ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A REVIEW PAPER ON ASSOCIATION RULES FOR PRIVACY-PRESERVING MINING FROM VARIOUS ENCRYPTION TECHNIQUES

Mr. Haibat Jadhav*, Prof. Pankaj Chandre
*Student Pursuing M.E. in Computer Networks from FIT,Khopi,Pune, India
Prof. at Department of Computer Engineering, FIT,Khopi,Pune, India

ABSTRACT

A new consequence has been significant in the Cloud computing in the example of data mining as a service, a company missing in proficiency or computational its mining needs to a third party service provider, the items and the association rules of the database are measured private property of the company. To care for company confidentiality, the data owner transforms its data and ships it to the server, sends mining queries to the server, and recovers the true patterns from the extracted patterns expected from the server. This paper, study the difficulty of the association rule mining task within a corporate privacy preserving framework. An attack model based on backdrop information and devises an idea for confidentiality preserve removal. It ensure that each altered item is the same with respect to the attackers backdrop information, from at smallest amount k1 other altered items. This paper will be efficient and keep confidentiality.

Keywords: Privacy preserving data mining, Transaction database, association rule mining.

INTRODUCTION

This paper is alert on giving confidentiality and mining of association rules. The main security issues is that the server has right to use to important data of the owner and may learn sensitive information from it. For example look at the transactions, the server otherwise an intruder who gain right of entry to the server can gain knowledge of which items are always co-purchased, the transactions and the mined pattern are the property of the data owner. So, it should remain protected from the server. This problem of protecting important secret information of companies is referred to as corporate confidentiality. Not like personal privacy, corporate privacy requires that both the entity items and the patterns of the set of data items. They are regard as corporate and thus must be protected. When there is attack during pattern mining must be protecting incorporate privacy-preserving framework. In particular, when the server possesses backdrop knowledge and conduct attacks on that basis, it should not be able to presumption the correct candidate item or item set matching to a given cipher item or item set with a possibility above a known threshold.

The progress of cloud computing also facilitate provider of services, particularly for the computational exhaustive tasks.The beneficial to achieve complicated investigation on great volumes of data in a cost-effective way, there exist numerous grave security issues of the data-mining as-a-service model.Assume a traditional frequency-based attack model in which the server know the accurate set of items in the owners data and as well, it also know the accurate support of each item in the unique data, the idea of use bogus items to secure against the frequency-based attack. The missing a formal theoretical investigation of privacy guarantee, and has been exposed to be faulty very recently in [5], where a method for breaking the future encryption is specified. The objective is to plan an encryption method which enables prescribed privacy guarantee to be proved, and to validate this model over large-scale real-life transaction databases. The client/owner encrypts its data by means of encrypt/decrypt Encryption or Decryption module, which can be basically treated as a black box from its perception. It is conscientious for transforming the input data into an encrypted database. The server conducts data mining and sends the encrypted patterns to the owner.

The encryption method has the property that the return supports are not true supports. The Encryption or Decryption module recover the true uniqueness of the returned patterns as well their true supports. It is minor to show that if the data are encrypted using 11 substitution [1] ciphers without

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449

using fake transactions, many ciphers and hence the transactions and patterns can be busted by the server with a high probability by initiation the frequency-based attack. Thus, the major focus of this paper is to plan encryption schemes such that formal privacy guarantee can be established against attacks conduct by the server using backdrop knowledge. The rest of the paper organized as follows Section 2. Related work Section 3. Literature survey Section 4. Assignment of pattern mining section 5. Confientiality model section 6.Encryption and decryption method section 7. Conclusion

RELATED WORK

Examine of PPDM [1] has fixed much interest recently;the main model here is that private

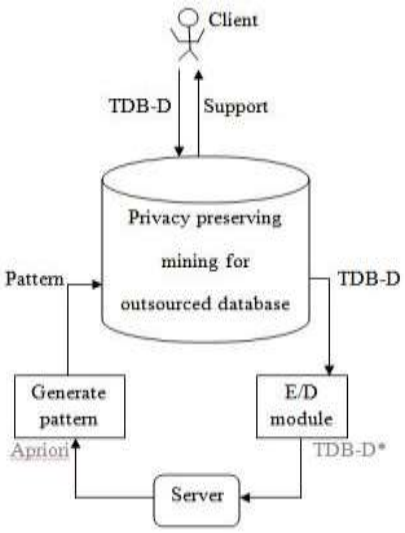

Fig. 1. Architecture of mining as service model[1]

data is composed from a number of collectors for consolidate the data and conduct mining. The collector is not trusted with defensive the privacy,so data are subjected to an arbitrary perturbation.Techniques [5][10]have been developed for perturbing the data so as to conserve privacy while ensure the mined patterns or other investigative properties are satisfactorily close to the patterns mined from unique data. R. Agrawal and R. Srikant was pioneer [12] and have been followed up by S. J. Rizvi and J. R. Haritsa [13].

PPDM [1] is not suited for corporate confidentiality in that some investigative properties are disclosed; another related problem is safe multiparty mining over distributed datasets,data on which mining is to be performed is partitioned,horizontally or vertically, and distributed among numerous parties. The partitioned data cannot be mutual and must remain private but the results of mining on the unification of the data are mutual among the participants, by means of multiparty safe protocols [14] [6][8]. This come close to partially apply corporate isolation, as local databases are reserved private,but it is too weak for this paper difficulty, as the ensuing patterns are disclose to multiple parties. The fussy difficulty attacked in this paper is of pattern mining within a corporate privacy-preserving structure. A key difference between this difficulty and the abovementioned PPDM harms is that, in this paper setting, not only the underlying data but also the mined consequences are not future for distribution and must remain confidential. In fussy when the server possesses backdrop information and conduct attacks on that basis, it should not be capable to deduction the correct candidate item or item set matching to a given cipher item or item set with a probability higher than a given threshold. The works that are nearly all associated by C. Tai, P. S. Yu, and M. Chen [12].They guess that the opposition possesses previous knowledge of the occurrence of items or item sets, which can be used to attempt to reidentify the encrypted items. The occupation utilizes a one-to-n item map together with nondeterministic addition of cipher items to care for the classification of entity items. The indoctrination system in [15] can be busted without using context-specific information. The victory of the attacks in [9] mainly relies on the existence of exceptional, frequent, and fake items, clear in [15]; this idea does not create any such items, and the attacks in [9] are not appropriate to this paper method. Assumed the attacker know exact frequency of single items. Using a similar privacy model [2]which require that each real item must have the same and also frequency calculate as k1 supplementary items in the dataset.

Table 1:TDB

TDB
Paper
Pencil Paper
Paper Pen
Pencil Sharpener
Pencil Eraser

Table 2:Item support table[1]

Item	Support
Paper	3
Pencil	3
Pen	1
Sharpener	1
Eraser	1

LITERATURE SURVEY

Data mining is a recently emerging field, connecting the three worlds of Databases, Artificial Intelligence and Statistics.The information age has enabled many organizations to gather large volumes of data. However, the usefulness of this data is negligible if meaningful information or knowledge cannot be extracted from it. Data mining, otherwise known as knowledge discovery, attempts to answer this need. In contrast to standard statistical methods[3], data mining techniques search for interesting information without demanding a priori hypotheses. The field of privacy has seen rapid advances in recent years because of the increases in the ability to store data. In particular, recent advances in the data mining field have lead to increased concerns about privacy. While the topic of privacy has been traditionally studied in the context of cryptography and information-hiding, recent emphasis on data mining has lead to renewed interest in the field Fosca Giannotti et.al[1]advantages like that the data owner produces the transaction as a stream and they have limited to maintain. If there are multiple transaction all transactions is send to provider for mining association rule that are local to individual store or global rule for organization there fore effective data mining for distributed owner is done and the disadvantages is the data perturbation is less attractive and gives only approximate result. Molloy et.al [2] A frequency analysis based attack that breaks a state of the art algorithm for outsourcing association the security approach to be insufficient to with stand known frequency attacks, and propose alternatives. The advantages is Association Rule Mining can be used to develop more secure schemes and data mining has potential of reducing the computation and software cost for the data owner. The disadvantages rule mining may not be practical, if the data owner concerned with the confidentiality.

M. Kantarcioglu and C. Clifton [3] Data mining can extract important knowledge from large data collections but sometimes these collections are split among various parties.Privacy concerns may prevent the parties from directly sharing the data, and some types of information about the data. The advantages are the data is securely shared among multiple parties in a network and data is secured by association rule, after distribution of original data also. The disadvantages is that the each time encryption decryption done at every site. So that approximate value generation[7] is done, at that time there are some changes done at every time which results loss in original data. Each party is responsible to choose their level of security, but as security level increases the cost also increases.

Yu, and M. Chen et al [4] for any service, privacy is a major concern. This paper focuses on frequent item set mining and examines the issue on how to protect privacy against the case where the attackers have precise knowledge on the supports of some items. The advantages is the various approach can provide more candidates for k-support anonymity with limits fake item as only the leaf nodes not the internal nodes of the taxonomy tree need to appear in transaction. And also internal node doesnt directly appear in the transaction but gets support from a set of transactions inferred through the taxonomy. The disadvantages are an approach is they do not offer the efficient result in case of additional item occurrence. And this approach is too weak for this paper problem, as the resulting patterns are disclosed to multiple parties.

MATHEMATICAL MODEL

Consent to D indicate the unique TDB that the owner has to protect the identification of entity items an apply an encryption function to D and transforms it to D*,the encrypted database. The notions of plain item sets and their cipher matching part are distinct and use o to denote the set of plain objects and E to refer to the set of cipher items. O=set of objects O = o1; :::; in D = t1; ; tm D is set of transaction. Patteren Mining Task is reqD(S) = suppD(S)/|D| jwhere, suppD(o) is individual support. f reqD(o) is frequency of o.

A]Attack Model [1] Item Based attack is guessing the plain item corresponding to the cipher item e with

probability prob(e):prob(e) = 1= j Cand(e) j where, e is cipher item.Cand(e) is set of candidate plain items Cand(e)O.Set Based Attack is guessing the plain item corresponding to the cipher item e with probability prob(e).prob(e) =1/|Cand(e)| where, E is cipher item set.C and E is a set of candidate plain item sets.

B. k-privacy model[5]- The various methods can be used encryption method in replacing every plain item in D by a 1-1 substitution cipher and K-Grouping the every item e there are at smallest amount others k-1 enciphered items with same support with adding fake transactions. In decryption method allow calculate the real support of each pattern. RobFrugal k-Private Grouping method the scheme obtaining Robust k-groups unsupported in D RobFrugal Grouping known the TDB D and its item support table in decreasing order of support:Step1 : Grouping together cipher items into groups of k contiguous objects Obtaining G =(G1, ...,Gm) Frugal Grouping Step2 : Modifying the groups of G by exchange operations, until no group of items is support in D.

Apriori Algorithm can be used simply counts item occurrences to determine the large 1-itemsets.

Step 1:Lp:Set of common itemsets of sizep(with min support)

Step 2:Cp:Set of candidate itemset of size p(potentially common itemsets)

Step 3:L1 = {common objects};

Step 4:for(p=1;Lp!=Ø;p++) do

Step 5:Cp+1=candidates created from Lp;

Step 6:for every transaction t in database do

Step 7:By increment the count of all candidates in Cp+1 that are contained in t

Step 8:Lp+1=candidates in Cp+1 with minimum support

Step 9:return ÙpLp;

CONCLUSION

This paper explored the problem of privacy-preserving mining of frequent patterns on an encrypted outsourced TDB.Traditional 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. This paper proposed an encryption scheme, called RobFrugal[1] that is based on 1-1 substitution ciphers for items and adding fake transactions to make each cipher item share the same frequency and makes use of a compact synopsis of the fake transactions from which the true support of mined patterns from the server can be efficiently recovered. This paper studies the complexity of the association rule mining task within a corporate privacypreserve structure[1][6]. An attack model based on backdrop information and devises an idea for confidentiality preserve removal. It ensure that each altered item is the same with respect to the attackers backdrop information, from at smallest amount k1 other altered items. This paper will be efficient and keep confidentiality, it will examine encryption scheme that can oppose such isolation vulnerabilities. It also involved in explore to improve the RobFrugal algorithm to be reduce the number of fake pattern.

REFERENCES

1. A Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang,“Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases ”, IEEE SYSTEMS JOURNAL , VOL. 7, NO. 3, SEPTEMBER 2013 .
2. V. Ciriani, S. D. C. di Vimercati, S. Foresti, and P. Samarati,“kanonymity in Proc. Secure Data anage of Decentralized System ”, 2007,pp. 323-353.
3. Qiu, Y. Li, and X. Wu,“Protecting business intelligence and customer privacy while outsourcing data mining tasks,”Knowledge Inform System ,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. M. Kantarcioglu and C. Clifton,“Privacy-preserving distributed mining of association rules on horizontally partition data,”IEEE Trans.Knowledge Data Eng., vol. 16, no. 9, pp. 1026-1037, Sep. 2004.
6. P. K. Prasad and C. P. Rangan,“Privacy preserving birch algorithm for clustering over arbitrarily partition databases,”in Proc. Adv. Data Mining Appl., 2007, pp. 146-157.
7. C. Tai, P. S. Yu, and M. Chen,“K-support anonymity based on pseudo taxonomy for outsourcing of numerous itemset mining,” in Proc. Int.Knowledge Discovery Data Mining, 2010, pp. 473-482

8. R. Agrawal and R. Srikant, Fast algorithm for mining association rules, in Proc. Int. Conf. Very Large Data Bases, 1994, pp. 487-499
9. P. Samarati,"Defending respondents identities in microdata release,"IEEE Trans. Knowledge Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov.2001.
10. B. Gilburd, A. Schuster, and R. Wolff,"k-ttp: A new privacy model for large scale distributed environment,"in Proc. Int. Conf. Very Large Data Bases, 2005, pp. 563-568.
11. 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. 411-436.
12. R. Agrawal and R. Srikant,"Privacy-preserving data mining,"in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 439-451.
13. S. J. Rizvi and J. R. Haritsa,"Maintain data privacy in association rule mining,"in Proc. Int. Conf. Very Large Data Bases, 2002, pp. 682-693.
14. M. Kantarcioglu and C. Clifton,"Privacy-preserving distributed mining of association rules on horizontally partition data,"IEEE Trans.Knowledge Data Eng., vol. 16, no. 9, pp. 1026-1037, Sep. 2004.
15. W. K. Wong and et al.,"Security in outsourcing of association rule mining,"in Proc. Int. Conf.Very Large Data Bases, 2007, pp. 111-122.