Providing Security to Outsourced Data Using Association Rule Mining on Vertically Partitioned Databases

Korra Joshna
M.Tech Student
Department of Computer Science Engineering
AU College of Engineering (A), Visakhapatnam.

Dr. A.Mary Sowjanya
Assistant Professor
Department of Computer Science & Systems
& Systems Engineering
AU College of Engineering (A), Andhra University,
Andhra University, Visakhapatnam

I. Introduction :

Data mining is the extraction of hidden predictive information from large databases. Data mining tools can be used to predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. They can also solve problems that are traditionally time consuming to resolve, and search databases for hidden patterns, to find hidden information that experts also may miss because it may lies outside their expectations.

Privacy Preservation in data mining involves securing confidential or important data from unauthorized persons. It has become important recently because of the increasing ability to store users personal data, and corporate data of institutes.

The privacy of outsourced databases is also important because it involves a third party which provides a mechanism to allow customers to create, store and access databases. Outsourced databases help reduce hardware equipment cost, system building and cost of the personnel also. But, when all the data is placed with an outsourced database service provider, the provider can not be trusted .

The major requirements for providing security in outsourced databases are confidentiality, privacy, integrity, access control in multi-user environment, availability and query authentication. Security mechanisms like access control based approach, order preserving encryption based approach, hardware based encryption approach, fake tuple based approach, secret sharing approach, authenticated data structure approach, attributes based approach, combined fragmentation and encryption based approach can be used for this purpose.

II. Existing System :

Most of the previous privacy-preserving data mining frameworks mine patterns from data intended to be shared with parties other than the data owner. A conservative frequency-based attack model was one of the early works for defending against the frequency-based attack in the data mining outsourcing scenario. It introduced the idea of using fake items to defend against the frequency-based attack, but lacks privacy guarantee.

III. Proposed System :

In our proposed system not only the underlying data but also the mined results are not intended for sharing so must remain private. Association rule mining algorithms have been developed to gather all data into a central site, and then run algorithms against that data. The Whirlpool algorithm is used for discovering frequent item sets without revealing individual transaction values distributed across different sources. Also a cloud-aided privacy preserving frequent item set mining solution is designed in which data owners can outsource their encrypted data to multiple parties.

IV. Related Work :

R. Cramer, R. Gennaro, and B. Schoenmakers present a new multi-authority encryption scheme that guarantees privacy, universal verifiability, and robustness. An interesting property of the scheme is that the time and communication complexity for the user is independent of the number of authorities. A user can simply posts a single encrypted message accompanied by a compact proof that it contains a valid data. [vol. 8, no. 5, pp. 481–490, 1997].

R. Agrawal and R. Srikant have dealt with the problem of discovering association rules between items in a large database of sales transactions and combined the best features into a hybrid algorithm called Apriori Hybrid. Apriori Hybrid has excellent scale up properties with respect to the transaction size and the number of items in the database.[*Proc. VLDB*, 1994, pp. 1–13].

W. K. Wong D. W. Cheung, E. Hung, propose a substitution cipher techniques in the encryption of transactional data for outsourcing association rule mining. [*Proc. VLDB*,2007, pp. 111–122].

When two sites want to engage in an association

rule mining, data privacy concerns are raised. Homomorphic encryption based solutions give more accurate results than data perturbation. M. G. Kaosar, R. Paulet, proposed a secure comparison technique based on state-of-the-art fully homomorphic encryption scheme, by which they built secure two-party association rule mining protocol.Which preserves complete privacy of both parties.[*ACSW-AISC*, 2011, pp. 15–22].

J.-L. Lin and J. Y.-C. Liu in their investigate the problem of privacy-preserving mining of association rules. A fake transaction randomization method is presented to ensure the privacy of data by mixing real transactions with fake transactions. The algorithm uses any off-the-shelf tool to mine frequent itemsets without rewriting their codes, making it easy to implement.[Mar. 2007, pp. 375–379].

V. Methodology :

The proposed approach provides security to outsourced data using security mechanisms. Association rule mining is done within the cloud so that a privacy preserving frequent itemset mining solution is generated. As such the data owners can outsource their encrypted data to other parties without worrying about security and privacy issues.

Whirlpool algorithm is used for encryption and also decryption. The interested parties can now decrypt the encrypted data set and use the frequent itemsets generated previously to recover the real frequent itmesets. Only the authorized parties having the key word can do so.

STEPS :

1. Pre-Processing

1.1 Data owner inserts fictious transcations to the original dataset inorder to midicate frequency analysis attacks.

1.2 Data items are encrypted using Whirlpool algorithm.

1.3 The encrypted database is outsourced to the cloud.

2. Mining

2.1 In the cloud association rule mining is performed to mine association rules from the encrypted database.

2.2 The cloud detects original data items from the fictious transcations with data owners help.

2.3 The cloud outsources the encrypted database to the interested parties.

3. Post-Processing

3.1 Only the authorized parties having a key word given by the cloud can make use of the outsourced data.

3.2 That party can recover real frequent itemsets from the frequent itemsets generated in the mining phase and also decrypt the encrypted dataset.

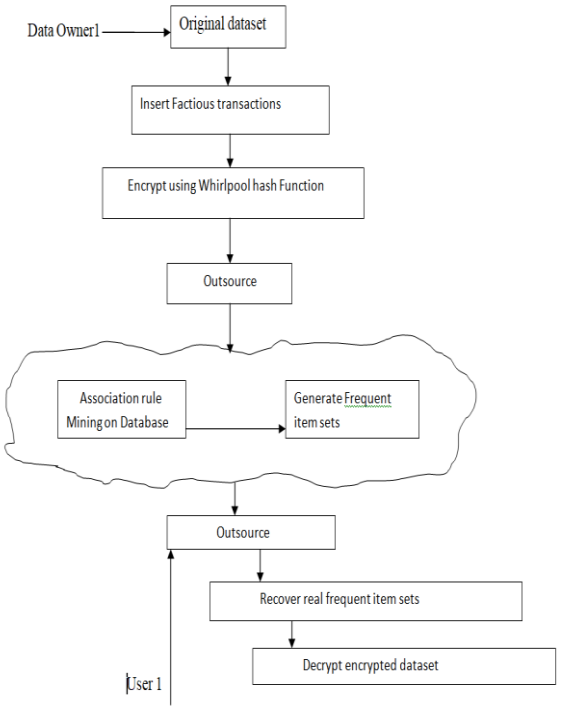
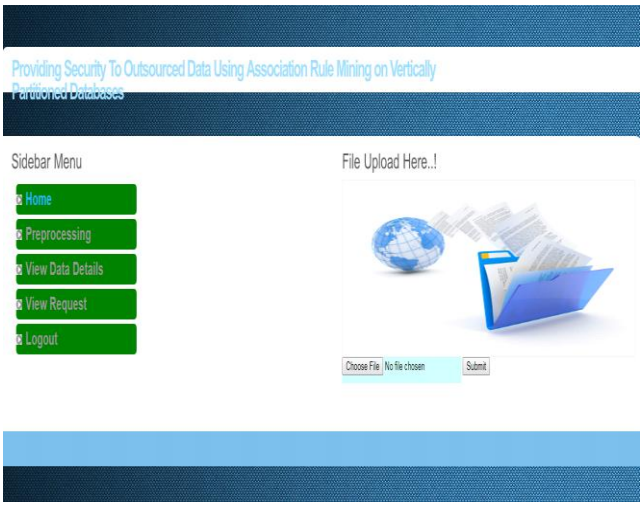


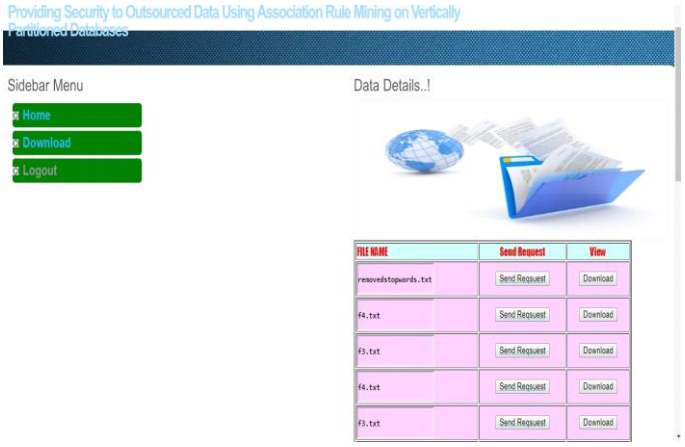
Figure : Block Diagram

RESULTS :

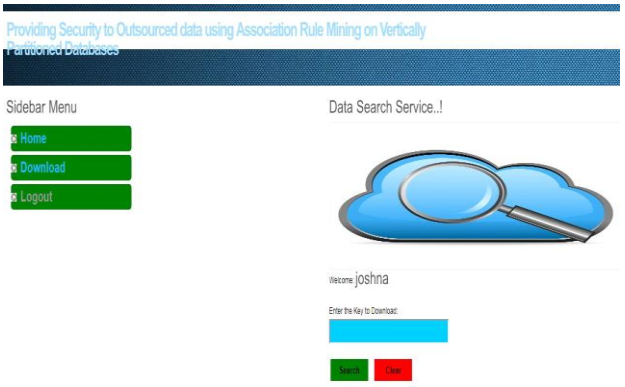


Description: The admin logins by his/her login id and password and uploads the files by choosing the file from the selected folder and browse the image then uploads the file.

Description: After uploading the file, the cloud displays this welcome page.



Description: The uploaded files are displayed in this picture.



Description: The key is entered by the cloud to authorized user generated to their e-mail or mobile number to download the required files.

VI. Conclusion :

In this paper we have proposed an approach to provide security to outsourced data using association rule mining. Data owners can securely outsource their data to interested parties through a cloud. encryption and association rule mining algorithms are used for that purpose. Only authorized parties can decrypt and use the outsourced data. In future more robust encryption schemes can be used in Whirlpool algorithm. Also instead of only key words a more secure method can be used to identify parties.

REFERENCES :

[1] Privacy-Preserving Outsourced Association rule mining on Vertically partioned databases. [Lichun Li, Rongxing Lu]

[2] A Secure and optimally efficient multi-authority encryption scheme. [vol. 8, no. 5, pp. 481–490, 1997].

[3] Fast algorithms for mining association rules [*VLDB*, 1994, pp. 1–13].

[4] Privacy-preserving mining of association rules from outsourced transaction databases[*IEEE Syst. J.*, vol. 7, no. 3, pp. 385–395,Sep. 2013].

[5] Secure two-party association rule mining *[ACSW-AISC*, 2011, pp.15 22].

[6] Privacy preserving itemset mining through fake transactions [Mar. 2007, pp. 375–379.

[7] A fast secure dot product protocol with application to privacy preserving association rule mining [May 2014, pp. 606–617].

[8] Privacy-preserving distributed association rule mining based on the secret sharing technique [Jun. 2010, pp. 345–350].

[9] Towards semantically secure outsourcing of association rule mining on categorical data [vol. 267, pp. 267–286, May 2014].

[10] Privacy preserving association rule mining over distributed databases using genetic algorithm [vol. 22, no. 1, pp. 351–364, 2013.