An Tree Based Accuracy Constrained Privacy Preserving Access Control Mechanism for Relational Data

BANDI PRADEEP RAJU¹, M. KAMALA KUMARI²

¹PG Scholar, Dept of CSE, Adikavi Nannya University, AP, India, E-mail: pradeep7teja@gmail.com.
²Assistant Professor, Dept of CSE, Adikavi Nannya University, AP, India, E-mail: kmarepalli@yahoo.com.

Abstract: To shield susceptible information from unauthorized users we are put forward an Access control mechanisms. The proposed system is an integrated framework of achieving both privacy and security is proposed though the integration of Access Control method to make much efficient, reliable data with Privacy Preservation Technique to prevent the authorized user from exploitation the sensitive information with some accuracy loss of data due to Access Control Mechanisms (ACM) is used to protect that only authorized information is available to users. The privacy is achieved at the cost of accuracy and imprecision is introduced in access control policy the authorized information is much secured. The present project focuses on an accuracy-constrained privacy-maintaining to control access framework. While satisfying the privacy requirement, k-anonymity or l-diversity, to roll the access controls policies and selection process. An additional constraint that needs to be satisfied by PPM is the imprecision bound for each selection predicate. The proposed system applies the application specific anonymization. Top Down Selection Mondrian (TDSM) algorithm is used for query workload-based anonymization. The access control mechanism allows only authorized query predicates on sensitive data. The privacy freeze module anonymizes the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism. The purpose of the present project is to propose heuristics for anonymization algorithms and to show the availability of the proposed approach for analytically satisfying the imprecision bounds for more permission.

Keywords: k-Anonymity, l-Diversity, Imprecision Bounds, Access Control.

I. INTRODUCTION

The two of the databases Data mining and knowledge discovery are new research areas that explore, the extraction of previously unknown patterns from huge amounts of data. Recent search in data collection, data distribution and related technologies have initiated a new era of research where existing data mining algorithms should be reconsidered from a different point of view, this of privacy preservation. In data mining and statistical databases privacy preserving data mining is a novel research direction, where data mining algorithms are analyzed for the side-effects they suffer in data privacy. Twofold is the main deliberation in privacy preserving data mining. First, sensitive raw data like identifiers, names, addresses and the like should be altered from the original database, in order for the recipient of the data not to be able to arrangement another person’s privacy. Second, sensitive knowledge which can be mined from a database by using data mining algorithms should also be preclude, because such knowledge can equally well compromise data privacy. The main objective is to develop privacy preserving data mining algorithms for modifying the original data in some way, so that the private data and private intelligence remain private even after the mining process, so authorized information only get to that user. The problem that happen when confidential information can be derived from released data by unauthorized users is also frequently called the “database inference” problem. AS organizations escalation their approval of database systems as the key data management technology for day-to-day process and decision making, the security of data handled by these systems becomes typical. Damage and misuse of data affect not only a specific user or application, but may have disastrous reaction on the entire organization.

The recent rapid propagation of Web based applications and information systems have further became every impressive the risk exposure of databases and, thus, data stability is today more crucial than ever. It is also essential to acknowledge that data needs to be protected not only from external threats, but also from insider threats, the proposed system uses the concept of deception bound for each acceptance to define a threshold on the amount of imprecision that can be tolerated. Existing work aware Anonymization techniques. In this proposed system the focus is on a static relational data compressed that we are going to make anonymized only once. To exemplify the proposed 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. We use the conception of inexactness sure for every permission to define a threshold on the quantity of inexactness which will be tolerated. Existing workload- aware anonymization techniques minimize the inexactness mixture for all queries and also the inexactness else to every permission/query within the anonymized small information isn't glorious.
Creating the privacy requirement a lot of demanding (e.g., increasing the worth of k or l) leads to further inexactness for queries. However, the matter of satisfying accuracy constraints for respective approval in an exceedingly policy/workload has not been studied before. The heuristics projected during this paper for accuracy-constrained privacy-preserving access management also are relevant within the context of workload-aware anonymization. The anonymization for continuous information publishing has been studied in literature [3]. during this paper the main focus is on a static relative table that's anonymized just one occasion.

II. RELATED WORK


A. Problem Statement

Organizations collect and analyze shopper information to enhance their services. Access control Mechanisms (ACM) square measure accustomed make sure that solely approved data is obtainable to users. However, sensitive data will still be used by approved users to compromise the privacy of customers. The idea of privacy-preservation for sensitive information will need the social control of privacy policies or the protection against identity revelation by satisfying some privacy needs. The access control mechanism allows only authorized query predicates on sensitive data. The privacy preserving module anonymizes the data to meet data.

B. System Model

The level of anonymity is based on application specific anonymization (Degree of privacy protection module). The sensitive table and privacy requirement will check the degree level of application from the privacy protection module to be anonymized from the anonymized table. The reference monitor will gets the permission from the privacy protection module with the reference of imprecision bound level and give the exact result to the user. By selecting the level of anonymity based on application, we can solve the k-PIB (k-anonymous partitioning with imprecision bound) problem and we can gain the large amount of information from the microdata. The county epidemiologist will not lose their information as shown in Fig.1. The existing methods focus on a universal approach that exerts the same amount of preservation for all persons, without catering for their concrete needs. The consequence is that the system may be offering insufficient protection to a subset of people, while applying excessive privacy control to another subset. Contributed technique performs the minimum generalization for satisfying everybody’s requirements, and thus, retains the largest amount of information from the microdata. The core of the solution is the concept of application specific anonymity, i.e., a Administrator can specify the degree of privacy protection for her/his sensitive values as shown in Fig.2.

C. Techniques

K- anonymity: k-anonymity is a property consume by certain anonymized data. In this situation of k-anonymization problems, a database is a table with n rows and m columns. Each row of the table produce a record present to a specific member of a people and the entries in the various rows need not be particular. The values in the discrete columns are the values of attributes correlate with the members of the population.

Fig 3. Before k-anonymity.
An Tree Based Accuracy Constrained Privacy Preserving Access Control Mechanism for Relational Data

There are 6 attributes and 10 records in this data. There are two common methods for achieving k-anonymity for some value of k.

- **Suppression:** In this method, certain values of the attributes are replaced by an asterisk '*'. All or some values of a column may be replaced by '*' . In the anonymized table below, we have replaced the values in the 'Name' attribute and all the values in the 'Religion' attribute have been replaced by a '*'.

- **Generalization:** In this method, individual values of attributes are replaced by a broader category. For example, the value '19' of the attribute 'Age' may be replaced by 'l ≤ 20', the value '23' by '20 < Age ≤ 30', etc.

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Gender</th>
<th>State of domicile</th>
<th>Religion</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>20 ≤ Age ≤ 30</td>
<td>Female</td>
<td>Tamil Nadu</td>
<td>*</td>
<td>Viral infection</td>
</tr>
<tr>
<td>*</td>
<td>20 ≤ Age ≤ 30</td>
<td>Female</td>
<td>Kerala</td>
<td>*</td>
<td>TB</td>
</tr>
<tr>
<td>*</td>
<td>20 ≤ Age ≤ 30</td>
<td>Male</td>
<td>Karnataka</td>
<td>*</td>
<td>No illness</td>
</tr>
<tr>
<td>*</td>
<td>20 ≤ Age ≤ 30</td>
<td>Female</td>
<td>Kerala</td>
<td>*</td>
<td>Heart-related</td>
</tr>
<tr>
<td>*</td>
<td>Age ≤ 20</td>
<td>Male</td>
<td>Karnataka</td>
<td>*</td>
<td>Cancer</td>
</tr>
<tr>
<td>*</td>
<td>Age ≤ 20</td>
<td>Male</td>
<td>Karnataka</td>
<td>*</td>
<td>Heart-related</td>
</tr>
<tr>
<td>*</td>
<td>Age ≤ 20</td>
<td>Male</td>
<td>Karnataka</td>
<td>*</td>
<td>Viral infection</td>
</tr>
</tbody>
</table>

Fig 4. After k-anonymity.

This data has 2-anonymity with respect to the attributes 'Age', 'Gender' and 'State of domicile' since for any combination of these attributes found in any row of the table there are always at least 2 rows with those exact attributes. The attributes available to an adversary are called "quasi-identifiers". Each "quasi-identifier" tuple occurs in at least k records for a dataset with k-anonymity.

### III. ATTACKS OF K-ANONYMITY

While k-anonymity is a promising approach to take for group based anonymization given its simplicity and wide array of algorithms that perform it, it is however susceptible to many attacks. Such attacks include,

- **Homogeneity Attack:** This attack leverages the case where all the values for a sensitive value within a set of k records are identical. In such cases, even though the data has been k-anonymized, the sensitive value for the set of k records may be exactly predicted.

- **Background Knowledge Attack:** This attack leverages an association between one or more quasi-identifier attributes with the sensitive attribute to reduce the set of possible values for the sensitive attribute. For example, Machanavajjhala, Kifer, Gehrke, and Venkitasubramaniam (2007) showed that knowing that heart attacks occur at a reduced rate in Japanese patients could be used to narrow the range of values for a sensitive attribute of a patient's disease.

- **l-Diversity:** l-diversity is a form of group based anonymization that is used to preserve privacy in data sets by reducing the granularity of a data representation. This reduction is a trade off that results in some loss of effectiveness of data management or mining algorithms in order to gain some privacy.

- **The l-diversity Principle:** An equivalence class is said to have l-diversity if there are at least l “well-represented” values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the table has l-diversity.

- **Distinct l-diversity:** The simplest definition ensures that at least l distinct values for the sensitive field in each equivalence class.

- **Entropy l-diversity:** The most complex definition defines entropy of an equivalent class E to be the negation of summation of s across the domain of the sensitive attribute of p(E,s)log(p(E,s)) where p(E,s) is the fraction of records in E that have the sensitive value s. A table has entropy l-diversity when for every equivalent class E, Entropy(E) ≥ log(l).

- **Recursive (c-l) diversity:** A compromise definition that ensures the most common value does not appear too often while less common values are ensured to not appear too infrequently.

### A. Predicate Evaluation and Imprecision

In this module, the question predicate analysis linguistics are mentioned. For question predicate analysis over a table, say T, a tuple is enclosed within the result if all the attribute values satisfy the question predicate. Here, planned system solely considers conjunctive queries (The adversative queries will be expressed as a union of conjunctive queries), wherever every question will be expressed as a d-dimensional hyper-rectangle.

### B. Anonymization With Imprecision Bounds

For anonymization the proposed system selects the, quasi identifier 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. And selects the corresponding selected attribute vector values, and set the upper bounds and the lower bound for the corresponding vector values for the selected attribute.

### C. Accuracy-Constrained Privacy-Preserving Mechanism

The exact tuple values in a very relation are replaced by the generalized values when the anonymization. during this case, access management social control over the generalized knowledge must be outlined.

- **Relaxed:** Use overlap linguistics to permit access to all or any partitions that are overlapping the permission.

- **Strict:** Use fenced in linguistics to permit access to solely those partitions that are absolutely fenced in by the permission.

### IV. PROPOSED SYSTEM

International Journal of Advanced Technology and Innovative Research
Volume. 08, IssueNo.07, July-2016, Pages: 1308-1311
A. Heuristics for Partitioning

Starting with the whole tuple space the nodes in the kd-tree are resourceful divided till the partition size is between k and 2k. The leaf nodes of the kd-tree are the output segregation that are mapped to equivalence classes in the given table. In the partitions are split along the median. Consider a segregation that overlaps a query. If the median also falls inside the query then even after splitting the partition, the exaggeration for that query will not change as both the new segregation still overlap the query. In this heuristic, the proposed system proposes to split the partition along the query cut and then choose the dimension along which the Exaggeration is minimum for all queries. The intuition behind this decision is that the queries with smaller bounds have lower tolerance for error and such a partition split ensures the decrease in imprecision for the query with the smallest imprecision bound. If no feasible cut satisfying the privacy requirement is found, then the next query in the sorted list is used to check for partition split. If none of the queries allow partition split, then that partition is split along the median and the resulting partitions are added to the output after compaction.

B. Improving the Number of Queries Satisfying the Imprecision Bound

In module, the query imprecision slack is defined as the difference between the query bound and query imprecision. This query deception slack can help satisfy queries that violate the bounds by only a small surplus by increasing the imprecision of the queries having more slack. The margin by which queries violate the bounds. In this repartitioning step, it considers only the first two groups of queries that fall within 10 percent and 10-25 percent of the bound only and these queries are added to the Candidate Query set (CQ), while all queries satisfying the bounds are added to the query set SQ. The output segregation are all the leaf nodes in the kd-tree. For repartitioning, it only considers those pairs of partitions from the output that are siblings in the kd-tree and have imprecision greater than zero for the queries in the candidate query set.

VI. CONCLUSION

The proposed system proposes an accuracy-constrained privacy-preserving access control framework for relational data has been proposed. K-Anonymity model is accommodating with minimum deception based data access control mechanism. Privacy preserved data access control mechanism is improved with additional mining model and cell level access control. The system reduces the deception rate in query processing. Access control mechanism is adapted for incremental mining model. Time complexity is reduced in the system. The system provides the dynamic policy management mechanism. The proposed system proposes the application specific anonymization. The proposed system plan to extend the proposed privacy-preserving access control to incremental data and cell level access control.

Future Work: Our future work is going to make data preserving access control to increase the level of data efficient control. We can make anonymity that is mapped to equivalence partition of data. For future work, it plan to extend the proposed privacy-preserving access control to cell level access control and can use the l-diversity instead of k-anonymity method.

VI. REFERENCES


Author’s Details:
Mr Bandi Pradeep Raju is a student of Adikavi Nannaya University, Rajahmundry. Presently he is pursuing his M.Tech [Computer Science & Engineering] from this college and he received his B.Tech from Chaitanya Institute of Engineering Technological Rajahmundry, affiliated to JNT University, Kakinada in the year 2014. His area of interest includes Data Ware Housing and Data Mining and its objectives including Object Oriented Programming Languages, all current trends and techniques in Computer Science.

Mrs M.Kamala Kumari, is working as Associate Professor, Department of Computer Science & Engineering Adikavi Nannaya University Rajahmundry, her area of interest includes Computer Networks, Data Warehouse and Data Mining, Mobile Computing.

International Journal of Advanced Technology and Innovative Research
Volume. 08, IssueNo.07, July-2016, Pages: 1308-1311