ACCURACY–ACCESS CONTROL MECHANISM FOR RELATIONAL DATABASE

R.Saikrishna 1 and Mrs.J.Aruna Shanthi 2

1Scholar, 2Assistant Professor, Department of IT Mahatma Gandhi Institute of Technology, Hyderabad, India.

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. A PPM can use suppression and generalization of relational data to anonymize and satisfy privacy requirements, e.g., k-anonymity and l-diversity, against identity and attribute disclosure. However, privacy is achieved at the cost of precision of authorized information. In this paper, we propose accuracy–constrained privacy-preserving access control framework.

Keyword – Access Control, Privacy Protection Mechanism, Privacy, Security.

INTRODUCTION
Organizations collect and analyze consumer data to improve their services. Access Control Mechanisms (ACM) are used to ensure that only authorized information is available to users. However, sensitive information can still be misused by authorized users to compromise the privacy of consumers. The concept of privacy-preservation for sensitive data can require the enforcement of privacy policies or the protection against identity disclosure by satisfying some privacy requirements. 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. Anonymization algorithms use suppression and generalization of records to satisfy privacy requirements with minimal distortion of micro data. The anonymity techniques can be used with an access control mechanism to ensure both security and privacy of the sensitive information. The privacy is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an access control policy.

Every organization keeps a set of databases to store their information and there may be several situations to share those information with others. As we are living in the information age there are large sources of data around us. To improve the services the organizations collect and analyze the data. The Confidentiality, Integrity and Availability are termed as the [CIA-triad] designed to enable the information security within the organization. They are considered to be the essential components of the security. To ensure that only the authorized information are available only to the authorized users and access control mechanism is
implemented in the databases. However, there may happen the misuse of sensitive information by the authorized users to compromise the privacy of the customers. For the enhancement of the protection against the identity disclosure and enforcing the privacy policies, the concept of privacy preservation of sensitive data introduced by satisfying some privacy requirements.

**BACKGROUND**

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 $T = \{A_1; A_2; \ldots A_n\}$ where $A_i$ 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.

To define tuple-level permissions fine-grained access control like Oracle VPD and SQL are introduced in relational databases. Truman model is introduced for evaluating user queries. A user query is modified by the access control mechanism and only authorized tuples are returned in this model. Column level access control allows queries to execute on the authorized column of the In relational data column level access control mechanisms allow queries to execute on authorized column, by replacing the unauthorized cell values by NULL values cell level access control for relational data is achieved. For defining permissions on objects based on roles in an organization a Role-based Access Control (RBAC) was introduced. An RBAC policy configuration includes a set of Users (U), a set of Roles (R), and a set of Permissions (P). We assume that the selection predicates on the QI attributes define a permission for the relational RBAC model. UA is a user-to-role (U $\rightarrow$ R) assignment relation and PA is a role to-permission (R $\rightarrow$ P) assignment relation.

**Privacy definitions**: Here, 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.

**PREDICATE EVALUATION AND IMPRECISION**

A In this section the query predicate evaluation semantics have been discussed. For query predicate evaluation over a table, says $T$, a tuple is included in the result if all the attribute values satisfy the query predicate. Here, we only consider conjunctive queries (The disjunctive queries can be expressed as a union of conjunctive queries), where each query can be expressed as a d-dimensional hyper-rectangle. The semantics for query evaluation on an
anonymized table $T^*$ needs to be defined. When the equivalence class partition (Each equivalence class can be represented as a d-dimensional hyper-rectangle) is fully enclosed inside the query region, all tuples in the equivalence class are part of the query result. Uncertainty in query evaluation arises when a partition overlaps the query region but is not fully enclosed. In this case, there can be many possible semantics. We discuss the following three choices

1. Uniform:- Assuming the uniform distribution of tuples in the overlapping partitions, include tuples from all partitions according to the ratio of overlap between the query and the partition. Query evaluation under this option might under-count or over-count the query result depending upon the original distribution of tuples in the partition region. Most of the literature uses this uniform distribution semantics to compare anonymity techniques over selection tasks. However, the choice of the sensitive attribute value for the selected tuples from an overlapping partition is not defined under uniform semantics. For access control, a tuple’s QI attribute values along with the sensitive attribute value need to be returned.

2. Overlap:- Include all tuples in all partitions that overlap the query region. This option will add false positives to the original query result.

3. Enclosed:- Discard all tuples in all partitions that partially overlap the query region. This option will have false negatives with respect to the original query result.

ANONYMIZATION WITH IMPRECISION BOUNDS

In this section, we develop the difficulty of $k-$anonymous dividing with imprecision Bounds and present an Efficient access control mechanism. Query imprecision Bound – $B_{Qi}$ 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 diagram. Solid lines with shaded rectangles are queries and dashed lines with rectangles are regions.

![Figure: 4. Anonymization satisfying imprecision bounds.](image)

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 - $s_{Qi}$ means Query accuracy slack for a Query, say $Qi$, is explain the difference between the actual query imprecision and the query imprecision bound.
\[ s_{Qi} = \begin{cases} B_{Qi} - imp_{Qi}, & \text{if } imp_{Qi} \leq B_{Qi}, \\ 0, & \text{otherwise}. \end{cases} \]

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 \((bjQi)\) is the end of the query interval are considered to split a partition along the \( j \)th dimension.

Example. Median cut and query cut comparisons 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.

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 separate \((\text{tuple}; \text{count})\) 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.

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 \( Q1 \) attributes.
The policy administrator defines the authorizations along with the accuracy bound for each permission, user-to-role assignments, and role-to-permission assignments. 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 $I_{Qj}$. Then, the expected accuracy for a query $Q_j$ is:

$$E(I_{Qj}) \leq \left( \prod_{i=1}^{d} \left[ i_i^{2j} + 1 \right] \right) \cdot |P_i| - |Q_j|.$$

Theorem 4.3: Query imprecision was denoted by Non-negative random variable $I_{Qj}$. An independent Poisson trial is denoted by $X_1; \ldots; X_n$, where $X_i$ is a random variable that is equal to 1 if a query, say $Q_i$, breaks the accuracy bound $B_{Q_i}$ otherwise is equal to 0

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

**CONCLUSION**

An Efficient Access control mechanism for relational database has been proposed. The framework is a combination of ACM and PPM. The AC Mallows 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 develop this situation as the problem of k-anonymous dividing with accuracy bounds (k-PIB). We provide hardness out come 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


Z. Wu, Lu Jiang, Q. Zheng, J.Liu, “Learning , to surface Deep Web content”. In proceedings of 24th AAAI conference on Artificial Intelligence, AAAI-10
R. SaiKrishna, pursuing Master of Technology in Software Engineering in Department of IT at Mahatma Gandhi Institute of Technology, Hyderabad. He has received B Tech degree in IT from Jawaharlal Nehru University, Hyderabad.

J. Aruna Shanthi working as Assistant Professor in the department of Information Technology at MGIT, Hyderabad. She received the B.E. degree in Information technology from the Nagarjuna University, Guntur, India and M.Tech degree in Computer science engineering from the JNTU, Hyderabad. She has 10 years of experience in teaching. She has published various papers in National and International conferences.

**Corresponding Address:**

R. SaiKrishna,
Information Technology (IT) Department
Mahatma Gandhi Institute of Technology,
Kokapet (Vill), Rajendranagar (Mdl)
Chaitanya Bharati Post, Gandipet
HYDERABAD - 500 075
PH: 9000848359.