JES
Journal of Engineering Sciences

Vol 10, Issue 9, Sept /2019
ISSN NO:0377-9254

EFFICIENTLY EVALUATING IIDPS FORMULA ON ENCRYPTED CLOUD DATA

[1] Syeda Sadia Banu, [2]Nasira Mahjabeen
[1]PG Scholar, MTech, Dept of CSE, Shadan Women's College of Engineering and Technology HYD, T.S.
sadia.banu17@gmail.com
[2]Asst Professor, Dept of CSE, Shadan Women's College of Engineering and Technology HYD, T.S.

ABSTRACT Frequent itemset mining [FIM], is the activity in affiliation standard mining. The sensational increment on the range of datasets accumulated and set away with cloud benefits, it is promising to sustain this calculation escalated citation process. A measure of exertion likewise moved the inexact mining scheming to accurate calculation, where such techniques not simply expand the exactness additionally plan to upgrade the productivity. While mining information placed away on open mists, it presents security worries on delicate datasets. Extra system is established for implementing safety of FIM, where information is assembled and dug in an encoded structure in an exposed cloud facility. Three secure FIM conventions are planned over this structure. To certify information protection and evaluation productivity, two distinctive homomorphic encryption (HE) plans are claim and structure a safe and viable examination plot. Initial convention accomplishes progressively productive mining execution while subsequent convention gives additional grounded security ensure. To further streamline the exhibition of convention, we influence a minor exchange off protection to produce third convention. Finally, assess the appearance of agreements with extensive tests, the outcomes illustration that conventions clearly outflank past arrangements in execution over a comparable refuge level

Index Terms—Cloud computing, privacy-preserving, FIM, efficiency.

I INTRODUCTION

With the raise of enormous data, data mining has turned out to be a standout amongst the most rising methods for data analytics on massive datasets. Specifically, given a transaction database, where each tuple in this database is indicated as a transaction, frequent itemset mining can find popular thing sets and their potential intriguing associations from this transaction database. For instance, Netflix or RedBox can perform frequent itemset mining to find whether individuals watched Fast and Furious will also watch Mission Impossible so as to prescribe new-release films to costumers.

The easy accesses and low costs of open cloud administrations can significantly save mining costs on massive datasets. All the more importantly, since large-scale data from numerous and variable data sources are gathered by cloud administrations, for example, Google, mining on cloud data can also dramatically improve the accuracy and effectiveness of mining. Then again, since many datasets gathered by cloud administrations are touchy, mining these confidential data inevitably brings important privacy issues.

So as to safeguard privacy of frequent itemset mining in open cloud benefits, some randomization-based approaches[3],[4] have been proposed. Unfortunately, randomization-based approaches dramatically scarify the accuracy and utility of frequent itemset mining with constrained privacy guarantee. Pushing a stage ahead, Yi et al. [16]as of late proposed a privacy-safeguarding frequent itemset mining convention in open mists, where all the encoded transactions are centralized to the cloud and diggers delegate all the mining tasks to the cloud. Be that as it may, to enable frequent itemset mining, this work requires n (where $n \geq 2$) aided semi-legit servers to perform appropriated unscrambling amid the evaluation on scrambled data. In addition, the need of these additional n aided semi-genuine servers hinders the running time of frequent itemset mining, and presents tremendous interactions and communication overheads.

In this paper, a privacy-protecting framework is proposed for secure frequent itemset mining on encoded data, where just a single aided server (alluded to as Evaluator) is required other than the Cloud Service Provider (CSP). The CSP gathers encoded transactions and maintains a scrambled transaction database, and the Evaluator assists the CSP to carry mining on encoded data. Based on this framework, three conventions are designed. A comparison between conventions and this past work is exhibited in Table 1. The details of our contributions are listed as below:

- In the conventions, we guarantee that the Evaluator obtains no extra information except the frequent results (i.e., 1s or 0s) via blinding and adding dummy sets. While for CSP, we formulate three security levels based on different leakage functions in our conventions, where Privacy Level 1 (PL-1) includes the mining itemset, its support and the frequent result, Privacy Level 2 (PL-2) contains the number of the mining items and the frequent result, and Privacy Level 3 (PL-3) reveals just the frequent result.
- We present three practical privacy-preserving frequent item-set mining conventions in the paper. Protocol 1 is extremely efficient with the

PL-1, while second convention achieves a higher privacy level PL-3. In convention 2, a new efficient comparison convention is leverage to achieve more feasible efficiency. Furthermore, some anonymization methods to obtain Protocol 3, which is an enhanced version of Protocol 2 with PL-2 privacy.

- Experimental results demonstrate that our conventions are practical and compared to the most-related previous work the implementation of the results demonstrate that our conventions are more efficient than conventions with the similar privacy level.

II. RELATED WORK

The problem of frequent itemset mining was originally described by Agrawal et al.[10],[11], and has been extensively studied in recent years. Evfimievski et al.[12] proposed corresponding solution with association to vary randomization operators based on Apriori algorithm. The existing privacy preserving solutions can be classified into two categories: randomization-based solutions and cryptography-based solutions.

Randomization-based solutions. Several studies leveraged perturbation approaches to achieve privacy-preserving association rule mining[12],[15]. The main idea in these approaches is to sanitize all the samples prior to their release to achieve privacy preservation. Specifically, using the information on the distribution of the random data or some generalization techniques to distort the original data, accordingly it can generate an approximation to the original data distribution without revealing the original data values. The most popular techniques include adding "noise" into a transaction database, and leveraging k-support anonymity which re-quires that each item has a similar support with at least k-1 items. Another main approach is to use k-privacy, which requires each itemset (a set of items) cannot be distinguished from at least k-1 item sets with the same size. However, the randomization-based approaches have significant drawbacks, where it can just provide an approximate result and limited privacy levels.

Cryptography-based solution. Cryptography-based methods provide a well-defined security level and an exact mining result for privacy-preserving frequent itemset mining. Presented the corresponding conventions with data being vertically partitioned among different sites using secure scalar item conventions. In[4], the authors proposed a solution for the complementary situation where data is horizontally partitioned. With the prevalent of large-scale datasets, both of the two solutions are not efficient to satisfy the practical requirement. Then efficient conventions are proposed by leveraging Bloom filters. However, similar to randomized-based solutions, using Bloom filters would sacrifice the accuracy of mining results. Vaidya et al. transferred the problem of association rule mining to private set intersection, and successfully utilized a secure set intersection cardinality convention to achieve rule mining. Lai et al. provided a semantic security secure solution for outsourcing association rule mining with the two data privacy and mining result privacy, while the efficiency is not enough for the practical requirements. Later, Yi et al[16]. also proposed a secure association rule mining scheme in the same outsourcing distributed computing environment with a distributed ElGmal homomorphic encryption, while their scheme needs extra n(≥ 2) aided servers to collaborate with the cloud, and huge amounts of communication overhead are generated during this process.

III. PROBLEM STATEMENT

System Model

The framework model is shown in Fig. 1. There are n clients {U1, ..., Un}, an information excavator, a Cloud Service Provider (CSP), and an Evaluator. Every client transfers an exchange (or different exchanges) to the CSP. The CSP fundamental maintains an exchange database, which incorporates countless contributed from various clients. The information digger can submit visit itemset mining inquiries to the CSP, and we influence an Evaluator to help the CSP to perform visit itemset mining proficiently on the exchange database. Toward the finish of the executions, the CSP will restore a Boolean outcome (e.g., visit or not) of a mining question to the information digger.

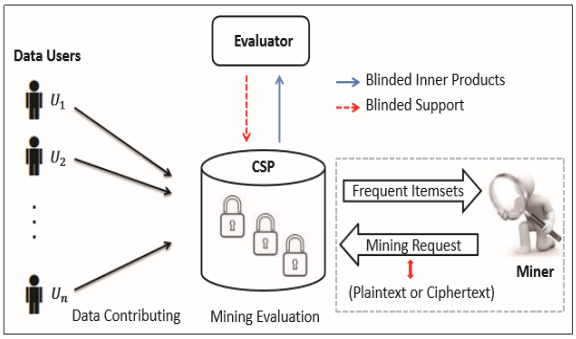


Fig:1 System model

Threat Model

In our model, clients don't completely confide in the CSP with the security of their exchanges. Consequently, every one of their exchanges are encoded before adding to the CSP. Both the CSP and the Evaluator are straightforward however inquisitive, which demonstrates them two can register information effectively by following conventions yet are interested

about information in the exchange database and consequences of successive itemset mining. Moreover, we expect that the CSP and the Evaluator don't connive. The information digger can submit visit itemset mining inquiries either in plaintext or in ciphertext because of various security necessities (i.e., regardless of whether the information excavator needs to uncover its questions to the CSP or not). Note that this non-conspiring model between the CSP and the Evaluator has been broadly utilized in numerous ongoing papers to effectively empower propelled calculations on enormous scrambled information.

	Model		Privacy	Efficiency	
	Servers	Miner		Computation	Interactions (Servers)
YRB's Protocol [10]	$n \geq 3$	Data Owner	Database Privacy	Slow	more than n
Our Protocol 1	two	Any Users	Item Privacy	Very Fast	One
Our Protocol 2	two	Any Users	Database Privacy	Slightly Fast	Two
Our Protocol 3	two	Any Users	Database Privacy	Fast	Two

Table 1: Comparison among our protocols and YRB's protocol

Privacy Objectives

In our paper, other than asking the CSP and the Evaluator to figure visit itemset mining productively and accurately, another principle target of our own is to protect the security of exchanges and the mining result. Casually, in the wake of performing successive itemset mining, with the exception of realizing whether a question is visit or not, the CSP or the Evaluator can't get familiar with the crude information in the exchange database. The substance of the incessant itemset mining inquiry ought to likewise be protected from the CSP or the Evaluator if the information digger presents its question secretly. Besides, precise backings or confidences of the mining itemset ought to be hidden in that capacity data can be utilized to gather some data about the crude information. Since various security necessities perhaps happen in our framework, we therefore define three dimensions of protection dependent on a spillage work L which catches all the data that are permitted to spill amid the assessment on the encoded information in a convention.

IV. SECURE FREQUENT ITEMSET MINING

The CSP holds a lot of scrambled exchanges, and a Miner might want to mine frequent itemset on this encoded exchange database. Every exchange is represented as a binary vector, and a mining question as another vector. We state an exchange fulfills a mining question if the internal result of this exchange and this mining inquiry is equivalent to the quantity of 1s in the mining inquiry. The CSP and the Evaluator can intuitively figure and choose the help of a mining inquiry on an exchange database. In view of various security necessities, we presently portray three conventions in the accompanying part to perform visit itemset mining on scrambled information, where the main convention centers around productivity while the second one is solid as far as protection insurance. At that point, we further enhance and improve the presentation of our second convention to get the third one with a minor tradeoff as far as security.

1. Protocol 1

We initially consider Protocol 1 where a Miner presents its mining question in plaintext while the exchanges are still encoded. CSP gets the scrambled exchanges from information clients and the plaintext mining inquiry from the Miner, and after that process the inward results of them. In this way, the added substance homomorphism Paillier encryption is enough in Protocol 1. Subtleties of this convention are appeared in Fig. 3.

Subtleties. In particular, the Evaluator initially produces a couple of open/private keys {PK, SK} for Paillier in Step 1. At that point, every client encodes its exchanges with the open key of the Evaluator and contributes its scrambled exchanges to the CSP in Step 2. Given a mining inquiry, the CSP registers a scrambled inward result of each encoded exchange and the mining question by calling Inne ProPC presented in Subsection 4.3 as a subroutine. So as to make sense of whether each encoded inward item fulfills the prerequisite, the CSP associates with the Evaluator by utilizing blinding. In the meantime, to shroud the genuine help of this mining question on the scrambled exchange database from the Evaluator, the CSP additionally haphazardly creates some sham encoded exchanges, assesses scrambled internal results of these sham encoded exchanges with the mining inquiry, rearranges these encoded inward items with the un dummied ones and sends them all together to the Evaluator. The Evaluator restores a dummied help to the CSP.

2. Protocol 2

To accomplish a higher security, we presently plan our second convention, which enables the Miner to present an encoded mining inquiry. Subsequently, the CSP and the Evaluator will cooperatively assess the encoded mining support through computing inward items over the scrambled information, and afterward choose whether a mining question is visit by assessing examination on encoded information. We influence InnerProCC and Compare presented in Sec. 4 to satisfy these undertakings.

Step 1: System Setup. The Evaluator generates two key pairs for data encryption and support evaluation as

$$\{\mathsf{PK}_{\mathsf{Enc}}, \mathsf{SK}_{\mathsf{Enc}}\} \leftarrow \mathsf{BGN.KeyGen}(1^{\lambda}),$$

$$\{\mathsf{PK}_{\mathsf{Eva}}, \mathsf{SK}_{\mathsf{Eva}}\} \leftarrow \mathsf{Paillier.KeyGen}(1^{\lambda}).$$

and then outputs $\mathsf{PK}_{\mathsf{Enc}}, \mathsf{PK}_{\mathsf{Eva}}$ and keeps $\mathsf{SK}_{\mathsf{Enc}}, \mathsf{SK}_{\mathsf{Eva}}$ private.

Step 2: Data Processing. Each user U_i encrypts its transactions $t_i = (e_{i,1}, \cdots, e_{i,n})$ with BGN cryptosystem as

$$c_{i,j} = \|e_{i,j}\|_{\mathsf{PK}_{\mathsf{Enc}}},$$

where $i \in \{1, \cdots, m\}$ is the number of transactions, and $j \in \{1, \cdots, n\}$ is the dimension of each transaction, and gets $C_i = (c_{i,1}, \cdots, c_{i,n})$, then sends $C = \{C_1, \cdots, C_m\}$ to CSP.

Step 3: Computation.

- Given a mining itemset $q = (s_1, \cdots, s_n)$, the Miner computes
$$r_j = \|s_j\|_{\mathsf{PK}_{\mathsf{Enc}}}, \qquad z = \|\|q\|_1\|_{\mathsf{PK}_{\mathsf{Enc}}},$$
and sends $\{Q, z\}$ to CSP, where $Q = (r_1, \cdots, r_n)$ and $||q||_1$ is the length of $q = (s_1, \cdots, s_n)$.
- Similar to Protocol 1, CSP follows InnerProCC to calculate
$$x_i = \|q \circ t_i\|_{\mathsf{PK}_{\mathsf{Enc}}}$$
with the inputs Q and C, and then selects a dummy transaction set D to compute
$$y_j = \|q \circ d_j\|_{\mathsf{PK}_{\mathsf{Enc}}}$$
with the additive homomorphism of BGN.
- As Protocol 1 in Step 3, CSP calculates a randomized set $W = \{\omega_1, \cdots, \omega_{m+k}\}$ and computes
$$W' = \pi(W)$$
with a secret permutation function π. Finally CSP sends $W' = \{\omega'_1, \cdots, \omega'_{m+k}\}$ to the Evaluator.

Step 4: Evaluation.

- Given W', the Evaluator computes (under Paillier)
$$v'_i = \begin{cases} [1]_{\mathsf{PK}_{\mathsf{Eva}}}, & \mathsf{BGN.Dec}(\mathsf{SK}_{\mathsf{Enc}}, \omega'_i) = 0, \\ [0]_{\mathsf{PK}_{\mathsf{Eva}}}, & \text{otherwise}, \end{cases}$$
and then returns $V' = \{v'_1, \cdots, v'_{m+k}\}$ to CSP.
- Once receiving V', CSP computes
$$V = \pi^{-1}(V'),$$
where $V = \{v_1, \cdots, v_{m+k}\}$, and then removes the dummy results and calculates
$$[supp(q)]_{\mathsf{PK}_{\mathsf{Eva}}} = \prod_{i=1}^{m} v_i.$$

Step 5: Comparison. Receiving the encrypted minimum support from the Miner, CSP and Evaluator together follow Compare on two encrypted values $[supp(q)]_{\mathsf{PK}_{\mathsf{Eva}}}$ and $[minup]_{\mathsf{PK}_{\mathsf{Eva}}}$. Finally, CSP outputs 1 if
$$supp(q) \geq minup,$$
otherwise outputs 0.

Inside Intrusion Detection and Protection System (IIDPS)

The IIDPS, consists of an SC monitor and filter, a mining server, a detection server, a local computational grid, and three repositories, including user log files, user profiles, and an attacker profile. Our enhanced protocol achieves more efficient mining performance with stronger privacy guarantee.

Algorithm: The algorithm for generating a user habit file by using frequent item-set mining

Input: U's log file where u is a user of the underlying system

Output: U's habit file

```
G=|log file|-|Sliding window|;
for(i=0;i<=G-1;i++) {
for(j=i+1;j<=G;j++) {
for (each of the logging done by the user)
{
If (user opens specific link)
{
then (user action==1)
}
else
{
(user action==0)
}
If(the identified pattern already exists in the habit file)
Increase the count of the pattern by one;
}
else
Insert the U's pattern in to the file with count=1;}}}
```

V. IMPLEMENTATION

MODULE EXPLANATION

- **User Interface Design**

To attach with server user must give their username and password. client as of now exits straight forwardly can sign in, else client must enrol their subtleties. The record made in database. Marking in is normally to enter a page. It will glance over the inquiry and demonstrate the request.

- **Cloud service Provider (Data Miner):**

This module will speak to as a database, the material which we are transferring should store in CSP. The customer moves the report that will be secured in a prearranged association in the veritable method for the endeavor workspace. The CSP with the assistance of Evaluator accumulates assorted trades and keeps encoded database in certifiable manner, to dispatch taking out on prearranged data.

Journal of Engineering Sciences
Vol 10, Issue 9, Sept /2019
ISSN NO:0377-9254

➢ **User:**

This is the third module here, Coordinator, BDM, and Employee goes about by means of a client. The BDM statistics as noteworthy data and applying encryption shows arranged data moved by him. BDM can move 'n' quantity of archives.

➢ **Encryption & Decryption Phase:**

The records moved by the BDM mixed and set away which keeps up trade database. Clients don't completely assurance of CSP with the protection of exchanges. Every one of their exchanges is encoded before adding in CSP. CSP & the Evaluator are straightforward however inquisitive, which shows both can register information accurately by following conventions yet are attracted in guidance in trading database & aftereffects of FIM.

➢ **Frequent Itemset mining:**

Information digger will effort to get the outcomes from database (DB). This is a champion midst the most basic data mining strategies, & it discover relations within different things in monstrous datasets. we will get FIM with the settlements 1 &2 as '0's &'1's.

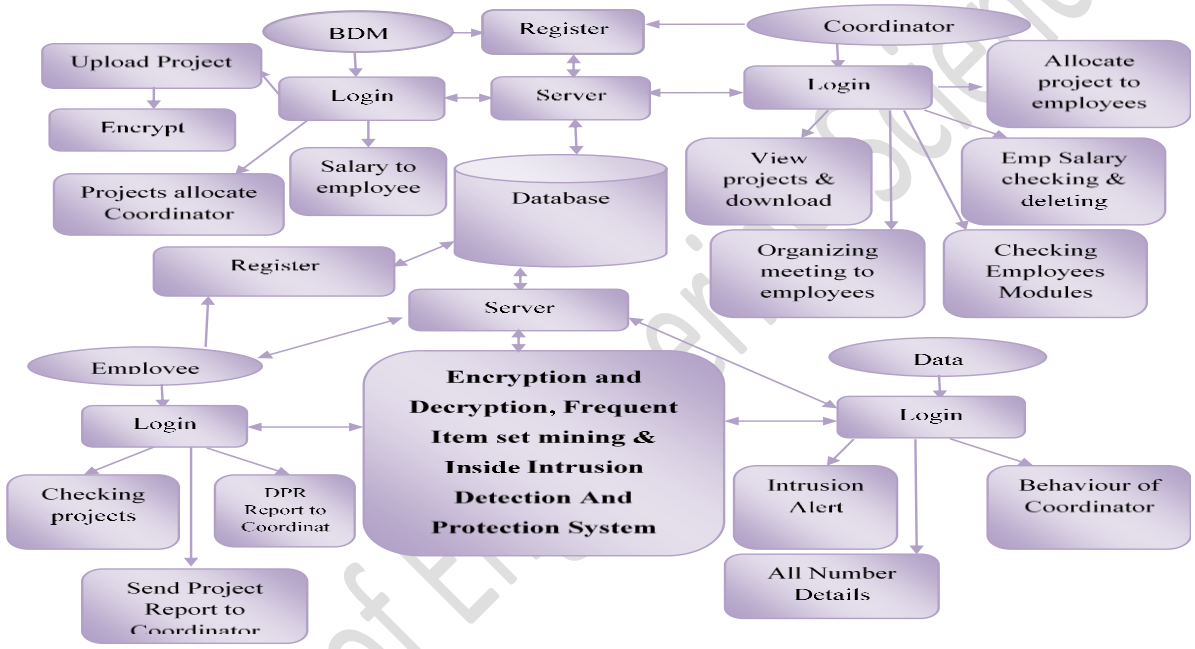
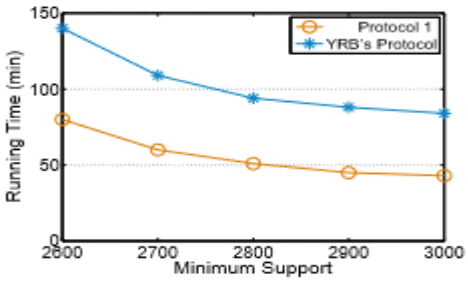


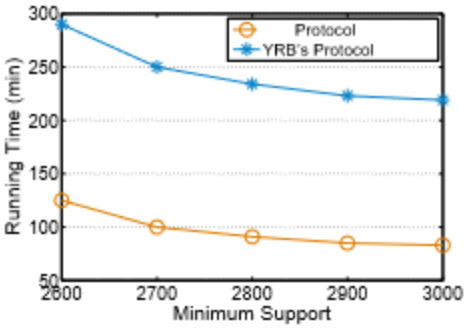
Fig 2: System Architecture

Coordinator, BDM, and Employee goes about as a client. The BDM material is essential data and applying encryption conventions. Clients don't completely confess CSP through the secrecy. Subsequently, everyone exchanges are scrambled before providing to the CSP. CSP and Evaluator are straightforward however inquisitive, which demonstrates both can figure knowledge effectively by following conventions yet are interested about documentation in the interchange DB & aftereffects of FIM. Data digger will get the result from the DB by FIM. It is comprehensively castoff to realize relations amid things in monstrous informational indexes.

VI. RESULTS:



Item privacy



Database privacy

Comparison the performance between our protocols and YRB's protocols with different security levels, where the size of dummy transactions in both two protocols are m/2.

VII. CONCLUSION AND FUTURE WORK

In this paper three common-sense safety saving FIM conventions are anticipated on scrambled cloud information. Protocol 1 accomplishes an amazingly higher mining execution while Protocol 2 gives a grounded protection ensure. Toward increasingly effectiveness, we upgrade the exhibition of our subsequent convention by exploiting a minor spillage of security to realize Protocol 3. Exploratory outcomes demonstrate that conventions are considerably more operative than earlier work over a analogous security level.

In our future work, Prominence on further refining the regulation of FIM on greater scale DB. Furthermore, due to the discrete assets of streaming data and its extensively applied application and concentration on the learning of the well-organized discretion- preserving FIM algorithm on stream data.

REFERENCES

[1] S. Brin, R. Motwani, J. D. Ullman, & S. Tsur, "Dynamic itemset counting & implication rules for market basket data," in ACM SIGMOD Record, vol. 26, no. 2. ACM, 1997, pp. 255–264.

[2] B. Mobasher, N. Jain, E.-H. Han, & J. Srivastava, "Web mining: Pattern discovery from world wide web transactions," Technical Report TR96- 050, Department of Computer Science, University of Minnesota, Tech. Rep., 1996.

[3] J. Vaidya & C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery & data mining. ACM, 2002, pp. 639–644.

[4] M. Kantarcioglu & C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Transactions on Knowledge & Data Engineering, no. 9, pp. 1026–1037, 2004.

[5] J. Vaidya & C. Clifton, "Secure set intersection cardinality with application to association rule mining," Journal of Computer Security, vol. 13, no. 4, pp. 593–622, 2005.

[6] T. Tassa, "Secure mining of association rules in horizontally distributed databases," Knowledge & Data Engineering, IEEE Transactions on, vol. 26, no. 4, pp. 970–983, 2014.

[7] J. Lai, Y. Li, R. H. Deng, J. Weng, C. Guan, & Q. Yan, "Towards semantically secure outsourcing of association rule mining on categorical data," Information Sciences, vol. 267, pp. 267–286, 2014.

[8] R. Curtmola, J. A. Garay, S. Kamara, & R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions & Efficient Constructions," in Proc. of ACM CCS'06, 2006.

[9] X. Yi, F.-Y. Rao, E. Bertino, and A. Bouguettaya, "Privacy-preserving association rule mining in cloud computing," in Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security. ACM, 2015, pp. 439–450.

[10] R. Agrawal, T. Imieli´nski, and A. Swami, "Mining association rules between sets of items in large databases," in ACM SIGMOD Record, vol. 22, no. 2. ACM, 1993, pp. 207–216.

[11] R. Agrawal, R. Srikant et al., "Fast algorithms for mining association rules," in Proc. 20th int. conf. very large data bases, VLDB, vol. 1215, 1994, pp. 487–499.

[12] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," Information Systems, vol. 29, no. 4, pp. 343–364, 2004.

[13] S. R. Oliveira and O. R. Zaiane, "Privacy preserving frequent itemset mining," in Proceedings of the IEEE international conference on Privacy, security and data mining-Volume14. Australian Computer Society, Inc., 2002, pp. 43–54.

[14] Y. Saygin, V. S. Verykios, and A. K. Elmagarmid, "Privacy preserving association rule mining," in Research Issues in Data Engineering: Engineering E-Commerce/E-Business Systems, 2002. RIDE-2EC 2002. Proceedings. Twelfth International Workshop on. IEEE, 2002, pp. 151– 158.

[15] S. J. Rizvi and J. R. Haritsa, "Maintaining data privacy in association rule mining," in Proceedings of the 28th international conference on Very Large Data Bases. VLDB Endowment, 2002, pp. 682–693..

[16] X. Yi, F.-Y. Rao, E. Bertino, and A. Bouguettaya, "Privacy-preserving association rule mining in cloud computing," in Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security. ACM, 2015, pp. 439–450.

AUTHOR'S PROFILE

Ms. SYEDA SADIA BANU has completed her B.E from ISL women's engineering college, OU Hyderabad. Presently, she is pursuing her Masters in Computer Science from Shadan Women's College of Engineering and technology, Hyderabad, TS. India.

Ms. NASIRA MAHJABEEN has completed B.Tech (CSE) from Dr. V. R. K college of engineering and technology, JNTUH University, Hyderabad, M.Tech (CSE) from Shadan women's college of engineering and technology, JNTUH University, Hyderabad , Currently she is working as an Assistant Professor of CSE Department in Shadan women's college of Engineering and technology, Hyderabad, TS. India.