ENCRYPTED DATA MANAGEMENT WITHDEDUPLICATION IN CLOUD COMPUTING

Poonam Chandwadkar[1], Priyanka Dhage[2], Manasi Sarode[3],Rashmi Sonawane[4],Prof.S.N.Bhadane[5]

[1,2,3,4]Student of IT Department Engineering,PVG's COE Nashik,Maharashtra,India

[5]Assistant Professor,Dept.Information Technology,PVG's COE Nashik,Maharashtra,India.

[1]poonamchandwadkar@gmail.com

[2] priyankadhage31894@gmail.com

[3] manasi.sarode121@gmail.com

[4]sonawane.rashmi4@gmail.com

***Abstract*—**Cloud computing, often referred to as simply the cloud, is the delivery of on-demandcomputing resourceseverything from applications to data centersover the internet on a pay-for-use basis.Cloud computing allows users and enterprises with various computingcapabilities to store and process data either in a privately-owned cloud, or ona third-party server located in a data center – thus making data-accessing mechanismsmore efficient and reliable. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility. Cloud computingplays an important role in supporting data storage, processing, and management inthe 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 areencrypted under different encryption schemes could be stored in the cloud, whichgreatly decreases the utilization rate of storage resources, especially for big data.Several data deduplication schemes have recently been proposed. However, mostof them suffer from security weakness and lack of flexibility to support secure dataaccess control. Therefore, few can be deployed in practice. We proposes a schemebased on attribute-based encryption (ABE) to deduplicate encrypted data stored inthe cloud while also supporting secure data access control.

***Keywords*—Cloud computing, Encryption, Access control, Internet of Things, authorisation, cryptography, data privacy.**

I.INTRODUCTION

Cloud computing is a method for delivering information technology (IT) services inwhich resources are retrieved from the Internet through web-based tools and applications,as opposed to a direct connection to a server. Rather than keeping files on aproprietary hard drive or local storage device,cloud-based storage makes it possible to save them to a remote database. As long as an electronic device has access to theweb, it has access to the data and the software programs to run it.It is called cloudcomputing because the information being accessed is found in "the cloud" and doesnot require a user to be in a specific place to gain access to it. This type of system allowsemployees to work remotely. Companies providing cloud services enable usersto store files and applications on remote servers, and then access all the data via the internet. In real practice, it is hard to allow data holders to manage deduplication due to a number of reasons.

First, the data holder may not be always online or availablefor such a management, which could cause a big storage delay.Second, the designedsystem for deduplication could be very complicated in terms of communications and computations.Third, it may intrude the privacy of data holder in the process ofdiscovering the duplicated data by the CSP. Fourth, the data holder has no idea how to issue access rights or deduplication keys to a user in some situations when he/shedoes not know other data holders due to data super-distribution. Therefore, CSP cannotcooperate with the data holder on data storage deduplication in many situations.However, current industrial deduplication solutions cant handle encrypted data.

Existing solutions for deduplication are vulnerable to brute-force attacks and cant flexibly support data access control and revocation .Few existingschemes for cloud data access control support data deduplication simultaneously,and few can ensure flexibility and security with sound performance for cloud datadeduplication that data owners control directly.

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.

II.EXISTING SYSTEM

The same data in an encrypted form is only saved once at the cloud but it can be accessed by different users based on th data owner's 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. Few existing schemes for cloud data access control support data deduplication simultaneously, and few canensure flexibility and security with sound performance for cloud data deduplicationthat data owners control

directly. Analysis and implementation demonstrate that our scheme is secure, effective, and efficient.

Zheng Yan,Wenxiu Ding and Haiqi Zhu proposed a practical scheme to manage the encrypted data in cloud with deduplication based on PRE. Proposed scheme can flexibly support data update and sharing with dedu-plication even when the data holder is offline. The encrypted data can be securely accessed because only authorized data holders can obtain the symmetric key used for data decryption.

III.PROPOSED SYSTEM

Data deduplication brings a lot of benefits, security and privacy concerns arise asusers sensitive data are susceptible to both insider and outsider attacks. Traditionalencryption, while providing data confidentiality, is incompatible with data deduplication.Specifically, traditional encryption requires different users to encrypt theirdata with their own keys. Thus, identical data copies of different users will lead todifferent ciphertexts, making deduplication impossible.
The idea helps to solve the issue of de-duplication in the situation where thedata holder is not available or hard to be involved.

Convergent encryption has been proposed to enforce data confidentiality whilemaking deduplication feasible. It encrypts/decrypts a data copy with a convergentkey, which is obtained by computing the cryptographic hash value of the content ofthe data copy. After key generation and data encryption,users retain the keys andsend the ciphertext to the cloud. Since the encryption operation is deterministic andis derived from the data content, identical data copies will generate the same convergentkey and hence the same ciphertext.

IV.SYSTEM DESIGN

There are 3 important modules:

- Data owner.
- Data holder.
- Cloud service provider.

Data owner:a data owner that stores its data at the CSP (we assume there is only one data owner for one data M).
CSP:CSP offers a storage service and performs honestly on data storage andmanagement to gain commercial profit but cant be fully trusted since its curiousabout the contents of stored data.
Encryption and hashing algorithm:user credentials including(PKu, SKu)and(pku,sku).This process is conducted at user u.
Data depduplication:user u is a data owner who saves data M at the CSP,using DEKuto protect the data,while user u is a data holder who tries to save the same data at that CSP.

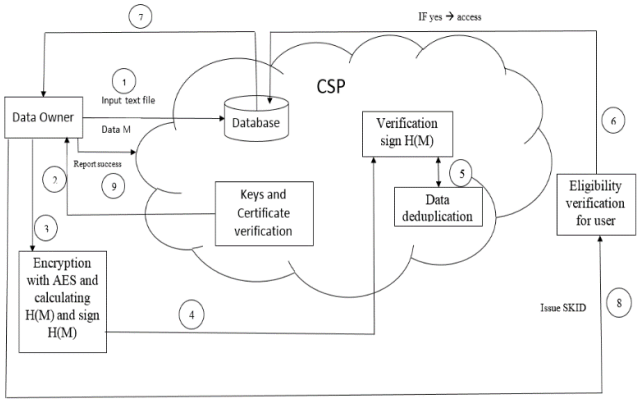

Fig.1.System architecture.

Verification of sign:After receiving DPu, the CSP verifies Cert(PKu)and Cert(pku).If the verification is positive,the CSP uses verify Sign(H(M), sku) to check ifduplicate data is saved by finding whether the same H(M) is in its storage.Ifthe check is negative, the CSP saves DPu.If the check is positive and the prestoreddata is from the same user, the CSP notifies that user.If a different useris storing the same data,the CSPperforms deduplication.
Eligibility:User u verifies the eligibility of u.If verification is positive,user u calls IssueIDSK(ID,SKu, PKu)to generateSKID(u,u)and issues SKID(u,u) tou to allow it to access M.
Zheng Yan, Mingjun Wang, and Yuxiang Li, Encypted Data Management with Deduplication in
Cloud Computing Cloud Computing, IEEE, vol. 3, issue. 2, 2016, pp. 28-35.

Output:

Fig.2.Home page.

Fig.3.Admin

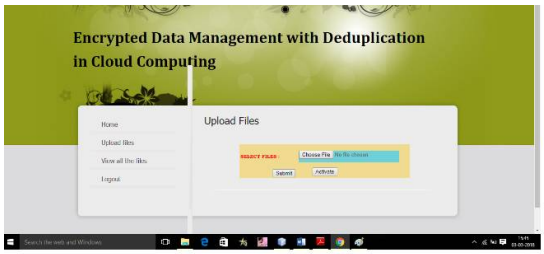

Fig.4.Upload data.

V.CONCLUSION

Deduplication of the encrypted data will be avoided which will be beneficial for saving storage space on the cloud, reduce bandwidth effects into faster access.Managing encrypted data with deduplication is significant in practice for running a secure, dependable, and green cloud storage service, especially for big data processes.

We proposed a practical scheme to manage the encrypted data in cloud with deduplication. Our proposed scheme can flexibly support data update and sharing with deduplication. The encrypted data can be securely accessed because only authorized data holders can obtain the symmetric key used for data decryption.

REFERENCES

[1] Zheng Yan, Mingjun Wang, and Yuxiang Li, Encypted Data Management with deduplication in Cloud Computing ,IEEE, vol. 3, issue. 2, 2016, pp. 28-35.

[2] [2] D.T. Meyer and W.J. Bolosky, A Study of Practical Deduplication, ACM Trans. Storage, vol. 7,no. 4, 2012, pp. 120.

[3] [3] M. Bellare, S. Keelveedhi, and T. Ristenpart, Message-Locked Encryption and Secure Deduplication,Advances in Cryptology (EUROCRYPT 13), LNCS 7881, 2013, pp. 296312.

[4] J.Li et al., A Hybrid Cloud Approach for Secure Authorized Deduplication, IEEE Trans. ParallelDistributed Systems, vol. 26, no. 5, 2015, pp. 12061216.

[5] Z. Yan,W. Ding, and H. Zhu, Manage Encrypted Data Storage with Deduplication in Cloud, Proc.Intl Conf. Algorithms and Architectures for Parallel Processing (ICA3PP), 2015, pp. 547561.

[6] Opendedup, http://opendedup.org/

[7] Meyer, D.T., Bolosky, W.J.: A Study of Practical Deduplication. ACM Transactions on Storage,vol.7, 2012, pp.1-20.

[8] Wilcox Z.O.: Convergent Encryption Reconsidered, 2011. http://www.mailarchive.com/cryptography@metzdowd.com/msg08949.html

[9] Puzio, P., Molva, R., Onen, M., Loureiro, S.: ClouDedup: Secure Deduplication with En-cryptedData for Cloud Storage. In: 5th International Conference on Cloud Computing Technology andScience, 2013, pp. 363-370.

[10] Fu, M., Feng, D., Hua, Y., He, X., Chen, Z.N., Xia, W., Huang, F., Liu, Q.: AcceleratingRestore and Garbage Collection in Deduplication-based Backup Systems via Exploiting HistoricalInformation. In: 2014 USENIX Annual Technical Conference. 2014, pp. 181-192.

[11] Kaczmarczyk, M., Barczynski, M., Kilian, W., Dubnicki, C.: Reducing Impact of Data FragmentationCaused by In-line Deduplication. In: 5th Annual International Systems and StorageConference. 2012, pp. 15:1-15:12.

[12] Lillibridge, M., Eshghi, K., Bhagwat, D.: Improving Restore Speed for Backup Systems thatUse Inline Chunk-based Deduplication. In: FAST, 2012, pp. 183-198.

[13] http://www.investopedia.com/terms/c/cloud-computing.asp

[14] Sunil S, A Ananda Shankar : Fast and Efficient Cloud Data Utilization with Deduplication. InInternational Journal of Emerging Research in Management & Technology, June 2017, pp. 68-72.

[15] http://www.unixwiz.net/techtips/iguide-crypto-hashes.html