International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 3 | ISSN : 2456-3307

# Privacy-Preserving Mining of Outsourced Transaction Databases for Association Rules Generation Using Paillier Encryption

**Chaitali C. Khandate, Prof. Antara Bhattacharya**
Department of Computer Science and Engineering, G. H. Raisoni institute of Technology and Engineering, Nagpur, Maharashtra, India

## ABSTRACT

Cloud computing or Distributed Network uses the ideal model of information mining-as-an organization, using these it is in every way a prominent choice for associations sparing cash on the cost of adding to secure, supervise and keep up an IT establishment. An affiliation/store debilitated in mining limit can outsource its mining needs to expert center on a cloud server. In any case, both the alliance rules and thing set of the outsourced database are viewed as private property of the affiliation. The data proprietor encodes the data and sends to the server to shield the corporate security. Client sends mining inquiries to server, and after that server conducts data mining and sends mixed case to the client. To get honest to goodness illustration client unscrambles mixed case. In this paper, we consider the issue of outsourcing the connection oversees mining task inside a corporate security safeguarding structure. Therefore Privacy Preserving Data Mining is an examination domain stressed with the security chose from eventually identifiable information when considered for data mining. The Rob Frugal encryption strategy is familiar with beat the security vulnerabilities of outsourced information, which is focused on adjusted substitution figures for things likewise, including fake cases for database. In any case, it contains diverse fake illustrations which augment the breaking point overhead. To beat this issue, the proposed procedure fuses extension of weighted support in remarkable support of things to diminish the amount of fake illustrations and to overhaul the security level for outsourced information with less multifaceted nature. The fake trade table data is changed over into grid arrangement to decrease the limit overhead. Moreover the estimating attack and man in the middle ambush are possible on fundamental Rob modest count. To vanquish these strikes we utilize Pallier Encryption on after Rob Frugal encryption plot with a particular true objective to give protection saving outsourced mining. In our proposed work we improved the security as thing and thing set build ambush are unfeasible in light of the structure; additionally we diminish the dealing with time.

**Keywords :** Cloud Computing, Association rule mining, Privacy-preserving outsourcing, , Rob Frugal.

## I. INTRODUCTION

Distributed computing will handle in which inconceivable social affairs of remote servers are masterminded to allow joined data stockpiling and online access to PC organizations or resources. With the passage of conveyed registering and its model for IT organizations in perspective of the web and gigantic server cultivates, the outsourcing of data and handling organizations is picking up a novel importance, which is decidedly required to take off within the near future. In business, outsourcing incorporates the contracting out of a business system to another get-together. Outsourcing arrangements to give an organization in a corporate protection safeguarding structure. Protection confirmation is the essential issue in data mining. Relationship, overall, would lean toward not to confer their own private data to various associations. The contemplation is that data is disseminated by Client for the benefit of allowing specialists to mine encoded outlines from the mixed database. As a portrayal, the esteem based database from different affiliations can be transported to an untouchable which gives mining organizations. The affiliation organization would slant toward not to utilize an in-house social event of data mining stars. Likewise, irregularly data is sent to the server or expert community who is responsible for keeping up the encoded data and coordinating mining on it in view of requesting from association inspectors of the association organization. The data proprietor is a client and the server is insinuated as the pro center. One of the fundamental issues with this standard is that the server has area to essential information of the proprietor and may uncover sensitive information from the data. For example, by looking trade database, the server can deduce or uncover which things or things are co-procured and in this way, the mined mixed cases that delineate the affiliation customers' inconspicuous components.

In this paper, we focus the issue of outsourcing the association represent mining task inside a corporate security safeguarding structure [7]. Thusly, Privacy Preserving Data Mining is an examination run stressed with the security chose from truly identifiable data when considered for data mining. In this novel situation, both the arrangement trade database and the mined encoded illustrations and each one of the purposes of enthusiasm of the association that can be isolated from the data are the property of the association organization and should remain safe from the server and whatever other aggressor. In reality the data mined from the data can be used from the association organization in basic promoting decisions to upgrade their organizations. An association or data proprietor needs their data to be puzzle however an association does not have satisfactory burrowing authority for data mining, for this we make the going with duties. We develop an encryption plot, called Improved RobFrugal in that the Encrypt/Decrypt module can use to change client data before it is conveyed to the server. Second, to allow the E/D module to recover the bona fide cases and their correct support, we prescribe that it makes and keeps a limited structure, called plot. Third, we show extension of weighted support in remarkable support of things and system advancement of fake trade to decrease the limit overhead. Fourth with a particular true objective to give security safeguarding outsourced mining, we utilize Pallier Encryption after Improve Rob Frugal encryption plot. With use of ECDH estimation theorizing ambush and man in the inside strike are improbable on our proposed structure. Fifth, for better execution Enhance FP-Growth figuring is used instead of Apriori estimation [13] for association control time. At last, we organize test examination of our illustration utilizing a broad honest to goodness dataset, our results demonstrate that our encryption piece is convincing, versatile, and accomplish the pined for level of security.

## II. LITERATURE SURVEY

### A. Substitution cipher techniques:

W. K. Wong et al. [1] proposed substitution figure techniques in the encryption of significant worth based data for outsourcing alliance run mining. In the wake of seeing the non-irrelevant dangers to the unmistakable composed thing mapping substitution figure, we propose a more secure encryption mastermind in perspective of a one-to-n thing mapping that progressions trades non-deterministically, yet guarantees change unscrambling. They develop a fruitful and beneficial encryption estimation dependant on this procedure.

### B. Data Perturbation:

There are two systems that can secure unstable information. Is to an encryption limit that progressions the primary data to an absolutely new association [4] [2]. The second could be to apply data bothering, which adjusts the principal unrefined data erratically [3]. The trouble approach is less engaging since it could simply give deduced comes to fruition; incidentally, the work of encryption grants comparative rule that they are recovered.

### C. k-support anonymity

The establishment learning, for instance, the support of ceaseless thing sets can be utilized to gain protection information in the outsourcing of general thing set mining. In this paper [5], C. Tai, P. Yu proposed k-reinforce anonymity to give security against an informed assailant with right support information. To finish k-reinforce mystery, they show a pseudo-logical order tree and host the third assembling mine the summed up unending thing sets. The work of the pseudo-logical grouping tree empowers stowing ceaselessly of the principal things and limits the fake things displayed in the encoded database. The test comes to fruition exhibited that the procedures for k-support mystery fulfill incredible security insurance with direct stockpiling overhead.

### D. Corporate privacy preserving mining

In this paper [6], F. Giannotti, A. Monreale concentrated the issue of protection safeguarding mining of constant cases on an encoded outsourced trade database. We recognize a dynamic model where the adversary knows the zone of things and their right repeat and can utilize this making sense of how to recognize figure things and figure thing sets. We proposed an encryption plot, called Rob Frugal that relies on upon 1-1 substitution figures for things and adding fake trades to make each figure thing offer a comparative repeat. It uses the diminished summation of the fake trades from which the honest to goodness support of mined cases from the server can be adequately recovered.

E. Association rule mining by Evmievski

Evmievski et al. [8] proposed an approach, for coordinating the security safeguarding alliance oversee mining. Kargupta et al. [9] proposed a procedure in perspective of subjective cross section awful isolating to recover extraordinary data from the irritated data. Huang et al. [10] proposed help, the two data propagation procedures, first PCA-DR and second, MLE-DR.

F. Randomized Response

The key individual to propose the Randomized Response (RR) was Warner [14].The RR plan was at initially made in the bits of knowledge gathering. It used to accumulate the information from individuals with the true objective that, the review examiners and the data processors don't know which of the two choice request are respondent have answered. In data mining, the technique for randomization is a fundamental strategy, can be easily associated at data gathering time. It was a profitable technique for disguising singular data in security holding data mining. The randomization method is more beneficial [11]. In any case, it achieves high information incident. The composition on Privacy-Preserving Mining of Association standards can be orchestrated into Pattern mining task, security illustrate, ultimately Encryption/Decryption plot.

## III. RELATED WORK

A. Access Control Based Approach

Information classification, respectability, and security of the customers' data are ensured by this methodology. Among different administrations of distributed computing, empowering secure access to outsourced information establishes a strong framework for data administration and different operations. Be that as it may, more research endeavours are expected to accomplish adaptable access control to vast scale dynamic information. In this environment, the information can be upgraded just by the first proprietor. In the meantime, end clients with various access rights need to peruse the data in a productive and secure way. Both information and client elements must be appropriately taken care of to save the execution and wellbeing of the outsourced stockpiling framework.

In "Secure and Efficient Access to Outsourced Data", Weichao Wang, Zhiwei Li,Rodney Owens, Bharat Bhargava proposed their methods that incorporate:-(1)The proposed approach gives fine grained access control to outsourced information with adaptable and productive administration. The information proprietor needs to keep up just a couple of privileged insights for key induction. (2)It does not have to get to the capacity server with the exception of information redesigns. They propose complete systems to handle flow in client access rights and redesigns to outsourced information.

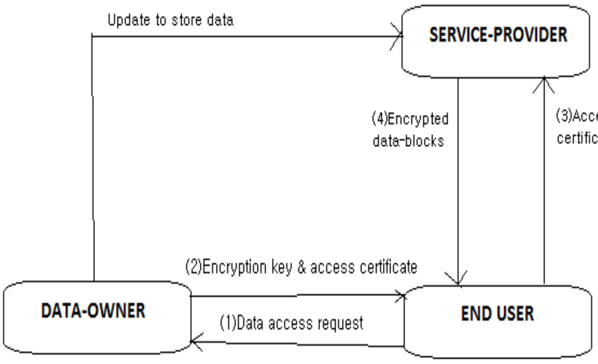

**Figure 1.** Illustration of the application situation

In this way, the proposed methodology is strong against conniving assaults if the hash capacity is viewed as protected. Examination demonstrates that the key determination system in view of hash capacities will present extremely constrained overhead. They propose to use over-encryption and/or apathetic repudiation to keep denied clients from accessing upgraded information pieces. The primary advantage of this methodology is extremely restricted overhead, maintain a strategic distance from deceitful assaults. The check plan of PKI is utilized for keeping up the uprightness information access and the correspondence accomplished for asset sharing. The responsibility is likewise bolstered in this methodology by following the client demand for information utilizing the timestamp. The downside of this framework does not have the strength regarding specialist recuperation. The methodology does not bolster the versatility for procuring extensive number of customers.

### B. Quality Based Access Control Approach

To accomplish Confidentiality, Accountability, Access Control Attribute based access control methodology is utilized as a part of which the entrance structure is identified with the arrangement of qualities of the client. In "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing", Shucheng Yu, Cong Wang, KuiRen, and Wenjing Louaddress the open issue and propose a protected and adaptable fine-grained information access control plan for distributed computing. They proposed plan in which every information record can be connected with an arrangement of qualities which are important with regards to intrigue. The entrance structure of every client can therefore be characterized as an interesting coherent expression over these credits to mirror the extent of information records that the client is permitted to get to. As the legitimate expression can speak to any coveted information document set, fine-graininess of information access control is accomplished. To uphold these entrance structures, they characterize an open key part for every trait. Information documents are encoded utilizing open key parts relating to their traits. Client mystery keys are characterized to mirror their entrance structures so that a client can decode a figure content if and just if the information document properties fulfill his entrance structure. Here accomplished these all Security prerequisite:

1. Fine-graininess of Access Control
2. Client Access Privilege Confidentiality
3. Client Secret Key Accountability
4. Information Confidentiality

The advantage of this strategy is that calculation and correspondence cost brought about for denial is less. It experiences one shortcoming. The characteristics connected with the clients are put in Attribute Authority. The denied client can degenerate this power by overhauling their own particular mystery key likewise the mystery key of non-repudiated clients.

### C. Fake Tuple Insertion Based Approach

Fake tuple based methodology is for the most part utilized as a part of outsourcing exchange database for the primary object is to befuddle the administration supplier which might be aggressor furthermore the security administrations like to trustworthiness and protection. Due to the fake tuple administration supplier can't locate the first backing of the things in the dataset. The addition of fake tuple based methodology is received in, and to give the uprightness administrations. It predominantly incorporates two methodologies as probabilistic methodology and deterministic methodology. In probabilistic strategy "Integrity evaluating of outsourced information", M. Xie, H. Wang, J. Yin, and X. Meng proposed the fake tuples are made and embedded into the database. For confirming the question honesty, the inquiry is let go against the database server which contains both the genuine and fake tuples as the predicates. The server gives back the inquiry comes about. These outcomes are confirmed by the customer who knows all the fake tuples in the database. The customer assesses the fake tuples returned by server through result and the tuples dictated by him. On the off chance that tuples from server and from customer are discovered to appear as something else, then the server is considered as deceptive and it is pronounced that the information has been altered; else if tuples from both customer and server are same, then it can be guaranteed that fulfillment is accomplished i.e. respectability of the information is kept up. As of now specified, the customer ought to know about the fake tuples. The customer needs to keep up the duplicate of late tuples. If there should arise an occurrence of expansive databases, a nearby database of fake tuples must be kept up which causes additional capacity overhead on customer and it is against the idea of outsourcing. Freshness is ensured by utilizing the fake redesign operation. The customer erases and embeds the fake tuples and break down the outcomes got by the server and assesses the freshness.

## IV. METHODOLOGY

We showed Homomorphic Paillier encryption and FP-Growth association run creation procedures for protection driving forward mining of alliance gauges from outsourced trade database. The execution purposes of enthusiasm of proposed structure are shown in Figure 1.

We grasp a direct repeat based attack show in which the server knows the right course of action of things in the proprietor's data and in addition, it in like manner knows the right support of steadily thing in the primary data. It was one of the early wears down ensuring against the repeat based strike in the data mining outsourcing circumstance. It has been exhibited using fake things to shield against the recurrence based strike it was inadequate with respect to a formal theoretical examination of security affirmations and has been seemed, by all accounts, to be faulty starting late in, the strategy for breaking the proposed encryption is given. Thusly, in our past and preliminary work, we proposed to handle this issue by using k-assurance, i.e. in that everything in the outsourced dataset should be undefined from at any rate k − 1 things concerning their support. The building behind our model is spoken to in Figure 1.

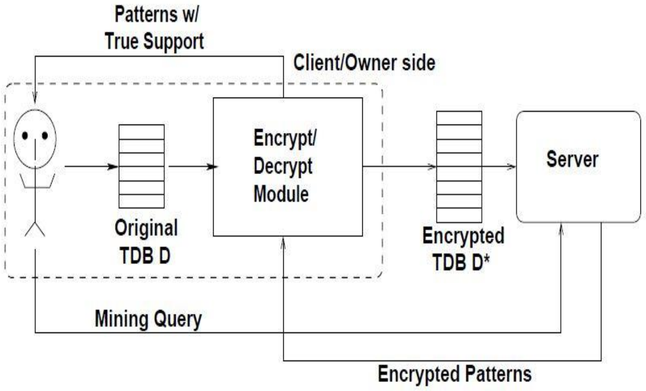

**Figure 2.** System Architecture

The client scrambles its data using an encode/unscramble module in insurance sparing, this module can be in a general sense viewed as a "black box" from its perspective. The server conducts data mining and sends the (encoded) cases to the proprietor. The propose encryption plot has the property that the returned support are not authentic sponsorships. In the propose structure the E/D module recovers the honest to goodness identity of the returned plans too their genuine support. The (E/D) module inconsequential to show that if the data is mixed using 1-1 substitution figures, In the figure content many figures and in this manner the trades and cases can be broken by the server with a high probability by driving the repeat based attack. In the propose system devise encryption arranges with the ultimate objective that formal insurance affirmations can be shown against strikes drove by the server using establishment data. At first, we formally portray an ambush show for the foe and make correct the establishment taking in the enemy may have. Our concept of security requires that for each figure message thing, there are in any occasion k−1 unmistakable figure things that are unclear from the thing concerning their sponsorships Second, we make an encryption scheme, called RobFrugal that the E/D module can use to change client data before it is sent to the server. Third, to allow the E/D module to recover the honest to goodness cases and their correct support of data thing, we recommend that it makes and keeps a littler structure, called outline. We also give the E/D module with a viable framework for incrementally keeping up the diagram against updates as affixss.

## V. ALGORITHMS

A. Rob Frugal Encryption

1) 1 to 1 substitution cipher :

The method which transformed original transaction database D into its encrypted version $D^*$. To improve the security fake transaction are added with encrypted database. Table 1(a) shows original transaction while Table 1(b) shows transaction after one to one substitution (encrypted)

| TDB |
|---|
| Soda Nuts |
| Soda Milk |
| Milk Soda |
| Nuts Milk |
| Soda Dates |
| Nuts Soda |
| Soda Egg |
| Nuts Cake |
| Cake |

| TDB* |
|---|
| e6 e5 |
| e6 e4 |
| e4 e6 |
| e5 e4 |
| e6 e2 |
| e5 e6 |
| e6 e3 |
| e5 e1 |
| e1 |

Table1 (a): TDB Table1 (b): TDB*

2) Support Calculation :

This approach was started with calculation of support of the items. Support count is the number of time the items occurred in the original transaction database.

3) Frugal Grouping :

| Item | Support |
|---|---|
| e6 | 6 |
| e5 | 4 |
| e4 | 3 |
| e1 | 2 |
| e3 | 1 |
| e2 | 1 |

Table 2: Descending order of items based on their item support

4) Robust k-Grouping method (Rob Frugal Grouping)

Where k be the group size (i.e 2 or 3), Here we consider the group size as 2.

Given the items support table, from a group of size k such that no two items from any original transaction comes adjacent to each other i.e. we can't group e6,e5 or e6,e2 as they occurs adjacent in original transaction. After K-grouping method we get output as:

| Item | Support | Noise (Difference) |
|---|---|---|
| e6 | 6 | 0 |
| e1 | 2 | 4 |
| | | |
| e5 | 4 | 0 |
| e3 | 1 | 3 |
| | | |
| e4 | 3 | 0 |
| e2 | 1 | 2 |

Table 3: Rob Frugal with K-robust grouping

5) Fake Transaction Construction :

We can construct Fake transaction by adding Noise in to original transaction i.e. we can add e1 4 times, similarly e3 3times and e2 2 times in original transaction, so we generate transaction from given noise as {e1, e3, e2}, {e1, e3}, {e1}

B. Paillier Encryption

1) Key generation :

a) Select two large prime numbers a and b arbitrary and independent of each other such that gcd(n, Φ (n)) = 1, where Φ (n) is Euler Function and n=ab.
b) Calculate RSA modulus n = ab and Carmichael's function is given by λ = LCM (a-1, b-1).
c) Select g called generator where $g \in \mathbb{Z}^*_{n2}$ Select α and $\beta$ randomly from a set $\mathbb{Z}_n^*$ then calculate $g = (\alpha n + 1)\, \beta^n mod\ n^2$.
d) Compute the following modular multiplicative inverse $\mu$ = (L ($g^\lambda$mod $n^2)^{-1}$ mod n. Where the function L is defined as L (u) = $(u{-}1)/n$.

The public (encryption) key is (n and g).
The private (decryption) key is (λ and μ).

2) Encryption:

a. Let mess be a message to be encrypted where mess $\in \mathbb{Z}_{n.}$
b. Select random r where r $\in \mathbb{Z}^*_{n2.}$
c. The cipher text can be calculated as:

$$\text{Cipher} = g^{mess} \cdot r^n.\text{mod } n^2.$$

3) Decryption:

a. Cipher text c $\in \mathbb{Z}^*_n{}^2$

Original message: mess = L (cipher$^\lambda$ mod $n^2$).μ mod n.

C. Association Rule Generation (FP-Growth)

```
Input: Built FP-tree
Output: complete set of frequent patterns
Method: Call FP-growth (FP-tree, null).
Procedure FP-growth (Tree, α)
{
If the event that Tree contains a single path P then
For each β = comb. of nodes in P do
pattern = β ∪ α
sup= min (sup of the nodes in β )
else
for each $a_i$ in the header of Tree do {
generate pattern = β ∪ α
sup= $a_i$.support
construct β's conditional pattern base
FPTree = construct β's conditional FP-tree
If Tree β = null
Then call FP-growth (Tree β, β)}
}
```

## VI. MATHEMATICAL MODEL

Let I = {i1, i2... in} be a set of n binary attributes called items.
Let D = {t1, t2... tm} be a set of transactions called the database.
Each transaction in D has a unique transaction ID and contains a subset of the items in I.
TRA1 = {bread}
TRA2= {milk, bread}
TRA3= {bread, milk}
TRA4 = {water, milk}
TRA5 = {bread, meat}

TRA6 = {bread, egg}
TRA7 = {water}
Where, I = {milk, bread, butter, meat}

The support supp(X) of an itemset X is defined as the proportion of transactions in the data set which contain the itemset.

For ex, the itemset milk, water has a support of 1/7= 0.1

The confidence of a rule is defined conf(X=> Y) = supp(X U Y) / supp(X)

## VII. EXPERIMENTAL RESULT

Following are the results obtained during the implementation phase:

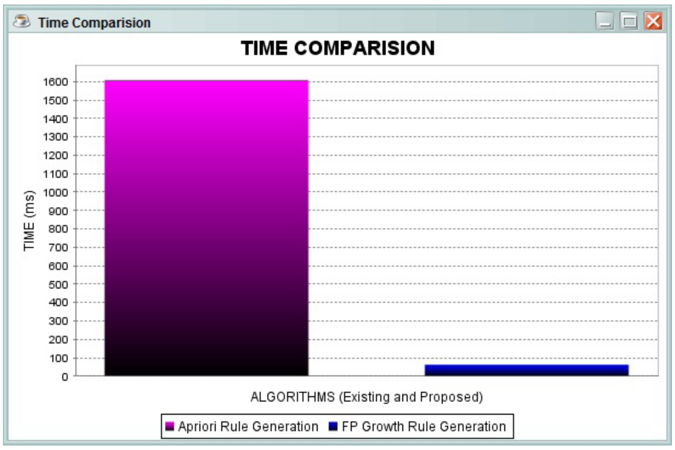

**Figure 3.** Time Comparison for Association Rule Generation

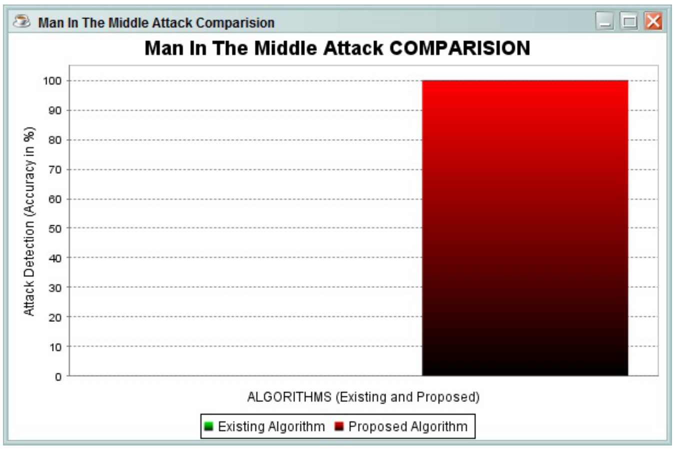

**Figure 4.** Detection of Man in Middle Attack

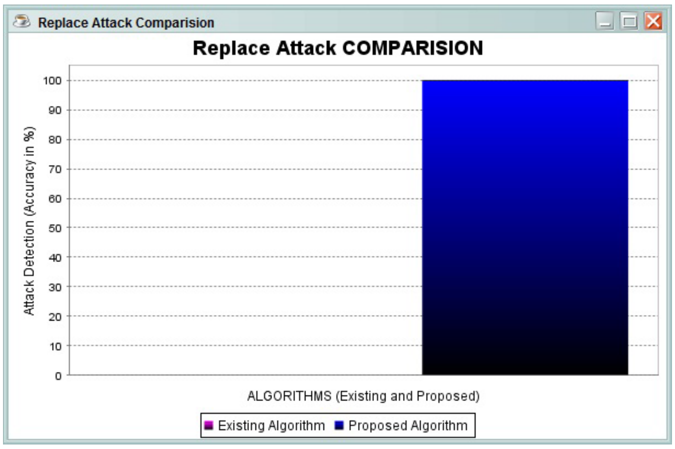

**Figure 5.** Detection of Replace Attack

Fig 2 compares the performance between FP-Growth, and Apriori Algorithm. Graphs show the execution time of implementations over the various instances. Figure 3 & 4 represents the detection of Man In Middle attack & replace attack respectively.

## VIII. CONCLUSION

Proposed Mechanism speaks to an arrangement of encryption techniques for encryption methodologies for Transactional databases that are suitable for outsourcing connection administer mining. Starting from a clear adjusted substitution figure, which is weak to attacks, we utilize Paillier Homomorphic encryption count which gives ideal security over existing plunder thrifty estimation. In like manner for association oversee time FP-Growth estimation is used which has preferable execution over Apriori. Speaks to about show that our encryption framework is greatly enthusiastic to attacks rather than essential facilitated figure, which can be easily broken with the help of establishment data. Furthermore man in the inside strike and hypothesizing attack are doubtful as structure uses Paillier encryption systems. Finally, through experimentation the proposed system has better execution with respect to time and security and lead time.

## IX. REFERENCES

[1] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, “Security in outsourcing of association rule mining,” in Proc. Int. Conf. Very Large Data Bases, 2007, pp. 111-122.

[2] G. I. Davida, D. L. Wells, and J. B. Kam. “ A database encryption system with sub keys.” ACM TODS, 6(2):312-328, 1981.

[3] J. He and M. Wang. Cryptography and relational database management systems. In IDEAS, 2001.

[4] B. Iyer, S. Mehrotra, E. Mykletun, G. Tsudik, and Y. Wu. A framework for efficient storage security in RDBMS. In EDBT, 2004.

[5] C. Tai, P. S. Yu, and M. Chen, “K-support anonymity based on pseudo taxonomy for outsourcing of frequent item set mining,” in Proc. Int. Knowledge Discovery Data Mining, 2010, pp. 473-482.

[6] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H.Wang, “Privacy preserving data mining from outsourced databases,” in Proc. SPCC2010 Conjunction with CPDP, 2010, pp. 411-426.

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

[8] S. J. Rizvi and J. R. Haritsa, “Maintaining data privacy in association rule mining”, in Proc. Int. Conf. Very Large Data Bases, 2002.

[9] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke, “Privacy Preserving Mining of Association Rules”, Information System, 2004.

[10] H. Kargupta, S. Datta, Q. Wang, K. Sivakumar, “On the Privacy Preserving Properties of Random Data Perturbation Techniques”, In Proceedings of the 3rd International Conference on Data Mining, 2003.

[11] Z. Huang, W. Du, B. Chen, “Deriving Private Information from Randomized Data”, In Proceedings of the ACM SIGMOD Conference on Management of Data, 2005.

[12] S. L. Warner, “Randomized Response: A Survey Technique for Eliminating 999999999999Evasive Answer Bias”, J. Am. Stat. Assoc., 1965.

[13] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrk, “Privacy Preserving Mining of Association Rules”, In Proceedings the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining, 2002.

[14] Ram Ratan Ahirwal, Manoj Ahke Samrat Ashok, “Elliptic Curve Diffie- Hellman Key Exchange Algorithm for Securing Hypertext Information on Wide Area Network” et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 363 368