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

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

Research Article

A Survey on Secure Distributed Deduplication Systems for Improved Reliability

Mahesh Janardhan Pawar†* and Pankaj R. Chandre†

†Dept. of Computer Engineering, Flora Institute of Technology, Pune, India

Accepted 02 Jan 2016, Available online 06 Jan 2016, **Vol.6, No.1 (Feb 2016)**

Abstract

Data de-duplication is one of the most important technique used for removing the identical copies of repeating data and it is used in the cloud storage for the purpose of reducing the storage space. However, there is only one copy for each file stored in cloud even if such file is owned by a huge number of users. Keeping the multiple data copies with similar content de-duplication eliminates redundant data by keeping only one physical copy and refer other redundant data to that copy. Data de-duplication can be file level or block level. The duplicate copies of identical file eliminate by file level de-duplication. And block level de-duplication eliminates duplicate blocks of data that occur in non-identical files. To maintain integrity we are providing the Third Party Auditor scheme which makes the audit of the file stored at cloud and notifies the data owner about file status stored at cloud server. This system supports security challenges such as an authorized duplicate check, integrity, data confidentiality and reliability.

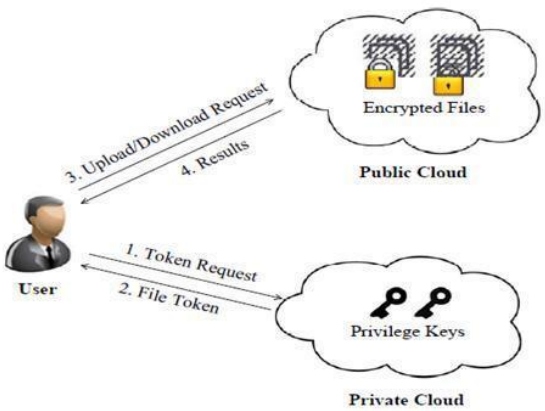
Keyword*: Data deduplication, Convergent encryption, Confidentiality, Hybrid cloud, Authorized Duplicate check.*

Introduction

Cloud computing provides unlimited virtualized recourse to a user as services across the whole internet while hiding the platform and implementing details. To make data management scalable in cloud computing, de-duplication has been invented as a conventional technique. Data De-duplication technique is used for eliminating the duplicate copies of repeated data in cloud storage and to reduce the data duplication. This technique is used to improve storage utilization and also be applied to network data transfers to reduce the number of bytes that must be sent. Keeping multiple data copies with the similar content, de-duplication eliminates redundant data by keeping only one physical copy and refer other redundant data to that copy. Data de-duplication occurs file level as well as block level. The duplicate copies of identical file eliminate by file level de-duplication. For the block level duplication which eliminates duplicates blocks of data that occur in non-identical files. Although data de-duplication takes a lot of benefits, security, as well as privacy concerns, arise as user's sensitive data are capable to both insider and outsider attacks. In the traditional encryption providing data confidentiality, is contradictory with data de-duplication. Traditional encryption requires different users to encrypt their data with own keys.

*Corresponding author: **Mahesh Janardhan Pawar**

Architecture



Explanation

In Proposed system, Convergent encryption has been used to enforce data confidentiality. Data copy is encrypted under a key derived by hashing the data itself. This convergent key is used for encrypting and decrypt a data copy. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP (storage cloud service provider). Security analysis demonstrates that that system is secure in terms of the definitions specified in the proposed security model.

• Designation is registered by admin or owner of the company based on his user id and password employees of the company able to perform operations such as file

upload download and duplicate checks on the files based on his privileges. There are three entities define in hybrid cloud architecture of authorized deduplication.

• Data Users: A user is an entity that wants to outsource data storage to the S-CSP(storage cloud service provider) and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. Each file is protected with the convergent encryption key and privilege keys to realizing the authorized deduplication with differential privileges.

• Private Cloud: This is a new entity for facilitating users secure use of cloud services. The private keys for privileges are managed by private cloud, which provides the file token to users. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between a user and the public cloud.

• S-CSP (storage cloud service provider): This is an entity that provides a data storage service in the public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

Literature Survey

1. A Hybrid Cloud Approach for Secure Authorized Deduplication (G.Prashanthi et al, 2015)

From this paper, we referred- Several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, 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. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct tested experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

2. A Hybrid Cloud Approach for Secure Authorized Deduplication (Gaurav Kakariya et al, 2014)

From this paper, we referred- Cloud computing has reached a maturity that leads it into a productive phase. This means that most of the main issues with cloud computing have been addressed to a degree that clouds have become interesting for full commercial exploitation. This, however, does not mean that all the problems listed above have actually been solved, only that the according risks can be tolerated to a certain degree. Cloud computing is therefore still as much a research topic, as it is a market offering. For better confidentiality and security in cloud computing, we have proposed new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. The proposed system includes proof of data owner so it will help to implement better security issues in cloud computing.

3. A Hybrid Cloud Approach for Secure Authorized Deduplication (Jin Li et al, 2010)

From this paper, we referred-In this paper, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, 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. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

4. Secure Deduplication And Data Security With Efficient and Reliable CEKM (N.O.Agrawal et al, 2014)

From this paper, we referred- The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We posit that secure deduplication services can be implement given additional security features insider attacker on Deduplication and outsider attacker by using the detection of masquerade activity. The confusion of the attacker and the additional costs incurred to distinguish real from bogus information, and the deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication.

5. A Hybrid Cloud Approach for Secure Authorized Deduplication (N.B. Kadu et al, 2015)

From this paper, we referred- It excludes the security problems that may arise in the practical deployment of

the present model. Also, it increases the national security. It saves the memory by deduplication the data and thus provides us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data.

6. Implementation of Hybrid Cloud Approach for Secure Authorized Deduplication (Jadapalli Nandini et al, 2015)

From this paper we referred- The notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, 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. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test-bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

7. A Hybrid Cloud Approach for Secure Authorized Deduplication (Jagadish et al, 2012)

From this paper, we referred- In this project, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. In this project, we perform several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. As a proof of concept in this project, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. From this project, we show that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

8. A Study on Authorized Deduplication Techniques in Cloud Computing (Bhushan Choudhary et al, 2014)

From this paper, we referred- The thought of authorized information deduplication was proposed to ensure the information security by counting differential benefits of clients in the duplicate copy check. The presentation of a few new deduplication developments supporting authorized duplicate copy in hybrid cloud architecture, in that the duplicate check tokens of documents are produced by the private cloud server having private keys. Security check exhibits that the methods are secure regarding insider and outsider assaults detailed in the proposed security model. As an issue verification of idea, the developed model of the proposed authorized duplicate copy check method and tested the model. That showed the authorized duplicate copy check method experience minimum overhead comparing convergent encryption and data transfer.

9. Secure Authorized Deduplication on Cloud using Hybrid Cloud Approach (Ankita Mahajan)

From this paper, we referred- We also presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, 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. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

10. Secured Authorized Deduplication based Hybrid Cloud (Rajashree Shivshankar, 2014)

From this paper, we referred- Data deduplication is an important technique for eliminating redundant data.Instead of taking no. of same files, it stores only single copy of the file. In most organizations, storage system contains many pieces of duplicate data. . For example, the same file may be saved in several different places by different users. 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 is data compression technique for improving the bandwidth efficiency and storage utilization. Data deduplication most widely used in cloud computing. It makes data management scalable and storage problem in cloud computing. Data deduplication protects the confidentiality of sensitive data. Data deduplication works with convergent encryption technique to encrypt the data before uploading. . Companies frequently use deduplication in backup and disaster recovery applications. In this paper we attempt authorized deduplication check, combine with convergent encryption for providing security to sensitive data using hybrid cloud computing.

Conclusion

In this Project, the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. In this project, we perform several new deduplication constructions supporting authorized duplicate check in hybrid cloud

architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. As a proof of concept in this project, we implement a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. From this project, we show that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer. Futures work: It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provides us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data.

References

G.Prashanthi, Z.Shobarani (2015), A Hybrid Cloud Approach for Secure Authorized Deduplication, International Journal of Innovative Research in Computer and Communication Engineering Vol.3, Special Issue 4.

GauravKakariya, Prof. SonaliRangdale (2014), A hybrid cloud approach for secure authorized deduplication, International Journal of Computer Engineering and Applications, Volume VIII.

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou (2010), A Hybrid Cloud Approach for Secure Authorized Deduplicatio, IEEE transactions on parallel and distributed system Vol: PP NO:99 .

N.O.Agrawal, S.S.Kulkarni (2014), Secure Deduplication And Data Security With Efficient And Reliable CEKM, International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 11.

N.B. Kadu, AmitTickoo (2015), A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal of Scientific and Research Publications, Volume 5, Issue 4, 1 ISSN 2250-3153

Jadapalli Nandini, Ramireddy Navateja Reddy (2015), Implementation of hybrid cloud approach for secure authorized deduplication, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 02 Issue: 03.

Jagadish, Dr.Suvarna Nandyal (2012), A Hybrid Cloud Approach for Secure Authorized Deduplication, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

BhushanChoudhary, Amit Dravid (2014), A Study on Authorized Deduplication Techniques in Cloud Computing, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 12.

AnkitaMahajan, Secure Authorized Deduplication on Cloud Using Hybrid Cloud Approach, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 12.

Rajashree Shivshankar Walunj (2014), Secured Authorized Deduplication Based Hybrid Cloud, The International Journal Of Engineering And Science (IJES), Volume 3, Issue 11, Pages 34-39