ISSN:0975-9646

Efficient Relational Data Management Using Access Control and Privacy-Preserving Mechanism

CH. Srinivasa Reddy[1] V.vasanthi[2] Ravuri Daniel[3] K.V.N. Rajesh[4]

Department of Information Technology, Vignan's Institute of Information Technology, Visakhapatnam, India[1, 2, 3,4]

Abstract-In an organization or agency the authorized users misuse the sensitive information. Sensitive information is data that must be protected from unauthorized access. Data confidentiality ,Availability, and Integrity are more necessary part of Relational data base security. Access control mechanism protect sensitive information but there is a lack of privacy-preserving when authorized users misuse the sensitive information. We can use both Access control and privacy-preserving mechanisms. The Access control policies define selection that implies to roles available in the system. The privacy-preserving can be achieved through anonymization like generalization and suppression. We introduce heuristics for anonymization algorithms. and also satisfy the privacy requirements e.g. k-anonymity and L-diversity. In addition to that we need to satisfy imprecision bounds for more permission and has lower total imprecision. We have proved that the total imprecision is low by using the anonymization algorithms which are Top-Down Heuristic 1(TDH 1) and Top-Down Heuristic 3(TDH3).TDH3 total imprecision is low when compared to TDH1. Finally total imprecision is low by which the privacy can be achieved

Keywords: Access control, privacy, k-anonymity, query evaluation.

I INTRODUCTION

Sensitive information refers to confidential or proprietary information that only definite people are allowed to see and that is therefore not available to everyone. If sensitive information is lost or used in any way other than planned, the result can be severe damage to the people or organization to which that belongs. The concept of privacy-preservation for sensitive data can require the enforcement of privacy policies or the protection against individuality disclosure by satisfying some privacy requirements. consider privacy-preservation from the anonymity aspect. The sensitive information, even after the deletion of identify attributes, is still vulnerable to linking attacks by the authorized users. Reduce the imprecision aggregate for all queries. The imprecision added to each permission/query in the anonymized data is not known. Not filling accuracy constraints for individual permissions in a policy/workload. The heuristics proposed for Efficient relational data management using Access control and privacy-preserving mechanisms. The structure is a combination of access control and privacy protection mechanisms. The access control policies define selection that implies to roles available in the system**.** The privacy preserving module anonymizes the data to convene privacy requirements and imprecision constraints on predicates set by the access control mechanism.

The remaining part of paper organized as follows: the literature review of the proposed work described in Section-II. Section-III explains the proposed architecture and methodology of anonymization algorithm. The analysis of the results are briefly described in Section-IV. The conclusions are concluded in Section-V.

II LITERATURE REVIEW

E. Bertino and R. Sandhu, [5] wrote "Database Security-Concepts, Approaches, and Challenges," it defines As organizations increase their reliance on, possibly shared, information systems for daily business, they become more accessible to security breaches even as they increase productivity and efficiency advantages. Though there are lots of techniques, such as encryption and electroni c signatures, are directly available to protect data when transmitted across sites, a truly comprehensive way for data protection must also include mechanisms for enforcing access control policies based on data contents, characteristics, subject qualifications and other relevant contextual information, such as time. It is well implies today that the semantics of data must be taken into account in order to specify effective access control policies. Also, the approach for data integrity and availability specifically tailored to database systems must be accepted. In this regard, over the years the database security community has developed a number of different techniques and approaches to persuade data confidentiality, integrity, and availability.

P. Samarati, [12] wrote "Protecting Respondents' Identities in Micro data Release," it defines Today's globally networked society places high demand on the dissemination and sharing of information. While in the previously released information was mostly in tabular and statistical form, many situations call today for the release of specific data (micro data). In order to protect the anonymity of the entities (called respondents)to which information refers, data holders to often erase or encrypt explicit identifiers such as names, addresses, and phone numbers.

B. Fung, K. Wang, R. Chen, and P. Yu [3] wrote "Privacy-Preserving Data Publishing: A Survey of Recent Developments," it defines The set of collected digital information by individuals, corporations, and governments has created tremendous opportunities for information-based

knowledge and decision making, compelled by mutual benefits, or by regulations that require certain data to be published, there is an appeal for the exchange and publication of data among different parties. Data in its original form, however, typically contains delicate information about individuals, and publishing such data will violate individual privacy.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam [1] wrote "L-Diversity: Privacy Beyond k-anonymity," it defines Publishing data about individuals without exposing the sensitive information about them is an important problem. In recent years the current definition of privacy called k-anonymity has gained popularity. In a k-anonymized dataset, each record is identical from at least k − 1 other records with respect to certain "identifying" attributes .

K .Le Fevre, D. DeWitt, and R. Ramakrishna [9] wrote "Workload-Aware Anonymization Techniques for Large-Scale Datasets," it defines Taking care of an individual's privacy is a serious problem in micro data distribution and publishing .Anonymization algorithms typically desired to satisfy certain privacy definitions with minimal impact the quality of the resulting data. While much of the previous literature has consistent quality through simple one-size-fits-all measures, we argue that quality is tough to judge with respect to the workload for which the data will ultimately be used. This article furnishes a set of anonymization algorithms that incorporate a target class of workloads, consisting of one or more data mining tasks likewise selection predicates.

Based on reference papers Access control mechanism protect sensitive information but there is a lack of privacy-preserving when authorized users misuse the sensitive information. We can use both Access control and privacy-preserving mechanisms. The Access control policies define selection that implies to roles available in the system**.** The privacy-preserving can be achieved through anonymization like generalization and suppression. We introduce heuristics for anonymization algorithms. and also satisfy the privacy requirements e.g. k-anonymity and L-diversity.

III PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY:

ARCHITECTURE

In the Fig.1 proposed architecture that the Architecture of access control and privacy preserving mechanisms. In access control mechanism Access control policies define selection that implies to roles available in the system. The privacy preserving mechanism contains privacy protection module and privacy requirements. The privacy protection module contains the generalization and suppression mechanism for generating anonymization . the privacy protection module take the sensitive information for generating the anonymization information.

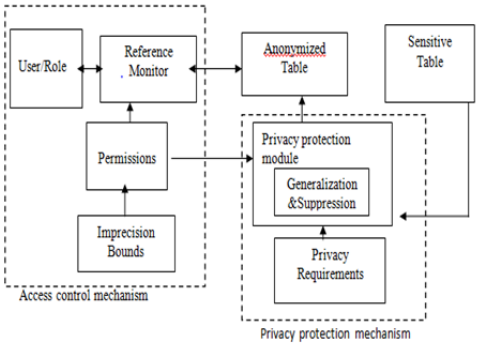

Fig.1 Architecture of Access control and privacy preserving.

Methodology

Top Down Selection Mondrian(TDSM) , the partitions are split beside the median. Consider a partition that overlaps a query. If the median as well falls inside the query then even after splitting the partition, the imprecision for that query will not change as both the new partitions still overlap the query . In this heuristic, we suggest to split the partition along the query cut and then choose the dimension beside which the imprecision is minimum for all queries.

Algorithm-1:Top-DownHeuristic (TDH1)

Input: T, K, Query and BQj.
Output: Partition (P).
Step 1: Initialize set of candidate partitions.
Step 2: for (CPi ϵ CP) do
Step 3: Search the set of queries that overlap candidate partitions.
Step 4: Place the queries in increasing order of BQj
Step 5: while (feasible cut is not found) do
Step 6: Choose query from QO.
Step 7: Create query cuts according dimension.
Step 8: Select dimension and cut having lower in imprecision.
Step 9: Check feasible cut found or not if feasible cut then
Step 10:. Create new partitions and add to CP
Step 11: otherwise
Step 12: Split candidate partitions recursively along median upto anonymity requirement is satisfied .
Step 13: Minimize new partitions and add to P.
Step 14: return (P)

Algorithm-2:Top-DownHeuristic3(TDH3)

Input: T; K; Query and BQj.
Output: Query set and Result set.
Step 1: Initialize candidate partitions.
Step 2: for each cpi in cp {
Search the queries which gets the overlap
Results and
insert it into Query set.
choose the Query from Query set with small
Result set.
Generate query cut with each dimension.
Select Imprecision for all queries into query set.
If (feasible cut found)
if(check with user secured attributes)
insert it into CP
Else
do recursively until anonymity requirement is satisfly.
Add into Result set.

Flow of System

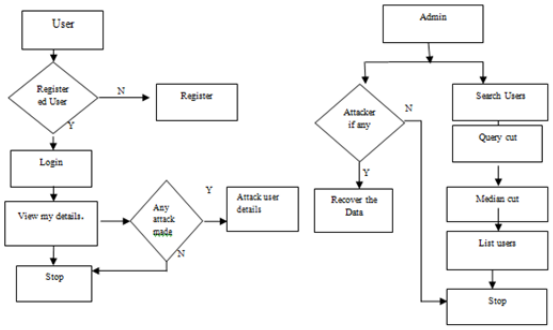

Fig : 2 Flow Diagram for User and Admin Modules

In user module there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like attack user details, view my details and logout. If user clicks on my details button, then the server will give response to the user with their tags such as user ID, name, mobile no, address, pin code and email ID. The user want attack the particular user information, then click on attack user details button, then enter user name to attack and submit. The server will display the user details, and then you can edit the user information, submit and server will give response to user. After modifying a data, the user will be considered as an attacker. The attacker details will be stored in an attacker module. In Admin module the Admin has to login by using valid user name and password. After login successful he can do some operations such as search users, query cut, median cut, list users, view attackers, data recovery and logout.

V RESULTS AND DISCUSSION

Consider one data set for the empirical evaluation of the proposed heuristics. The data set is the Medical data set defacto benchmark for k-anonymity research. The attributes in the Medical data set are: User Id, Name, Password, Blood Group, Email Id, Mobile no., Location, Date of Birth, age, Address, Gender, Disease name, Pin code.

UID	Username	Blood Group	Disease Name	E-Mail	Mobile	Location	DOB	Age	Address	Gender	Pincode
24	mallum	B+	fine	mallum.t@gmail.com	8125996634	Nellore	15/06/1992	24	kotapetnellore	Female	560002
25	swetha	AB+	Fever	swetha.r@yahoo.com	8265478652	Vizag	3/4/1992	24	GajuwakaVizag	Female	560003
26								0			0
27	ravindra	A+	Diarrhea	ravindra4002@gmail.com	8500143556	Nellore	14/7/1992	24	Nellore	Male	560003
28	ch.vasantha	AB+	fine	vasanthi21@gmail.com	8125996634	hyderabad	01/06/1992	24	HYDERABAD	Female	560005
1	rakshath	A+	Diarrhea	rakshath.st@gmail.com	9738352279	Bangalore	02/08/1990	13	Mandapeta, Tanikur	Male	560043
2	rudresh	A+	Fever	rudresh@gmail.com	9535866270	Bangalore	02/08/1990	24	Tanikur	Male	560016
3	xyz	A+	Fever	xyz@gmail.com	9535866270	Bangalore	02/08/1990	19	Mandapeta, Tanikur	Male	560040
4	santhosh	A+	fine	santhosh@gmail.com	9535866270	Bangalore	02/08/1990	18	Mandapeta, Tanikur	Male	560040
5	aaa	A+	fine	rakshath.st@gmail.com	9738352279	Bangalore	02/08/1990	42	Mandapeta, Tanikur	Male	560040
6	bbb	A+	Fever	bbb@gmail.com	9535866270	Bangalore	02/08/1990	46	Mandapeta, Tanikur	Male	560040
7	ccc	A+	Diarrhea	ccc@gmail.com	9535866270	Bangalore	02/08/1990	62	Mandapeta, Tanikur	Male	560041
8	ddd	A+	Fever	ddd@gmail.com	9535866270	Bangalore	02/08/1990	64	Mandapeta, Tanikur	Male	560040

Fig: 3 List of Users Data in Admin

Median cut

In this module, the admin can search the diseases based on the age and blood group, then server will mine the all data and send the related data to particular user.

Query Cut

A query cut is define as the splitting of a partition beside the query interval values. For a query cut using Query Qi, both the begin of the query interval (aQij) and the ending of the query interval (b Qij) are considered to split a partition along the jth dimension. Example the comparison of median cut and query cut is given in Fig.6.2 The rectangle with thick lines represents Query Q1. While, the rectangles with scattered lines represent partitions. In Fig. a the tuples are partitioned according to the median cut and even after dividing the tuple space into four partitions there is no decrease in imprecision for the Query Q1. However, for query cuts in Fig. b the imprecision is compact to zero as partitions are either non-overlapping or fully with this inside the query region.

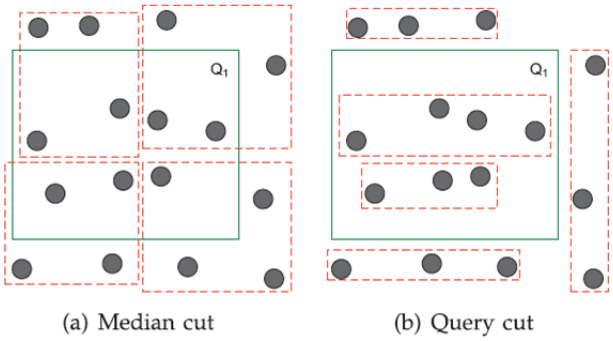

Fig : 5 Comparison of median and query cut

Query Imprecision Bound:

It is the total imprecision suitable for a query predicate and is preset by the access control administrator. The query imprecision bound, denote by BQi , is the total imprecision acceptable for a query predicate Qi and is preset by the access control administrator.

.

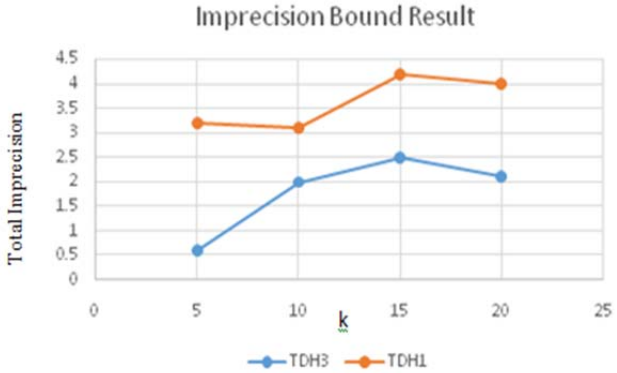

Fig: 4 Imprecision Bound Result

The imprecision bound result is shown in the above figure. The x-axis contain the k-values and y-axis contain the total imprecision values. Anonymization algorithms are TDH1 and TDH3 which will be considered for the imprecision values . TDH3 will give the lower imprecision value when compared to TDH1.

CONCLUSION

The proposed system discussed about the Access control and Privacy –preserving mechanisms. The Access control policies defines selection that implies to roles available in the system**.** The privacy-preserving can be achieved through anonymization like generalization and suppression. We introduced heuristics for anonymization algorithms. and also satisfy the privacy requirements e.g. k-anonymity and L-diversity. We have proved that the total imprecision is low by using the anonymization algorithms which are Top-Down Heuristic 1(TDH 1) and Top-Down Heuristic 3(TDH3).TDH3 total imprecision is low when compared to TDH1. Finally total imprecision is low by which the privacy can be achieved.

FUTURE WORK

For future work, we plan to extend the proposed privacy-preserving access control to incremental data and cell level access control.

REFERENCES

[1] A. Machanavajjhala, D. Kifer, J. Gehrke, and M Venkitasubramaniam, "L-Diversity: Privacy Beyond k-anonymity," ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article 3, 2007.

[2] A.Rask, D.Rubin, and B.Neumann,"Implementing Row-and Cell-Level Securitya in Classified Databases Using SQL Server 2005," MS SQL Server Technical Center 2005.

[3] B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys, vol. 42, no. 4, article 14, 2010.

[4] D. Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn, and R. Chandramouli, Proposed NIST Standard for Role-Based Access Control," ACM Trans. Information and System Security, vol. 4, no. 3, pp. 224- 274, 2001.

[5] E. Bertino and R. Sandhu, "Database Security- Concepts, Approaches, an Challenges," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[6] J. Buehler, A. Sonricker, M. Paladini, P. Soper, and F. Mostashari, "Syndromic Surveillance Practice in the United States: Findings from a Survey of State,Territorial, and Selected Local Health Departments,"Advances in Disease Surveillance, vol. 6, no. 3, pp. 1- 20, 2008.

[7] J.Friedman,J.Bentley,and R. Finkel,"Analgorithm for Finding Best Matches in Logarithmic Expected Time," ACM Trans. Mathematical Software, vol. 3, no. 3, pp. 209- 226, 1977.

[8] K.Browder and M. Davidson, "The Virtual Private Database in oracle 9ir2,"Oracle TechnicalWhite Paper, vol. 500, 2002

[9] K.LeFevre, D. DeWitt, and R. Ramakrishnan, "Workload-Aware Anonymization Techniques for Large-Scale Datasets,"ACM Trans. Database Systems, vol. 33, no.3, pp. 1-47, 2008.

[10] K. LeFevre, D. DeWitt, and R. Ramakrishnan, 22nd Int'l Conf. Data Eng., pp. 25- 25, 2006.

[11] K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, and D. DeWitt, " Limiting Disclosure in Hippocratic Databases," Proc. 30th Int'l Conf. Very Large Data Bases, pp. 108-119, 2004.

[12] P. Samarati, "Protecting Respondents' Identites Micro data Release," IEEE Trans. Knwledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov. 2001.

[13] S.Chaudhuri, T. Dutta, and S. Sudarshan, "Fine Grained Authorization through Predicated Grants," Proc. IEEE 23rd Int'l Conf. Data Eng.,

[14] S.Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy, "Extending Query Rewriting Techniques for Fine-Grained Access Control," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 551-562, 2004.

[15] T.Iwuchukwu and J. Naughton, "K- Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc. 33rd Int'l Conf. Very Large Data Bases, pp. 746-7-57, 2007