Security of Private Data Extracted from Outsource Database in Association Rule Mining

Madhuri B. Bagdane
PG Student
Comp Engg, PVPIT
SavitribaiPhule Pune University

N. D. Kale
HOD
Comp Engg, PVPIT
SavitribaiPhule Pune University

ABSTRACT
Spurred by using trends inclusive of cloud computing, there has been enormous current hobby inside the paradigm of Data mining-as-a-carrier. A business enterprise (statistics proprietor) lacking in expertise or computational resources can outsource its mining desires to provider issuer (server). However, both the gadgets and the affiliation rules of the outsourced database are considered private property of the organization (facts proprietor). To guard corporate privacy, the data proprietor transforms its facts and ships it to the server, sends mining queries to the server, and recovers the real patterns from the extracted patterns obtained from the server. In this paper, look at the trouble of outsourcing the affiliation rule mining undertaking within a company privacy-keeping framework. Advise an attack model based totally on history expertise and devise a scheme for privacy maintaining outsourced mining. Another interesting heading is to unwind our presumptions about the attacker by permitting him to know the subtle elements of encryption calculations and/or the recurrence of thing sets and the dissemination of exchange lengths. Our present system expects that the attacker does not have such information. Our scheme ensures that every transformed item is indistinguishable with appreciate to the attacker's background expertise, from as a minimum k−1 other converted items. Our comprehensive experiments on a completely big and actual transaction database display that our techniques are powerful, scalable, and protect privacy.

Keyword
Association rule mining, privacy-preserving outsourcing.

1. INTRODUCTION
The initiation of cloud computing and its model for IT administration in light of the web and huge information focuses, the outsourcing of information and registering administrations is in advance a novel pertinence, which is relied upon to soar in the not so secluded future. Business approaching and information confession administrations, for example, progressed examination in view of information mining innovations, are required to be among the administrations agreeable to be externalized on the cloud, because of their information serious nature, and in addition the intricacy of information mining calculations. In this manner, the worldview of mining and management of records as administration will probably broaden as ubiquity of distributed computing develops [1]. that is the records mining-as-an management worldview, long past for empowering institutions with restricted computational property and/or facts mining capability to outsource their information mining necessities to an intruder management provider [2], [3].

Regardless of the fact that it's far valuable to accomplish cutting-edge exam on massive volumes of statistics in a savvy manner, there exist some true protection problems of the facts-mining as-a-carrier worldview. One of the precept security issues is that the server has entry to profitable records of the owner and can soak up delicate facts from it. As an example, via taking a gander on the exchanges, the server (or a gatecrasher who accesses the server) can comprehend which matters is dependably co purchased. Nonetheless, both the exchanges and the mined examples are the assets of the records proprietor and must stay safe from the server. This issue of making sure important non-public facts of associations/companies is alluded to as corporate safety [4]. Distinctive to man or woman security, which just considers the insurance of the individual statistics recorded approximately people, company protection calls for that each the singular matters and the examples of the collection of information things are regarded as company resources and accordingly must be secured

2. LITERATURE SURVEY

2.1 Defining privacy for Data mining
Protection saving Data mining getting legitimate data mining results without taking in the fundamental information values has been accepting consideration in the examination group and past. It is indistinct what security saving means. This paper gives a structure and measurements to examining the importance of security saving information mining, as an establishment for further research in this field.

Method : Secure multiparty computation

Disadvantage: It have low efficiency for providing privacy to data.

2.2 Privacy-Preserving in Outsourced Transaction Databases from Association Rules Mining
Authors proposed an attack version in light of past information and devise a plan for protection saving outsourced statistics mining. Authors plan guarantees that modified statistics is diverse as for the assailant's beyond records. The trial consequences beyond records. The trial consequences on proper alternate database exhibit that our processes are adaptable, efficient and comfortable protection.

Method: Rob Frugal encryption scheme

Disadvantage: It Provide onside privacy protection.

2.3 Maintaining Data Privacy in Association Rule Mining
Data mining administrations require precise data information for their outcomes to be significant, however protection concerns might in clients to provide spurious data. We examine here, concerning mining affiliation rules, whether clients can be urged to give right data by guaranteeing that the

mining process can't, with any sensible degree of sureness, abuse their protection. We present a plan, taking into account probabilistic distortion of client information that can all the while give a high level of security to the client furthermore, hold an abnormal state of precision in the mining results. The execution of the plan is approved against agent genuine and synthetic datasets.

Method: MASK algorithm(Mining Associations with Secrecy Constraints)

Disadvantage: This algorithm was not re-interrogating the distorted database.

2.4 Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data

Data mining can remove imperative learning from vast information accumulations – however now and again these accumulations are part among different gatherings. Security concerns might keep the gatherings from straightforwardly sharing the information, and a few sorts of data about the data. This paper addresses secure mining of affiliation rules over on a level plane divided data. The techniques consolidate cryptographic systems to minimize the data shared, while adding minimal overhead to the mining task.

Method : Secure association rule mining algorithm

Disadvantage: The algorithm given is for three or more parties—the difficulty with the two-party case

2.5 kTTP: A New Privacy Model for Large Scale Distributed Environments

Secure multiparty calculation permits gatherings to together compute an element of their private inputs without uncovering anything other than the yield. Hypothetical results give a general development of such conventions for any capacity. This paper is to define another protection model k-protection by method for an inventive, yet normal speculation of the acknowledged trusted outsider model. This permits executing cryptographically secure efficient primitives for certifiable expansive scale dispersed framework.

Method: K-Privacy algorithm

Disadvantage: The least size of a group for which our algorithm allows learning combined statistics

3. EXISTING SYSTEM

3.1 Existing Method

An attack model was generated based on following criteria such as based on assumption that the service provider (who can be an attacker) is semi honest in the sense that although he does not know the details of the encryption algorithm, he can be interested and thus can use his background knowledge to make inferences on the encrypted transactions. It has been assumed that the attacker always returns (encrypted) item sets together with their precise support. Rob Frugal algorithm helps to give the privacy for the database on server, who ships the data to client for association rule mining. By Rob Frugal algorithm, true support of mined patterns can be improved. Rob Frugal algorithm involves one to one substitution, k grouping methods and Fake transactions. Rob Frugal encryption transforms a Plain Transaction Database (TDB) into the encrypted Database $D*$. At the time of pattern mining the patterns are generated for given query with the high potential of imitation that probably degrades the accuracy of generated patterns.

Disadvantage of Existing Method

- Rob Frugal algorithm considers only cipher text attacks only
- If the attackers provide any background knowledge, it may break our encryption scheme.
- Server may send encrypted transaction database to third parties for getting plain text

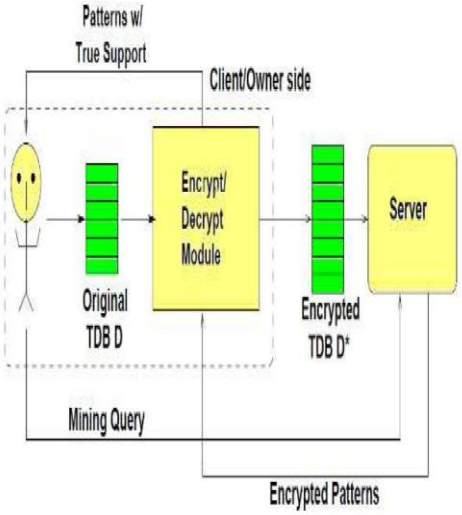

Fig1. Architecture of Existing system

The Fig. 1. Show architecture of mining as service paradigm. The client/owner encrypts its data using encrypt/decrypt (E/D) module, which can be basically treated as a black box from its perspective. It is responsible for transforming the input data into an encrypted database. The server conducts data mining and sends the (encrypted) patterns to the owner. Our encryption scheme has the property that the returned supports are not true supports. The E/D module recovers the true identity of the returned patterns as well their true supports. It is trivial to show that if the data are encrypted using 1–1 substitution ciphers (without using fake transactions), many ciphers and hence the transactions and patterns can be broken by the server with a high probability by launching the frequency-based attack. Thus, the major focus of this paper is to devise encryption schemes such that formal privacy guarantees can be proven against attacks conducted by the server using background knowledge, while keeping the resource requirements under control.

3.2 Proposed System

The attacker always returns (encrypted) item sets together with their exact support. The data owner considers the true identity of every cipher item, every cipher transaction and every cipher frequent pattern as the intellectual property which should be protected. Item based attack is having attack on all sets of a specific group. Set Based Attack is having Attacks on specific set alone. System enhance the existing work with ECDH (Elliptic Curve Diffie Hellman) key exchange algorithm was proposed after Rob frugal encryption scheme in order to provide privacy preserving outsourced mining. ECDH algorithm provides bidirectional encryption of client and server which protect against the forging the contents of the communication.

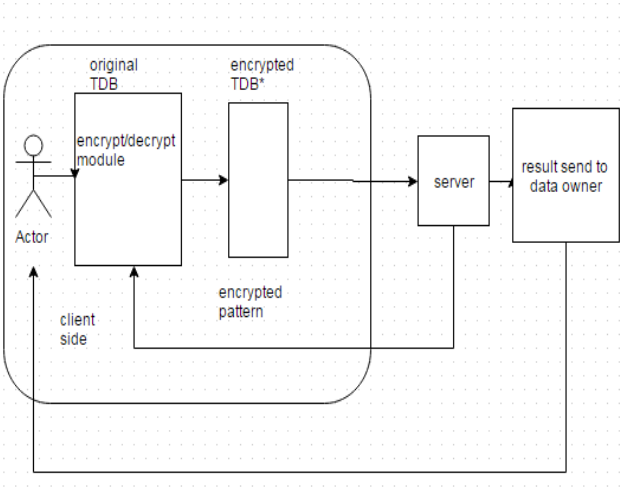

Fig 2. Architecture of proposed system

Advantage of Proposed System
ECDH algorithm provides bidirectional encryption of client and server which protect against the forging the contents of the communication and helps to prevent man in the middle attack.

4. OBJECTIVE

The objectives of our project are given below

- Privacy preserving issues for transferring a data
- Provide security to transaction database using Encryption scheme
- Achieving privacy bidirectional way using E/D Module to the transaction database
- Achieving accuracy and improve time management using ECDH Algorithm

5. MATHEMATICAL MODEL

Let S is the Whole System Consists:

S= {I,P,O}

I= Input

I= {TDB}

TDB= Transaction Database

P= Process

P= {E, W, K , N, D}

E=Encryption

To transfer original transaction database used 1-1 substitution cipher method in encryption scheme

Encryption (TDB) = TDB*

TDB* =Encrypted data

W= Weighted Support Construction

Calculate weighted support value

K= K-Grouping Method

Applying K grouping method on list of cipher item

e1, e2, e3…….en

N= Noise table construction

Calculate noise construction into database

N (e) = Noise value of cipher item

D=Decryption

Decryption (TDB*) =TDB

O= Output

Output: Secure Transaction data base.

6. USED ALGORITHM

1. Encrytion

Encryption scheme uses 1-1 substitution cipher method which transformed original transaction database D into its encrypted version D*. To improve the security fake transaction are added with encrypted database.

2. Fake Transaction Construction

The fake transaction was constructed based on the adding noise value to the original transaction database. In Fake transaction construction use Weighted support construction this approach was started with calculation of support of the items. Support count is the number of time the items occurred in the original transaction database. Given the items weighted support table, a few strategies were followed to group the items into groups of size k. This started from a basic grouping method. The output of group partitioning technique can be represented as the noise table. It expands the item weighted support table with an additional column "Noise value" representing, for each one cipher item e, the difference among the weighted support of the most successive cipher item in e's gathering and the weighted support of e itself, as reported in the item weighted support table. Adding longer fake transactions technically does not form privacy protection. However, for added protection, the system can decrease the lengths of the added fake transactions so that they are in line with the transaction lengths in transaction database D.

3. Decryption
When the client requests for the implementation of a pattern mining query to the server, for that indicate a minimum support threshold, then the server returns the compute frequent patterns from $D*$.

7. RESULT ANALYSIS

Performance of Encryption/Decryption Overhead

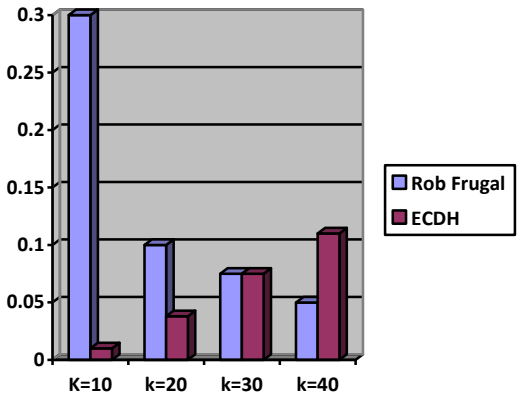

Fig. 3. Graph of Encryption/Decryption Overhead

Table 1. Value of comparison between Algorithms

K Values	RobFrugal	ECDH
10	0.3(sec)	0.01(sec)
20	0.1(sec)	0.038(sec)
30	0.075(sec)	0.075(sec)
40	0.05(sec)	0.11(sec)

The table shows the comparison between Rob Frugal and ECDH algorithm it shows encryption /decryption overhead values. Fig 3 graph shows ECDH is better than Rob frugal algorithm.

8. CONCLUSION AND FUTURE WORK

In this paper, focused on the difficulties of protection and providing safe guarding mining of regular examples on a scrambled outsourced TDB. In this proposed system an encryption plan, known as Rob Frugal that relies upon on 1–1 substitution figures for matters and including fake transaction to make each determine aspect offer the equal recurrence as ≥ k−1 others. It makes utilization of a smaller outline of the fake transaction from which the genuine backing of mined examples from the server may be correctly reoccupied. We assumed that a conservative model where the adversary knows the domain of items and their exact frequency and can use this knowledge to identify cipher items and cipher item sets. Distinct to past works, formally tested that our approach is lively towards sick-disposed attack in mild of the precise matters and their continuous backing. We also proposed ECDH (Elliptic Curve Diffie Hellman) key exchange algorithm was proposed after Rob frugal encryption scheme in order to provide privacy preserving outsourced mining. ECDH algorithm provides bidirectional encryption of client and server which protect against the forging the contents of the communication. The limitation of ECDH algorithm is not generating association rule properly. In Future we use FP-Growth algorithm in future for improving the generation of association rule.

9. ACKNOWLEDGMENT

It gives me a great pleasure and immense satisfaction to present this special topic on “Security of Private Data Extracted from Outsource Database in Association Rule Mining”, The success of this topic has throughout depended upon an exact blend of hard work and unending co-operation and guidance, extended to me.

10. REFERENCES

[1] R. Buyya, C. S. Yeo, and S. Venugopal, “Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities,” in *Proc. IEEE Conf. High Performance Comput. Commun.*, Sep. 2008, pp. 5–13.

[2] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, “Security in outsourcing of association rule mining,” in *Proc. Int. Conf. Very Large Data Bases*, 2007, pp. 111–122.

[3] L. Qiu, Y. Li, and X. Wu, “Protecting business intelligence and customer privacy while outsourcing data mining tasks,” *Knowledge Inform. Syst.*, vol. 17, no. 1, pp. 99–120, 2008.

[4] C. Clifton, M. Kantarcioglu, and J. Vaidya, “Defining privacy for data mining,” in *Proc. Nat. Sci. Found. Workshop Next Generation Data Mining*, 2002, pp. 126–133.

[5] I. Molloy, N. Li, and T. Li, “On the (in)security and (im)practicality of outsourcing precise association rule mining,” in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 872–877.

[6] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, “Privacy-preserving data mining from outsourced databases,” in *Proc. SPCC2010 Conjunction with CPDP*, 2010, pp. 411–426.

[7] Nilesh B. Prajapati, Krupali H. Shah, Information Technology Department, B.V.M.,V.V.Nagar, GTU, INDIA.Privacy Preserving in Association Rule mining"

[8] Cong Wang, Member, IEEE, Sherman S.M. Chow, Qian Wang, Member, IEEE, Kui Ren, Senior Member, IEEE, and Wenjing Lou, Senior Member “Privacy-Preserving Public Auditing for Secure Cloud Storage:”, IEEE, IEEE Transactions on computers, VOL. 62, NO. 2, FEBRUARY 2013.

[9] Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Lou, Department of ECE, Worcester Polytechnic Institute,Department of ECE, Illinois Institute of Technology,.

[10] Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data" .