ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

Cloud Based Deduplication on Encrypted Data

Ankush R. Deshmukh[1], Prof. R. V. Mante[2], Dr. P N. Chatur[3]

PG Scholar, Department of Computer Science and Engineering, Government College of Engineering, Amravati, India[1]

Assistant Professor, Department of Information Technology, Government College of Engineering, Amravati, India[2]

Associate Professor, Department of Computer Science and Engineering, Government College of Engineering, Amravati, India[3]

ABSTRACT: Nowaday we regularly use cloud services in our daily life. There are various services provided by cloud such as Software as a service, Platform as a service, Infrastructure as a service. We used to keep our data, documents, files on cloud. The data that we store may be personal, private, secret data. So we must be very sure that whatever the cloud service we use that must be secure. At the same time with security we have to think of price that we have to pay to cloud. For these problems, in this paper there is ABE, Attribute based encryption, Scheme is used. To ensure data privacy, existing research proposes to outsource only encrypted data to CSPs. However, the same or different users could save duplicated data under different encryption schemes at the cloud. Although cloud storage space is huge, this kind of duplication wastes networking resources, consumes excess power, and complicates data management. At the same time, data owners want CSPs to protect their personal data from unauthorized access. CSPs should therefore perform access control based on the data owner's expectations. In addition, data owners want to control not only data access but also its storage and usage. From a flexibility viewpoint, data deduplication should cooperate with data access control mechanisms. That is, the same data, although in an encrypted form, is only saved once at the cloud but can be accessed by different users based on the data owners' policies. However, current industrial deduplication solutions can't handle encrypted data. Existing solutions for deduplication are vulnerable to brute-force attacks and can't flexibly support data access control and revocation. We propose a scheme based on attribute-based encryption (ABE) to deduplicate encrypted data stored in the cloud and support secure data access control at the same time. Analysis and implementation demonstrate that our scheme is secure, effective, and efficient.

KEYWORDS: Access control, deduplication, cloud computing, proxy re-encryption.

I. INTRODUCTION

Cloud offers various services to the user. Data storage service provided by cloud is the most commonly used service given by cloud.User used to upload data onto the cloud and let cloud to manage that data.Data may be in the form of files which can be personal or private.As user is storing data onto the cloud, user has to pay rent for storing data. Data storage rent may vary from different cloud service provider but as user is paying rent and if user is storing the same copy of data on cloud multiple times then the rent will go on increasing.In this paper, cloud based deduplication scheme is proposed.In which user will store the data on cloud only onceSame copy of data will not get stored again. Deduplication mechanism is proposed in this paper.As user is storing data onto cloud then user must be requiring security to data. For the security purpose data is get stored on the cloud in the encrypted format. So deduplication checking has to be performed on the encrypted data. User will store the data on the cloud in encrypted format and then it is checked that whether that data is already present on the cloud or not. If the data that user want to store onto the cloud is already present then duplication occur. User will also store data in encrypted format duplication check will also be performed on the data present on the cloud which is also the encrypted data. It is complicated for the data holder to check the deduplication on encrypted data. In this scheme the data ownership challenge and proxy re-encryption is used to manage encrypted data storage with deduplication. By using this technique the rent that user has to pay for data will

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

reduced as the same copy of data will not be allowed to store onto the cloud. With reducing the cost it also provide security to the user as data will store in encrypted format.

The paper is organized as follows. Section II presents related work. The details of methods used for Deduplication described in section III. The section IV presents system architecture and conclusion in sectionV.

II. RELATED WORK

Pasquale Puzio, Refik Molva, Melek Onen, Sergio Loureiro[2] proposed a system in which they are using block level deduplication and providing data confidentiality at the same time.Aim of the system is toidentify identical data and store them only once. The result of encryption is to make encrypted data copy which cannot be distinguishable after being encrypted. In the process of deduplication it is difficult to identify the same data segment.So they have used convergent encryption in which encryption key is usually the result of hash of data.In this processtheyassures block-level deduplication anddata confidentiality. Using Block-level deduplication makes the system more flexible and efficient.ClouDedup preserves confidentiality and privacy even against potentially malicious cloud storage providers thanks to an additional layer of encryption. ClouDedup offers an efficient key management solution through the metadata manager; The new architecture defines several different components and a single component cannot compromise the whole system without colluding with other components. ClouDedup works transparently with existing cloud fully compatible with standard storage APIs and any cloud storage provider can be easily integrated in this architecture.

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee[3], and Wenjing Lou proposed a system, in which they have used convergent encryption technique to encrypt the data, Data duplication is performed with rendering confidentiality of data. They also present several new deduplication schemes to perform duplicate check in a hybrid cloud. In their system,they have usedhybrid cloud architecture consisting of a public cloud and a private cloud. The private cloud is involved as a proxy toallow data owner to securely perform duplicate check with differential privileges. The data owners only outsource their data storage by utilizingpublic cloud while the data operation is managed inprivate cloud. A new deduplication system supporting differentialduplicate check is proposed under this hybridcloud architecture where the S-CSP resides in the publiccloud. The user is only allowed to perform the duplicatecheck for files marked with the corresponding privileges.

Z. Sun, J. Shen, and J. M. Yong[4] proposed a system, which consist of a front-end deduplication applicationand Hadoop Distributed File System At the front end, it has a deduplication application. At the back end,there are two main components, which are HDFSused as a mass storage system and HBase,used as a fast index.Promising results were obtained from simulation usingVMware to simulate a cloud environment and execute theapplication on the cloud environment.

Mihir Bellare1, Sriram Keelveedhi2, Thomas Ristenpart3[5] proposed, formalize a new cryptographic primitive, Message-Locked Encryption (MLE), where the keyunder which encryption and decryption are performed is itself derived from the message. On the practical side, They provide ROM security analyses of a natural familyof MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and they make connections with deterministic encryption, hash functions secure oncorrelated inputs and the sample-then-extract paradigm to deliver schemes under different assumptionsand for different classes of message sources.

T. Y. Wu, J. S. Pan, and C. F. Lin[6] proposed a system,in which to reduce the workload due to duplicate file,proposed the index name server(INS) tomanage not only file storage, data de-duplication, optimized nodeselection, and server load balancing, but also file compression,chunk matching, real-time feedback control, IP information,and busy level index monitoring. To manage and optimize thestorage nodes based on the client-side transmission status by proposed INS, all nodes must give optimal performance andoffer suitable resources to clients. In this way, not only can theperformance of the storage system be improved, but the files can also be reasonably distributed, decreasing the workload of thestorage nodes.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

III. THEORETICAL BACKGROUND

Cloud computing offers a new way of service provision by re-arranging various resources over the Internet. The most important and popular cloud service is data storage. In order to preserve the privacy of data holders, data are often stored in cloud in an encrypted form. However, encrypted data introduce new challenges for cloud data de-duplication, which becomes crucial for cloud data storage and processing in cloud. Traditional de-duplication schemes cannot work on encrypted data. Existing solutions of encrypted data de-duplication suffer from security weakness.They cannot flexibly support data access control and revocation. Therefore, few of them can be readily deployed in practice. In this, we propose a scheme to de-duplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption.

Objectives

To develop a cloud based application in which documents will be uploaded only once to reduce cloud server space as well as platform rent. To implement security of document on cloud server with the help of encryption techniquesTo implement document destruction technique to enhance security of important documents as well as to reduce the rent of cloud platform. To integrate cloud data de-duplication with access control.Existing solutions of encrypted data de-duplication suffer from security weakness. They cannot flexibly support data access control and revocation. Therefore, few of them can be readily deployed in practice. Here the scheme is to de-duplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption. It integrates cloud data de-duplication with access control.

System Model

Two clouds are used in this system model. Access permission cloud and cloud service provider forstoring encrypted data. Data holder are the users that wish to store data onto cloud. Data holders includes data owner.If we consider architecture of company having employees and the boss who is the data owner but in the organisation if boss wants to give access permission to employees for data storage then these access permission are stored on the access permission cloud. Actual data will be stored in the cloud service provider's cloud.

Figure 1: System Model

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

IV. SYSTEM WORK FLOW

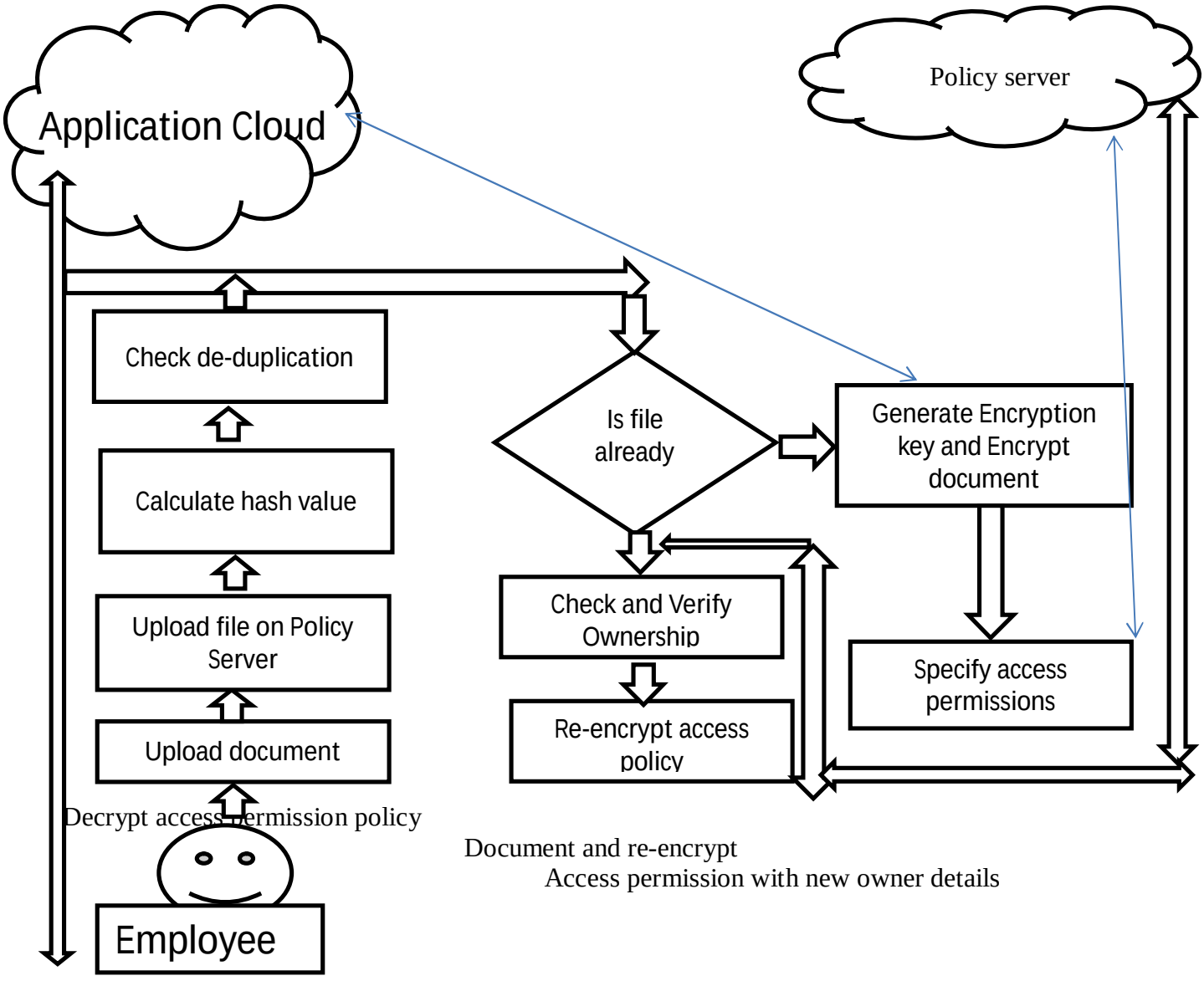

Figure 2: System flow

A. DE-DUPLICATION

De-duplication technique is used to check whether the uploaded document is already exist on cloud server or not?If uploaded file is already exist, access policy server will check and verify user ownership or access permission. If user ownership is verified, the access policy stored on access policy server will be re-encrypted. In our proposed system, we

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 1, January 2017

International Conference on Recent Trends in Engineering and Science (ICRTES 2017)

20th-21st January 2017

Organized by

Research Development Cell, Government College of Engineering, Jalagon (M. S), India

will check de-duplication of document with the help of hash value of particular document. Hash value is unique representation of any document.

b.Access Control Management

Access permissions will be of two types

- Decryption Access permission
- De-duplication access permission

Access permission Details will be maintained on access policy server. At the time of file upload user have to specify access permission to decrypt the document.

Here we are maintaining two clouds, application cloud and the policy server cloud. Policy server cloud stores the access permission of users. There are two types of access permission that we have used in our project Deduplication access permission and Decryption access permission. First user upload the document then its hash value is calculated by using SHA algorithm. Each document which is not the same has unique hash value. If the deduplication has occurred then the

deduplication access permission of user is verified. If the user has Deduplication access permission, then the same data is not stored onto the cloud otherwise data will get stored. For example, there are two branches of company first branch want to maintain their own data separately then they will not provide deduplication access permission to other branch. So that data that may be even duplicated will get stored onto cloud. But if they provide Deduplication access permission to each other then their employees will not be able to save same data on cloud

V. CONCLUSION

In this survey paper, the different deduplication techniques for cloud based storage are discussed. Technique uses access permission mechanism for deduplication and decryption. Paper shows two access permission mechanism as deduplication access and decryption access permission. It reduces the cost for data storage in cloud as duplicate data is not get stored onto cloud. This survey paper shows implementation security of document on cloud server with the help of encryption techniques.

REFERENCES

[1] "Deduplication on Encrypted Big Data in Cloud", Zheng Yan, Senior Member, IEEE, Wenxiu Ding, Xixun Yu, Haiqi Zhu, and Robert H. Deng, Fellow, IEEE, IEEE TRANSACTIONS ON BIG DATA, VOL. 2, NO. 2, APRIL-JUNE 2016

[2] "ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage", Pasquale Puzio, Refik Molva, Melek Onen, Sergio Loureiro, 2013 IEEE International Conference on Cloud Computing Technology and Science

[3] A Hybrid Cloud Approach for Secure Authorized Deduplication Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Wenjing Lou IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 26, NO. 5, MAY 2015.

[4] Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355, doi:10.1109/CSCWD.2011.5960097

[5] "Message-Locked Encryption and Secure Deduplication", Mihir Bellare1, Sriram Keelveedhi2, Thomas Ristenpart3, proceedings of Eurocrypt 2013.

[6] T. Y. Wu, J. S. Pan, and C. F. Lin, "Improving accessing efficiency of cloud storage using de-duplication and feedback schemes,"IEEE Syst. J., vol. 8, no. 1, pp. 208–218, Mar. 2014,doi:10.1109/JSYST.2013.2256715.

[7] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocolsin cloud storage," in Proc 27th Annu. ACM Symp. Appl. Comput.,2012, pp. 441–446.

[8] Z. C. Wen, J. M. Luo, H. J. Chen, J. X. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc.Int. Conf. Intell. Netw. Collaborative Syst., 2014, pp. 85–90, doi:10.1109/INCoS.2014.111.

[9] C. Y. Liu, X. J. Liu, and L. Wan, "Policy-based deduplication in secure cloud storage," in Proc. Trustworthy Comput. Serv., 2013, pp. 250–262, doi:10.1007/978-3-642-35795-4_32.