An Efficient Mechanism For Secure Data Mining In Distributed Databases Based On Rule Prediction.

[1] *Prof. Harshal N. Kolhe,Asst. Professor, Information Technology, PVG COE Nashik, Maharashtra, India*
[2] *Jyoti S. Patil, Information Technology, PVG COE Nashik, Maharashtra, India*
[3] *Kshitija D. Sonawane, Information Technology, PVG COE Nashik, Maharashtra, India*
[4] *Monali V. Shinkar, Information Technology, PVG COE Nashik, Maharashtra, India*
[5] *Rutuja A. Sonawane, Information Technology, PVG COE Nashik, Maharashtra, India*

ABSTRACT

Data mining is the technique of extracting the relevant data from large database Prediction rule mining is the popular mining technique that identifies interesting correlations between database attributes. In this paper we proposed a protocol for secure mining of prediction rules in distributed databases .Our protocol is based on Encrypted hierarchical index search(EHI) which is indexed based technique, metric preserving transformation(MPT) ,flexible distance-based hashing (FDH)which is more secured than FDM and Apriori algorithm. The main ingredients of our protocol is secure mining the appropriate data form large database without data leakage while mining the data. Secure multiparty computation all users can get their own data. Privacy preserving can be done by providing only authorized access to database. This protocol can reduced computation and communication cost and avoid data leakage while mining the database. This proposed architecture covers almost all the disadvantages occurred in the previous algorithm that has been implemented in previous system.

Keyword - *Data mining, Prediction Rules, Distributed Databases, Encrypted Hierarchical Index.*

1. INTRODUCTION

Data mining is technique to retrieve the relevant data from large database. It's searching is based on most consistent and relevant pattern and the and/or relationships that are systematic between the variables and then to validate by applying the detected patterns for findings the appropriate data is securely mined with less time computation.

We study here the problem of secure mining in distributed database. In distributed database hold homogenous database in several entities. Homogenous database means the database that share the same schema but hold information on different s entities. The goal is to find all prediction rule with support size's' & confidence level 'c', that hold in unified database. The information that we would like to protect is not only for individual transaction in different database, but also for global users transaction and apply various prediction rule to find the relevant data.

The main goal is to solve the problem of secure multiparty calculation.in such issue the many users that want to transaction from same database, then the data is securely mined without the appearance of trusted outsider. It is accustomed have privacy protective distributed data processing. In secure multiparty calculation was used for analysis the administrated for secure cooperative data processing. The thought of polynomials and privacy protective protocol were utilized in and severally. So that every client should gate their own data .It is a similar and administrated with less communication price. Several researchers experimented with 2 players for secure and distributed data processing as explored in. while mining in distributed database the privacy is maintain, so that the unauthorized access is not provided to users. For this fast searching algorithm are developed which required less time. And Encryption algorithm (AES) provide more security. We later posed in excess information only to small number (three) possible coaliation, unlike the other protocol that discloses information for single user only. We provide security so that excess information is not leaked while mining the data early protocol leak the information while mining.

Early the protocol reduced computational and communication cost but privacy is less because, it leak the some data was the issue so the protocol that we proposed, reduced the computation and communication cost and data leakage is avoided.as database is large and there are 'n' number of users so, union and intersection of database is done while mining the data.so that only relevant data is mined from large database and transfer the data to appropriate users.

2. LITERATURE SURVEY

1. Privacy Preserving Data Mining:

AUTHORS: R. Agrawal and R.Srikant. A useful way is provided for future data mining, research is done to be develop the techniques that incorporate privacy concerns. Since the primary task in data mining is the development of models that relate to same data .We are concentrating to case of building a decision-tree classifier from training data in which the values of individual records. The output of data records look very different from the original records and the distribution of data values is also very different from the original distribution. While it is not possible to accurately define original values in individual data records, we propose a novel reconstruction procedure to accurately define the distribution of original data values. By using these rebuild distributions. We are able to construct objects whose accuracy is compared to the accuracy of objects built with the original data.

2. Keying Hash Functions for Message Authentication:

AUTHORS: M. Bellare, R. Canetti, and H. Krawczyk. Cryptographic hash functions such like MD5 or SHA-1 for any data or message authentication is done. It is the main approach in many applications, particularly Internet security protocols. As they very easy to implement these techniques are usually based on ad-hoc techniques that lack a sound security analysis. We present new, simple, and practical building of message authentication schemes based on a cryptographic hash function. Our schemes, NMAC and HMAC, are proven to be secure as long as the underlying hash function has some reasonable cryptographic strengths. Moreover we show, in a best way, that the schemes maintain almost all the security of the underlying hash function. The performance of our schemes is

efficient that of the underlying hash function. Hash function is widely available library code or hardware can be used to develop them in a simple way, and irreplaceability of the hash function is easily supported.

3. Secret Sharing Homomorphism's:

Keeping Shares of a Secret AUTHORS: J.C. Benaloh In 1979,Blackley and Shamir independently proposed schemes by which a secret can be segmented into many shares which can be distributed to appropriate users. This property decreased the need for trust among agents and allows secret sharing to be applied to several new problems. One application described here gives procedure of verifiable secret sharing which is much simple and used by more client than previous schemes.

4. Privacy-Preserving Graph Algorithms in the Semi-Honest Model:

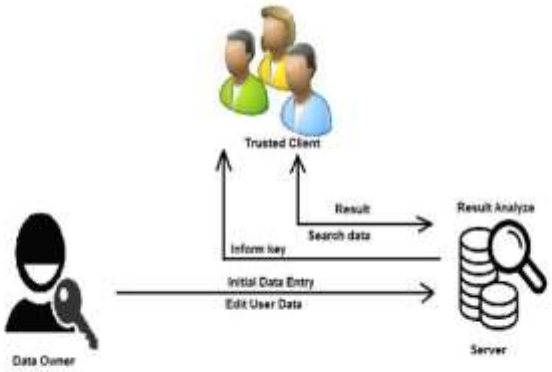
AUTHORS: J. Brickell and V. Shmatikov, This system consider scenarios in which two client, each in possession of a graph, wish to compute some algorithm on their joint graph in a privacy-preserving manner, that is, without leaking any information about their inputs except that information is provided by the algorithms output by using secure multiparty concept. We present new algorithms for privacy-preserving computation of APSD (all pairs shortest distance) and SSSD (single source shortest distance), as well as two new algorithms for privacy-preserving set union. Our algorithms are more efficient than previous generic construction. In previous work on privacy-preserving data mining, we prove that our algorithms are secure and the clients are honest, but curious.

3. PROPOSED SYSTEM

Here we implemented an alternative protocol for the secure computation of data form large database. The implemented protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol depend on encryption of data while storing the data in database. As well as data in only editable to data owner by which privacy is maintain of client data. It has simplicity because it can use any field so any non-technical client can use it efficiently.so which contributes towards much reduced communication and computational costs. Our protocol can be useful in place where existing system, the excess information may leak the excess information leaked .Because our protocol provides authorization of client only register client can retrieved there data which can improved the privacy of the system. The protocol that we implemented here computes a parameterized family of functions, which we call threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets the problem of secure multiparty computation is solve in this protocol.

4. IMPLEMENTATION

System mainly consists of three main modules:



System Architecture

1) Data owner:

Data owner has a higher authority for handling an database of any system.it has an authority to entry and updation of an particular user data.

2) Trusted client:

The client who has register to a particular system can only retrieve the data. The client can decrypt the data on the basis of key provided by the system. Any unauthorized client cannot access the other client data so any unauthorized client has to be register to access to the system.

3) Server:

The main role of server is to encrypt and decrypt the data that is store in database the encryption and decryption is done by AES (advanced encryption standard) algorithm. Every client data verification and validation is done by server. It maintains all the records of client.

4) Prediction rule:

Prediction rule is based upon conditions the condition such as ANDing & ORing by selecting the particular data item ,it can be of two or three possible combination, which is useful to retrieved the relevant data and it provide accessing data very fast.

Advantages of proposed system:

1. This system proposed a protocol for secure mining of Prediction rules in fully distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency.

2. The main ingredient in our proposed protocol is a novel secure multiparty protocol for computing the union (or intersection) of private subsets that each of the interacting players holds.

3. 3. It provides more security then the existing system

4. PROPOSED ALGORITHM

1) Encrypted Hierarchical Index Search:-

1: request the server for the (encrypted) root node Loot;

2: H:=new min-heap; pnn:=NULL;

3: ϒ:=mine2Lrootmaxdist(q, e); . derive NN distance bound

4: for each entry e ϵ Lroot such that mindist(q; e) < ϒ do

5: insert the entry (e, mindist(q, e)) into H;

6: while H is not empty and its top entry's key < ϒ do

7: pop next λ entries from H and insert them into a set S;

8: request the server for each (encrypted) child node of S;

9: for each retrieved node Lcur do

10: if Lcur is a leaf node then .check for closerobjects

11: update ϒ and pnn by using objects in Lcur;

12: else . expand the entries of Lcur

13: ϒ:=minf{ϒ; mine2Lcurmaxdist(q, e)};

14: for each e ϵ Lcur such that mindist(q; e) }<ϒ do

15: insert the entry (e; mindist(q; e)) into H;

16: return pnn as the result;

2) MPT building algorithm

1:use a heuristic to select a set of A anchor objects from P;

2: Integer B:=[|P|/A];

3: use a heuristic to assign each data object of P to an anchor object, subject to the capacity constraint B; 4: for i:=1 to A do

5: let ai be the i-th anchor object;

6: let ai:S be the set of objects assigned to the anchor ai;

7: ri:=maxpai:Sdist(ai, p); compute covering radius

8: for each object p ϵ ai:S do

9: send the tuple {p:id;OPE(dist(ai,p)) ECR(p; CK)} to the server;

3) Frequency Distance-Based Hashing:-

1: for i := 1 to A do . key generation

2: choose an object randomly from P as an anchor object ai;

3: find the distance value ri such that half of objects P ϵ P satisfy dist(ai; p) < ri;

4: for each object P ϵ P do

5: compute the encryption ECR(p;CK);

6: compute BM(p);

7: send the tuple (p:id;BM(p); ECR(p; CK)) to the server;

5. RESULT ANALYSIS

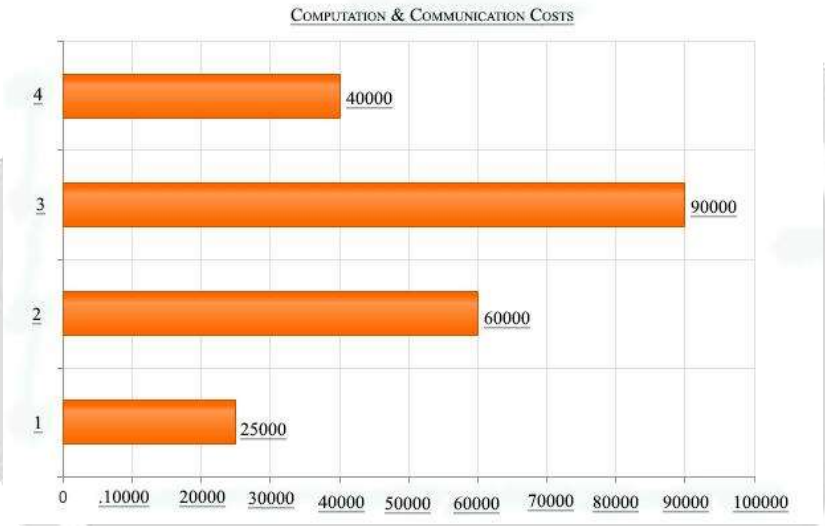


Fig - 1: Computation & Communication Costs

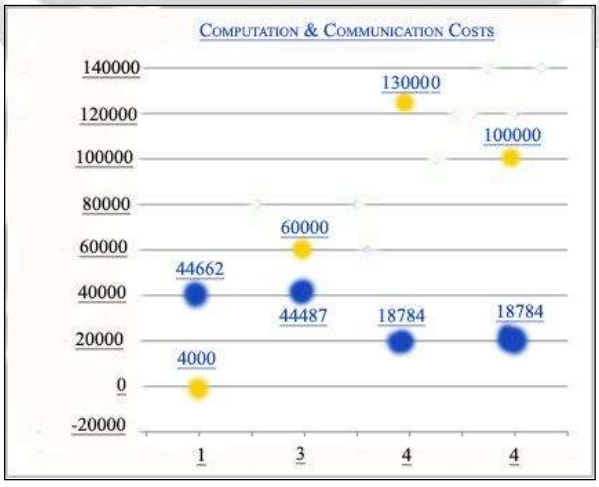


Fig - 2: Computation & Communication Costs

X - Axis: No of users

Y - Axis: Experiment set

Dataset: For implementation any organization dataset D is used. So that by using that prediction rules are generated as this is very fundamental part of data mining. System shows the expected results which provide the user data in which the user is interested and in minimum time. It takes the minimum computation time with this use.

Where,

M: No of Users

N: Experiment set

N= 25000,6000,9000,4000

M= 4

6. CONCLUSION

Mining the data prediction rules is one of the data mining techniques which are very useful for making well informed decisions. In this paper we study secure mining of prediction rules. Thus the statistical measures can be used to know how the rules are useful. We proposed a protocol for secure mining of prediction rules in distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency. In future we improve the prototype and make it a useful tool for mining the relevant data from large database that are handled by any organization, hospital, school, government sector and various other sector.

7. REFERENCES

[1] Santhana Joyce M PG Scholar, Department of IT Sathyabama University Chennai, Tamilnadu, IndiaPrivacy in Distributed Databases Based on Prediction Rules

[2] R. Agrawal and R. Srikant, "Fast algorithms for mining Prediction rules in large databases", in VLDB, page 487–499, 1994.

[3] R. Agrawal and R. Srikant, "Privacy-preserving data mining", in SIGMOD Conference,

[4] Kantarcioglu M, Clifton C, "Privacy-Preserving distributed mining of Prediction rules on ly partitioned data", in IEEE Transactions on Knowledge and Data Engineering Journal, IEEE Press, Vol 16 (9), pp.1026-1037, 2004.

[5] N V Muthu Lakshmi, Dr. K Sandhya Rani, "Privacy Preserving Prediction Rule Mining in ly Partitioned Databases Using Cryptography Techniques", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 3 (1), PP. 3176 – 3182, 2012.