CLIENT SIDE SECURE DE-DUPLICATION SCHEME IN CLOUD STORAGE ENVIRONMENT

Mahejuba Soudagar[1], Rajashekhar D. Salagar[2]

[1]M.Tech Student, Department of Computer Science and Engineering, BLDEA's V.P. Dr.P.G.Halakatti College of Engineering & Technology Vijayapur, Karnataka, India
[2]Assistant Professor, Department of Computer Science and Engineering, BLDEA's V.P. Dr.P.G.Halakatti College of Engineering & Technology Vijayapur, Karnataka, India

---***---

Abstract: *During the last few years, cloud computing technology becomes an eye-catching development of leveraging cloud based services for large scale content storage, processing and distribution. Thus, data de-duplication becomes more and more a essential for cloud service provider. Data de-duplication is a technique for sinking the amount of storage space an institution needs to save its data. Also the top concerns for public cloud are security and privacy. Hence by keeping in mind these security challenges, we propose and implement an Open Stack Swift, a new client-side de-duplication scheme for securely storing and sharing outsourced data via the public cloud. In this technique we have concentrated mainly on two things. First, it ensures better confidentiality towards the users who are not authorized. That is, every client computes as per data key to encrypt the data that he intends to store in the cloud. As such, the data access is managed by the data owner. Second, by integrating access rights in metadata file, an authorized user can decipher an encrypted file only with his private key.*

Key Words**: Cloud Storage, Deduplication, Integrity, Security, Privacy.**

1. INTRODUCTION

Cloud computing provides a low-cost, scalable, location-independent infrastructure for data management and storage, 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. For saving resources consumption in network bandwidth and storage capacities, many cloud services, namely Drop box, wuala and Memopal, apply client side deduplication. As we know the cloud offers many important advantages in various fields of technology like the resource saving, data storage etc, it also provides very ease of usage to the users because they need not worry about their software hardware and the privacy of their data, because cloud provides good privacy to the data stored within it.

Among the endless and countless storage space provided by the cloud service providers, the users can use as much space as they want and the vendors or servers will be continuously trying to keep the unnecessary data to the lowest and minimum and also to maximize the storage space. Technology is changing every day and organizations are expected to adapt to the changes and transform enterprise IT with self-service, charge back, service catalogs, resource orchestration, complete application provisioning hybrid IT, reservations, etc. One of the famous techniques which has been used worldwide is the technique of deduplication. Deduplication defines the technique which stores only a single copy of the file on the storage server provided by the cloud service providers regardless of how many number of clients request to store that particular file in the cloud.

1.1 Types of Cloud Deployment Model.

1)Private Cloud: - Private Cloud as the name indicates is a cloud infrastructure which is exclusively operated for a single organization, it can be maintained by the cloud itself or by a separate third-party. Private cloud is hosted either internally or externally on the outside[1] Undertaking a private cloud project requires a significant level and degree of commitment to virtualized the business atmosphere, and requires the organization to re-examine decisions about present possessions.

2)Public Cloud: - A cloud is called a "public cloud" in which the services offered by the cloud are available over the network that is open and free for public use. Public clouds are usually free of cost and charge[2]. When we see technically there is very little or no difference between the architecture of public cloud and private cloud, but most importantly the security measures are significantly different for both mainly for the services like storage, applications and other resources which are made available by the cloud service provider so that public viewers and users can use them efficiently, it is also used when communication is effected over a non-trusted group.

3)Hybrid Cloud: - Hybrid Cloud is a mixture or composition of two or more same or different clouds which maybe private, public or community, these will remain separate and discrete entities but are still bound together. Hence it offers endless benefits of multiple operation models. Offering the benefits of Hybrid cloud can also mean the ability to connect collocation, managed and/or committed services with cloud resources.

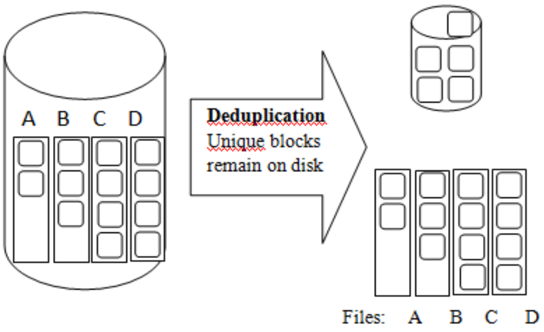

Fig 1: Hybrid Cloud

The data deduplication technique is considered as one of the most important and promising solution in cloud technology because it keeps only a single copy of the repeated or duplicate data. It is very important because it helps in reducing storage space and also improves experience of users by saving the backup time of dropping and also the network bandwidth. De-duplication [3, 4] can take place at either the file level or the block level. In the file level de-duplication, duplicate copies of the same file are eliminated or removed. De-duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. De-duplication can be performed at different locations

Depending on the participating machines and steps in the customized de-duplication process, it is either performed on the client machine (source-side) or near the final Storage-server (target-side). Despite these significant advantages in saving resources, client data de-duplication brings many security issues many efforts have been proposed under different security models, these schemes are called Proof of Ownership systems (PoW).

2. PROOF OF OWNERSHIP

Enable the users to provide their ownership of data copies to the storage server we choose proof of ownership. Proof of ownership is implemented as an interactive algorithm run by a power and verifier. From a data copy of M, the verifier derives a short value Φ(M). To prove the ownership of the data copy M, the user needs to send Φ to the verifier such that. The formal security definition for PoW roughly follows the threat model in a content distribution network, where an attacker does not know the entire file, but has accomplices who have the file. The accomplices follow the "bounded retrieval model", such that they can help the attacker obtain the file, subject to the constraint that they must send fewer bits than the initial min-entropy of the file to the attacker [5].

This paper introduces a new cryptographic method for secure Proof of Ownership (PoW), based on the joint use of convergent and the Merkle-based Tree, for improving data security in cloud storage systems, providing dynamic sharing between users and ensuring efficient data deduplication

3. LITERATURE SURVEY

In 2002, Douceur et al. [6] studied the problem of deduplication in multi-tenant environment. The authors proposed the use of the convergent encryption, i.e., deriving keys from the hash of plaintext. Then, Storer et al. [13] 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.

In order to prevent private data leakage, Halevi et al. [7] proposed the concept of Proof of Ownership (PoW), while introducing three different constructions, in terms of security and performances. These schemes involve the server challenging the client to present valid sibling paths for a subset of a Merkle tree leaves [8].

Recently, Ng et al. [9] propose a PoW scheme over encrypted data. That is, the file is divided into fixed-size blocks, where each block has a unique commitment. The hash-tree proof is then built, using the data commitments. Hence, the owner has to prove the possession of a data chunk of a precise commitment, with no need to reveal any secret information. However, this scheme introduces a high computation cost, as requiring generation of all commitments, in every challenging proof request.

4. PROPOSED SYSTEM

Figure 2 illustrates descriptive network architecture for cloud storage. It relies on the following entities for the good management of client data:

- Cloud Service Provider (CSP): The Cloud service provider has many significant and important resources to manage and maintain the distributed cloud storage servers. It also manages its own database servers. CSP also hosts many application services with the help of virtual storage infrastructure
- Client: Client makes use of provider's resources to store, retrieve and share data with multiple users. A client can be either an individual or an enterprise.
- Users: The end users will be able to obtain and access the contents stored within the cloud depending on the access rights which are the authorizations provided by the client, the authorizations like the rights to read the data, write the data or re-store the modified data within the cloud. These access rights serve to specify several groups of users.

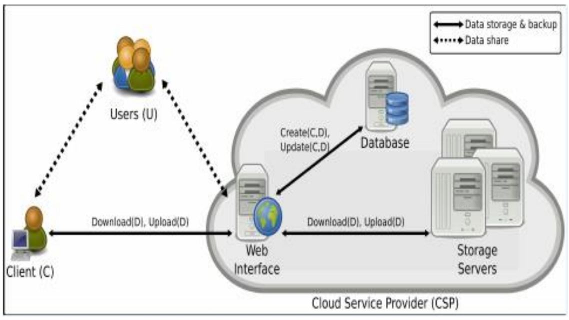

Fig. 2: Architecture of cloud data storage

Advantages of proposed System:

- A new cryptographic method was introduced for secure Proof of Ownership (PoW), based on the joint use of convergent encryption and the Merkle-based Tree for improving data security in cloud storage systems.
- This PoW identifier helps to check that whether the same data is already available in remote cloud serves
- It is used to ensure efficient access control in dynamic sharing scenarios.
- Dynamic sharing between users and ensuring efficient data deduplication.
- The multi-owner data possession challenges. For instance, several attacks target either the bandwidth consumption or the confidentiality
- 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.

5. CONCLUSION

The growing need for secure cloud storage services and the attractive properties of the convergent cryptography lead us to combine them, thus, defining an innovative solution to the data outsourcing security and efficiency issues. Our solution is based on a cryptographic usage of symmetric encryption used for enciphering the data file and asymmetric encryption for metadata files, due to the highest sensibility of this information towards several intrusions.

REFERENCES

[1] Attention, shoppers : Store is tracking your cell, New York Times. References

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

[3] Iuon-Chang and Po-Ching Chien, "Data Deduplication Scheme for Cloud storage". IJ3C, Vol. 1, No. 2 (2012)

[4] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. "Secure data deduplication". In Proc. of StorageSS, 2008.

[5] S. Quinlan and S. Dorward. Venti: "A new approach to archival storage". In *Proc. USENIX FAST*, Jan 2002.

[6] 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 In Proceedings of 22nd International Conference on Distributed Computing Systems (ICDCS, 2002.)

[7] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Proceedings of the 18th ACM conference on Computer and communications security, CCS '11, pages 491–500, New York, NY, USA, 2011. ACM.

[8] R. C. Merkle. A digital signature based on a conventional encryption function. In A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology, CRYPTO '87, pages 369–378, London, UK, UK, 1988. Springer-Verlag.

[9] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC '12, pages 441–446, New York, NY, USA, 2012. ACM.