ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Elliptic Curve Cryptography Based Algorithm for Privacy Preserving in Data Mining

Krishna Pratap Rao[1], Adesh chaudhary[1] , Prashant johri[1]

School of Computing Science and Engineering, Galgotias University, India

Abstract: - Data mining is used to extract or "mine" knowledge from large amounts of data base. However, data is often distributed among different locations. In terms of privacy and security issues data mining techniques are recently investigated with the conclusion that they reveal data or information to other parties involved to find global results. Recently many cryptographic techniques have been found to address privacy problem in data mining. However the methods are too far too inefficient and impractical for computing large databases. In this paper, we propose efficient algorithm to mine association rule using elliptic curve cryptography technique over horizontally partitioned data. Here we consider unsecured environment. Our proposed algorithm provides security against involving parties and intruder who can read the unsecured channel and algorithm provides authentication between involving parties. We discussed the privacy and security provided by our proposed algorithm.

Keywords: - Data mining, privacy, cryptography, Elliptic Curve, Association Rule Mining

I. INTRODUCTION

Most of the organizations today have multiple data sources distributed at different locations, which need to be analyzed to generate interesting patterns and rules. A very effective way to deal with multiple data sources (where data to be mined is distributed among several relations on different database management systems (DBMSs) is to mine the association rules at different sources and forward the rules to a centralized system rather than sending the data to be mined which is likely to be very large.

Data mining promises to discover unknown data. If those data are personal or corporate data, it offers the potential to reveal what others regard as private. This problem is called Privacy Preserving Data Mining (PPDM). In order to perform privacy preserving data mining these techniques such (1) as kanonymity classification, clustering and association rule mining have been suggested in recent years. As data mining become more pervasive, privacy concerns are increasing. Organizations obtain information about individuals for their own specific needs. Different units within an organization themselves may find it necessary to share information. Each organization must be sure that the privacy of the individual is not violated or that sensitive business information is not revealed. In this paper a model which is exactly that of multi-party computation. So, there exists a secure protocol for any probabilistic polynomial-time functionality. These solutions are very inefficient, when large inputs and complex algorithms are involved. We assume that there is unsecured environment. In any multi-party computation setting, malicious adversary can always alter its input. For the data- mining setting, this fact can be very damaging since the adversary can define its input to be the empty database.

II. LITERATURE SURVEY

Research has addressed classification using Bayesian Networks in vertically partitioned data [2], and situations where the distribution is itself interesting with respect to what is learned [3]. Shenoy et al. proposed an efficient algorithm for vertically mining association rules [4]. Finally, data mining algorithms that partition the data into subsets have been developed [5]. However, none of this work has directly addressed privacy issues and concerns.

PRIVACY PRESERVING ALGORITHM: - A lot of implementations of the confidentiality of data and knowledge are applied in association rule mining process. Basically privacy preserving association rule mining algorithms commonly can be divided into three categories.

(A) HEURISTIC-BASED TECHNIQUES: - Heuristic-based techniques are used to resolve how to select the appropriate data sets for data modification. Optimal selective data modification or sanitization is an NP-Hard problem , heuristics can be used to address the complexity issues.

(B) RECONSTRUCTION-BASED ASSOCIATION

RULE: - Agrawal et al. in [6] first proposed the method of distribution reconstruction on numeric data which is disturbed by Bayesian algorithm in 2000. Then, Dakshi and Charu in [7] improve the work over the Bayesian-based reconstruction procedure by using an Expectation Maximization (EM) algorithm for distribution reconstruction.

(C) CRYPTOGRAPHY-BASED TECHNIQUES:- Many Cryptography-based approaches have been proposed in the context of privacy preserving data mining algorithms. Cryptography-based approaches like Secure Multi-party Computation (SMC) are secure at the end of the computations.

(D) SECURE MULTI-PARTY COMPUTATION:- Goldwasser defined the SMC problem as a problem that deals with computing any function on any input, in a distributed network where each participant holds one of the inputs, while ensuring that no more information is revealed to a participant in the computation than the information that can be inferred from that participant's input and output [8].

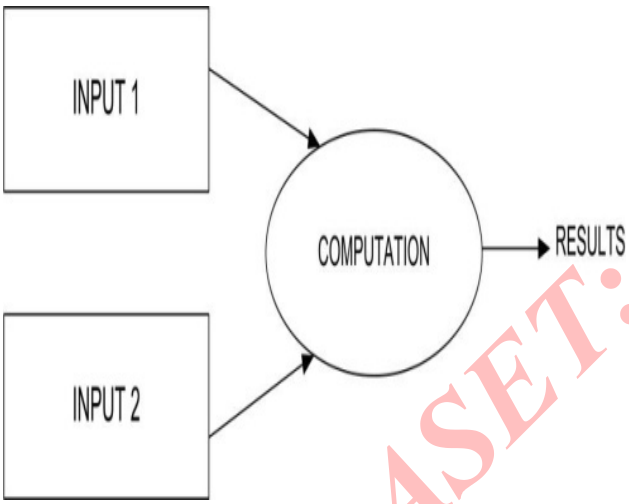

Figure 1. Transformation of a multi-input computation model to a secure multiparty computation model.

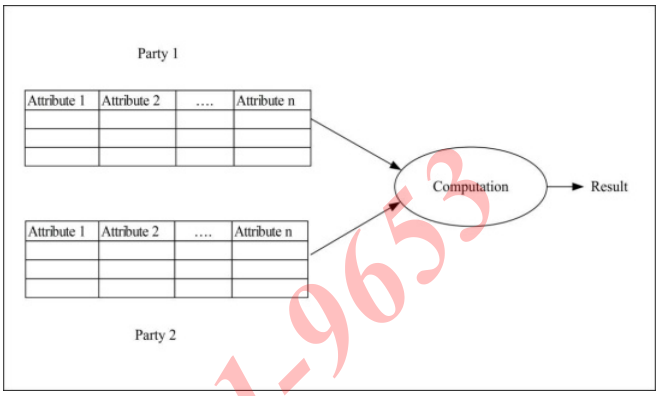

Figure 2 . Transformation of a single-input computation

Model to a homogeneous secure multi-party computation model.

III. MOTIVATION

Computers have promised us a fountain of wisdom but delivered a flood of information. This huge amount of information is distributed among the different parties and distributed data makes it crucial to develop tools to discover hidden knowledge (association rule). These tools are called data mining tools and promise to discover hidden knowledge, but if that hidden knowledge is sensitive and then owners would not be happy if this knowledge were exposed to the public, involving parties or to adversaries. This problem motivates research to develop techniques, algorithms and protocols to assure data owners that privacy is protected while satisfying their need to share data among involving parties in distributed environment.

IV. PROBLEM DESCRIPTION

We consider scenario where more than two parties want to corporate for computing association rule on union on their databases. Although databases are private to party and they don't want to reveal their private information to other parties or to the intruder.

We show how involved parties find global association rule efficiently without revealing personal information to other involving parties or intruder. Involving parties cannot learn anything other than global association rule.

In proposed approach, we used Apriori algorithm to generator frequent item set and for authentication and security we used ECDSA (Elliptic Curve based Digital Signature Algorithm) and ECIES (Elliptic Curve Integrated Encryption Scheme) under elliptic curve cryptography.

V. PROPOSED APPROACH

In this paper the aim is to prevent disclosure of association rule of individual parties (from adversary or malicious user).The main characteristics of proposed algorithm for distributed database listed

below.

Objective

- Privacy and Security: Proposed algorithm provides privacy against involving parties in distributed environment and also provides security and privacy against adversary who is be part of mining process in distributed environment.
- Efficiency: Finding association rule with less iteration and less time consuming.
- Communication and Computation Cost: The communication and computation cost are reasonable for small database.

ALGORITHM

In this section we describe our proposed elliptic curve cryptography based parallel algorithm for privacy preserving mining of association rule in distributed environment. However we consider only horizontally partitioned data. First we will discuss the basic concepts which are used in our proposed algorithm. Then we describe proposed communication protocol and privacy preserving association rule mining algorithm in detail.

(A). *Elliptic Curve Integrated Encryption Scheme (ECIES) for encryption:*

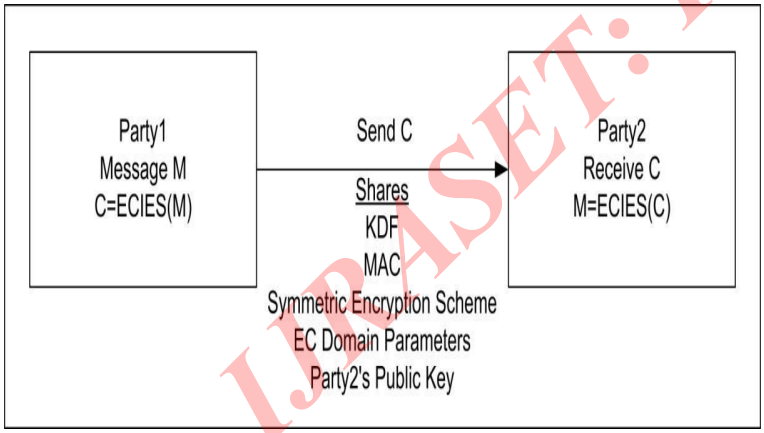

Figure 3. Use of ECIES algorithm between two parties**.**

(B). Elliptic curve based digital signature algorithm (ECDSA) for authentication and verification:

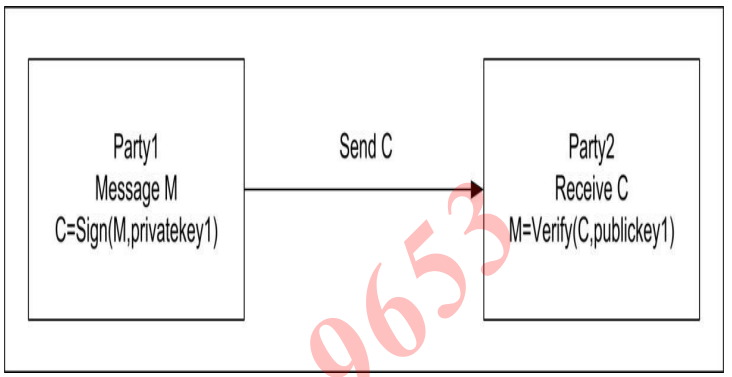

Figure 4. Use of ECDSA algorithm between two parties.

(C). Use of ECIES and ECDSA in our protocol :

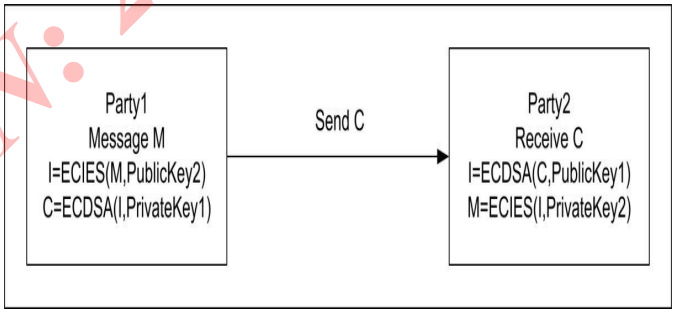

Figure5 .Use of ECIES-ECDSA in our algorithm

VI. RESULTS AND SIMULATION

We tested our protocol on different datasets and recorded results. Here we are comparing our results with the old method [9] with respect to different database [10][11][12]. Analysis of our proposed method is shown in next subsection.

In proposed algorithm initiator party adds random number, signs and encrypts the total count and send to another party. Another party cannot guess the value easily and cannot predict original value and this way provides privacy against involving parties. Parties cannot read communication channel because massage is passing in encrypted form. Thus using this protocol we provide privacy against involving parties.

Table 1: Experimental results of chess.dat database

Database	Support	Algorithm	3-parties	4-parties	5-parties	6-parties
Chess.dat	0.95	New	4.6 sec	5.7 sec	6.7 sec	7.1 sec
		Old	4.7 sec	6 sec	7.3 sec	7.5 sec
	0.9	New	8.6 sec	8.8 sec	9.6 sec	11.9 sec
		Old	14.8 sec	13.1 sec	15.4 sec	17.1 sec
	0.85	New	19.8 sec	23.7 sec	24.8 sec	28.1 sec
		Old	52.4 sec	56.3 sec	57.6 sec	61.2 sec
	0.8	New	72.8 sec	75.6 sec	82.2 sec	87.9 sec
		Old	151.5 sec	169.8 sec	175.5 sec	179.8 sec

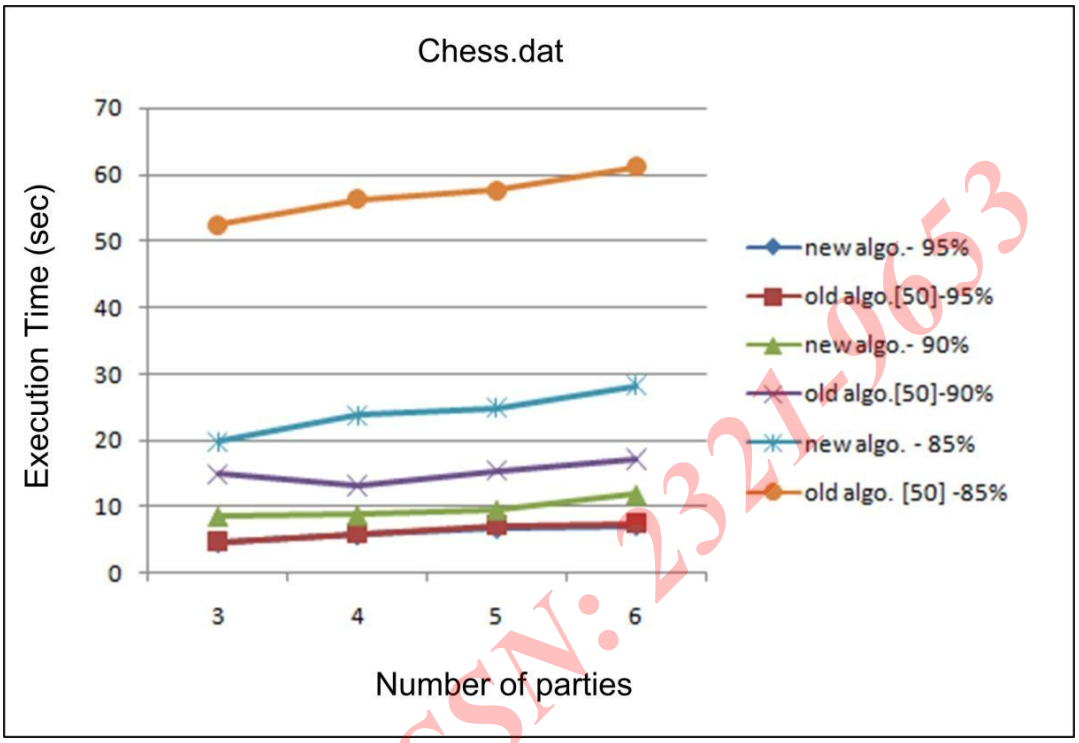

Figure 6. Comparison between old and new algorithm on chess.dat database for different parties and different support

VII. CONCLUSION

Here we proposed a novel approach for mining association rules on horizontally partitioned data in unsecured environment using elliptic curve cryptography. We also show that the global association rule can be generated without loss of privacy to the involving parties or intruder. The communication cost and computational cost also be reasonable for small amount of databases.

REFERENCES

[1] Ashishkumar C. Patel , Elliptic Curve Cryptography Based Algorithm for Privacy Preserving in Data Mining, 2011-12.

[2] R. Chen, K. Sivakumar, and H. Kargupta. "Distributed web mining using bayesian networks from multiple data streams", In N. Cercone, T. Young Lin, and X. Wu, editors, Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01), pages 75–82, San Jose, California, USA, November 2001. IEEE Computer Society.

[3] R. Wirth, M. Borth, and J. Hipp. "When distribution is part of the semantics: A new problem class for distributed knowledge discovery", In Ubiquitous Data Mining for Mobile and Distributed

Environments workshop associated with the Joint 12th European Conference on Machine earning (ECML'01) and 5th European Conference on Principles and Practice of Knowledge Discovery in databases (PKDD'01),pages 3–7, Freiburg, Germany, September 2001. ACM Press

[4] P. Shenoy, J. R. Haritsa, S. Sundarshan, G. Bhalotia, M. Bawa, and D. Shah. "Turbo-charging vertical mining of large

databases", In Proceedings of ACM SIGMOD Record, volume 29(2), pages 22–33, Dallas, Texas,
USA, June 2000. ACM Press

[5] A. Savasere, E. Omiecinski, and S. B. Navathe. "An efficient algorithm for mining association rules in large databases", In Proceedings of the 21st International Conference on Very Large Data Bases (VLDB'95),pages 432–444, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

[6] Rakesh Agrawal and Ramakrishnan Srikant, "Privacy-preserving data mining", In Proceedings of the ACM SIGMOD Conference on Management of Data (2000), pp.439–450.

[7] Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke, "Privacy preserving mining of association rules", In Proceedings of the 8th ACM SIGKDDD International Conference on Knowledge Discovery and Data Mining (2002).

[8] S. Goldwasser. "Multi-party computations: Past and present", In Proceedings of the 16th Annual ACM Symposium on the Principles of Distributed Computing, pages 1–6, Santa Barbara, California, USA, 1997. ACM Press.

[9] Modi Chirag, Rao Udai Pratap, Patel Dhiren R, "Elliptic Curve Cryptography Based Mining of Privacy Preserving Association Rules in Unsecured Distributed Environment", In proceedings of the International Conference on Advances in Communication, Network, and Computing, IEEE Society,pp 94-98,2010.

[10]. "Frequent Itemset Mining Implementations Repository", http://fimi.ua.ac.be/data/chess.dat [Accessed: Feb. 5, 2012].

[11]."Frequent Itemset Mining Implementations Repository", http://fimi.ua.ac.be/data/T10I4D100K.dat [Accessed: Feb. 5, 2012].

[12]. "Frequent Itemset Mining Implementations Repository", http://fimi.ua.ac.be/data/mushroom.dat [Accessed:Feb. 5, 2012].