A Novel Approach for Protecting Privacy in Cloud Storage based Database Applications

AMBIKA PAWAR, Dr. AJAY DANI
Department of Computer Science and Information Technology
Symbiosis Institute of Technology
Symbiosis International University
Symbiosis Knowledge Village, Lavale, Pune
INDIA
ambikap@sitpune.edu.in, ardani_123@rediffmail.com

Abstract: - Cloud computing technology is appealing to many organizations and has become a popular approach to reduce the cost of services. Adoption of cloud storage is limited due to privacy problems while storing and retrieving sensitive information. This paper presents a solution to the privacy by splitting table containing identity and sensitive information into two separate tables, one with identity information and another stores only sensitive information. It provides k-anonymity while performing inserts, select, update and delete operations. The degree of privacy increases with the values of k. In this paper we carry out performance analysis of our proposed scheme for different values of k. Performance analysis shows that the system performance do not degrade significantly for higher values of k. The Regression models of relation between dependent variables like response time, local computation cost, local communication cost and independent variable k have been developed in this paper. This help system designer to identify the degree of privacy required and predict the time required for different database operations. The analysis of results of residual shows that the developed models are good fit for the experimental data.

Key-Words: - cloud storage, privacy, k-anonymity, database, splitting, shuffling, insert, retrieve, update, delete, query, sensitive.

1 Introduction

Cloud computing technology provides economical and scalable service models such as IaaS, PaaS, SaaS and DaaS. Adoption of any cloud service model allows organizations to focus on their core activities and use computing services provided by the cloud service providers. Although cloud computing technology provides computing services, security and privacy are significant obstacles preventing direct adoption of cloud services. Solutions to security and privacy will lead to wider adoption of cloud computing technology and services.

In case of DaaS and SaaS model, organizations using these services have to shift their database to third party cloud service provider. Organizations are reluctant to release their sensitive data to semi trusted service providers as data is important resource. As cloud services cannot be completely trusted, there is need of protecting sensitive data from third party service providers.

Anonymization is a privacy enhancing technique for outsourcing sensitive data and still it allows the usage of data. In data anonymization original data is changed using some anonymity technique and changed data can be stored on a cloud and processed privately.

K-Anonymity is a privacy model proposed by L. Sweeny [2]. It also measures the degree of privacy preservations. The goal to anonymize records is to make any record in database/table uniquely unidentifiable i.e. for any record there must be at least k-1 other records that match the attribute values. In this case k ($\geq$1) is a positive integer. As the value of k increases degree of privacy preservation increases. The higher value of k ensures that the probability of uniquely identifying a record decreases. However size of k can impact the performance of application.

This paper aims to solve an individual and query privacy problem for cloud based database applications. The design of privacy technique and its

application is discussed in this paper. During storage, sensitive and identifiable data are protected by storing them in separate tables. While retrieval, sends anonymous requests to protect the privacy. We also analyze the impact of value of k on the performance of queries. We also attempt to build predictive model which can predict query response time using k as independent variable. This can help system designer to find out impact of value of k on query response time and make appropriate design trade off. We also carry out residual analysis to validate the model.

The rest of the paper is organized as follows. In Section 2, we discuss the related work. In Section 3, definition of the privacy problem is presented. In Section 4, the design of proposed anonymization technique is presented. In Section 5, the performance of proposed scheme and statistical models are discussed. We conclude in section 6.

2. Related Work

The problem of data privacy has been explored in different research fields as follows.

2.1 Privacy preserving data publishing

This area of research deals with private publishing of sensitive databases. In this research area various techniques and tools are invented to privately publish database. The primary objective is to protect data subject's privacy e.g. in case of medical database information belongs to patients so we call patient as a data subject. The privacy in this case means identification of an individual's record present in published database. After publishing data one must not be able to identify presence of an individual's record in published data.

Researchers have proposed many approaches for different scenarios of data publishing. Data publishing has three different phases: data collection, data sanitization and data publishing. In data collection phase data owner collects data e.g. hospital collects patient records, data sanitization - data owner sanitizes original data to preserve data privacy of record. Data publication - data owner publishes data for data mining or other purpose to public or third party service providers like cloud DaaS. Hybrid solution using statistical analysis and encryption for privacy preserving medical data sharing in the cloud environment has been proposed in [1].

Various data sanitization techniques have been proposed in literature. Generalization [2-4], suppression [4,5], swapping [6,7] and randomization [8] are widely used. All these techniques handle record owner's privacy using sanitization.

2.2 Privacy preserving database outsourcing

Hacigumus et al. [9] designed private database outsourcing model for protecting records of clients from third party. Many techniques have been proposed using encryption to protect data privacy [10]. The querying and processing on encrypted data is not very easy. Another approach is sending multiple records as a result of clients query and filtering of result on client side [11-18]. These approaches fail to preserve data owner's privacy since they reveal additional information to client. CryptDB using various encryptions in order to support different types of SQL queries on encrypted data is proposed in [19]. It is prone to attack as fully dependent on CryptDB component.

A secret sharing scheme using multiple server architecture has been proposed in [20]. This scheme assumes the existence of non colluding servers. In addition to it, it has significant communication and computation overhead.

A survey of existing privacy techniques and indexing to different approaches based on query support, level of confidentiality and performance has been proposed in [21].

2.3 Privacy preserving continuous data publishing

Continuous data publishing techniques proposed in [22] and [23] support only insert operation. The insert and delete operations are supported in [24]. The insert, delete and update operations using anonymization approach for continuous publishing are supported in [25] and [26]. The group id is exposed in [26] and sensitive information related to records in specific group can also be exposed. It uses temporary insert and update record table and encryption to store records in these temporary tables.

A graph based anonymization algorithm has been proposed in [27]. The existing approaches are modifying data using generalization, suppression for making it anonymous. It can cause information loss. Our approach is different; we are anonymizing data without generalization and suppression. Our model supports private bulk data upload, insert, select, update and delete operations.

2.4 Privacy preserving query processing

This research area focuses on privacy of clients query while outsourcing database services. Since clients concern is to privately query the database lot of work has been carried out in this research area. As third party cloud server can infer about client's interest with the query access pattern, it is expected to provide query privacy. In order to query outsourced data privately many searchable encryption based solutions are proposed.

Homomorphic encryption is leading cryptographic technique in providing privacy preserving query processing on encrypted data. A framework to prevent the information leakage and support secure query processing has been proposed in [28].

Although extensive research has been aimed to solve the privacy problem, there is challenge of overcoming high computational and communication cost. After solving this problem small scale companies with smaller IT infrastructure will also be able to use the cloud for their application. Unless the above mentioned limitations are overcome, we consider the problem of preserving data privacy in cloud based applications is still open.

2.5 Encryption Techniques

Public key encryption (PKE) and Symmetric key encryption (SKE) are widely used cryptographic approaches. Homomorphic encryption, attribute based encryption; Identity based encryption, proxy re-encryption, searchable encryption these are other encryptions.

Public Key Encryption (PKE): PKE is asymmetric key encryption i.e. it requires two keys one public key for encryption and other private key for decryption. PKE cannot be widely used in cloud environment due to key management challenges.

PKE used with Symmetric Key Encryption as a hybrid approach, i.e. encrypting data using symmetric key and securing these symmetric key using PKE. PKE uses Elliptic Curve Cryptography and RSA techniques for generating public/private keys.

Symmetric Key Encryption: SKE uses only one secrete key to encrypt and decrypt. SKE are effective in securing data. The SKE based algorithm is widely adopted and Advanced Encryption Standard (AES) algorithm is widely used. AES is declared as a standard by National Institute of Standards and Technology. RC4 and A5/1 are stream ciphers which are commonly used.

Attribute based Encryption (ABE): Attribute based encryption was introduced in [33]. In ABE encryption and decryption keys are generated using attributes of users. It provides fine grain access control but has heavy computational overhead. Further ABE work was extended by researchers to generate new cryptosystems, such as ciphertext policy Attribute based Encryption (CP-ABE), Key Policy Attribute Based Encryption (KP-ABE) and Multi-Authority Attribute Based Encryption (MA-ABE).

Searchable Encryption: It enables search operation on encrypted data without disclosing contents and is used to query servers which are not trusted. Similar to other cryptographic approaches it suffers from computational cost and also functional usability.

Identity Based Encryption (IBE): The concept of IBE was introduced in [34]. IBE work, then extended to make IBE practically applicable [35]. IBE uses user identities such as email id and name for generating public key and corresponding secrete keys get issued by trusted third party.

Fully homomorphic encryption (FHE): Homomorphic encryption is great innovation in the field of cryptography. It supports computations (query execution) directly on encrypted data and returns result in encrypted form. Arithmetic operations such as addition and multiplication on encrypted numbers without decrypting, has been proposed in [36].

Cryptographic approaches provide strong data security but have computational overhead.

AES is the suitable technique for the design proposed in this paper, to store the identity information of a person as cipher text. This will prevent identification of person in the database. After retrieving group of k records to make anonymous retrieval of data, only one required record from k will be decrypted. Since the encryption is one time and also performed on small size data, it will incur less additional computational cost. Decryption is also required to be done on small size data. This will lead to little increase in computational overhead. This approach can be further extended by using different encryption techniques which support processing on cipher text. This Paper limits operation to the select query using record id.

2.6 Anonymity

K-anonymity classifies attributes as key attributes which allows unique identification of an individual. E.g. name, SSN etc. Key attributes should be removed from tables before outsourcing. The second attribute class is quasi identifier these can be

grouped together to uniquely identify an individual with external database which stores values for similar set of quasi identifiers e.g. age, gender, zip code etc. These types of attributes can be generalized or suppressed before outsourcing to anonymize records. Third class is that of sensitive attributes. Sensitive data should be de-linked from individual identification (QI) attributes. Efficient full domain k-anonymity is proposed in [3].

A generalized version of table1 is shown in table 2. Consider that the hospital outsource table 2 instead of original table 1. The cloud service provider can no longer uniquely decide individual's sensitive information. In this example adversary cannot uniquely identify Bob's record from table 2, since first 2 records can belong to Bob.

Table 1 Original Table Patient Microdata

Name	Age	Gender	Zip code	Disease
Bob	21	M	1200	Heart
Smith	23	M	1400	Diabetes
Alice	34	M	1800	Flue
Andy	37	M	2000	cold

Table 2 2-anonymous Patient Microdata

Name	Age	Gender	Zip code	Disease
Bob	[21-30]	M	[10k-15k]	Heart
Smith	[21-30]	M	[10k-15k]	Diabetes
Alice	[31-40	M	[16k-20k]	Flue
Andy	[31-40	M	[16k-20k]	cold

L-diversity model proposed [29] and also stated that k-anonymized dataset is prone to two types of attacks; first, attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes and second, attacker often has background knowledge which may disclose sensitive values. In above example if Bob and Smith have same disease then it does not satisfy l-diversity and disclose sensitive information. A novel privacy notion has been proposed [32], which requires the distribution of a sensitive attribute in any equivalence class close to the distribution of the attribute in the database table.

Anatomy for publishing sensitive data that obey l-diversity has been proposed in [30]. Anatomy releases quasi-identifiers and sensitive data separately in two different tables. Combined with grouping mechanism, it protects privacy as table supports l-diversity i.e. in each QI group, at most 1/l rows posses the most frequent sensitive value. This work also overcomes the limitation of generalization in aggregate analysis. Anatomy to relational database and encrypted keys between identifiable data and sensitive data has been proposed in [31]. In this case only client has keys and he can reconstruct the original identifiable data. In this way the privacy of data is protected. Further this work is extended in [26] to support private write operations such as insert, delete and update on anatomized data.

Anatomy technique proposed in [30] is very effective but exposes group id and sensitive information related to records in specific group to an advisory. Due to this the adversary can make a probabilistic guess about associating identity information. The concept of temporary insert record table, update record table and encryption to store records in these temporary tables has been proposed in [26]. The use of encryption can increase client side computation cost.

Our proposed technique uses shuffling and grouping techniques. It overcomes the drawbacks of [30] as group level relation between QI and sensitive attributes are hidden from adversary. It uses temporary insert and update without encryption reduces the client computational overhead as compared to [26]. In the literature we studied, we could not find the contribution which mentions about private retrieval of specific record. Our scheme supports single record retrieval privately. Existing research efforts and solutions still do not provide comprehensive solution in cloud computing environment to protect privacy of the customer information and queries run by customers to retrieve the data. The current cyber law is more focused on the intrusion protection rather than protecting users from business collecting private information and the user's health data. Many companies have commercial interests in collecting private information of users and cloud databases are good source to provide this information. Therefore, it is crucial to protect the user information stored on the cloud servers as well as prevent malicious cloud servers from gaining the information by inspecting the queries used by clients.

The cloud computing technologies allow performing intensive computing on cloud servers by removing the burden of computation from the client devices. However, it is a great challenge to achieve this effectively and efficiently, while protecting the privacy of individual and query data. There are very few solutions which can ensure the privacy of user data stored on semi trusted cloud servers and prevent cloud servers from inferring sensitive information by monitoring the user's query content.

3. Proposed Solution and Problem Definition

In this section we discuss the problem definition, assumptions and threat model and proposed solution.

3.1 Problem Definition

For any database application containing sensitive data, it will have the data privacy issue in the cloud computing environment. For example in case of CRM applications, customer database which consists of customer's highly sensitive data there is need to protect the data privacy.

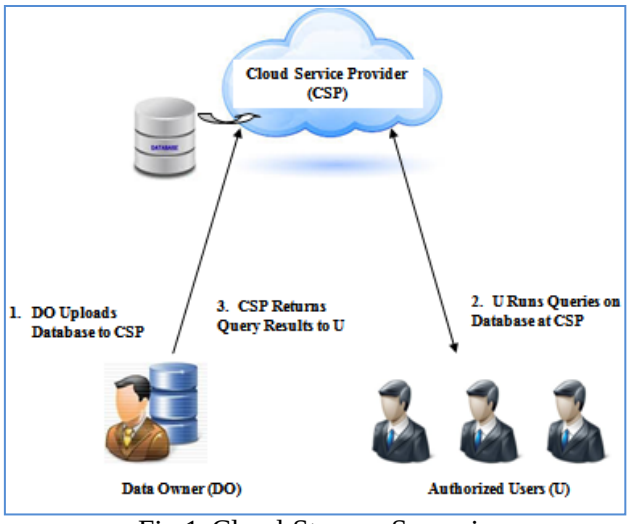


Fig.1 Cloud Storage Scenario

Figure 1 presents the architecture of typical application that uses cloud storage. In the cloud there are three different entities such as: i) Cloud Service Provider (CSP), ii) Data Owner (DO), and iii) Users (U). Data owner first uploads database to CSP. Authorized users then query on uploaded data at CSP. CSP answers to the user queries.

As shown in table 3, table has different attributes such as Name, Gender, Mobile number and annual income. We can categorize these attributes as identifying attributes and sensitive attributes using the technique described in [9] or as per customer's privacy requirement. Identity attributes helps to uniquely identify a person in a given database. In this paper, the term identifiable attribute is used for the quasi identifiers

In cloud, databases that contain person specific sensitive information get stored and processed in the cloud environment. This stored sensitive data will be retrieved from authorized users. Thus cloud service providers (CSP) who host a service for data storage and retrieval are untrustworthy. To earn revenue they may sell users private information. So, it is crucial to protect sensitive data during storage and retrieval on the cloud servers.

As shown in Table 3 person specific sensitive information can be classified as Identity Attributes (IA) and Sensitive attributes (SA). Identity attributes values enable identification of an individual and sensitive attributes store sensitive information e.g. in table 3 annual income can be categorized as sensitive information.

Table 3 Attribute Classification

Identity Attributes			Sensitive Attribute
Name	**Date of birth**	**Zip code**	**Income**
Smith	12-02-1984	411028	5 Lac
Bob	03-10-1953	411025	7 Lac
Andy	29-09-1986	411023	10 Lac

Let us assume that a database consists of attributes A_1, A_2,…, A_n, A_{s1}, $A_{s2,\ldots..}$ A_{sm}. Attributes classified as A_1, A_2,…, A_n as identifiable attributes and A_{s1}, $A_{s2,\ldots..}$ A_{sm} as sensitive attributes.

The privacy can be defined as a "Probability of not able to associate sensitive data to a particular person or to his/her identity information". Increasing the probability of disassociation of sensitive data, results in the maximization of the privacy of an individual. Sensitive and identifiable data during storage can be protected by storing it in separate tables. During the retrieval to protect the privacy it is essential to make the data requests anonymous.

3.2 Assumptions and Threat Model

In cloud DaaS model, data owner first anonymizes the database such that individual identifiable information (attributes) and sensitive information are not linkable. Data owner then stores this anonymized data on cloud server. We assume cloud server is semi honest but curious and interested in knowing sensitive information. Client queries the data stored on cloud server to delegate data processing.

Privacy is achieved with partial query processing at data owners end. This will preserve individual's (data subject's) privacy as well as client's query

privacy as all client queries will pass through data owner and get modified before sending it to cloud server. Adversary server will never be able learn about particular client's query pattern. Access control will be implemented on application server. We assume that data owner site application server is highly secure, all networks and communication channels are also secure.

3.3 Proposed Solution

The proposed anonymity technique performs the split operation on table containing sensitive information before storing database to cloud servers. Split operation generates two tables identifying information table and other sensitive information table.

Splitting alone will not solve privacy issue. Identity information and sensitive information can be easily linked/mapped. Further we shuffle either table based on less computation requirements. During Shuffling link table is generated to store links between identity and sensitive information. Only client can link it using link tables. This protects an individual privacy during storage.

Private retrieval is done using grouping of k records. K records are grouped without using generalization and suppression. User query get modified to make retrieval k-anonymous.

4 Design of Anonymization Technique

In this section first we describe the challenges and then we explain the design of anonymization technique with example.

4.1 Challenges with Privately Querying Outsourced and Transformed Database

This section demonstrates challenges while executing write queries e.g. insert, update and delete on sensitive patient database stored using our privacy model.

Knowing the table before and after query execution will allow cloud server to learn about the sensitive information. E.g. the delete query, execution of Delete from T where PID=1, if we delete record simultaneously from T1 and T2. Server will easily infer Sim's mobile no. and his annual income.

Logical deletion is the simplest and effective approach to delete single record without revealing any information to the cloud server. We update link table on data owner's site to indicate record status as deleted record. Thus it will not change any information on cloud server and protect record privacy.

Single record insert query will simultaneously insert information to T1 and T2. Thus server will understand the person and his/her sensitive information.

In case of update query, it will reveal the sensitive information and identity of person if query simultaneously updates both the tables as in case of insert. Thus all the write queries reveal the sensitive information due to simultaneous updates in T1 and T2.

To solve this issue we store updated information in temporary tables. Separate temporary tables for insert and update records are stored with data owner. We upload updated/inserted records in group of size k to the cloud server. Once we have k records in temporary insert /update tables, we make data k anonymous. We use split and upload method as explained in section 1.2, to delink the associations of identifying information and sensitive information.

Thus after performing delayed insert or update on database will not reveal direct association of sensitive information to an individual. But still server can relate sensitive information with 1/k probability.

Grouping of records in temporary tables can be done using l-diversity [26] or t-closeness [29] to provide strong privacy since k-anonymity is weak as compare to above two privacy measures. The value of k is critical for privacy. In case of higher value of k, the probability of identifying correct record decreases.

4.2 Example

Any business organization's customer database can be classified on the basis of type of attributes as Identifying Attributes (IA) and Sensitive Non-identifiable attributes (SA). The IA and SA are stored on the cloud storage in separate tables to break direct association of sensitive attributes with the identifying information. Consider an example of person specific and sensitive information table given in the Table 4.We consider two stages of data in cloud storage, first storage and second retrieval. Store Operation: To protect the privacy, the proposed approach splits the database into two tables as shown in Table 5 and 6.

After the split operation, one table holds all identity attributes and other holds all sensitive attributes. The proposed approach, stores the

sensitive data separately which is unknown to others.

Standard shuffling algorithm is applied to the sensitive attribute table in the Table 6. The algorithm shuffles records and place records in random order. Table 7 is a shuffled sensitive information table. Table 8 is link or mapping table to map identity record from table 5 to sensitive record in shuffled sensitive info table i.e. table 7.

Table 4 Person specific and sensitive information table

R ID	Name	Gender	Mobile No.	Annual Income(Lac)
1	Sim	F	9042341234	6
2	Kim	F	7823452323	7
3	Jon	M	8834563423	5
4	Jim	M	9432456778	8
5	Tom	M	7823134257	10
6	Vita	F	8842414514	9
7	Ayuri	F	7324569708	11
8	Ram	M	7634567800	4

Table 5 Person specific information table

R ID	Name	Gender
1	Sim	F
2	Kim	F
3	Jon	M
4	Jim	M
5	Tom	M
6	Vita	F
7	Ayuri	F
8	Ram	M

Table 6 Person sensitive information table

S ID	Mobile No.	Annual Income(Lac)
1	9042341234	6
2	7823452323	7
3	8834563423	5
4	9432456778	8
5	7823134257	10
6	8842414514	9
7	7324569708	11
8	7634567800	4

Table 7 Shuffled sensitive information table

S ID	Mobile No.	Annual Income(Lac)
1	9432456778	8
2	7634567800	4
3	7324569708	11
4	7823452323	7
5	8834563423	5
6	8842414514	9
7	9042341234	6
8	7823134257	10

Table 8 Link or Mapping table

G ID	P ID	S ID
1	1	7
1	2	4
1	3	5
1	4	1
2	5	8
2	6	6
2	7	3
2	8	2

Retrieve Operation: Table 8, the link table will be stored on fully secure application server for one to one mapping between record in identifying information table and sensitive information table. A request to retrieve k records from identifying information table and one record from sensitive information table makes it k anonymous.

User can request k record from identity table and one sensitive record to make data k anonymous. The value of k can be selected on the basis of query processing time and sensitivity of the information.

A group of k can be formed with different sensitive information (l-diversity and t-closeness) i.e. in case of customer records with distinct income can be grouped together. In this way the privacy user information is protected and controlled by data owner's even though they use third-party cloud storage.

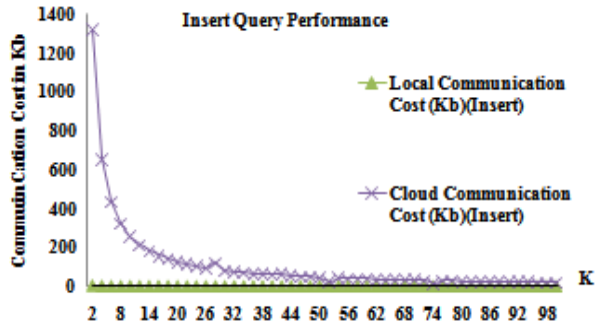
5 Performances and Statistical Analysis

In this section we present the performance analysis of proposed privacy model and also present statistical models to predict response time required for different database operations for varying values of k. To measure the efficiency of the proposed approach, in subsections 5.1 we present the performance of the proposed approach in terms of communication and computation cost. In section 5.2 we present query response time for varying values of k. Section 5.3 presents statistical analysis of parameters considered in performance analysis.

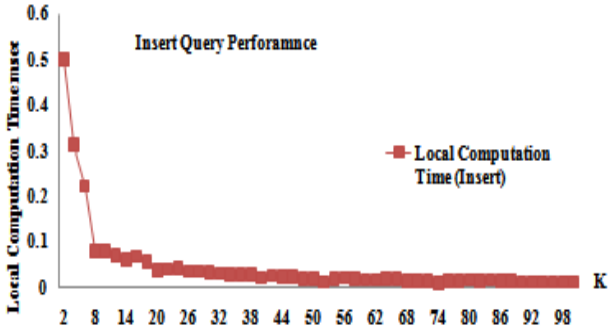
5.1 Insert, Select and Update Query Performance

In experiment ‘k’ numbers of queries are executed to complete all ‘k’ records insertion and update cycle. The average time and data exchange of all types of queries are calculated. For Insert, select and update queries, we vary the value of k = 2,4,6,…,100.

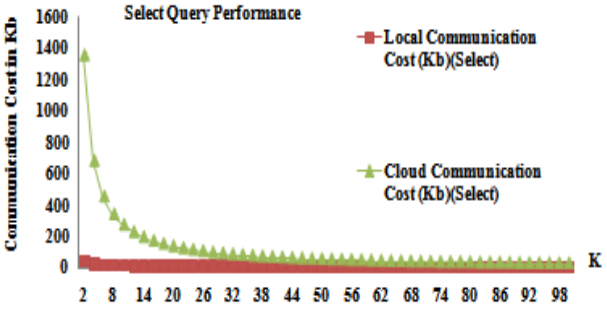
As shown in Fig. 2, data owner side local computation time and communication cost decreases as value of k increases. A similar trend is observed in the cloud communications cost. This justifies that the designed privacy scheme performs well for increased values of k.



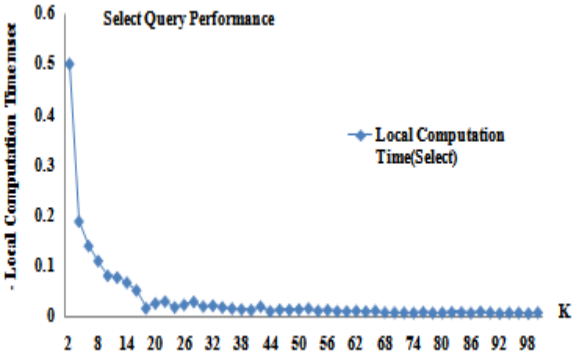
(a)



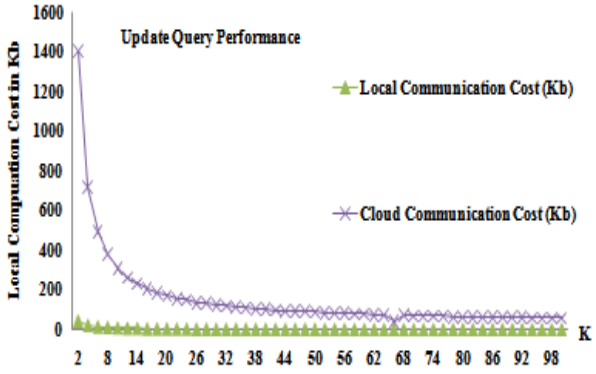
(b)



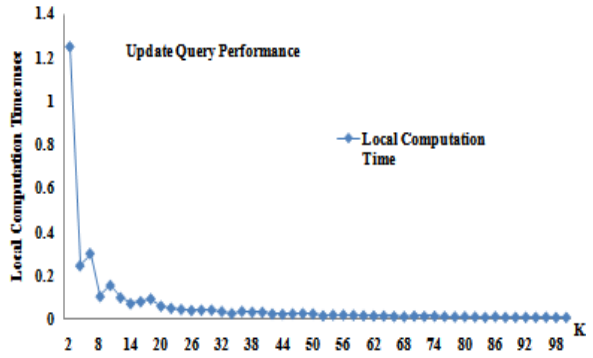
(c)



(d)



(e)



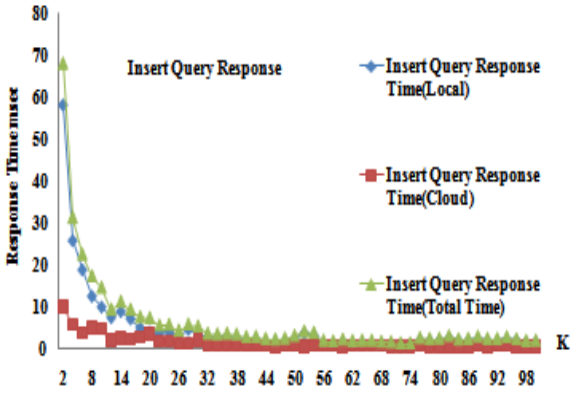
(f)

Fig. 2 (a) Insert Query Performance, (b) Insert Query Performance - Local Computation Time, (c) Select Query Performance (d) Select Query Performance - Local Computation Time, (e) Update Query Performance (f) Update Query Performance - Local Computation Time

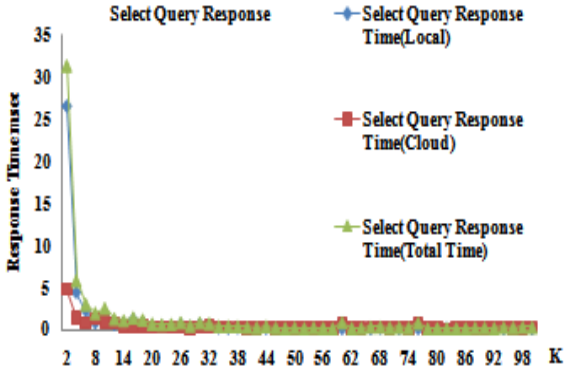
5.2 Insert, Select and Update Query Response Time

The graphs in Fig. 3 show the query response time for insert select and update queries. The response time decreases as the value of k increases.

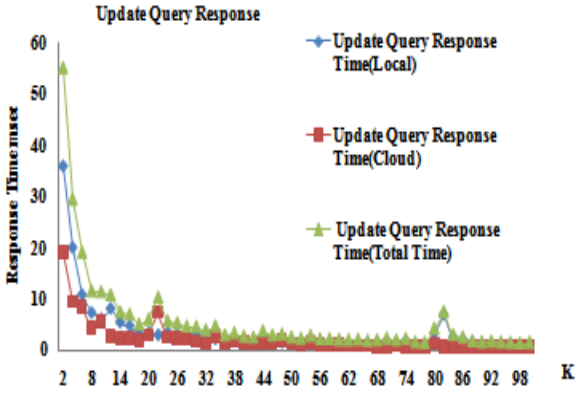
The performance of this scheme is optimal since the insert and update queries are processed in group of size k. The group processing reduces average time required i.e. time for single insert or update query. Also the select query performance achieved using caching this adds local storage and maintenance overhead. Due to caching reduces the select query processing time if data/record is already available on local cache.



(a)



(b)



(c)

Fig. 3 (a) Insert Query Response, (b) Select Query Response Time, (c) Update Query Response Time

5.3 Statistical Regression Models

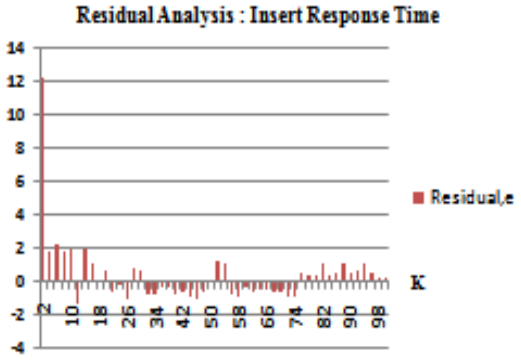
Statistical regression analysis is done using SPSS tool for estimating the relationship among variables. In our experiment k is independent variable which is a measure of privacy, maximum value of k maximizes privacy but at the same time increases local cache and storage requirements. Since our objectives are reduce communication and computation overhead on cloud user site while using cloud services and also improve query response time. All these objectives are achieved using proposed technique.

Table 9 contains list of dependent variables(Y) and equations showing relation of dependent variables with independent variable X=k obtained using regression analysis. In all models we are getting promising value of coefficient of determination i.e. R^2 which is above 0.8.

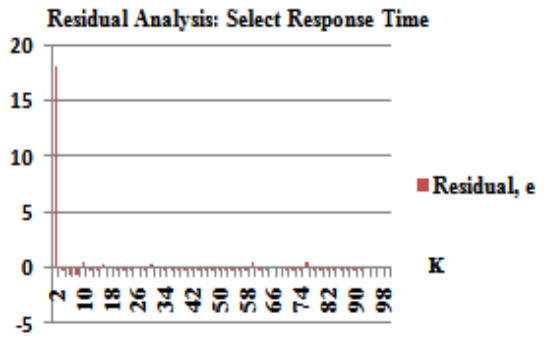
Fig 4(a-i) presents plots of residual analysis of regression models. Residual analysis is done for validation of regression models. Residual is the difference between the observed value of dependent variable (Y) and predicted variable (Y'). For each data point we calculate the residual and then plot the graph. All the residual plots show the random patterns thus they are positive and proves the fitness of regression models for experimental data.

Table 9 Regression Models of System Variables

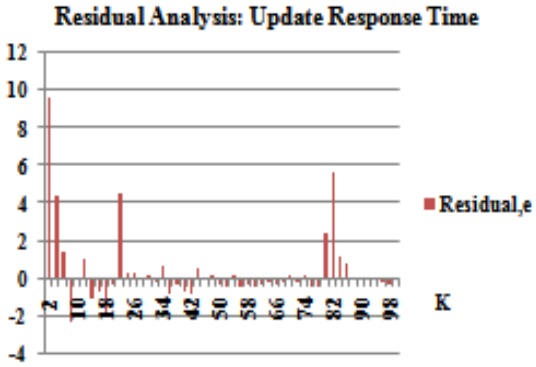
Dependent Variable(Y)	R Square	Equation
Insert Response Time	0.898	$Y=106.52X^{-0.928}$
Select Response Time	0.835	$Y=30.447X^{-1.198}$
Update Response Time	0.87	$Y=82.486X^{-0.857}$
Insert Local Computation Cost	0.96	$Y=0.915X^{-0.991}$
Select Local Computation Cost	0.957	$Y=0.954X^{-1.126}$
Update Local Computation Cost	0.972	$Y=1.47X^{-1.023}$
Insert Local Communication Cost	0.99	$Y=7.297X^{-1.014}$
Select Local Communication Cost	1	$Y=73.983X^{-0.994}$
Update Local Communication Cost	1	$Y=88.149X^{-1.019}$



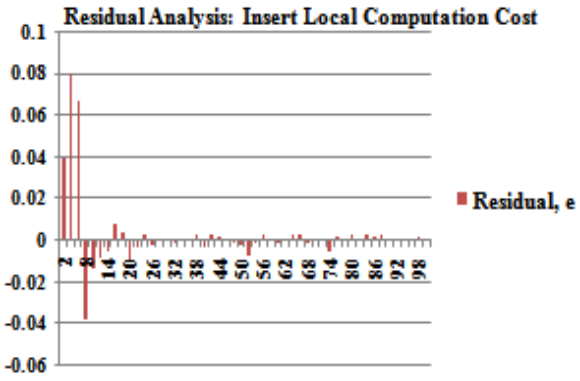
(a)



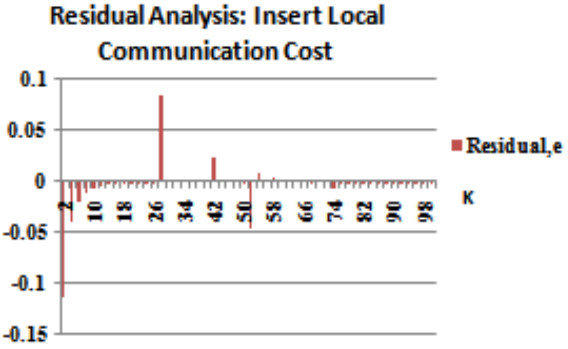
(b)



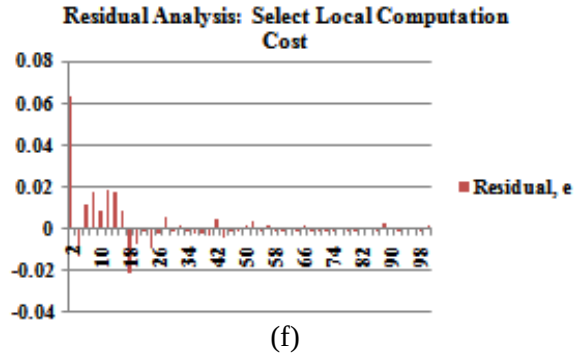
(c)



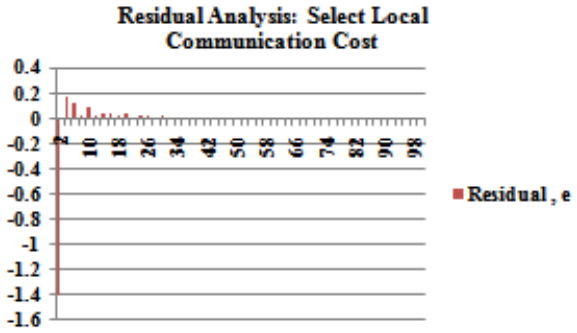
(d)



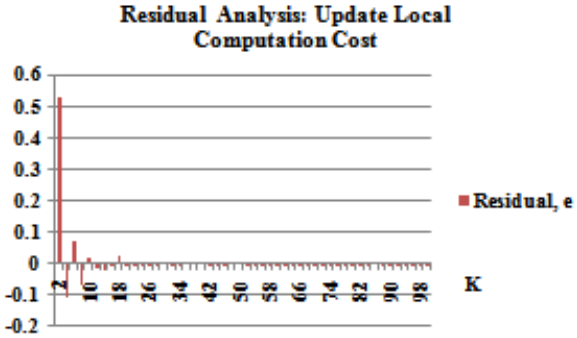
(e)



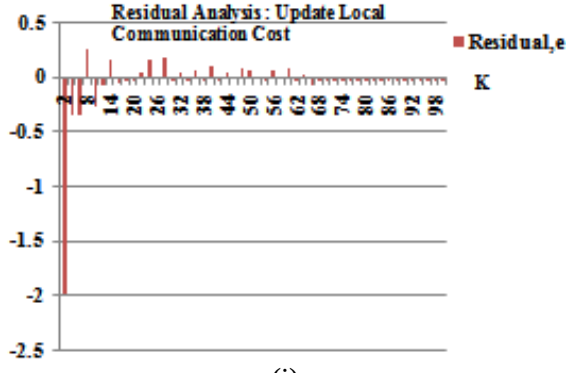
(f)



(g)



(h)



(i)

Fig. 4 Residual Analysis (a) Insert Response Time (b) Select Response Time (c) Update Response Time (d) Insert Local Computation Cost (e) Insert Local Communication Cost (f) Select Computation Cost (g)Select (h) Update Computation Cost (i) Update Communication Cost.

6. Conclusion and Future Work

This paper presents a privacy protection model for storing private sensitive database on cloud. In this paper, we systematically presented the problem of data privacy in cloud storage environment. We presented an efficient privacy protection anonymity technique for cloud based database applications. Also proposed solutions to execute write queries on outsourced database without violating individual privacy. This is the first approach that may not use encryption and also no need to store original database with data owner.

Our experiments and performance analysis shows that the proposed technique is not only feasible but also efficient and robust under various parameter settings. The statistical analysis presents different prediction models. Assessment of models to prove the appropriateness is also given. We believe this work steps towards practical application of designing technique to protect data privacy in the cloud environment.

Advantages of proposed scheme are it is simple to implement and efficient in terms of response time, communication and computation cost. Also it scales well for increasing/more privacy i.e. increased value of k. Limitation of presented scheme is requirement of cashing and temporary local storage, also central application server required at data owner site. The optimal value of k depends on query frequencies i.e. k increases with increase query frequencies and query patterns, so k cannot be generalised. Optimal value of k for specific application can be derived using proposed prediction models.

There are still open opportunities such as query optimization can be done for existing model. We have implemented our model for k anonymous random groups. It can be implemented using strong privacy measures such as l-diversity and t-closeness. In presented work we split table into two tables this can be further improved to provide stronger privacy by splitting table into multiple tables and supporting solution using distributed database management systems and multi cloud distributed storage.

References:

[1] J. J. Yang, J. Q. Li & Y. Niu, A hybrid solution for privacy preserving medical data sharing in the cloud environment. *Future Generation Computer Systems*, 43,2015,pp. 74-86.

[2] L. Sweeney, k-Anonymity: A Model for Protecting Privacy, In *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, volume 10*, 2002, pp. 557–570.

[3] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, Efficient Full-Domain k-Anonymity. *In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, 2005, pp. 49–60.

[4] R. J. Bayardo and R. Agrawal, Data Privacy through Optimal k-Anonymization. *In Proceedings of the 21st International Conference on Data Engineering*, 2005, pp. 217–228.

[5] A. Meyerson and R. Williams, On The Complexity of Optimal k- Anonymity. *In Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2004, pp. 223–228.

[6] S. P. Reiss, M. J. Post, and T. Dalenius, Non-Reversible Privacy Transformations. *In Proceedings of the 1st ACM SIGACT-SIGMOD Symposium on Principles of Database Systems*, 1982, pp. 139–146.

[7] S. P. Reiss, Practical Data-Swapping: The First Steps. *ACM Trans. Database Syst*., 1984, pp. 20–37.

[8] R. Agrawal and R. Srikant, Privacy-Preserving Data Mining. *In Proceedings of the 2000 ACM SIGMOD International Conference on Managementof Data, SIGMOD*, 2000, pp. 439–450.

[9] H. Hacigumus, B. R. Iyer, C. Li, and S. Mehrotra, Executing SQL Over Encrypted Data in The Database-Service-Provider Model. *In SIGMOD Conference*, 2002, pp. 216–227.

[10] R. Sion, Secure Data Outsourcing. In *Proceedings of the conference on Very large data bases*, 2007, pp. 1431–1432.

[11] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, Order Preserving Encryption for Numeric Data, *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, 2004, pp. 563–574.

[12] M. Kantarcioglu and C. Clifton, Security Issues in Querying Encrypted Data, In *Proceedings of the 19th annual IFIP WG 11.3 Working Conference on Data and Applications Security, DBSe*, 2005, pp. 325–337.

[13] J. Li and E. R. Omiecinski, Efficiency and Security Trade-off in Supporting Range Queries on Encrypted Databases. In *Proceedings of the 19th Annual IFIP WG 11.3 Working Conference on Data and Applications*

Security, DBSec, Springer-Verlag, 2005, pp. 69–83.

[14] E. Shmueli, R. Waisenberg, Y. Elovici, and E. Gudes, Designing Secure Indexes for Encrypted Databases, 2005, pp. 54–68.

[15] Z. Yang, S. Zhong and R. N. Wright, Privacy-Preserving Queries on Encrypted Data. In *Proceedings of the 11th European Conference on Research in Computer Security*, 2006, pp. 479–495.

[16] E. Damiani and et al., Balancing Confidentiality and Efficiency in Untrusted Relational DBMSs. In *Proceedings of the 10th ACM Conference on Computer and Communications Security* ACM, 2003, pp. 93–102.

[17] B. Hore, S. Mehrotra, and G. Tsudik, A Privacy-Preserving Index for Range Queries. In *Proceedings of the 13th International Conference on Very Large Databases - Volume 30, VLDB*, 2004, pp. 720–731.

[18] V. Ciriani and et al., Keep A Few: Outsourcing Data While Maintaining Confidentiality. In *Proceedings of the 14th European Conference on Research in Computer Security Springer-Verlag*, 2009, pp. 440–455.

[19] R. Popa, C. Redfield, N. Zeldovich, and H. Balakrishnan, CryptDB: Protecting Confidentiality with Encrypted Query Processing. In *Proceedings of the 23rd ACM Symposium on Operating Systems Principles*, 2011, pp. 85–100.

[20] D. Agrawal, A. El Abbadi, F. Emekci, & A. Metwally, Database management as a service: Challenges and opportunities. In *Data Engineering, ICDE'09. IEEE 25th International Conference*, 2009, pp. 1709-1716.

[21] J. Köhler, K. Jünemann & H. Hartenstein, Confidential database-as-a-service approaches: taxonomy and survey. *Journal of Cloud Computing: Advances, Systems and Applications*, 4(1), 2015, 1.

[22] J. Pei, J. Xu, Z. Wang, W. Wang, and K. Wang, Maintaining k-anonymity against incremental updates. In *Scientific and Statistical Database Management 19th International Conference on*. IEEE, 2007, pp. 5-5.

[23] K. Wang and B. C. M. Fung, Anonymizing sequential releases, *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 414-423.

[24] X. Xiao and Y. Tao, M-invariance: towards privacy preserving re-publication of dynamic datasets. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, 2007, pp. 689-700.

[25] Y. Bu, A. W. C. Fu, R. C. W. Wong, L. Chen, and J. Li, Privacy preserving serial data publishing by role composition, *Proceedings of the VLDB Endowment* 1.1, 2008, pp. 845-856.

[26] A. E. Nergiz, C. Clifton, & Q. M. Malluhi, Updating outsourced anatomized private databases, In *Proceedings of the 16th International Conference on Extending Database Technology*, 2013, pp. 179-190.

[27] Y. He, S. Barman, and J. Naughton. Preventing equivalence attacks in updated, anonymized data. . In *Data Engineering (ICDE), IEEE 27th International Conference on* IEEE, 2011, pp. 529-540.

[28] B. K. Samanthula, Y. Elmehdw,G. Howser & S. Madria. A secure data sharing and query processing framework via federation of cloud computing. *Information Systems*, 48,2015,pp. 196-212.

[29] A. Machanavajjhala, D. Kife, J. Gehrke, & M. Venkitasubramaniam, l-diversity: Privacy beyond k-anonymity, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2007, pp. 3-3.

[30] X. Xiao, & Y. Tao, Anatomy: Simple and effective privacy preservation, In *Proceedings of the 32nd international conference on Very large data bases*, 2006, pp. 139-150.

[31] A. E. Nergiz and C. Clifton, Query processing in private data outsourcing using anonymization, *Data and Applications Security and Privacy XXV., Springer Berlin Heidelberg*, 2011, pp. 138-153.

[32] N. Li, T. Li, and S. Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity, In *ICDE. Vol. 7*, 2007, pp. 106-115

[33] A. Sahai, B. Waters, Fuzzy Identity-Based Encryption, in *Proc. 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Aarhus, 2005, pp. 22-26.

[34] A. Shamir, Identity-Based Cryptosystems and Signature Schemes, in *Proc of CRYPTO,19*85,pp.47-53.

[35] D. Bohen and M. K. Franklin, Identity-based encryption from the weil pairing, In *Proc. CRYPTO*, 2001, pp. 213-229.

[36] C. Gentry and Silverberg, Hierarchical ID-based cryptography, in *Proc. Adv. Cryptography*, 2002, pp. 548-566.