International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Privacy Preserving Data Mining With Multi Keyword Ranked Search for Medical Data

P.Subhashree, G.Gunasekaran

Abstract: ***Privacy Preserving Data Mining (PPDM) maintains the privacy of data stored in cloud. This work aims to protect outsourced data in cloud, and also permit multi keyword search over the encrypted data in a secure way by NLP process without downloading and decrypting all files. Different methods for privacy preservation were analyzed and randomization for multilevel trust is proposed along with an efficient method for keyword search in cloud.***

Keywords : Cloud, Privacy preservation, multi keyword search, NLP.

I. INTRODUCTION

Data mining involves knowledge discovery from databases. Many data mining techniques are available for privacy preservation and still research is on progress. Different state of art methods has their own advantages and disadvantages. In single-level trust, even though one perturbed copy is available, data miner can create accurate reconstruction than permitted by data owner (diversity attack). In multi-level trust, different amount of noise is added to data that allows data miners present at higher trust levels to get higher accurate data.Multi Level Trust is an important concept in PPDM where different types of diversified attacks are prevented. The main challenge is to prevent the data miners from merging copies at different trust levels, which can be dealt with by correlating noise across copies at different trust levels, where data owners are allowed to create perturbed data at different trust levels.

II. LITERATURE SURVEY

With many state of art privacy preservation algorithms already available, research is still going on to obtain information in many application areas. Information retrieval is the process of finding a relevant document from a huge sets of documents based on user query. The various techniques for PPDM is provided in Table 1.When information from sensitive medical data sets has to be retrieved, only specific level of information has to be disclosed considering anonymity. In[1] various techniques of PPDM is analyzed as shown in Table 1.

In [2], data Perturbation for PPDM is analyzed and in [3] various attacks in PPDM is discussed. In [4,5], fuzzy based PPDM is proposed and in [6,7], PPDM is proposed. In [8,9], multilevel trust party is proposed and in [10,11] review on PPDM algorithm is done with focus on scalability for future use.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
P. Subhashree*, M.E, Anna University of Technology, Trichy
G.Gunasekaran, Principal J.N.N Institute of Engineering Chennai, India.

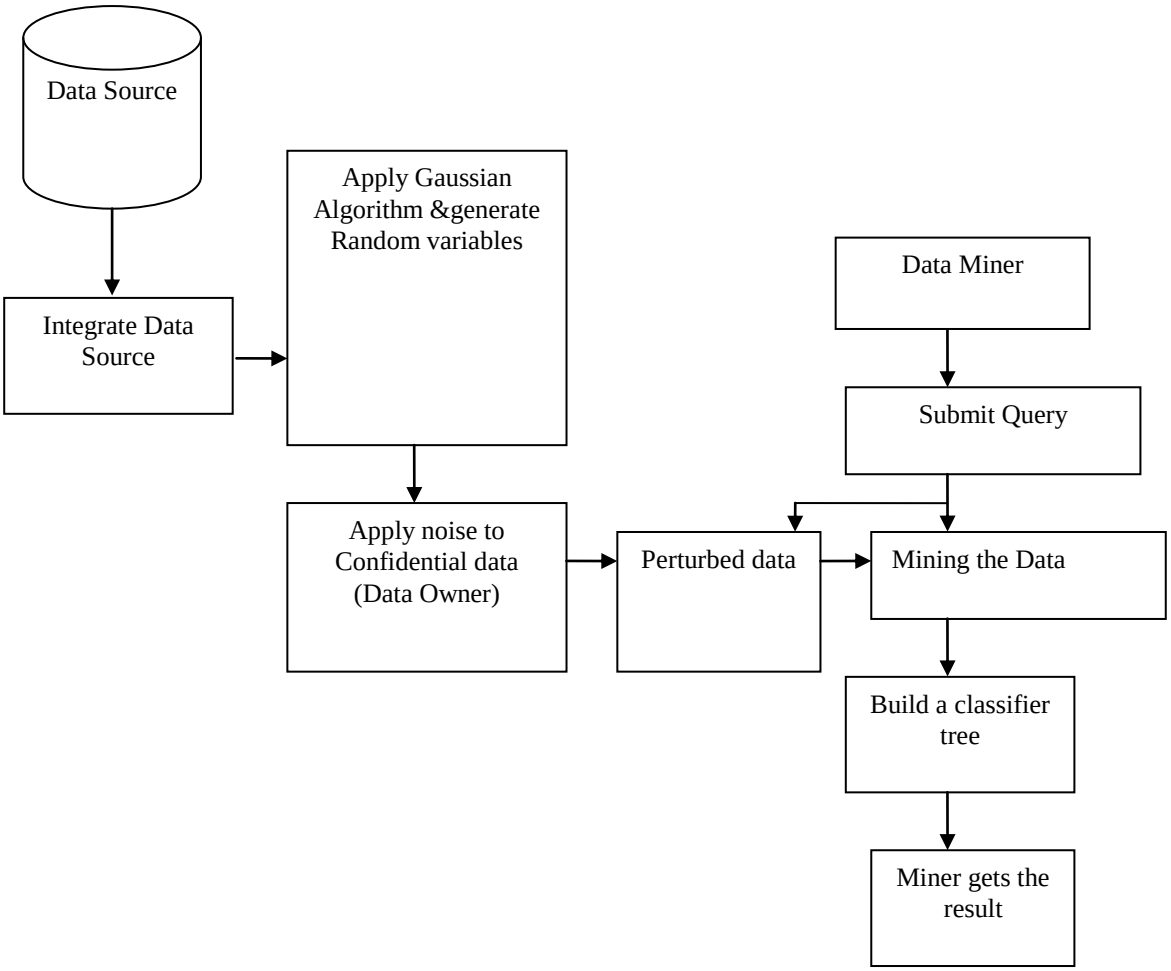
Table1 : Techniques of PPDM

Techniques of PPDM	Merits	Demerits
ANONYMIZATION	This method is used to protect respondents' identities while releasing truthful information. While *k*-anonymity protects against identity disclosure, it does not provide sufficient protection against attribute disclosure.	There are two attacks: the homogeneity attack and the background knowledge attack. Because the limitations of the *k*-anonymity model stem from the two assumptions. First, it may be very hard for the owner of a database to determine which of the attributes are or are not available in external tables. The second limitation is that the *k*-anonymity model assumes a certain method of attack, while in real scenarios there is no reason why the attacker should not try other methods.
PERTURBATION	Independent treatment of the different attributes by the perturbation approach.	The method does not reconstruct the original data values, but only distribution, new algorithms have been developed which uses these reconstructed distributions to carry out mining of the data available.
RANDOMIZED RESPONSE	The randomization method is a simple technique which can be easily implemented at data collection time. It has been shown to be a useful technique for hiding individual data in privacy preserving data mining. The randomization method is more efficient. However, it results in high information loss.	Randomized Response technique is not for multiple attribute databases.
CONDENSATION	This approach works with pseudo-data rather than with modifications of original data, this helps in better preservation of privacy than techniques which simply use modifications of the original data.	The use of pseudo-data no longer necessitates the redesign of data mining algorithms, since they have the same format as the original data.
CRYPTOGRAPHIC	Cryptography offers a well-defined model for privacy, which includes methodologies for proving and quantifying it. There exists a vast toolset of cryptographic algorithms and constructs to implement privacy-preserving data mining algorithms.	This approach is especially difficult to scale when more than a few parties are involved. Also, it does not address the question of whether the disclosure of the final data mining result may breach the privacy of individual records.

Multiple data source are integrated which is applied to Gaussian algorithm to generate random variables. The data owner applies noise to confidential data generating the perturbed data. When a data miner submits a query, mining the data at different trust levels is done from which a classifier tree is built and mining is done as shown in Figure 1.

Retrieval Number: B7375129219/2020©BEIESP
DOI: 10.35940/ijitee.B7375.019320
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)



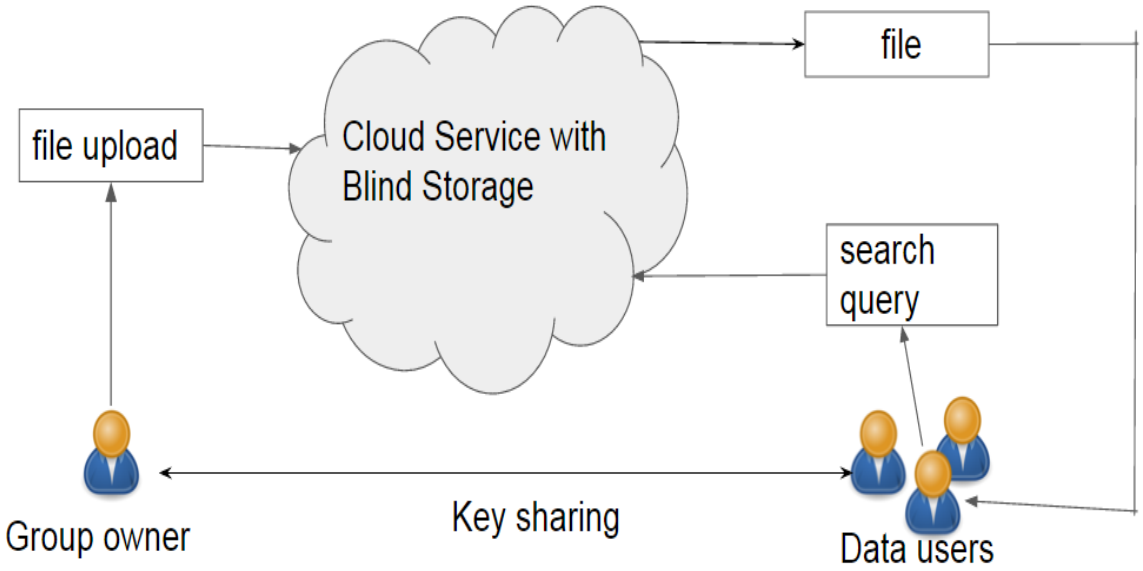


Figure 2 Keyword search

In order to search in cloud, important requirement is searching the encrypted data using single/multi keywords and providing the relevant result to the users.

Encryption, when done in cloud server makes it difficult as Multi keyword search is impossible on encrypted cloud data. Here downloading is done on encrypted file after which decryption is done and searched.

- The first step is creating a group, having group owners and group users.
- The second step is about text mining process done on the documents.
- The third step is blind storage where data is stored on the cloud
- The fourth step is query search to obtain the relevant file.

Group creation

Data owner should form a group and users will register with owner to add. When data owner accepts, user can retrieve the required documents rom the data owner.

Text mining process

Before uploading the files in cloud, the keywords are extracted using NLP technique which is later used for searching.

Blind Storage

The files are encrypted in gateway and stored in cloud after being split into chunks of equal blocks. These chunks are randomly stored and it becomes blind storage as encrypted content is only visble.

Query search

Data user search with keywords and cloud server map it with stored index and search. Filename to user is provided by cloud server and for viewing it the user should get the keys from data owner after which data access is done.

Multiple group creation, each group is having owner and multiple users

- We can give access control to each file for separate user.
- To encrypt data using Asymmetric algorithm (RSA) and key re-encryption.
- Using NLP technique and word net tool for text mining process.
- Index file generation on cloud.

III. EXPERIMENTAL RESULTS

The implementation of the proposed keyword search contains the following steps

1.Registering as a group owner or user
2. Register the group owner
3. Register the group user
4.Login for owner
5. Add user to group
6.User successfully added to group
7.Group owner uploading file as public
8.Keywords from file are listed after NLP process
9. Uploading file as private and giving read and write access to specific user
10.Keywords retrieved and NLP process
11. User searching for the file
12.Results of file search
13. Reading specific file

Figure 3 Keyword search implementation

Database Queries implementation with add, delete and append are given below

```
Connection connection=null;
ServletContext context;
String sql="";
dbconnection dbc=new dbconnection();

public Connection getConnection()
{

     connection=dbc.getConnection();
     return connection;
}
public String keytable()
{
     sql="insert into keystable(Username,publickey,privkey)
values(?,?,?)";
     return sql;
}
public String checkkeys(String uname)
{
     System.out.println("cjjjjjjjjjjjj"+uname);
     sql="select publickey,privkey from  keystable where
Username='"+uname+"'";
     return sql;
}
public String getlist(String uname)
{
     sql="select groupname from register where
Username='"+uname+"'";
     return sql;
}
public String getusergnam(String uname)
{
     sql="select gpnam from userregister where
username='"+uname+"'";
     return sql;
}
public String getusername(String gpname)
{
```

```
        {
              sql="insert into usercontrol
(username,filename,accesscontrol) values(?,?,?)";
              return sql;
        }
        public String requestacceptdetails(String uname)
        {
              sql="select ownername from acceptrequest where
username='"+uname+"'";
              return sql;
        }
        public String ratecheck()
        {
              //sql="select ownername from acceptrequest where
username='"+uname+"'";
              return sql;
        }
        public String getownname(String gpname)
        {
              sql="select ownername from acceptrequest where
groupname='"+gpname+"'";
              return sql;
        }
        public String getuserprivkey(String unam)
        {
              sql="select privkey from keystable where
Username='"+unam+"'";
              return sql;
        }
        public String userpubkey(String unam)
        {
              sql="select publickey from keystable where
Username='"+unam+"'";
              System.out.println("sqllllllllllll"+sql);
              return sql;
        }
        public String acceptcheckingdetails(String unam)
        {
              sql="select * from acceptrequest where username='"+unam+"'";
              System.out.println("sqllllllllllll"+sql);
```

Figure 4 depicts the time for search in cloud server

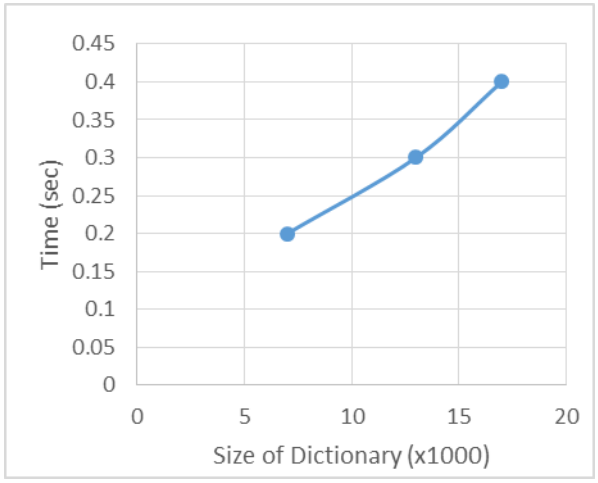


Figure 4 Time for search in cloud server

IV.CONCLUSION

In this paper , multi keyword ranked search scheme is proposed to enable accurate efficient search over encrypted cloud data. The search schemes allow multi-keyword query and provide result which is ranked according to relevance, which is also demonstrated in our proposed system. Effective data retrieval instead of undifferentiated data retrieval through ranked search is being performed. Ranked search helps in retrieving data quickly from huge documents that are stored in cloud. For future work we will be investigating on authentication and access control issues in searchable encryption technique.

REFERENCES

1. Gayatri Nayak et al "A survey on privacy preserving Data mining: Approaches and Techniques", International Journal of Engineering Science and Technology, Vol. 3 No. 3 March 2011
2. A. Hussien, N. Hamza and H. Hefny, 2013 , "Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing, Journal of Information Security", Vol. 4 No. 2, 2013, pp. 101-112. doi:10.4236/jis.2013..
3. Kasugai H, Kawano A, Honda, K, Notsu A. 2013, "A study on applicability of fuzzy k-member clustering to privacy preserving pattern recognition", IEEE International Conference on Fuzzy Systems (FUZZ), 2013, pp:1-6 .
4. Kokkinos Y, Margaritis K. 2013, "Distributed privacy preserving P2P data mining via probabilistic neural network committee machines", Fourth International Conference on Information, Intelligence, Systems and Applications (IISA), 2013, pp: 1-4.
5. Li Yaping Chen Minghua, Li Qiwei, Zhang, Wei. 2012, "Enabling Multilevel Trust in Privacy Preserving Data Mining, Knowledge and Data Engineering", IEEE Transactionson (Volume:24, Issue: 9), pp: 1598 – 1612.

ABOUT AUTHORS

Mrs.P.Subhashree has done M.E at Anna University of Technology, Trichy and B.Tech at JJCET Trichy. She had 6+ years of teaching experience. Pursuing Ph.D. at Sathyabama Institute of Science and Technology. Chennai. Attended two national conferences and one International Conference and published a SCOPUS journal.

Dr.G.Gunasekaran, Principal J.N.N Institute of Engineering Chennai, has obtained his B.E from National Engineering College and M.E from Jadavpur University and Ph.D. also from the same University and he has published 13 International Journals and has attended 7 International Conferences. He has got a profound experience of 27 years.