International Journal of Scientific Research in Computer Science, Engineering and Information Technology

© 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307

Privacy-Preserving Outsourced Association Rule Mining on Vertically Partitioned Databases

K Geetha[1], K Gurunadha Guptha[2], S N V A S R K Prasad[3]

*[1] Assistant Professor, Department of Computer Science and Engineering, Sri Indu College of Engineering & Technology, Telangana, India
[2,3]Assistant Professor, Department of Computer Science and Engineering, Sri Indu College of Engineering & Technology, Telangana, India

ABSTRACT

Data Analysis techniques that are Association manage mining and Frequent thing set mining are two prominent and broadly utilized for different applications. The conventional framework concentrated independently on vertically parceled database and on a level plane apportioned databases on the premise of this presenting a framework which concentrate on both on a level plane and vertically divided databases cooperatively with protection safeguarding component. Information proprietors need to know the continuous thing sets or affiliation rules from an aggregate information set and unveil or uncover as few data about their crude information as could reasonably be expected to other information proprietors and outsiders. To guarantee information protection a Symmetric Encryption Technique is utilized to show signs of improvement result. Cloud supported successive thing set mining arrangement used to exhibit an affiliation govern mining arrangement. The subsequent arrangements are intended for outsourced databases that permit various information proprietors to proficiently share their information safely without trading off on information protection. Information security is one of the key procedures in outsourcing information to different outside clients. Customarily Fast Distribution Mining calculation was proposed for securing conveyed information. These business locales an issue by secure affiliation governs over parceled information in both even and vertical. A Frequent thing sets calculation and Distributed affiliation administer digging calculation is used for doing above method adequately in divided information, which incorporates administrations of the information in outsourcing process for disseminated databases. This work keeps up or keeps up proficient security over vertical and flat perspective of representation in secure mining applications.

Keywords: Association Rules Mining, Frequent Item Set Mining, Privacy-Preserving Data Mining, Partitioned Data

I. INTRODUCTION

Frequent item set mining and association rule mining, two widely used data analysis techniques, are generally used for discovering frequently co-occurring data items and interesting association relationships between data items respectively in large transaction databases. These two techniques have been employed in applications such as market basket analysis, health care, web usage mining, bioinformatics and prediction. A transaction database is a set of transactions, and each transaction is a set of data items with a unique TID (Transaction ID). An item set Z is regarded frequent if and only if *Supp (Z)* $\geq$ *Ts*, where *Ts* is a threshold specified by the data miner. *Supp (Z)* is Z's support, which is defined as Z's occurrence count in the database. An association rule is expressed using $X \Rightarrow Y$, where X and Y are two disjoint item sets. $X \Rightarrow Y$ indicates that X's occurrence implies *Y's* occurrence in the same transaction with a certain confidence. A supermarket's transaction database as an example, where a transaction is some customer's shopping list. A customer buying ―bread‖ and ―butter‖ will also buy ―milk‖. Then {bread, butter} ⇒ milk is a possible association rule. $X \Rightarrow Y$ is meaningful and useful if the confidence is high and $X \cup Y$ is frequent. More specifically, $X \Rightarrow Y$ is regarded as an association rule if and only if *Supp*$(X \cup Y) \geq$ *Ts* and *Conf* $(X \Rightarrow Y \geq$ *Tc*. *Conf* $(X \Rightarrow Y)$ as the confidence of $X \Rightarrow Y$. The latter is the probability of *Y's* occurrence given X's occurrence (i.e. *Conf* $(X \Rightarrow Y)$ = *Supp*$(X \cup Y)$/*Supp*(X)). *Tc* denotes the threshold specified by the data miner. The values of *Ts* and *Tc* are generally

CSEIT172483 | Received : 18 July 2017 | Accepted : 29 July 2017 | July-August-2017 [(2)4: 280-286]

configured based on the type of transactions, the usage of the mining result, the size of database, etc. It is easy to mine association rules after mining frequent item sets and obtaining their supports. Most association rule mining algorithms are built based on frequent item set mining algorithms.

If each data owner has one or more rows (i.e. transactions) in the joint database, we say that the database is *horizontally partitioned*. If each data owner has one or more columns in the joint database, the database is considered *vertically partitioned*. This work focuses on both vertically partitioned databases and horizontally partitioned database. In this work, proposing a cloud-aided privacy-preserving frequent item set mining solution for partitioned databases, which is then used to build a privacy-preserving association rule mining solution. Both solutions are designed for applications where data owners have a high level of privacy requirement. The solutions are also suitable for data owners looking to outsource data storage i.e. data owners can outsource their encrypted data and mining task to a semi-trusted (i.e. curious but honest) cloud in a privacy preserving manner. To the best of our knowledge, this is the first work on outsourced association rule mining and frequent item set mining for vertically and horizontally partitioned databases.

The key underlying techniques in these solutions are an efficient enhanced secure encryption scheme and a secure outsourced comparison scheme.

II. LITERATURE REVIEW

Lichun Li, Rongxing Lu , Kim-Kwang Raymond Choo, Anwitaman Datta, and Jun Shao[1] has proposed Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases‖ which focus on privacy-preserving mining on vertically partitioned databases. In such a scenario, data owners wish to learn the association rules or frequent itemsets from a collective data set and disclose as little information about their (sensitive) raw data as possible to other data owners and third parties.

J. Vaidya and C. Clifton [2], has proposed the first work to identify and address privacy issues in vertically partitioned databases, a secure scalar product protocol is presented and used to build a privacy-preserving frequent itemset mining solution. Association rules can then be found given frequent itemsets and their supports. Since the publication of this seminal work, a number of privacy preserving association rule mining or frequent itemset mining solutions have been published.

B. Rozenberg and E. Gudes [3] has proposed the all existing solutions, with the exception of do not utilize a third-party server to compute the mining result. Some solutions use asymmetric encryption to compute the supports of itemsets, while other solutions use a secure scalar product protocol, a set intersection cardinality protocol or a secret sharing scheme to perform these computations. A majority of these solutions expose exact supports to all data owners, resulting in the leakage of information about the data owners’ raw data.

S. Zhong, [4], has proposed in this: There are two privacy-preserving solutions for frequent itemset mining. The first solution exposes exact supports, which is not desirable. The second solution does not expose exact supports. However, association rules cannot be mined based on the result of second solution because confidences cannot be computed without the exact supports. In addition, this solution’s method cannot be used to mine association rules because securely computing confidence is more complicated than computing support. In comparison with this solution, our frequent itemset mining solution’s computational complexity is significantly lower.

F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang[5] In existing solutions, the data owner outsources their data and the mining task to the cloud, but at the same time, wish to keep the raw data secret from the cloud. Generally, data items in the database are encrypted using a substitution cipher prior to outsourcing.

W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis [6] proposed a solution to counter frequency analysis attack on substitution cipher. However, a later work demonstrated that [19]’s solution is not secure. Giannotti et al. proposed a solution based on k-anonymity frequency.

J. Lai, Y. Li, R. H. Deng, J. Weng, C. Guan, and Q. Yan[7]Another recent work proposed a privacy-preserving outsourced association rule mining solution based on predicate encryption. This solution is resilient to chosen-plaintext attacks on encrypted items, but it is vulnerable to frequency analysis attacks. Applying this solution to vertically partitioned databases will also result in the leakage of the exact supports to data owners. In this paper, our adversary model is different. We assume the cloud has knowledge of the item frequencies instead of chosen plaintext-ciphertext pairs, and our solutions are resilient to frequency analysis attacks.

M. Kantarcioglu and C. Clifton [8] various works have been proposed to modify the statistical distribution of data by implementing substitution ciphers and adding fake transactions. In this following encoding steps were proposed:

- ’1-to-n’ mapping of individual original items.
- Additional unique and common items are added to theTransaction-level mappings.
- Addition of fake items to each transaction these proposed algorithms are straightforward and provides a good level of practical privacy protection.

However, the processing, and especially storage overheads are quite considerable. The ’1-ton’mappings means that the data storage required is increased by quite a number of folds.

O. Goldreich[9] Furthermore, confirmed an attack which showed that the scheme was no more secure than ’1-to-1’ mappings. The paper then proposed a stricter notion of security based on perfect secrecy from the works of Shannon. However, the paper then stipulates that the achievement of such strong notions of privacy is extremely impractical as the resources required to meet them quickly exceeds the owner having to perform the mining themselves.To improve the privacy notions three practical implementation changes were proposed.

H. Grosskreutz, B. Lemmen and S. Rüping, In [10], a confidentiality notion known as k-anonymity was proposed, in which each susceptible item in a transaction is protected by k-1 other items or records of similar support. In this manner an attacker is unable to distinguish between the different items of k-size, thus preventing re-identification. This was relaxed using a concept known as probabilistic k-anonymity, in which the correct reidentification of data by an attacker is at most 1/k. This was achieved through the partitioning of transactions and data permutations.

Can Xiang, Chunming Tang, In [11] present two secure outsourcing schemes for modular exponentiations, which enable users to strongly outsource modular exponentiations to a single untrusted cloud server and detect the dishonest behavior of untrusted cloud server. The first one is a sheltered outsourcing scheme for variable base-variable exponent modular exponentiation, while the second is for synchronized modular exponentiations.

Xinjing Ge, Li Yan, Jianming Zhu, Wenjie Shi In [12] explores the issue of privacy preserving distributed association rule mining in vertically partitioned data among multiple parties, and based on the Shamir’s secret sharing technique a collusion-resistant algorithm of distributed association rule mining is proposed, which prevents effectively the collusive behaviors and conducts the computations across the parties without compromising their data privacy.

Xuan Canh Nguyen, Hoai Bac Le, Tung Anh Cao In [13] propose an Enhanced M.Hussein et al.’s Scheme (EMHS) for privacy-preserving association rules mining on horizontally distributed databases. EMHS is based on the M.Hussein et al.’s Scheme (MHS) proposed in 2008 and improves privacy and performance when increasing the number of sites. EMHS uses two servers, Initiator and Combiner, combined with MFI approach to generate candidate set and homomorphic Paillier cryptosystem to compute global supports. Experimental results demonstrate that the performance of EMHS is better than MHS in specific databases when increasing the number of sites.

Mahmoud Hussein, Ashraf El-Sisi, Nabil Ismail[14] proposed a change to privacy preserving association rule mining on distributed homogenous database environment.

The algorithm used is faster than previous one which modified with privacy and better results. The algorithm is based on a semi-honest model with collision probability. The flexibility to extend to any number of sites without modification in implementation. And also any increase does not add more time to algorithm because clients sites perform the mining operations in the same time so the overhead in execution time only. The total bit-communication cost for algorithm is function in sites.

Moez Waddey, Pascal Poncelet, Sadok Ben Yahia,[15] proposed a new algorithm for assess closed frequent itemsets in distributed environment, using encryption to ensure privacy concerns. We address secure mining of association rules over partitioned data horizontally.

J. Vaidya, Clifton [16] Propose the association rule mining where transactions are distributed over sources. Every site holds a few characteristics of every site, and the locales wish to work together to recognize all inclusive substantial association rules. Be that as it may, the locales must not uncover singular site information. We exhibit a two-party calculation for effectively finding incessant itemsets with least bolster levels, without either site uncovering singular exchange values.

N. V. Muthu Lakshmi1 & K. Sandhya Rani[17] new model is proposed to find association rules by satisfying the privacy constraints for vertically partitioned databases at n number of sites along with data miner. This model adopts different cryptography techniques such as encryption, decryption and scalar product technique to find association rules efficiently and securely for vertically partitioned databases.

Zhu Yu- quan, Tang Yang, Chen Geng [18] For resolving the problem that the existing protocol of secure two-party vector dot product computation has the low efficiency and may disclose the privacy data, propose a method which is effective to find frequent item sets on vertically distributed data is put forward. The method uses semi-honest third party to participate in the calculation, put the converted data of the parties to a third party to calculate. The results show that compared to the original Vector dot product algorithm, the method can obviously improve the algorithm efficiency and accuracy of the results at the precondition that assured the data privacy of all parties.

III. PROBLEM STATEMENT

In conventional framework, work is done on regular thing set database dividing. In any case, these works were not made in a point of view for disseminated outsourcing of information mining. Our intend to outline a framework to dispersed mining outsourcing utilizing a Horizontal and vertical circulations of the database to numerous servers.

IV. TOPIC INITIATIVES

In conventional framework, work is done on continuous thing set database apportioning. Notwithstanding, these works were not made in a point of view for conveyed outsourcing of information mining. Visit affiliation administer mining is one key process in information outsourcing continuously information dynamic representation. Generally more number of specialized consents is accomplished to create protection safeguarding on successive thing set era crosswise over information base preparing in both vertically and on a level plane parceled information. For this need to build up a safe recurrence numbering convention, proposed convention guarantees privacy to respondent's information. Likewise need to plan convention to use for any data mining model empowered by recurrence.

1. Our work is identified with protection saving outsourced visit thing set mining arrangement. A portion of the fundamental destinations are: Design a security protecting outsourced visit thing set digging answer for level and vertically divided databases.
2. Allowing the information proprietors to outsource mining undertaking on their joint information in a security safeguarding way.
3. Building a security saving outsourced affiliation manage for parceled databases.
4. Apply Symmetric Encryption Technique to shield information proprietor's crude information from other information proprietors and the cloud.
5. Designing the utility of the proposed Hybrid outsourced conspires.
6. Deploying framework on appropriated environment.
7. Analysis with most existing arrangements, with framework concerning check spillage data about the information proprietors' crude information.

V. OUTLINE OF PROPOSED WORK

The framework model is contained at least two information proprietors and a cloud. Every information proprietor has a private database, and the information proprietors encrypt their private databases preceding outsourcing the encoded databases to the cloud. Information proprietors can likewise ask for the cloud to mine affiliation administers or regular item-sets from the joint database for their benefit. The cloud is entrusted with the ordering and putting away of databases got from various information proprietors, the mining of affiliation principles or regular item-sets for information proprietors, and the sending of the mining result to pertinent information proprietors.

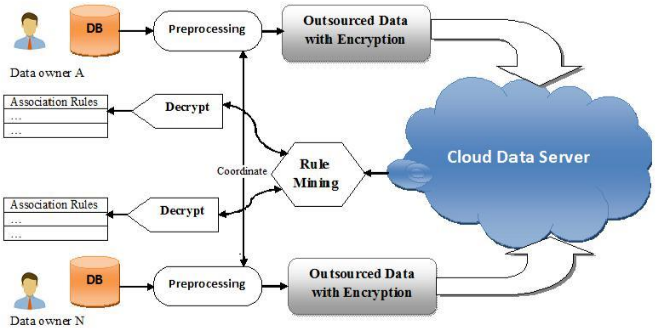

Figure 1. System Architecture

Above framework is proposed to have the following modules along with useful requirements. In this framework affiliation control mining arrangement, every information proprietor possesses a private database, and information proprietors cooperatively mine their joint database's affiliation rules with the help of the cloud. The System affiliation control mining arrangement comprises of taking after different stages.

Preprocessing Stage

In the preprocessing stage, data owners and the cloud collaborate to generate an encrypted joint database at the cloud's end and some auxiliary data for privacy-preserving mining. Each data owner inserts fictitious transactions to his private database, and encrypts items in the database with a Symmetric Encryption technique or substitution cipher. The fictitious transactions are used to mitigate frequency analysis attacks. Once the databases have been encrypted, they are outsourced to the cloud as part of the joint database maintained by the cloud. To allow the cloud to accurately mine the database (which has fictitious transactions), Data owners tag each transaction in their outsourced databases and joint database with an encrypted realness value (ERV) using our customized encryption scheme. A realness value (RV for short) is either 0 or 1, which indicates that the transaction is fictitious or real, respectively. All ERVs are sent to the cloud. Please note that the cloud is still unable to determine whether a transaction is fictitious or not, even having ERVs.

Mining Stage

In the mining stage, the cloud mines association rules for information proprietors in a protection safeguarding way. The cloud mines affiliation lead hopefuls from the scrambled joint database. Due to the presence of imaginary exchanges, a few applicants will be "false positives". To permit information proprietors distinguishing false positives, the cloud confirms competitors in a protection saving way. The cloud figures every competitor's encoded confirming result from the ERVs, using encryption and secure correlation plans. The cloud gives back all competitors and their encoded checking results to the information proprietors. At last, information proprietors decode the encoded confirming results and affiliation lead contender to recoup the genuine affiliation rules. The principle thought of our incessant itemset mining arrangement is comparable, and the main contrasts are in the mining stage. In the mining stage, the cloud mines visit itemset applicants (i.e. the apparently visit itemsets are characterized later) rather than affiliation run hopefuls. The information proprietors then decode the encoded verifying results and frequent itemset candidates to recover the real frequent itemsets.

Distributed Association Rule Mining

Distributed Association Rule Mining algorithm is used to compute confidence and support of a given candidate itemset. By considering values of the attributes finding whether a particular itemset is frequent, Considering number of records (count) where the values for all the attributes in the itemset are not null, then the candidate itemset is declared as the frequent itemset.

VI. CONCLUSION

The issues of privacy preserving association rule mining are addressed here. In particular, privacy preserving algorithms over horizontal and vertical partitioned databases are discussed. Here, we proposed model- based privacy preservation technique i.e. privacy preserving outsourced association rule mining for partitioned database (PPOARMD). In PPOARMD, we observed that the system allows data owners to mining task on their joint data in the privacy preserving manner.

This can be done by using a different rule mining method. Also we will introduce hybrid outsourced scheme which is the combination of two techniques, which provides the best way for data owners to outsource their data on partitioned databases with the less data leakage and high level of privacy without compromising performance.

VII. REFERENCES

[1]. Lichun Li, Rongxing Lu, Kim-Kwang Raymond Choo, Anwitaman Datta, and Jun Shao, Privacy Preserving-Outsourced Association Rule Mining on Vertically Partitioned

[2]. Databases IEEE Transactions on Information Forensics and

[3]. Security, Vol. 11, No. 8, August 2016.

[4]. J. Vaidya and C. Clifton, Privacy preserving association rule mining in vertically partitioned data, in Proc. SIGKDD, 2002, pp. 639–644.

[5]. B. Rozenberg and E. Gudes, Association rules mining in vertically partitioned databases, Data Knowl. Eng., vol. 59, no. 2, pp. 378–396.

[6]. S. Zhong, Privacy-preserving algorithms for distributed mining of frequent itemsets, Inf. Sci., vol. 177, no. 2, pp. 490–503.

[7]. F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, Privacy-preserving mining of association rules from outsourced transaction databases, IEEE Syst. J., vol. 7, no. 3, pp. 385–395,Sep. 2013.

[8]. W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N.

[9]. Mamoulis,Security in outsourcing of association rule mining, in Proc. VLDB,2007, pp. 111–122.

[10]. J. Lai, Y. Li, R. H. Deng, J. Weng, C. Guan, and Q. Yan, Towards semantically secure outsourcing of association rule mining on categorical data, Inf. Sci., vol. 267, pp. 267–286, May 2014.

[11]. M. Kantarcioglu and C. Clifton, Privacy-preserving distributed mining of association rules on horizontally partitioned data, IEEE transactions on knowledge and data engineering, vol. 16, no. 9, pp. 1026-1037.

[12]. O. Goldreich, Encryption schemes, working draft, (2003) March.

[13]. H. Grosskreutz, B. Lemmen and S. Rüping, Secure

[14]. Distributed Subgroup Discovery in Horizontally Partitioned Data, Transactions on Data Privacy, vol. 4 no. 3, (2011), pp. 147-165.

[15]. Can Xiang, Chunming Tang Efficient outsourcing schemes of modular exponentiations with checkability for untrusted cloud server J Ambient Intell Human Comput (2015) 6:131–139.

[16]. Xinjing Ge, Li Yan, Jianming Zhu, Wenjie Shi Privacy-Preserving Distributed Association Rule Mining Based on the Secret Sharing Technique, 2009 IEEE.

[17]. Xuan Canh Nguyen, Hoai Bac Le, Tung Anh Cao An enhanced scheme for privacy-preserving association rules mining on horizontally distributed databases 978-1-4673-0309-5/12, 2012 IEEE

[18]. Mahmoud Hussein, Ashraf El-Sisi, Nabil Ismail, Fast Cryptographic Privacy Preserving Association Rules Mining on

[19]. Distributed Homogenous DataBase, Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, Volume 5178/2008, pp. 607 -- 616 (2008).

[20]. Moez Waddey , Pascal Poncelet, Sadok Ben Yahia, Novel

[21]. Approach For Privacy Mining Of Generic Basic Association

[22]. Rules,In PAVLAD'09, November 6, 2009, Hong Kong, China, 2009 ACM.

[23]. J. Vaidya, Clifton. Privacy preserving association rule mining in vertically partitioned data In: Proceedings of the Eighth

[24]. N. V. Muthu Lakshmi1 & K. Sandhya Rani, Privacy Preserving Association Rule Mining in Vertically Partitioned Databases, In International Journal of Computer Applications (0975 – 8887) Volume 39– No.13, February 2012.

[25]. Zhu Yu- quan, Tang Yang, Chen Geng, A Privacy Preserving Algorithm for Mining Distributed Association Rules, 19-21 May 2011.

[26]. Boxiang Dong, Ruilin Liu, and Hui (Wendy) Wang Trust-but-Verify: Verifying Result

Correctness of Outsourced Frequent Itemset Mining in Data-Mining-As-a-Service Paradigm IEEE Transactions On Services Computing, Vol. 9, No. 1, January/February 2016.

[27]. Md. Golam Kaosar, Russell Paulet, Xun Yi Optimized Two Party Privacy Preserving Association Rule Mining Using Fully Homomorphic Encryption Springer 2011.

[28]. Kaingade, Rasika M., and Hemant A. Tirmare. "Personalization of Web Search based on privacy protected and auto-constructed user profile." Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE, 2015.

[29]. Parkar, Mr Vishal Vijaykumar. "Enhancing Security Audit for Web Applications with Dynamic Modeling Approach." (2016).