**INTRODUCTION**

FREQUENT itemset mining and association rule mining are most significantly used data analysis techniques. These techniques are generally used for discovering frequently co-occurring data items and association relationships between data items in large transaction databases. Frequent pattern means frequently occurring data pattern. The different type of patterns are frequent itemset,frequent subsequences and substructures. Frequent itemset is a itemsets that appear together in a transactional dataset. For example milk and bread, these are brought together in a market by customer. Association rule mining is used to find the frequent patterns, correlation, associations or structures from datasets of relation database, transactional database or other forms of data source.

These two techniques are used in applications such as market basket analysis, health care, web usage mining, bio informatics and prediction. A transaction database is a set of transactions, and each transaction is a set of data items with a unique Transaction ID (TID). It is easy to mine association rules after mining frequent itemsets and obtaining their supports. Most association rule mining algorithms are built based on frequent itemset mining algorithms. Classic frequent itemset mining and association rule mining algorithms, such as Apriori, Eclat and FP-growth, were designed for a centralized database setting where the raw data is stored in the central site for mining. Privacy concerns were not considered in this setting. Vaidya and Clifton and Kantarcioglu and Clifton are the first to identify and address privacy issues in horizontally / vertically partitioned databases. Due to an increased understanding of the importance of data privacy,a number of privacy-preserving mining solutions have been proposed in recent times.

**Problem Statement**

Earlier frequent itemset mining and association rule mining algorithms like Apriori,Eclat and FP-growth were used only for centralized database, where data is stored in central site and privacy issue was not concerned. Due to increased importance of data many privacy preserving mining have been proposed but data owners who are having multiple data are not willing to send their data to central site due to privacy concerns. By providing cloud-aided privacy-preserving frequent itemset mining solution for vertically partitioned database, build a privacy-preserving association rule mining solution. This solution is useful when data owners want high level privacy requirement and also for data owners looking for outsourced data storage- i.e, data owners can outsource their encrypted data and mining task to a semi-trusted cloud in a privacy preserving manner.

**Existing System**

Privacy-preserving association rule mining and frequent itemset mining on vertically partitioned databases the first work to identify and address privacy issues in vertically partitioned databases, a secure scalar product protocol is presented and used to built privacy-preserving frequent item set mining solution. Association rules can then be found given frequent itemsets and their support. All existing solutions, with the expection of,do not utilize a third-party server to compute the mining result. Some solutions use asymmetric encryption to compute the supports of itemsets, while other solutions use a secure scalar product protocol, a set intersection cardinality protocol or a secret sharing scheme to perform these computations. Other than the setting of vertically partitioned databases and third party-aided mining, privacy-preserving frequent itemset mining and association rule mining have been studied in the setting of horizontally partitioned databases data publishing and differential privacy.

**Proposed System**

Efficient homomorphic encryption scheme is proposed here. Using the proposed homomorphic encryption scheme, secure outsourced comparison scheme is constructed. Both scheme will then serve as the basis of our privacy-preserving mining solutions. Existing homomorphic encryption schemes are generally asymmetric. In this, propose a symmetric homomorphic encryption scheme,which is significantly more efficient than asymmetric schemes.The scheme support many homomorphic addition and limited number of homomorphic multiplications, and comprises the following three algorithms key generation algorithm, Decryption algorithm.

**Objective**

* Provide cloud-aided privacy-preserving frequent itemset mining solutions for vertically partitioned databases, which help to build a privacy-preserving association rule mining solution.
* It helps to build privacy for association rule mining solution, where these solutions designed for application and data owners have high level privacy.
* Privacy preserving mining solution are more efficient due to use of homomorphic encryption scheme.
* By designing efficient homomorphic encryption can get secure outsourced computation of confidence and secure outsourced comparison sheme for comparing confidence with thresholds.
* Ciphertext tag approach is used to avoid fictious data’s effect on mining result.

**LITERATURE SURVEY**

* Zhiqiang Yang and Rebecca N[1] explained about privacy preserving on vertically partitioned data using Bayesian network. Bayesian network is an acyclic graph of random variables and their condition. There were many data mining techniques designed in the centralize model in which all data is collected and available in one central sites. As data increased,that are carried out using computer and computer network and more this become difficult to store more sensitive data.The privacy preserving protocol is used to give solution by using distributed data mining algorithm in which data is protected. if two clients have confidential database and want to learn Bayesian network on the combination of their database without leaking their data to each other.By using Bayesian network can get privacy,efficiency,privacy and accuracy than MSK.Bayesian network can be computed with constant overhead.Bayesian network used only acylic graph what to do when we have cyclic graph.
* Jaideep Vaidya and Chris Clifton[2] describes privacy preserving association rule mining in vertically partitioned data. here mainly considering the privacy conservation.They mainly addressing the problem of data rule algorithm.Data is distributed over all the sites,each site consisting different types of attributes and then need to consider all the attributes.Two partition algorithm is used.It shows the individual transactions. clustering algorithm is used partitioning of the data. Vertically partitioned data is analyzed by using market based techniques.Market based technique consisting mainly two parts grocery purchase and clothing purchase like it shows the relationship between the grocery purchase and clothing purchase like using the ATM cards. Examples are tires,electronic products and car equipments,etc. It gives the better privacy setting. Large data can be transmitted.By using this method there are some problems like It leads to the homogeneous population, Duplication can occur, unable to detect cross-site correlation and value for a single data can split across other site it leads to confusion.
* Fosca Giannotti,Laks and Anna[3] proposed privacy preserving mining of association rules from transactin databases.It mainly describes the third person correlated data and how to access that data using simple example, the company owners owns the data from the outsourced but it is not efficient so it contact the third person for the data or patterns and all the data will be private data. It is going to store the data in the different servers for privacy purpose. It will be done by privacy preserving framework to protecting the private data and then some of the data can be used. Mining algorithm can provide good security and scalability for private data.Encryption schema is used to identify the cipher data or item from correlated data easily. It works like adding one-one substitution cipher for items and adding fake transaction to wake each cipher item share the same frequency as>=k-1 others.In this method cryptographic notation schema for analyzing the privacy splitting.Easily identify the cipher iteams from the exact patterns.provides better security. Some disadvantages like complex networking analysis.difficult for understanding, because it contains the more AND/OR operations and it only consider the encryption items.
* Mohmoud Hussein[4] explained about performance turing of steganography algorithm for privacy preserving association rule mining. Mainly focus on privately mining association rules in vertically partitioned data and the problem is to reduce privately computing Boolean scalar products. They made a modification of steganography-based multiparty protocols for this problem. There are some existing techniques they use for solving the problem of private scalar product but they have some problem like increasing the running time of the scalar product computing. In early methods the main concentration is to make better privacy preserving with high performance. When the data base become large then adding the privacy will decrease the performance, so some algorithms are proposed to solve this problem of mining very large databases. To gain high performance with acceptable level of privacy in large database, they proposed fast technique for computing scalar product. By reducing the computation time of computing scalar product, helps smaller matrix to hide the vectors used in computing the scalar product. In this work they propose a modification of steganography-based multiparty protocol for computing scalar product.This modification gives suitable solution for tradeoff between the performance and privacy.The proposed modification fine tune the performance to be faster in case of very large database, with acceptable level of reduction in privacy. It will prevent the discovery of sensible information.
* Madhuri N. Kumbhar[5] proposed privacy preserving for association rule on horizontal and vertically partitioned database.The main aim of privacy preserving data mining is to preserve the individual site information.Many Privacy Preserving Association Rule Mining [PPARM] algorithms are proposed for different partitioning methods to satisfy privacy constraints. The privacy constraints in vertically partitioned databases, algorithm based on cryptography techniques, Homomorphic encryption, Secure Scalar product and Shamir’s secret sharing technique are used. For horizontal Partitioned databases, algorithm that combines advantage of both RSA public key cryptosystem and Homomorphic encryption scheme and algorithm that uses Paillier cryptosystem to compute global supports are used.Data miner is used to initiate the process by sending support threshold and public key. Data miner also used in encryption and decryption process for frequent item sets in order to protect individual sites information.This scheme is not secure to maintain privacy of master sites. Boolean association rule mining are designed for semi-honest model.

**Methodology**

* **Substitution Cipher and Frequency Analysis**

A substitution cipher encrypts a message by substituting the units of the message with ciphertext units according to a substitution alphabet. Substitution cipher has been used in outsourced association rule mining and frequent itemset mining. Substitution cipher is subject to frequency analysis attack if the frequencies of message units are different. Frequency analysis, the analysis of frequencies of ciphertext units or unit groups, has been used to break classical ciphers such as substitution ciphers.

Attackers with some knowledge of the frequencies of message units or unit groups can recover some plaintext through frequency analysis. For example, if an attacker knows that bread and milk are the most and second most frequent items in a transaction database, the attacker can infer that the most and second most frequent ciphertext units in the encrypted database correspond to bread and milk,respectively. To counter frequency analysis attack, fictitious items or transactions can be added to hide item frequency.

* **Homomorphic Encryption Scheme** Existing homomorphic encryption schemes are generally asymmetric. Here by focus on a symmetric homomorphic encryption scheme (using only modular additions and multiplications), which is significantly more efficient than asymmetric schemes. The scheme supports many homomorphic additions and limited number of homomorphic multiplications, and comprises the following three algorithms:

• **Key generation algorithm**

KeyGen() (s, q, p) ← KeyGen(λ) The key generation algorithm KeyGen() is a probabilistic algorithm, which takes a security parameter λ as input and outputs a secret key SK = (s, q) and a public parameter p. Both p and q are big primes, and p \_ q. The bit length of q depends on the security parameter, and s is a random number from Z∗p.

• **Encryption algorithm** E()

E(SK,m, d) = sd (rq + m) mod p The encryption algorithm E() is a probabilistic algorithm, which takes a secret key SK, a plaintext m ∈ Fq and a parameter d as inputs. The algorithm outputs a ciphertext c ← E(SK,m, d). The parameter d is a small positive integer called ciphertext degree, and we say the ciphertext is a d-degree ciphertext. Let r denote a big random positive integer, and the bit length of r , |r |, satisfies |r| + |q| < |p|. We say r is the random ingredient of c. The encryption of a plaintext m is denoted by E(m) for short.

• **Decryption algorithm** D()

D(SK, c, d) = (c × s−d mod p) mod q The decryption algorithm D() is a deterministic algorithm, which takes a secret key SK, a ciphertext c ∈ Fp and the ciphertext’s degree d as inputs. The algorithm outputs a plaintext m ← D(SK, c, d). Let s−d denote the multiplicative inverse of sd in the field Fp.

* **Privacy Preserving Outsourced Mining**

In Association rule mining solution each data owner owns a private database, and data owners collaboratively mine their joint database’s association rules with the assistance of the cloud. Our association rule mining solution includes two stages, namely: preprocessing and mining. In the preprocessing stage, data owners and the cloud collaborate to generate an encrypted joint database at the cloud’s end and some auxiliary data for privacy-preserving mining. Each data owner inserts fictitious transactions to his private database, and encrypts items in the database with a substitution cipher. The fictitious transactions are used to mitigate frequency analysis attacks.

**REFERENCES**

[1] Privacy-Preserving Computation of Bayesian Networks on Vertically Partitioned Data

Zhiqiang Yang and Rebecca N. Wright, Member, IEEE.

[2] Privacy preserving association rule mining in vertically partitioned data. Jaideep Vaidya and Chris Clifton,purdure University west Lafayette,Indiana.

[3] Privacy-preserving Mining of Association Rules from Outsourced Transaction Databases

Fosca Giannotti, Laks V.S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang.

[4] Performance Tuning of Steganography Algorithm For Privacy Preserving Association Rule Mining in Heterogeneous Data Base Mahmoud Hussein, Ashraf El-Sisi and Nabil Ismail Computer Science Department , Faculty of computers and Information, Menofyia University, Shebin Elkom 32511, Egypt.

[5] Privacy Preserving Mining of Association Rules on Horizontally and Vertically Partitioned Data: A Review Paper

[6] M. J. Zaki, “Scalable algorithms for association mining,” IEEE Trans. Knowl. Data Eng., vol. 12, no. 3, pp. 372–390, May/Jun. 2000.

[7] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” in Proc. ACM SIGMOD, pp. 1–12, 2000.

[8] J. Vaidya and C. Clifton, “Privacy preserving association rule mining in vertically partitioned data,” in Proc. SIGKDD, 2002, pp. 639–644.

[9] M. Kantarcioglu and C. Clifton, “Privacy-preserving distributed mining of association rules on horizontally partitioned data,” IEEE Trans.Knowl. Data Eng., vol. 16, no. 9, pp. 1026–1037, Sep. 2004.