ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Privacy Preserved Association Rule Mining For Attack Detection and Prevention

V.Ragunath[1], C.R.Dhivya[2]

P.G Scholar, Department of Computer Science and Engineering, Nandha College of Technology, Erode, Tamilnadu[1]

Assistant Professor, Department of Computer Science and Engineering, Nandha College of Technology, Erode, Tamilnadu[2]

Abstract: In this project, a company to protect the corporate privacy, the data owner transforms its data and shifts it to the server and recovers the true patterns from the extracted patterns received from the server. The client owner encrypts its data using an (E/D) module. Our encryption scheme has the property that the returned supports are not true supports. The encrypt/decrypt 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. The data owner recover true pattern from the E/D module by using incremental maintenance. The privacy guarantees of our method in case of known-plaintext attacks, chosen-plaintext attacks and chosen-cipher text attacks. Finally which one has access the data by unauthorized permission that one can be detected and removed from particular group.

Index Terms: Association rule mining, privacy-preserving, outsourcing

I. INTRODUCTION

Association rule mining, one of the most important and well researched techniques of data mining, was first introduced in. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc. Various association mining techniques and algorithms will be briefly introduced and compared later. Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub problems. One is to find those item sets whose occurrences exceed a predefined threshold in the database; those item sets are called frequent or large item sets. The second problem is to generate association rules from those large item sets with the constraints of minimal confidence. Suppose one of the large item sets is Lk, Lk = {I1, I2,…,Ik}, association rules with this item sets are generated in the following way: the first rule is{I1,I2,…,Ik-1}⇒ {Ik}, by checking the this rule can be determined as interesting or not.

Then other rule are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty. Since the second sub problem is quite straight forward, most of the researches focus on the first sub problem.

The first sub-problem can be further divided into two sub-problems: candidate large item sets generation process and frequent item sets generation process. Call those item sets whose support exceed the support threshold as large or frequent item sets, those item sets that are expected or have the hope to be large or frequent are called candidate item sets.are the protocols to calculate the shortest path. In the distance vector routing, each router sends a vector of distances to

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

its neighbor node. The vector contains distances to all nodes. In the link state routing, each router sends a vector of distances to all nodes. The vector contains distances to the neighbor node. Privacy preserving data mining has the potential to increase the reach and benefits of data mining technology. Privacy-preserving data mining considers the problem of running data mining algorithms on confidential data that is not supposed to be revealed even to the party running the algorithm. First, sensitive raw data like identifiers, names, addresses and so on, should be modified or trimmed out from the original database, in order for the recipient of the data not to be able to compromise another person's privacy. Second, sensitive knowledge which can be mined from a database by using data mining algorithms, should also be excluded, because such a knowledge can equally well compromise data privacy. So, privacy preservation occurs in two major dimensions: users' personal information and information concerning their collective activity. Most legal efforts have been directed to protecting data of the individual.

II. RELATED WORK

The main model here that is corporate data is collected from a number of sources by a collector for the purpose of consolidating the data and conducting mining. The collector is trusted with protecting the privacy, so data are subjected to a random perturbation as it is collected. This approach is suitable for corporate privacy, in that some analytical properties are discloser. This approach partially implements corporate privacy, as local databases are kept private, but it is too weak for our outsourcing problem, as the resulting patterns are disclosed to multiple parties.

TDB
Bread
Milk Bread
Bread Milk
Water Milk
Bread Beer
Bread Eggs
Water

(a)

Item	Sup
Bread	5
Milk	3
Water	2
Beer	1
Eggs	1

(b)

The particular problem attacked in our paper is outsourcing of pattern mining within a corporate privacy-preserving framework. Our empirical study also shows that in practice, due to specific characteristics of the real transaction datasets (e.g., the power-law distribution of items), even the privacy-preserving methods for less-strict privacy models can enjoy a relatively high level of privacy in practice.

We propose an incremental method for updating the compact synopsis maintained by the owner against updates to the database.

III. PROPOSED SYSTEM

Personal privacy guarantees can be proven against attacks conducted by the server using background knowledge, while keeping the resource requirements under control. The main idea of privacy requires that, for each cipher text item, there are at least k−1 distinct cipher items that are indistinguishable from the item regarding their supports.

We proposed an encryption scheme, called Rob Frugal, 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 and we introduce a

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

strategy for incremental maintenance of the synopsis against updates consisting of appends and dropping of old transaction batches .This method is robust against an adversarial attack based on the original items and their exact support.

We conduct an analysis source based on our attack model and prove that the probability that an individual item a transaction, or a pattern can be broken by the server can always be controlled to be below a threshold chosen by the owner, by setting the anonymity threshold k.

We are using modules in Preprocessing, Encryption, Decryption and Constructing Fake Transactions.

IV. SYSTEM MODEL

A. Preprocessing

We select Coop stores database for processing. Data Preprocessing is a Computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures and visualization.

B. Encryption Overhead

Which transforms a TDB D into its encrypted version D. Our scheme is parametric with respect to k > 0 and consists of three main steps: 1) using 1-1 substitution ciphers for each plain item; 2) using a specific item k-grouping method; and 3) using a method for adding new fake transactions for achieving k-privacy. The constructed fake transactions are added to D (once items are replaced by cipher items) to form D, and transmitted to the server. A record of the fake transactions, i.e., DF = D *\D, is stored by the E/D module in the form of a compact synopsis, as discussed in Sections V-C and V-D.

C. Decryption Overhead

When the client requests the execution of a pattern mining query to the server, specifying a minimum support threshold the server returns the computed frequent patterns from D*. Clearly, for every item set S and its corresponding cipher item set E, we have that suppD(S) =¡suppD* (E). For each cipher pattern E returned by the server together with suppD*\D (E), the E/D module recovers the corresponding plain pattern S. It needs to reconstruct the exact support of S in D and decide on this basis if S is a frequent pattern. To achieve this goal, the E/D module adjusts the support of E by removing the effect of the fake transactions. suppD(S)= suppD.(E). suppD*\D(E).

D. Constructing Fake Transactions

Given a noise table specifying the noise N (e) needed for each cipher item e, we generate the fake transactions as follows. First, we drop the rows with zero noise, corresponding to the most frequent items of each group or to other items with support equal to the maximum support of a group. Second, we sort the remaining rows in descending order of noise.

E. Attack prevention

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

As the sophistication of cyber criminals continues to increase, their methods and targets have also evolved. Instead of building the large Internet worms that have become so familiar, these criminals are now spending more time concentrating on wealth gathering crimes, including fraud and data theft. An online article from Cyber Media India Online Ltd. , suggests that because home users often have the poorest security measures in place , they have become the most widely targeted Although home users may not feel like they are connected to a network , any activity on the Internet can be considered " networked activity. " Therefore, protection measures employed by networks may also benefit the home user. Routers and firewalls can help control access to a home computer, but more specific steps may be utilized group.

V. PRIVACY PRESERVING TECHNIQUE

A. K-anonymity

Anonymity is used to prevent identification of individual records in the given data set. The database is said to be K-anonymous where attributes are generalized until each row is identical with at least k-1 rows. K-Anonymity guarantees that the data released is accurate. The two different techniques used by k-anonymity [14] method are, generalization and suppression. To protect respondent's identity when releasing data, data operators often remove explicit identifiers like names, social security numbers etc. One of the interesting aspects of k-anonymity is its association with protection techniques that preserve originality of the data in each record. Basic approach towards privacy protection in data mining has to perturb the data before it is mined.

The guarantee given by k-anonymity [11] is that no information can be linked to groups of less than k individuals. In Generalization for k-anonymity which losses considerable amount of information, for higher-dimensionality data set. K-anonymity model for multiple sensitive attributes consist of three kinds of information disclosure:

1. Identity Disclosure: An individual who can link to a particular record in the published data set is known as identity disclosure.
2. Attribute Disclosure: When the sensitive information regarding particular individual revealed is known as attribute disclosure.
3. Membership Disclosure: The information regarding individual belongs from data set is presented or not revealed is a membership disclosure.

Anonymity refers to a state where data does not show its identity. A dataset which satisfies k-anonymity if every record in the dataset is not distinguished from at least k-1 other records with respect to every set of quasi-identifier attributes is known as k-anonymity dataset.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose of (corporate) privacy-preserving mining of frequent patterns (from which association rules can easily be computed) on an encrypted outsourced TDB (Transaction Database). We proposed an encryption scheme. That is based on 1–1 substitution ciphers for items and adding fake transactions. It could be interesting to consider other attack models where the attacker knows some pairs of items and their cipher values. We could study the privacy guarantees of our method in case of known-plaintext attacks (where the adversary knows some item, cipher item pairs). Another interesting direction is to relax our assumptions about the attacker by allowing him to know the details of encryption algorithms and/or the frequency of item sets and the distribution of transaction lengths.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

REFERENCES

[1] Agrawal R and Srikant R, "Privacy-Preserving Data Mining", ACM SIGMOD International Conference on Manage Data, pp. 439–450, 2000.

[2] Agrawal R and Srikant R, "Fast Algorithms for Mining Association Rules", International Conference on Very Large Data Bases, pp. 487–499, 1994.

[3] Buyya R, Yeo C S, and Venugopal S, "Market-Oriented Cloud Computing: Vision, Hype and Reality for Delivering Services as Computing Utilities", IEEE Conference on High Performance Communication, pp. 5–13, September 2008.

[4] Ciriani V, di Vimercati C, Foresti S, and Samarati P, "K-Anonymity", Secure Data Manage Decentralized System, pp. 323–353, 2007.

[5] Clifton C, Kantarcioglu M, and Vaidya J, "Defining Privacy for Data Mining", National Science Found Workshop on Next Generation Data Mining, pp. 126–133, 2002.

[6] Cormen H T, Leiserson E C, Rivest L R, and Stein C, "Introduction to Algorithms", Cambridge MA, MIT Press, pp. 555–561, 2001.

[7] Feder T, Aggarwal G, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu, "Approximation Algorithms for K-Anonymity", Journal of Privacy Technology, pp. 112-120, 2005.

[8] Giannotti F, Lakshmanan V L, Monreale A, Pedreschi D and Wang H, "Privacy-Preserving Data Mining from Outsourced Databases", IEEE Transactions on Knowledge Data Engineering, pp. 411–426. 2010.

[9] Gilburd B, Schuster A, and Wolff R, "A New Privacy Model for Large Scale Distributed Environments", International Conference on Very Large Databases, pp. 563–568, 2005.

[10] Huang L J, Chuang K T, and M.-S. Chen, "Power-Law Relationship and Self-Similarity in the Item Set Support Distribution: Analysis and Applications", International Conference on Very Large Databases, volume 17, no. 5, pp.1121–1141, 2008.

[11] Kantarcioglu M and Clifton C, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge Data Engineering, volume 16, no. 9, pp.1026–1037, 2004.

[12] Lakshmanan V, Giannotti F, Monreale A, Pedreschi D, and Wang H, "Privacy-Preserving Outsourcing of Association Rule Mining", ISTI-CNR, Pisa, Italy, Technology Representation, 2009.

[13] Liu K, Kargupta H and Ryan J, "Random Projection Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining", IEEE Transactions on Knowledge Data Engineering, 2006.

[14] Molloy I, Li N, and Li T, "On The Security And Practicality of Outsourcing Precise Association Rule Mining", IEEE International Conference on Data mining, pp. 872–877, December 2009.

[15] Prasad K P and Rangan P C, "Privacy Preserving BIRCH Algorithm for Clustering over Arbitrarily Partitioned Databases", Advantages of Data Mining Application, pp. 146–157, 2007.

[16] Qiu L, Li Y, and Wu X, "Protecting Business Intelligence and Customer Privacy While Outsourcing Data Mining Tasks", International Conference on Knowledge Information System, volume 17, no.1, pp. 99–120, 2008.