HANDLING PRIVACY PRESERVING OF DATA USING AES ALGORITHM

P.Suganthi,**P. Thiruselvan,Dr. C. Balasubramanian**

**PG Scholar, Department of CSE*

P.S.R.Rengasamy College of Engineering for Women

Sivakasi,Tamilnadu

***Asst Professor,Department ofCSE*

P.S.R.Rengasamy College of Engineering for Women

Sivakasi,Tamilnadu

****Prof and Head of CSE*

P.S.R.Rengasamy College of Engineering for women

Sivakasi,Tamilnadu

ABSTRACT

The sensitive data are patient record details, organization record which are handled only administrative member. Preservation of privacy is a major feature of data mining and individual's privacy without losing of confidential data. The objective of privacy preserving data mining is to mine significant information from large amounts of dataset. To prevent the misuse of sensitive information such as patient record by the system administrator. The database administrator collects patient's data which are maintained in secured manner. The sensitive information has patient's name, Address, Disease, Date of Visit. The privacy preservation mechanism can be carried out through anonymization techniques. In order to preserve the sensitive data, the system administrator modifying the patient details. Privacy preserving techniques used to handle the sensitive information from unsanctioned users. The proposal system is original patient record should be transformed into Anonymized table which is identified only by administrative. To keep anonymized table in cluster form using k-means clustering algorithm. To keep anonymized table in cluster form using k-means clustering algorithm. To partition anonymized table into cluster, and each cluster has similar data object which is based on anonymized data attribute. After that the grouped similar data to be protected using AES encryption algorithm. The aim of this paper is to provide more accuracy and better level of privacy of sensitive attribute..

Keywords-Privacy preserving, Anonymized Table, Clustering, AES encryption algorithm.

INTRODUCTION

Data mining deals with extraction of required information from large volume of data such as medical research, customer relationship management. The large volume of individual information are collected

and analyzed with the help of data mining. The objective of privacy preserving [3]is to modify the original data using algorithm, so the private data should be protected.

The unauthorized user are not access the private data when privacy techniques [5] is used. The privacy preservation for sensitive data can require the privacy policies. The Access control mechanism also important notation in public information system.

In particular, medical records in the form of text which have patient's detail such as name, address, age, zip code, diseases. The sensitive data are name, disease. Here privacy techniques are used to protect the sensitive data from unauthorized users. On the other hand, analyzing data by third party means a new threat of misusing of that privacy. As data contains sensitive information about patient and medical organization. The cryptographic algorithm [11] constructs to implement privacy preserving data mining algorithm. The cryptographic techniques are used where multiple parties are involved and to determine the non sensitive mining results.

RELATED WORK

The field of privacy has been advances in recent years because of the increasing the ability to sensitive data. The security of organization and governmental databases provide privacy of data.

G. Ghinita.[8] proposed an additive proposed system for constructing decision tree classifiers. The original data are randomized by adding noise data. These noise data are chosen independently by Gaussian distribution.

Lin [6] described data protection by perturbation method. The data matrix is vertically partitioned into several sub-matrix which are held by different authorized users. Each dataholder can choose a rotation matrix randomly. They showed the random rotation matrix.

A.Machanavajjhala [9] discussed k-Anonymity model for sharing sensitive information. This goal is achieved by using generalization and suppression without revealing identity of a person. This method used to protect the sensitive information.

G.Ghinita [4] works on access control systems which describe the key access control models. Access control provides limits on who can do what with objects on the computer. It ensures that all direct access to the objects are authorized manner.

Chaudhuri[5] provide a broad discussion of access control with privacy mechanism. The privacy requirement are represent in terms of k-anonymity. Anonymization is to protect individual privacy from attackers.

Helger Lipmaa [11] conclude that AES algorithm is faster than DES. When the transmission of data provide irrelevant difference in performance of different key symmetric key.

SYSTEM DESCRIPTION

Using K-Means Clustering algorithm for Handling Data Precision is a combination of access control and privacy protection mechanisms.

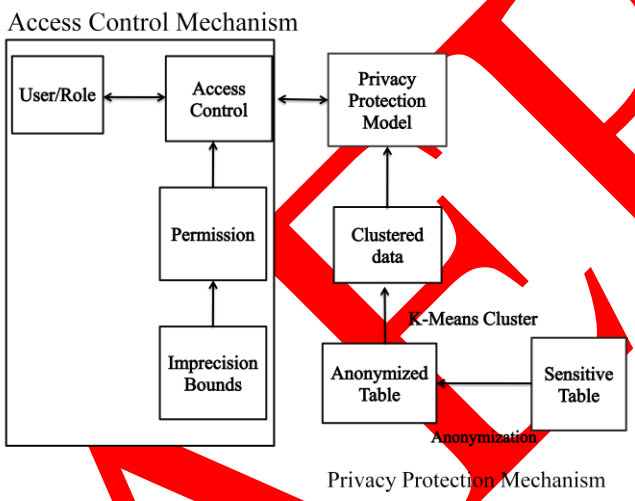

Fig 1:System Description

The access control specifies the authorized access of the system. The privacy protection mechanism ensure privacy of sensitive data accomplish by using AES algorithm. The access control mechanism define following things:

- Subject-User or Process
- Request for access
- Reference Monitor/Access control
- Reference Monitor Grants(or)Denied Access Request
- Object-Files or Data

The policy administrator defines the permission with imprecision bound for each role in the system. It also describe user-to-role assignments, and role-to-permission assignments. Here imprecision bound is the total imprecision acceptable for a query predicate and which is predetermined by the access control administrator.

Role Based Access Control protects the sensitive data with minimum value of imprecision bound values. The imprecision bound are not shared with the users. The privacy protection mechanism provide privacy requirement and it reduces the imprecision rate.

IMPLEMENTATION

This section describes the implementation of proposed work. The proposed system implemented with the following modules:

- Role Management
- Anonymization
- Clustering Analysis
- Privacy preserving Techniques

A)Role Management

User details and their access permissions are maintained in the role management process.

Admin

- Create patient record
- Create RBAC registration

The patient registration is created by admin. The patient record called as sensitive table which includes patient name, age, gender, disease, address, city, zip code, state, country. The sensitive table is maintained by administrative of medical association.

Role based access control defines the access rights depend on the role of the person.

Role-based security model is a tuples represent by

RBAC=(S,Q ,RL,P)

Where S->Set of users of a system

Q->Set of objects

RL->Set of role

P->Set of access rigits

Roles denote instead of person determine access rights. Subjects have different roles according to their rights

B) Anonymization

Anonymization techniques are used to the individual person's information. This approach describe remove the identify attribute from original data. Patient details are represented in single line such as tuples which stored in the original table. If the medical database anonymous and it is unfeasible to discover patient record. The attribute are classified as follows:

Quasi-Identifier (QI):

It contains attributes, example: zipcode, birth of date, gender, which are identify an individual based on other information.

Sensitive attribute:

It contains attribute which are related to unique individual person information. example, disease, salary.

There are two anonymization methods are used.

1)Suppression-based anonymization

In suppression based anonymization, we mask value with special value by *.The Quasi identifier are masked with special value when suppression method is used. After that third party could not predict the patient details.

2) Generalization-based anonoymization

In Generalization method, a set of value mapped with more general value.

Generalization and Suppression Algorithm:

Input: Medical Dataset PT[A1,A2,….An] and having attribute such as name, age, zipcode are sensitive attribute.

Output: Anonymized table g(T).

Steps:

1.Load Patient table.

2.Remove the identifier from the table.

3.Create Anonymize table.

4.Copy contents of Patient table to

Anonymized Table.

Suppression

5.Assign id=Patient ID

check id valid or not

if id valid{

convert id into '**'

}

6.Assign Name value to '******'

else if ID already exist ,choose another ID

generalization

7.Assign k=age

8.if k>Min and k <=Max{

age=Min-Max }

9.zip=zipcode

convert zip into substring(0,2)

end.

Table I-Original dataset

ID	NAME	AGE	DISEASE	ZIP
1	Anitha	26	Flu	626123
2	Ashik	30	Flu	626125
3	Kala	40	Fever	565435
4	Shalini	29	Asthma	652489
5	Mani	75	Brain tumour	974678

Table II-Anonymized Table

ID	**NAME**	**AGE**	**DISEASE**	**ZIP**
1	*****	20-30	Flu	626***
2	*****	30-40	Flu	626***
3	*****	40-50	Fever	565***
4	*****	20-30	Asthma	652**
5	*****	70-80	Brain tumour	974**

Table I represents original information of the patient which are stored in the form of tuples.

Table II shows that, after applying suppression and anonymization based method, the original datasets are anonymized and it generate the anonymized table. The quasi-identifier are changed into special value '*'.The result of generalization shows replace individual attribute with a border category displayed.

C) Clustering Analysis

The process of partitioning a set of objects in the dataset into subsets (cluster) called as Clustering. Each subset is a cluster, such that attribute in a clusters are similar to one another. K-means clustering algorithm for finding accuracy of anonymized table. Clustering analysis has broad application such as machine learning, spatial database technology, web search, customer segmentation.

In K-means clustering algorithm, the term K represents number of clusters. This algorithm partitioning the anonymized dataset into subsets called as clusters. The centroid of a cluster is its center point. The mean value of its cluster is calculated based on number of similar objects within

the cluster. Using Weka(Waikato Environment for Knowledge Analysis) tool for finding grouping of similar attribute in the dataset.

K-Means Clustering Algorithm

1.Input:

- k,the number of clusters,
- D:Anonymized dataset containing n attribute.

2.Selecting the first centers $x_{1,\ldots}x_k$

3.Repeat

For each attribute a_i

a.Calculate distance between a_i and different center.

b.Find minimal distance.

c.Assign the attribute a_i to the nearest center.

End for

For each nearest cluster

a.Calculate the new center:$y_1,\ldots y_k$

b.$x_1=y_1,\ldots.,x_k=y_k$

End for

4.Until no change between old and new centers

5.Ouput:A set of K clusters.

The confusion matrix is a useful tool for analyzing the classifier can recognize tuples of different classes. There are four terms used in computing evaluation measures.

True positives(TP):The positive tuples that were correctly labeled by the classifier.

True negatives(TN):The negative tuples that were correctly labeled by the classifier.

False positives(FP): The negative tuples that were incorrectly labeled as positive.

False negatives(FN): The positive tuples that were mislabeled as negative.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \qquad (1)$$

TP-Instance belongs to a,classified as a

FP-Instance belongs to others,classified as a

FN-Instance belong to a,classified to other

TN-Instance belongs to others,classified as others.

$$TN = \text{Total instance} - (TP + FP + FN) \qquad (2)$$

Fig 2.Clustering of Anonymized Dataset

D) Privacy Preserving Techniques

The term privacy which means the protected Information should be confidential or private should not be distributed or publicly known. Cryptograph techniques are used in this proposed system. The cryptographic algorithm constructs to implement the privacy preserving. It only shares the non sensitive medical data using AES algorithm.

The advantage of cryptography technique is transformed data are more precise and which also protected.

AES algorithm:

Step1:

Get the Anonymized Table which have non sensitive data.

Step2:

The activation code is created for each entity in the anonymized table. The key is only generated by the system/database administrator.

Step3:

RBAC request to admin when view the anonymized table.

Step4:

Admin accept or reject the request. Encrypt the data using shared secret key by AES algorithm.

Step5:

Compare the key, if the key match means, the anonymized table are displayed to RBAC.

Otherwise

Reject the Request of RBAC.

RESULT AND ANALYSIS

This section describes the result and analysis of proposed method work.

Table III-Time Taken To Form Cluster

Algorithm	L-diversity Dataset(sec)	Anonymized Dataset(sec)
K-Means	0.04	0.02
EM	0.34	0.31
Cobweb	0.05	0.03
Hierarchical	0.03	0.02
Optics	0.2	0.14

Table III describes the time taken to form a cluster by using different clustering algorithm. It shows that the comparison of time taken to form a cluster using two different dataset such as l-diversity and anonymized table.

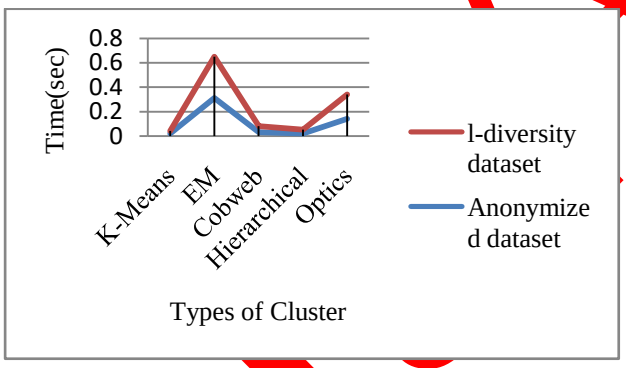

Fig 3-Graphical Representation of Time Taken To Form Cluster

Table IV shows that the comparison of accuracy of anonymized dataset and l-diversity dataset for sensitive and nonsensitive attribute. Figure 4 shows that better accuracy of anonymized dataset compare than other dataset.

Table IV-Accuracy Measure

Attribute	Anonymized dataset	l-diversity dataset
Gender	88%	88%
Age	68%	84%
Disease	68%	62%

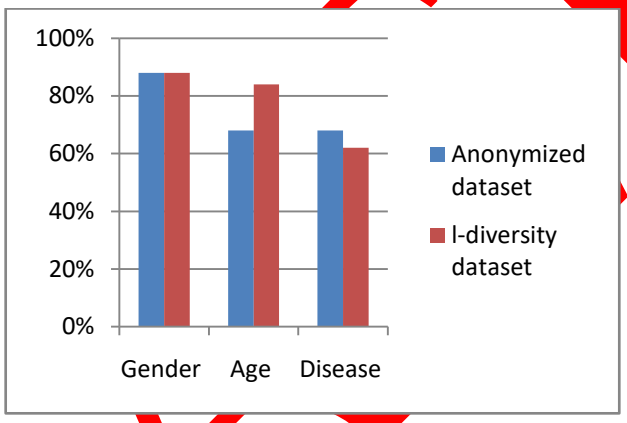

Fig 4-Graphical representation of accuracy measure

In Figure 5 represent the privacy guarantee values. X-axis represents dataset types and Y-axis represents privacy guarantee values. Anonymized dataset has better privacy values compare to other tecchniques.

Table V- Privacy Measure

Dataset Types	Privacy level
L-diversity	50%
Anonymized	66.67%

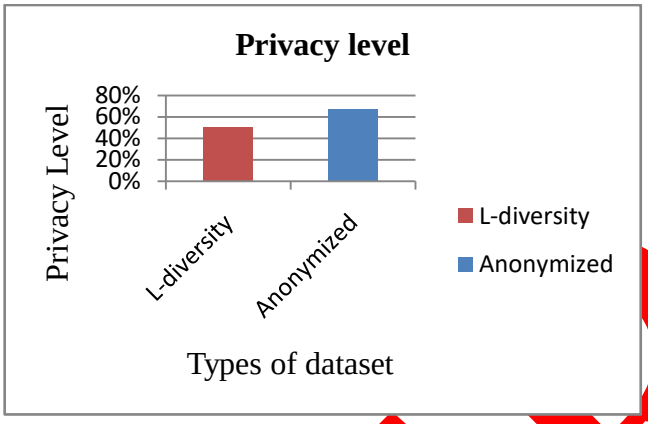

Figure 5-Comparison of privacy of different dataset

CONCLUSION

The privacy preserving algorithm provides protecting sensitive data item which are in the form of anonymized table. The privacy preserving techniques anonymizes the data to acquire requirements. In this, we use cryptography based anonymization which ensure that the resulting details should be anonymous.

REFERENCES

[1] Zahid Pervaiz, Walid G. Aref, "Accuracy Constrained Privacy-Preserving Access Control Mechanism for RelationalData",vol26,idno:10.1109/TKDE.2013.71,2014.

[2] N. Li, W. Qardaji, and D. Su,"Provably Private Data Anonymization:Or, k-Anonymity Meets Differential Privacy,"Arxiv preprintar Xiv:1101.2604,2011.

[3] X. Xiao, G. Bender, M. Hay, andJ. Gehrke, "Ireduct: Differential Privacy with Reduced Relative Errors," Proc. ACM SIGMOD Int'lConf Management of Data, 2011.

[4] G.Ghinita, P. Karras, P. Kalnis, and N.Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems,vol.34, no. 2, article 9,2009.

[5] S.Chaudhuri, R.Kaushik,"Database Access Control Privacy:Is There a Common Ground",Proc.Fifth Biennial Conf.Innovative Data Systems Research,pp.96-103,2011.

[6] Lin J.L., Wei M. C., "An efficient clustering method for k-anonymization," In: proceedings of the International Workshop on Privacy and Anonymity in Information Society, 2008.

[7] B.Fung, K. Wang, R. Chen, and P. Yu, "Privacy Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, no.4, article 14, 2010.

[8] G. Ghinita, P. Karras and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, no. 2, article 9, 2009.

[9] A.Machanavajjhala, D. Kifer, J. Gehrke, and M.Venkitasubramaniam, "l-diversity: Privacy beyond k- anonymity," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 1, no. 1, pp. 3–es, 2007.

[10] N.Mohammed and C. Lee, "Centralized and Distributed Anonymization for High-Dimensional Healthcare Data," ACM Trans. Knowledge Discovery from Data,Oct. 2010.

[11] Helger Lipmaa," Cryptographic Techniques in Privacy- Preserving Data Mining", University College London, Estonian Tutorial 2007.

[12] Helger Lipmaa," Cryptographic Techniques in Privacy- Preserving Data Mining", University College London, Estonian Tutorial 2007.