SECURE DATA DEDUPLICATION SYSTEM WITH TAG CONSISTENCY IN CLOUD DATA STORAGE

[1]Pramod G Patil, [2]Aditya Dixit, [3]Aman Sharma, [4]Prashant Mahale, [5]Mayur Jadhav
[1]Assistant professor, [2]Final Year Student, [3]Final Year Student, [4]Final Year Student, [5]Final Year Student
[1]Computer Department,
[1]Sandip Institute of Technology and Research Centre (SITRC).Mahiravni, Nashik, India

Abstract: Cloud computing technique that is most generally used nowadays, in that clouding up is completed over giant communication network like web and provides massive space for storing in all sectors like administrative unit, private enterprises etc. and additionally store personal information on cloud. But the foremost necessary drawback in cloud is that giant quantity of space for storing is needed and conjointly duplicate copies of knowledge is store on cloud .There are a unit several techniques that is employed for eliminating duplicate copies of continuation information .From that one in all the necessary technique is information deduplication, Information Deduplication is specializes data compression techniques for removing duplicate copies of continuation data and has been wide utilized in cloud storage to chop back the quantity of space for storing and save system of measurement. To safeguard the confidentiality of sensitive data on cloud, the confluent encryption technique is utilized to cipher the data before outsourcing. In projected system used secrete sharing theme. In that files unit of measurement divided into vary of blocks referred to as shares and this shares store on the assorted vary of nodes. By victimization recover technique varies of shares square measure combine into line. Which suggests victimization this scheme provide data security, confidentiality, data responsibility.

Index Terms: **- Cloud storage, Deduplication, Network, encryption.**

I. INTRODUCTION

Due to duplicated information on public cloud, its storage efficiency and information measure efficiency isn't used properly and this is often main focus of development. To avoid duplication, information on public cloud is compared to the uploaded information. However main drawback is encrypted information on cloud cannot compared thence main task to match encrypted information. The deduplication check needn't to be the user specific it got to check the de-duplication at cloud storage level information is shared among multiple users thus there got to be some constraint that prohibits the de-duplication check for unauthorized user. The software will be generating convergent keys for encryption and maintaining tag consistency. It will also deduplicate the files in the given cloud using Block Deduplication. It will also handle the dynamic ownership management for the cloud service. The software will encrypt the files using converging keys alongside deduplication. It will increase the speed of service and reduce the bandwidth use. Application of the Software: Cloud Service Providers. And methodologies used to solve the problem and efficiency issues are: Symmetric encryption, Convergent encryption, Proof of ownership and Shamir secret sharing, using all the above together will increase the tag consistency in the current available systems and will also increase the efficiency. After applying this software to the cloud service the duplicated files will be deleted and the files will be encrypted with convergent keys according to hast value of file. The files will have consistence tags to increasing throughput. This will also help in levitating the speed of cloud service.

II. LITERATURE SURVEY

[1] In order to keep data privacy against inside cloud server as well as outside challengers, users may want their data encrypted. However, conventional encryption under different users' keys makes cross-user de-duplication impossible, since the cloud server would always see different ciphertexts, even if the data are the same, regardless of whether the encryption algorithm is deterministic. Douceur [2] introduces Convergent Encryption, which is the promising solution to this problem.

[2] Bellare [3] introduces an idea of message-locked encryption (MLE), with its security approach to solving the problem of CE. He also proposed randomized convergent encryption (RCE) as one application of MLE which provides a technique to achieve secure de- duplication. In RCE, initial uploader encrypts a message using a random encryption key and it results into a ciphertext refer as C1. This message encryption key is again encrypted along with a key encrypted key (KEK) which is derived from the message by using hash function and results into a ciphertext refer as C2. Here message tag is generated from the KEK.

[3] Xu[4] proposes a leakage-resilient de-duplication scheme to solve the data integrity issue. It addressed a vital security concern in cross-user client-side de-duplication of encrypted files in the cloud storage: privacy of users' sensitive files against both outside challengers and the honestbut-curious cloud storage server in the bounded leakage model. Instead of encrypting the convergent keys on a per-user basis, Dekey builds secret shares on the original convergent keys (that are in plain) and assigns the shares over various KM- CSPs. If many users share the identical block, they can access the same corresponding

convergent key. This significantly decreases the storage overhead for convergent keys. In addition, this method provides fault tolerance and allows the convergent keys to remains accessible even if any subset of KM-CSPs fails.

III. EXISTING SYSTEM

a. No Security Concerns in the existing system.
b. In existing system, a single uploaded file into cloud can be uploaded 'n' number of times into the server.
c. Makes the server with duplicate copy of file and it will attack by anyone.
d. Confidential level is decreased.

IV. PROPOSED SYSTEM

a. The proposed system uses hash function to avoid the duplication in cloud.
b. High efficient.
c. Elliptic Curve Cryptographic (ECC) algorithm is used for encryption and decryption process.
d. Proposed system proposes an efficient group key management protocol in distributed group communication.
e. ECC algorithm provides high end security. Avoid duplication in cloud.

V. ARCHITECTURE DIAGRAM

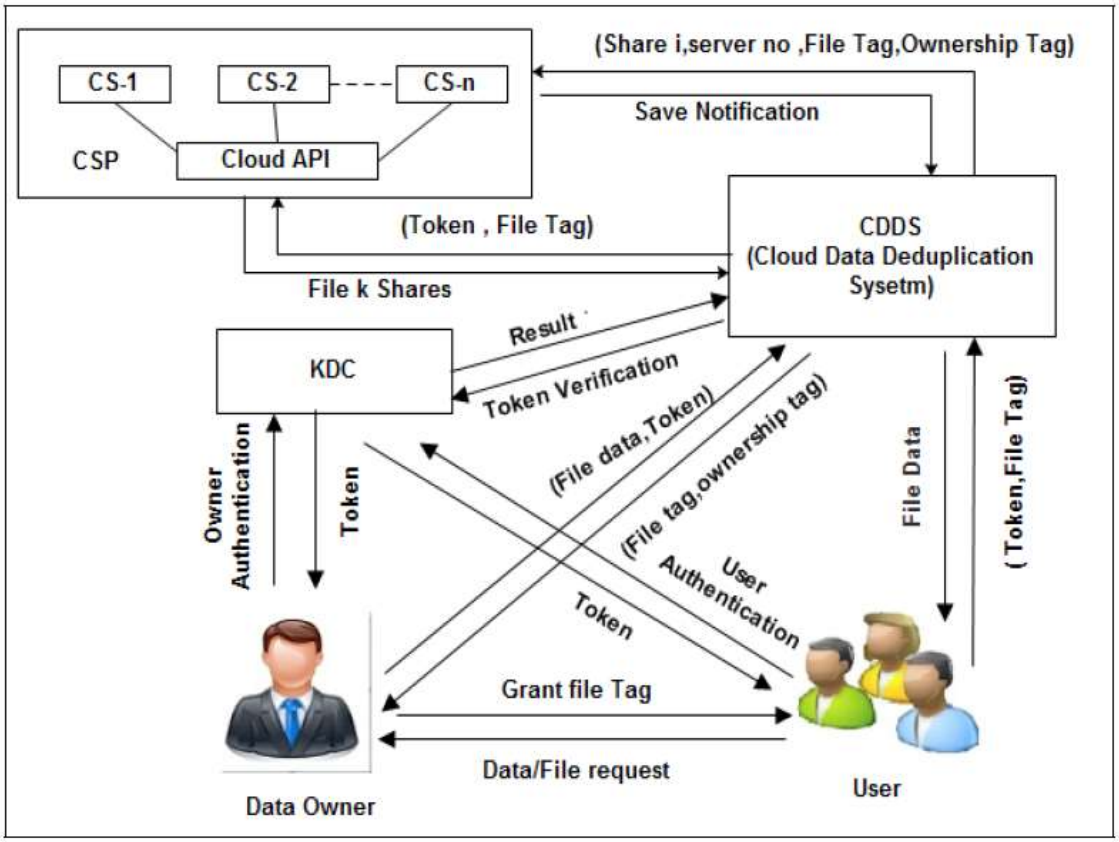

Figure.1 Architecture diagram

VI. APPLICATIONS

The software will be primarily applied by cloud service providers who are providing cloud services to the big number of users. E.g. Amazon cloud, Drop-Box, Google Drive

VII. CONCLUSION

In this paper, we have reviewed different data deduplication techniques over encrypted data that are used in the cloud computing for secure data storage. Traditional encryption makes deduplication impossible because of the randomization property of encryption.

Recently, several deduplication schemes are proposed to solve this issue by allowing each owner to share the same encryption key for the same data. Convergent encryption has different encryption variants for secure deduplication which was formalized as MLE later in. Though, CE suffers from security flaws with regard to tag consistency and ownership revocation.

ACKNOWLEDGMENT

Gives us great pleasure in presenting the preliminary paper on "Secure Data Deduplication System with Tag Consistency in Cloud Data Storage", we would like to take this opportunity to convey our gratitude to internal guide Prof. P. G. Patil for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. Dr. A. D. Potgantwar, Head of Computer Engineering Department, SITRC for his indispensable support, suggestions.

REFERENCES

[1] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. Hassan, and A. Alelaiwi, "Secure Distributed Deduplication Systems withImproved Reliability," IEEE Transactions on Computer, Vol. 64, No. 2, pp. 3569–3579, 2015.
[2] R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, "Reclaiming space from duplicate files in a server less distributed file system," Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," Proc. Eurocrypt 2013, LNCS 7881, pp. 296–312, 2013. Cryptology ePrint Archive, Report 2012/631, 2012.
[4] Xu, E. Chang, and J. Zhou, "Leakage-resilient client-side deduplication of encrypted data in cloud storage," ePrint, IACR, http://eprint.iacr.org/2011/538.
[5] Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 6, 2014.
[6] Pramod Gorakh Patil, Vijay Kumar Verma,"A Recent Survey On Different Symmetric Key Based Cryptographic Algorithms".IJCAT Feb 2016
[7] X. Jin, L. Wei, M.Yu, N.Yu and J. Sun, "Anonymous deduplication of encrypted data with proof of ownership in cloud storage," Proc. IEEE Conf. Communications in China (ICCC), pp.224-229, 2013.
[8] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.