International Journal For Technological Research In Engineering
Volume 5, Issue 10, June-2018 *ISSN (Online): 2347 - 4718*

ENCRYPTION OF DATA DEDUPLICATION IN CLOUD COMPUTING

Pritam Prasad Lata
Assistant Professor, Department Of Computer Science
Shekhawati Institute of Technology, Sikar, Rajasthan, India

ABSTRACT: Cloud computing plays an important role in supporting data storage, processing, and management in the Internet of Things (IoT). To preserve cloud data confidentiality and user privacy, cloud data are often stored in an encrypted form. However, duplicated data that are encrypted under different encryption schemes could be stored in the cloud, which greatly decreases the utilization rate of storage resources, especially for big data. Several data deduplication schemes have recently been proposed. However, most of them suffer from security weakness and lack of flexibility to support secure data access control. Therefore, few can be deployed in practice. This article proposes a scheme based on attribute-based encryption (ABE) to deduplicate encrypted data stored in the cloud while also supporting secure data access control. The authors evaluate the scheme's performance based on analysis and implementation. Results show the efficiency, effectiveness, and scalability of the scheme for potential practical deployment.

I. INTRODUCTION

CSPs provide desirable service properties, such as scalability, elasticity, fault tolerance, and pay per use. Thus, cloud computing has become a promising service paradigm to support IoT applications and IoT system deployment. 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. Thus, saving storage is becoming a crucial task for CSPs. Deduplication can achieve high space and cost savings, reducing up to 90 to 95 percent of storage needs for backup applications (http://opendedup.org) and up to 68 percent in standard file systems.1 Obviously, the savings, which can be passed back directly or indirectly to cloud users, are significant to the economics of cloud business. 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 attacks2 and can't flexibly support data access control and revocation (see the "Related Work in Data Deduplication" sidebar for a discussion of some other work in this area).3 Few existing schemes for cloud data access control support data deduplication simultaneously, 4 and few can ensure flexibility and security with sound performance for cloud data deduplication that data owners control directly.5–7 We propose a scheme based on attribute-based encryption (ABE) to deduplicate encrypted data stored in the cloud and support secure data access ontrol at the same time. Analysis and implementation demonstrate that our scheme is secure, effective, and efficient.

II. EXISTING SYSTEM

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. Existing solutions for deduplication are vulnerable to brute-force attacks2 and can't flexibly support data access control and revocation (see the "Related Work in Data Deduplication" sidebar for a discussion of some other work in this area). Existing industrial solutions fail in encrypted data deduplication.

DISADVANTAGES

Deduplication technology has become quite the staple in many data storage environments. But what makes it a good fit in one data center, may not be the case in another. This E-Guide from SearchStorage.com is designed to help you determine what you're trying to solve with deduplication technology. It then outlines: The advantages and disadvantages of dedupe backup Dedupe misconceptions Howdedupe and compression on primary storage can reduce your data footprint.

PROPOSED SYSTEM

In this application we propose 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. intra-userdeduplication and interdeduplication.6 In their scheme, the ciphertext C of convergent encryption is further encrypted with a user key and transferred to the servers. However, it doesn't deal with data sharing after deduplication among different users.

Advantages:

The scheme can easily realize data access control by introducing control policies into AP when calling EncryptKey(DEKu, AP, PKIDu) by updating AP to support

both deduplication and access control based on practical demands. Our scheme can also support digital rights management based on the data owner's expectations. Second, the scheme saves CSP storage space since it only stores one copy of the same data. torage-based data deduplication reduces the amount of storage needed for a given set of files. It is most effective in applications where many copies of very similar or even identical data are stored on a single disk—a surprisingly common scenario. In the case of data backups, which routinely are performed to protect against data loss, most data in a given backup remain unchanged from the previous backup.

III. MODULE DESCRIPTION

Modules:

Data Owner

In this module, the data owner uploads their data in the cloud server. For the security purpose the data owner encrypts the file and the index name and then store in the cloud. The data encryptor can have capable deleting of a specific file. And also he can view the transactions based on the files he uploaded to cloud.

Data User

In this module, user logs in by using his/her user name and password. After Login user requests search control to cloud and will Search for files based on the index keyword with the Score of the searched file and downloads the file. User can view the search ratio of the files and also the top k documents.

Cloud Server

The cloud server manages a cloud to provide data storage service. Data owners encrypt their data files and store them in the cloud for sharing with Remote User. To access the shared data files, data consumers download encrypted data files of their interest from the cloud and then decrypt them.

The cloud server authorizes the data owner and the data user and provides the search requests sent from the users. Also in this module it shows personalized search model and the interest search model. Can view all the file attackers.

IV. CONCLUSION

In this paper, we explore the problem of secure multi-keyword search for multiple data owners and multiple data users in the cloud computing environment. Different from prior works, our schemes enable authenticated data users to achieve secure, convenient, and efficient searches over multiple data owners' data. To efficiently authenticate data users and detect attackers who steal the secret key and perform illegal searches, we propose a novel dynamic secret key generation protocol and a new data user authentication protocol. To enable the cloud server to perform secure search among multiple owners' data encrypted with different secret keys, we systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel Additive Order and Privacy Preserving Function family. Moreover, we show that our approach is computationally efficient, even for large data and keyword sets. As our future work, on one hand, we will consider the problem of secure fuzzy keyword search in a multi-owner paradigm. On the other hand, we plan to implement our scheme on the commercial clouds.

REFERENCES

[1] D.T. Meyer and W.J. Bolosky, "A Study of Practical Deduplication," ACM Trans. Storage, vol. 7, no. 4, 2012, pp. 1–20.

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-Locked Encryption and Secure Deduplication," Advances in Cryptology (EUROCRYPT 13), LNCS 7881, 2013, pp. 296–312.

[3] J. Li et al., "A Hybrid Cloud Approach for Secure Authorized Deduplication," IEEE Trans. ParallelDistributed Systems, vol. 26, no. 5, 2015, pp. 1206–1216.

[4] Z. Wan, J. Liu, and R.H. Deng, "HASBE: A Hierarchical Attribute-Based Solution for Flexible and Scalable Access Control in Cloud Computing," IEEE Trans. Information Forensics and Security, vol. 7, no. 2, 2012, pp. 743–754.

[5] M. Fu et al., "Accelerating Restore and Garbage Collection in Deduplication-Based Backup Systems via Exploiting Historical Information," Proc.Usenix Ann. Technical Conf., 2014, pp. 181–192.

[6] M. Kaczmarczyk et al., "Reducing Impact of Data Fragmentation Caused by In-Line Deduplication," Proc. 5th Ann. Int'l Systems and StorageConf., 2012, pp. 1–12.

[7] M. Lillibridge, K. Eshghi, and D. Bhagwat, "Improving Restore Speed for Backup Systems That Use Inline Chunk-Based Deduplication," Proc.11th Usenix Conf. File and Storage Technologies, 2013, pp. 183–198.

[8] Z. Yan and M.J. Wang, "Protect Pervasive Social Networking Based on Two Dimensional Trust Levels," IEEE Systems J., Sept. 2014, pp. 1–12; doi: 10.1109/JSYST.2014.2347259.

[9] Z. Yan, W. Ding, and H. Zhu, "Manage Encrypted Data Storage with Deduplication in Cloud," Proc. Int'l Conf. Algorithms and Architecturesfor Parallel Processing (ICA3PP), 2015, pp. 547–561.

