Attribute-Based Storage Supporting Secure Verifiable Storage Deduplication Scheme on Big Data in Cloud

Yenumala Sankara Rao[1], Anumula Bhavani[2]
Department of MCA
[1]Associate Professor,St. Mary's Group of Institutions, Guntur, Andhra Pradesh, India
[2]PG Student, St. Mary's Group of Institutions, Guntur, Andhra Pradesh, India

Abstract*-Cloud Computing is a emerging technology where we can get platform as a service, Software as a service, infrastructure as a service, while getting storage as a service from cloud. Memory management are very most important in cloud computing. In existing paper user can upload text files only and not a secured. attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage In existing paper we propose a user can upload all type of files (jpg, gif, png, audio, video, pdf, ppt) using deduplication technology. We propose a scheme to deduplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption. It is commonly used to reduce the space and bandwidth requirements of service by eliminating redundant data and storing only a single copy of file using cloud service provider. Here we propose secure deduplication to avoid the same files uploaded multiple times.*

Keywords-Attribute-based encryption(ABE),Deduplication, Big Data, Cloud Computing, Access Control, Proxy Re-Encryption, Memory Management

I. INTRODUCTION

Cloud computing is an innovative service mode. Through the network to the required resources (hardware, platform, and software), virtual integration into a reliable and high performance of computing platform. In cloud computing, all user data are stored in the cloud resources like cloud nodes. The Cloud computing is the use of computing resources that are delivered as a service over a network. The resources are made available on the internet as managed third-party services. One of the most fundamental services offered by cloud providers is data storage. The data is available at any time because the cloud computing is an internet based computing. The cloud service providers are providing various services to the users. The conclusion distribute to the user through the network when the user needed the security. Although cloud computing has aim to mature service model, and have large commercial, cloud computing is still facing many problems[8]. Cloud Computing still facing the number of major challenges: Safety, Stability and Performance issue. Including the security and memory management problem concerns the most.

A real cloud computing security incidents have profound acknowledge the urgency of cloud security and authentication issues, example for users cannot access to their email and other personal data. More important, due to the technical personnel not to be make backups of their data, conclusion of Microsoft cannot recover data. Although the cloud storage service can realize multiple copy of files fault tolerance and backup automatically, it is also guarantee 100 percent security and authentication. Deduplication has proved to achieve high cost savings storage needs for backup applications. In this paper, we propose a scheme based on data ownership challenge and Proxy Re-Encryption (PRE) to manage encrypted data storage with deduplication[5][8]. Specifically, the contributions of this paper can be summarized below:

1. Our scheme can flexibly support data sharing with deduplication even when the data holders or offline.
2. We propose verify data ownership and check duplicate storage with secure challenge and bigdata support.
3. We integrate cloud data deduplication with access control.
4. We propose the security and assess the performance of the proposed data deduplication scheme. The result shows the efficiency, effectiveness and applicability.

II. RELATED WORKS

Attribute-Based Encryption. Sahai and Waters [6] introduced the notion of attribute-based encryption (ABE), and then Goyal et al. [16] formulated key-policy ABE (KP-ABE) and ciphertext-policy ABE (CP-ABE) as two complimentary forms of ABE. The first KP-ABE construction given in [16] realized the monotonic access structures, the first KP-ABE system supporting the expression of non-monotone formulas was presented in [17] to enable more viable access poli-cies, and the first large class KP-ABE system was presented by in the standard model in [18]. Nevertheless, we believe that KP-ABE is less flexible than CP-ABE because the access policy is determined once the user's attribute private key is issued. Bethencourt, Sahai and Waters [19] proposed the first CP-ABE construction, but it is secure under the generic group model. Cheung and Newport [20] presented a CPABE scheme that is proved to be secure under the standard model, but it only supports the AND access structures.

A CP-ABE system under more advanced access structures is proposed by Goyal et al. [21] based on the number theoretic assumption. In order to overcome the limitation that the size of the attribute space is polynomially bounded in the security parameter and the attributes are fixed ahead, Rouselakis and Waters [22] built a large universe CP-ABE system under the prime-order group. In this paper, the Rouselakis-Waters system is taken as the underlying scheme for the concrete construction. Hindong Wu[2] presented a Data Mining with Big Data concern large-volume, complex, growing data sets with the multiple, autonomous sources. Big data with the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly increasing in all science and engineering domains, including Physical, Biological and Biomedical sciences. This paper presents a HACE theorem that characters of the features of Big Data revolution process, and we proposes a Big Data processing model, from the data mining perspective. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.M. Bellare, S. Keelveedhi, and T. Ristenpart [4] presented a DupLESS: Server aided encryption for deduplicated storage which describes Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. When the fast development of the networking, data storage, and the data collection capacity, Big Data are now increase in all arts and science and engineering domains, including physical [1], biological and biomedical sciences. This paper presents a HACE theorem that characters the features of the Big Data revolution, and proposes a Big Data processing model is most important of the data mining perspective. This data-driven model involves demand driven aggregation of information sources, mining and analysis, user interest modeling, security and authentication considerations. We analyze the important challenging issues in the data driven model and also the Big Data revolution process.

Mozy, Mozy[5] presented A File-storage and Sharing Service which explains the concept of smart grid has emerged as a convergence of traditional power system of engineering and information and communication technology. In this paper, we propose an efficient and privacy-preserving aggregation scheme, named EPPA, for smart grid communications. EPPA uses a super expanding the sequence to structure multidimensional data and an encrypt the structured data by only the homomorphic Paillier cryptosystem technique. For data communications from user to smart grid computing operation center, data aggregation is performed the directly on cipher text (original data) at local gateways without decryption, and aggregation result of the original file can be obtained at the operation center. Through extensive analysis, we expose that EPPA resists various security threats and preserve the user privacy, and has significantly less computation and communication overhead than existing competing approaches.Z. O. Wilcox [6] presented Convergent encryption which describes Cloud computing is envisioned as the next generation architecture of IT enterprises. While this subcontract storage and computing beau ideal (paradigm) can potentially bring high economical savings for data owners and users. We present a general techniques for using searchable encryption or decryption techniques, which allows encrypted data to be searched only the users without loss of information about the data itself and. In particularly, we discuss the three desirable functionalities of most usable search operations: supporting result ranking, similarity search, and search over structured data thus are the important functionalities. For each of them, we explaining the approaches to design efficient privacy-assured searchable encryption schemes, which are based only several recent symmetric-key encryption primitives. We analyze their advantages and disadvantages, and important of the future challenges that need to be solved to make such secure searchable cloud data service a reality.

M. Bellare, S. Keelveedhi, and T. Ristenpart[3] presented Message-locked encryption and secure deduplication which explains the pervasiveness of smart phones and the advance of wireless connection of body sensor networks (BSNs), and the mobile Healthcare (m-Healthcare), which use the operation of Healthcare provider into a pervasive environment for better health monitoring, has the attracted with considerable interest recently. However, the flourish of m-Healthcare facing many challenges: one of the Information Security (IS), privacy preservation. In this paper propose a secure and privacy-preserving computing subcontracting, called SPOC, for m-Healthcare emergency[4]. The SPOC, smart phone resources including computing power and energy can be gathered to process the computing intensive personal health information (PHI) during m-Healthcare emergency with minimum privacy disclosure. In specific, to authority of PHI privacy disclosure and the more reliability of PHI process and transmission in m-Healthcare emergency, we in introduce an efficient user-centric privacy access control in SPOC concepts, which is based only a attribute-based access control and a new privacy-preserving scalar product computation (PPSPC) method, and allows a medical user to make up your mind who can involve in the opportunistic computing to support in processing his overpowering PHI data. Detailed security analysis shows that the proposed SPOC framework can efficiently achieve user-centric privacy access control of the m-Healthcare emergency. In additional, performance evaluations via extensive simulations demonstrate the SPOC's effectiveness in term of providing high-reliable-PHI process and transmission while minimize the privacy disclosure during the m-Healthcare emergency

III. PROBLEM STATEMENTS

A. System and Security Model

We propose the system high security is provided. The file can be uploaded only one time. User can‟t download without admin permission. It improves storage capacity in the cloud.

Any types of file can be uploaded using encryption and decryption algorithm. The Fig 1 shows the overall deduplication process [10]. The user can upload the files in cloud computing nodes and check this files already there in the database. If already there that file cannot uploaded if else uploaded the file using encryption algorithm (AES, DES, SHA)

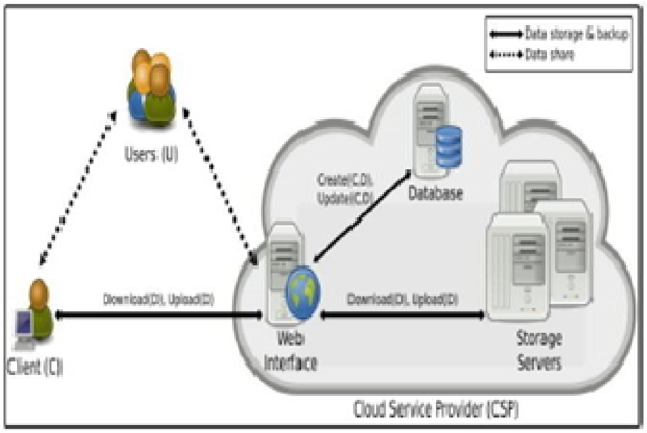


Figure1. System Model

Cloud Service provider (CSP)

CSP can offer storage services and cannot be fully trusted.

User

User can upload the files using encryption algorithm and before store the files to check database because of avoid the duplicate files.

Clients

Clients get the permission to the user and provide the token to access the files in database for security purpose for providing token. If the clients already have a token directly access the files in database.

SCHEME File split

In this module we create user interface page from this page user can create our own profile, once the user registered that securely stored in server database. After user can login and upload our preferred document to server database, that document must be spit multipart before store in actual database.

Generate Token

In this module we generate token to all spitted part of data‟s. Data checker is a intermediate server between user and CSP (Cloud service provider), job of the data checker is find de-duplication. For that it generate token to all parts of incoming data and then check if that token is already exist or not if yes send “duplicate ” response to user or request forwarded to Key generator for encrypt data.

Encrypt Data

In this module we encrypt user uploaded data. In our project we implemented two Cryptographic algorithm DES and AES, Encryption mode based on user request (EX : DES or AES).

DES

The Data Encryption Standard (DES) is a symmetric-key block cipher DES is an implementation of a Feistel Cipher structure. It uses 16 round of the Feistel structure. The block size of DES is 64-bit. Though, key length of DES is 64-bit, DES has an effective key length of 56 bits, since 8 of the 64 bits of the key are not used by the encryption algorithm.

AES

The Advanced Encryption Standard is the more popular and most adopted symmetric encryption algorithm like Advanced Encryption Standard (AES). It is found at least six times faster than 3DES. Nowadays AES is an iterative with Feistel cipher structure. It is based only a „substitution–permutation network‟. It comprises of a series operations, some of which involving replace the inputs by specific outputs (substitutions) and others involve shuffling bits around. All encrypted values are stored in database as key-value pair (key is token number another value is document part).

Encrypted Data Update

In this module used to update the encrypted data in cloud by the data owner. Only authorized user can access this files. This module provide the high security and avoid the redundancy.

Share Document’s

In this module user can view all uploaded document and also share our document to community users. One of the major advantages of out project is Data lineage. All stored data must be based on Data Lineage concept. Data Lineage means share one copy of data to all users and also maintain all accessed consumer information in dataset. In this way we can avoid Duplications and easy to identify data leakage. Effectively manage Database memory.

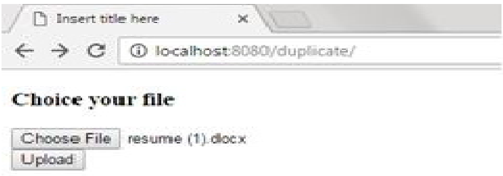
Thus the above modules are explained avoid the redundancy of data or files.

C. Design Goals

The file can be uploaded only one time. User can‟t download without admin permission. It improves storage capacity in the cloud. Any types of file can be uploaded using encryption and decryption algorithm.

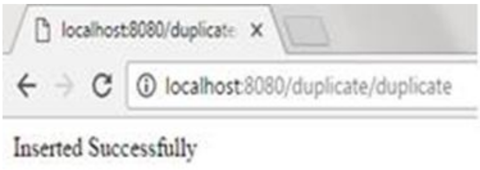
D. Sample Outputs

The given below diagrams shows how to work the deduplication process. The user can upload the files in cloud computing nodes and check this files already there in the database. If already there that file cannot uploaded if else uploaded the file using encryption algorithm.



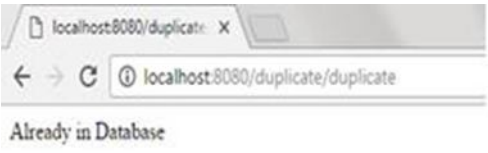
Choice your file

Figure2. Choose file



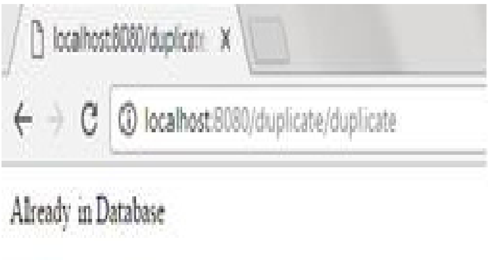
Inserted Successfully

Figure3. Insert file



Already in Database

Figure4. Upload file



Already in Database

Figure5. Identify Deduplication

The above diagrams shows the avoid the uploading repeated files.

IV. CONCLUSION

Managing encrypted data with deduplication is important and significant in practice for achieving a successful cloud storage service, especially for big data storage. In this paper, we proposed a practical scheme to manage the encrypted big data in cloud with deduplication only based on ownership challenge and Proxy Re-Encryption. Our scheme can mostly support flexible data update and sharing with deduplication even when the data holders are offline. Here we propose secure deduplication to avoid the same files uploaded multiple times and the files are secured based on encryption. Encrypted data can be securely accessed because only authorized data holders can obtain the symmetric keys used for data decryption.

REFERENCES

[1] Zheng Yan "Deduplication on Encrypted Big Data in Cloud" ieee Trans on bigdata year: 2016, pp.138-150

[2] Hindong Wu "Data Mining with Big Data" IEEE Trans on Knowledge year: 2014,pp.97-107

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart"Message-locked encryption and secure deduplication" in Proc.CryptologyEURO-CRYPT, year:2013, pp.296-312.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.

[5] Mozy. Mozy: A File Storage and sharing Service. (2016). Online].Available: http://mozy.com

[6] Z. O Wilcox, "Convergent encryption reconsidered", 2011. Online].Available: http: //www.mailarchive.com

[7] Dropbox, A file-storage and sharing service. (2016). Online]. Available: http://www.dropbox.com

[8] G.Ateniese, K. Fu, M. Green, and S.Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inform. Syst. Secure., vol. 9, no. 1, pp. 1–30, 2006.

[9] Opendedup. (2016).Online]. Available: http://opendedup.org/

[10] D. T. Meyer and W. J Bolosky, "A study of practical deduplication," ACM Trans. Storage, vol. 7, no. 4, pp. 1–20,2012.