A Tool for Protecting Electronic Data in Centralized Database using Improved Advance Encryption Standard (AES) and Secure Hash Algorithm (SHA)

Amrita Malviya
School of Information Technology
UTD, RGPV, Bhopal, M.P, India

Roopam Gupta, Ph.D
School of Information Technology
UIT, RGVP, Bhopal, M.P, India

Sanjeev Sharma, Ph.D
School of Information Technology
UTD, RGPV, Bhopal, M.P, India

ABSTRACT
In recent years, there is a wide amount of data available; which can be any one's personal data or any organization's important data. The main problem is to hide sensitive information, which include personal information, fact or even pattern which is generated by any data mining algorithm. Hence the desire to keep our own sensitive data secret in multiparty storage system is an important challenge of privacy preserving. The traditional privacy protection methods cannot do well, facing an urgent need for privacy protection in data mining, when they protect sensitive data. This paper propose a technique that protect electronic data in centralized database using Improved Advance Encryption Standard(AES) and Secure Hash Algorithm (SHA) This method is suitable for any type of data like- text, image and video.

Keywords
Protecting privacy: data mining: sensitive data.

1. INTRODUCTION
With the popularity & advancement of data storage capabilities of computer, variety of new data mining algorithm has been proposed. Too much information can be obtained from all the personal organization, government organization and social organization. The data received for storage is very huge amount data, our job is to store these data & also protect this data. First find the most sensitive data from these huge data and then secure it. Second find that the data received is in the original form or whether revised or cut from the original data. The purpose of doing this is to prevent individual data from others. The data received can be stored in two forms-centralized or distributed. Distributed data scenarios can be classified as horizontal data distribution and vertical data distribution. Data can store in both the forms and can provide security for individual data. Because the traditional data mining technology collects all individual's data together to process, it will cause the individual data abused or misused easily. There for use this new technology that protect individual data from others. In the traditional centralized database system they collect all the data and process for providing security, but in that way some data or may be most important data is don't secure properly or may be read by some other party. For example:-suppose someone buys a product online, there is so much data shown like-user id, password, card number, product number, product type, amount etc. From all these data find that the card number and password are most important data and have to provide more security to that data.

Some traditional data mining techniques that are used to provide security are-

TABLE 1

Technique	Description
Data Distribution	Distributed data consists of vertically partitioned data and horizontall partitioned data. Vertically partitioned data each database record attribute value in different sites. In Horizontally partitioned data different database records in different sites.
Data Distortion	Modify original database record before release, so as to achieve privacy protection purpose. This method includes-perturbation, blocking, sampling, aggregation, merging.
Data Mining Algorithm	Classification mining, association rule mining, clustering and Bayesian network etc.
Data or Rules Hidden	Hide original data or rules of original data.
Privacy Protection	(1) Modify data based on adaptive heuristics methods and only modify the selected value but not all value, which makes information loss of data is minimized. (2) Encryption technologies, such as secure multiparty computation. If each site knows only their input and input, but nothing about the others. (3) Data reconstruction method can reconstruct of original data distribution of random data.

Encryption is a powerful technique for protecting confidential data stored on untrusted server, such as a cloud computing. One limitation of encrypting confidential data is that the data must usually be decrypted for processing by any application-which required trusting the server running the application.

2. LITERATURE REVIEW

Xinjun Qi et.al [1] proposed a method with the advance of huge data storage capability of a computer all the data of personal or social organizations were stored in a computer. So, there is an urgent need to secure this data. The traditional privacy protection methods have not done well when protects this sensitive information. So there are some new improved methods to protect important information's. It consists two aspects, first the sensitive information such as-card number, password, address was not enclosed to any other person other than the owner of it, second the data we receive is no revised or cut from the original data. The essential purpose of privacy protection mining is to revise the original data in some way and store this data in a database. There are some data protection methods that can be used to secure the data such as- Data Distortion, Data Distribution, Data Mining algorithm, Data or Rule Hidden, Privacy Protection.

Archana Tomar, et.al.[2] described the method in which, the main problem with data mining is to hide sensitive information from others. To overcome this problem, the authors introduced a new algorithm called improve privacy preserving algorithm using association rule mining which based on random perturbation technique. The association rule mining has received a great deal of attention. Association rule mining is two a step process- finding all frequent item set; finding strong association rules from the item sets. The purpose of privacy preserving is to discover accurate pattern without access to the original data. The algorithm of association rule mining to minine the association rule based on minimum support and minimum confidence. The most useful method to hide the association rule is to reduce the support and confidence below the minimum support and minimum confidence. At the present privacy preserving association rule mining algorithm divided into three categories- Heuristic based, Reconstruction based, cryptography based. The whole privacy preserving algorithm has five phases-(i) check for authentication(ii) encode the data by using random perturbation technique(iii) decrypt the data to read the transformation on the basis of decryption key(iv) perform pruning(v) generate association rules, on the basis of decryption key.

Raluca Ada Popa, et.al.[3]described in Order preserving encryption a sort order of cipher text matches the sort order of the corresponding plain text. The ideal securities guarantee for order preserving encryption that the cipher text reveals no information about the plain texts beside the order. There are so many schema were proposed, these entire schema leaks more information than order. The technique is mutable cipher text means cipher text for the small number of plain text values will change. Mutable cipher text is used for ideal security. The model of order preserving encoding (OPE) consist an OPE client and OPE server that interact with each other. The client is the main party or the owner of data that is to be encrypted so that the client is a trusted party. The OPE server is an untrusted party; so it can be Passive server or Active server. The Passive server performe correctly and return the correct answer to client but it tries to learn more information beyond the order about the data. The Active server can misbehave in any way, such as provide the important data to the third party, or return the wrong answer to the client. This two threats leak only the order of the data.

Larry A. Dunning,et.al.[5]described that the popularity of the internet as a communication medium for personal or business use depends on its support for anonymous communication. An algorithm for anonymous sharing of personal information among N numbers of the parties is developed. In this technique we assign a unique ID number to each node ranging from 1 to N. This unique assignment to each node is anonymous in that the identity of one node is unique to the member of the other node. These assignments of serial number allow more complex data to be shared between two parties. This work deals with a very efficient algorithm for assigning identifiers (ID) to the node in a network in such a way that the ID's are anonymous using a distributed computation with no central authority. To differentiate anonymous ID assignment from anonymous communication, let's take a situation where N parties try to display their information collectively, but anonymously in N slots on the third party site. The ID's can be used to assign N slots to users, while the anonymous communication can allow the parties to hide their identities from the third party.

Geetha Mary A, et.al.[7] proposed that protecting privacy is a very major anxiety when dealing with the real datasets. Data perturbation is one of the useful PPDM techniques, which deal with the numerical data and heed on the statistical analysis of the data. Perturbation is two types, multiplicative perturbation and additive perturbation, when randomly generated data either multiplied or added to the data, which results in random modified data. In this paper they proposed a model in which perturbation is done by randomization, in which data is generated in intervals based on the level of privacy specified by each customer. This model applied classification algorithm in the perturbed data set and the accuracy is still remain same.

Privacy preserving data mining process is divided into three tiers, first is a Data Provider tier-where data collection take place, second Data Warehouse tier-in which data is converted into OLAP to easier processed data like- sum, aggregates, average etc., third tier or top tier is-Data Mining Server-where analysis is done according to the requirement. The main tier is Data Provider tier where the collection of data happens.

K. Srinivasa Rao, et.al.[8] present that the recent advance in communication, information, data mining and security technology have given rise to a new era of research, called as Privacy Preserving Data Mining. So many data mining, which incorporate privacy preserving mechanism, have been developed that allow one to extract relevant knowledge from the huge amount of data, and hide sensitive data or information from disclosure.

In Privacy Preserving Data Mining we usually take one of the three approaches; Data Hiding, Rule Hiding, and Secure Multiparty Computation.

Purushothama B.R. et.al[9] conclude that privacy preserving data publishing address the difficulty of publishing the data collected from the owner of the data by the holder of the data or publisher so that personal sensitive or private information of the individual is preserved and the published data is very useful. The anonymization technique like-suppression, bucketization, generalization, swapping, and randomization suffer from individual identity disclosures or the loss information which reduce the usefulness of the data. In this paper author present a novel privacy preserving data

publishing schema based on tuple duplication. They introduced the notion of semantically equivalent attribute values for sensitive attributes and the reputation loss by discloses to hide the sensitive information of an individual in the published data. The trapdoor attribute value for sensitive attribute are defined which helps in recovering the original dataset from the published dataset.

Table 2

Author	Techniques	Disadvantages
Geetha Mary A et.al.[7]	Random Perturbation Technique for data encryption	Data perturbation is one of the useful PPDM techniques, which deal with only numerical data.
Raluca Ada Popa et.al.[3]	Order-preserving encryption	The model of order preserving encoding (OPE) consist an OPE client and OPE server that interact with each other. The client is the main party or the owner of the data that is to be encrypted so that the client is a trusted party. The OPE server is an untrusted party. It can misbehave in any way, like provide the important data to the third party, or return the wrong answer to the client, and it learns more information beyond the order about the data.
Nirali Nanavati et.al. [4]	Global cyclic association rule	The type of data distribution for protecting the data are-horizontally partitioned data and vertically partitioned data. This technology is limited to only horizontal partitioned homogenous models.
K. Srinivasa Rao et.al. [8]	Privacy preserving data mining methods	The privacy preserving methods are-data modification, data perturbation, data hiding; however all are in fact related to the use of some type of technique to modifying the original data so that private data and knowledge remain private even after the mining process. In data hiding sensitive raw data like-name, address, identifier etc. were changed. In "Secure Multiparty Computation" distributed data are encoded. So there needs a single terminology that can be applied in place of data hiding, rule hiding, SMC etc.
urushothama B. R. et.al. [9]	Sanitization method for preserving privacy	Proximity Attack: if some sensitive values occur frequently within a range, then the attacker could still confidently infer the sum range in the group. The guessing probability would be 50%.

3. PROPOSED WORK

The proposed work defines a technique name protecting electronic data using Improved Advance Encryption Standard (AES) and Secure Hash Algorithm (SHA). The entire system architecture consists of five phases: (1) Authentication process in which database is created using user's information and also verify user from database (2) Received data from user (3) Encrypt the data using Modified AES algorithm and also find the hash value of the data using SHA algorithm (4) Store the data in database in the encrypted form (5) Decrypt the data using Improved AES algorithm and also find the hash value of the data using SHA algorithm.

The authentication procedure is shown in figure1. For providing authentication, first take the basic information from the user for database creation. After that user's information is store in the database in encrypted form using Improved AES algorithm and also find the hash value using SHA algorithm. When the user wants to store their data they first enter their username and password and the system verify it from the database. If it is an authorized user then the access is granted to store the data, and if not an authorized user then access is denied.

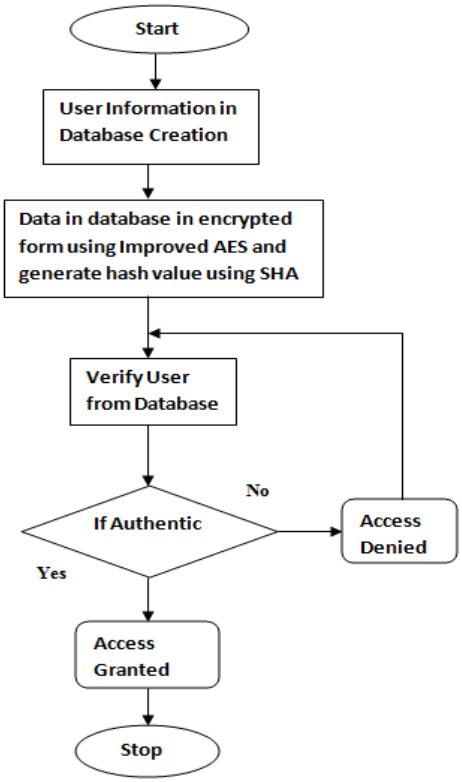

Fig.1AuthenticationProcess

Fig 2 shows the encryption process. In encryption process first fetched the data from user to be stored in database. In this paper for encryption improved AES algorithm is used and for each data value corresponding hash value is generated using SHA algorithm. Some modification has been used to enhance the performance of AES algorithm in terms of speed and security. First modification is decreasing the number of rounds to one while the second modification is replacing the S-box with new S-box. In this paper a single S-box is used for both encryption and decryption instead of two S-box used in initial AES algorithm where this lead to reduce the ROM used. Advance Encryption Standard (AES) is a famous and strong encryption algorithm and embedded with this Secure Hash algorithm (SHA) algorithm for hash value generation is a strong combination for data security. After using AES and SHA algorithm store the encrypted data with hash value in database.

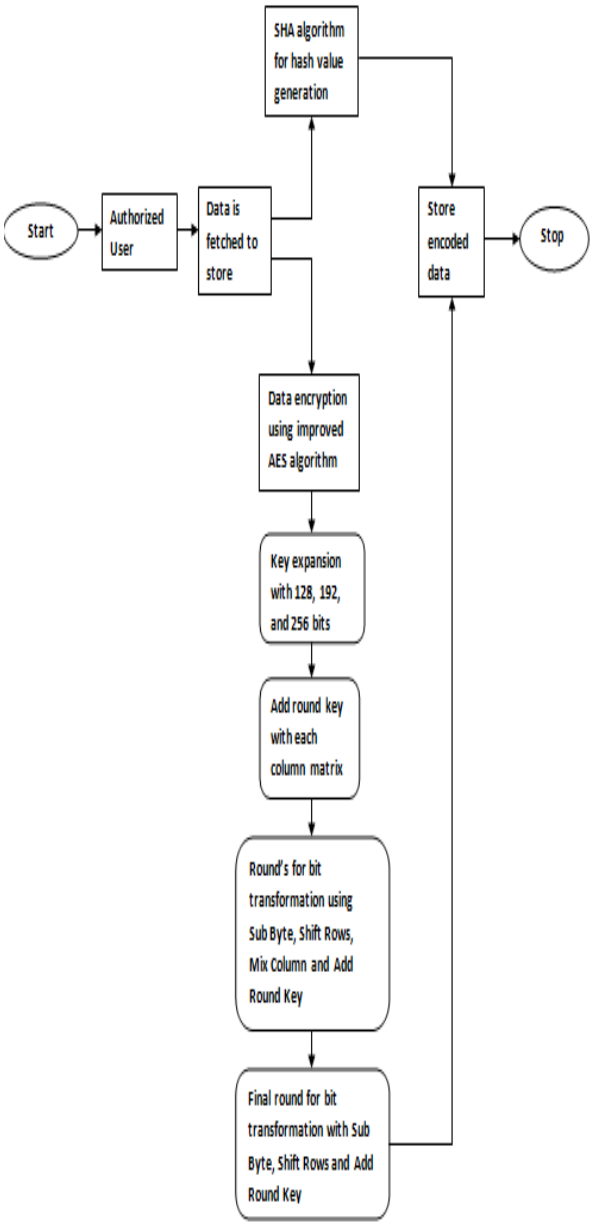

Fig.2 Encryption Process

Fig 3 shows the decryption process. First verify authentic user from database, after that fetched the encrypted data from database. For decryption process using improved AES algorithm, here we used same S-box for decryption that is used for encryption this leads to reduce the ROM used and the number of rounds is also one for decryption as in encryption process. Find the hash value for decrypted data using SHA algorithm and match this value with previously generated encrypted hash value. If both the values are same then the security of data is not compromised. This process provides the user's data back to original form with no integrity loss.

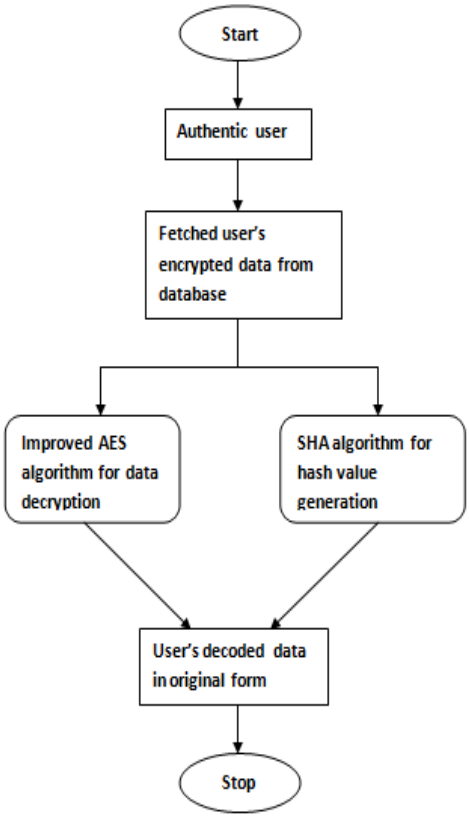

Fig 3 Decryption Process

Algorithm:

Input:

1. Source database (text, image)
2. Key for finding the authentication
3. Using improved AES encrypt the data
4. Find hash value using SHA

Output: Decrypted data

Begin

Step1: Provide authentication and verification

a. With the help of user information database created

b. Data store in encrypted form in the database

Step2: If user wants to store their data

a. Verify user from database and if authentic user access granted

b. If not an authentic user access denied

Step3: Data encryption and hash value generation

a. Encrypt user's data using improved AES algorithm

Number of rounds one and

a single S-box for both encryption and decryption

b. Find hash value of data using SHA algorithm

Step4: Encrypted data store in database

a. All the user's precious data store in database

in encrypted form and no one can access the data

Step5: Decrypted user data in original form

a. When user want their data back decrypt the data using improved AES algorithm

and also find the hash value of data using SHA algorithm

Finish

Processing Step's of Improved AES Algorithm

Step 1: Key Expansion:- To create round keys for each round, AES uses a key expansion process. If the number of rounds is Nr, the key expansion routine creates Nr+1, 128-bit round keys from one single 128-bit cipher key.

Step 2: Initial Round:-

(i) Add Round Key:- Add round key proceeds one column at a time. Add round key adds a round key word with each state column matrix.

Step 3: Rounds:-

(i) Sub Bytes:- Sub Bytes transformation is a byte substitution that operates independently on each byte of the state using a substitution table called S-box. S-box is derived from multiplicative inverse. The modified AES uses a single S-box for both encryption and decryption.

(ii) Shift Rows:- In shift rows transformation, the bytes in the last three rows of the state are cyclically shifted over different number of bytes. The first row is not shifted.

(iii) Mix Columns:- Mix column transformation operates at the column level; it transforms each column of the state to a new column.

(iv) Add Round Key:- Add round key proceeds one column at a time. Add round key adds a round key word with each state column matrix.

Step 4: Final Round:-

(i) Sub Bytes:- Sub Bytes transformation is a byte substitution that operates independently on each byte of the state using a substitution table called S-box. S-box is derived from multiplicative inverse. The modified AES uses a single S-box for both encryption and decryption.

(ii) Shift Rows:- In shift rows transformation, the bytes in the last three rows of the state are cyclically shifted over different number of bytes. The first row is not shifted.

(iii) Add Round Key:- Add round key proceeds one column at a time. Add round key adds a round key word with each state column matrix.

4. CONCLUSION

This paper focused on privacy protection involves data mining. Here explain a technique that can be applied to encrypt text, image and video data that is stored in centralizes manner and also maintain the privacy of individual parties. This paper used improved AES algorithm and SHA algorithm for data encryption and decryption. Advance Encryption Standard (AES) is a well known and strong encryption algorithm which has several advantages in data ciphering. However, AES suffer from some drawbacks such as high computation time for data processing. Some modifications have been proposed to enhance the performance of AES algorithm in terms of time ciphering. This paper used the improved AES algorithm with SHA algorithm for data encryption. Data became more secure when using AES algorithm with SHA algorithm for data encryption because after encryption hash value is also generated for finding that the data is not alter or changed by any one. Two modifications has been used in this paper, first is number of rounds reduced to one and a single S-box is used for both encryption and decryption, that enhance the processing time and reduced the ROM used.

By using improved AES and SHA technique for encryption and decryption it can improve computational efficiency and processing time. It provides a single data mining algorithm that can be applied to any type of data.

5. REFERENCES

[1] Xinjun Qi, Mingkui Zong, "An Overview of Privacy Preserving Data Mining", SciVerse ScienceDirect, Procdia Environmental Science 12(2012) 1341-1347, ICESE(2011).

[2] Archana Tomar, Vineet Richhariya, and Mahendra Ku. Mishra, "A Improve Privacy Preserving Algorithm using Association Rule Mining In Centralized Database", ISSN No.:2250-3536, IJATER(2012).

[3] Raluca Ada Popa, Frank H. Li and Nickolai Zeldovich, "An Ideal-Security Protocol for Order-Preserving Encoding", 2013 IEEE Symposium on Security and Privacy.

[4] Nirali Nanavati and Devesh Jinwala " Privacy Preservation for Global Cyclic Associations in Distributed Database", SciVerse ScienceDirect, procedia Technology 6(2012) 962-969, 2nd (ICCCS-2012).

[5] Larry A. Dunning and Ray Kresman "Privacy Preserving Data Sharing With Anonymous ID Assingnment", IEEE Transactions on Information Forensics and Security, 1556-6013, 2013.

[6] Mafruz Zaman Ashrafi, David Taniar and Kate Smith "ODAM: An Optimized Distribution Association rule Mining Algorithm", IEEE Distributed System Online 1541-4922, 2004.

[7] Geetha Mary A and N.Ch.S.N.Iyengar"Non-Additive Random Data Perturbation for Real World Data", SciVersa Sciencedirect, procedia Technology 4(2012)350-354.

[8] K.Srinivasa Rao and B. Srinivasa Rao"An Insight in to Privacy Preserving Data Mining Methods",SIJ Transactions on CSEA,No.3 july-august2013.

[9] Purushotham B.R. and B.B.Amberker"Duplication with Trapdoor Sensitive Attribute Values:A New Approach for Privacy Preserving Data Publishing", SciVersa Sciencedirect, procedia Technology 6(2012) 970-977.

[10] E.C. Laskari, E.C. Meletiou, D.K. Tasoulis, and M.N. Vrahatis"Privacy Preserving Electronic Data Gathering", Sciencedirect Mathematical and Computer Modelling 42(2005) 739-746.

[11] Da-Wei Wang, Churn-Jung Liau, Tsan-sheng Hsu and Jeremy K.-P. Chen"Value versus Damage of Information Release: A Data Privacy Perspective", Sciencedirect International Journal of Approximate Reasoning 43(2006) 179-201

[12] Salim M. Wadi and Nashruddin Zainal "Rapid Encryption Method Based on AES algorithmfor GGrey Scale HD Image Encryption", ScienceDirect11(2013)51-56