Heterogeneous Data Storage Management Technique including De-duplication in Cloud

A.Saikumar[1], J. Srikanth[2], M. Ashok[3]
[1]*CSE Department,Sri Chaitanya Technical Campus,Ibrahimpatnam,Telangana,India*
[2]*CSE Department,Sri Chaitanya Technical Campus,Ibrahimpatnam,Telangana,India*
[3]*Asst. Prof, IT Department, St Martin's Engineering College, Secunderabad-14, Telangana, India*

Abstract: The Cloud storage is the very popular and important services in the cloud computing. It can support cloud users to overcome the rules of given resources and without enhancing their systems it can elabotare its storage[5]. It will give guarantee the security and privacy of cloud users. Data can be outsourced in encryption form for always. The encrypted data could be incurring more waste of cloud storage and it is very complicated when data sharing among authorized users. Till now we are facing problems with the deduplication[3] on encrypted form of data storing and also the management. Our existed deduplication methods are consontrating on specific application scenarios only, this deduplication is totally controlled by either authorized persons or cloud servers. Then they will not satisfy several demands of the data owners corresponding to the data sensitivity level. In this paper, we suggest the homogeneous data storage management, It can offers flexibly both the deduplication management and also the control access simultaneously over the multi-Cloud Service Providers[2][4]. We are evaluating its performance with the analysis of security for implementation and comparison. The results show its efficiency, security and efficacy towards usage of potential practical.

Keywords: *Cloud Computing, De-duplication, Heterogeneous*

I. INTRODUCTION

Cloud computing provides data storing in centralizing manner and it can also provides online accessibility for the computer services. It is giving modern IT services by rearranging the several resources and it is providing them based on their demands to the users. Cloud computing having the properties are scalability, fault tolerance, pay per use and elasticity. Data storage service is mostly used in world wide for consuming the cloud service. Cloud user can access the data from the any place in the world from the cloud. Cloud users are getting the benefits fully when they store lot of data in the cloud. Cloud computing service providers can offers the storaging the data in cloud even having the problems in cloud. Different type of data storaging in the cloud giving the dissimilar security depending on the sensitivity of the data which we store. Generally people can store the data in cloud having several ways, they are personal data storing, sharing the data with the public and also can share the data with in the specific groups. So, they need protection to their data which is storing in the cloud. And also need the data deduplication[3]. The technique deduplication is useful to detect and reducing the duplicate data. It is most useful technique for users of the cloud.

Pliable deduplication technique in cloud with the control access of the data is still an open issue. Mimeographed or reproduced data might be stored in the form of encrypt in cloud with the selfsame user or may different, in same or dissimilar providers. From the perspective of similarity, it is very hugely anticipated that the data similarity may amalgamate with the accessibility and Control. This copied data is only stored once in the cloud, but it can be accessed by the unauthorized users based on the policies of data holders. Eventhough can have storage space is huge, redundant data storage could greatly waste network resources, consume plenty of power energy, increase operation costs, and making the datamanagement complicated[9]. Economic capacity will incredibly profit CSPs by diminishing their activity costs and contrarily advantage cloud clients with decreased administration expenses. Certainly, data deduplication would be particularly important for bigdata and it's storage management. Whatever be the literature is still having lot of studies on the formable cloud data deduplication over the multi cloud service providers. Existing technique or scheme cannot suggest the comprehensive solution to abutment both the deduplication and also the control access in comfortable and uniforming way over the cloud.

II. RELATED WORK

A. Access control on encrypted data

Existing researches are proposed beyond outsourced it to the cloud in concerning to anticipate the data confidentiality from being invaded in the cloud service provider. Access control on data which is encrypted then it requesting that only authorized substance can decrypt the encrypted data. An absolute accession is each data to encrypt once and problem related keys to authorized entities one time only. Anyhow cause of to the changeability of trusted association, key management will be difficult due to constant key update. Lists of Access Control are applied to arrange security to data in a distrusted or semi trusted party. Thereby, this approach is impractical to be applied in many real applications where the trusted association among the different users changes frequently.

B. Encrypted Data Deduplication

This is a warmful investigation matter to accommodate deduplication and client side encryption. Existing industrialized methods are failed to perform deduplication on encrypt data for example drop box, google drive, and mozy.

Message Locked Encryption is proposed to conclused this tension. Convergent Encryption(CE) is the most embossed appearance of MLE, was introduced. As a result, CE can ensure high security only when the underlying data is drawn from a large space that is too big to exhaust. What's more, CE can't bolster information get to controlled by information proprietors, and additionally other approved gatherings. They are generated depends on the data with an oblivious Pseudo Random Function (PRF) protocol. The KSis separated from a Storage Service. Users secure themselves to the KS without leaking any of their data. Thus, high security can be assured if the KS is not Accessible to attackers. Eventhoughboth KS and SS are compromised, Duple Scan still preserve the security of stored data depends on the guarantee of MLE. But some data owners do not like to authorize a third party like KS to control their data, since in some specific situations they prefer to manage the storage and access of their data by themselves and keep trace of data storage and usage status. However, DupLESS cannot support this desirable feature. And distributing its shares among multiple servers[6][7]. However, it still cannot avoid the innate drawbacks of CE.Wen et al. constructing the session key depends on the convergent key management method and the convergent key sharing method to clarify the probloms that encrypted data will blocks and the data ownership is frequently changed. But this work requests all data owners to communicate with each other to managing their session key. The CE problem still occurs. Even though this scheme requesting the data owner and it is available always online for data ownership check and deduplication[7]. This approach does'nt handles the problems that the authorized person of the data is not available, in practically it is common. Cross CSP was not discussed in this work. The above Schemes cannot flexibly manage data deduplication in various situations and across many CSP's. They cannot solve the issues as described in the introduction. Neither can they support the management of digital rights.

III. RESULT ANALYSIS

Our scheme security is relies based on the ABE technique, and the PRE technique, symmetric key technique (for encryption pepose) and PKC technique. In the existed work PRE and ABE security provided. But this will giving the flexible control access and encrypted dat for the security perpose. These two thesis are plays the vital role in data encryption. These two thesis are much enough to give the security for long period. We analyses the security of this scheme regarding to the data ownership verification and data deduplication.

Person	This paper	Previous Paper
Authorized person/Data owner	O(n)	O(1)
CSP	O(n)	O(n)
Data Holder	O(1)	O(1)

When we are comparing the our work with the previous scheme. Majorly we identify one thing that is our scheme is providing the security to the data stored in the clod. And also data access and controlivity, we can call combiningly access control. This access control is managed by the data owner or by the third parties when the data owner is busy. But in the previous scheme it will not giving the protection and security to cloud data of the data users. And also access control is not good, it does not providing the easy data access and access control. It is not better to big data storages in the cloud. When the people are storing large amount of data in the cloud then previous scheme does not support, our scheme will support for cloud storage. We can see that the proposed method is a heterogeneous solution[7][8]. It can realize the both fine grained and offline access control also, that's why it has better flexibility than previous work. In addition, the random hash code challenge is applied to verifying the data ownership, which will give guarantee that the data holders really have the original data rather than its hash code[4]. Though possession proof is achieved in by applying Elliptic Curve Cryptography, hash code set employed in this paper is also very efficient if we make challenged part of data is small.

IV. CONCLUSION

Data deduplication is the very important and significant aspect in the cloud data storage, mainly for Big Data[8]. Here we are proposing the heterogeneous scheme for data storage management and it offers the pliable to the cloud data for deduplication and also the access control. Our scheme can follow to several application synopsis and arrogations and economic offersfor big data storage management over the many cloud service providers[3]. It can achieve the access control and data deduplication with the various security requirements. Security analysis, comparison with existing thesis and implementation depends on the evaluation of the performance that our thesis is very secure, advanced and most efficient.

V. REFERENCES

[1]. Sujansky, Walter (August 2001). "Heterogeneous Database Integration in Biomedicine". Journal of Biomedical Informatics. **34** (4): 285–298. doi:10.1006/jbin.2001.1024. Retrieved 30 July2012.
[2]. "The World's Technological Capacity to Store, Communicate, and Compute Information". MartinHilbert.net. Retrieved 13 April 2016.
[3]. "Data, data everywhere". The Economist. 25 February 2010. Retrieved 9 December 2012.
[4]. "Understanding Data Deduplication" Druva, 2009. Retrieved 2013-2-13
[5]. *CHRISTIAN CACHIN; MATTHIAS SCHUNTER (December 2011).*
[6]. "It's Time to Explore the Benefits of Cloud-Based Disaster Recovery". Dell.com. Retrieved 2012-03-26.
[7]. Winkler, Vic (2011). Securing the Cloud: Cloud Computer Security Techniques and Tactics. Waltham, MA USA: Elsevier. pp. 65, 68, 72, 81, 218–219, 231, 240. ISBN 978-1-59749-592-9.
[8]. Gartner Market Guide to CASB
[9]. https://www.bitglass.com