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. 4, Issue 5, May 2016

Secure Fast Distribution Mining of Association Rules in Horizontal Database

Zameena.R

M.Tech Student, Dept. of CSE, Marian Engineering College, Trivandrum, Kerala, India

ABSTRACT Privacy in data mining prevents the parties from directly sharing the data, and some types of information about the data. We introduce a protocol for secure mining of association rules in horizontally distributed databases using FDM, which an unsecured distributed version of the Apriori algorithm. Distributed Associated rule based algorithm produce some relationships between locally and globally large item set using FDM, which generates a small number of candidate sets and substantially reduces the number of messages to be passed to customers. We extended our work with Horizontal Aggregation that give rise to multiple row output, which transform rows to column using CASE, SPJ or PIVOT operators depending on the input. We introduced security and privacy through cryptography and level based slicing. Level Based Slicing introduce different techniques such as Generalized Data, Bucketized Data, Multiset-based Generalization Data, One-attribute-per-Column Slicing Data, Sliced Data .Our result shows that proposed method have higher performance other sequential algorithms**.**

KEYWORDS: privacy preserving data mining, Advanced Encryption Standard, Association Rules, Level Based Slicing, Horizontal Aggregation

I. INTRODUCTION

Distributed computing plays an important role in the Data Mining process for several reasons. First, Data Mining often requires huge amounts of resources in storage space and computation time. To make systems scalable, it is important to develop mechanisms that distribute the work load among several sites in a flexible way. Second, data is often inherently distributed into several databases, making a centralized processing of this data very inefficient and prone to security risks

This paper studies the problem of association rules mining in horizontally distributed databases. In the distributed databases, there are several players that hold homogeneous databases which share the same schema but hold information on different entities. The goal is to find all association rules with support s and confidence to minimize the information disclosed about the private databases held by those players. The information that we would like to protect in this context is not only individual transactions in the different databases, but also more global information such as what association rules are supported locally in each of those databases [1].

Kantarcioglu and Clifton studied the problem where more suitable security definitions that allow parties to choose their desired level of security are needed, to allow effective solutions that maintain the desired security [2].The main part of that protocol is sub protocol for secure computation of the union of private subsets that are held by the different players. It makes the protocol costly and its implementation depends upon encryption primitive's methods, and also the leakage of information renders the protocol not perfectly secure [1]. This paper solved the difficulty by using Cryptographic Technology that provides Hash Function for Encryption and Decryption. Privacy can be implemented through Level Based Slicing. Slicing preserves better data utility than generalization and also prevents membership disclosure.

II. RELATED WORK

The paper [1] studied the problem of secure mining of association rules in horizontally partitioned databases. Tamir Tassa proposed here a protocol Fast Distributed Mining algorithm (FDM) for mining of association rules in horizontally distributed databases. The main idea is that the player's finds their locally s-frequent itemsets then the players check each of them to find out globally s-frequent item set. Paper assumes that the players are semi

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. 4, Issue 5, May 2016

honest; information. Hence the player compute the encryption of their private database together by applying commutative encryption .Paper shows that their protocol offers better privacy and is significantly more efficient in terms of communication cost and computational cost while the solution is still not perfectly secure cause it leaks excess information [1].

The paper [3] deals with the problems of association rule mining. The problems can be divided as data hiding and knowledge hiding. Data hiding is defined as the trial of removing confidential or private information from the data before its disclosure. Knowledge hiding, on the other hand, concerns the information, or else the knowledge, that a data mining method may discover after having analyzed the data. This paper reviews the methods of privacy preserving and proposed an improvement of sensitive rule hiding to make it more accurate and secured. The secure multiparty computation (SMC) is used to find global support and confidence without data leakage. To provide privacy to the database Tiny Encryption Algorithm (TEA) is used

The paper [4] and [5] also uses Apriori algorithm for generating association rules and player cipher technique is used to transfer that generated rules. This paper defines two parts of association rule; Antecedent is the item found in database and consequent, found in combination with the first.

The paper [6] Proposed data anonymization technique called slicing to improve the current state of threat. Slicing partitions the data set both vertically and horizontally. Vertical partitioning is done by grouping attributes into columns based on the correlations among the attributes. Each column contains a subset of attributes that are highly correlated. Horizontal partitioning is done by grouping tuples into buckets. Finally, within each bucket, values in each column are randomly permutated (or sorted) to break the linking between different columns

ADVANTAGES

- The proposed system ensures the anonymity requirement.
- Improves the accuracy of the system and performance.
- Trade off between column generalization and tuple partitioning is implemented effectively

III. PROPOSED ALGORITHM

A. *Design Considerations:*

- Two users are present
  Admin
  Client
- File creation of particular request is stored in Database.
- MD5, Hash function is used for Key Generation.
- AES Algorithm is used for Encryption and Decryption of client and server.
- FDM Algorithm is used For Association Rule Generation.
- PIVOT,CASE,SPJ operators used for Horizontal Aggregation

B. *Description of the Proposed Algorithm:*

Enhanced FDM Algorithm. This includes FDM with CASE, PIVOT, SPJ operators.

The proposed algorithm is consists of three main steps.

Step 1: Cryptographic Primitive Selection:

-Player selects hash function to apply on all itemsets prior to encryption

- Player compute lookup table with hash values to find preimage of given hash values.

Step 2: All itemsets Encryption

Step 3: Itemset Merging

-Each odd player sends his encrypted set to player 1.

- Each even player sends his encrypted set to player 2.

- Player 1 unifies all sets that were sent by the odd players and removes duplicates

Step 4: Horizontal Aggregation

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. 4, Issue 5, May 2016

IV. PSEUDO CODE

Step 1: Based on the request Global Itemset is generated.
Step 2: Firstly Fs set of all s-frequent are calculated.
Step 3: Then calculate non-frequent global itemset for all transaction.
Step 4: IF S is an element of C, then delete non frequent Item set.
Table. Add
Step 6: Table.GetTransaction,k=3.
Step 7: For all transaction t is an element of Table.
Step 8: While C! =null
For all k subset s of t
k++;
Step 9: Generate Candidate Item set k+1 pass
Step 10: End.

V. SIMULATION RESULTS

Based on our studies security in implemented through advanced encryption standard AES. Hash functions are a fundamental elementary in the field of cryptography. A hash function is an efficient function mapping binary strings of arbitrary length to binary strings of fixed length (e.g. 256 bits), called the hash-value or digest. A hash function H is an algorithm it takes as input an arbitrary length message M and a fixed-length key K , and makes a fixed-length output D called the message digest. One of the hash function isMD5 (Message Digest5), it is used cryptographic hash function with a 256-bit hash value. It processes a variable-size message into a fixed-length output of 256 bits. The more popular and widely adopted symmetric encryption algorithm is Advanced Encryption Standard (AES). It is found at least six times faster than triple DES. AES is widely adopted and supported in both hardware and software. AES has built-in flexibility of key length, which allows a degree of 'future-proofing' against progress in the ability to perform exhaustive key searches. In AES the input message is divided into chunks of 512-bitblocks; then the message is padded for making its length divisible by 512. In this sender use the public key of the receiver to encrypt the message and receiver use its private key to decrypt the message. In this experimental performance analysis of the given algorithms on the basis of the following parameters on local system at different input size. In this section describes the experimental parameters, platforms and key management of experimental algorithms.

a) Evaluation Parameters.

1. Encryption Time: The encryption time considered the time that an encryption algorithm takes to produces a cipher text from a plain text.

2. Decryption Time: The decryption time considered the time that a decryption algorithm takes to produces a plain text from a cipher text.

b) Evaluation Platforms Performance of encryption algorithm is evaluated considering the following system configuration.
1. Software Speciation: Experimental evaluation Visual Studio 2010, Windows 8.1 Pro 64 bit Operating System.
2. Hardware Speciation: All the algorithms are tested on Intel Core i5 (2.40 GHz) fourth generation processor with 4GB of RAM with 1 TB-HDD.

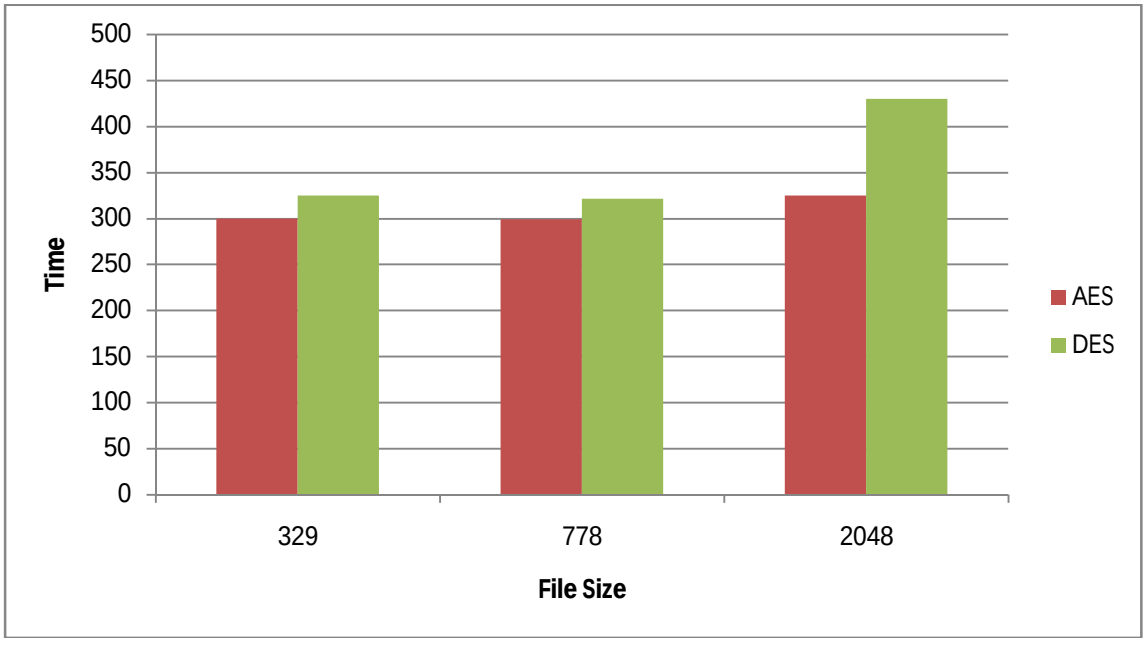
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Sly No:	Algorithm	Key size (bit)	File Size (bytes)	Encryption time (milliseconds)	Decryption time (milliseconds)
1	AES	256	329	300	310
			778	299	304
			2048	325	330
2	DES	56	329	325	300
			778	321	308
			2048	430	502



VI. Conclusion and Future Work

We proposed a protocol for to reduce data leakage in horizontally distributed database using association rules that improves significantly upon current leading protocol in terms of privacy and efficiency. One of the main ingredients in our proposed protocol is novel secure multi party protocol for computing the union of private subsets

In this paper, we use the horizontal aggregation methods CASE, PIVOT and SPJ with direct and indirect method. As we perform CASE method directly from fact table, then it makes code more compact but it is time consuming process comparing with indirect CASE evaluation where, horizontal aggregation is computed after vertical aggregation. Query optimization is most challenging task in the horizontal aggregation. We can try to achieve better query optimization. We can also use the horizontal aggregation for the further extended for Association Rules by applying Apriori Algorithm.

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. 4, Issue 5, May 2016

REFERENCES

1. Tamir Tassa,"Secure mining of association rule in horizontally distributed databases", IEEE Trans. Knowledge and Data Engg, Vol. 26, no.2, April 2014
2. M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, Sept. 2004.
3. Prof. Geetika. Narang, Anjum Shaikh, Arti Sonawane, Kanchan Shegar, Madhuri Andhale," Preservation Of Privacy in Mining Using Association Rule Technique", International Journal of Scientific & Technology Research, Volume 2, Issue 3, March 2013.
4. Meera Treesa Mathews, Manju E.V," Extended Distributed RKSecure Sum Protocol in Apriori Algorithm For Privacy Preserving", International Journal of Engineering and Advanced Technology (IJEAT), Volume-3, Issue-4, April 2014
5. P.Jagannadha Varma, Amruthaseshadri,.M. Priyanka, M.Ajay Kumar, B.L.Bharadwaj Varma, " Association Rule Mining with Security Based on Playfair Cipher Technique" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014.
6. Tiancheng Li, Ninghui Li, Senior Member, IEEE, Jia Zhang, Member, IEEE, and Ian Molloy "Slicing: A New Approach for Privacy Preserving Data Publishing" Proc. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3 MARCH 2012.
7. Carlos Ordonez, Zhibo Chen. "Horizontal Aggregations in SQL to Prepare Data Sets for Data Mining Analysis," IEEE Transactions on Knowledge and Data Engineering, Digital Object Identifier 10.1109/TKDE.2011.16, April 2012.
8. Han, J., Pei, J., Yin, Y. and Mao, R. (2004) ―Mining frequent pattern without candidate generation: a frequent pattern tree approach,‖ *Data Mining and Knowledge Discovery*, Vol. 8, No. 1, pp.53-87.

BIOGRAPHY

Zameena.R is an M.Tech Student, Dept. of CSE, Marian Engineering College Trivandrum, and Under Kerala University. During the academic Year 2014-2016.This Paper work is my Thesis.