iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2017 **Issue:** onferende **Month of publication:** September 15, 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Hiding-Securing Restore to Databases Using Third Force

Mr. Anup R. Maurya[1], Ms. Amruta Mhatre[2]
[1]*Assistant Professor, Computer Engineering, Vishwatmak Om Gurudev College of Engineering, Thane [M.S.], India*
[2]*Assistant Professor, Computer Engineering, Atharva College of Engineering, Malad [M.S.], India*

Abstract: *Database has been widely studied and applied into many fields such as business development but* ***advanced the increased risk of privacy disclosure also so for that it is important to protect privacy of user.*** *In addition to extracting useful information and revealing patterns from large amounts of data, Database using third force also protects private and sensitive data from disclosure without the permission of data owners or The literature paper discusses various privacy preserving data mining algorithms and provide a wide analyses for the representative techniques for privacy preserving data base along with their merits and demerits.* ***Suppose there exists an unidentified database (e.g Hospital records) then the objectives include how to still retreat the data while updates are being made into the current unidentified database by secure the isolation of the user and hiding of the database. It is required to ensure that the database is still unidentified after the update. In this paper, we propose two methods solving this problem on oppression and inference based k-anonymous and hiding database but also a new protocol called Third Force which is directly connected with user and a database system.***
Index Terms: Data collection, database system, Hiding database security.

I. INTRODUCTION

Today there is an increased concerrn for privacy. The availability of huge numbers of databases recording a large variety of information about individuals makes it possible to discover information about specific individuals by simply correlating all the available databases. Although confidentiality and privacy are often used as synonyms, theyare different concern. Hence, the database confidentiality and privacy of user is a big concern. In this paper, a method is proposed by which it is possible to maintain the privacy of each and every individual and simultaneously devise a method to preserve the confidentiality. Privacy is the data that can be securely shown to the valid owner without leaking the sensitive information from the database. Data confidentiality is the difficulty experienced by the third party to know any sensitive information stored in the database. Privacy is an essential issue in case of transpose sensitive information from one location to another location through internet. This issue is arising in different areas such as census, medical, financial transactions, governmental organizations and industries etc. Confidentiality can be termed as the preservation of information against unauthorized disclosure and limiting data access to authorized users. Data confidentiality is the nondisclosure of certain information except to authorize person. So problem arises at this point where database needs to be updated. When tuple is to be inserted in the database problem occurs relating to privacy and confidentiality that is database owner decide that whether database preserve privacy without knowing what new tuple to be inserted. To carry out task of privacy, confidentiality to anonymous database, two approaches can be used. One is Suppression and the other is Generalization with a new protocol called TPA (Third Party Access).

II. LITERATURE REVIEW

A. Security-Control Methods for Statistical

1) Databases: A Comparative It 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. This suggested in the literature are classified into four general their performance with respect the identified evaluation criteria. A detailed omparative analysis of the most promising methods fo protecting dynamic-onlin statistical databases is also presented [1]. To date no single security-control method prevents both exact and partial disclosures. There are, however, a few perturbation-based methods that prevent exact disclosure and enable the database administrator to exercise ―statistical disclosure control.‖ Some of these methods, however introduce bias into query responses or suffer from the O/l query set-size problem (i.e., partial disclosure is possible in case of null query set or a query set of size 1).it recommend directing future research efforts toward developing new methods that prevent exact disclosure and provide statistical-disclosure control, while at the same time do not suffer from the bias problem and the O/l query-set-size problem. Furthermore, efforts directed toward developing a bias-correction mechanism and solving the general problem of small query-set-size would help salvage a few of the current perturbation based methods.

B. K-Anonymity: a model for protecting privacy

It was carried in 2002 by L.Sweeney**,** The solution provided in this includes a formal protection model named *k*-anonymity and a set of accompanying policies for deployment. A release provides *k*-anonymity protection if the information for each person contained in the release cannot be distinguished from atleast *k*-1 individuals whose information also appears in the release. This examines re-identification attacks that can be realized on releases that adhere to *a*nonymity unless accompanying policies are respected [2]. This paper has presented the K-anonymity protection model, explored related attacks and provides the way in which attacks can be thwarted.

C. Anonymizing Sequential Releases

It was carried in 2006 by K. Wang and B. Fung, An organization makes a new release as new information become available, releases a tailored view for each data request, and releases sensitive information and identifying information separately. The availability of related releases sharpens the identification of individuals by a global quasi-identifier consisting of attributes from related releases. Since it is not an option to anonymize previously released data, the current release must be anonymized to ensure that a global quasi identifier is not effective for identification.In this sequential anonymization problem is studied under Assumption that current release should be anonymized to ensure that a global quasi-identifier is not effective for identification. The issue here is how to anonymized current release so that it cannot link to previous releases yet it releases, and propose a scalable and practical solution [7].

D. Continuous Privacy Preserving Publishing of Data Streams

It was carried in 2008 by Y. Han, J. Pei, B. Jiang, Y. Tao, and Y. Jia,, In this study of an emerging problem of continuous privacy preserving publishing of data streams which cannot be solved by any straightforward extensions of the existing privacy preserving publishing methods on static data. To tackle the problem, method has developed a novel approach which considers both the distribution of the data entries to be published and the statistical distribution of the data stream. An extensive performance study using both real data sets and synthetic data sets verifies the effectiveness and the efficiency of our methods[8].

E. Public Key Encryption with keyword Search

It was carried in 2004 by D. Boneh, G. diCrescenzo, R. Ostrowsky, and G. Persiano**,** in this paper we study the problem of searching on data that is encrypted using a public key system. The paper defines define the concept of public key encryption with keyword search and give several constructions [9].

F.Anonymous Connections and Onion Routing

it was carried in 1998 by M. Reed, P. Syverson, and D. Goldschlag, this paper describe Onion routing, it provide infrastructure for providing private communication through public network and also provide anonymonous connection that provide strong resistant to eavesdropping and traffic analysis. This describes anonymous connections and their implementation using Onion routing [10].

G. K-anonymity

The k-anonymity model requires that within any equivalence class of the micro-data there are at least k records. The protection k-anonymity provides is simple and easy to understand. K-anonymity cannot provide a safeguard against attribute disclosure in all Homogeneity attack and the Background knowledge attack are identified when using K-anonymity.

H. L-Diversity

From the limitation of k-anonymity l-diversity can be introduced. L-diversity tries to put constraints on minimum number of distinct values seen within an equivalence class for any sensitive attribute.

An equivalence class has l-diversity if there is l or more well-represented values for the sensitive attribute.

A table is said to be l-diverse if each equivalence class of the table is l-diverse.

1) 1 Limitation of L-diversity: While the _-diversity principle represents an important step beyond k-anonymity in protecting against attribute disclosure, it has several shortcomings.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

a) ‘l - Diversity may be difficult to achieve and may not provide sufficient privacy protection.: Suppose that the original data have only one sensitive attribute: the test result for a particular virus. It takes two values: positive and negative. Further, suppose that there are 10,000 records, with 99 percent of them being negative, and only 1 percent being positive. Then, the two values have very different degrees of sensitivity. One would not mind being known to be tested negative, because then one is the same as 99 percent of the population, but one would not want to be known/considered to be tested positive. In this case, 2-diversity does not provide sufficient privacy protection for an equivalence class that contains only records that are negative. In order to have a distinct 2-diverse table, there can be at most 10;000 * 1% = 100 equivalence classes and the information loss would be large. Also, observe that because the entropy of the sensitive attribute in the overall table is very small, if one uses entropy _l-diversity, l must be set to a small value.

b) ‘l-diversity is insufficient to prevent attribute disclosure.

2) Attacks on l-diversity

a) *Skewness attack:* When the overall distribution is skewed, satisfying that _-diversity does not prevent attribute disclosure. Suppose that one equivalence class has an equal number of positive records and negative records. It satisfies distinct 2-diversity, entropy 2-diversity, and any recursive (c, 2)-diversity requirement that can be imposed. However, this presents a serious privacy risk, because anyone in the class would be considered to have 50 percent possibility of being positive, as compared with the 1 percent of the overall population. Now, consider an equivalence class that has 49 positive records and only 1 negative record. It would be distinct 2- diverse and has higher entropy than the overall table (and thus, satisfies any Entropy _l-diversity that one can impose), even though anyone in the equivalence class would be considered 98 percent positive, rather than 1 percent. In fact, this equivalence class has exactly the same diversity as a class that has 1 positive and 49 negative record, even though the two classes present very different levels of privacy risks.

III. DRAWBACK

The existing methods need to perform manual pre-processing, i.e., generation of a domain generalization taxonomy to define the hierarchy of the categorical attribute values involving prior knowledge about the domain. The domain tree should be prepared separately for every domain. Moreover, there might be disagreements between domain experts about the correct structure of the taxonomy tree, which may lead to differences in the results.

IV. PROPOSED SYSTEM

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 oppression 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 Inference 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 replaced by the general value then this record will replaced by the general value and these general values being inserted into table.

V. METHODOLOGY

A. There are three methods will be used:-i) Oppression-based unidentified database

In oppression 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. It allows the owner of DB to properly unidentified the tuple t, without gaining any useful knowledge on its contents and without having to send to t‘s owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. In order to

perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.
In particular, when using a suppression-based anonymization method, we mask with the special value *, the value deployed by person for the anonymization.
Where sensitive information and all information that allows the inference of sensitive information are simply not released.

B. Inference-based unidentified database

The second protocol is aimed at inference-based unidentified databases, In Inference 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 replaced by the general value then this record will replaced by the general value and these general values being inserted into table and it relies on a secure set intersection protocol, to support privacy-preserving updates on a Inference-based k-anonymous DB.
When using a inference-based unidentification method, original values are replaced by more general ones, according to a priori established value inference hierarchies (VGHs).
It will rely on a secure set intersection protocol, to support privacy-preserving updates on a k-anonymous database.

C. Third Force

Third Force has a communication between the user and the database. The user directly contact to the Third Force for information there is no need for the user to check the encrypted tuples again and again. Same for the loader to unidentified the tuple for crypto.
A channel is a bidirectional communication connection between a program on third protocol and a participant.

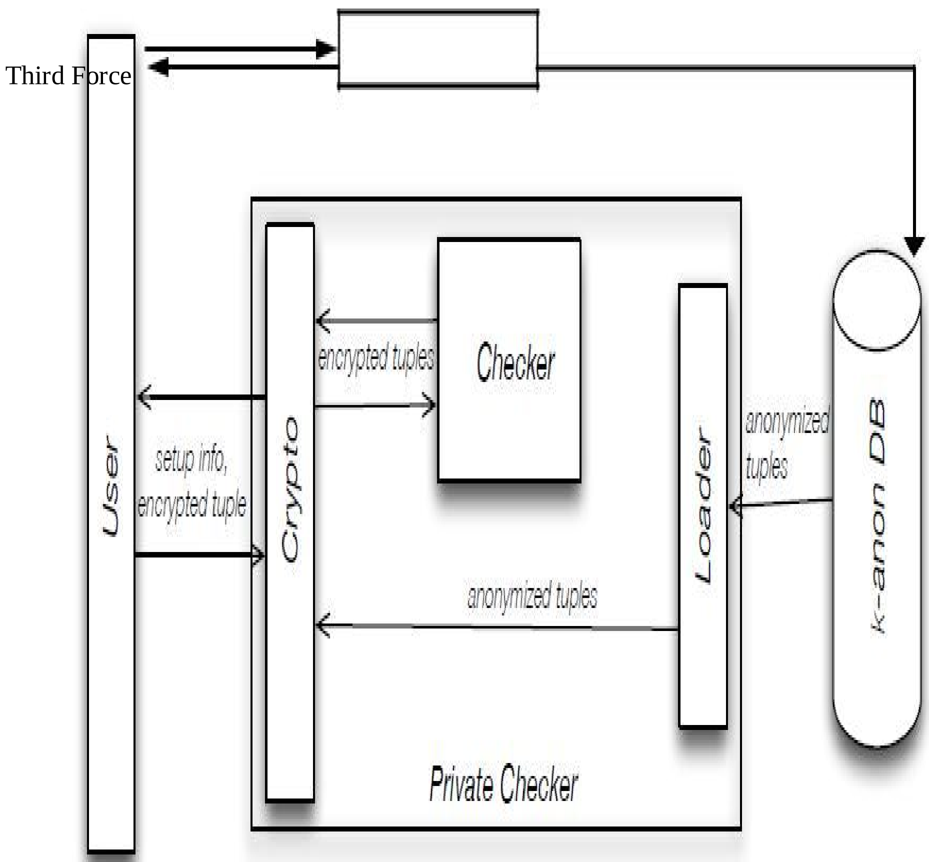

Fig 1 Proposed System Architecture

VI. CONCLUSION & FUTURE WORK

The generalization and suppression methods help to preserve data for confidential databases and also help to maintain data privacy. This method used to verify that if new record is being inserted to the database, it does not affect anonymity of database using AES technique.The objective of devising a private update technique other than k-anonymity could be obtained by using the concept of Third Force. The concept of anonymisation ensures that only authorized users can view the sensitive information and to other users the database appears in the unidentified form.

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 the 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:

A. Implement database for invalid entries.

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

C. Implement real world database system

REFERENCES

[1] Luo Yongcheng, Le Jiajin and Wang Jian ―Survey of Anonymity Techniques for Privacy Preserving‖, Proc. of CSIT 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009), Volume1, IACSIT Press, Singapore, pp. 248-252.

[2] E. Bertino, R. Sandhu ―Database Security - Concepts, Approaches, AndChallenges‖, Dependable and Secure Computing, IEEE Transactions, Volume 2, Issue 1, ISSN: 1545-5971, pp. 2-19, Jan-March 2005

[3] S.Vijayarani, Dr.A.Tamilarasi, N.Murugesh ―Comparative Analysis ofMasking Techniques in Privacy Preserving Data Mining‖, International Journal of Computer Science & Applications (IJCSA), ISSN: 0974-0767, Issue 2, pp.51-55, July 2012

[4] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati―Microdata Protection‖, Springer US, Advances in Information Security (2007)

[5] Charu C. Aggarwal, S.Y Philip ―Privacy-Preserving Data Mining: Models and Algorithms‖, Advances in Database Systems, Volume 34, 2008

[6] Nikita Patel, Saurabh Upadhyay, ―Study of Various Decision Tree PruningMethods with their Empirica Comparison in WEKA‖, International Journal of Computer Applications (0975 – 8887) Volume 60, No.12, December 2012

[7] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann 1993.

[8] J. Daemen and V. Rijmen, ―AES Proposal: Rijndael, AES algorithmsubmission‖, available: http://www.nist.gov/CryptoToolkit, September 3, 1999.

[9] L. Sweeney, ―k-Anonymity: A Model for Protecting Privacy,‖ Int‘l J.bUncertainty, Fuzziness and Knowledge Based Systems, vol. 10, no. 5, pp. 557-570, 2002

[10] A.Machanavajjhala, J.Gehrke, et al., ℓ-diversity: Privacy beyond k-anonymity, In Proc. of ICDE, Apr.2006.B

[11] N. Li, T. Li, and S. Venkatasubramanian, t-Closeness: Privacy Beyond k-anonymity and l-Diversity, In Proc. Of ICDE, 2007, pp. 106-115.

[12] P.Samarati, ―Protecting Respondent‘s Privacy in Microdata Release‖, IEEE Trans.Knowledge and Data Eng.,vol.13,no.6,pp. 1010-1027, Nov./Dec. 2001. W. and Marchionini, G. 1997

[13] Xiaokui Xiao, Yufei Tao ―Personalized Privacy Preservation‖, SIGMO 2006, June 27–29, 2006, Chicago, Illinois, USA. Copyright 2006 ACM 1595932569/ 06/0006

[14] Alberto Trombetta ,Wei Jaing,Elisa Bertino and Lorenzo Bossi, ―Privacy Preserving Updates to anonymous and Confidential database‖ IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 8, NO. 4, JULY/AUGUST 2011.