# Anonymous and Confidential Database Security through Privacy Preserving Updates

Poonam Joshi[1], Prashant Jawade[2]

[1]*Information Technology Department, Thakur College of Engineering and Technology, Mumbai University, India*
[2]*Information Technology Department,Thakur College of Engineering and Technology, Mumbai University, India*

[1]pnmjoshi2@gmail.com
[2]prashant.jawade@thakureducation.com

*Abstract*— **Information Security has become crucial issue since the information sharing have a become need of present technology. Privacy of information is a necessary to avoid fraud and theft for economic growth. New advances methods in data mining and knowledge discovery allow extraction of hidden knowledge in an enormous amount of data impose new threats on the seamless integration of information.Anonymization means identifying information is removed from original data to preserve private information. Data anonymization can be performed in different ways but in this paper k-anonymization approach is used. The data owners directly read the contents of the database simply it breaks the privacy of the user's data. If the users access the database content directly then the confidentiality of the data owners has been violated. So both privacy and confidentiality of the database are considered to be a major problem. In the existing system the privacy information gets lost in large amount and does not provide any security mechanism. It is impossible to consider every possible inference and also drastically reduces the quality of the data. The proposed system considers the privacy information is more valuable both in research and business areas. So the data sharing is common in order to provide the data remain k-anonymous even after the updates. In this paper, we propose two methods solving this problem on suppression-based and generalization-based k-anonymous and confidential databases. We are introducing concept of non colluding third party for dealing with case of malicious parties.**

*Keywords*— **Privacy, Confidentiality, Anonymous, Suppression, Generalization.**

## I. INTRODUCTION

Database is important and critical thing for application so their security is very important .Data in the databases has its own relevant value. For example medical data collected by over the history of patients over years is an invaluable asset, which needs to be secured and can be used by people in various related areas of work [13]. Nowadays, privacy accidents have become common problem in the information systems. For example, a hospital may have record of all the patients with various diseases critical and non-critical. If the hospital wishes to reveal the data to any pharmaceutical company or online market services, it should not be able to infer with particularity of patients with those diseases. It can give as a statistical view or just the superficial information such that privacy is not detained. There are huge numbers of databases that hold number of confidential information's such that people access those data correlating various information from various databases. Disclosure of confidential information to unauthorized persons may lead to data insecurity leading to dissatisfaction to users. For example, there was a company which sold health products online that also revealed the customer names phone numbers credit card numbers etc on the website. It leads to huge loss of information and breach of privacy. There was another issue when a researcher was enabled to retrieve health records from anonymous databases of insurance claims of employees. Privacy relates to what data can be safely disclosed without leaking information regarding the legitimate owner [14]. Thus, if one asks whether confidentiality is still required once data have been anonymized, the reply is yes if the anonymous data have business value for the party owning them or the unauthorized disclosure of such anonymous data may damage the party owning the data or other parties. The term anonymized or anonymization means identifying information is removed from the original data to protect personal or private information. There are many ways to perform data anonymization. We only focus on the k-anonymization approach. To better understand the difference between confidentiality and anonymity, consider the case of a medical facility connected with a research institution. Suppose that all patients treated at the facility are asked before leaving the facility to donate their personal health care records and medical histories (under the condition that each patient's privacy is protected) to the research institution, which collects the records in a research database. To guarantee privacy to each patient, the medical facility only sends to the research database an anonymized version of the patient record. Once this anonymized record is stored in the research database, the nonanonymized version of the record is removed from the system of the medical facility. Thus, the research database used by the researchers is anonymous. While k-anonymity protects against identity disclosure, it is insufficient to prevent attribute disclosure. Generalization and suppression such technique provides privacy by modifying data in such a way that it gives the same result for more than two tuples. So the problems of confidentiality and anonymization are different. 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 reply is yes if the anonymous data have business value for the party owning them or the unauthorized disclosure of such anonymous data may damage the party owning the data or other parties. The term anonymized or anonymization means identifying information is removed from the original data to protect personal or private information. There are many ways to perform data anonymization. We only focus on the k-anonymization approach [2]. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting Data Provider and Database owner know the contents of the database and the tuple, respectively. In this paper, we propose two protocols solving this problem on suppression-based and generalization-based k-anonymous and confidential databases.

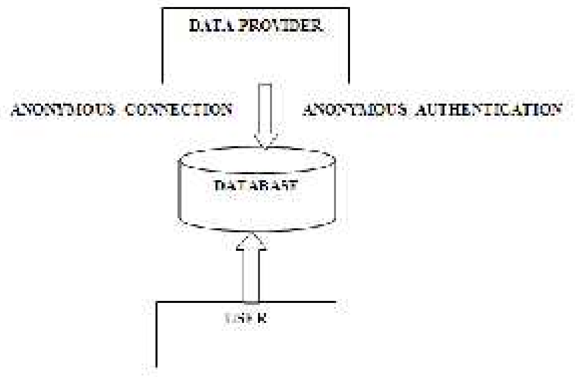


Fig.1 Anonymous Database

Fig.1 shows technique to update anonymous database. Suppose User A is user who owns the database and data provider that is User B who wants to insert his own tuple.So it is necessary to check whether it is possible to update database or insert tuple without knowledge of User B so that privacy of User B cannot be violated and confidentiality of database does not violated if User B have access to the contents of database.

## II. LITERATURE REVIEW

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]. K-anonymity was carried in 2002 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]. 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 tuples 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]. 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 was 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, can be outsourced to external entities [6] . 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]. 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].

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]. 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]. 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]. Anonymous Connections and Onion Routing was carried in 1998 by 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].

## III. PROBLEM STATEMENT

Suppose hospital has some private data which has important details of every patient. Now hospital want to send these details to research institute for some specific purpose with a primary concern that privacy of each patient should not violate by disclosing his data to researchers. If we permit each patient to add data directly into the database then database confidentiality will be broken and if we permit researchers to read every detail of patient then privacy of patient will break. So to preserve privacy and confidentiality we have proposed approach namely suppression anonymous based method and generalization based method so as to maintain the privacy of the patient. The meaning of anonymity is to remove identifying entity from the database. Beside we are dealing with the case of malicious parties by the introduction of non-colluding third party.

## IV. PROPOSED SYSTEM

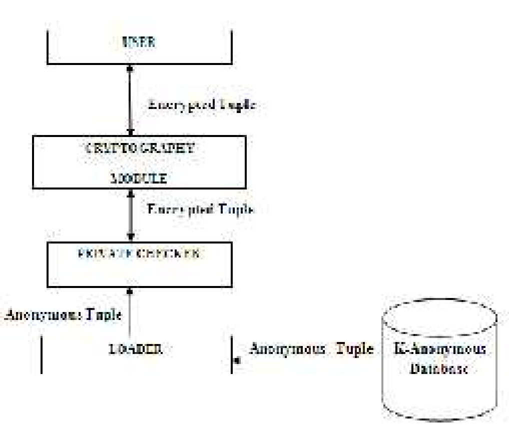


Fig.2 Proposed System

The proposed system consists of User, Cryptographic module, Private Checker module, Loader, and k-anonymous Database. User enters data is stored into encryption module which is responsible for encrypting all tuples exchanged between User and the private Updater. Loader module reads data or tuple from anonymous database .Checker module that performs all the controls that is to checks whether the inserted data is matched with the data's in the anonymous database using generalization method or suppression method. The main concept behind private checker is to check whether insertion is possible into the k anonymous database.

## V. SYSTEM ARCHITECTURE

K-Anonymity is a method for providing privacy preservation by ensuring that data cannot be displayed to an individual. The main purpose is to protect individual privacy. In a k-anonymous dataset, if any identifying information is found in the original dataset with k tuples then first we identify quasi-identifiers i.e. the tuple that clearly distinguish the given tuple in database. Then we are applying for suppression based algorithm. In this algorithm we are identifying quasi-identifiers and we are computing a k-partition which is a collection of disjoint subsets of rows in which each subset contains at least k rows and the union of these subsets is the entire table. And next we are replacing each record having with. In suppression based approach we are applying DES (Data Encryption Standard) algorithm to encrypt and decrypt data by using the shared key. In this approach we are dealing with encrypted data not directly with the original data. When user enters his information then we are encrypting his information by using DES and we are also encrypting all data in table using same algorithm. If information from user matches with table information this tuple will decrypted and inserted into table. In Generalization based Approach we are replacing the value in table with the more general values. If the data entered by the user matches with the value being replaced by the general value then this record will replaced by the general value and these general values being inserted into table.

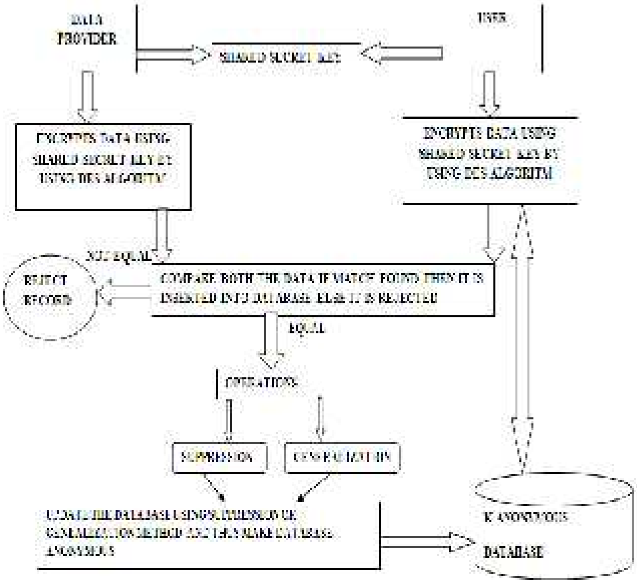

DATA PROVIDER
SHARED SECRET KEY
USER
ENCRYPTS DATA USING SHARED SECRET KEY BY USING DES ALGORITM
ENCRYPTS DATA USING SHARED SECRET KEY BY USING DES ALGORITM
NOT EQUAL
REJECT RECORD
COMPARE BOTH THE DATA IF MATCH FOUND THEN IT IS INSERTED INTO DATABASE ELSE IT IS REJECTED
EQUAL
OPERATIONS
SUPPRESSION
GENERALIZATION
UPDATE THE DATABASE USING SUPPRESSION OR GENERALIZATION METHOD AND THEN MAKE DATABASE ANONYMOUS
K ANONYMOUS DATABASE

Fig. 3 System Architecture

Overview of System Architecture is in fig.3which 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 cryptographic technique is used to secure data 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.

## VI. SUPPRESSION BASED METHOD

Consider Table T = {t1… tn} over the attribute set A. The idea of this algorithm is mask some attributes by special value *, the value employed by User A for the anonymization.In suppression based method, every attribute is suppressed by *.So third party cannot differentiate between any tuples. Here, k-anonymity indicates that is each row in the table cannot be distinguished from at least other k-1 rows by only looking a set of attributes. We assume that the database is anonymized using suppression based method.

The protocol works as follows:

**Step1**: User A sends User B an encrypted version containing only the s non-suppressed attributes.

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

**Steps3**: User A 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.

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.

## VII. GENERALIZATION BASED METHOD

For generalization-based anonymization, we assume that each attribute value can be mapped to a more general value. The main step in most generalization based k-anonymity protocols is to replace a specific value with a more general value.

The protocol works as follows:

**Step 1:** User A randomly chooses a δ ε $T_w$ (Witness Set).

**Step 2:** User A computes γ = GetSpec (δ).

**Step 3:** User A and User B collaboratively compute s= SSI (γ, τ).

**Step 4:** If s=u then t's generalized form can be safely inserted to T.

**Step 5:** Otherwise, User A repeats the above procedures until either s=u or witness set is empty.

Let t is User B'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 A. 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 γ denotes to *GetSpec* ($\delta$).Now, User B 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(γ, $\tau$) as a secure protocol which computes cardinality of γ ∩$\tau$ .On the receiving first request User A chooses random tuple from table T. User A computes function γ = GetSpec ($\delta$). User A and User B individually compute SSI (γ,$\tau$).next step to compare SSI (γ, $\tau$) with u.If both are equal then t in generalized form can be inserted in database. Otherwise it again get computes until we get both values same.

## VIII. RESULT AND DISCUSSION

Suppression based k-anonymity approach to provide privacy 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 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 result windows involved.

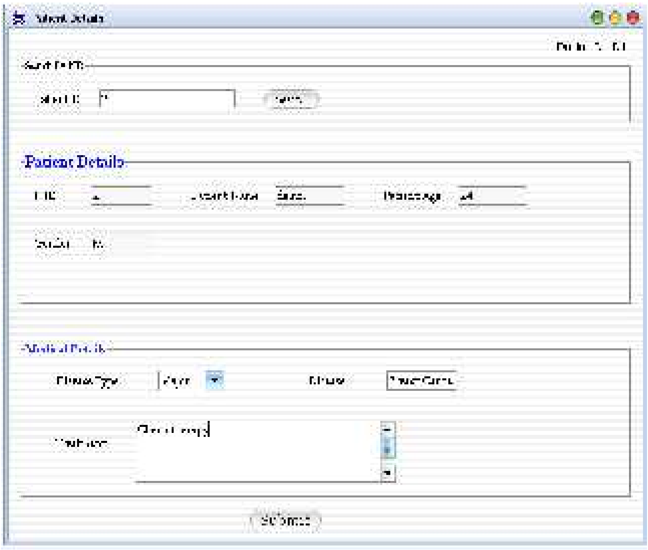


Fig. 4 Patient Details Form

Patient Detail Form is shown in Fig.4. It is used to register patient details.

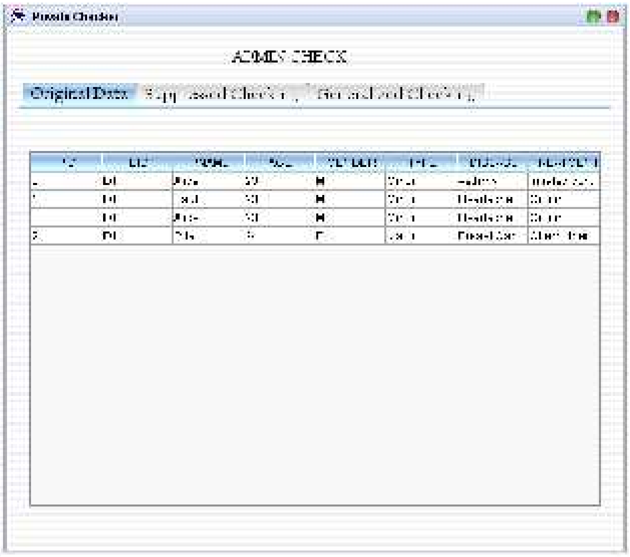
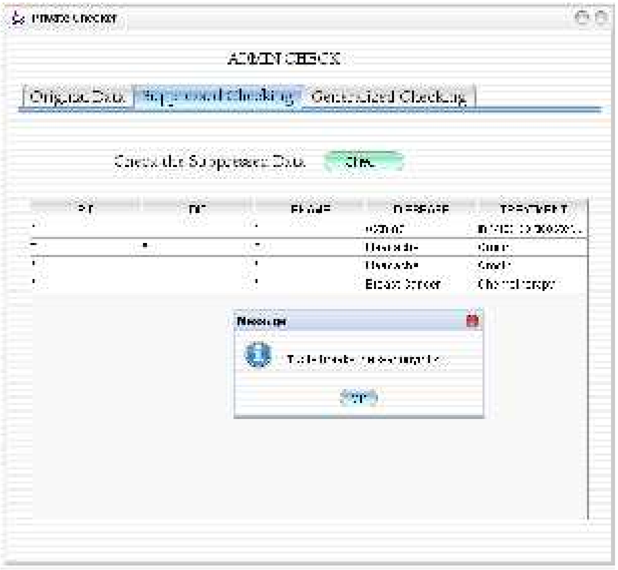


Fig.5 Private Checker Form Original Data View

Private Checker Form Original Data View provide only the administrator to acess original data of the database.An administrator have right to acess all views of the private checker such as original, suppressed and generalized . In the original view all contains are shown. In the suppresssed view key fields such as patientid, doctorid, patientname are suppressed with special symbol'*" so that the patient details are not displayed . In generalized view we get general view of the data so that privacy is not violated .Thus generalized and suppressed views help to maintain privacy. Private Checker form Original Data View is shown in Fig.5.



Private Checker

ADMIN CHECK

Original Data | Suppressed Checking | Generalized Checking

Check the Suppressed Data [Check]

| PID | DID | PNAME | DISEASE | TREATMENT |
|---|---|---|---|---|
| * | | * | [illegible] | [illegible] |
| * | * | * | Headache | [illegible] |
| * | | * | Headache | [illegible] |
| * | | * | Breast Cancer | Chemotherapy |

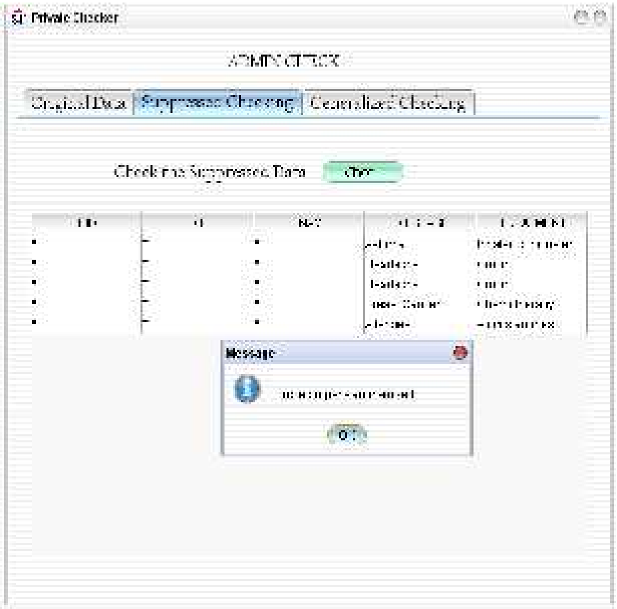
Message

[illegible]

[Ok]

Fig.6 Private Checker Form Suppressed Data View
With Tuple break k-anonymity

Private Checker Form Suppressed Data View allows user entered data to be checked in original table if it is valid then it replaces original values with suppressed values and database will be updated, otherwise if it is not valid then the user input values are rejected and database are not updated. We can make conclusion that if all values entered by data provider are correct then database will be updated successfully otherwise tuple will not be inserted in the database. Only administrator will be able to acess original data of the database. In Fig.6 User input values are breaking the k-anonymity so database are not updated.



Private Checker

ADMIN CHECK

Original Data | Suppressed Checking | Generalized Checking

Check the Suppressed Data [Check]

Message

[illegible]

[OK]

Fig.7 Private Checker Form Suppressed Data View
With Tuple properly k-anonymized

In Fig.7 User input values are properly k-anonymized so database is updated and we get suppressed values.

## IX. Conclusion

In this paper, we have proposed secure method 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 method. This is useful in medical application. If insertion of record satisfies the k-anonymity then such record is inserted in table and suppressed the sensitive information attribute by * to maintain the k-anonymity in database. Thus, by making such k-anonymity in table that makes unauthorized user too difficult to identify the record. In Generalization method we have shown that original values are replaced by more general values so that attacker cannot identify correct values. This is particularly applicable in military application or health care system. Beside we are dealing with the case of malicious parties by the introduction of non-colluding third party. The future work can be enhancing the redundancy of operations and also introducing new private updates to databases that support notions other than k-anonymity.

## 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.

## Reference

[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 Mining 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]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.
[14] E. Bertino and R. Sandhu, “Database Security—Concepts, Approaches and Challenges,” IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.