ISSN:2320-0790

An International Journal of Advanced Computer Technology

Association Rules in Horizontally Distributed Databases with Enhanced Secure Mining

[1]Sonal Patil, [2]Harshad S Patil
[1]Asst. Prof, GHRIEM, Jalgaon
[2]Research Scholar, GHRIEM, Jalgaon

Abstract: Recent developments in information technology have made possible the collection and analysis of millions of transactions containing personal data. These data include shopping habits, criminal records, medical histories and credit records among others. In the term of distributed database, distributed database is a database in which storage devices are not all attached to a common processing unit such as the CPU controlled by a distributed database management system (together sometimes called a distributed database system). It may be stored in multiple computers located in the same physical location or may be dispersed over a network of interconnected computers. A protocol has been proposed for secure mining of association rules in horizontally distributed databases. This protocol is optimized than the Fast Distributed Mining (FDM) algorithm which is an unsecured distributed version of the Apriori algorithm. The main purpose of this protocol is to remove the problem of mining generalized association rules that affects the existing system. This protocol offers more enhanced privacy with respect to previous protocols. In addition it is simpler and is optimized in terms of communication rounds, communication cost and computational cost than other protocols.

Keyword: Advanced Encryption Standard (ASE), Secured Searching of Valuable Data in Database, Association Rule, Apriori algorithm.

I. INTRODUCTION

We are study here the problem of secure mining of association rules in horizontally partitioned databases. In that there are several places, several parties and several player that hold homogeneous databases, i.e., databases that share the same schema but hold information on different entities [1][5]. The goal is to minimizing 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. In our problem, the inputs are the partial databases, and the required output is the list of association rules that hold in the unified database with support and confidence no smaller than the given thresholds *s* and *c*, respectively [1]. As the above mentioned generic solutions rely upon a description of the function *f* as a Boolean circuit, they can be applied only to small inputs and functions which are realizable by simple circuits. In more complex settings, such as ours, other methods are required for carrying out this computation. In such cases, some relaxations of the notion of perfect security might be inevitable when looking for practical protocols, provided that the excess information is deemed benign (see examples of such protocols in e.g. [3], [4], [5], [6], [7]).

II. SYSTEM ARCHITECTURE

The below diagram shows how the protocol is work and following are phases:

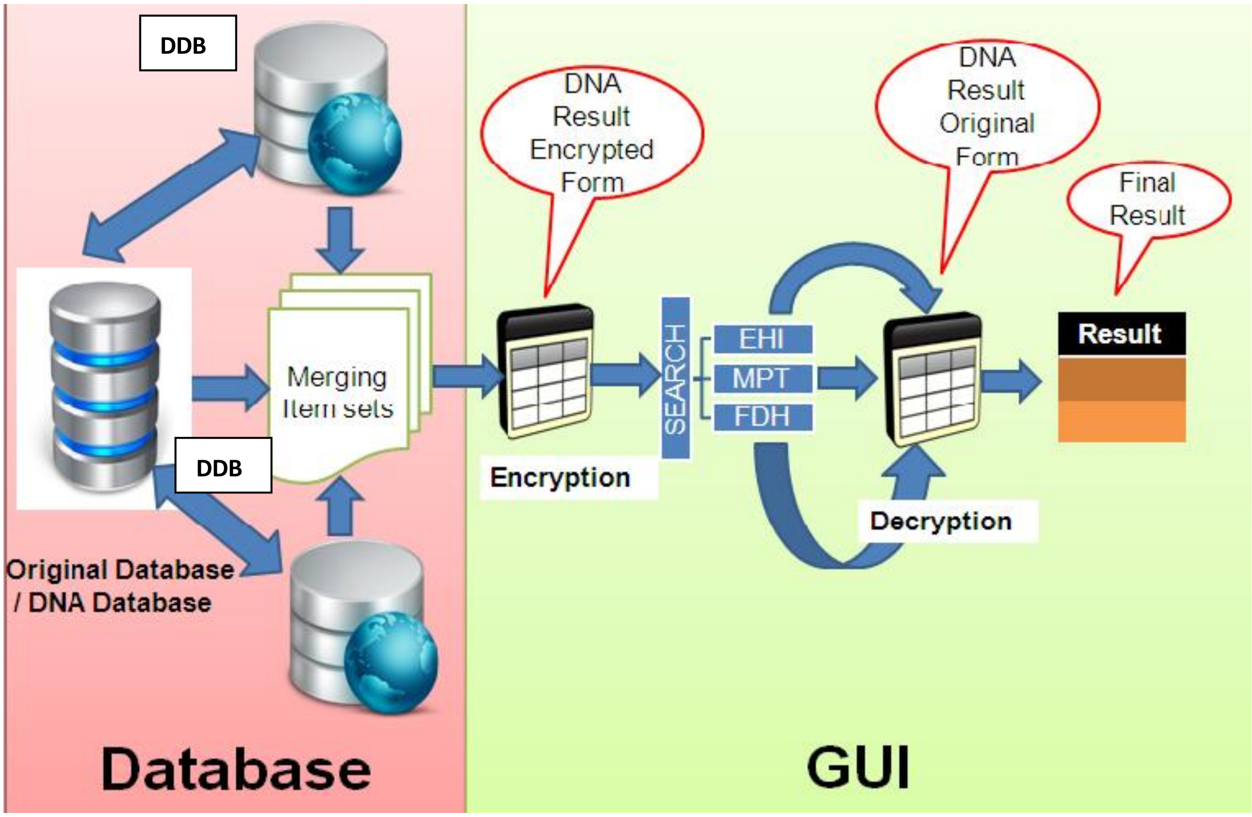


Fig 1: System architecture for horizontal database with enhances secure mining.

Phase 0: Getting started.
Phase 1: Encrypt DNA Result.
Phase 2: Searching special DNA Test on Database using one of the following searching Techniques.
Phase 3: Merging itemsets.
Phase 4: Generated result Decrypt by using decrypt key.

III. DETAILED DESCRIPTION OF PHASES

Phase 1 and 4:
Advanced Encryption Standard (AES):

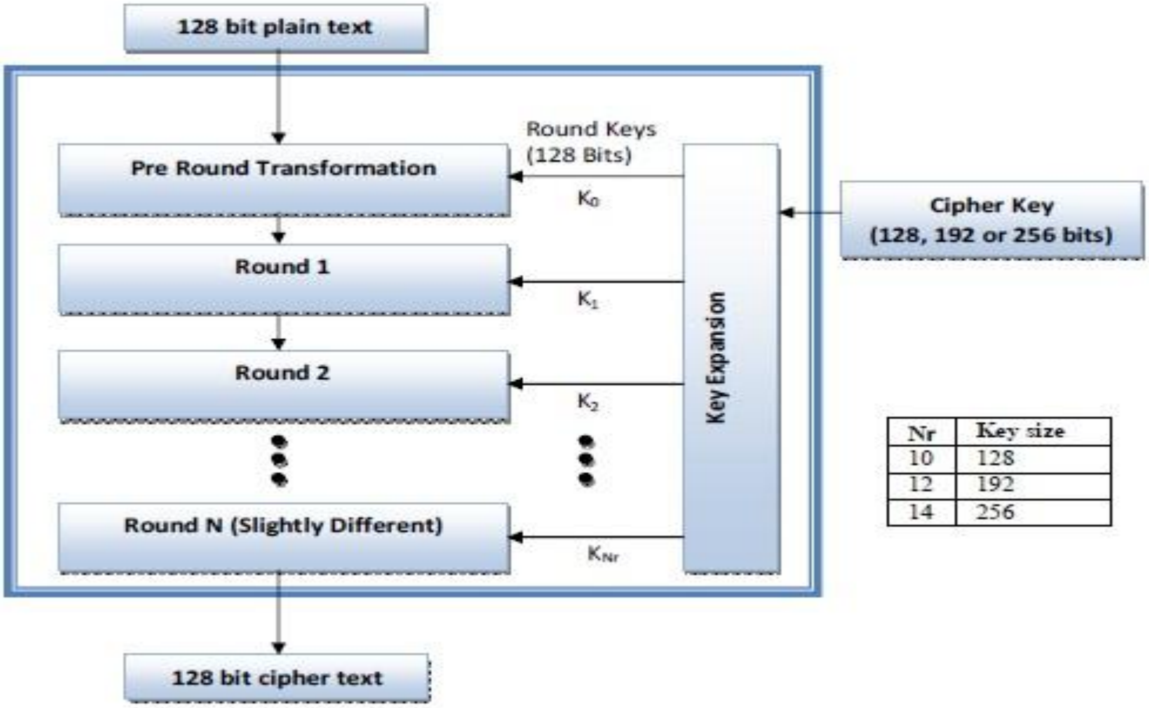
The Advanced Encryption Standard (AES) is a symmetric key encryption standard adopted by the U.S. government. The standard comprises three block ciphers AES-128 AES-192 and AES-256 adopted from a larger collection originally published as Rijndael. Each of these ciphers has a 128-bit block size with key sizes of 128, 192 and 256 bits respectively.

The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input called the plaintext into the final output called the cipher text. The numbers of cycles of repetition are as follows:

- 10 cycles of repetition for 128-bit keys.
- 12 cycles of repetition for 192-bit keys.
- 14 cycles of repetition for 256-bit keys.

Phase: 2
Secured searching of valuable data in database:

The aim is collecting the similarity queries from various users and stored in the database. Here data owner server trusted clients are used. Here it is able to maintain data confidentiality with respect to untrusted parties including the service provider. Data owner and service provider and trusted client are used. Data owner is one who stores the data in the database. Here server is the database or third party who maintains the data in the database. Trusted client is one who needs the data from the database. In this project the data owner provide the privacy to the sensitive information [2]. Here we use DNA Test related information so we collected the DNA Test related data and stored in my database. Only authorized users are allowed to access the data. Nobody else should be able to view the data. So that data will be kept private. Based on the queries it will be revealed to the trusted users alone.



Nr	Key size
10	128
12	192
14	256

Fig 2: General design of AES

This transformation technique offers perfect data privacy for the data owner but it gives the final result at multiple rounds of communication. This technique also provides an interesting trade-off between query cost and accuracy. Existing solutions either offer query efficiency at no privacy or they offer complete data privacy while sacrificing query efficiency. But the proposed methods are very secure and efficient.

1 Encrypted Hierarchical Index Based Search
2 Metric Preserving Transformation
3 Flexible Distance-Based Dynamic Hashing [8]

Phase 3:
Apriori Algorithm:

Apriori is designed to operate on databases containing transactions [11]. The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data.

IV. RESULT

Here one result is exiting protocol result and second one is propose system result. Here show three comparisons first graph show Computation Time second graph show Time to Generate Unify Items and third graph show Message Size. Results are shown below.

Computation Time:

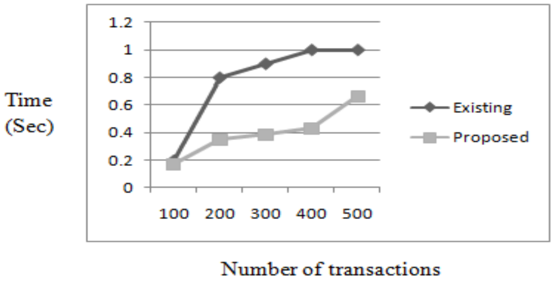


Fig 3: Graph for Computation Time Time to Generate Unify Items

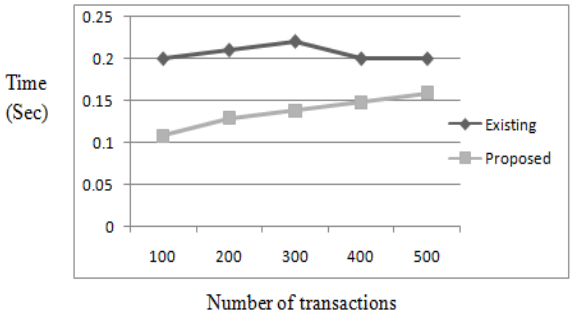


Fig 4: Graph for Time to Generate Unify Items

Message Size

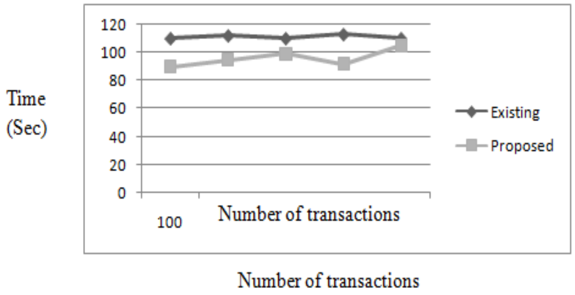


Fig 5: Graph for Message Size

V. CONCLUSION

Data mining describes a class of applications that look for hidden knowledge or patterns in large amounts of data. In this project a new efficient method has been proposed in order to keep confidentiality of data in database. A proposed protocol for secure mining of association rules in horizontally distributed databases improves significantly the current leading protocol in terms of privacy and efficiency. That data mining has a very important role in our life we use and handle it regularly. So security and privacy should be provided to that database.

This project concept is useful in medical system or more place [Example: Banking Sector etc.] because people or user don't want show their medical record to other person or third party person so they secure their report here or if anyone try to access their medical record so it found in encrypted format means third party person cannot read their data and also if owner and user want to access that record in database so by using searching method they can easily access particular test record in less time and securely.

VI. REFERENCES

[1] Tamir Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", 2013 IEEE.

[2] r. Rathika1, dr. K. Raja2 "secured searching of valuable data in a metric space based on similarity measure" ijcsmc, vol. 2, issue. 4, april 2013, pg.507 – 512.

[3]. M. Kantarcioglu and C. Clifton. Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering, 16:1026-1037, 2004.

[4]. T. Tassa and D. Cohen. Anonymization of centralized and distributed social networks by sequential clustering. IEEE Transactions on Knowledge and Data Engineering, 2012.

[5]. T. Tassa and D. Cohen. Anonymization of centralized and distributed social networks by sequential clustering. IEEE Transactions on Knowledge and Data Engineering, 2012.

[6]. J. Vaidya and C. Clifton. Privacy preserving association rule mining in vertically partitioned data. In KDD, pages 639-644, 2002.

[7]. S. Zhong, Z. Yang, and R.N. Wright. Privacy-enhancing kanonymization of customer data. In PODS, pages 139-147, 2005.

[8] Stepan Kozak, David Novak, and Pavel Zezula, "Secure Metric-Based Index for Similarity Cloud" Springer-Verlag Berlin Heidelberg 2012.

[9] R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In VLDB, pages 487–499, 1994.

[10]. M. Kantarcioglu and C. Clifton. Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering, 16:1026–1037, 2004.

[11]. A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy preserving mining of association rules. In KDD, pages 217–228, 2002.