International Journal of Security and Its Applications
Vol. 9, No. 3 (2015), pp. 41-48
http://dx.doi.org/10.14257/ijsia.2015.9.3.05

Data Outsourcing based on Secure Association Rule Mining Processes

V. Sujatha[1], Debnath Bhattacharyya[2], P. Silpa Chaitanya[3] and Tai-hoon Kim[4]

[1]Department of Computer Science and Engineering,
Vignan's Nirula ITW Palakaluru,
Guntur, AP, India
sujiekkurthi@gmail.com
[2]Department of Computer Science and Engineering,
Vignan's Institute of Information Technology,
Duvvada, Visakhapatnam, India
debnathb@gmail.com
[3] Department of Computer Science and Engineering,
Vignan's Nirula ITW Palakaluru,
Guntur, AP, India
silpam86@gmail.com
[4] Department of Convergence Security, Sungshin Women's University,
249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea
taihoonn@daum.net
(Corresponding Author)

Abstract

Data mining is the process of extracting information from data warehousing applications. Data outsourcing is the major task in present days, for accessing services and other features of the database processing. But sometimes this process may achieve to split among various parties with recommended data items in analyzing of the data. Data security is one of the key processes in outsourcing data to various outside users. Traditionally Fast Distribution Mining algorithm was proposed for securing distributed data. This paper addresses a problem by secure association rules over partitioned data in both horizontal and vertical representation. A secure frequency developed algorithm is used for doing above process efficiently in partitioned data, which includes services of the data in outsourcing process. Frequent item sets are used to access services in outsourcing data in recent application development data mining. Our proposed work maintains efficient security over vertical and horizontal view of representation in secure mining applications. The result shows that algorithm timing is desirable for big size data for security considerations using association rule mining operations in real time application development.

Keywords: *Association Rules, Cryptography, Distributed Database, Distributed Computation, Privacy Preserving Data Mining*

1. Introduction

Data mining is an interdisciplinary subfield in computer applications with computational process of discovering patterns in large data sets involving methods at interaction of some

ISSN: 1738-9976 IJSIA

machine learning and database system process applications. Data mining technology is an emerging process of identifying patterns from large quantities of data with relevant features in semantic data. Mostly used scenario in database applications is distributed data mining as shown in Figure 1, in this mining process can perform operations in distributed process based on sharing users over partition of data.

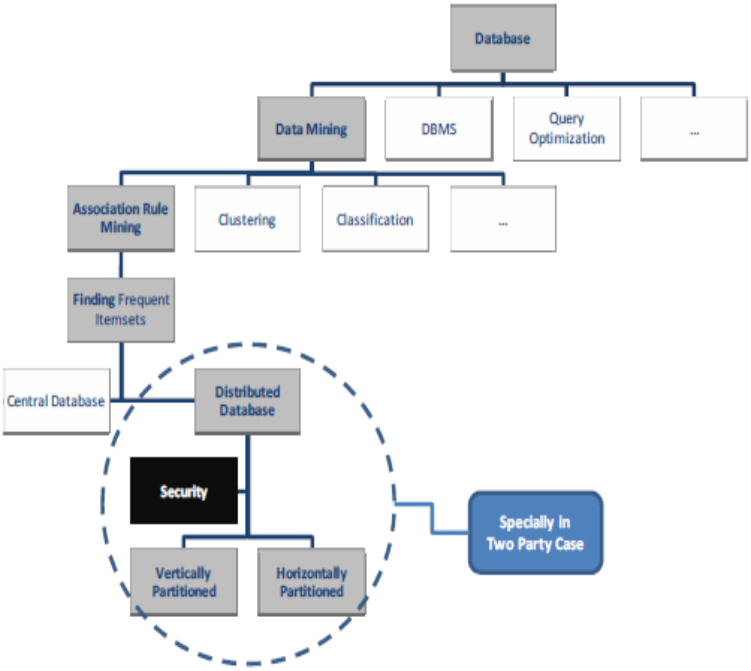

Figure 1. Data Processing over Distributed Data Mining

In data mining applications, association rule mining aims to discover item sets with co-occurrence of frequently performed transactions on database. In this paper, the problem on secure association rule mining in both horizontal and vertical database process, the process of association rule mining will calculate minimum support 's' and minimum confidence 'c' that hold in unified data sets, while minimizing the information disclosed about the private databases held by those users in process applications [2]. Privacy concerns are the major important scenario in recent approaches because each party may not want to reveal in their own partitioned database that exists relative data efficiency in commercial services. Using some protocol hierarchy in data out sourcing there is a relation of processing the data into other episodes to users present in the progressive manner. Some standard association algorithms were used to describe the identification of frequent item sets in real time progression [8]. Based on the frequent item set representation of each data outsourcing between multi parties in distributed data mining operations, Vertical data out sourcing is the major process in recent application development for providing security our data sources. This provides less security which constitutes reliable data out sourcing, the solution for this aspect is we provide security both horizontal and vertical data outsourcing for more security in frequent item set representation in real time calculations [1]. Traditional developed algorithms provide security based on data process which considers the relative data events with services in data outsourcing. So in this paper we propose to develop efficient secure frequency counting algorithm for both vertical and horizontal representation of partitioning data. The main focusing term of proposed algorithm gives sufficient neutrality for privacy

preserving mining of both vertical and horizontal representation of partitioned data. This algorithm should be very efficient than the existing solutions. In addition, we want to design algorithm that works not only for three parties and above but also for multi-party accessing.

Remaining of this paper proposes efficient data security religions as follows: Section 2 defines existing application progression with features development in relative data outsourcing and describes advantages and disadvantages. Section 3 defines back ground work for the proposed approach in data outsourcing operations with secure association rule mining operations. Section 4 defines proposed approach with consecutive terminology with technique development and Section 5 defines experimental result specification in developed application process on data outsourcing. Section 6 defines conclusion process of the overall paper experiments.

2. Related Work

Consider the data mining knowledge of the each research user specification there is a relative data outsourcing solutions were introduced by each research user. Cliften and his students worked on privacy preserving distributed data mining of frequent item sets [6], they are using some complex terminology in association rule mining algorithm specifications, these user specifications are accessed services with relevant data features. Vaidya and Cliften gave algebraic solution for secure association rule mining operations in secure data outsourcing in vertical way process. After some years in recent working technology there is a process that includes data security in linear of each participant. And then Goldriech was proposed secure combination for highly amount of data out sourcing, these considerations are accessed services in distributed data mining [5]. Privacy preserving distributed data mining first addresses secure association process on secure data outsourcing [11]. The state-of art additetively homomorphic encryption schemas used for securing data outsourcing but these techniques were not accessed data sourcing are slower process than multiparty control accessing services, but these solutions are represented with relative data accuracy for providing security in both vertical and horizontal data assurance. These algorithms were not presented solutions with nearest data processing utilities in secure association rule mining operations.

3. Background Work

Due to the process of secure mining of association rule mining in both vertical and horizontal data base access with homogeneous and heterogeneous databases [7]. For secure mining association in distributed data mining operations Fast Distribution algorithm was proposed for during above process efficiently. This algorithm is an unsecured version of the Apriori algorithm, adding some additional protocols to this process then FDA algorithm may perform efficient operations on each contributed data efficiently [4]. This protocol increases efficiency in providing security in real time application development processes. This algorithm does not depend on the view of processing encryption and decryption techniques for privacy. The main idea of this algorithm achieves s-frequent item set must be locally s-frequent in at least one of the developed site. Then in order to find all the globally s-frequent item sets each user represents locally s- frequent item sets.

Step 1: Initialization of each data item, with proposed process of calculation of item sets which includes s- frequent item sets.
Step 2: **User set generation:** Each user computes the set of frequent item sets that are locally and globally frequent item sets
Step 3: **Local Pruning:** For each data item x computes the relation of item sets

that locally s-frequent operations.
Step 4: **Unifying user item set:** Each user broadcasts his C*k,m s* and then all players compute C*k* ∪ *s* := M*m* =1 C*k,m*.
Step 5: **Compute local supports:** All players compute the local supports of all item sets in C*k s*.
Step 6: **Broadcast Mining Results:** Each player broadcasts the local supports that he computed. From that, everyone can compute the global support of every item set in C*k s*.

Algorithm 1. Fast Distribution Secure Mining

The above algorithm violates security in two stages: First where the users broadcast the item sets that are locally frequent in their private data bases, and where the broadcast sizes of local supports and then other feasibilities with user possibilities in item set selection [10]. In the first iteration when k= 1 then the item set computes the process in overall frequent data item sets Fast distribution algorithm finds all the item sets that globally s-frequent and then find all the item sets locally with s-frequent item set representation in data outsourcing applications process. The length of item set is K, then in the $(K+1)^{th}$ iteration of the FDM it will find no (K + 1) item sets that are globally s-frequent, in which case it terminates [12]. The main process of the prescribed events in data sourcing is possible for horizontal distributed data bases.

4. Proposed Methodology

Data mining achieves machine learning and statistics in supporting data outsourcing in prom antic data combinations. Many organizations stores warehousing operations with relative data representation and stores data amounts in the terabyte range and high speed computing to other sources present in data mining application development.

4.1. Processing State

Association rule mining process constitutes the relation of each item set that contain each process in required format. Processing work is proposed using services of data outsourcing in real time products [2, 3]. For taking the transactional data base with followed support threshold value s and confidence threshold value c using the item set transaction representation, the final representation. Using these resources of the transactional data base we perform association rule mining operations based on the threshold value of transactional data set with representation support and confidence as percentages in transactional data base. Then apply association rules, for example if a rule is supported globally and locally on data items, this particular rule was achieved data item in real time progression of processing item set.

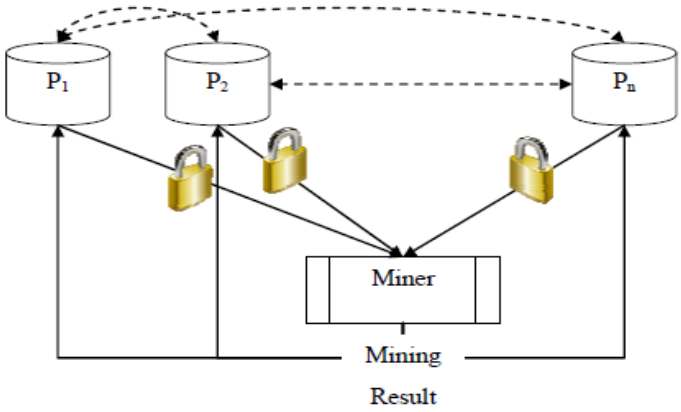

Figure 2. Process State for Executing Real Time Progression

The Figure 2 shows transactional data processes are organized in partitioned data in both vertically and horizontally [2]. Vertical partition means columns will be arrange in columns of matrix data. The development of our methodology will achieve the properties of the following protocols like secure distributed frequency counting protocol, secure association rule mining in vertical data partitioning and horizontal partitioning data in real time processing data items.

4.2. Secure Association Rule Mining Over Vertical Data Partitioned Data Bases

Vertical association rules in partitioned data items based on support count of the item set representation. This vertical partitioning can be developed using the following example the are some data sets from hospital and then some data sets from super market but there is a relation from people verification from two data sets [6, 9]. By taking good association rule mining in their prescribed data, for example we will find a rule {beef meat, sugar} =>{Diabetes} that means most people who consume beef, meat, sugar suffer diabetes in this case we have vertical partitioned data Because each site's dataset is different with others, but they have a relational field that join their data together.

4.3. Secure Association Rule Mining Over Horizontal Data Partitioned Database

In horizontal distributed data sets, all the transactions distributing among number of item sets. In that we are calculating global item set is equal to sum of local item sets. An itemset X is globally supported if the global support count of X is bigger than s% of the total transaction database size. A k-itemset is called a globally large k-itemset if it is globally supported [8]. In this way we proceed to develop efficient progression in commercial data set representation in data out sourcing.

5. Results and Analysis

In this section we proposed to develop a simple cryptographic method that allows data to compute frequencies of values in respondent's data. The proposed contains cryptographic data item sets like protocol that assures cryptographic methodology which designed secure frequency counting data items. We assume the scenario of the large number of components which establish the services [2, 3]. Consider the example with followed results that achieves as follows: The scenario consists n respondents U1,…..Un, each respondent has a value a, this

value is special frequent data item in whole data set. Then sum of all the data items $a=\sum_{i=1}^{n} ai$, without relieving each a_i.

5.1. Secure Frequency Counting Protocol

This protocol based on homomorphism property in ElGamal encryption ensures privacy of the protocol. This process consist a group G with representative in developed approach.

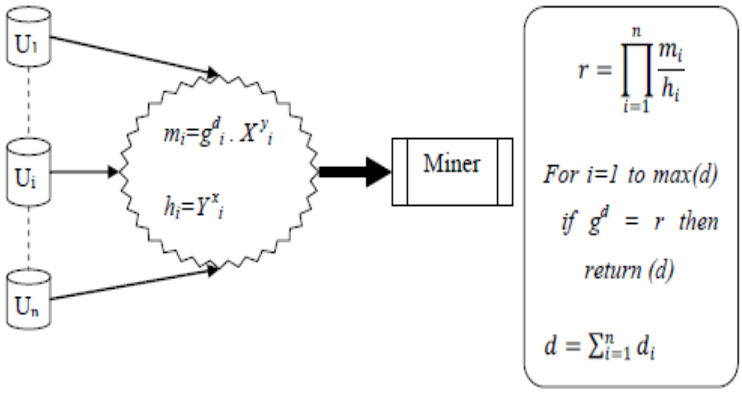

Figure 3. Secure Frequency Counting Architecture for Processing

As shown in the Figure 4 each respondent holds the value of a_i and miner with total number of values. First each respondent U_i try to encrypt itself value a_i using private key assurance like x_i,y_i that shared public key X,Y process in ElGamal cryptographic system. Then the miner will use *r* for decrypt that particular file in recommended format. All the cryptographic operations use open SSL libraries. These library simulations are used to perform security operations, for example we choose 512 bits as the length of the cryptographic system then our developed algorithm hierarchy affects and divide those materials into number of respondents then perform our protocol operations in each regular format for specified religions. In our developed application consists time comparisons in main factors that reflect in religion factors in data outsouring. Therefore we ignore all the time results based on the factors presented based on respondents. In this experiment, the respondents then the range of a value can be indicated with equal number of respondents.

We use 5, 10, 20, 50 and 100 respondents in our experiment and the earned time is based on the average of five protocol runs. As you can see, the time offers a linear behavior related to number of respondents. The total time of protocol is shown in Figure 4. For example, miner computation takes 498 milliseconds for 100 respondents.

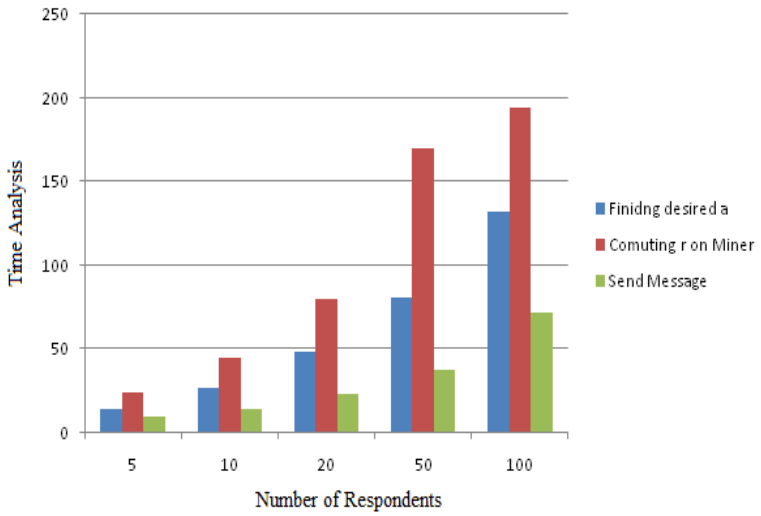

Figure 4. Total Time of Frequency Counting Protocol Related to Number of Respondents

6. Conclusion

Frequent association rule mining is one key process in data outsourcing in real time data progressive representation. Traditionally more number of technical permissions is achieved to develop privacy preserving on frequent item set generation across data base processing in both vertically and horizontally partitioned data. In this we propose to develop a secure frequency counting protocol, our proposed protocol ensures confidentiality to respondent's data. Our proposed protocol used for any data mining model enabled by frequency computation for both theoretical analysis and proof in addition to experimental results show that the protocol is very efficient and runs in desirable time.

References

[1] T. Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 4, **(2014)** April, pp. 970-983.
[2] A. R. Kenari and M. N. M. Sap, "Cryptographic Technique for Association Rule Mining in Multi Party Databases", 20th Annual ACM Symposium on the Theory of Computing, **(2009)**, pp. 11–19.
[3] W. W. Kit, D. W. Cheung, E. Hung, B. Kao and N. Mamoulis, "Security in outsourcing of association rule mining", 33rd international conference on Very large data bases, VLDB Endowment, **(2007)**, pp. 111-122.
[4] 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, **(2004)**, pp. 1026-1037.
[5] I. Ioannidis and A. Grama, "An efficient protocol for Yao's millionaires' problem", 36th Annual Hawaii International Conference on System Sciences, **(2003)** January 6-9, pp. 6.
[6] O. Goldreich, "Encryption schemes", working draft, **(2003)** March.
[7] E. Mykletun, M. Narasimha and G. Tsudik, "Authentication and integrity in outsourced databases", ACM Transactions on Storage (TOS), vol. 2, no. 2, **(2006)**, pp. 107-138.
[8] H. Grosskreutz, B. Lemmen and S. Rüping, "Secure Distributed Subgroup Discovery in Horizontally Partitioned Data", Transactions on Data Privacy, vol. 4 no. 3, **(2011)**, pp. 147-165.
[9] T. Tassa and D. J. Cohen, "Anonymization of Centralized and Distributed Social Networks by Sequential Clustering", IEEE Transactions on Knowledge and Data Engineering, vol. 25, Issue 2, **(2013)**, pp. 311-324.
[10] T. Tassa and E. Gudes, "Secure distributed computation of anonymized views of shared databases", ACM Transactions on Database Systems (TODS), vol. 37, no. 2, **(2012)**, pp. 11.

[11] K. S. Rao, V. N. Mandhala, D. Bhattacharyya and T.-H. Kim, "An association rule hiding algorithm for privacy preserving data mining", International Journal of Control and Automation, IJCA, vol. 7, no. 10, **(2014)** October, pp. 393-404.
[12] A. Ben-David, N. Nisan and B. Pinka, "FairplayMP: a system for secure multi-party computation", Proceedings of the 15th ACM conference on Computer and communications security, New York, NY, USA, **(2008)**, pp. 257-266.

