International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

*Review Article*

# Review paper on Authorized Duplication Checker in Hybrid C cloud

**Priyanka K. Shinde† and Avinash P. Wadhe†**

†Department of Computer Science& Engineering, G.H. Raisoni College of Engineering & Management, Amravati, India

Accepted 20 March 2015, Available online 25 March 2015, **Vol.5, No.2 (April 2015)**

***Abstract***

*A hybrid cloud is a combination of public and private clouds bound together by either standardized or proprietary technology that enables data and application portability. Proposed system aiming to efficiently solving the problem of deduplication with differential privileges in cloud computing. A hybrid cloud architecture consisting of a public cloud and a private cloud and the data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. To make data management scalable in cloud computing, deduplication has been a very well-known technique recently is use. Deduplication reduces your bandwidth requirements, speeds up the data transfers, and it keeps your cloud storage needs to a minimum. Proposed system present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. To maintain the confidentiality of data the convergent encryption technique has been used to encrypt the data before outsourcing. Authorized deduplication system support differential authorization duplicate check. As a proof of concept, a prototype is implemented in authorized duplicate check scheme and conduct test bed experiments using prototype, authorized duplicate check scheme incurs minimal overhead compared to normal operations.*

***Keywords:*** *Deduplication, Authorized duplicate check, hybrid cloud.*

## 1. Introduction

Cloud computing enables new business models and cost effective resource usage. Instead of maintaining their own data center, companies can concentrate on their core business and purchase resources when it will needed. Especially when combining publicly accessible clouds with a privately maintained virtual infrastructure in a hybrid cloud, the hybrid cloud technology can open up new opportunities for businesses. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and the data shared by different users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data on cloud. To make the data management scalable in cloud computing, deduplication QoS (Rahumed, H. C. H. Chen, *et al,* 2011).has been a well-known technique recently use. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes. Instead of keeping multiple data copies with same content, deduplication eliminates the redundant data by keeping only one physical copy and referring other redundant data to that copy. Data deduplication brings a lot of benefits, though security and privacy concerns arise as users sensitive data are susceptible to both insider and outsider attacks. Convergent encryption [3] has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts and decrypts a data copy with a convergent key and which is obtained by computing the cryptographic hash value of the content of the data copy. After the key generation and data encryption, users retain the keys and send ciphertext to the cloud. Since the encryption operation is deterministic and it is derived from the data content, the same convergent key is generated by identical data copies and hence the same ciphertext. To prevent unauthorized access, a secure proof of ownership protocol QoS (S. Halevi, D. Harnik,, *et al,* 2011) is also needed to provides the proof that the user indeed owns the same file when the file duplicate is found and After the proof, subsequent users with the same file provide a pointer from the server without needing to upload the same file. A user can able to 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 will allow the cloud to perform deduplication on the ciphertexts and the proof of ownership prevents the unauthorized user to access the file.

However, the previous deduplication systems cannot support to differential authorization and duplicate check, which is very important in many

*Corresponding author: **Priyanka K. Shinde**

applications. In an authorized deduplication system each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is bounded by a set of privileges which specify which kind of users is allowed to perform the duplicate check and access right of the files. Before submitting his duplicate check request for some file, the user needs to take his file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

## 2. Literature Survey

QoS (Elhadj Benkhelifa and Dayan Fernando *et al.* 2013**)** proposed a novel hybrid solution for increased security to be implemented as part of a real business case project. The project is concerned with highly sensitive data, hence a more complex security approach is needed. The proposed hybrid solution, Single-Sign-On and two-factor authentication, is accepted by the project consortium and end-users to be a state-of-the-art and highly secure authentication approach.

Several deduplication schemes have been proposed by QoS( J. R. Douceur *et al.* 2002 )showing how deduplication allows very appealing reductions in the usage of storage resources. The problems of coalescing and identifying identical files in the distributed file system, for the purpose of reclaiming storage space consumed by incidentally redundant content. Recently QoS (Jan Stanek *et al.* 2013)proposed an encryption scheme that guarantees semantic security for unpopular data and provides weaker security and better storage and bandwidth benefits for popular data. For popular data that is not particularly sensitive, the traditional conventional encryption is performed. For unpopular data another two-layered encryption scheme with stronger security while supporting deduplication is proposed. In this way, they achieved better tradeoff between the efficiency and security of the outsourced data.

Convergent encryption is a cryptographic primitive introduced by QoS( J. R. Douceur *et al.* 2002), attempting to combine data confidentiality with the possibility of data deduplication. Convergent encryption of a message consists of encrypting the plaintext using a deterministic (symmetric) encryption scheme with a key which is deterministically derived from the plaintext and when two users independently attempt to encrypt the same file, they will generate the same ciphertext which can be easily deduplicated but convergent encryption does not provide semantic security as it is vulnerable to content guessing attacks. Later, QoS (Bellare M., Keelveedhi *et. al* 2013) formalized convergent encryption under the name message-locked encryption. As expected, the security analysis presented in QoS (Bellare M., Keelveedhi *et. al* 2013) highlights that message-locked encryption offers confidentiality for unpredictable messages only, clearly failing to achieve semantic security.

QoS( Jia Xu. Chang, E.C., Zhou, J) present a PoW scheme allowing client-side deduplication in a bounded leakage setting. They provide a security proof in a random oracle model for their solution, but do not address the problem of low min-entropy files. Recently, QoS (M. Bellare *et. al* 2013) presented DupLESS a server-aided encryption for deduplicated storage which uses a modified convergent encryption scheme with the aid of a secure component for key generation.

QoS (Hongwei Li *et. al* 2009) presented an identity based authentication for cloud computing which is based on the identity-based hierarchical model for cloud computing (IBHMCC) and corresponding signature and encryption schemes. Being certificate-free and the authentication protocol aligned well with demands of cloud computing. Performance analysis indicates that the authentication protocol is more efficient and lightweight. Recently QoS,( S. Bugiel et. QoS (K. Zhang *et. al* 2011*)* also presented the hybrid cloud techniques to support privacy-aware data-intensive computing. Proposed system consider to address the authorized deduplication problem over data in public cloud.

Security and privacy are among top concerns for the public cloud environments. Towards these security challenges, QoS (Nesrine Kaaniche *et. al* 2012 ) proposed a new client-side deduplication scheme for securely storing and sharing outsourced data via the public cloud. It ensures better confidentiality towards unauthorized users. Every client computes a per data key to encrypt the data that he wants to store in the cloud. As such, the data access is managed by the data owner and also introduces a new cryptographic method for secure Proof of Ownership (PoW), for improving data security in cloud storage systems, providing dynamic sharing between users and ensuring efficient data deduplication.

## 3. Data duplication problem in cloud

Storage efficiency functions such as deduplication afford storage providers better utilization of their storage backends and the ability to serve more customers with the same infrastructure. It is the process by which a storage provider only stores a single copy of a file owned by several of its users and there are four different deduplication strategies, depending on whether deduplication happens at the client side (i.e. before the upload) or at the server side, and whether deduplication happens at a file level or at a block level. Deduplication is most rewarding when it is triggered at the client side, as it also saves upload bandwidth but For these reasons, deduplication is a critical enabler for a number of popular and successful storage services which offers a cheap, remote storage to the broad public by performing client-side deduplication, thus it will saving both the network bandwidth and storage costs. Indeed, data deduplication is arguably one of the main reasons why the prices for cloud storage and cloud backup services have dropped so sharply.

As the world moves to digital storage for archival purposes, there is an increasing demand for systems that can provide a secure data storage in a cost-effective manner. By identifying the common chunks of data both within and between files and storing them only once, by this deduplication can yield cost savings by increasing the utility of a given amount of storage but Unfortunately, deduplication exploits identical content, while encryption attempts to make all content appear random, when the same content encrypted with two different keys results in very different ciphertext. Thus, in encryption combining the space efficiency of deduplication with the secrecy aspects is problematic. Although data deduplication brings a lot of benefits to cloud user, security and privacy concerns arise as users sensitive data are susceptible to both insider and outsider attacks. While Traditional encryption, providing data confidentiality, is incompatible with data deduplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to a different ciphertexts, which makes deduplication impossible. Thus Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible.

## 4. Security issues in cloud

The security will be analyzed in terms of two aspects, that is, the confidentiality of data and the authorization of duplicate check . We suppose that all the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under this assumption, two kinds of adversaries are considered, that is, adversaries which aim to extract secret information as much as possible from both public cloud and private cloud, and internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. The data will be encrypted in our deduplication system before outsourcing to the storage cloud to maintain the confidentiality of data. The data is encrypted with the traditional encryption scheme and The data encrypted with such encryption method which guarantees the security of data. System address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for Differential Authorization and Authorized Duplicate Check. Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any unathorised user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server. 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 the duplicate check directly and tells the user if there is any duplicate.

The security requirements considered in two folds, including the security of data files and security of file token. For the security of file token . Unauthorized users without appropriate privileges or file prevented from getting or generating the file tokens for duplicate check of any file stored at the Storage cloud. The users are not allowed to collude with the public cloud server. It requires that any user without querying the private cloud server for some file token, he cannot able to get any useful information from the token, which includes the privalege or the file information information and to maintain the data confidentiality unauthorized users without appropriate privileges or files, prevented from access to the underlying plaintext stored at Storage cloud.

## 5. Propose authorised duplication checker for clouds

There are three entities defined in system, that is, users, private cloud and storage cloud servise provider in public cloud as shown in Fig. 1. The Storage cloud performs deduplication by checking if the contents of two files are the same and stores only one of them and The access right to a file is defined based on a set of privileges. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denotes the tag with specified privileges. A user computes and sends duplicate-check tokens to the public cloud for authorized duplicate check. While Users have access to the private cloud server, a semitrusted third party which perform duplicable encryption by generating file tokens for the requesting users.

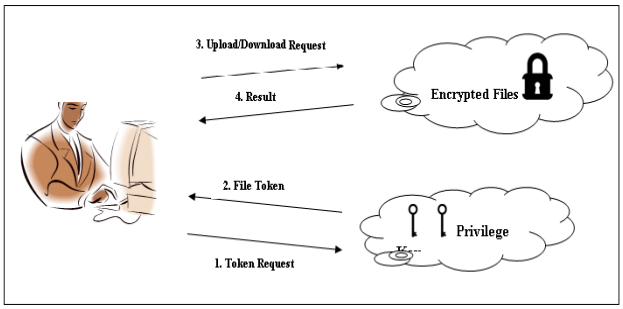


**Fig. 1** Architecture for Authorized Deduplication

### *5.1 Storage Cloud*

This is an entity that provides a data storage service in public cloud. The storage cloud service provider provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the storage cloud eliminates the storage of redundant data via deduplication and keeps only unique data.

### *5.2 Data User*

A user is an entity that wants to outsource data storage to the S-CSP and access the data later when needed. In a storage system supporting deduplication, to save the

upload bandwidth the user can only uploads unique data but does not upload any duplicate data, which may be owned by the same user or the different users. In authorized deduplication system, each user is issued a set of privileges in the setup of the system and Each file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

*5.3 Private Cloud*

The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users and this interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

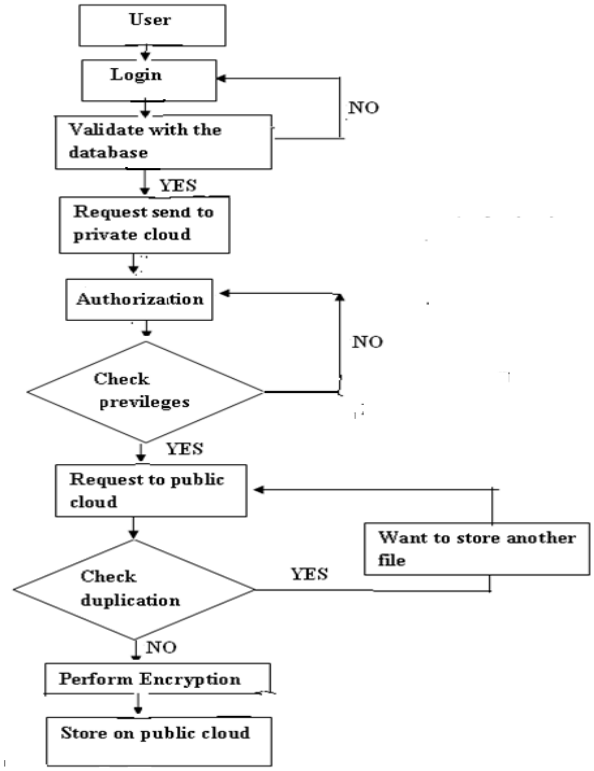


**Fig. 2** Flow Diagram of Proposed Method

In deduplication system, a hybrid cloud architecture is introduced to solve the problem of unauthorized deduplication of file. The private keys for privileges will not be issued to users directly, which will be kept and managed by the private cloud server. The user needs to send a request to the private cloud server to get a file token. The user needs to get the file token from the private cloud server to perform the duplicate check for some file. The private cloud server also check the user's identity before issuing the corresponding file token to the user. The user perform the authorized duplicate check for this file with the public cloud before uploading this file. The user either uploads this file or prove their ownership based on the results of duplicate check. If a file duplicate is found, the user needs to run the Proof of ownership protocol with the cloud storage service provider to prove the file ownership. Otherwise, if no duplicate is found then the data owner performs an identification to prove its identity with private key. If it is passed, the private cloud server will find the corresponding privileges of the user from its stored table list and send to the user then user can upload his files. The same way user can download his file from storage cloud.

## Conclusion

Hybrid clouds offer a greater flexibility to businesses while offering choice in terms of keeping control and security. Hybrid clouds are usually deployed by the organizations willing to push part of their workloads to public clouds either for cloud bursting purposes or for projects requiring faster implementation Because hybrid clouds vary based on company needs and structure of implementation. In proposed system authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check system presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, the duplicate-check tokens of files are generated by the private cloud server with private keys. Proposed system is secure in terms of insider and outsider attacks specified in the proposed security model. The proposed authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

## References

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, (2013) ,A Hybrid cloud approach for secure authorised deduplication, *IEEE Transactions on Parallel and Distributed Systems*.

Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui.,(2011) ,A secure cloud backup system with assured deletion and version control, *In 3rd International Workshop on Security in Cloud Computing.*

J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.(2002), Reclaiming space from duplicate files in a serverless distributed file system.,*In ICDCS*.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg.(2011), Proofs of ownership in remote storage systems. *ACM*.

Elhadj Benkhelifa , Dayan Fernando.( 2013) A Novel cloud hybrid access mechanism for highly sensitive data exchange, *The Fourth International Conference on Cloud Computing*.

J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl.(2013), A secure data deduplication scheme for cloud storage, *In Technical Report*.

Bellare, M., Keelveedhi, S., Ristenpart, T.( *2013)* Message-locked encryption and secure deduplication, *In: Advances in Cryptology.*

Xu, J., Chang, E.C., Zhou, J., Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. *In: 8th ACM SIGSAC symposium.*

Bellare, M., Keelveedhi, S., Ristenpart, T. (2013), DupLESS: server-aided encryption for deduplicated storage*. In: 22nd USENIX conference on Security.*

Hongwei Li, Yuanshun Dai1,, Ling Tian, and HaomiaoYang(2009), Identity-Based Authentication for Cloud Computing. *In Cloud Com*.

S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider.(2011), Twin clouds: An architecture for secure cloud computing. *In Workshop on Cryptography and Security in Clouds,.*

K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan.(2011), Sedic: privacyaware data intensive computing on hybrid clouds. *In Proceedings of the 18th ACM conference on Computer and communications security, USA.*

M. Bellare, C. Namprempre, and G. Neven.(2009), Security proofs for identity-based identification and signature schemes.

R. D. Pietro and A. Sorniotti. (2012), Boosting efficiency and security in proof of ownership for deduplication. *ACM*.

Nesrine Kaaniche, Maryline Laurent(2014), A Secure Client Side Deduplication Scheme in Cloud Storage Environments. *6th international conference on new technologies,mobility and security.*