Grenze International Journal of Engineering and Technology, January Issue

Cloud-based Privacy Preservation using AES and Data Recovery for Decision Tree Learning

Smita Shahajirao Ghorapade[1], Dr. Sandeep Gajanan Sutar[2] and Mr. Prafull Mahadev Kumbhar[3]
[1]Department of Computer Science & Engineering, Annasaheb Dange College of Engineering and Technology, Ashta, Sangli, India
Email: smita3090@gmail.com
[2]Associate Professor, Dr. D. Y. Patil Pratishthan's College of Engineering, Salokhenagar, Kolhapur, India
Email: sutarsandeep07@gmail.com
[3]Dean (Academic), Appasaheb Alias Sa. Re. Patil Institute of Technology, Shirol, Kolhapur, India
Email: engr.kumbharp@gmail.com

***Abstract*— Privacy preservation is important for machine learning and data mining, but measures designed to protect private information often result in a trade-off: reduced utility of the training samples. This paper introduces a privacy preserving approach that can be applied to decision tree learning, without concomitant loss of accuracy. The Algorithm we use in our project is the AES (Advanced Encryption Standard) Algorithm so that we can encrypt the collected data and it will be stored in cloud. It describes an approach to the preservation of the privacy of collected data samples in cases where information from the sample database has been partially lost. Data loss can have severe consequences for businesses, making data backup and recovery practices critical for protecting valuable information. Cloud storage has emerged as a powerful tool in the fight against data loss, offering secure off-site storage and robust recovery solutions.**

***Index Terms*—Encryption, Decryption, Data Storage, Security, Cloud computing, Privacy, Recovery.**

I. INTRODUCTION

Cloud and remote storage, which has become the preferred method for reducing the security risk and safeguarding our data, is widely used for backups or outsourcing. all the problems related to cloud computing we focus on user privacy issues so that data is stored securely and data is secure.

Cloud computing emerges as a paradigm for new computers and other business data protection platforms. Most of the data will be stored in the cloud after encryption so that data can be accessed anywhere and anytime without data cuts.

The cloud services that will be provided for data storage, data storage is not just an issue. planning for tedious infrastructure management activities and reducing development costs and maintaining them.

However, keeping user privacy concerns should be considered important when designing security and privacy guidelines because there is a probable for misuse of data.

Cloud computing has emerged as a successful and ubiquitous paradigm for service-oriented computing and deploying web applications.

Grenze ID: 01.GIJET.11.1.805

However, since the administrators, privileged users and malicious users with root or system rights have the full access rights to the server and data, we must encrypt the private and sensitive data before storing them in this untrusted server. The encrypted data, which is known as “ciphertext”, are inaccessible without the correct key or password. The encryption of data will reduce the risks of essential data suffering the numerous painful consequences of a hack, breach, or leakage.
AES is the preferred encryption method because it excels in many key performance metrics. A few benefits of AES include:
Security: Even the lowest level of AES encryption, AES-128, would take an estimated 1 billion billion years (no, that extra billion isn’t a typo) to crack if using a brute force method.
Cost: AES encryption is available for free, as it was originally developed to be released on a royalty-free basis.
Ease of use: The AES algorithm is easy to implement across a multitude of applications and is known for its simplicity and ability to conform across hardware and software platforms.
To implement the proposed system we are using symmetric key algorithm such as AES encryption scheme where the same key is used for encryption and decryption.
Data recovery is the process of restoring lost, corrupted accidentally deleted or otherwise inaccessible data to its server, computer, mobile device or storage device. The cloud has benefited people by providing unlimited, secure storage without any high cost of investment in hardware. Cloud storage also aids in the security of your data backup, allowing you to access your files and documents even if your hardware is lost or damaged.

II. MOTIVATION

Motivated by privacy issues, advancement in mobile technology and wild success of cloud service model, we propose to build privacy into e-health care system. Since ehealthcare system have been gaining popularity, large amount of health data are involved for medical purpose and people start to realize that they would completely lose control over their personal data once it enters to the cyberspace. Some of the good reasons to keep health data private, the company person may refuse to hire someone with serious diseases and an insurance company may refuse to provide life insurance for the person with certain diseases. In order to overcome these privacy issues we propose to build privacy preserving health data storage and retrieval system.

III. RELATED WORK

[1] This paper introduces a review on the use of the AES Algorithm AES block cipher. The key size can be 128/192/256 bits. Encrypts data in 128-bit blocks each. That means it takes 128 bits as input and extends 128 bits of encrypted cipher text as output. AES relies on the network-licensed exchange policy which means it is performed using a series of linked functions that include retrieving and pushing input data. A set of AES commands are now integrated with the CPU (provides GB / s output) to improve the speed and security of applications that use AES encryption and decryption.
In [2] introduce the method to solve the challenging keyword search problem over the encryption data in the multi data owner architecture. Thus any data owner could security share the data with the user without the secure channel to transport trapdoor. We redefine the searchable asymmetric encryption while introducing the data owner’s key pair.
Ref. [3] [4] [5] we have reviewed is the survey paper which is comparative study of various secret sharing algorithms. In the area of cloud computing several cloud resources and services are being offered to the customers based on a pay-as-you-go model. Information outsourcing is a new way in which a third party organization offers storage services. This is very cost effective way for the customer as here is no need of buying expensive hardware and software systems for information storage. The customer is also relieved from maintenance activities of software version updates and management. Before information outsourcing can viable, the provider that is third party organization needs to ensure that the information is secure.
Ref. [6] In today's world diabetes has become one of the most life threatening and at the same time most common diseases not only in India but around the world. Diabetes is seen in all age groups these days and they are attributed to lifestyle, genetic, stress and age factor. Whatever be the reasons for diabetics, the outcome could be severe if left unnoticed. Currently various methods are being used to predict diabetes and diabetic inflicted diseases. In the proposed work, we have used the Machine Learning algorithms Support Vector Machine (SVM) & Random Forest (RF) that would help to identify the potential chances of getting affected by Diabetes Related Diseases. After pre-processing the data, features which influences the prediction are selected by implementing step forward and backward feature selection. The Principle Component Analysis (PCA) dimensionality reduction

method is analysed after the selection of specific features and the accuracy of the prediction is 83% implementing Random Forest (RF) which is significant in comparison with Support Vector Machine (SVM) with accuracy of 81.4%.
Ref. [7] discusses that in cloud computing, failing in data recovery and backup management leads to many problems for organizations and businesses. Most cloud service providers charge a premium for data backup on their premises, which consumers and small businesses may not be able to afford. The research presents a straightforward solution for this challenging task in the form of cloud-based disaster recovery and data backup methods. This technique lowers the solution cost, protects data from calamity, and allows switching from one cloud service provider to another more easily. This method relieves the consumer dependency on service providers and reduces data backup costs. Even though the cloud is an extraordinarily diverse and fast-supplied solution, it currently needs a backup solution. We recommend that data be backed up on the consumer premises. Consumers can get peace of mind at a low cost with cloud services. Without a backup plan, no solution is complete. For any business, business continuity and recovery are crucial. The lack of cloud backup capability must be addressed by business organizations so that there will be no negative impact on businesses later.

IV. OBJECTIVES

The main objective of our project is,

- To preserve the privacy to save those files in cloud and to recover if it gets lost.
- Using Weka tool to classify using Decision tree
- To implement the machine learning algorithm.
- To predict the diabetic disease effectively.
- To enhance the performance analysis.

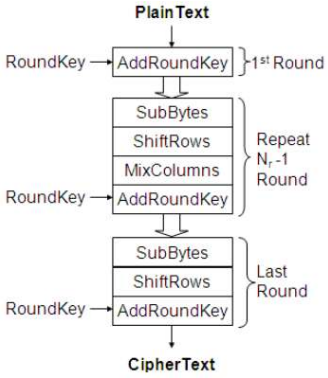
V. METHODOLOGY

The Advanced Encryption Standard (AES) is a symmetric key block cipher published by the National Institute of Standards and Technology (NIST) in December 2001. AES is used to encrypt and decrypt data block of 128 bits. It uses 128,192 or 256 bits depending on number of rounds. The code for the AES-128 bit version of this design is shown below.

```
state = M
AddRoundKey(state, &w[0])
for i = 1 step 1 to 9
    SubBytes(state)
    ShiftRows(state)
    MixColumns(state)
    AddRoundKey(state, &w[i*4])
end for
SubBytes(state)
ShiftRows(state)
AddRoundKey(state, &w[40])
```

Figure.1. Algorithm for AES encryption scheme

Figure 1 shows the AES encryption algorithm for 128 bit secret key. In the AES algorithm plain text is divided into blocks of 128 bits and intern block is divided into 16 bytes that is 4 by 4 matrix called state, on which operations are performed. Initially state is filled with input data and XOR operation is performed with encryption key. The various rounds are SubByte, ShiftRow, MixColumn, AddRoundKey.
Figure 2 shows data flow of encryption scheme. The main function of ShiftRow, SubByte, MixColumn are based byte oriented arithmetic operation and AddRoundKey is 128 bit wise XOR operation. SubByte is used at the encryption side to substitute a byte as two hexadecimal digits. ShiftRow is transformation operation operated at row level and MixColumn is transformation operation operated at column level. AddRoundKey is matrix addition and it proceeds one column at a time and adds the round key with each state column matrix.



Nr	Key size
10	128
12	192
14	256

Relationship between number of rounds and cipher key size

Figure 2. Flow Diagram for AES encryption scheme

The single block is 16 bytes, a 4x4 matrix holds the data in a single block, with each cell holding a single byte of information.

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

The matrix shown in the image above is known as a state array. Similarly, the key being used initially is expanded into (n+1) keys, with n being the number of rounds to be followed in the encryption process. So for a 128-bit key, the number of rounds is 16, with no. of keys to be generated being 10+1, which is a total of 11 keys. The mentioned steps are to be followed for every block sequentially. Upon successfully encrypting the individual blocks, it joins them together to form the final ciphertext. The steps are as follows:

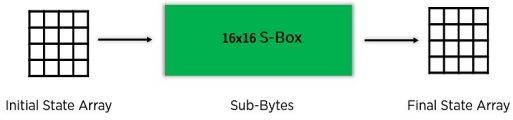
- Add Round Key: You pass the block data stored in the state array through an XOR function with the first key generated (K0). It passes the resultant state array on as input to the next step.

1	2	3	4
6	7	8	5
11	12	9	10
16	13	14	15

XOR

K_0	K_1	K_2	K_3
K_4	K_5	K_6	K_7
K_8	K_9	K_{10}	K_{11}
K_{12}	K_{13}	K_{14}	K_{15}

- Sub-Bytes: In this step, it converts each byte of the state array into hexadecimal, divided into two equal parts. These parts are the rows and columns, mapped with a substitution box (S-Box) to generate new values for the final state array.

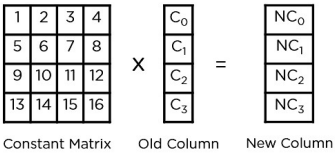


- Shift Rows: It swaps the row elements among each other. It skips the first row. It shifts the elements in the second row, one position to the left. It also shifts the elements from the third row two consecutive positions to the left, and it shifts the last row three positions to the left.

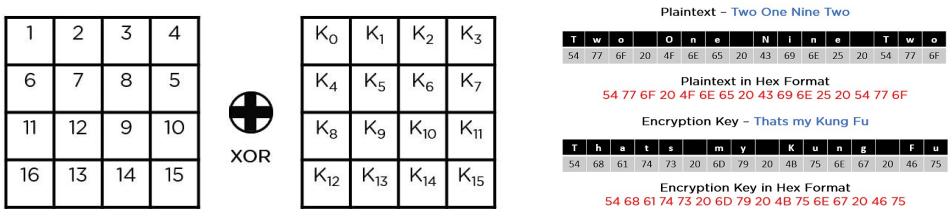
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1	2	3	4
6	7	8	5
11	12	9	10
16	13	14	15

- Mix Columns: It multiplies a constant matrix with each column in the state array to get a new column for the subsequent state array. Once all the columns are multiplied with the same constant matrix, you get your state array for the next step. This particular step is not to be done in the last round.



- Add Round Key: The respective key for the round is XOR'd with the state array is obtained in the previous step. If this is the last round, the resultant state array becomes the ciphertext for the specific block; else, it passes as the new state array input for the next round.



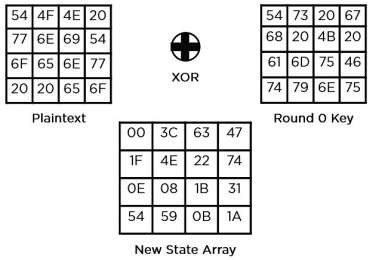
In the image above, the plaintext and encryption convert keys to hex format before the operations begin. Accordingly, you can generate the keys for the next ten rounds, as you can see below.

Keys generated for every round

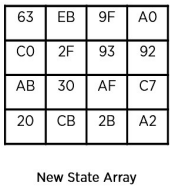
- Round 0: 54 68 61 74 73 20 6D 79 20 4B 75 6E 67 20 46 75
- Round 1: E2 32 FC F1 91 12 91 88 B1 59 E4 E6 D6 79 A2 93
- Round 2: 56 08 20 07 C7 1A B1 8F 76 43 55 69 A0 3A F7 FA
- Round 3: D2 60 0D E7 15 7A BC 68 63 39 E9 01 C3 03 1E FB
- Round 4: A1 12 02 C9 B4 68 BE A1 D7 51 57 A0 14 52 49 5B
- Round 5: B1 29 3B 33 05 41 85 92 D2 10 D2 32 C6 42 9B 69
- Round 6: BD 3D C2 B7 B8 7C 47 15 6A 6C 95 27 AC 2E 0E 4E
- Round 7: CC 96 ED 16 74 EA AA 03 1E 86 3F 24 B2 A8 31 6A
- Round 8: 8E 51 EF 21 FA BB 45 22 E4 3D 7A 06 56 95 4B 6C
- Round 9: BF E2 BF 90 45 59 FA B2 A1 64 80 B4 F7 F1 CB D8
- Round 10: 28 FD DE F8 6D A4 24 4A CC C0 A4 FE 3B 31 6F 26

You need to follow the same steps explained above, sequentially extracting the state array and passing it off as input to the next round. The steps are as follows:

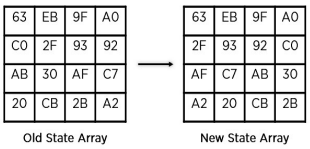
- Add Round Key



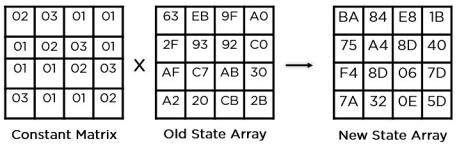
- Sub-Bytes: It passes the elements through a 16x16 S-Box to get a completely new state array.



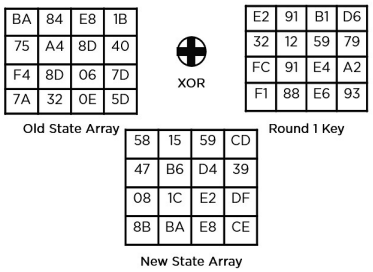
- Shift Rows



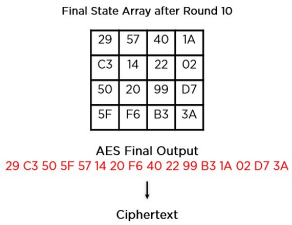
- Mix Columns



- Add Round Key



This state array is now the final ciphertext for this particular round. This becomes the input for the next round. Depending on the key length, you repeat the above steps until you complete round 10, after which you receive the final ciphertext.



The decryption process is the encryption process done in reverse in decryption the rounds can be easily undone as these stages have an opposite to it which when performed reverts the changes. Each 128 blocks goes through the 10,12 or 14 rounds depending on the key size.

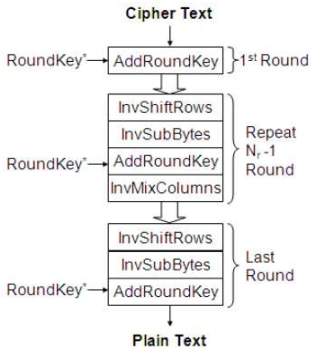


Figure 3. Flow Diagram for AES decryption scheme

Data recovery is the process of restoring lost, corrupted accidentally deleted or otherwise inaccessible data to its server, computer, mobile device or storage device. Cloud backup and recovery preserves copies of data in a secondary, offsite storage location. The cloud provider offers access to the storage and possibly additional managed backup and recovery services. The main purpose of a backup is to create a copy of data that can be recovered in the event of a failure in the primary data copy.

VI. RESULTS

Based on cloud data acquisition, Data will be encrypted for security purposes to minimize data misuse and it is done with the help of the AES algorithm to become the most efficient and accurate encryption algorithm.

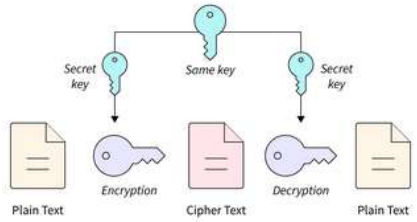


Figure 4. AES Encryption and Decryption process

This paper was successfully completed the implementation of encryption and decryption for AES algorithm. This implementation is use for build privacy preserving health data storage and retrieval system.
Figure 5. shows the Encryption simulation was successfully completed by use of key expansion and transformation of shift rows, sub-bytes, mix columns, add rows round keys.

Figure 5. Encryption Result

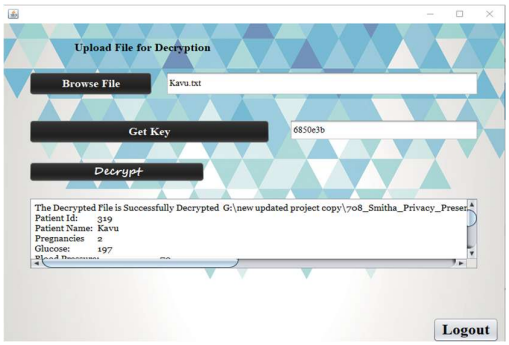


Figure 6. Decryption Result

Figure 6. show the Decryption simulation was successfully completed by the use of key expansion and transformations of inverse of shift rows, inverse sub-bytes, inverse mix columns, inverse add round keys.
Data recovery is the process of restoring data that has been lost, accidentally deleted, corrupted or made inaccessible.

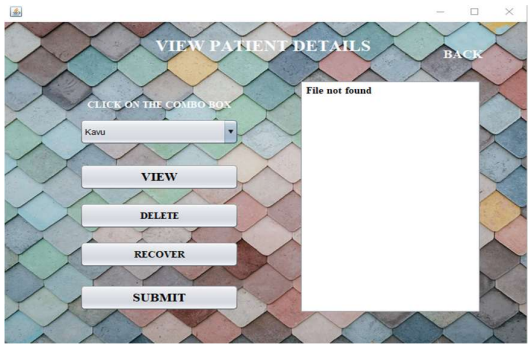


Figure 7. Data Loss Result

A backup is to create a copy of data that can be recovered in the event of a failure in the primary data copy.

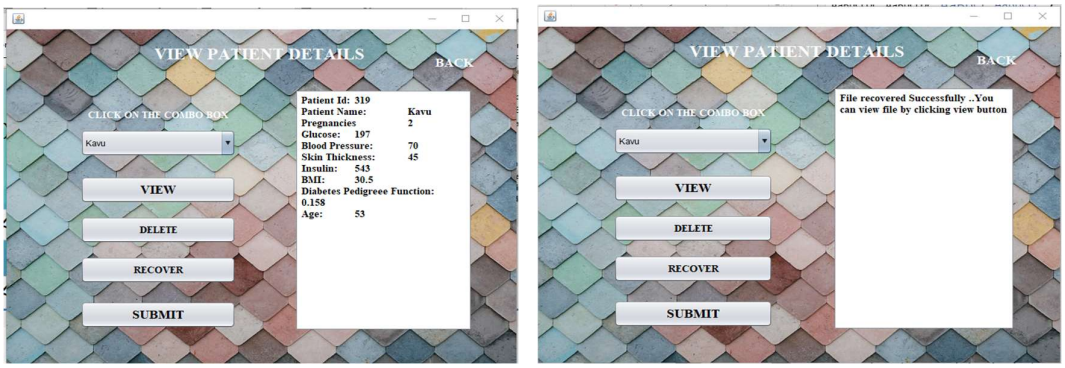


Figure 8. Data Recovery Result

VII. CONCLUSIONS

The proposed system is to build privacy in diabetic prediction system with the help of private cloud. We provided privacy preserving data storage by encrypting the data at client side before storing it on remote cloud server. For our implementation we have chosen AES algorithm for encryption /decryption purpose. Since it is symmetric key encryption were using same key for encryption/decryption. Also Proposed scheme helps to achieve secure and efficient dynamic operations on blocks of data as well as recovery of corrupted data or lost data can be achieved by replicating data on backup servers.

REFERENCES

[1] C. Wang, B. Zhang, K. Ren, J. Roveda, C. Chen, Z. Xu, "Healthcare monitoring system has helped to increase privacy and oppressive behavior," at INFOCOM'14, Toronto, Canada, 2014.

[2] "Jianyi Zhang, Chenggen Song, Zhiqiang Wang, Tao Yang and Wenming Ma"Efficient and Provable Security Searchable Asymmetric Encryption in the Cloud at IEEE 2018

[3] Z. Yang, C.Wang, K. Ren, and W. Lou N. Cao, "Privacy-preserving query over encrypted graph-structured data in cloud computing," in IEEE Int. Conf. Distrib. Comput. Syst, Jun. 2011,

[4] X. Zhu, C. Zhang, and Y. Fang J. Sun, "HCPP: Cryptography based secure EHR system for patient privacy and emergency healthcar," in IEEE Int. Conf. Distrib. Comput. Syst, Jun. 2011

[5] Chen RJ, Lin BS Tseng FK, "iPEKS: Fast and secure cloud data retrieval from the public-key encryption with keyword search," in InternationalConference on Trust. Security and Privacy in Computing and Communications, IEEE, 2013

[6] S Sivaranjani,S Ananya,J Aravinth "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction"2021

[7] V. Javaraiah, "Backup for cloud and disaster recovery for consumers and SMBs," 2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS), Bangalore, India,2011,

[8] S. Gokulakrishnan Dr. J. M. Gnanasekar "Data Integrity and Recovery Management in Cloud Systems" IEEE 2020

[9] A. Botta, W. De Donato, V. Persico, and A. Pescapé, ''Integration of cloud computing and Internet of Things: A survey,'' Future Gener. Comput. Syst., vol. 56, pp. 684–700, May 2016.

[10] Sandeep Sutar, G. A. Patil, "Privacy Management in Cloud by making use of Homomorphic Functions", International Journal of Computer Applications in the edition of January 2012.

[11] Sandeep G. Sutar, Dr. Manjunathswamy BE, Neha Surywanshi, "Secure Big Data Storage on Multi-Cloud Servers with Enhanced Attribute-Based Access Control", International Journal for Research in Applied Science and Engineering Technology, Volume 11 Issue VIII Aug 2023