Enhancing Security in Database Outsourcing using Cryptographic Techniques

Ms. Bhoomi G. Patel[1], Prof. Manoj B. Patel[2]

[1, 2]Department of Computer Engineering

[1, 2] Alpha College of Engineering &Technology

Abstract- *Database outsourcing is the one disciplinary in the field of the data-mining in which more than one different data-owners sends their valuable or precious data at the third party service provider's sites by paying some predefined cost where service provider provides the different services related to the database management system like data-owner can create, delete, update or manage the database at the service-provider's site so data-owner has no burden of the data at their sites. In the database outsourcing process data-owners or client can easily fire the query and get the output from the original database. But this information is sometimes important in the market value analysis or they can be used to predict something about the product. In this work research is done in to preserve the privacy of the data-owner's data from the any attacker or service-provider and the term is called as the "Corporate Privacy".*

Keywords- Access Control, Confidentiality, Freshness, Integrity, Outsourced Databases, Query Authentication, Security mechanisms

I. INTRODUCTION

Data-mining is the computational process of finding the large pattern from the database and this information may be important for business point of view or industrial point of view so it must be require to keep private from the others so privacy preservation in the data-mining is nothing but help to achieve data-mining goals without affecting the privacy of the data. Database outsourcing is the one disciplinary in the field of the data-mining in which more than one different data-owners sends their valuable or precious data at the third party service provider's sites by paying some predefined cost where service provider provides the different services related to the database management system like data-owner can create, delete, update or manage the database at the service-provider's site so data-owner has no burden of the data at their sites.

In the database outsourcing process data-owners or client can easily fire the query and get the output from the original database. But this information are sometimes important in the market value analysis or use for the predict something about the products so this database is very important part of the data-owners. The service-provider who provides the service is not always the trusted person so privacy preservation of the data-owner's data is required or it may be possible that any adversary brake the security and hake the original database. So privacy preservation in the database outsourcing is nothing but hide the original database from the service-provider who may be an adversary. So privacy preservation in the database outsourcing is current research topic and in this work research is done in to preserve the privacy of the data-owner's data from the any attacker or service-provider and the term is called as the "Corporate Privacy". Corporate privacy define as not only the private information of the particular one person but it is a whole organization's data which is very valuable or important data which must be require to keep private from the unauthorized person.

As the example of getting corporate or organization privacy is Super-market chain data management in which the operational transactional database from different shops of a super-market gets it services by the agent providing services on help regarding the data mining services. Considering here the case of super-market, the consumer here is the one owning the data whereas the agent providing the services is termed as the servers. Major problem with this the agent providing the services can get the information all confidential data of the agent who owns the data and if it is not properly secure then the server can access the information considering the yahoo.com store the password of the user registered to it but it uses the hash indexing to store to them if they are straight away stored into database then the server can and may see any other users account hence the privacy is in the picture and the main concern within it keeping the track.

II. THEORITICAL BACKGROUND AND LITERATURE SURVEY

2.1 Theoretical Background

This part defines the concept of database as a service and benefits and architecture of outsourced transaction database model then different security services require for the database.

2.1.1 Database as a service

Database as a service (DBaas) is an important approach in which it provides different functionalities to the IT providers to deploy their important data for multiple projects for development, testing, production, and failover environment [1].

2.1.2 Architecture of Outsourced Transaction Database Model

There are mainly three entities participated in the Outsourced Transaction Database Model. Three entities are:

1. Data Owner
2. Service Provider
3. Clients

The architecture is look like given below of Outsourced Transaction Database Model:-Usually, the consumers are considered as the persons while service-provider is not genuine or attacker in perspective of revealing data in an illegal manner.

The **Data owner** has authority for update, insert, delete, modify and access databases. Entity owing the data can authorize or stop the consumer retrieving data of data-base. The entity proving the services as provides the services it holds the rights regarding the data and maintains it. Even both software and hardware are at the entity holding service tasks that Service-provider performs are given below:-

- Deliver Data-base as a Facility
- Preservation &controlling
- Managing transaction
- Recovery and Rollback Scheme
- Failure adaptability
- Scaling
- Data-base Accessibility
- Adversity Protection
- Effective Query-Processing

Three types of the outsourcing models are available which names are single data-owner and service-provider then multiple client model and multiple data-owner model.

In single data-owner model single entity is available for do the all different operations which is used in the [2] [3] [4].

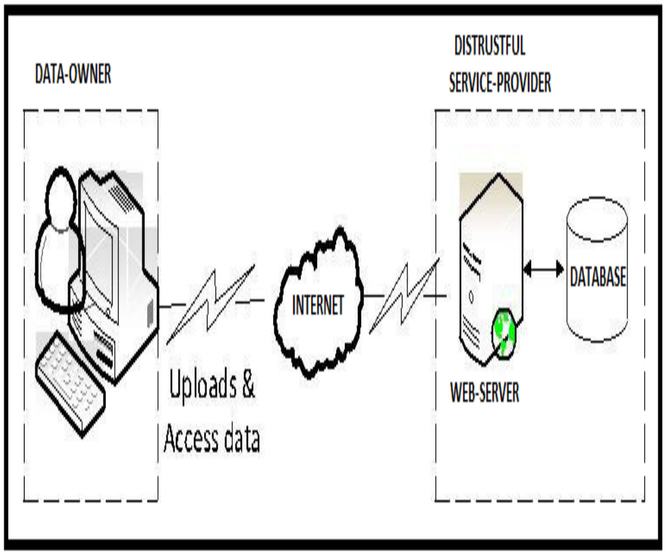


Figure 2.1.Single Data-owner and Service-Provider (Unified client model)

In the second type of the multiple client model in which multiple clients have the authority to read and access the data which were upload by the data-owner at the service-provider's site. In figure we can see the architecture of the multiple clients model which is used in [2] [3] [5].

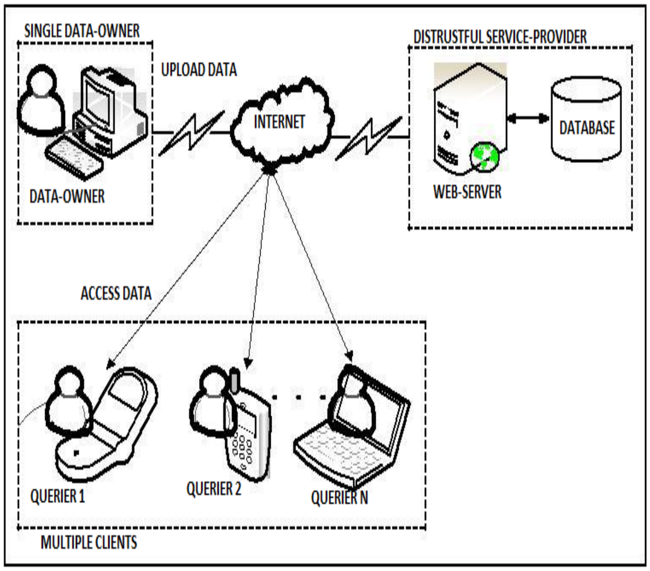


Fig. 2.2 Single Data-owner Multiple clients and Service-provider(Multiple Client Model)

In the third type of the multiple data-owner model more than one data-owner has the authority to use the database management services which is provided by the third party service provider.

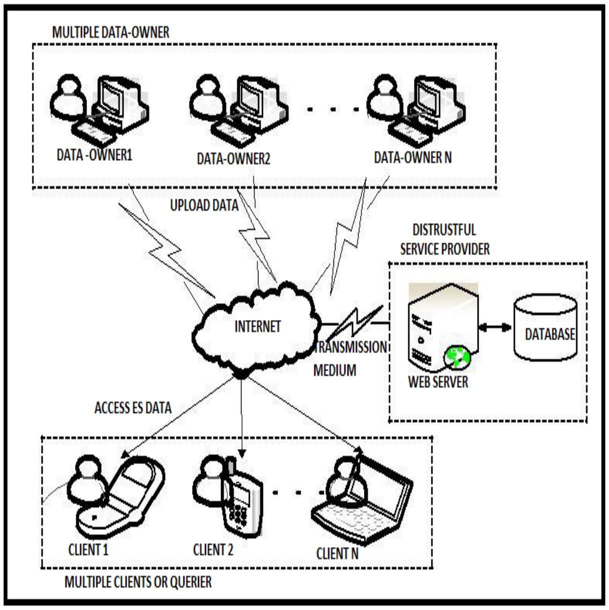


Fig. 2.3 Multiple Data-owner Multiple Clients and Service Provider (Multiple Data Owner Model)

2.1.3 Security Requirements

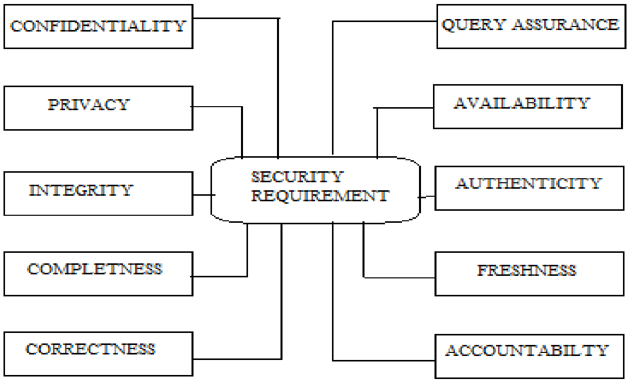


Fig. 2.4 Security Requirements for Database Outsourcing.

Figure 2.4 describes the different security services which all are needed for the database outsourcing.

2.1.3.1 Confidentiality

Confidentiality means the data becomes in unintelligible when the data is in stored mode or in the transmission state so it is the most important security service in the database outsourcing.

2.1.3.2 Privacy

Privacy is also considered as maintaining the confidentiality throughout the network to source and destination. User privacy and access privacy are the types of the privacy to hide the identity of the person and give the particular access privilege.

2.1.3.3 Integrity

Integrity is the assurance of the data that either in the transmission process or stored process data is not changed or altered. It is the combination of two parameters like completeness and correctness.

2.1.3.4 Completeness

Completeness means when user fire the query over the database he/she get the complete result from the all whole database.

2.1.3.5 Correctness

Correctness means when client fire the query over the database he/she get the correct or the genuine result from the database.

2.1.3.6 Query Assurance

Query Assurance means client believe that he/she get the result after executing the query is original means query is executed only on the genuine database.

2.1.3.7 Availability

Availability means the service of service-provider is all time available for the data-owner or client in any situation.

2.1.3.8 Authenticity

Authenticity means all the entities must be validated for authenticity before given any access or privilege of data.

2.1.3.9 Freshness

Freshness gives the guaranty to the data-owner that he/she got the result of the executing query is fired on the latest or recent version of the data.

2.1.3.10 Accountability

Accountability means the tasks performed by each entity are accountable for that entity only. *Access control* is referred to as allowing only the authorized users to access the protected data they are permitted to. *Access control* can be realized by following the three steps viz. *Identification*, *Authentication* and *Authorization*.

2.2 Literature Survey

Sr.No.	Research Work Name	Achieved Security Services	Benefis	Drawbacks
1.	**Using Multi Shares for Ensuring Privacy in Database-as-a-Service [6]**	-Privacy	-Main advantage of this work is high cost of encryption decryption is reduced here and preserve privacy of the data from distrustful server at the time of the query retrieval process.	- The drawback of the system is that it works only for numeric data and small size data for query processing.
2.	**Efficient audit service outsourcing for data integrity in clouds [7]**	-Integrity -Confidentiality	- The main benefits of this approach are TPA can audit without downloading the data, ensures verification correctness, privacy preserving and give the high performance with minimum overheads in storage, communication and computation.	- The drawback is only that one extra person TPA is required to count the all computations related to audit system.
3.	**Outsourced Private Set Intersection Using Homo-morphic Encryption [8]**	-Privacy -Correctness -Completeness -Integrity	- The benefits of this approach is that it is very efficient techniques to transfer or count the data of the clients and server in outsourcing approach very securely.	- The drawback of this system that this approach compute intersection between clients and server but this approach is not used for compute intersection between clients and service-provider so this is the limitation for this work it will be a future work for that work.
4.	**RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases [9]**	-Correctness	- The main benefits of this approach is that it provides correctness for range queries in the database outsourcing.	- The limitation is here not possible to allow the data owner and authorized users update the encrypted data without undermining the security.
5.	**Privacy Enhanced Data Outsourcing in the Cloud [10]**	-Secrecy -Access Control	-The main benefit of this approach is to provides flexible key-management functionalities at different access level.	-The drawback of this approach is two keys are used for decryption.

6.	**De-Linkability: a Privacy-Preserving Constraint forSafely Outsourcing Multimedia Documents [11]**	-Privacy	-The main benefit of this work is to preserve the privacy in the multimedia files of every person and its give the guaranty of safety of data.	-The disadvantage of this work is they provides sanitizing algorithm to protect the content of individual so it becomes some complex nature of the system.
7.	**GEODAC: A Data Assurance Policy Specification and Enforcement Framework for Outsourced Services [12]**	-Integrity	- There are many benefits of this system that it provides privacy protection, data migration, data retention, data confidentiality, and integrity and data usage control.	- The limitation of this work is it’s not work for distributed enforcement of data assurance policies.
8.	**Integrity Auditing of Outsourced Data [13]**	-Integrity	- The main benefit of this approach is that client all time verify the result first so he or she always be aware of the dishonest server which provides freshness security services.	- The drawback of the system is that the client has to maintain the copy of recent tuples. In case of large databases, a local database of fake tuples has to be maintained which causes extra storage overhead on client and it is against the concept of database outsourcing.
9.	**Privacy-Preserving Mining of Association Rules from Outsourced Transaction Databases [14]**	-Privacy -Integrity	- The main benefit of this system is to original supports values are hides from the service-provider and service-provider can't see the original database.	- The drawback of this system is that this approach provides only one-way security and mining result was not protected by any method or algorithm.
10.	**Improvising Technique of Privacy Preserving in Outsourced Transaction Database [15]**	-Privacy -Integrity	- The main advantage of this approach that it provides two layer security services in the original pattern and also in the mining result by using public key cryptography algorithm.	- The drawback of the system is that the decryption time is more compare to the previous work.

III. PROPOSED ALGORITHM AND IMPLEMENTATION METHOD

3.1 Proposed Methodology Framework

3.1.1 Transformation of original Transaction Database

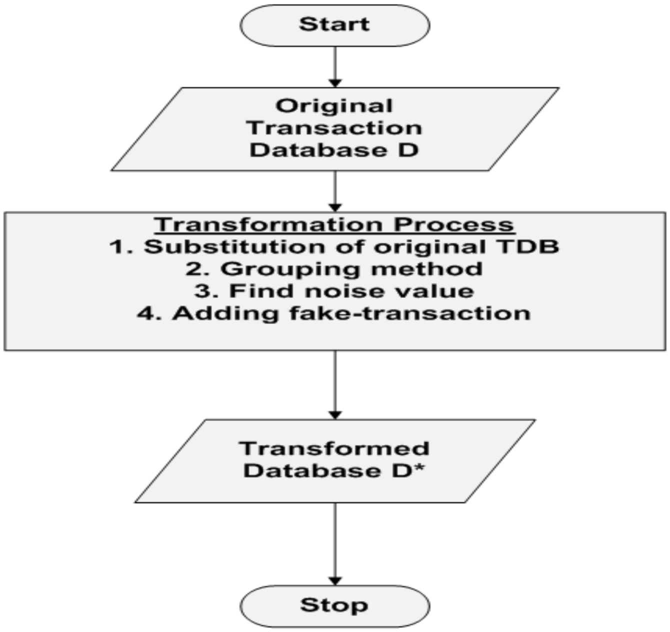


Figure 3.1 Flowchart of Encryption of Original TDB

3.1.2. Finding True Result from Transformed Database

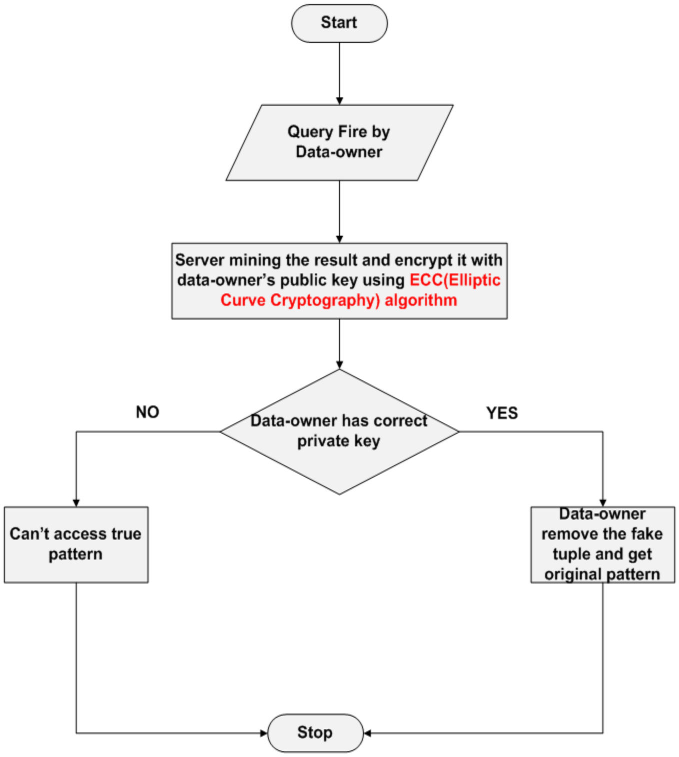

Start
Query Fire by Data-owner
Server mining the result and encrypt it with data-owner's public key using ECC(Elliptic Curve Cryptography) algorithm
NO
Data-owner has correct private key
YES
Can't access true pattern
Data-owner remove the fake tuple and get original pattern
Stop

Figure 3.2 Flow chart of True Result Mining Task

3.2 Theoretical analysis

3.2.1 Working of Existing Encryption Scheme [15]

Here we describe the existing work architecture which looks like given in below figure:-

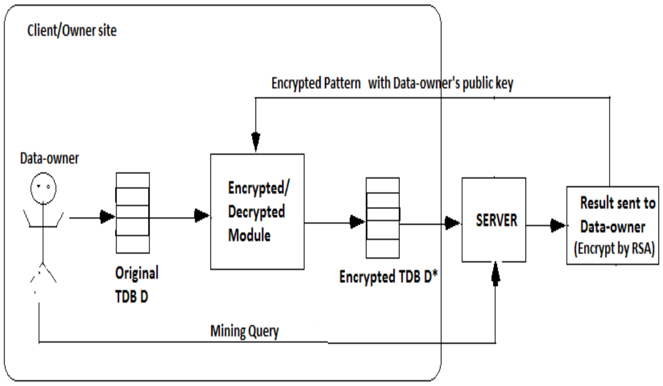

Client/Owner site
Encrypted Pattern with Data-owner's public key
Data-owner
Original TDB D
Encrypted/ Decrypted Module
Encrypted TDB D*
SERVER
Result sent to Data-owner (Encrypt by RSA)
Mining Query

Figure 3.3 Basic Architecture of Existing Work [15]

Introduction of RSA Algorithm [23]:-

RSA is one of the public-key crypto-systems and is widely used for secure data transmission. In such a cryptosystem, the encryption key is public and differs from the decryption key which is kept secret. A user of RSA creates and then publishes a public key based on the two large prime numbers, along with an auxiliary value. The prime numbers must be kept secret.

The RSA algorithm involves three steps: **key generation, encryption and decryption**.

(i) Key generation:-

RSA involves a **public key**and a **private key**. The public key can be known by everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted in a reasonable amount of time using the private key.

The keys for the RSA algorithm are generated the following way:

1. Choose two distinct prime numbers ***p***and***q***.
   For security purposes, the integer's ***p***and***q***should be chosen at random, and should be ofsimilar bit-length.
2. Compute **n= pq**.
3. Compute $\varphi(n) = \varphi(p) * \varphi(q) = (p-1) * (q-1) = n - (p + q$
   Where, φ is Euler's totient function.

4. Choose an integer e such that **$1 < e < \varphi(n)$** and **GCD (e, $\varphi(n)$) = 1** i.e. **e** and **$\varphi(n)$** are co-prime.***e***is released as the public key exponent.
5. Determine d as **$d \equiv e^{-1}$ (mod $\varphi(n)$)** ; i.e: **d** is the multiplicative inverse of **e (modulo $\varphi(n)$)**.

This is more clearly stated as: solve for ***d***given***d·e***≡ **1 (mod $\varphi(n)$).**
dis kept as the private key exponent.The*public key* consists of the modulus ***n***and the public (or encryption) exponent ***e***. The *private key* consists of the modulus ***n***and the private (or decryption) exponent ***d***, which must be kept secret. ***p, q,*** and **$\varphi(n)$** must also be kept secret because they can be used to calculate ***d***.

(ii) Encryption:-

Alice transmits her public key **(*n, e*)** to Bob and keeps the private key **d** secret. Bob then wishes to send message **M** to Alice. He first turns **M** into an integer **m**, such that $0 \leq m < n$ by using an agreed-upon reversible protocol known as a padding scheme. He then computes the cipher-text c corresponding to

$$C \equiv m^e \pmod n$$

This can be done efficiently, even for 500-bit numbers, using Modular exponentiation. Bob then transmits **c** to Alice.

(iii) Decryption:-

Alice can recover **m** from **c** by using her private key exponent **d** via computing
$$m \equiv c^d \pmod n$$

Given m, she can recover the original message M.

In the existing encryption/decryption scheme first they have taken original transaction database and then apply encryption scheme in different steps given below:-

Encryption of Original TDB:-

Input: -Original TDB D ***Output:*** -Encrypted TDB D*
Step:-
1. Take original TDB and arrange all TDB in tabular form with items and their corresponding support.

2. Apply 1-1 substitution method in order of alphabetically of every item.

3. Arrange table of items in decreasing order of support.

4. Do grouping using columnar matrix method which is a type of transposition technique. If k=10 then in our scheme the items divide in 10 different columns with the help of columnar matrix technique and then put in 10 different groups and find noise value for adding fake transactions.

5. Adding fake transaction in following way:

   a. Put “0” value of the noise column in which item has maximum support in the group.

   b. Find noise value corresponding to items with maximum support in the particular group in table.

   c. Count noise value for every items using equation **N(ei) =Max support of Item – Support of (ei)**.

   d. Discard all rows whose noise value are “0” and arrange all rows in decreasing order of their noise values.

   e. Create hash table to store value of noise or frequency of occurrence of fakely occurred in TDB with <ei, Timesi, occursi> where, ei = Num of item in TDB, times*i*represents the number of times that the fake transaction {*e*1, *e*2, . . . , *ei*} occurs in the set of fake transactions, and occ*i*is the number of times that *ei*occurs altogether in the future fake transactions after the transaction {*e*1, *e*2, . . . , *ei*}, the *i*th entry of a hash table HT containing the item *ei*has**times*i*= *N*(*ei*) − *N*(*ei*+1)***and* **occurs*i*= N(ei) - timesi** where g is the number of items in the current group.

   f. Do these all steps till added all fake transaction in all group.

6. Then finally add these fake-transactions in the original database and sends to the third party service-provider.

END

Decryption (True Pattern Mining Task):-

***Input:* -** Query ***Output:* -** True Pattern

Step:-
1. Data-owner fire query or give minimum threshold value of support for mining particular pattern.

2. Servers mining result from the encrypted pattern and send mined result with encryption using data-owner’s public-key using RSA.

3. Data-owner first decrypts mining result.

4. Then after data-owner removes fake transaction with the help of below equation

Support(S) = Supp D*(E) – (Supp D*(E) - Supp D (E))

Where, for every item set S and its corresponding cipher item set E, we have that supp$D(S) \leq$ supp$D*$ (E).
S = support of item set in TDB
D*(E) = Encrypted TDB with fake support
D (E) = Encrypted pattern with original support
5. Finally Data-owner get true pattern from fake-transaction.

END

3.2.2 Working of Proposed System

Here we have proposed our basic architecture of the our system.

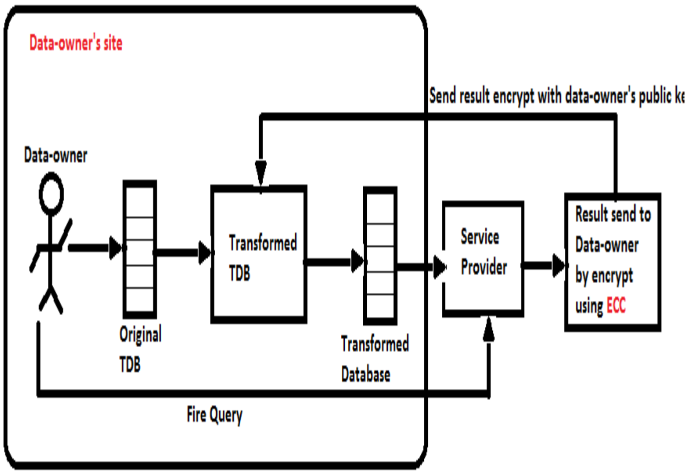


Figure 3.4 Basic Architecture of The Proposed Work

Introduction to Elliptic Curve Cryptography Algorithm [16]:-

First we have defined some basic parameters which are require in the algorithm:

The equation of an elliptic curve is given as,

$$y^2 = x^3 + ax + b$$

Few terms that will be used,
E -> Elliptic Curve
P -> Point on the curve
n -> Maximum limit (This should be a prime number)

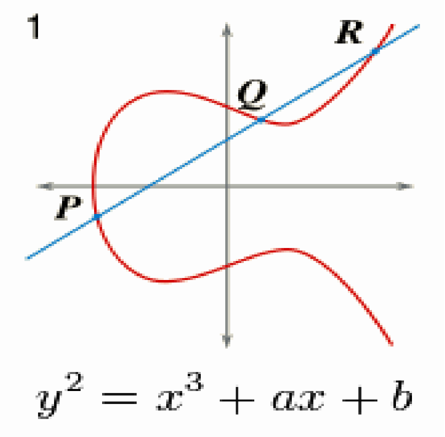


Figure 3.5Simple Elliptic Curve [16]

(i) Key Generation:-

Key generation is an important part where we have to generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key.

Now, we have to select a number **'d'**within the range of **'n'**. Using the following equation we can generate the public key

Q = d * P

d = The random number that we have selected within the range of (**1 to n-1**). **P** is the point on the curve.

'Q' is the public key and **'d'** is the private key**.**

(ii) Encryption:-

Let 'm' be the message that we are sending. We have to represent this message on the curve.

Consider *'m'* has the point *'M'* on the curve *'E'*. Randomly select 'k' from [1 – (n-1)].

Two cipher texts will be generated let it be **C1** and **C2**.

C1 = k*P
C2 = M + k*Q

C1 and C2 will be send.

(iii) Decryption:-

We have to get back the message 'm' that was send to us,

M = C2 – d * C1

M is the original message that we have send.

Mathematical Proof of Algorithm:-

Here the algorithm is proven with example in table 3.2.2.1 we can see the original TDB.

Table 3.1 Example of TDB and its support table. (a) TDB. (b) Item support table.

TDB	
Bread	
Milk	Bread
Bread	Milk
Water	Milk
Bread	Beer
Bread	Eggs
Water	

(a)

ITEM	SUPPORT
Bread	5
Milk	3
Water	2
Beer	1
Eggs	1

(b)

Step 1 Apply 1-1 substitution method in order of alphabetically of every items.

Table 3.2 Encrypted TDB

Items	Support
e1	1
e2	5
e3	1
e4	3
e5	2

Step 2 Arrange tables of items in decreasing order of support.

Table 3.3 Encrypted TDB in decreasing order of support

Items	Support
e2	5
e4	3
e5	2
e1	1
e3	1

Step 3 Do grouping with novel method using transposition technique (columnar matrix method).

Plain Text: e2 e4 e5 e1 e3 group k=2 where k is the number of group

Encryption: Table 3.4.Grouping of encrypted TDB

K-Grouping	k1	k2
	e2	e4
	e5	e1
	e3	

Cipher text: e2 e5 e3 e4 e1

After Grouping we can get output which is defined below k1 = {e2, e5, e3}k2 = {e4, e1}

Table 3.5 Encrypted TDB after Grouping

Items	Support
e2	5
e5	2
e3	1
e4	3
e1	1

Step 4 Adding fake transactions

A. Find noise value corresponding maximum support of an item.

Table 3.6 Noise table of TDB

Items	Support	Noise
e2	5	0
e5	2	3
e3	1	4
e4	3	0
e1	1	2

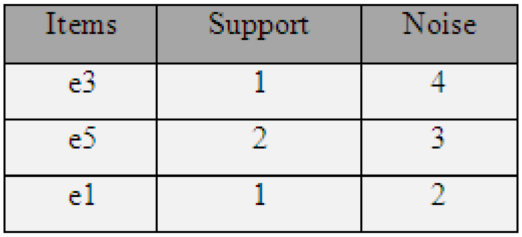
B. Discard the row which has noise value is “0”

Table 3.7 Noise table after discarded rows of “0” value

Items	Support	Noise
e5	2	3
e3	1	4
e1	1	2

C. Arrange the rows in decreasing order of noise

Table 3.8 Noise table of decreasing order of noise



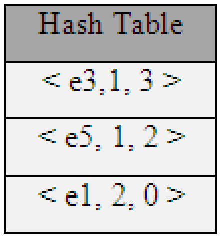
Items	Support	Noise
e3	1	4
e5	2	3
e1	1	2

D. Create Hash table to store the value of noise or frequency of occurrence of fakely in original TDB using <Ei, Timei, Occursi> In general, the *i*th entry of a hash table HT containing the item *ei* has times*i*= *N*(*ei*) − *N*(*ei*+1)

occur *i*= N(ei) – Time si where g is the number of items in the current group.

Here, i=3 N(e3)=4 times3=N(e3) – N(e5)
= 4 – 3=1
Occurs of e3 = N(e3) – Time si = 4 – 1 = 3

Table 3.9 Hash Table



Hash Table
< e3,1, 3 >
< e5, 1, 2 >
< e1, 2, 0 >

ei, time s*i*, occ*i*, where time s*i* represents the number of times that the fake transaction {*e*1, *e*2,.,*ei*} occurs inthe set of fake transactions, and occurs *i* is the number of times that *ei* occurs altogether in the future fake transactions after the transaction {*e*1, *e*2, . . . , *ei*}.

So, in this way data-owner send the encrypted TDB D* with adding fake transaction to the server.

IV. RESULT ANALYSIS

4.1 Execution time of CHESS DATASET by applying RSA and ECC algorithms.

Here I have been done experiment on chess-dataset to find the total execution time for RSA and ECC and get less execution time in ECC than RSA. which are shown in below figure(graph).

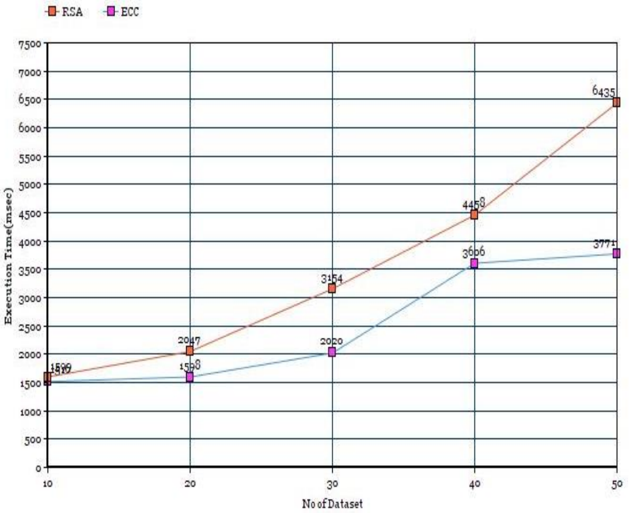


Figure 4.22 Graph for CHESS DATASET

4.2 Execution time of RETAIL DATASET by applying RSA and ECC algorithms.

Here I have been done experiment on Retail-dataset to find the total execution time for RSA and ECC and get less execution time in ECC than RSA. which are shown in below figure(graph).

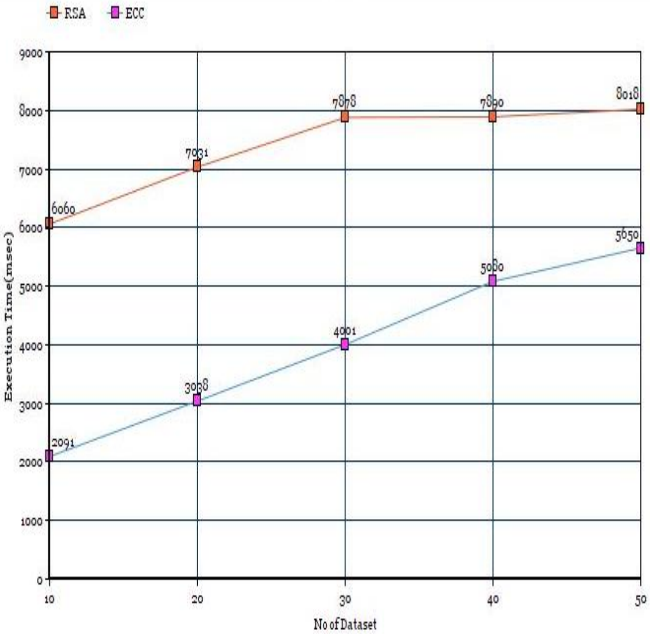


Figure 4.23 Graph for RETAIL DATASET

V. CONCLUSION

I have mainly studied the different security techniques used for the database outsourcing and analysed that which technique provides which level of the security. I have try to enhanced the security in the outsourced transaction database by using the ECC (Elliptic Curve Cryptography) algorithm because discrete logarithm problem is very harder or we can say that this problem is unsolvable till and decryption time is less of ECC algorithm compare to existing work. So by using this approach security or privacy is defiantly enhanced provides high level of security in the mining result from the third party service provider in the database outsourcing and the title of dissertation " Enhancing security in database outsourcing using cryptographic techniques" is justified.

REFERENCES

[1] "www.oracle.com/technetwork/topics/.../oes-refarch-dbaas 508111.pdf"

[2] M. Narasimha, and G. Tsudik, E. Mykletun Authentication and integrity in outsourced databases, In Proc. of ACM Trans. On Storage, vol. 2, 2006, pp. 107-138.

[3] M. Xie, H. Wang, J. Yin, and X. Meng, Integrity auditing of outsourced data,"VLDB 2007, pp. 782-793.

[4] Ai-Guo Tang, Zheng - Fei Wang Implementation of Encrypted Data for Outsourced Database, In Proc. of Second International Conference on Computational Intelligence and Natural Computing (CINC), IEEE, 2010, pp. 150-153.

[5] Marios H, George K,Feifei, Dynamic Authenticated Index Structures for Outsourced Databas**e,** In Proc. of ACM SIGMOD'06. Chicago, Illinois, 2006, pp. 121-132

[6] MA Alzain, E pardede ,Using Multi Shares for Ensuring Privacy in Database-as-a-Service, System science 2011 IEEE**.**

[7] Y Zhu, H Hu, GJ Ahn, SS Yau, Efficient audit service outsourcing for data integrity in clouds, Journal of system and software, 2012 Elsevier.

[8] F Kerschnbaum, Outsourced Private Set Intersection Using Homo-morphic Encryption, 2012 acm.

[9] K Chen, R kavuluru, S Gou,RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases, 2011 acm.

[10] M Zou, Y Mu, W Sushilo, J Yan, L Dong, Privacy Enhanced Data Outsourcing in the Cloud, Journal of network and computer, 2012 Elsevier.

[11] B AI Bouna, EJ Raad, C Elia, R Chbeir, De-Linkability: "a Privacy-Preserving Constraint forSafely Outsourcing Multimedia Documents",2013acm.

[12] J Li, B Stephenson, GEODAC: A Data Assurance Policy Specification and Enforcement Framework for Outsourced Services, 2011 IEEE.

[13] M Xie, H Wang, J Yin, X Meng, Integrity Auditing of Outsourced Data, 2011 ACM.

[14] Laks V. S. Lakshmanan, FoscaGiannotti, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases, IEEE SYSTEMS JOURNAL, VOL. 7, NO. 3, SEPTEMBER 2012.

[15] J Parmar, V Gupta, Improvising Technique of Privacy Preserving in Outsourced Transaction Database, Emerging research communication, information, commutaion application,ERCICA-2015, SPRINGER.

[16] https://bithin.wordpress.com/2012/02/22/simple-explanation- for- elliptic- curve -cryptography-ecc/.

[17] http://www.thestudymaterial.com/presentation-seminar/electronics-presentation/248-elliptic-curve-cryptography.html?start=5.