct then the data will be sent to the corresponding user and also will check the file name, end user name and secret key. If all are true then it will send to the corresponding user or he will be captured as attacker.

Cloud Data Backup

Cloud Data backup is nothing but the Backup Database, The data backup process starts when the client requests for retrieving the data previously stored in the cloud. The data backup process includes the following messages:

ClientRequestBackup: it contains the URI of the requested data that the client wants to retrieve. Upon receiving this client request, the CSP verifies the client ownership of the claimed file and generates a ResponseBackup message.

ResponseBackup-

In his response, the CSP includes the encrypted outsourced data. Upon receiving the ResponseBackup message, the client first retrieve the file metadata and deciphers the data decrypting key, using his secret key. Then, he uses the derived key to decrypt the request data file.

Data Consumer (End User)

The data consumer is nothing but the end user who will request and gets file contents response from the corresponding cloud servers. If the file name and secret key, access permission is correct then the end is getting the file response from the cloud or else he will be considered as an attacker and also he will be blocked in corresponding cloud. If he wants to access the file after blocking he wants to UN block from the cloud.

Attacker-

Attacker is one who is integrating the cloud file by adding malicious data to the corresponding cloud.

Privacy violation – In this module the attacker modifies the owner name, therefore sensitive data leakage is a critical challenge. That is, cloud users should have an efficient way to ensure that remote servers are unable to access outsourced data or to build user profiles.

Poison attack – when a data file D is encrypted on the client side, relying on a randomly chosen encryption key, the cloud server is unable to verify consistency between the uploaded file and the proof value hash. In fact, given a pair, the storage server cannot verify, if there is an original data file, that provides a hash value hash. As such, a malicious user can replace a valid enciphered file with a poisoned file. So, a subsequent user loses his original copy of file, while retrieving the poisoned version.

ISSN: 2278 – 1323

V.CONCLUSION

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 these information towards several intrusions. In addition, thanks to the Merkle tree properties, this proposal is shown to support data deduplication, as it employs an pre-verification of data existence, in cloud servers, which is useful for saving bandwidth.

REFERENCES

1. Zheng, Q., Xu, S.: Secure and efficient proof of storage with deduplication. In: Proc. CODASPY 2012, pp. 1–12 (2012)
2. Marques, L., Costa, C.J.: Secure deduplication on mobile devices. In: Proc. OSDOC 2011, pp. 16–29 (2011)
3. Storer, M.W., Greenan, K., Long, D.D., Miller, E.L.: Secure data deduplication. In: Proc. StorageSS 2008, pp. 1–10 (2008)
4. Xu, J., Chang, E., Zhou, J.: Secure Cloud Storage with Encrypted Data using File-Based Authentication. In: IACR (2011), http://eprint.iacr.org/2011/538.pdf
5. Rahumed, A., Chen, H.C.H., Tang, Y., Lee, P.P.C., Lui, J.C.S.: A secure cloud backup system with assured deletion and version control. In: Proc. ICPPW 2011, pp. 160–167 (2011)
6. Anderson, P., Zhang, L.: Fast and Secure Laptop Backups with Encrypted De-duplication. In: Proc. LISA 2010, pp. 29–40 (2010)
7. Douceur, J.R., Adya, A., Bolosky, W.J., Simon, D., Theimer, and M.: Reclaiming space from duplicate files in a serverless distributed file system. In: Proc. ICDCS 2002, pp. 617–624 (2002)
8. Harnik, D., Pinkas, B., Shulman-Peleg, and A.: Side channels in cloud services: deduplication in cloud storage. IEEE Security & Privacy 8(6), 40–47 (2010) CrossRef
9. Goldwasser, S., Micali, S.: Probabilistic encryption & how to play mental poker keeping secret all partial information. In: Annual ACM Symposium on Theory of Computing (1982)
10. Gantz, J.F., et al.: The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010. In: IDC (March 2007)

ISSN: 2278 – 1323