Secure Data Protection With Deduplication In Cloud Computing

Shital D. Kale[1], Laxmi P. Thite[2], Pooja V. Mohite[3], Prof. Inamdar I.Y.[4], Prof Pandhare S.D.[5]
[1, 2, 3, 4, 5] Dept of Computer Science & Engineering
[1, 2, 3, 4, 5] SMSMP Institute of Technology & Research, Akluj, India

Abstract- *Sharing encrypted data with different users via public cloud storage is an important research issue. This paper proposed the concept of Data Protection with Deduplication in Cloud Computing. In this model, the input plain text is placed into a block-128 in a specific manner, and key is calculated. By using this key, the plain text is transformed into intermediate cipher text. In the initial stage the scope of the project will be to provide Internet access through limited number of users and provide centralized management through a server. In addition, if user uploading the same file which is already uploaded by another user then it is not accepted by the system and it shows a massage aboutexisting file& owner of that file with link so user can access that file. So this process in our scheme solves the Deduplication problem of data sharing & save the storage and the bandwidth of network. Data owner can extract a key which includes cipher texts' indices, delegate's identity and expiration date of the key. The cloud server is obtains the identity of download-applicant from the key with public parameter and then controls download right. In order to achieve efficient and secure data sharing in dynamic cloud storage, the method should be proposed stable in expense, and should be leakage-resilient. Our scheme can satisfy both requirements.*

Keywords- Encryption, Key, cloud computing, data sharing, Deduplication

I. INTRODUCTION

The cloud computing is the widely used technology that enables the virtual storage, it means renting the space from unknown places. Also Mining technique is applied for checking the matching of contents presents on a cloud.

Cloud computing greatly facilitates data providers who want to outsource their data to the cloud without disclosing their sensitive data to external parties and would like users with certain credentials to be able to access the data. This requires data to be stored in encrypted forms with access control policies such that no one except users with attributes (or credentials) of specific forms can decrypt the encrypted data. An encryption technique which meets the Cipher Text attribute-based encryption (CP-ABE) a user's private key is associated with an attribute and a message is encrypted under an access structure over a set of attributes. Also, a user can decrypt a cipher text with his/her private key if his/her set of attributes satisfies the access policy associated with this cipher text. In spite of, the accepted ABE system decline to manage highly secure duplication, which is a to save storage, space, network, and bandwidth by eliminating redundant copies of the encrypted data which is stored in the cloud. To the best of our knowledge, the existing constructions for secure de-duplication are not built on attribute-based encryption. Nevertheless, since ABE and secure de-duplication have been widely applied in cloud computing, it would be desirable to design a cloud storage system possessing both properties.

With the help from Amazon S3 cloud, bulk amount of data (Big-data) can be handled for storing on it. Sometimes it may causing with storage running out properties based on the same data replication. In previous they are work with the avoidance of duplication based on the attributes of the particular data/file it may be name, extension of the file.

For the secure de-duplication we may go with ciphering techniques. Here the another concern for space after de-duplication is compression of files based on Huff-Mann compression technique. Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys [2]. Here in this paper security plays the major role for protecting the data/file over the Amazon s3 cloud.

II. MOTIVATION

We observe in research paper that with the rapid growth in database, networking, and computing technologies, such data can be Integrated and analyzed digitally. The one hand, this has led to the develop of data mining tools that aim to infer useful trends from this data. But, on the other hand, easy access to personal data causes data loss so it leads to their information loss, which typically refers to the amount of critical information preserved about the datasets after the perturbation. Thus, we need to work towards minimizing both privacy loss and information loss.

III. LITERATURE SURVEY

1.] Cheng Guo, Yingmo Jie (2017)." Key-Aggregate Authentication Cryptosystem for Data Sharing in Dynamic Cloud Storage**" ,**14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing.

The scheme solves the secret-key leakage problem by setting up an efficient identity authentication. The key-aggregate authentication cryptosystem scheme supports secure, efficient and flexible data sharing via cloud storage. The KAACscheme can be used in other scenario as patient-controlled encryption, searchable encryption via cloud storage and so on. Flexible and leakage-resilient delegation scheme with compact keys will have more and more prospects for use.

2.] Priyanka G. Masal ,B.M.Patil(2017)." Encrypted Big Data with Data Deduplication in Cloud", International Journal of Computer Applications (0975 – 8887).

This scheme can flexibly support data update and sharing with deduplication even when the data holders are offline. Only authorized data holders can use the symmetric keys used for data decryption by encrypted data can be securely accessed. The performance analysis and graphs showed that there scheme is secure under the described security model for big data deduplication. Discuss about to managing encrypted data with deduplication is for achieving a successful cloud storage service, for big data storage. In this paper, proposed a scheme to manage the encrypted big data in cloud with deduplication based on PRE and ownership challenge

3.] Chaur-Chin Chen. (2004). RSA scheme with MRF and ECC for data encryption.Multimedia and Expo, 2004.ICME '04. 2004 IEEE International Conference on, 2, 947-950.

This paper combines schemes of cryptography with steganography for hiding secret messages. Given secret messages, for example; an English sentence, OUT scheme J h t converts the messages to an 1cf x N binary image which is then covered by a biliary random texture synthesized from a 20 Ising Markov random field using the seed, a shared secret key, between the sender and the receiver, generated by the strategy of elliptic curve cryptography (ECC). The concealed messages are then encrypted based on the RSA scheme for transmission. An experiment shows that using an unauthorized key gets messages totally different from the original ones even the error key is very close to the authorized one.

4.]Som S., Banerjee M., (2013) “Cryptographic Technique by Square Matrix and Single Point Crossover on Binary Field”, 1stInternational Conference on Communications, Signal Processing, and their Applications (ICCSPA'13), IEEE Explorer, Print ISBN: 978-1-4673-2820-3, February 12 – 14, 2013, Sharjah, UAE.

In this paper new cryptographic algorithm is introduced. This technique uses three keys for encryption and decryption. A nearby square matrix with few column cells is used to place the input plain text in a unique manner. The leftdiagonals positional value will be the key-1 with that key intermediate cipher text is produced. A 7-digit random number is generated as key-2. According to the digits of key-2 the section division, the block division process and the crossover point is finalized. Uniform point crossover is applied on the binary field of intermediate cipher text to produce complex final cipher text.

5.]Yang, C., & Lin, Y. (2009). Reversible VQ Data Hiding Based on Locally Adaptive Coding and Recursive Walking. Computer Science and Its Applications, 2009.CSA '09. 2nd International Conference on, 1-6.

This paper present a new LSB-based steganography technique to embed a secret image in a cover image while keeping the perceptual degradation of the cover image to a minimum level so as to avoid visual attacks. To achieve high embedding capacity we propose to use the YIQ color space model and RGBA based cover image. RGB values of the secret image pixel’s is first converted into YIQ color space, which are then embedded into the least significant bits of the color pixels of the cover image, as well as in the alpha channel. Results obtained demonstrate that it is practically possible to hide an image in another image while maintaining acceptable image quality.

IV. PROPOSED METHODOLOGY

In the proposing system, we .eliminating duplicate copies of repeating data and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the privacy of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing .To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication.

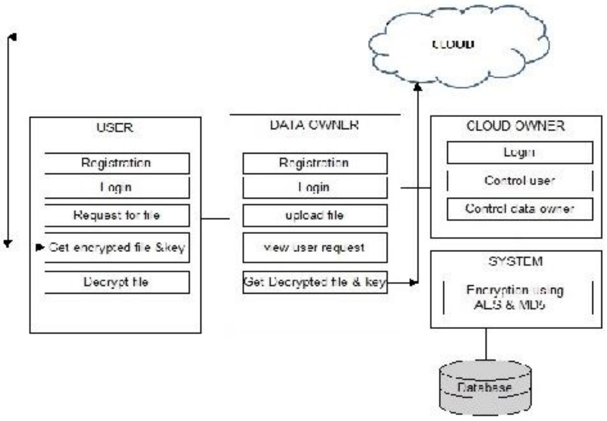

Fig: Architecture Diagram

An encryption technique that meets this requirement is called attribute-based encryption (ABE), where a users private key is associated with an attribute set, a message is encrypted under an access policy (or access structure) over a set of attributes, and a user can decrypt a ciphertext with his/her private key if his/her set of attributes satisfies the access policy associated with this ciphertext. However, the standard ABE system fails to achieve secure deduplication, which is a technique to save storage space and network bandwidth by eliminating redundant copies of the encrypted data stored in the cloud. On the other hand, to the best of our knowledge, existing constructions for secure deduplication are not built on attribute-based encryption. Nevertheless, since ABE and secure deduplication have been widely applied in cloud computing, it would be desirable to design a cloud storage system possessing both properties.

V. RESULT AND DISCUSSION

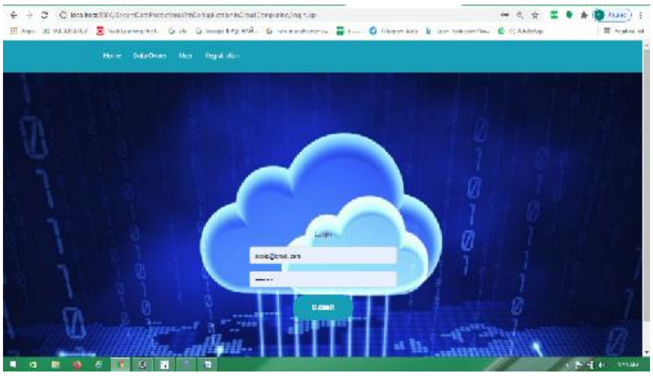

Fig. Data Owner Login

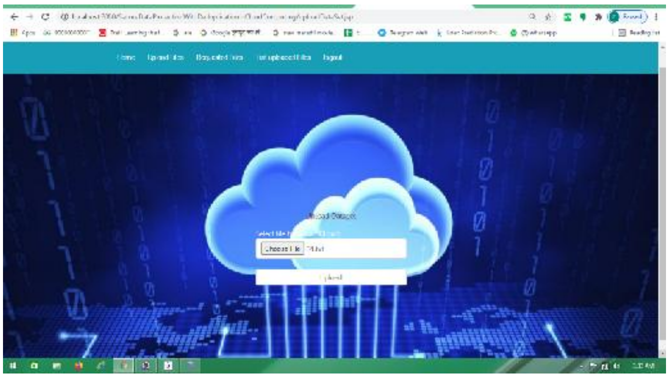

Fig. Upload file

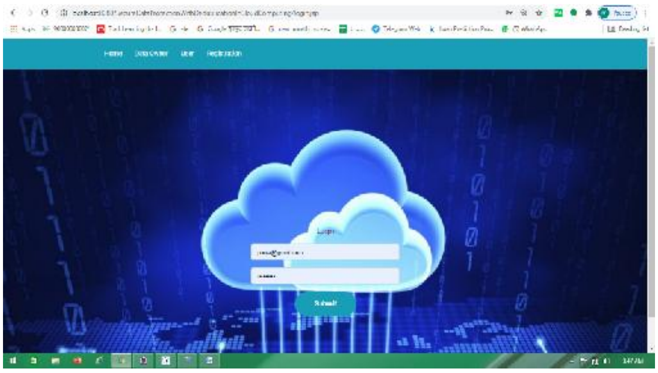

Fig. User Login

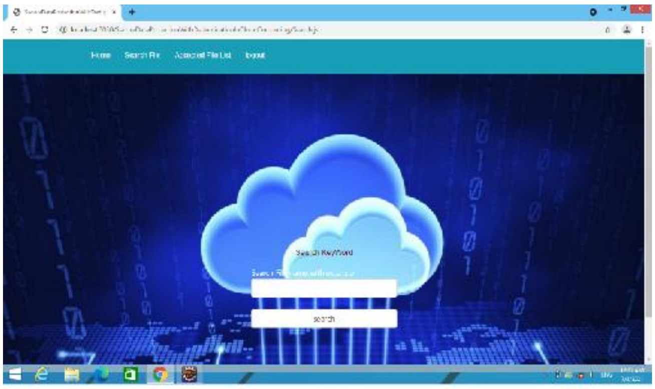

Fig. Search file

VI. CONCLUSUON

Here the Encryption and Decryption algorithm are publically available .The privacy of data transmission depends only on secret of key and while sharing data it is compromised .This Application widely used in cloud computing where data providers resource with their encrypted data to the cloud & share data with user. On other hand duplication of data is an important issue to save storage space & bandwidth of network .We dismisses the duplicate data so the duplicate data store only once in the cloud .It has been shown that how data is secured using the algorithms .We use encrypted data and extract useful information from the cipher texts. After proposing a frame-work for this purpose, a simple case study was proposed, illustrating how this frame-work may be employed for encrypted data classification. Recommendations have been presented to improve the security of the cryptosystems against the attack. In addition we avoid the duplicate file giving access to that file which user want to upload but already exists. so we reduce the deduplication of data & increase storage space and bandwidth of network.

VII. FUTURE SCOPE

In the initial stage the scope of the project will be to provide Internet access through limited number of users and provide centralized management through a server.

- This would provide excellent security to large amount of data for example Military, Patients data etc.
- Since this potentially provides a billion eyeballs for views.
- This type of encryption is provided to large organization.

REFERENCES

[1] Cheng Guo , Yingmo Jie (2017)"Key-Aggregate Authentication Cryptosystem for Data Sharing in Dynamic Cloud Storage" , 2017 14th International Symposium on Pervasive Systems, Algorithms and Networks & 2017 11th International Conference on Frontier of Computer Science and Technology & 2017 Third International Symposium of Creative Computing.

[2] Priyanka G. Masal , B. M. Patil (2017) "Encrypted Big Data with Data Deduplication in Cloud" , International Journal of Computer Applications (0975 – 8887) Volume 174 – No.6, September 2017.

[3] Som S., Banerjee M., (2013) “Cryptographic Technique by Square Matrix and Single Point Crossover on Binary Field”, 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA'13), IEEE Explorer, Print ISBN: 978-1-4673-2820-3, February 12 – 14, 2013, Sharjah, UAE.

[4] Yang, C., & Lin, Y. (2009). Reversible VQ Data Hiding Based on Locally Adaptive Coding and Recursive Walking. Computer Science and Its Applications, 2009.CSA '09. 2nd International Conference on, 1-6.

[5] Khadivi, P. &Momtazpour, M. (2009). Application of data mining in cryptanalysis. Communications and Information Technology, 2009.ISCIT 2009. 9th International Symposium on, 358-363.

[6] Chaur-Chin Chen. (2004). RSA scheme with MRF and ECC for data encryption. Multimedia and Expo, 2004.ICME '04. 2004 IEEE International Conference on, 2, 947-950.

[7] Farouk, H., & Saeb, M. (2005). An improved FPGA implementation of the modified hybrid hiding encryption algorithm (MHHEA) for data communication security. Design, Automation and Test in Europe, 2005. Proceedings, 76-81.

[8] Murugeshwari, B., Sarukesi, K., & Jayakumar, C. (2010). An Efficient Method for Knowledge Hiding Through Database Extension. Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on, 342-344

[9] Som S., Mitra D., Halder J., (2008) “Session Key Based Manipulated Iteration Encryption Technique (SKBMIET)”, IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE 2008), ISBN No.: 978-0-7695-3489- 3, pp: 694-698, 20-22, December 2008, Phuket, Thailand

[10] Shirali-Shahreza, M. (2008). Text Steganography by Changing Words Spelling. Advanced Communication Technology, 2008.ICACT 2008. 10th International Conference on, 3, 1912-1913.

[11] Hadhoud, M., Ismail, N., Shawkey, W., & Mohammed, A. (2004). Secure perceptual data hiding technique using information theory. Electrical, Electronic and Computer Engineering, 2004.ICEEC '04. 2004 International Conference on, 249-253.