Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Encrypted Association Rule Mining for Outsourced Data Mining

Swapnil C. Salunke
Dept. of Computer Engg,
Dr.D.Y.Patil SOET, Lohagaon Pune,
University of Pune, India

Prof. Roshani Raut (Ade)
Faculty at Dept. of Computer Engg,
Dr.D.Y.SOET, Lohagaon Pune,
University of Pune, India

Abstract - Spurred by developments such as cloud computing there has been considerable recent interest in the paradigm of data mining-as-a-service. A company (data owner) lacking in expertise or computational resources can outsource its mining needs to a third party service provider (server). However, both the items and the association rules of the outsourced database are considered private property of the corporation (data owner). To protect corporate privacy, 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 received from the server. In this paper, we study the problem of outsourcing the association rule mining task within a corporate privacy-preserving framework. We propose an attack model based on background knowledge and devise a scheme for privacy preserving outsourced mining.

Index Terms—Association rule mining, privacy-preserving outsourcing, data mining

I. INTRODUCTION

A. Background

With the advent of cloud computing and its model for IT services based on the internet and big data centers, the outsourcing of data and computing services isacquiring a novel relevance, which is expected to skyrocket in the near future. Business intelligence and knowledge discovery services, such as advanced analytics based on data mining technologies, are expected to be among the services amenable to be externalized on the cloud, due to their data intensive nature, as well as the complexity of data mining algorithms. Thus, the paradigm of mining and management of data as service will presumably grow as popularity of cloud computing grows. This is the data mining-as-a-service paradigm, aimed at enabling organizations with limited computational resources and/or data mining expertise to outsource their data mining needs to a third party service provider.

Although it is advantageous to achieve sophisticated analysis on tremendous volumes of data in a cost-effective way, there exist several serious security issues of the data-mining asa-service paradigm. One of the main security issues is that the server has access to valuable data of the owner and may learn sensitive information from it. For example, by looking at the transactions, the server (or an intruder who gains access to the server) can learn which items are always copurchased. However, both the transactions and the mined patterns are the property of the data owner and should remain safe from the server.

This problem of protecting important private information of organizations/companies is referred to as corporate privacy. Unlike personal privacy, which only considers the protection of the personal information recorded about individuals, corporate privacy requires that both the individual items and the patterns of the collection of data items are regarded as corporate assets and thus must be protected. In this paper, we study the problem of outsourcing the association rule mining task within a corporate privacy preserving framework. A substantial body of work has been done on privacy-preserving data mining (PPDM) in a variety of contexts. A common characteristic of most of the previously studied frameworks is that the patterns mined from the data (which may be distorted, encrypted, anonymzed, or otherwise transformed) are intended to be shared with parties other than the data owner. The key distinction between such bodies of work and our problem is that, in the latter, both the underlying data and the mined results are not intended for sharing and must remain private to the data owner.

B. Purpose

Data mining is an important approach to converting “data rich” to “knowledge rich” in strategic decision making. Outsourcing data mining tasks is a common practice widely adopted by many organizations. One example is that American Express outsourced its data mining tasks for the purpose of identifying high potential customers for crossselling products. When data mining tasks are outsourced, it is important to protect the following three elements from which business intelligence and customer privacy can be drilled down the source data which is the database of all transactions, the mining results which are frequent

itemsets as well as their supports in our context, and the mining requests which are itemsets of interests. In recent years, people have proposed various methods

for privacy preserving data mining. The proposed methods can be classified into two categories: data perturbation and secure multiparty computation. With data perturbation methods, the raw data is modified by adding random noise so that it no longer reveals privacy related information, while the statistical properties of the data are retained. These methods

may have unpredictable impacts on data mining precision. On the other hand, secure multiparty computation works only in distributed data sharing environment where multiple parties participate in a protocol so as to share their own pieces of private data, and cooperate to get the final results.

II. SYSTEM ARCHITECTURE

Architecture of System

The proposed system architecture is as shown in figure 1. Data of the user will be stored in Local DB. First this data will be encoded and will send this encoded data to the service provider. Before sending this data to service provider we will apply AES Encryption algorithm on data. On receiving this encrypted data service provider will apply modified Apriori Algorithm to find association rules and will generated results. These results will be in encrypted format which will be sent back to user. User will receive these results, encrypt it and then encode it to see the final results.

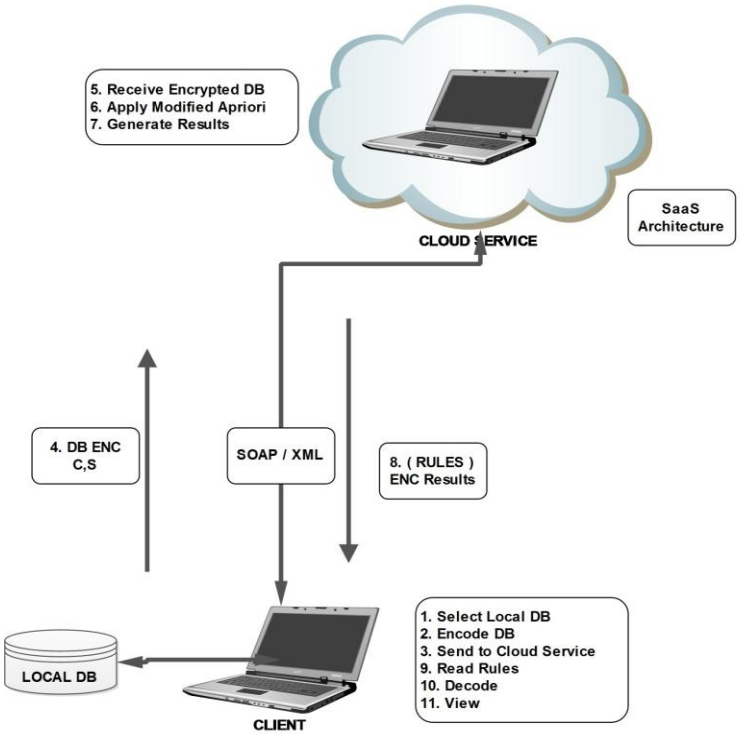

Figure 1. System Architecture

Problem Formulation

To develop a system where in the privacy of the outsourced data can be preserved.
Let I = {I1,i2,…..im} set of literals called items.
Given a set of transaction D, where each transaction T is a set of items such that T ⊆ I.
An association rule is an expression X ➔Y
Where X ⊆ I, Y ⊆ I and X ∩ Y = Ø
An example of such rule :- 90% of customers who buy hamburgers also buy coke.
90% here is called confidence of the rule.
90% of the transaction that contain X also contain Y = $\frac{|X \cup Y|}{|X|}$

The support of rule is the percentage of transactions that contain both X and Y, which is calculated as $\frac{|X \cup Y|}{|N|}$
The problem of mining association rules is to find all rules that are greater than the user specified minimum support and minimum confidence and hide/protect such rules from unauthorized persons.

III. PROPOSED METHODOLOGY

A. Mathematical Formulation

Let G be the global set
G = {U, T, S, F, D, R}
U is a set of all users using the server database and mining services from the server. And U = (U1 ,U2 ,U3, ...Un) where n = ∞
T is a set of all transactions and T = (T1 ,T2 ,T3, ...Tn) where n = ∞
S is the server component of the system. The server is responsible for registering, authenticating and providing associations to the end user.
S = (S1,S2 ,S3, ...Tk) where K ≠ ∞
F is a set of functions.

F = (F1 ,F2 ,F3, ...Fn) where $F \neq \infty$
D is a set of local database that a user owns. It consists of data tables having data items related to the products and their sales transactions.
D = (D1 ,D2)
R is a set of mining rules of Apriory that are applied on the input dataset provided by the client from his D.

B. Morphism

Stored to server ← Registration (Client details)
Access application ← Login (User ID, Password)
Stored to local database ← Add transactions (Transaction details)
EDS ← Encode data set (Data set, Indexing)
ENC(DS) ← Encrypt data set (EDS , key)
Export to S ← Export to server (ENC (DS), Key)
EDDS ← Decrypt data set (ENC (DS), Key)
<Rules>← Apply Apriory (EDDS, S, C)
E<Rules>← Encrypt rules (<Rules>, Key)
Client ← Export to client (E<Rules>)

C. Algorithm

1. In case of large dateset, Apriori algorithm produce large number of candidate itemsets. Algorithm scan database repeatedly for searching frequent itemsets, so more time and resources are required in large number of scans so it is inefficient in large data set.
2. Assumes transaction database is memory resident.
3. This algorithm requires large number of dataset scans.
4. It only explains the presence and absence of an item in transactional databases.
5. Apriori algorithm works on transactions dataset but modified Apriori works on indexes of dataset.

V = Call function wiener2 (Xi)
C1, C2, …CK = Call function kmeans (V, K)
For each cluster Ci
Cdn : Candidate itemset of size n
Ln: frequent itemset of size n
L1 = {frequent items};
For (n=1; Ln != ϕ ; n++)
Do begin
Cdn+1 = candidates generated from Ln;
For each transaction T in database do
Increment the count of all candidates in Cdn+1 which are contained in T
Ln+1= candidates in Cdn+1 with min_support
End
UnLn are the frequent itemsets generated
End

IV. CONCLUSION

The problem of (corporate) privacy-preserving mining of frequent patterns (from which association rules can easily be computed) on an encrypted outsourced TDB. Proposed technique executes an encryption scheme, called AES that is based on 1–1 substitution ciphers for items and adding fake transactions to make each cipher item share the same frequency as ≥ k−1 others. It 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 technique also proposed a strategy for incremental maintenance of the synopsis against updates consisting of appends and dropping of old transaction batches. Unlike previous works, proposed technique formally proved that our method is robust against an adversarial attack based on the original items and their exact support.

REFERENCES

[1] 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] Ling Qiu, Yingjiu Li, Xintao Wu "Protecting business intelligence and customer privacy while outsourcing data mining tasks" Springer-Verlag London Limited 2007
[3] Ian Molloy, Ninghui Li, and Tiancheng Li "On the (In)Security an (Im)Practicality of

Outsourcing Precise Association Rule Mining" 2009 Ninth IEEE International Conference on Data Mining

[4] Rajkumar Buyya, Chee Shin Yeo AND Srikumar Venugopal "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities" The 10th IEEE International Conference on High Performance Computing and Communications

[5] Ling QIU, Yingjiu LI, Xintao WU "An Approach to Outsourcing Data Mining Tasks while Protecting BusinessIntelligence and Customer Privacy" Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06)

[6] Murat Kantarcioglu and Chris Clifton"Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 16, NO. 9, SEPTEMBER 2004

[7] Shyue-Liang Wang, Yu-Huel Lee, Steven Billis, Ayat Jafari "Hiding Sensitive Items in Privacy Preserving Association Rule Mining" 2004 IEEE International Conference on Systems.

[8]]R. Agrawal and R. Srikant, "Privacy-preserving data mining," in Proc.ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 439–450.

