ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (415-418), Month: April - June 2015, Available at: www.researchpublish.com

Secure Authorized Deduplication by Hybrid Cloud Architecture

Rashmi Muttappanavar

M. Tech Computer Science and Engineering, S.T.J.I.T, Ranebennur, India

Abstract: **Data deduplication is a method of reducing storage need, which includes elimination of redundant data. Only one unique instance of the data is actually retained on storage media. Data deduplication is also known as "intelligent compression" or "single-instance-storage". It has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect confidentiality of sensitive data while supporting deduplication, the encryption technique has been proposed. To better protect data security, we made an attempt to formally address the problem of authorized data deduplication. Different from old deduplication systems, the differential privileges of users are further considered in duplicate check beside the data itself. We presented a new deduplication construction supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our schema is secure in terms of the definitions specified in the proposed security model. We show that our proposed schema of authorized deduplication incurs minimal overhead compared to normal operations.**

Keyword: **Deduplication, authorized duplicate check, hybrid cloud, confidentiality.**

I. INTRODUCTION

Cloud computing is an emerging style of computing where applications, data and resources are provided to users as services over the web. The services provided may be available globally, always on, low on cost, on demand, massively scalable, pay-as-you-grow. Cloud computing is a modern driven technology that provides configurable computing resources such as servers, networks, storage and applications as and when required with minimum effort over the internet services. Nowadays cloud computing is used more and more, there is increase in the amount of data being stored in the cloud and shared by number of different users. It is a big task to manage, ever increasing volume of data. Nowadays utilization of cloud storage capacity is an important issue. When it comes to security, there is an possibility where malicious users can penetrate the cloud by impersonating a legalize user, there by affecting the entire cloud, which further infects other customers who share infected cloud. We have one more problem related to multiple copies of same data, which will lead to waste of bandwidth and storage.

To deal with problems like utilization of cloud storage, security against malicious users, and duplication of data and to make data management scalable, we make use of technology known as deduplication [1]. Deduplication is mainly used by the cloud to reduce storage usage. It deals with duplicate copies of data in storage and avoids repeat ion of same data. It eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. We can perform deduplication either at file level or at block level. We make use of hybrid cloud architecture along with deduplication. Hybrid cloud is a combination of public cloud and private cloud. The hybrid cloud model is the use of both public and private clouds simultaneously and it is an intermediate step in the evaluation process. It offers the best of clouds, the scale and convenience of a public cloud and the control, security and reliability of private cloud.

A. Existing System

Existing system uses traditional encryption techniques. This technique provides data confidentiality but failed to support deduplication. Some old deduplication system use convergent encryption techniques and provides data confidentiality but not efficiently support differential authorization duplicate check. Private cloud only involved as proxy to allow owner/users to perform security check. Data owner only outsource data by public cloud, while data operations are managed in private cloud. Due to the presence of multiple copies of same data, deduplication becomes impossible.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (415-418), Month: April - June 2015, Available at: www.researchpublish.com

Existing systems are less secure and confidential and not support differential authorization duplicate check. Hence we need a system which provides more security, confidentiality and secure authorized deduplication.

B. Proposed System

We proposed a system which provides more security, data confidentiality and deduplication. Our system support differential duplicate check and we use hybrid cloud architecture. We provide security by performing security check and file encryption so that unauthorized user cannot able to decrypt the files. Convergent encryption [2] has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts/decrypts a data copy with a convergent key. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption allows the cloud to perform deduplication on the cipher texts. We also provide systematic way for authorized users to use the system. We added duplicate check process, which avoids unnecessary storage of same data and helps to reduce storage. We included number of security checks, which helps to identify the unauthorized users. We enhance our system in security. Security analysis demonstrates that our system is secure .The use of hybrid cloud architecture makes efficient usage of public and private cloud. Deduplication makes data management scalable.

C. Contributions

In our system, we are aiming at efficiently solving the problem of storage of multiple copy of identical data, providing more security and confidentiality for data in cloud computing. The use of hybrid cloud architecture provides features of both public and private cloud. Differential duplicate check is proposed under hybrid cloud architecture, where storage is provided by public cloud. We enhance our system in security and support security by encrypting file with differential privilege keys so that unauthorized user cannot decrypt the file. Security analysis says that our system is secure in term of definitions specified in proposed security model.

D. Organization

The rest of the paper proceeds as follows. In Section II, we propose system architecture for our deduplication system. In Section III, we discuss about literature review. Finally we draw conclusion in Section IV.

II. SYSTEM ARCITECTURE

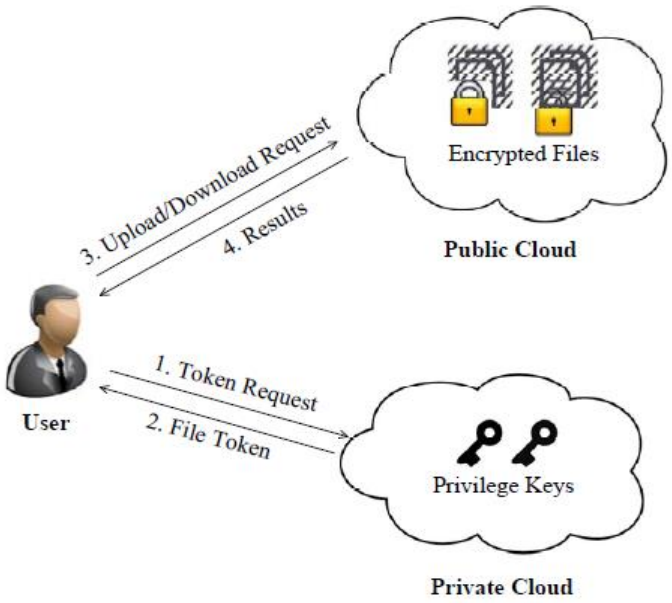

Fig. 1. Architecture of Authorized deduplication (Source: Li et al. [3])

There are three entities defined in our system, that is, users, public cloud and private cloud. The Fig. 1 shows Architecture for Authorized deduplication. Hybrid cloud including both public and private cloud provides storage by public cloud and security keys for users by private cloud. It performs deduplication by checking if the two files are same and stores only one of them. The access right to the file is defined based on set of privileges. We can perform deduplication at file level as well as at block level. In our system we will only consider the file level deduplication. Whenever a user wants to upload a file, then duplicate check is performed. It checks for existence of the file, if file already exits then it cannot upload the file and if not it will upload the files. Each data copy is associated with a token for the duplicate check. In case of download user need to get permission from private cloud in terms of keys, after download user can use the key to decrypt the file because files in public cloud are in encrypted form. The entities of system are

Users: User is one who either uploads the data or retrieves the data. Each user initially needs to be registered so that they can be authorized user. In a storage system to support deduplication, the user only uploads unique data but not upload any duplicate data. In authorized deduplication system, each user is issued with set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

Public cloud: Public cloud is an entity that provides the data storage. User can upload the data to public cloud or download the data from public cloud. To reduce the storage cost, the public cloud eliminates the storage of redundant data via deduplication and keeps only unique data. In our system, we assume that public cloud is always online and has abundant storage capacity and computation power.

Private cloud: Private cloud facilitates users secure usage of cloud service. It is able to provide data users/owner with an execution environment and work as interface between user and public cloud. The private key are managed by private cloud and it responses to all the requests made by users.

III. LITERATURE REVIEW

1. Message-locked encryption and secure deduplication [4]

This formalize a new cryptographic primitive that we call Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud storage providers.

2. Security proofs for identity-based identification and signature schemes [5]

This paper provides either security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly in existing literature. Underlying these is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work. We also analyse a generic folklore construction that in particular yields identity-based identification and signature schemes without random oracles.

3. Twin Clouds: Secure Cloud Computing with Low Latency [6]

We propose an architecture and protocols that accumulate slow secure computations over time and provide the possibility to query them in parallel on demand by leveraging the benefits of cloud computing. In our approach, the user communicates with a resource-constrained Trusted Cloud (either a private cloud or built from multiple secure hardware modules) which encrypts algorithms and data to be stored and later on queried in the powerful but untrusted Commodity Cloud. We split our protocols such that the Trusted Cloud performs security-critical precomputations in the setup phase, while the Commodity Cloud computes the time-critical query in parallel under encryption in the query phase.

4. Secure Data Deduplication [7]

We have developed a solution that provides both data security and space efficiency in single-server storage and distributed storage systems. Encryption keys are generated in a consistent manner from the chunk data; thus, identical chunks will always encrypt to the same cipher text. Furthermore, the keys cannot be deduced from the encrypted chunk data. Since the information each user needs to access and decrypt the chunks that make up a file is encrypted using a key known only to the user, even a full compromise of the system cannot reveal which chunks are used by which users.

5. Convergent Dispersal: Toward Storage-Efficient Security in a Cloud-of-Clouds [8]

Cloud-of-clouds storage exploits diversity of cloud storage vendors to provide fault tolerance and avoid vendor lock-ins. Its inherent diversity property also enables us to offer keyless data security via dispersal algorithms. However, the keyless security of existing dispersal algorithms relies on the embedded random information, which breaks data deduplication of the dispersed data. To simultaneously enable keyless security and deduplication, we propose a novel dispersal approach called convergent dispersal, which replaces original random information with deterministic cryptographic hash information that is derived from the original data but cannot be inferred by attackers without knowing the whole data.

IV. CONCLUSION

We proposed a new deduplication constructions supporting authorized duplicate check by using hybrid cloud architecture. In our system duplicate check tokens of files are generated with private keys by private cloud. Proposed system includes authorization for each user and security check for data retrieval. By doing security analysis we can say that our system is secure in terms of insider and outsider attacks. We showed that our authorized deduplication schema is more secure and helps to improve storage utilization.

V. ACKNOWLEDGMENT

With immense pleasure, we are publishing this paper as a part of the curriculum of M.E. Computer Science and Engineering. We feel great pleasure to acknowledge the guidance and assistance of all those people who have made my work on this paper pleasant endeavour. This paper being an unforgettable and educative experience. It gives us proud privilege to complete this paper work under the valuable guidance of our guide for providing all facilities and help for smooth progress of paper work. We would also like to thank all the Staff Members of Computer Science Department, Management, friends and family members, Who have directly or indirectly guided and helped me for the preparation of this paper and gives us an unending support right from the stage the idea was conceived.

REFERENCES

[1] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST , Jan 2002.

[2] 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 , pages 617–624, 2002.

[3] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. A Hybrid Cloud Approach for Secure Authorized Deduplication M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

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

[8] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

AUTHOR PROFILE:

Rashmi Muttappanavar received the B.E degree in Computer Science from S.I.E.T, Bijapur in 2013 and pursuing M.Tech in Computer Science and Engineering from S.T.I.J.T, Ranebennur.