*International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868*
*Foundation of Computer Science FCS, New York, USA*
*International Conference & workshop on Advanced Computing 2013 (ICWAC 2013) – www.ijais.org*

# Securing Anonymous and Confidential Database through Privacy Preserving Updates

Prashant Jawade
Assistant Professor
Thakur College of Engineering &Technology,
Mumbai University, Kandivali (E) Mum-101.

Poonam Joshi
ME Student
Thakur College of Engineering &Technology,
Mumbai University, Kandivali (E) Mum-101.

## ABSTRACT

Typically for economic gain, someone wrongfully obtains and uses another person's personal data in a way that involves fraud or deception. So Privacy of database is a necessity. In this paper we propose a system of updating the confidential database with preserving the privacy of it. Anonymization means identifying information is removed from original data to protect personal or private information. Data anonymization can be performed in different ways but in this paper k-anonymization approach is used. Suppose one person X having his own k-anonymous database and needs to determine whether database is still k-anonymous if tuple inserted by another person Y. For some applications (for example, Patient's record), database needs to be confidential, so access to the database is strictly controlled. The confidentiality of the database managed by the owner is violated once others have access to the contents of the database. Thus, problem is to check whether the database inserted with the tuple is still k-anonymous without letting the owner X and others (Y) to know the content of the tuple and database respectively. In this paper, we propose two methods solving this problem on suppression and generalization based k-anonymous and confidential database. Beside we are dealing with the case of malicious parties by the introduction of non-colluding third party.

## General Terms

Database, Security, Privacy , Confidentiality.

## Keywords

Suppression, Generalization, K-anonymity.

## 1. INTRODUCTION

Today society is growing in terms of data collection, containing personal details or some other confidential information. Database modeled as collection of data that can be accessed, updated and it enables user to retrieve data. To provide security to these databases is big issue. For example, Medical data of each patient should be protected for years. There are different approaches to protect database. The emphasis in database privacy should fall on a balance between confidentiality, integrity and availability of personal data, rather than on confidentiality alone. Database privacy concerns the protection of information about individuals that is stored in a database. Sometimes, without your knowledge, health records used by insurance companies, drug manufacturing companies. Because medical records may contain some sensitive information like effect of drug on patient it is important to keep this information private. Confidential data is personal information relating to a person, it could be reference to an identification number or other factors like social identity. To maintain confidentiality, unauthorized third parties must be prevented from accessing and viewing medical data. It is also essential to maintain database integrity while data is transferring from source to destination. Confidentiality is achieved by Cryptography methods or tools. Though data is anonymized, Confidentiality is necessary. Anonymized or anonymization means remove personal identifier to protect private information. There are many ways of anonymization but we will focus on k-anonymization approach only. Data anonymization enables transferring information between two organizations by converting text data into non human readable form using encryption method remove personal identifier to protect private information. There are many ways of anonymization but we will focus on k-anonymization approach only. Data anonymization enables transferring information between two organizations, by converting text data into non human readable form using encryption method[1].There have been lots of techniques developed k-anonymization is one of the approach[2].This technique protects privacy of original data by modification. So problem arises at this point where database needs to be updated. So when tuple is to be inserted in the database problems occurs relating to privacy and confidentiality that Is database owner decide that whether database preserve privacy without knowing what new tuple to be inserted k-anonymization is one of the approach [2].This technique protects privacy of original data by modification. So problem arises at this point where database needs to be updated. So when tuple is to be inserted in the database problems occurs relating to privacy and confidentiality that Is database owner decide that whether database preserve privacy without knowing what new tuple to be inserted.

Consider a data owner, such as a hospital or a bank that has a privately held collection of person-specific, field structured data. Suppose the data owner wants to share a version of the data with researchers. How can a data owner release a version of its private data with scientific guarantees that the individuals who are the subjects of the data cannot be re- identified while the data remain practically useful? The aim of this project is to preserve the user sensitive data from misuse, by including a formal protection model named k-anonymity and the connection is kept anonymous. Based on the generalization and suppression method data is grouped and the anonymization is done. The original data is encrypted using a cryptography method and stored for the user data backup. Anonymous connection method is also used to prevent user identification though its IP address [12].

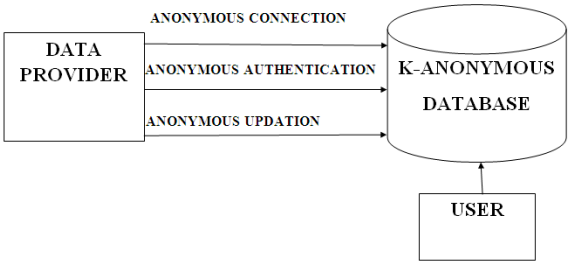


**Fig 1: Anonymous Database**

The problem occurs when it comes to the updating of the database. When the tuple is to be inserted into the database, there are two problems: Is updated database still maintains privacy? And owner of the database really know data to be inserted? Figure1 shows approach to update anonymous database. Data provider wants to insert data in the database and Database owner is user who owns the database. Database should be updated in such manner that data provider privacy is protected and confidentially of database is not violated. Here Anonymous connection and Anonymous authentication is established between Data provider and Database owner so that they are not aware of each other data and thus privacy of Data provider is preserved and confidentially of database is maintained.

## 2. RELATED THEORY

A number of techniques have been developed to provide privacy to the databases such as, randomization, secure multi-party computation and k- anonymity.

### 2.1 Randomization

The randomization method provides effective way of preventing the user from learning sensitive data which can be easily implemented because the noise added to the given record is independent from the other records. The amount of noise is large enough to smear original values, so individual record cannot be recovered. The randomization method is simple as compare to other methods because it does not require to knowledge of other records. Large randomization increases the uncertainty and the personal privacy of the user's .They claim that approaches may lose information as well as not provide privacy by introducing random noise to the data by using random matrix properties, [13]. It successfully separates the data from the random noise and subsequently discloses the original data.

### 2.2 Secure multi-party computing

Goal of secure party computation is to compute function when each party has some input. It generally deals with problems of function computation with distributed inputs. In this protocol, parties have security properties e.g., privacy and correctness regarding privacy a secure protocol must not reveal any information other than output of the function. *K-anonymity* and SMC are used in privacy-preserving data mining, but they are quite different in terms of efficiency, accuracy, security and privacy [15]. The goal of methods for secure multi-party computation is to enable parties to jointly compute a function over their inputs, while at the same time keeping these inputs private**.**

### 2.3 K-anonymous

A large number of privacy models were developed most of which are based on the *k*-anonymity property. The K-anonymity model was proposed to deal with the possibility of indirect identification of records form public databases-anonymity means each released record has at least (k-1) other records in the release whose values are indistinct [2].

## 3. LITERATURE REVIEW

K-anonymity was carried in2002 by L.Sweeney**,** The solution provided in this paper includes a formal protection model named *k*-anonymity and a set of accompanying policies for deployment. [2]. Security Security-Control Methods for Statistical Databases: A Comparative Study was carried in 1989 by N.R. Adam and J.C. Wortmann deals with algorithms for Database anonymization. Here idea of protecting database through data suppression or data perturbation has been extensively investigated [1]. The Role of Cryptography in Database Security was carried in 2004 by E. Bertino and R. Sandhu, the problem of defining and achieving security in a context where the database is not fully trusted, i.e., when the users must be protected against a potentially malicious database [5]. Executing SQL over Encrypted Data in Database Service Provider Model **w**as carried in 2002 by H. Hacigu¨mu¨ s, B. Iyer, C. Li, and S. Mehrotra**,** In this research direction is related to query processing techniques for encrypted data These approaches do not address the k-anonymity problem since their goal is to encrypt data, so that their management can be outsourced to external entities [6] . Privacy-Enhancing k-Anonymization of Customer Data was carried in 2005 by S. Zhong, Z. Yang, and R.N. Wright**,** The problem of computing a k-anonymization of a set of tuple while maintaining the confidentiality of their content. However, these proposals do not deal with the problem of private updates to k-anonymous databases [3].Privacy Preserving Incremental Data Dissemination was carried in 2009 by J.W. Byun, T. Li, E. Bertino, N. Li, and Y. Sohn.The problem of protecting the privacy of time varying data has recently spurred intense research activity [4]. Crowds: Anonymity for Web Transactions was carried in 1998 by Michael K. Reiter and Aviel D. Rubin AT&T Labs Research. It introduces a system called Crowds for protecting users' anonymity on the world-wide-web [8]. Anonymzing Sequential Releases was carried in 2006 by K. Wang and B. Fung, The issue here is how to anonymized current release so that it cannot link to previous releases yet it remains useful to its own release purpose [10]. Practical Techniques for Searches on Encrypted Data was carried in 2000 by D.X. Song, D. Wagner, and A. Perrig, which consist of description various cryptographic schemes for the problem of searching on encrypted data and provides proofs of security for the resulting crypto systems approach [7]. Information Sharing across Private Databases was carried in 2003 by R. Agrawal, A. Evfimievski, and R. Srikant, in this information on integration across databases tacitly assumes that the data in each database can be revealed to the other databases [9]. Anonymous Connections and Onion Routing was carried in1998by M. Reed, P. Syverson, and D. Goldschlag, this research describe Onion routing, it provide infrastructure for providing private communication through public network and also provide anonymonous connection [12]. Public Key Encryption with keyword Search was carried in 2004 by D. Boneh, G. diCrescenzo, R. Ostrowsky, and G. Persiano .In this study of problem of searching on data that is encrypted

using a public key system is done. It defines the concept of public key encryption with keyword search and gives several constructions [11].

## 4. PROBLEM STATEMENT

Various methods for anonymity of Database have been developed. Existing works are still subject to some drawbacks. Number of approach has been proposed but existing works has some serious limitations, in that they do not support generalization-based updates, which is the main strategy adopted for data anonymization Therefore, if the database is not anonymous with respect to a tuple to be inserted, the insertion cannot be performed. To achieve the objective to check whether the database inserted with the tuple is still k-anonymous, without letting admin and user know the contents of the tuple and the database. Also to ensure Privacy Preserving and Confidentiality during Updating of Database we propose two methods solving this problem that are suppression-based and generalization-based algorithms. Beside we are dealing with the case of malicious parties by the introduction of, non-colluding third party.

## 5. PROPOSED METHOD

The proposed system relies on the fact that anonymity of database does not affected by inserting tuple if the information contained in tuple is properly anonymized and is already contained in database. The problem arise of privately checking whether there is a match between tuple to be inserted and tuple that already contained in database. To solve this problem we propose two algorithm that is suppression based and generalization based algorithm.

### 5.1 Proposed System

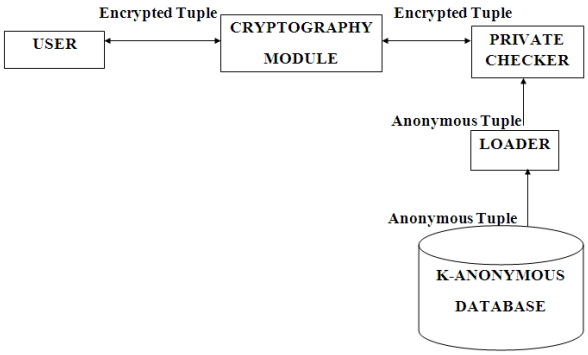


**Fig 2: Proposed System Block Diagram**

In figure2, Proposed System is made of Cryptography module, Private checker, Loader module. User enters data is stored into Cryptography module which is responsible for encrypting all tuples exchanged between User and the private Updater. Loader module reads data or tuple from anonymous database. Private Checker module performs all the controls that is to checks whether the inserted data is matched with the data's in the anonymous database using suppression and generalization method. The main concept behind Private checker is to check whether insertion is possible into database. Proposed system in the figure2 compares existing data and the updates and make sure there is no redundancy and helps to analyses the data in database. K-Anonymization allows database to maintain a suppressed and generalized form of data such that data is much secured. The cryptography technique is used to secure the saved data in database safely such that the information is encrypted, stored and can be retrieved and decrypted back to original with specific authorization. The cryptographic scheme used in this paper is DES (Data Encryption Standard) algorithm along with hash encoding.

### 5.2 System Architecture

The figure 3 shows the flow of steps followed in the system. It starts with authentication of user. Each user is provided with username and password registered in system already. The authentication user has access to the database and system has particular access rights for each user. The anonymous database suppresses and generalizes the data according to data value. The database can be accessed by research centers forgathering statistical data regarding particular medicines, the percentage of curable medicines. The internal or private information of the patients are not revealed to the research centre computation. The research people can see the data's send by the database according to its access right. And allocate research peoples to each research data. And forward the data to research people. Here research people can't do any changes or modifications in patient database they only can use the database for reference purpose. Information of the patients is not revealed to the research centre computation. The research people can see the data's send by the database according to its access right. And allocate research peoples to each research data. And forward the data to research people. Here research people can't do any changes or modifications in patient database they only can use the database for reference purpose. The anonymous database can forward information to research centre which has permissions to access the information. The research centre access has its own restrictions for use of the data. They can access only superficial data and whatever data that they have access to. They cannot access the patient details as shown in figure 3.

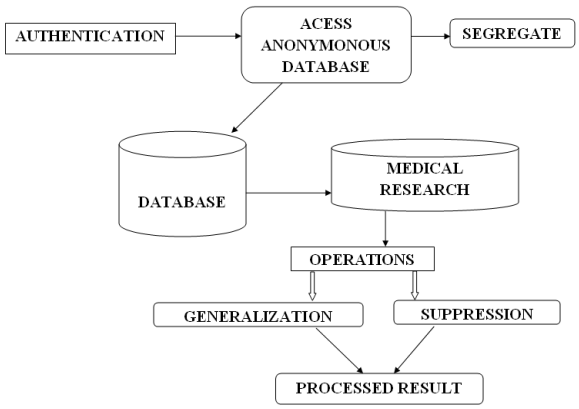


**Fig 3: System Architecture**

### 5.3 Suppression Based Algorithm

The idea used in the Suppression algorithm is to mask some attributes by special value *, In suppression algorithm t stands for private tuple provided by Data provider, T stands for Anonymous database, QI stands for Quasi-Identifier which consist of set of attributes that can be used with certain external information to identify a specific individual.

The suppression algorithm works as follows:

Step1: User X sends User Y an encrypted version containing only the s non-suppressed attributes.

Step2: User Y encrypts the information received from User X and sends it to her, along with encrypted version of each value in his tuple t.

Steps 3-4: User X examines if the non suppressed QI attributes is equal to those of t. If true, t can be inserted to table T. Otherwise, when inserted to T, t breaks k- anonymity.

## 5.4 Generalization Based Algorithm

In generalization algorithm attributes are replaced with general value based on Value Hierarchy Graph (VGH) [14].

The protocol works as follows:
Step 1: User X randomly chooses a $\delta \varepsilon T_w$ (Witness Set).
Step 2: User X computes $\gamma$ = GetSpec ($\delta$).
Step 3: User X and User Y collaboratively compute s= SSI ($\gamma$, $\tau$).
Step 4: If s=u then t's generalized form can be safely inserted to T.
Step 5: Otherwise, User X repeats the above procedures until either s=u or witness set is empty.

Let t is User Y's private tuple from table T containing anonymous attributes ,so User B can generate $\tau$ which holds corresponding values t[A1],…,t[Au];Let u (Size of anonymous tuple)be disjoint value Generalization hierarchies corresponding to anonymous attributes known to User X. Let $\delta \in T$ and let Getspec ($\delta$) be specific value that is bottom of VGH (Value Graph Hierarchy) related to each anonymous attribute [14]. Function $\gamma$ denotes to *GetSpec* ($\delta$).Now, User Y generates a set $\tau$ containing corresponding values to tuple t.We use Secure Set Intersection (SSI) protocol to compute cardinality of set. Here we denote SSI ($\gamma$, $\tau$) as a secure protocol which computes cardinality of $\gamma \cap \tau$ .On the receiving first request User X chooses random tuple from table T. User X computes function $\gamma$ = GetSpec ($\delta$). User Y and User X individually compute SSI ($\gamma$, $\tau$).next is step to compare SSI ($\gamma$, $\tau$) with u.If both are equal then t is in generalized form and can be inserted in database. Otherwise it again get computes until we get both values same.

# 6. RESULT AND DISCUSSIONS

Suppression based k-anonymity approach to provide privacy preserving updates to confidential database is designed. Data entered by the user directly replaced by special value '*' and these values being inserted into table. To carry out this task, we have made separate table for original values. When user enters data it checks value in original table if it is valid then it replaces original value with suppressed values. Based on this outcome data will get inserted or rejected. .We can make result that if all values entered by data provider is correct then database will be updated successfully otherwise tuple will not be inserted to the database. Thus we can say that database successfully updated while preserving privacy and k-anonymity. Generalization based k-anonymity approach to provide privacy preserving updates to confidential database is designed. If the data entered by the user matches with the general value then this record will replaced by the general value and these general values being inserted into table. To carry out this task, we have made separate table for original values and general values. When user enters data it checks value in original table if it is valid then it matches with the values of generalized table. Based on this outcome data will get inserted or rejected. As a result if we enter all values correct then database will be updated successful otherwise tuple will not be inserted to the database. Thus we can say that database successfully updated while preserving privacy and k-anonymity. Figures below shows various windows involved.

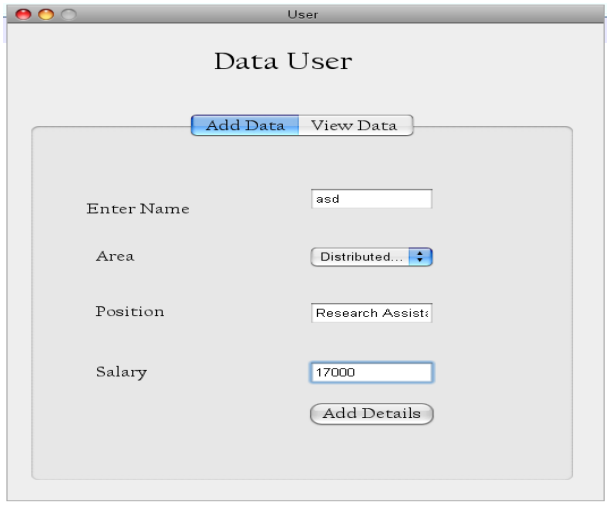


**Fig 4: Add User Detail Window**

Add User Detail Windows is used to add information about the user in our case we enter information such as Name, Area, Position and Salary of college staffs as shown in figure 4.

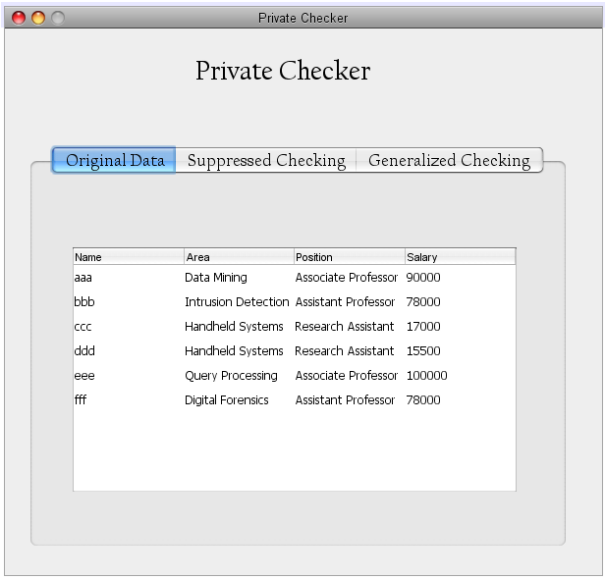


**Fig 5: Private Checker Window with Original Data**

Private Checker Window with Original Data as shown in figure 5 provides information about the original Data in Database and is displayed only to the administrator.

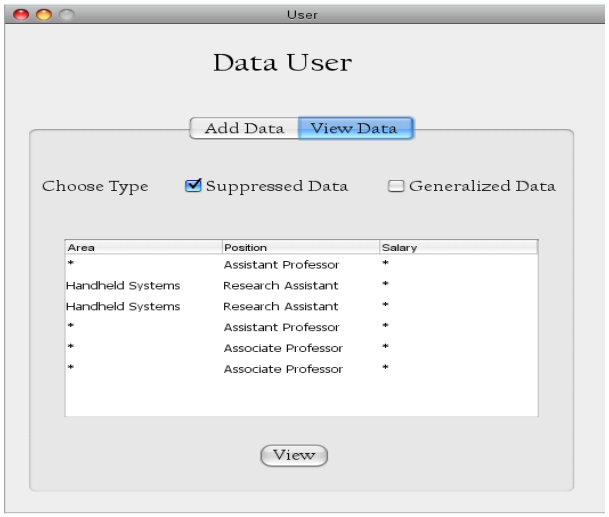


**Fig 6: User Suppressed Data View**

User Suppressed Data View is module used to provide suppressed view to critical data, the original value is replaced by '*'. In figure 6 suppressed view of data is selected and attributes such as Area and Salary are masked with special value '*' Thus suppressed value preserved private information of individuals from displaying.

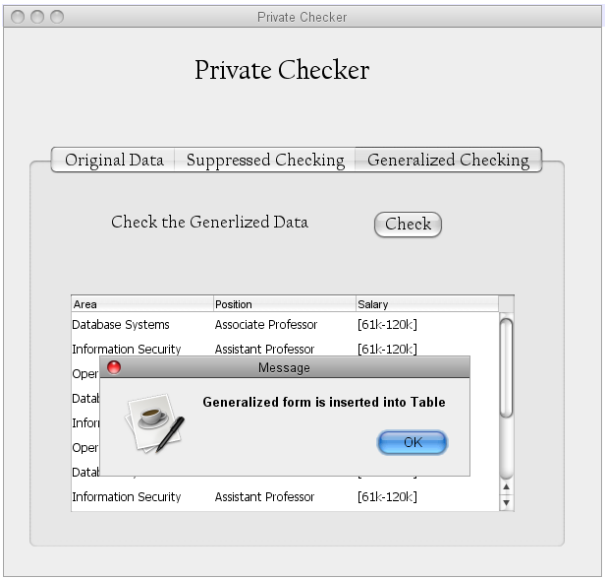


**Fig 7: Private Checker Window with Generalized Data**

Private Checker Window with Generalized Data is used to generalize the original Data. The attributes are replaced with more general value for example Area field original value Data mining is replaced with Database System and Salary field original value is replaced by range as shown in figure 7

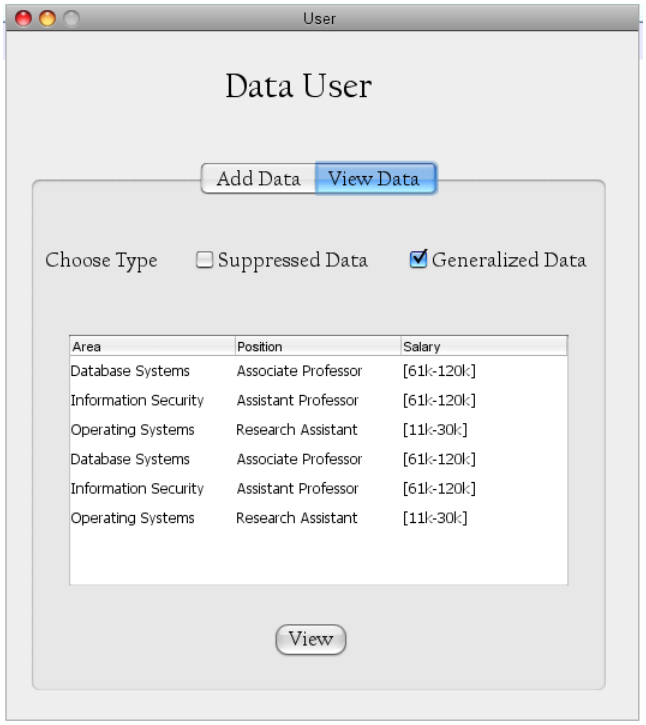


**Fig 8: User Generalized Data View**

User Generalized Data View is module used to provide generalized view to critical data; the original value is replaced by general value according to VGH [14]. In figure8 generalized view is selected. The attributes are replaced with more general value for example Area field original value Data mining is replaced with Database System and Salary field original value is replaced by range.

## 7. CONCLUSION

In this paper two secure methods Suppression and Generalization are proposed that is used to check that if new tuple is being inserted to the database, it does not affect anonymity of database. It means when new tuple get introduced, k-anonymous database retains its anonymity. Database updates has been carried out properly using proposed methods. Execution shows that once system verifies user tuple, it can be safely inserted to the database without violating k-anonymity. Only user required to send non-suppressed attributes to the k-anonymous database. Thus the database is updated properly using the proposed methods. The data provider's privacy cannot be violated if user updates a table. If updating any record in database violate the k-anonymity then such updating or insertion of record in table is restricted. If insertion of record satisfies the k-anonymity then such record is inserted in table and suppressed sensitive information attribute used to maintain the k-anonymity in database. Thus such k-anonymity in table makes difficult for unauthorized user to identify the record.

The important issues in future will be resolved:

1) Implement database for invalid entries.

2) Improving efficiency of protocol in terms of number of messages exchanged between user and database.

3) Implement real world database system.

4) Devising private update techniques to database systems that supports notions of anonymity different than k-anonymity

## 8. ACKNOWLEDGEMENT

I must thank, first and foremost, my guide Mr. Prashant Jawade, who gave me opportunity to write this paper without his guidance and patience, this dissertation would not be possible. Finally, I thank him to review my paper and his invaluable suggestion that make to improve quality of my paper.

## 9. REFERENCES

[1] N.R. Adam and J.C. Wortmann, "Security-Control Methods for Statistical Databases: A Comparative Study," ACM Computing Surveys, vol. 21, no. 4, pp. 515-556, 1989.

[2] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[3] S. Zhong, Z. Yang, and R.N. Wright, "Privacy-Enhancing k-Anonymization of Customer Data," Proc. ACM Symp.Principles of Database Systems (PODS), 2005.

[4] J.W. Byun, T. Li, E. Bertino, N. Li, and Y. Sohn, "Privacy-Preserving Incremental Data Dissemination," J. Computer Security,vol. 17, no. 1, pp. 43-68, 2009.

[5] U. Maurer, "The Role of Cryptography in Database Security,"Proc. ACM SIGMOD Int'l Conf. Management of Data, 2004.

[6] H. Hacigu¨mu¨ s¸, B. Iyer, C. Li, and S. Mehrotra, "Executing SQLover Encrypted Data in the Database-Service-Provider Model,"Proc. ACM SIGMOD Int'l Conf. Management of Data, 2002.

[7] D.X. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp.Security and Privacy, 2000.

[8] M.K. Reiter and A. Rubin, "Crowds: Anonymity with Web Transactions," ACM Trans. Information and System Security (TISSEC), vol. 1, no. 1, pp. 66-92, 1998.

[9] R. Agrawal, A. Evfimievski, and R. Srikant, "Information Sharing across Private databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2003.

[10] K. Wang and B. Fung, "Anonymizing Sequential Releases," Proc. ACM Knowledge Discovery and Data Ming Conf. (KDD), 2006.

[11] D. Boneh, G. di Crescenzo, R. Ostrowsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Euro crypt Conf., 2004.

[12] M. Reed, P. Ryerson, and D. Goldschlag, "Anonymous Connections and Onion Routing," IEEE J. Selected Areas in Comm., vol. 16, no. 4, pp. 482-494, May 1998.

[13] Kargupta H. Datta, S. Q. Wang and K. Sivakumar,"*On the privacy preserving properties of random perturbation techniques*"IEEEICDM, 2003.

[14] Privacy –Preserving Updates to Anonymous and Confidential Databases, Alberto Trombetta, Wei Jiang, Elisa Bertino and Lorenzo Bossi Department of Computer Science and Communication, University of Insure, Italy 2011.

[15] Current Developments of *k*-Anonymous Data Releasing, Jiuyong Li, Hua Wang, Huidong Jin, Jianming Yong, School of Computer and Information Science, University of South Australia, Mawson Lakes Adelaide, Australia, 5095.