Efficient Access Control Mechanism For Relational Database

K.Pavithra, E.Elakiya, M.Thenmozhi.
Department of Computer Science and Engineering,
Panimalar Engineering College, Chennai.

Guide Name- Mrs.C.Jackulin M.E, chin.jackulin@gmail.com

Abstract—**Access control mechanisms preserve the sensitive datas from unauthenticated users. However, when sensitive datas is given and a Privacy Protection Mechanism (PPM) is not in place, an authorised user can still surrender the privacy of a person leading to unique revelation. A PPM can use suppression and generalization of relational data to anonymize and satisfy privacy necessity, e.g., k-anonymity and l-diversity, against unique and attribute revelation. However, privacy is achieved at the cost of precision of authorized data. In this paper, we propose an efficient access control method. The privacy preserving methods define selection items available to roles while the privacy necessity is to satisfy the k-anonymity or l-diversity. An additional thing that needs to be satisfied by the PPM is the imprecision bound for each selection item. The techniques for workload-aware anonymization for particular selection items have been discussed in the papers. The problem of satisfying the accuracy constraints for multiple roles has not been studied before. We propose method for anonymization algorithms and show experimentation that the proposed approach satisfies imprecision bounds for more authorisation.**

Index Terms—Access control, privacy, k-anonymity, query evaluation.

I. INTRODUCTION

COMPANIES gather and resolve user information to develope their quality services. Access Control Mechanisms (ACM) is used to protect the sensitive information from the unauthorized users. However, sensitive information can still be used improperly by correct users to compromise the privacy of users. The theme of privacy protection for important data can require the compulsion of privacy rules. In this paper, we inquire privacy preservation from the anonymity view. The important information, even after the deleting of sensitive attributes, is still infectible to linking attacks by the correct users [2]. This difficulty has been studied widely in the area of micro data publishing [3] and privacy definitions, e.g., k-anonymity [2], l-diversity [4], and variance diversity [5]. Anonymization algorithms use suppression and generalization methods to satisfy privacy necessity with minimal problem of micro data. The anonymity techniques can be used with an ACM to ensure both guarantee and protection of the important data. The protection is done at the cost of exactness and imprecision is introduced in the authorized data under an access control rule. We use the method of imprecision bound for each prohibition to define a threshold on the amount of imprecision that can be tolerated. Existing workload aware anonymization techniques [5], [6] reduce the imprecision aggregate for all questions and the imprecision added to each question in the anonymized micro data is not known. Developing the protection needs more stringent returns in extra imprecision for questions. However, the solution of fulfill accuracy constraints for individual prohibition in a workload has not been known before. The methods proposed in this paper for efficient access control mechanism are also related in the background of workload-aware anonymization. The anonymization for related data publishing has been studied in other papers [3]. In this paper the aim is on a fixed relational table that is anonymized one time only. To show our approach, role-based access control is considered. However, the concept of accuracy constraints for prohibition can be addressed to any security rules, e.g., discretionary access control. The impact of this paper is as follows. First, we develope the guaranteed and privacy restrictions as the difficulties of k-anonymous Partitioning with Imprecision Bounds (k-PIB) and give hardness outcome. Second, we provide introduction the theme of efficient access control mechanism for relational database. Third, we propose methods to the solution of the k-PIB problem.

II. EXISTING SYSTEM

COMPANIES gather and resolve user information to develope their quality services. Access Control Mechanisms (ACM) are used to protect the sensitive information from the unauthorized users. However, sensitive information can still be used improperly by correct users to compromise the privacy of users. Sensitive information can still be used improperly by correct users to compromise the privacy of users. This difficulty has been studied widely in the area of micro data publishing and privacy definitions. Disadvantage occurs in this method that is important datas can still be abused by authorized users to surrender the privacy of users.

III. PROPOSED SYSTEM

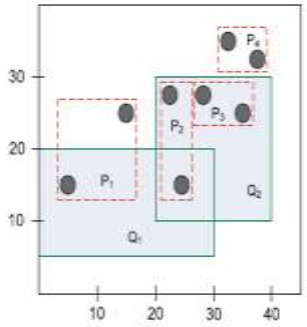
In this paper, we inquire privacy and protection from the anonymity view. The important data, even after the deleting of sensitive variable, is still infectible to linking attacks by the correct users. This problem has been studied widely in the area of micro data publishing and privacy definitions, e.g., k-anonymity, l-diversity, and variance diversity .Suppression and generalization methods are used by anonymization algorithm to satisfy privacy necessity with low mistakes of micro data. The ACM permits only authorized predicates on important data. We give tough outcome for the k-PIB problem and present method for dividing the data to the satisfy the privacy variable.

IV. METHODOLOGY DESCRIPTION

A. ANONYMIZATION WITH IMPRECISION BOUNDS

In this section, we develope the difficulty of k-anonymous dividing with Imprecision Bounds and present an Efficient access control mechanism. Query Imprecision Bound- BQi means Query imprecision bound, is the total accuracy capable for a query statement Qi and is preset by the access control administrator.

Examples Two types of queries are given in below diag. Solid lines with shaded rectangles are queries and dashed lines with rectangles are regions.



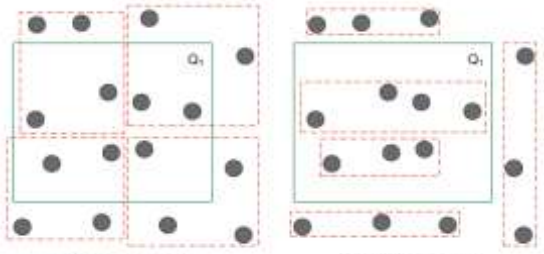
The accuracy bounds for Queries Q2 and Q1 are preset to 0 and 2. In above diagram the partitioning does not satisfy the accuracy bounds. In next diagram the dividing satisfies the bounds for Queries Q2 and Q1 as the imprecision for Q2 and Q1 is 2 and 0, respectively.

Query Imprecision Slack-sQi means Query accuracy slack for a Query, say Qi, is explain the difference between the actual query imprecision and the query imprecision bound.

$$s_{Q_i} = \begin{cases} B_{Q_i} - imp_{Q_i}, & \text{if } imp_{Q_i} \le B_{Q_i}, \\ 0, & \text{otherwise.} \end{cases}$$

The TDSM algorithm to divide a partition By using the median value along a dimension. In the proposed methods in Section 4, query distance is used to divide the partitions that are defined as query cuts.

Query Cut- A query cut means split the partition along the query interval value. For a query cut using Qi means query, both (ajQi) is the start of the query interval and the (bjQi) eis the end of the query interval are considered to split a partition along the jth dimension.
Example .Median cut and query cut comparisions is given in next figure for 3-anonymity.Query Q1 is represent as rectangle with solid lines. While, the partitions are represent as rectangles with dotted lines. In Fig. 4a the variables are partitioned according to the median cut and even after splitting the variable space into four partitions there is no reduction in accuracy for the Query Q1. However, for query cuts in Fig. 4b the imprecision is reduced to zero as partitions are either non-overlapping or fully enclosed inside the query region.



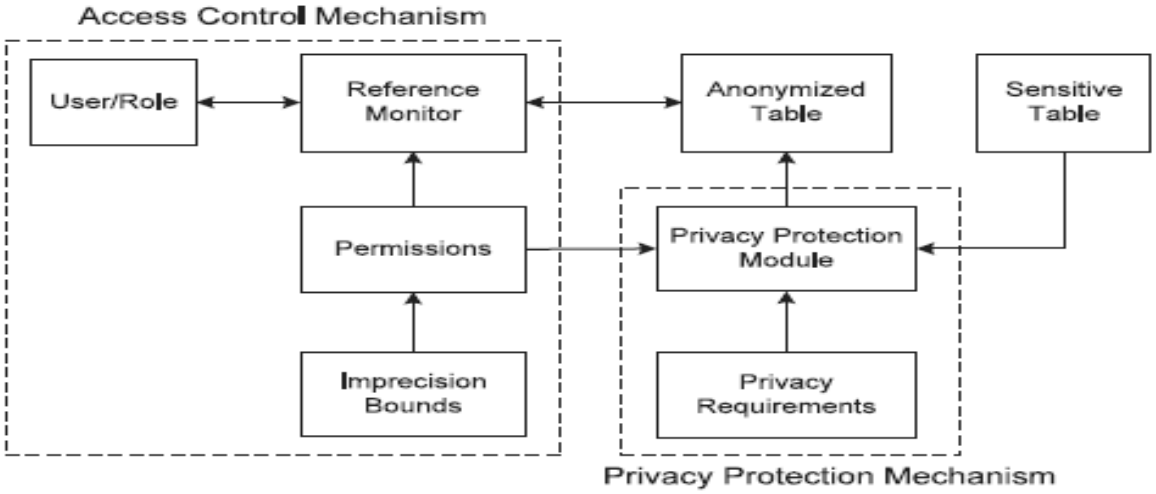
B. The k-PIB PROBLEM

We show that finding k-anonymous dividing that break accuracy bounds for small number of queries is also NP hard. A multiset of variables is changed into an equivalent set of seperate (tuple; count) pairs. The elements of Query Qi are the sum of count values of values occurring inside the query hyper-rectangle. An upper bound for the number of queries defined by constant qv that can violate the bounds. The choice version of the k-PIB problem is as follows:

Decisional k-anonymity with Imprecision Bounds.

Theorem 3.1. Decisional k-anonymity with Imprecision Bounds is NP-complete. Proof. Refer to Appendix, which can be found on the Computer Society Digital Library.

C. Efficient Access Control Mechanism



An Efficient access control mechanism, illustrated in Fig. 5 (arrows represent the Flow of direction), is developed. The PPM assures that the privacy and accuracy aims are met before the important information is available to the

ACM. The authorization in the access control rules are based on selection variables on the QI attributes. The policy administrator defines the authorizations along with the accuracy bound for each permission, user-to-role assignments, and role-to permission
assignments [18]. The requirements of the accuracy bound assure that the allowed data has the desired level of accuracy. The accuracy bound data is not shared with the consumers because knowing the accuracy bound can result in breaking the privacy requirement. The PPM is want to meet the privacy requirement along with the accuracy bound for each permission.

Lemma 4.2.Query imprecision was denoted by Non-negative random variable IQj.Then, the expected accuracy for a query Qj is

$$E(I_{Q_j}) \leq \left[\left(\prod_{i=1}^{d} \left\lfloor \frac{l_i^{Q_j} + l_i^{P_e}}{l_i^{P_e}} \right\rfloor \right) * |P_e| \right] - |Q_j|.$$

Theorem 4.3

Query imprecision was denoted by Non-negative random variable IQj. An independent Poisson trial is denoted by X1; . . .;Xn, where Xi is a random variable that is equal to 1 if a query, say Qi, breaks the accuracy bound BQi otherwise is equal to 0.

$$E[X] = \sum_{i=1}^{n} p_i \leq \sum_{i=1}^{n} \frac{E(I_{Q_i})}{(B_{Q_i} + 1)}.$$

V. HEURISTICS FOR PARTITIONING

A. Top-Down Heuristic (TDH)

In TDSM, the partitions are split along the median. Consider a partition that overlaps a query. If the median also 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 as illustrated in Fig. 4. In this heuristic, we propose to split the partition along the query cut and then choose the dimension along which the imprecision is minimum for all queries. If multiple queries overlap a partition, then the query to be used for the cut
needs to be selected.

The queries having imprecision greater than zero for the partition are sorted based on the imprecision bound and the query with minimum imprecision bound is selected.

The algorithm for TDH is listed in Algorithm 2. There are two differences compared to TDH1. First, the kd-tree traversal for the for loop in Lines 2-14 is preorder. Second, in Line 14, the query bounds are updated as the partitions are being added to the output (P). The time complexity of TDH2 is $O(d|Q|^2n^2)$, which is the same as that of TDH1. In Section 4.3, we propose changes to TDH2 that reduce the time complexity at the cost of increased query imprecision.

Algorithm 2: TDH2

Input : T, k, Q, and B_{Q_j}
Output: P

Initialize Set of Candidate Partitions($CP \leftarrow T$)
for ($CP_i \in CP$) **do**
    // Depth-first(preorder) traversal
    Find the set of queries QO that overlap CP_i such that $ic_{CP_i}^{QO_j} > 0$
    Sort queries QO in increasing order of B_{Q_j}
    while *(feasible cut is not found)* **do**
        Select query from QO
        Create query cuts in each dimension
        Select dimension and cut having least overall imprecision for all queries in Q
    if *(Feasible cut found)* **then**
        Create new partitions and add to CP
    else
        Split CP_i recursively along median till anonymity requirement is satisfied
        Compact new partitions and add to P
        Update B_{Q_j} according to $ic_{P_i}^{Q_j}$, $\forall Q_j \in Q$
return (P)

B. REPARTITIONING

Repartitioning step is equivalent to partitioning all the leaf nodes that in the worst case can take O (|Q|n) time for each candidate query set.

Algorithm 4: Repartitioning

Input : T, k, Q, P, and B_q
Output: P

Initialize SQ, CQ, and CP
Add $q \in Q$ satisfying bound to SQ
Add $q \in Q$ violating bound by 10% to Candidate Query set(CQ)
Add all sibling leaf node pairs having $\sum_{q \in CQ}(ic_{P_i}^{q_j} + ic_{P_{i+1}}^{q_j}) > 0$ to Candidate Partition(CP)
for ($CP_i \in CP$) **do**
    Merge the first pair CP_i and CP_{i+1}
    Select q from CQ with the least imprecision greater than the imprecision bound
    Create the candidate cuts in each dimension
    Select the cut and the dimension satisfying all $q \in SQ$ with the minimum imprecision $\forall q \in CQ$
    if *(feasible cut found)* **then**
        Remove CP_i and CP_{i+1} from CP and P
        Add new partitions to P
        for ($q \in CQ$) **do**
            if ($Imp_q < B_q$) **then**
                Remove q from CQ and add to SQ
return (P)

EXPECTED OUTPUT:

1) The original data of patients:

Name	DOB	age	Phone no	Disease	Place
Pavithra	07-06-1994	21	908776589	Flu	Theni
Karthik	08-01-1990	20	979969435	Cancer	Madurai
Raj	09-12-2010	5	954866535	Cancer	Chennai
Devi	04-11-1999	15	835676788	Flu	Coimbatore

In above table Important informations which was needed by doctors are age, disease and palce.So in our project we extract only those information from the database. In this table name, DOB, age and place are sensitive informations.So in our project we hide this information from the users.

2) Extract important information:

DOB	Age	Disease
07-06-1994	21	Flu
08-01-1990	20	Cancer
09-12-2010	5	Cancer
04-11-1999	15	Flu

3) Hide the sensitive information:

DOB	Age	Disease
07-06-19**	*0	Flu
08-01-19**	*0	Cancer
09-12-20**	*5	Cancer
04-11-19**	1*	Flu

4) Generalize the information for privacy:

Age	Disease
15-20	Flu
5-20	Cancer

FINAL OUPUT: If Authorized user search about Flu means the OUTPUT was:

Age	Disease
15-20	Flu

So Our Proposed system only show above details. So privacy is there and extract exact information from big database.

VI. CONCLUSION

An Efficient Access control mechanism for relational database has been proposed. The framework is a combination of ACM and PPM. The ACM allows only authorized query variables on important data. The PPM anonymizes the information to meet privacy needs and accuracy values on value set by the ACM. We develope this situation as the problem of k-anonymous dividing with accuracy Bounds (k-PIB). We provide hardness outcome for the k-PIB problem and present methods for dividing the information to the satisfy the privacy values and the accuracy bounds. In the current work, fixed access control and relational data model has been assumed. For future work, we plan to extend the proposed privacy preserving access control to more data and cell level access control.

REFERENCES

[1] 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.
[2] P. Samarati, "Protecting Respondents' Identities in Microdata Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov. 2001.

[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] 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.
[5] 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.
[6] 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. 746-757, 2007.
[7] 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.
[8] K. Browder and M. Davidson, "The Virtual Private Database in oracle9ir2," Oracle TechnicalWhite Paper, vol. 500, 2002.
[9] A. Rask, D. Rubin, and B. Neumann, "Implementing Row-and Cell-Level Security in Classified Databases Using SQL Server 2005," MS SQL Server Technical Center, 2005.
[10] 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.
[11] S. Chaudhuri, T. Dutta, and S. Sudarshan, "Fine Grained Authorization through Predicated Grants," Proc. IEEE 23rd Int'l Conf. Data Eng., pp. 1174-1183, 2007.
[12] 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.
[13] 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.
[14] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," Proc. 22nd Int'l Conf. Data Eng., pp. 25- 25, 2006.