ISSN 2395-1621

# Data Protection with Deduplication In Cloud Computing

#1Sonali P. Kabade, #2Dr.B.K.Sarkar

1kabadesonali29@gmail.com
2dr.bksarkar2003@yahoo.in

#12Department of Computer Engineering, Savitribai Phule Pune University,
TSSM's PVPIT, Bavdhan, Pune, Maharashtra, 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 about existing 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.**

ARTICLE INFO

**Article History**

Received: 15th July 2019

Received in revised form :

15th July 2019

Accepted: 18th July 2019

**Published online :**

**19th July 2019**

## I. INTRODUCTION

Over the past few years, there has been a tremendous growth in the amount of private data collected about individuals that can be collected and analyzed. This data comes from a variety of sources including medical, financial, library, telephone, and shopping records. With the rapid growth in database, networking, and computing technologies, such data can be Integrated and analysed 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 poses a threat to individual privacy the goal is to perform data mining operations on sets of data without disclosing the contents of the sensitive data. Since the results of the mining inform us something about the data, some information about the original data is leaked to the mining results. This leads to privacy loss. If the data is perturbed on the other hand for privacy concerns, 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.[3]

The main goal in cryptography is to conceal information from non-authorized parties. However, crypt-analysers try to find weaknesses in cryptography algorithms to obtain meaningful information from the encrypted data. With modern cryptography algorithms, even finding one bit of information is valuable from cryptanalysis perspective. Different forms of information around the world are entangled with known or unknown patterns and characteristics. Also, it is almost clear that the amount of information in the plain and cipher texts are equal.

Cryptosystems conceal information through data recoding. Hence, information and the related features still exist in the cipher texts. Data mining is the process of analysing the data from different viewpoints and generating useful information. It is the process of extracting hidden useful patterns from data. These patterns and information can then be used to improve our understanding of different economical, social, or engineering systems to increase the incomes, decrease the costs, or improve the behaviours[6]

Their a new algorithm for encryption and decryption is introduced . In this proposed 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 about existing 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 .In order to improve the security of cryptography systems we reduce deduplication of data and make the system more usable.

## II. LITERATURE SURVEY

1.] 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 is 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 their square matrix with few column cells is used to place the input plain text in a unique manner. The left diagonals 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.

2.] Encrypted Big Data with Data Deduplication in Cloud is proposed by Priyanka G. Masal .The proposed system discuss a safe K-NN classifier over center information on cloud. For this they uses vector base cosine similarity (vcs) algorithm to find out matching data. Their aim is to maximize space & minimize redundant data in cloud so the duplicate data storage only once in cloud. If user wants to upload a file in cloud which is already in system then providers shows owner list to the user. So they achieved cost savings & high space in the deduplication

3.] 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, 242-249

In this paper they introduce the scheme which 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 KAAC scheme 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.

## III. PROPOSED METHODOLOGY

### *A.* Problem Statement

We exhibit the execution of encryption and decoding algorithms data privacy, computational efficiency and effectiveness of the cloud storage system. We demonstrate novel approach of Data Protection with Deduplication in Cloud on huge data stored on cloud. We are providing Data Protection with Deduplication in Cloud Computing using key for more security and avoid duplicate data.

### *B.* Proposed System Overview

In this proposed 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.

This would provide excellent security to large amount of data for example Military, Patients data etc.

This would provide excellent security to large amount of data for example Military, Patients data etc
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 about existing 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.

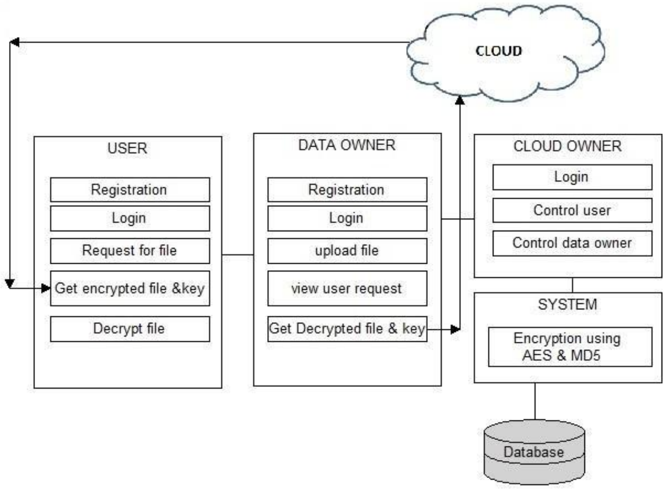


**Figure 1. Proposed System Architecture**

In this system architecture we have following modules:

- USER FUNCTIONS

1) Registration: The new user will have to register to the application.
2) login to application: The registered user thus can login to the application.
3) View file: After the process of login and registration is complete the user will be able to

download the file.
4) Download file: Now the user can download the file.
5) Decrypt file: Thus the user can decrypt the file for his purpose.

- ADMINISTRATIVE

1) Login to application: The admin need to login to the application.
2)Authentication of user: The admin will check if the user in know one.
3)Update details: Admin will update the user details.
4)Allow/Deny access: Admin will decide whether to allow the user or not.
5)View request of users: Admin will check the users list.

- DATA OWNER

1) Login to application: The data owner needs to login to the application.
2)Upload file: The owner will upload the file.
3)Encrypt file: The owner will encrypt the file.
4)View File:Owner will check the file.

### *C.* Algorithm

**Algorithm 1 : AES Algorithm**

Process:

The encryption process uses a set of specially derived keys called round keys. These are applied, along with other operations, on an array of data that holds exactly one block of data the data to be encrypted. This array we call the state array.

Step1 :We have following AES steps of encryption for a 128- bit block:
Step2 :Derive the set of round keys from the cipher key.
Step3 :Initialize the state array with the block data (plaintext). Step4 :Add the initial round key to the starting state array.
Step5 :Perform nine rounds of state manipulation.
Step6 :Perform the tenth and final round of state manipulation. Step7 :Copy the final state array out as the encrypted data (ciphertext).

**Algorithm 2 : MD5 Algorithm**

Process:

We begin by supposing that we have a b-bit message as input, and that we wish to find its message digest. Here b is an arbitrary non negative integer; b may be zero, it need not be a multiple of eight, and it may be arbitrarily large. We imagine the bits of the message written down as follows:

m_0 m_1 ... m_{b-1}

tep 1. Append Padding
Bits m_0 m_1 ... m_{b-
1}
Step 1. Append Padding Bits

Step 2. Append Length
Step 3. Initialize MD
Buffer
Step 4. Process Message in 16-Word
Blocks Step 5. Output

### Mathematical Model

Let _S‘ be the | universal final set
This will include user, resources,
system. S = {…………}
Identify the inputs
as I I = {F}
F = {F1, F2, F3, F4 …| _I‘ files to be
uploaded} Identify the outputs as O
O = {T}
K= { key …| _K‘ given key for
files} Identify the functions as F‘

S = {…
F = {F1 (), F2(), F3(), F4(), F5()}
F1 (I) = Upload file
F2 (I) = Request file to data
owner F3 (O) = Encryption
F4 (O)= Get authenticated
document F5 (O)=Decryption

## IV. RESULTS AND DISCUSSION

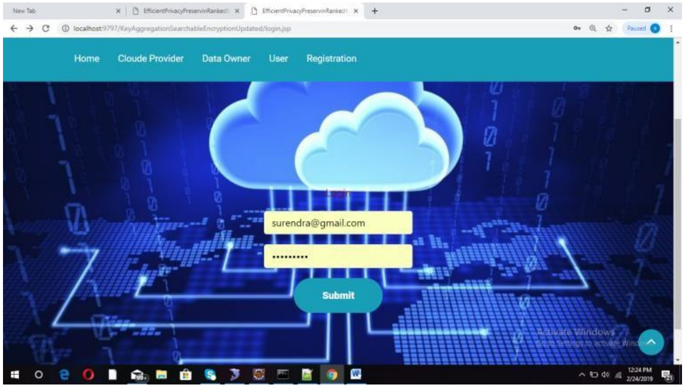
### A. Setup

Hardware and software of proposed system given below:
□ Software Technology:
1. Technology: Core Java
2. Tools: Eclipse
3. Operating System: Windows
□ 7/8/10 Hardware Technology
1. Processor: 1.0 GHz
2. RAM: 1 GB
3. Hard Disk: 730 GB

### B. Results

Results are based on the user inputs i.e. Data from the user to the cloud . Every time data need a secure authentication so here we are securing our cloud by username & password.

Every user must log-in the system every time



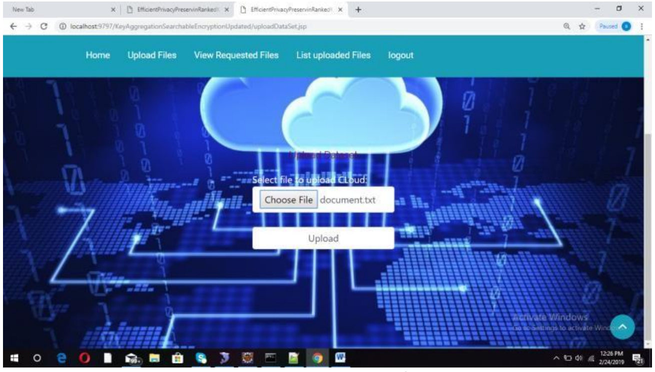
localhost:9797/KeyAggregationSearchableEncryptionUpdated/login.jsp

surendra@gmail.com

••••••••••

**Figure 2. Login Window**

After the successful login, user can upload file.



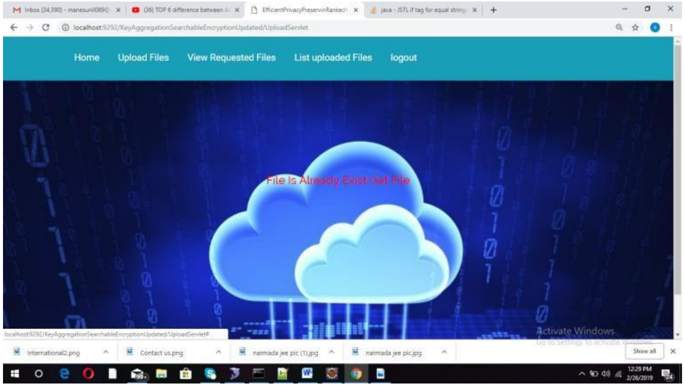
localhost:9797/KeyAggregationSearchableEncryptionUpdated/uploadDataSet.jsp

Upload Dataset

Select file to upload Cloud: document.txt

**Figure 3.Uploding  the file**

If another user wants to upload the file to cloud which is already present then user get message and link through which user can access that file. For this user get a link like "file is already exist get file".



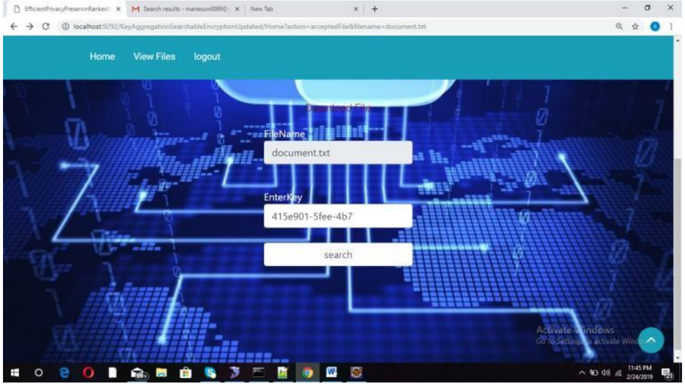
localhost:9292/KeyAggregationSearchableEncryptionUpdated/UploadServlet

File Is Already Exist Get File

localhost:9292/KeyAggregationSearchableEncryptionUpdated/UploadServlet#

**Figure 3. Access to existing file**

After Clicking on that link user can access that file using File Name & Encrypt Key.



localhost:9292/KeyAggregationSearchableEncryptionUpdated/Home?action=acceptedFile&filename=document.txt

Download File

FileName: document.txt

EnterKey: 415e901-5fee-4b7

**Figure 4. Accessing File through key**

## V. CONCLUSION

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.

## ACKNOWLEDGEMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance.

## REFRENCES

[1] 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

[2] 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.

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

[4] 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.

[5] 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.

[6] 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

[7] 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.

[8] 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.

[9] 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, 242-249.