# To Reduce Data Leakage in Horizontally Distributed Database Using Association Rules

S.Azarudeen

M.Tech(information technology)
SRM University-Kattankulathur, Chennai-India
Azarudeen0189@gmail.com

**Abstract— The mainstay of this project is to propose a protocol for to reduce data leakage in horizontally distributed database using association rule. The current leading protocol is that of kantarcioglu and Clifton in [1]. My protocol is based on the Fast Distributed Mining (FDM) algorithm, which is an unsecured distributed version of the Apriori algorithm. This protocol offers enhanced privacy with respect to the protocol. In addition it is simpler and is significantly more efficient in terms of communications rounds, communication cost and computational cost.**

**Index Terms— privacy preserving data mining, Advanced Encryption Standard, Association rules.**

## 1.INTRODUCTION

### *Purpose*

There are several sites (or players) that hold homogeneous databases, i.e., databases that share the same schema but hold information on different entities in [2], [3] . The goal is to find all association rules with support at least *s* and confidence at least *c*, for some given minimal support size *s* and confidence level *c*, that hold in the unified database, while 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 [3], [4].

### *Scope*

We propose a protocol for secure mining of association rules in horizontally distributed databases. Our protocol is based on the Fast Distributed Mining (FDM) algorithm in [1]. The main ingredients in our protocol are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the interacting players hold, and another that tests the inclusion of an element held by one player in a subset held by another. Our protocol offers enhanced privacy. In addition, it is simpler and is significantly more efficient in terms of communication rounds, communication cost and computational cost.

## 2. EXISTING SYSYEM

Consider there are M players that hold private inputs, x1….xM, and they wish to securely compute y = f(x1, . . . , xM) for some public function f. If there existed a trusted third party, the players could surrender to him their inputs and he would perform the function evaluation and send to them the resulting output in [4]. In the absence of such a trusted third party, it is needed to devise a protocol that the players can run on their own in order to arrive at the required output y.

Kantarcioglu and Clifton studied that problem and devised a protocol for its solution. The main part of the protocol is a sub-protocol for the secure computation of the union of private subsets that are held by the different players. That is the most costly part of the protocol and its implementation relies upon cryptographic primitives such as commutative encryption, oblivious transfer, and hash functions. This is also the only part in the protocol in which the players may extract from their view of the protocol information on other databases, beyond what is implied by the final output and their own input.

## 3. ADVANCED ENCRYPTION STANDARD

In this paper we mainly discuss about the security of data. Using Advanced Encryption Standard (AES) in [5] technique data will be encrypt and decrypt while inserting and retrieving the data. Advanced Encryption Standard (AES) takes only less amount of time to encrypt and decrypt the data.

## 4. ASSOCIATION RULE

Association rule is popular and well research method for discovering interesting Relation between variables in large database. It is intended to identify strong rules discovered in database using different measures of interestingness

## 5. PROPOSED SYSTEM

While extracting data from distributed database system more number of irrelevant data will occur. Irrelevant data is avoided by using the Apriori algorithm in [6]. Data leakage is more in Apriori algorithm. Encryption is done at the time of retrieving data from the database.

## 6. USER REGISTRATION

***Initial Registration***
To register in this application, you have to do initial registration. It's simple, all you have to do is, just give your mail id and enter the captcha correctly given there in the application. After that, registration link will be sent to your mail id.

***Creating Account***
Now, after receiving the registration, click on it or copy and paste it in your browser. Register a new account.

***Confirm Registration***
After registering, a confirmation link will be sent to your mailed. After confirming your registration, you're eligible to login into your employment office account.

## 7. REGISTRATION

***Registering academic details***
After successful login, you can register your academic details. Make sure that, you are providing valid certificate number in the asked field. Users from different circle can enter their records by correctly choosing their circle.

***Updating academic details***
Users can maintain their up to date buy updating their records till PG.

***Renewal***
Every user has an expiry date for their registered account. Users can renew it through renewal option. They can extend their validity up to 2 years in one renewal

## 8 ADMIN MONITOR

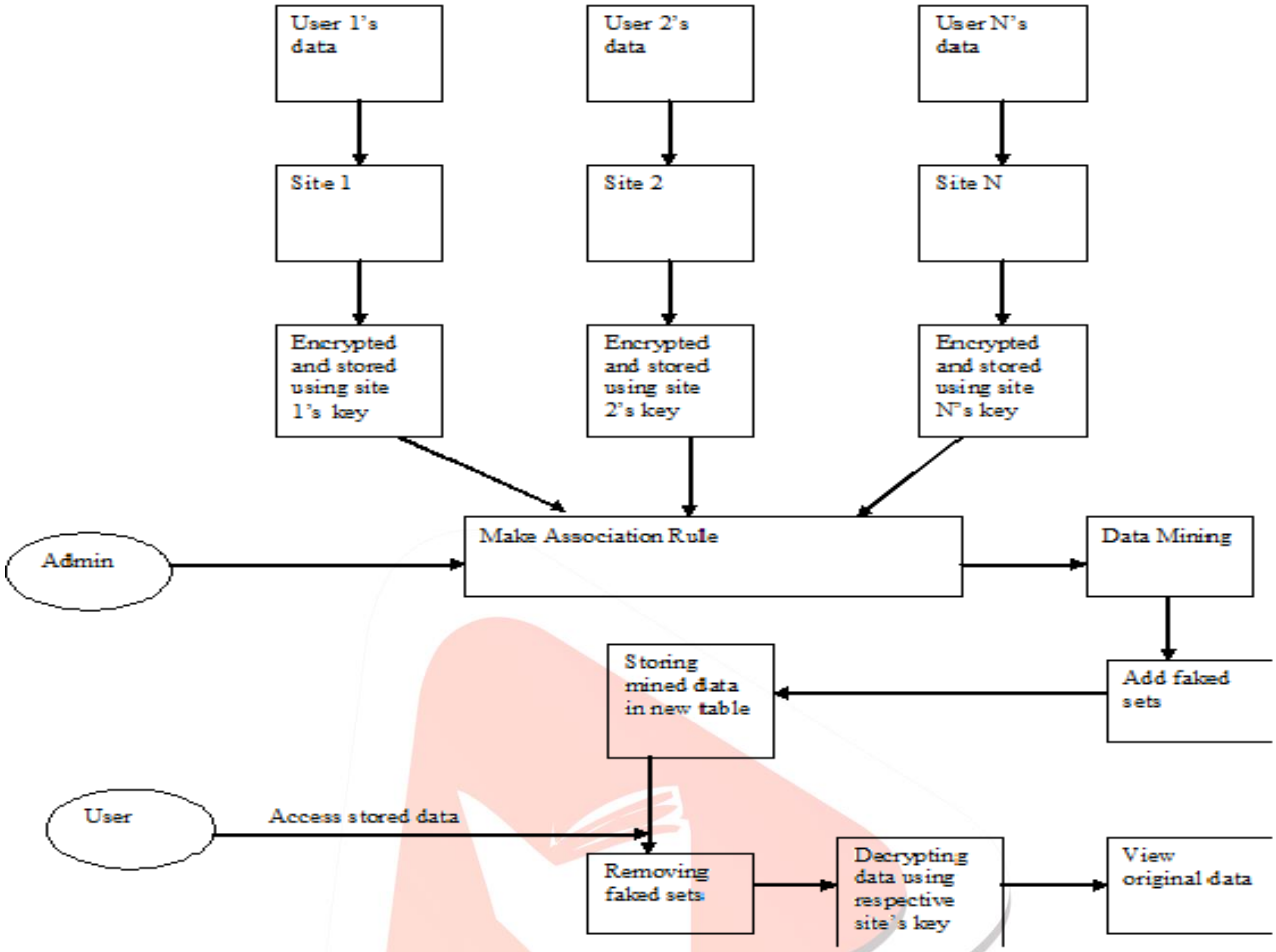
***Creating association rule***
For data mining, admin will follow FDM algorithm by making association rule. Union of all private subsets will be done by using association rule.(In our application association rule formed for the eligibility of a candidate for job.

***Securely storing private union of subset***
These subsets will be in encrypted earlier in the process when user does the (Registering academic details) process. It is made secure by adding faked item sets, so that the original count may not be revealed to any other third party. It will be stored in a different table (new table).

***Request to view securely mined data***
If the respective circles admin wants to view, mined data (shortlisted candidate) admin has to provide their circle's key to retrieve the data. While retrieving, the faked item sets that where during earlier process, will be removed.

***View original data***
After getting the orinal count, the data will be decrypted and admin will send mail to all candidates for interview. Users can also see this in the account by clicking my job tab.

**7. Architecture diagram**




**8. CONCLUSION**

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 subsetsthat each of intracting player hold. Those protocols exploits the fact that the underlying problem is of the interest only when the number of players is greater than two. In our protocol data leakage is reduced while retrieved the data from distributed database.

**REFERENCES**

[1] R.Agarwal and R.Srikant. fast algorithm for mining rules in large database. In VLDB, pages 487-499,1994
[2] R.Agarwal and R.Srikant. privacy preserving data mining.In SIGMOD conference, pages 439-450,2000.
[3] D. Beaver, S. Micali, and P. Rogaway. The round complexity of secure protocols. In *STOC*, pages 503–513, 1990.
[4] M. Bellare, R. Canetti, and H. Krawczyk. Keying hash functions for message authentication. In *Crypto*, pages 1–15, 1996.
[5] A. Ben-David, N. Nisan, and B. Pinkas. FairplayMP - A system for secure multi-party computation. In *CCS*, pages 257–266, 2008
[6] J.C. Benaloh. Secret sharing homomorphisms: Keeping shares of a secret secret. In *Crypto*, pages 251–260, 1986.
[7] J. Brickell and V. Shmatikov. Privacy-preserving graph algorithms in the semi-honest model. In *ASIACRYPT*, pages 236–252, 2005.
[8] D.W.L. Cheung, J. Han, V.T.Y. Ng, A.W.C. Fu, and Y. Fu. A fast distributed algorithm for mining association rules. In *PDIS*, pages 31–42.1996
[9] D.W.L Cheung, V.T.Y. Ng, A.W.C. Fu, and Y. Fu. Efficient mining of association rules in distributed databases. *IEEE Trans. Knowl. Data*