Privacy Preserving Data Mining Using Secure Multiparty Computation Based On Apriori and Fp-Tree Structure of Fp-Growth Algorithm

P.Yoganandhini, G.Prabakaran

Abstract: *In this work, a method is proposed to deal with secure multiparty computation (SMC) based problems. The computation is done on the grocery dataset collected from three various grocery shops. The privacy is maintained by generating the rules based on FP-Tree algorithm under Association Rule Mining (ARM). Privacy and correctness are the important requirements of SMC. In privacy requirement, the things apart from necessary are not learned. This implies that only output will be learned by the parties. Each party must receive correct output to ensure the correctness. In this work, secure auction is done using SMC and frequent item sets are computed to perform the association rule mining. The most familiar FP-growth schemes have the short fallings like former space complexity and latter time complexity. The performance of the algorithms has been enhanced by using APFT algorithm which is a combined version of FP-tree structure of FP-growth algorithm and Apriori algorithm. The conditional and sub conditional patterns are not generated continuously in APFT. The speed of the APFT is high when compared to Apriori algorithm and FP-growth.The correlated items are included by modifying APFT and non-correlated item sets are shaped by using APFT. This modification is used for FP-tree optimization. From the frequent item set, the loosely associated items are removed by using this modification. The system implemented is clearly described and its performance is evaluated. The results confirmed that the proposed scheme is extremely effective.*

Keywords: *Apriori algorithm, Association Rule Mining, Distributed Data Mining, Frequent Pattern Growth, Privacy Preserving Data Mining (PPDM), Secure Multiparty Computation (SMC).*

I. INTRODUCTION

Data mining is an as of late rising field, interfacing the three universes of databases, Artificial Intelligence and Statistics. The excess amount of information can be collected from the aged data. The data are distributed with various amount of security and privacy to the various parties for data mining works. Different methods like classification, Association Rule Mining (ARM), clustering exception discovery and regression are used in this filed. The statistical data analysis uses the K-Means clustering for obtaining better results. Mining of potentially helpful information of high volume data can be studied with help of data mining for various applications.

Data mining and discovery of knowledge in databases are two novel study fields that examine the automatic extraction of earlier unfamiliar patterns from huge volumes of data. Privacy Preserving Data Mining (PPDM) is a new study in statistical databases, data mining in which the results are scrutinized for the side effects. The twofold is a key in preserving privacy of the data in mining. The private data are trimmed out of the real dataset. It will enable receiver to not compromise on privacy. The data mining algorithms are used to extract the private information from dataset to disqualify it [1]. In PPDM, various techniques are implemented to secure the private data and sensitive data after mining process. The personal data are left confidential. Moreover, there is a necessity to guard insightful knowledge all through a data mining practice. This is known as PPDM. The resources of data are spread over various locals in Distributed Data Mining (DDM) model. Proficiently attainment of the mining can be done using various algorithms. The most of the proposed techniques are not considering security as an issue, mostly they concentrate on improving efficiency. The data can be distributed by vertically or horizontally partitioning it.

PPDM is a broad area of research [2] in public and private sector [3]. It is also discussed in various international conferences and workshops [4]. Current research concentrates on the progress of technical schemes like application of cryptography to enhance the privacy and security requirements of data mining approaches including categorization and classification.

While running the data mining schemes on a secured data, the data should not be known to the running party of the algorithm. This issue is considered in PPDM. It must be noted that there are two classes in PPDM. In first class, mining algorithm is run on the union of the database of the parties without knowing other parties data. Initially data is divided into two or more parties' data. Second class releases some statistical data. The statistical data consist of confidential data. This dataset is modified to ensure the privacy and to get a better results. First class of privacy-preserving is concentrated in this work.

PPDM can be applied in medical database mining, customer transaction analysis and in homeland safety. The applications involving medical and bio-terrorism will intersect in scope. The presented method uses the horizontally partitioned data to mine and the privacy are not violated. It also demonstrate how to use data mining techniques to preserve privacy. The sharing information is minimized by using cryptographic techniques. It is done by multiparty security model.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author

P.Yoganandhini*, Research Scholar, Department of Computer and Information Science, Annamalai University.
Dr.G.Prabakaran, Associate Professor, Department of Computer Science and Engineering, Annamalai University.

The proposed method concern about the use of horizontally partitioned data to compute secured mining of association rules. The proposed scheme, presume that few users are permitted to notice some of the data, just that no one is permitted to notice the entire data. The personal information has to be secured. This is the major aim of data privacy. The information can be linked to any individual person. If the individual persons data are under mining then that has to be secured. This paper implements an protection method which some patterns and trends.

II. RELATED WORK

The distorting of data can be done by randomization method. This method is completely dependent on the probability distribution. The distorting of data in surveys has a severe effect of privacy [5,6]. This randomization method can also be employed in the PPDM [7].

Consider a server with lot of users. All the users are having some amount of data. The server collects all the data and construct the aggregated data model by using data mining. The users can randomly disturb their data before sending it to server in randomization approach [6]. The users can take true information out and noise can be introduced. The aggregates required for data mining can be recovered by applying statistical estimation on those noisy data. The random values are multiplied or added with the real items to introduce a noise or it can be done by deleting some real values and introducing wrong values in the records [8,9]. It is possible to estimate the aggregate model with elevated accurateness by using proper amount of randomization and proper method.

The classical secrecy framework, revelation risk and damage measures in case of statistical databases has the definition of privacy in order characterize randomization [10, 11]. But it is defined well in recent frameworks [12-14]. The knowledge of the data miner is modelled to follow probability distribution in order to deal with randomization uncertainty. The major advantage is that,, it is enough to study the randomization algorithm to guarantee privacy and no need to learn about data mining operations. But the rules in this are approximate in the sense it requires high amount of randomized data to get a highly accrued results [15].

The techniques like suppression and generalization are used to reduce pseudo-identifiers granular representation in -anonymity techniques [16]. In case of generalization, attribute values are completely generalized for the purpose of lessening granularity of representation within a range. For instance, year of birth can be used to generalize date of birth to reduce risk in identification. The value of attributes are removed in suppression method.

The reduction of identification risk can be done by using public records but it reduces the application accuracy in transformed data. The sensitive data are suppressed before the occurrence of computation or disclosure in order to preserve privacy. This suppression process is difficult if the suppressions of data are dependent or there exist a dependency among suppressed and disclosed data. If the full access of the sensitive data is required by the data mining techniques, then it not possible to implement suppression. Certain statistical characteristics are protected against discovery by using suppression. It minimizes the data mining results of the other distortions. Most of the optimization problems are intractable computationally [17-19].

III. DISTRIBUTED DATA MINING

The important data from the collection of large data set can be extracted by using data mining techniques. This process involves large number of nodes. The data from various resources can be collected by a single authority by using the process called Data warehousing. This technique may increase the violation of privacy. The privacy concern makes user to not share the information directly on the network. Distributed data mining schemes are employed for the purpose of resolving this issue. In this technique, data will be disclosed only after it gets published.

IV. PROPOSED METHODOLOGY

The proposed methodology deals with the problem based on secure multiparty computation carried out with grocery detailed datasets of three different grocery shops and uses Apriori algorithm and FP-tree structure of FP-growth algorithm (APFT) under association rule mining to generate rules to generate rules and the final steps to maintain privacy.

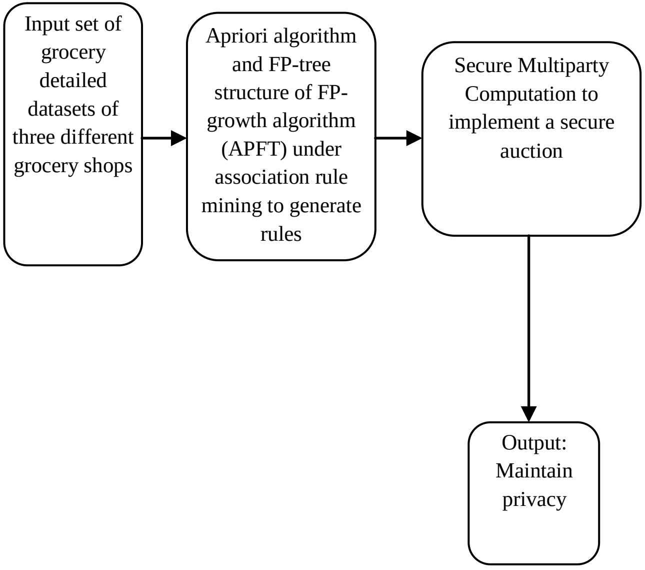

Input set of grocery detailed datasets of three different grocery shops
Apriori algorithm and FP-tree structure of FP-growth algorithm (APFT) under association rule mining to generate rules
Secure Multiparty Computation to implement a secure auction
Output: Maintain privacy

Figure 1: Methodology Architecture

Mining of some patterns in the large database frequently has a high impact in the broad applications and data mining tasks. Test methods and a priori-like candidate-generation are methods used by the existing approaches. The dataset with long and prolific patterns cannot be processed by using this methods.

Here, a class of efficient pattern-growth scheme is formulated for resolving the about said issue. Databases and mining tasks are decomposed by using divide-and-conquer approach.Pattern-growth methods like FP-growth are efficient and scalable. It is a fast frequent pattern mining scheme. Frequent pattern mining has a serious impact in data mining tasks like mining association rules. Following contributions are made.

Here, systematically formulated a pattern-growth scheme for frequent pattern mining. A novel scheme with the assistance of FP-growth is formulated to have high efficiency in reported problem.

A.Frequent Pattern Growth (FP-Growth) Algorithm

ARM depends on frequent

pattern mining. The items occurs more frequently above the threshold in a given set of transaction can be found by frequent pattern mining. FP-tree algorithm is a fast and famous algorithm infrequent pattern mining [24].

B.Fp-Tree Construction

A prefix tree of transaction is called FP-tree. Same prefix is shared by the every path of tree and thy indicates a collection of transactions. One item will be made correspond to one node. The nodes having same item are linked to make the process of finding transactions of a specific item. The items can be counted by list traversing. The list has head element to record the number of occurrence of a specific item in a dataset [24]. FP-tree is revealed in figure 2

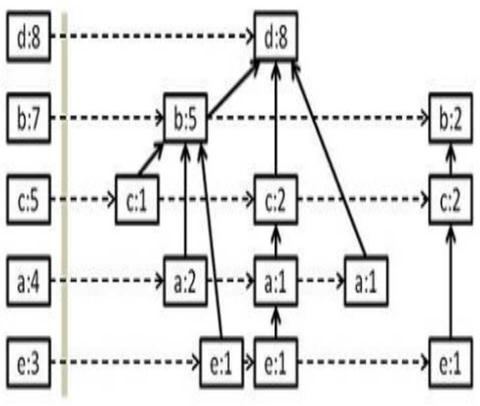


Figure 2: FP-Tree

Consider I as item set and database of transaction as B = {T1, T 2, …, T n }, in which T i indicates a transaction which has item set I. In case of a frequent pattern A, when A's support is not below a predefined minimum support threshold S. Provided aDB and S, computation of frequent patterns is given as frequent pattern mining problem.

C.Frequent Pattern Mining Problem

This section formally introduces association rule mining problem and defines the most importantterms in this field. It is defined as follows.

The rule mining is given as: consider set of n binary elements $I = \{i_1, i_2, \dots, i_n\}$ named as items and set of transactions D= {t 1, t 2, …, t n }$D = \{t_1, t_2, \dots, t_m\}$ named as database. Unique transaction ID is given to all transactions in D andit ha item subset in I. A rule is characterized as an inference of the form $X \Rightarrow Y$where $X, Y \subseteq I$, and $X \cap Y = \emptyset$. The collection of items X and Y are recognized as antecedent and consequent of the rule.

Take all items in transaction as $I = \{i_1, i_2, \dots, i_n\}$. Maximum number of items is given by n. A subset item is called as k-itemset if it not empty and has the item length as k. An association rule is an inference of the form $X \rightarrow Y$ where both X and Y areitemsets, and there exists no item that appears in both and formally,X, $Y \subseteq I$and $X \cap Y = \emptyset$. An association rule has two features: Confidence and support. Thesupport of the rule $X \rightarrow Y$ equals itemset XY's support. The confidence, denotedwith C, indicates the proportion of the transactions in the database including X and Y. It is regarded as a conditional probability, P(Y |X).

In order to lessen search space and for the purpose of discovering only those rules which can be attractive, two thresholds are established, minimum support and confidence thresholds. Consider, an itemset is frequent in case when its support surpasses a user-defined minimum support threshold, α min. Number of transactions Nis used to convert support and minimum support thresholdinto integer as they are given in percentage. It counts the transaction occurrence in itemset and integer minimum occurrence threshold can be compared with it in a simple way.

D.Formation of APFT algorithm

Han et al [21] mined frequent patterns by hyper structure mining and to have a high pattern growth mining efficiency, prefix tree structure with array-based implementation is used. APFT method is proposed to overcome the difficulties [22]. APFT uses two step procedure. Initially, FP-tree is built and Apriori algorithm is employed for mining frequent items. Here, APFT is included with correlation concept to reduce association rule with minimum support however also contain liner associations among them. The computational outcomes authenticate the superior performance of APFT with correlation algorithm.

E.Correlation Concept

It can be used on transaction databases with the subsequent changes. Consider an item 'a' is related with item 'b' when it convince the subsequent stipulations: P (ab) >P(a)P(b). At this point, P(ab) = probability of items 'a' and 'b' taking place jointly in the transaction database i.e. both 'a' and 'b' come about mutually/overall number of transactions. P(a) =The amount of transactions in which 'a' takes place/overall transactions. P(b) = The amount of transactions in which 'b' takes place/overall transactions [9]. As a result, the formula fundamentally corresponds to Observed probability > Expected Probability. This stipulation is said to be positive correlation among items 'a' and 'b'.

F.APFT with correlation

The frequent itemset will have the size of 2 in order to use correlation. The support of each branch is computed during the construction of calculation N Table. In this approach, tree need not be traversed again. This is a most efficient way of calculating support.

```
Algorithm APFT [ ]
Input: FP-tree, minimum support threshold ∋
Output: the entire frequent itemset
Apriori$(T, \epsilon)$
$L_1 \leftarrow \{large\ 1 - itemsets\}$
$k \leftarrow 2$
```

```
$while\ L_{k-1} \neq \emptyset$
$C_k \leftarrow \{a \cup \{b\} \,|\, a \epsilon L_{k-1} \wedge b \notin a\} - \{c \,|\, \{s\} s \subseteq c \bigwedge |s| = k-1\} \nsubseteq L_{k-1}\}$
for transaction $t \epsilon T$
$C_t \leftarrow \{c | c \epsilon C_k \wedge c \subseteq t\}$
for candidates $c \epsilon C_t$
$count\ [c] \leftarrow count\ [c] + 1$
$L_k \leftarrow \{c | c \epsilon C_k \wedge count\ [c] + 1$
$k \leftarrow k + 1$
$return \bigcup_k L_k$
```

G. Proposed SMC Protocol

A truly secure SMC protocol uses only input and output information or any information polynomial computable from it. Utmost care should be taken as it leads to privacy breach. Assume securecomputing of sum of local numbers. The input of other party can be calculated by sum value. As a result, it's essential for the purpose of investigating whether the input and output taken mutually possibly will disclose excessive information.

When a protocol is formulated, the system have to establish its security. For the reason that the entire interaction takes place through the messages transmitted and received, the system reproduces the views of the entire parties through simulating the consequent messages. When the system can reproduce these messages, subsequently the system can effortlessly simulate the complete protocol simply by running it.

The execution of protocol on same data multiple times will transfer various messages. This is based on the random choice adapted by the party. The generated message are indistinguishable with the messages generated by simulator then this protocol is said to be secured.

H.Multi-Party Computation

In Multi-Party Computation (MPC), a given number of participant's $p_1, p_2, \ldots, p_N$ each include a private data, correspondingly $d_1, d_2, \ldots, d_N$. The user need to work out the public function F value on N variables at the point $(d_1, d_2, \ldots, d_N)$.

If the users are not able to learn anything from public function then MPC protocol is said be more secured. The model used defines the amount of data can be learned from the data uploaded by same user.

The security of an MPC rely on:

- It can be either computational or unconditional.
- The model assumes the use of synchronized network. The channel between pair of users are secured
- Adversary may be passive or active which is controlled centrally.
- The same may be dynamic or static. The attainment of security in dynamic adversary is so difficult.
- Adversary may be complex of threshold structure.

Secure MPC offers solutions to real-life complications like contribution of signature or decryption functions, private information retrieval, circulated voting, private bidding and auctions, , , etc.,

I.Secure Multiparty Computation to maintain privacy

Secure Multiparty Computation (SMC) can successfully safeguard the susceptible data. SMC believes a collection of associates to jointly mine their data. The distributed PPDM complication can be decreased to the secure computation of a function dependent on distributed inputs and is consequently solved through using cryptographic shemes.

In case of SMC, every party contributes to the computation of the protected function by offering its confidential input. A protected cryptographic protocol that is implemented among the collaborating parties makes sure that the private input that is contributed through each party is not revealed to the others.

V. EXPERIMENTAL SETUP AND RESULTS

The process of mining association rules in case of the market basket data [20] is a core knowledge discovery action. It enables the methods to find the correlations between items belonging to customer transactions. The distributed mining of association rules has sites with standardized schema for records. It includes of transactions. Association rules are extracted by combining all the transactions and applying association-rule mining algorithm. This provides some amount of security but users may learn regarding the transactions of other users. But it is required to extract the association rules to ensure the confidence and support of parties' transactions [23]. By using global information, association rules can be computed without disclosing individual transactions.

An example from supermarket domain is used to demonstrate the concept. The item set is given by I = {milk, bread, butter, beer}, which corresponds to the rule that the customers buying milk and bread also likes to buy butter. This is a small example but practical cases, support from tons of transactions are needed to get a significant results. The dataset also has millions of transactions.

Measures of consequence and significance are employed to select the rules from available rules. The minimum threshold on support and confidence are most commonly used measures. The transactions in the data set which has item-set defines the support supp(X) of an itemset X. For example, consider an item-set with support of 1 / 5 = 0.2 which corresponds to the occurrence in 20% of transactions.

The confidence of a rule is given as

$$Conf\ (X \Rightarrow Y) = supp(X \cup Y) / supp(X)$$

VI. RESULTS AND DISCUSSIONS

Cryptographic tools is used to perform data mining. The developed project has provided processes to mine distributed association rules. It uses horizontally partitioned data.Two party cases pose difficulties. If the system have only two parties, one party part does not support global rule then the other party will support it.

The Table 1 shows that computation by secure multiparty is better for mining largedatabase than the cryptographic technique such as encryption and decryption. The FP TreeAlgorithm provides better result for the multiparty case than a two party case's. In this paper, the final output demonstrated that distributed ARM can be implemented resourcefully under reasonable security suppositions using secure multiparty computation.

Table 1: Comparison Of Cryptographic Techniques

Data Sets	**Apriori algorithm and FP-tree structure of FP-growth algorithm (APFT) under association rule mining to generate rules**			
	Time taken for rule generation		**Memory Utilization**	
	Encryption and Decryption (in seconds)	SMC (in seconds)	Encryptionand Decryption (in bytes)	SMC (in bytes)
Grocerydata 1 (Real time data from Chinthamani super market, Erode)	4.99	0.52	2492	258
Grocerydata 2 (Real time data from Sakthi super market, Erode)	5.06	0.17	2531	86
Market Basket (From UCIRepository, website)	2.89	0.16	1445	78

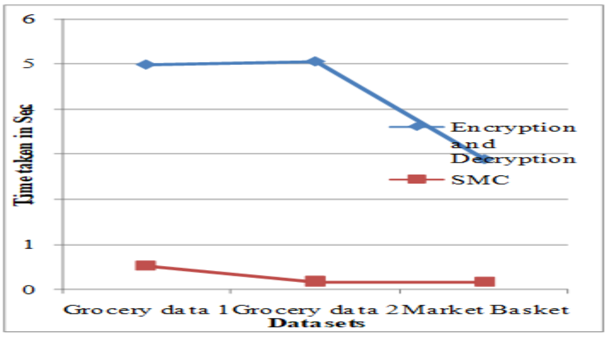


Figure 3: Comparison graph of Time taken for rule generation

From the Figure 3, it can be explained that the proposed Apriori algorithm and FP-tree structure of FP-growth algorithm (APFT) under association rule mining to generate rules based SMC approach provides a better performance by taking minimum time.

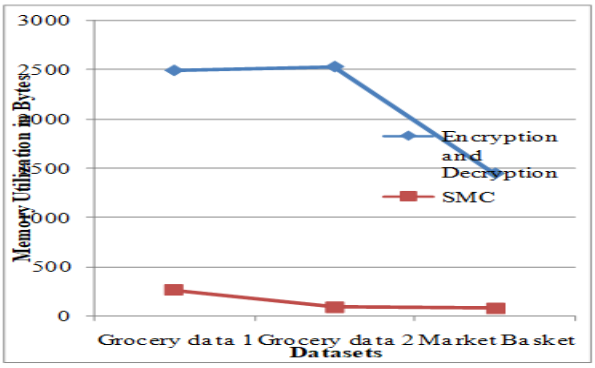
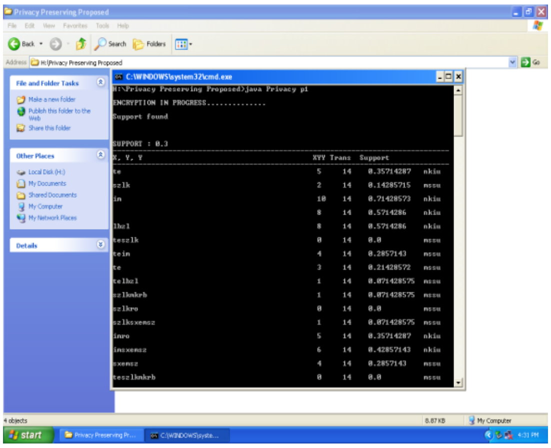


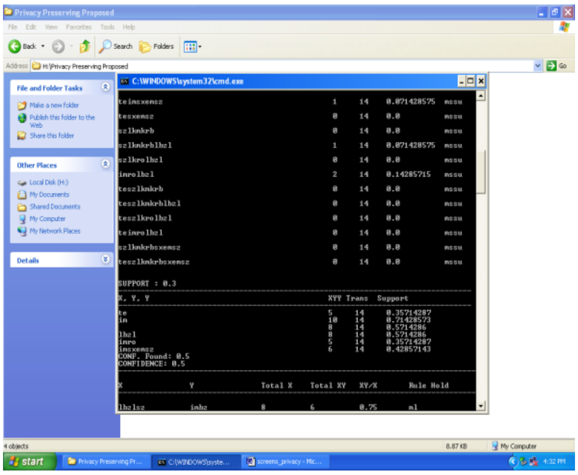
Figure 4: Comparison graph of Memory Utilization

From the Figure 4, it can be explained that the proposed APFT algorithm under association rule mining to generate rules based SMC ssapproach provides a better performance by using lesser amount of memory when compared with the existing approaches. The following figures shows the results of Grocery data 1.



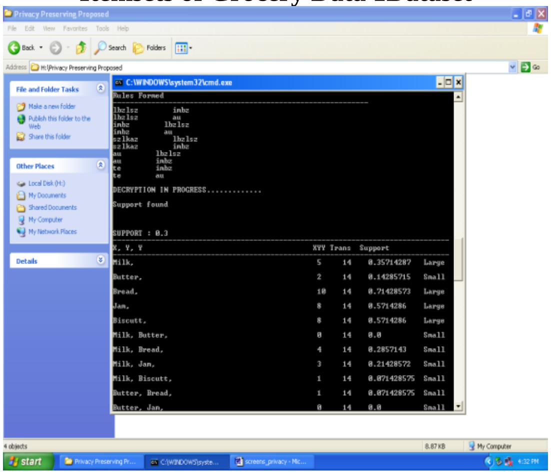
```
H:\Privacy Preserving Proposed>java Privacy p1
ENCRYPTION IN PROGRESS..............
Support found

SUPPORT : 0.3
---------------------------------------------------------------------------
X, Y, Y                                          XYY Trans  Support
---------------------------------------------------------------------------
te                                                5     14    0.35714287   nkiu
szlk                                              2     14    0.14285715   mssu
im                                                10    14    0.71428573   nkiu
                                                  8     14    0.5714286    nkiu
lbzl                                              8     14    0.5714286    nkiu
teszlk                                            0     14    0.0          mssu
teim                                              4     14    0.2857143    mssu
te                                                3     14    0.21428572   mssu
telbzl                                            1     14    0.071428575  mssu
szlkmkrb                                          1     14    0.071428575  mssu
szlkro                                            0     14    0.0          mssu
szlksxemsz                                        1     14    0.071428575  mssu
imro                                              5     14    0.35714287   nkiu
imsxemsz                                          6     14    0.42857143   nkiu
sxemsz                                            4     14    0.2857143    mssu
teszlkmkrb                                        0     14    0.0          mssu
```

Figure 5: Encryption Process of Grocery Data 1Dataset



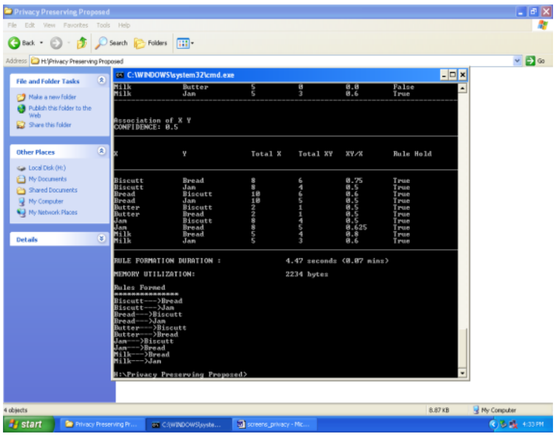
```
teimsxemsz                                        1     14    0.071428575  mssu
tesxemsz                                          0     14    0.0          mssu
szlkmkrb                                          0     14    0.0          mssu
szlkmkrblbzl                                      1     14    0.071428575  mssu
szlkrolbzl                                        0     14    0.0          mssu
imrolbzl                                          2     14    0.14285715   mssu
teszlkmkrb                                        0     14    0.0          mssu
teszlkmkrblbzl                                    0     14    0.0          mssu
teszlkrolbzl                                      0     14    0.0          mssu
teimrolbzl                                        0     14    0.0          mssu
szlkmkrbsxemsz                                    0     14    0.0          mssu
teszlkmkrbsxemsz                                  0     14    0.0          mssu

SUPPORT : 0.3
---------------------------------------------------------------------------
X, Y, Y                                          XYY Trans  Support
---------------------------------------------------------------------------
te                                                5     14    0.35714287
im                                                10    14    0.71428573
                                                  8     14    0.5714286
lbzl                                              8     14    0.5714286
imro                                              5     14    0.35714287
imsxemsz                                          6     14    0.42857143
CONF. Found: 0.5
CONFIDENCE: 0.5
---------------------------------------------------------------------------
X              Y              Total X    Total XY   XY/X      Rule Held
---------------------------------------------------------------------------
lbzlsz         imbz           8          6          0.75      ml
```

Figure 6: Support Value >= 0.3 to find Frequent Itemsets of Grocery Data 1Dataset



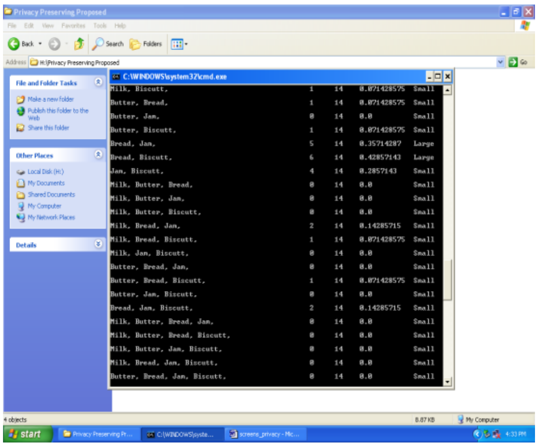
```
Rules Formed
-------------------------------------------------------------
lbzlsz         imbz
lbzlsz         au
imbz         lbzlsz
imbz         au
szlkaz         lbzlsz
szlkaz         imbz
au        lbzlsz
au        imbz
te        imbz
te        au

DECRYPTION IN PROGRESS............
Support found


SUPPORT : 0.3
---------------------------------------------------------------------------
X, Y, Y                                          XYY Trans  Support
---------------------------------------------------------------------------
Milk,                                             5     14    0.35714287   Large
Butter,                                           2     14    0.14285715   Small
Bread,                                            10    14    0.71428573   Large
Jam,                                              8     14    0.5714286    Large
Biscutt,                                          8     14    0.5714286    Large
Milk, Butter,                                     0     14    0.0          Small
Milk, Bread,                                      4     14    0.2857143    Small
Milk, Jam,                                        3     14    0.21428572   Small
Milk, Biscutt,                                    1     14    0.071428575  Small
Butter, Bread,                                    1     14    0.071428575  Small
Butter, Jam,                                      0     14    0.0          Small
```

Figure 7: Decryption Process of Grocery Data 1Dataset



```
Milk            Butter          5          0          0.0         False
Milk            Jam             5          3          0.6         True
-------------------------------------------------------------------------


Association of X Y
CONFIDENCE: 0.5


X               Y               Total X    Total XY   XY/X        Rule Hold


Biscutt         Bread           8          6          0.75        True
Biscutt         Jam             8          4          0.5         True
Bread           Biscutt         10         6          0.6         True
Bread           Jam             10         5          0.5         True
Butter          Biscutt         2          1          0.5         True
Butter          Bread           2          1          0.5         True
Jam             Biscutt         8          4          0.5         True
Jam             Bread           8          5          0.625       True
Milk            Bread           5          4          0.8         True
Milk            Jam             5          3          0.6         True

RULE FORMATION DURATION :                 4.47 seconds (0.07 mins)

MEMORY UTILIZATION:                       2234 bytes

Rules Formed
****************
Biscutt--->Bread
Biscutt--->Jam
Bread--->Biscutt
Bread--->Jam
Butter--->Biscutt
Butter--->Bread
Jam--->Biscutt
Jam--->Bread
Milk--->Bread
Milk--->Jam

H:\Privacy Preserving Proposed>
```

Figure 8: Rule Generation in Decryption Process



```
Milk, Biscutt,                    1     14    0.071428575  Small
Butter, Bread,                    1     14    0.071428575  Small
Butter, Jam,                      0     14    0.0          Small
Butter, Biscutt,                  1     14    0.071428575  Small
Bread, Jam,                       5     14    0.35714287   Large
Bread, Biscutt,                   6     14    0.42857143   Large
Jam, Biscutt,                     4     14    0.2857143    Small
Milk, Butter, Bread,              0     14    0.0          Small
Milk, Butter, Jam,                0     14    0.0          Small
Milk, Butter, Biscutt,            0     14    0.0          Small
Milk, Bread, Jam,                 2     14    0.14285715   Small
Milk, Bread, Biscutt,             1     14    0.071428575  Small
Milk, Jam, Biscutt,               0     14    0.0          Small
Butter, Bread, Jam,               0     14    0.0          Small
Butter, Bread, Biscutt,           1     14    0.071428575  Small
Butter, Jam, Biscutt,             0     14    0.0          Small
Bread, Jam, Biscutt,              2     14    0.14285715   Small
Milk, Butter, Bread, Jam,         0     14    0.0          Small
Milk, Butter, Bread, Biscutt,     0     14    0.0          Small
Milk, Butter, Jam, Biscutt,       0     14    0.0          Small
Milk, Bread, Jam, Biscutt,        0     14    0.0          Small
Butter, Bread, Jam, Biscutt,      0     14    0.0          Small
```

Figure 9: Displaying Transaction Database in SMC process of Grocery Data 1Dataset

It is evident from the results that the proposed scheme is extremely efficient when compared the existing systems.

VII. CONCLUSION

The better results for secured computation are obtained by using Cryptographic protocols. Functions can be computed securely by using genetic constructions. Specialized constructions can be used get more efficient computation. The proposed secure protocol (SMC) for computing a certain function is more effective than a naive protocol that does not provide any security.It is obtained that require for data mining in existence of privacy disquiets will augment. The proposed system (SMC) model experimentation includes knowledge discovery in the midst of intelligence services of several grocery stories and relationship among corporations without revealing deal secrets. Permitting error in the results possibly will enable more well-organized schemes with high level security. The secure multi-party computation definitions are very restrictive that are derived from cryptography domain. More suitable security definitions are needed to enable the parties to choose security level, allowing efficient solutions with required security. Designs of game theory and economics might be appropriate for this future enhancement.

REFERENCES:

1. Silverman B. W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, 1986.
2. Oliveira.Sathttp://www.cs.ualberta.ca/%7Eoliveira/psdm/pub_by_year.htmlorK.Liuat http://www.cs.umbc.edu/~kunliu1/research/privacy_review.html
3. IBM, http://www.almaden.ibm.com/software/disciplines/iis/
4. http://www.cs.ualberta.ca/%7Eoliveira/psdm/workshop.html at 2004.
5. Liew C. K., Choi U. J., Liew C. J. A data distortion by probability distribution. ACM TODS, 10(3):395-411, 1985.
6. Warner S. L. Randomized Response: A survey technique for eliminating evasive answer bias. Journal of American Statistical Association, 60(309):63–69, March 1965.
7. Agrawal R., Srikant R. Privacy-Preserving Data Mining. ACM SIGMOD Conference, 2000.
8. Evfimievski, A., Srikant, R., Agrawal, R., &Gehrke, J. (2002). Privacy preserving mining of association rules. Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02).
9. Rizvi, S. J., &Haritsa, J. R. (2002). Maintaining data privacy in association rule mining. Proceedings of the 28th International Conference on Very Large Data Bases (VLDB'02), Hong Kong, China.
10. Shannon, C. E. (1949). Communication theory of secrecy systems. Bell System Technical Journal, 28(4), 656-715.
11. Lambert, D. (1993). Measures of disclosure risk and harm. Journal of Official Statistics, 9(2), 313-331.
12. Blum, A., Dwork, C., McSherry, F., &Nissim, K. (2005). Practical privacy: The SuLQ framework. Proceedings of ACM Symposium on Principles of Database Systems (PODS'05), Baltimore.
13. Dinur, I., &Nissim, K. (2003). Revealing information while preserving privacy. Proceedings of ACM Symposium on Principles of Database Systems (PODS'03), San Diego, CA.
14. Evfimievski, A., Gehrke, J., &Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. Proceedings of ACM Symposium on Principles of Database Systems (PODS'03).
15. Evfimievski, A., Srikant, R., Agrawal, R., &Gehrke, J. (2002). Privacy preserving mining of association rules. Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02).
16. Samarati P.: Protecting Respondents' Identities in Micro data Release. IEEE Trans. Knowl. Data Eng. 13(6): 1010-1027 (2001).
17. Atallah, M., Elmagarmid, A., Ibrahim, M., Bertino, E., &Verykios, V. (1999). Disclosure limitation of sensitive rules. Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange.
18. Oliveira, S. R. M., &Zaiane, O. R. (2003). Protecting sensitive knowledge by data sanitization. Proceedings of the 3rd IEEE International Conference on Data Min¬ing (ICDM'03).
19. Pinkas B.: Cryptographic Techniques for Privacy-Preserving Data Mining. ACM SIGKDD Explorations, 4(2), 2002.
20. Agrawal R, Imielinski T and Swami A, Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD Conference on Management of Data, pages 207–216, New York, NY, USA, May 1993. ACM Press.
21. J. Pei, J. Han, and H. Lu. Hmine: Hyper-structure mining of frequent patterns in large databases. In ICDM, 2001, pp441–448.
22. BanuPriya.M*, Umarani. V —Enriching the Efficiency of Association Rule Mining using Enhanced IFP-Growth Algorithm‖, International Journal of Advanced Research in Computer Science and Software Engineering, January, 2013.
23. Kantarcioglu M and Clifton C. Privacy-preserving distributed mining of association rules on horizontally partitioned data. In IEEE Transactions on Knowledge and Data Engineering Journal, volume 16(9), pages 1026–1037, Piscataway, NJ, USA, September 2004. IEEE Educational Activities Department.

AUTHORS PROFILE

P.YOGANANDHINI,M.C.A.,M.Phil., working as Assistant Professor in Dept. of Computer Technology, Vellalar College for Women, Erode and currently pursuing her Ph.D in the Dept. of Computer and Information Sciences, Annamalai University, Annamalai Nagar. She has published one paper in an UGC refereed International Journal. she has five years of teaching experience and two and a half years of Industrial experience.

Dr.G.PRABAKARAN working as Associate Professor in Dept. of Computer Science and Engineering, Annamalai University, Annamalai Nagar. He has published two papers in National Journals and four papers in International refereed Jounals. He has published more than 8 papers in National Conference and 13 papers in International Conferences. He has 19 years of teaching experience and 4 years of Research experience. He is a member of CSI and ISTE. His main research work focuses on Image processing, computer graphics.