ISSN (Online) : 2454-4159

International Journal of Advanced Research in Science Management and Technology

IJARSMT

Volume 2, Issue 6, June 2016

# Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data

**Amol Phoke [1], Avinash Mahamune [2], Sunil Ambhore [3]**

UG Student, Dept. Of Computer Engg, Sinhgad Institute of Technology, Lonavala, Maharashtra, India[1,2,3]

***ABSTRACT***— Access control mechanisms protect sensitive information from unauthorized users. However, when sensitive information is shared and a Privacy Protection Mechanism (PPM) is not in place, an authorized user can still compromise the privacy of a person leading to identity disclosure. The technique of k-anonymization has been proposed in the literature as an alternative way to release public information, while ensuring both data privacy and data integrity. We prove that two general versions of optimal k-anonymization of relations are NP-hard, including the suppression version which amounts to choosing a minimum number of entries to delete from the relation.. The techniques for workload-aware anonymization for selection predicates have been discussed in the literature. we propose heuristics for anonymization algorithms and show empirically that the proposed approach satisfies imprecision bounds for more permissions and has lower total imprecision than the current state of the art.

***KEYWORDS*** – Access Control, K-annonymity, L-diversity, Query Imprecision, privacy, Data Confidentiality.

## I. INTRODUCTION

ORGANIZATIONS collect and analyse consumer data to improve their services. Access Control Mechanisms (ACM) are used to ensure that only authorized information is available to users. In this paper, we investigate privacy-preservation from the anonymity aspect. The sensitive information, even after the removal of identifying attributes, is still susceptible to linking attacks by the authorized users.

This problem has been studied extensively in the area of micro data publishing and privacy definitions, e.g., k-anonymity, l-diversity, and variance diversity. We use the concept of imprecision bound for each permission to define a threshold on the amount of imprecision that can be tolerated. To exemplify our approach, role-based access control is assumed. However, the concept of accuracy constraints for permissions can be applied to any privacy-preserving security policy, e.g., discretionary access control.

Organizations may release private data for the purposes of facilitating useful data analysis and research, for example, patients' medical records may be released by a clinic to aid a medical study. K-anonymity has been proposed as a means to preserving privacy in data releases.

Put simply, the private data set is modified so that each record is indistinguishable from at least k −1 other records. Indistinguishability is defined in terms of any set of attributes that can be used to uniquely identify an individual. This set of attributes has been called a quasi-identifier in the literature.

An example of a quasi-identifier is the set of attributes comprising Age, Sex and Zip code.

## II. BACKGROUND THEORY

In this section, role-based access control and privacy definitions based on anonymity are over-viewed. Query evaluation semantics, imprecision, and the Selection Mondrian algorithm are briefly explained. Given a relation T

={A1,A2,….,An}; . . .;Ang, where Ai is an attribute, T* is the anonymized version of the relation T. We assume that T is a static relational table. The attributes can be of the following types:

_ Identifier. Attributes, e.g., name and social security that can uniquely identify an individual. These attributes are completely removed from the anonymized relation.

_ Quasi-identifier (QI). Attributes, e.g., gender, zip code, birth date that can potentially identify an individual based on other information available to an adversary. QI attributes are generalized to satisfy the anonymity requirements.

_ Sensitive attribute. Attributes, e.g., disease or salary, that if associated to a unique individual will cause a privacy breach.

### 2.1 Access Control for Relational Data

Fine-grained access control for relational data allows to define tuple-level permissions, e.g., Oracle VPD and SQL . For evaluating user queries, most approaches assume a Truman model . In this model, a user query is modified by the access control mechanism and only the authorized tuples are returned. Column level access control allows queries to execute on the authorized column of the relational data only. Cell level access control for relational data is implemented by replacing the unauthorized cell values by NULL values.

Role-based Access Control (RBAC) allows defining permissions:

(U), a set of Roles (R), and a set of Permissions (P). For the relational RBAC model, we assume that the selection predicates on the QI attributes define a permission. a hyper-rectangle in the tuple space and all the tuples enclosed by this hyper-rectangle are authorized to the role assigned to the permission. In practice, when a user assigned to a role executes a query, the tuples satisfying the conjunction of the query predicate and the permission are returned.

### 2.2 Anonymity Definitions

In this section, privacy definitions related to anonymity are introduced.

Definition 1 (Equivalence Class (EC)). An equivalence class is a set of tuples having the same QI attribute values.

Definition 2 (k-anonymity Property). A table T_ satisfies the k-anonymity property if each equivalence class has k or more tuples.

| | Qi1 | Qi2 | S1 |
|---|---|---|---|
| ID | Age | Zip | Disease |
| 1 | 5 | 15 | Flu |
| 2 | 15 | 25 | Fever |
| 3 | 28 | 28 | Diarrhea |
| 4 | 25 | 15 | Fever |
| 5 | 22 | 28 | Flu |
| 6 | 32 | 35 | Fever |
| 7 | 38 | 32 | Flu |
| 8 | 35 | 25 | Diarrhea |

**Table 1. Sensitive Table**

|  | Qi1 | Qi2 | S1 |
|---|---|---|---|
| ID | Age | Zip | Disease |
| 1 | 5 | 15 | Flu |
| 2 | 15 | 25 | Fever |
| 3 | 28 | 28 | Diarrhea |
| 4 | 25 | 15 | Fever |
| 5 | 22 | 28 | Flu |
| 6 | 32 | 35 | Fever |
| 7 | 38 | 32 | Flu |
| 8 | 35 | 25 | Diarrhea |

**Table 2. 2-anonymous table**

Query Imprecision is defined as the difference between the number of tuples returned

by a query evaluated on an anonymized relation T_ and the number of tuples for the same query on the original relation T.

The imprecision for query Qi is denoted by $imp_{Qi}$ ,

$$imp_{Qi} = |Q_i (T^*)| - |Q_i (T)|, \text{ where}$$

$$|Q_i (T^*)| = \sum_{\text{EC overlaps Qi}} |EC|$$

ECj: (1)

The query Qi is evaluated over T_ by including all the tuples in the equivalence classes that overlap the query region.

Intuitively, this metric assigns to each tuple *t* in *V* a penalty, which is determined by the size of the equivalence class containing *t*.

As an alternative, we also propose the *normalized average equivalence class size metric*:

$C_{AVG} =($total_records/ total_equivalence_classes$)$

Let IQi be a non-negative random variable that denotes the query imprecision. Let X1; . . .;Xn be an independent Poisson trial, where Xi is a random variable that is equal to 1 if a query, say Qi, violates the imprecision bound BQi otherwise is equal to 0.

For $X=\sum^{n}_{i=1} x_i$ and $B_{Qi}>0$. We have

$$E|X| = \sum_{i=1} P_i < \sum_{i=1} \frac{E(I_{Qi})}{(B_{Qi}+1)}$$

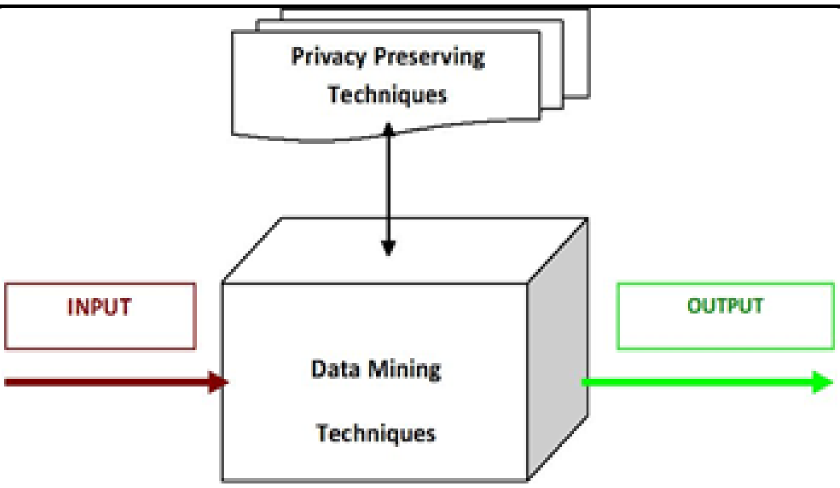

Fig. 1 Data Flow in privacy preserving

**Query Cut**:-A query cut is defined as the splitting of a partition along the query interval values. For a query cut using Query Qi, both the start of the query interval (aj Qi ) and the end of the query interval (bjQi) are considered to split a partition along the jth dimension.

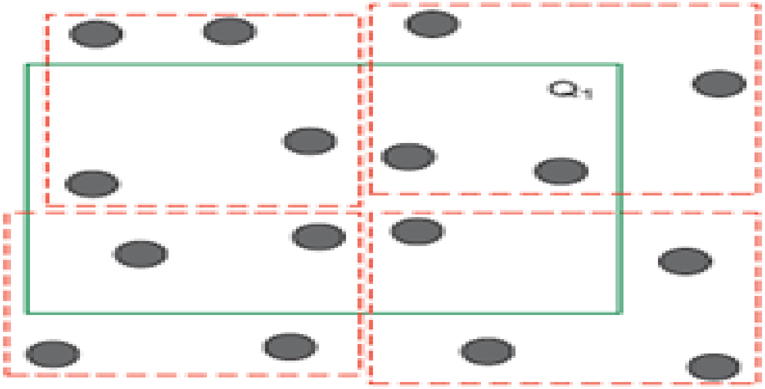

Fig. 2 Median Cut

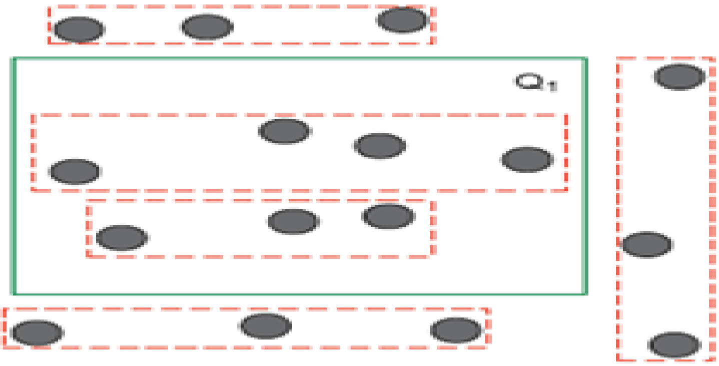

Fig. 3 Query Cut

The median cut generates a balanced tree with height lgn and the work done at each level is djQjn. The partitions created by TDSM have dimensions along the median of the parent partition.

## III. ANONYMIZATION WITH IMPRECISION BOUNDS

In this section, we formulate the problem of k anonymous Partitioning with Imprecision Bounds and present an accuracy-constrained privacy-preserving access control framework.

## IV. ARCHITECTURE

An accuracy-constrained privacy-preserving access control mechanism, illustrated in Fig.  (arrows represent the direction of information flow), is proposed. The privacy protection mechanism ensures that the privacy and accuracy goals are met before the sensitive data is available to the access control mechanism. The permissions in the access control policy are based on selection predicates on the QI attributes. The policy administrator defines the permissions along with the imprecision bound for each permission/query, user-to-role assignments, and role-to permission assignments. The specification of the imprecision bound ensures that the authorized data has the desired level of accuracy. The imprecision bound information is not shared with the users.

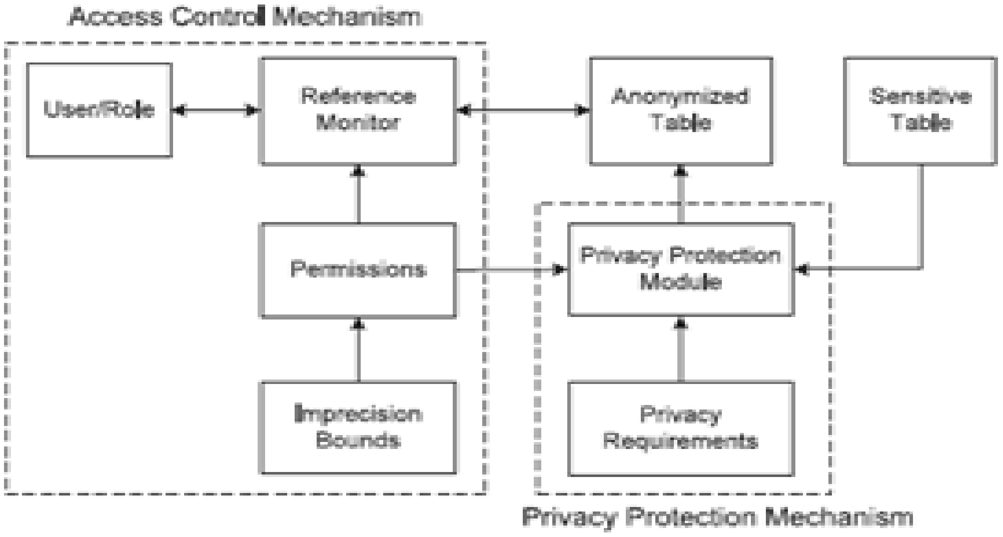

**Fig. 4 Architecture**

Access control mechanisms for databases allow queries only on the authorized part of the database. Predicate based fine-grained access control has further been proposed, where user authorization is limited to pre-defined predicates. Enforcement of access control and privacy policies have been studied in . However, studying the interaction between the access control mechanisms and the privacy protection mechanisms has been missing. Recently, Chaudhuri et al. have studied access control with privacy Mechanisms.

## V. CONCLUSION

In this paper, Enforcing k-anonymity means ensuring that the information recipient will not be able, even when linking information to external data, to associate each released tuple with less than k individuals. We formulate this interaction as the problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB).. For future work, we plan to extend the proposed privacy-preserving access control to incremental data and cell level access control.

## REFERENCES

**[1] R. Anderson, ªA Security Policy Model for Clinical Information Systems,º Proc. IEEE Symp. Security and Privacy, pp. 30-43, May 1996.**

[2] A. Davey and H.A. Priestley, Introduction to Lattices and Order. Cambridge Univ. Press, 1990.

[3] L. Sweeney, ªGuaranteeing Anonymity when Sharing Medical Data, the Datafly System,º Proc. J. Am. Medical Informatics Assoc., Washington, DC.: Hanley & Belfus, Inc., 1997.

[4] B. Woodward, ªThe Computer-Based Patient Record Confidentiality, The New England J. Medicine, vol. 333, no. 21, pp. 1419-1422, 1995.

[5] 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.

[6] R. Sandhu and Q. Munawer, “The Arbac99 Model for Administration of Roles,” Proc. 15th Ann. Computer Security Applications Conf., pp. 229-238, 1999.

[7] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi, “Extending Relational Database Systems to Automatically Enforce Privacy Poliies,” Proc. 21st Int’l Conf. Data Eng., pp. 1013-1022, 2005.

[8] C. Dwork, “Differential Privacy,” Proc. 33rd Int’l Colloquium Automata, Languages and Programming, pp. 1-12, 2006