IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

Providing security for Association Rules Mining in Distributed Databases Result analysis

Pankaj P. Joshi[1], Prof. R. M. Goudar[2]

Student, Computer Department, MITAOE, Alandi, Pune, India [1]

Asst. Prof., Computer Department, MITAOE, Alandi, Pune, India [2]

Abstract: Quick change in the field of information technology raises the significant issue about gigantic data storage. To build up the connection among set of data items, likewise gets to be troublesome test. To defeat these issues in large data, association rule mining does a vital part. As of late, the vast majority of the analyst's purpose of intrigue is to increase the effectiveness of association rule mining based algorithm and accordingly expanding the speed of mining. In existing framework, association rules are applied on horizontal partitioned database and such strategy is relying on upon IDMA, EMADS communication. In this paper, we consider both horizontal and vertical partitioned database, while applying association rule mining framework likewise attempting to minimize the communications overhead. Further to upgrade the security of framework, RC4 algorithm is utilized to secure the horizontal and vertical partitions of database. At last framework uses the protocol for perceiving the fake and duplicate frequent rules.

Keywords: Data mining, Apriori, Distributed databases, Frequent item set, association rules.

I. INTRODUCTION

Generating the knowledge from the data, which is embedded in the database, is one of the main task of data mining technology. Data mining procedure extract the information from the frequently growing the quantity of data and the transferring this data into useful knowledge. Therefore, this becomes core point of attraction of many researchers. The researchers have developed number of technique, for analysing the increasing volume of huge data. One of the most important types of data mining technique named as Association rule mining was introduced in 1993 by Agrawal. The term association rules mining refers to find out the correlation among large set of data items. Association rule mining plays a very important role in decision making algorithms. One of the characteristic instances of the association rule mining is the market base analysis.

Association rule mining performs the two major operations, which are listed here: 1) frequent item generation: This is the item set which satisfied the minimum thresholdsupport. 2) Rule generation: this is also called as the strongrules, since these rules have thehigh confidence rules which were found in the initial step of the association rule.

Database can be partitioned either horizontally or vertically.Security becomes majorconcern issue during data mining or partitioning. Here we focus on the issue occursduring the secure mining of association rules in the horizontal and vertical partitioneddatabase. There are numerous websites are available, who handles the uniform database, which share the same information but hold the data on the distinct schemas. The focus is to find out the association rules with the value of support and confidence level. Our focus is not only to secure the information from the entity transaction in distinct databases, but also provide more universal information like the association ruleswhich are supported in each of that database.

II. RELATED WORK

In [1], proposed an optimized distributed association rule mining algorithm for thebiologically dispersed data which is used in parallel and dispersed environment for reducingthe communication cost. In this approach the multiple nesting problems in XMLdata is handled appropriately for assuring the correctness of the results. The algorithmis used for the mining process in a parallel and distributed environment.

In [2], authors tracing out the recently trends in parallel and distributed Apriori algorithm.In this paper author review distinct parallel and dispersed association rule miningwhich are developed based on the Apriori algorithm.
In this paper discussed 10 distinctalgorithms of association rule mining which are dynamic hashing and pruning, frequentitem set mining, hybrid distribution, intelligent data distribution, partition andnon-partitioned, simply partitioned, and hashed partitioned Apriori, DMM and FDM.

In [3], identify and solve the issues of mining association rules on shared nothingmultiprocessor. Author presents three algorithms which discover a spectrum of trade-offsamong computation, communication, memory usage, synchronization, and the useof problem specific information.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

In [4], detailed use of association rule mining for extracting the patterns is explained,which occurred regularly within a dataset. Also, shows the execution of the Apriori algorithmfor mining association as rules from the dataset which contains the crime dataconcerning women.

The paper [5], represent a well-organized algorithm which produce all importantassociation rules among items in the database. The algorithm incorporates buffer managementand novel estimation and pruning techniques.

In [6], Matthias Klusch, Stefano Lodi, Gianluca Moro discussed one key fact aboutdeveloping the huge amount of independent and varied data sources in the Internet isnot only how to recover, gather and put together applicable information but to determinepreviously unknown, implied, and precious knowledge. In recent years, numerous approachesto dispersed data mining and knowledge discovery have been developed, butonly a few of them make use of intelligent agents.

In [7], Stolfo, Prodromidis, et al illustrates the JAM system, a dispersed, scalable,and moveable agent based data mining system which employs a universal approach forscaling data mining applications that called meta learning. JAM offers a set of learningprograms, executed either as JAVA applets or applications that calculate models overdata stored locally at a site. JAM also offered a set of meta-learning agents for mergemultiple models that were learned at distinct sites.

III.SYSTEM IMPLEMENTATION

A. Proposed System

Initially the distributed database i.e. the database from the distinct sites is send to thecentralized server. We have denoted the dispersed database as database 1, database 2and database 3.

- After this centralized server collects all the vertically partitioned dataset given bythe dispersed database.
- Database is partitioned in two ways horizontally and vertically partitioned data.
- After partitioning the database, encryption algorithm is used to encrypt each partition.The partitioned data are encrypted for providing the security over data.
- After providing the security to the partitioned data, encrypted data and minimumsupport value send to the distributed agent.
- Distributed agent receives the encrypted data and decrypts it.
- After decrypting the encrypted data, associated rules are generated by using theApriori algorithm. After generating the rules, distributed agent send these rules tothe centralized server.
- Centralized server receives result from all the agents, after receiving rules from the entire agents, centralized server remove duplicate and fake rules from the server.
- Finally, the global list is generated by the centralized server.

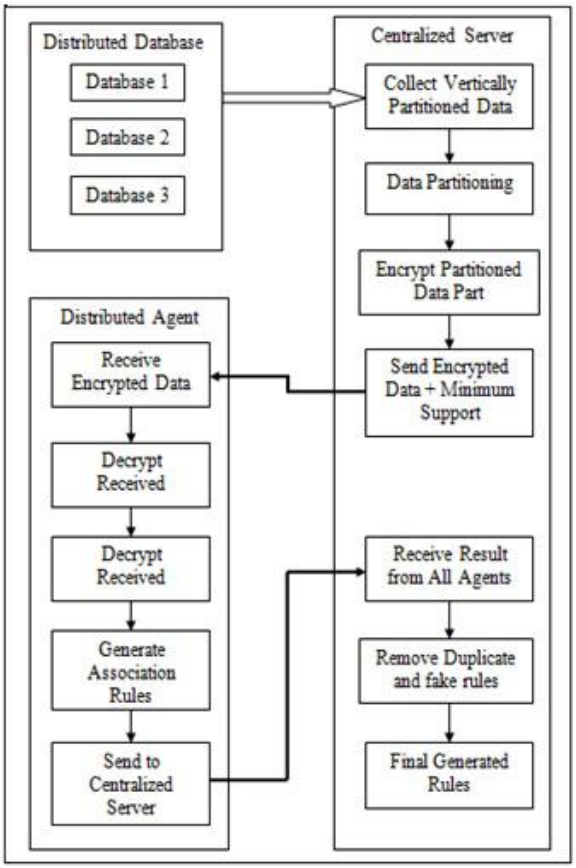

Fig. 1 System Architecture

B. Algorithm

Algorithm 1: Unifying lists of locally frequent item set Algorithm

1. Vertically partitioning of databases.
2. For centralized Server Gather partitioned data from database.
3. after combine data partitioned into no of parts.
4. for each agent
5. If(distributed agent is free)
6. Encrypt each part of data to send distributed processor agent.
7. End If;
8. End For
9. Send encrypts data to agent and minimum support value.
10. at each agent
11. Decrypt received data
12. Generate association rules base on support and confidence.
13.Send generated rules above the minimum support to the centralized server.
14.Merge all generated rules from different distributed agents.
15.Remove Duplicate and fake rules
16.Generate Final association rules

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 2, February 2017

C. Mathematical Equations

Encryption method for RC4:

For each message byte Mi

i = (i + 1) (mod 256)

j = (j + S[i]) (mod 256)

Swap(S[i], S[j])

t = (S[i] + S[j]) (mod 256)

Ci = Mi XOR S[t]

Equation for generating the Confidence value:

$$\text{Confidence} = \frac{\text{Support } (X \cup Y)}{\text{Supp } (X)}$$

Where x and y item sets.

Equation for generating the support value:

$$\text{Support} = \frac{\text{Numberofoccurenceofx}}{\text{Totalnumberoftransactions}}$$

Where x is the item sets.

IV. RESULT AND DISCUSSION

A. Graphs

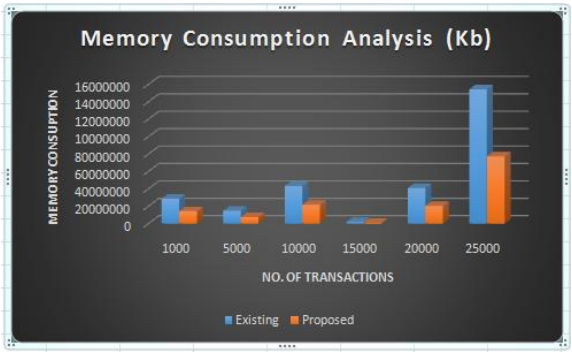

Fig: 2 Memory consumption comparison, Threshold 0.01

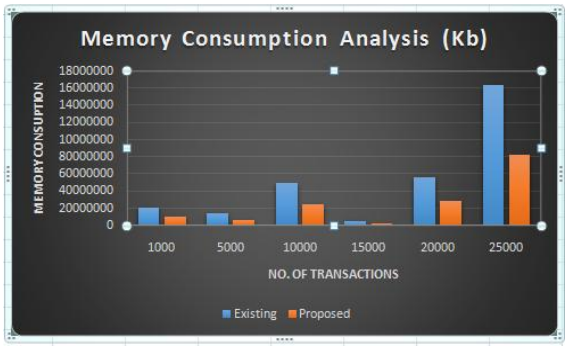

Fig:3 Memory consumption comparison, Threshold 0.02

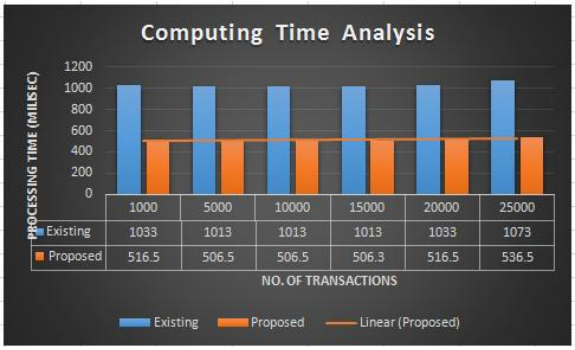

Fig: 4 Computing time comparison, Threshold 0.01

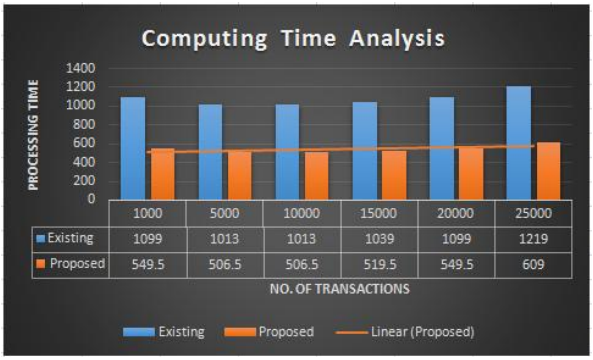

Fig:5 Computing time comparison,, Threshold 0.02

Above graphs show comparison of computing time and memory consumption comparison for reference values of threshold value between the existing and proposed system. The propose system consume less time than the existing system and increase the performance of a system.

V. CONCLUSION

Here proposed the technique for secure mining of associations rule in both horizontaland vertical partitioned database.We explain about the basic overview of dispersedassociation rule mining and, we discuss about the framework of agent based disperseddata mining. Here, the architecture for mobile agent based dispersed associationrule mining is designed. The architecture is used for dropping the communication overhead. This architecture ensures the security for the mobile agent. We also provide security to the proposed system by using the RC4 algorithm. By using this algorithm, the fake rules and duplicate rules generated in the framework have removed.

REFERENCES

[1] Dr (Mrs).Sujni Paul, "An Optimized Distributed Association Rule Mining Algorithmin Parallel and Distributed Data Mining With XML Data for Improved ResponseTime.", International Journal of Computer Science and Information Technology,Volume 2, Number 2, April 2010.

[2] T. AnuRadha, P. Lavanya, "Recent Trends in Parallel and Distributed Apriori Algorithm", International Journal of Engineering Research and Applications, Vol. 1, Issue 4, pp.1820-1822

[3] R. Agrawal and J. Shefer, "Parallel Mining of Association Rules", IEEE transaction6 December 1996

[4] Divya Bansal, LekhaBhambhu, "Execution of APRIORI Algorithm of Data MiningDirected Towards Tumultuous Crimes Concerning Women", International Journalof Advanced Research in Computer Science and Software Engineering.

[5] R. Agrawal, T. Imielinski, and A. Swami, "Mining Associations between Sets ofItems in Massive Databases," Proceedings of the ACM SIGMOD, Washington DC, 1993.

[6] Matthias Klusch, Stefano Lodi, Gianluca Moro, "Agent-Based Distributed DataMining: The KDEC Scheme".

[7] Stolfo, Prodromidis, et al "JAM: Java agents for meta-learning over distributeddatabases", in 1997.