A Survey on a Hybrid Cloud Mechanism for Secure Deduplication Storage System

Rushikesh Naiknaware[1], Omkar Deshpande[2], Abhishek Kangle[3], Himalay Koli[4], Dipmala Salunke[5]

Student, Dept. of IT, Rajarshi Shahu College of Engineering, Tathawade, Savitribai Phule Pune University, Pune, India[1234]

Prof. Dept. of IT, Rajarshi Shahu College of Engineering, Tathawade, Savitribai Phule Pune University, Pune, India[5]

ABSTRACT: Data deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this work makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. The work also presents several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, the work implement a prototype of proposed authorized duplicate check scheme and conduct tested experiments using the prototype. The work shows that the proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations. Data deduplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. 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. Companies frequently use deduplication in backup and disaster recovery applications, but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication

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

I. 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. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by 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. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage.
Deduplication can take place at either the file level or the block level for file level deduplication; it eliminates duplicate copies of the same file. Traditional encryption, while 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 different cipher texts making deduplication impossible.

Convergent encryption has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts/decrypts a data copy with a Convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. A Hybrid Cloud is a combined form of private clouds and public clouds in which some critical data resides in the enterprise's private cloud while other data is stored in and accessible from a public cloud. Hybrid clouds seek to deliver the advantages of scalability, reliability, rapid deployment and potential cost savings of public clouds with the security and increased control and management of private clouds.. In computing, data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. This technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

II.   LITERATURE SURVEY

1.   *A Hybrid Cloud Approach for Secure Authorized Deduplication [1]*
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 [2]*
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. 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 [3]*
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 [4]*
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 [6]*
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*
From this paper we Referred-

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 [7]*
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[8]*
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 [9]*
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 [10]*
From this paper we Referred-

Data deduplication is an important technique for eliminating redundant data.Instead of taking no. of same files, it store only single copy of 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 improve the bandwidth efficiency and storage utilization. Data deduplication most widely used in cloud computing. It make data management scalable and storage problem in cloud computing. Data deduplication protects the confidentiality of sensitive data. Data deduplication work 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.

III. 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 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.

REFERENCES

1. G.Prashanthi, Z.Shobarani "A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal of Innovative Research in Computer and Communication Engineering Vol.3, Special Issue 4, April 2015
2. GauravKakariya, Prof. SonaliRangdale "A HYBRID CLOUD APPROACH FOR SECURE AUTHORIZED DEDUPLICATION" International Journal of Computer Engineering and Applications, Volume VIII, Issue I, October 14
3. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou "A Hybrid Cloud Approach for Secure Authorized.
4. Deduplication" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEM VOL:PP NO:99 YEAR 2014
5. N.O.AGRAWAL1, Prof Mr. S.S.KULKARNI "Secure Deduplication And Data Security With Efficient And Reliable CEKM" International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 11, November 2014
6. Prof. N.B. Kadu*, Mr. AmitTickoo" A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015 1 ISSN 2250-3153
7. JADAPALLI NANDINI , RAMIREDDY NAVATEJA REDDY "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 | June-2015