ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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

Kumba Bala Subramanyam
P.G. Scholar (M. Tech),
Department of CSE,
Srinivasa Institute of Technology & Sciences,
Ukkayapalli, Kadapa, Andhra Pradesh.

K. Rajasekhar Reddy
Assistant Professor,
Department of CSE,
Srinivasa Institute of Technology & Sciences,
Ukkayapalli, Kadapa, Andhra Pradesh.

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 an accuracy-constrained privacy-preserving access control framework. The access control policies define selection predicates available to roles while the privacy requirement is to satisfy the k-anonymity or l-diversity. An additional constraint that needs to be satisfied by the PPM is the imprecision bound for each selection predicate. The techniques for workload-aware anonymization for selection predicates have been discussed in the literature. However, to the best of our knowledge, the problem of satisfying the accuracy constraints for multiple roles has not been studied before. In our formulation of the aforementioned problem, 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.

Index Terms:

Access control, privacy, k-anonymity, l-diversity, query evaluation, application specific anonymization.

I.INTRODUCTION:

Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data. The term is actually a misnomer. Thus, data mining should have been more appropriately named "knowledge mining from data," which is unfortunately somewhat long. "Knowledge mining," a shorter term, may not reflect the emphasis on mining from large amounts of data. Nevertheless, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material . Thus, such a misnomer that carries both "data" and "mining" became a popular choice. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. AS organizations increase their adoption of database systems as the key data management technology for day-to-day operations and decision making, the security of data managed by these systems becomes crucial. Damage and misuse of data affect not only a single user or application, but may have disastrous consequences on the entire organization. The recent rapid proliferation of Web based applications and information systems have further increased the risk exposure of databases and, thus, data protection is today more crucial than ever. It is also important to appreciate that data needs to be protected not only from external threats, but also from insider threats, the proposed system uses the concept of imprecision bound foreach permission to define a threshold on the amount of imprecision that can be tolerated. Existing workload aware Anonymization techniques. In this proposed system the focus is on a static relational table that is 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 [5], [6] minimize the inexactness mixture for all queries and also the inexactness else to every permission/query within the anonymized small information isn't glorious.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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 individual permissions 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 ano- nymization. The anonymization for continuous information publishing has been studied in literature .during this paper the main focus is on a static relative table that's anonymized just one occasion.

II .Related work :

Gabriel GhinitaData anonymization does not constrain the privacy of information .Surajit Chaudhuri Data privacy is inadequate for supporting data. Ninghui Li Privacy-preserving microdata does not provide access control for different roles. Elisa BertinoRole-based access control does not tells about scalability. ShariqRizvi Fine Grained Access Control does not support privacy related database. ZahidPervaiz[8] Accuracy and privacy interaction produces K-PIB problem such as lower imprecision bound.

III .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 privacy requirements and imprecision constraints on predicates set by the access control mechanism. It has been formulated this interaction as the problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB). It gives hardness results for the k-PIB problem and present heuristics for partitioning the data to the satisfy the privacy constraints and the imprecision bounds.

EXISTING SYSTEM:

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. Existing workload aware anonymization techniques minimize the imprecision aggregate for all queries and the imprecision added to each permission/query in the anonymized micro data is not known. Making the privacy requirement more stringent (e.g., increasing the value of k or l) results in additional imprecision for queries.

DIS-ADVANTAGES:

1.Their is no privacy for users
2.The sensitive information,even after the removal of identifying attributes,is still susceptible to linking attacks by the authorized users.

IV.PROPOSED SYSTEM:

The heuristics proposed in this paper foraccuracy-constrained privacy-preserving access controlare also relevant in the context of workload-aware anonymization. The anonymization for continuous data publishinghas been studied in literature. In this paperthe focus is on a static relational table that is anonymizedonly once. To exemplify our approach, role-basedaccess control is assumed. However, the concept of accuracyconstraints for permissions can be applied to anyprivacy-preserving security policy, e.g., discretionaryaccess control.

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 theprivacy protection module with the reference of imprecision bound level and give the exact result to the user.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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 amont of information from the microdata. The county epi-demiologist will not lose their information.

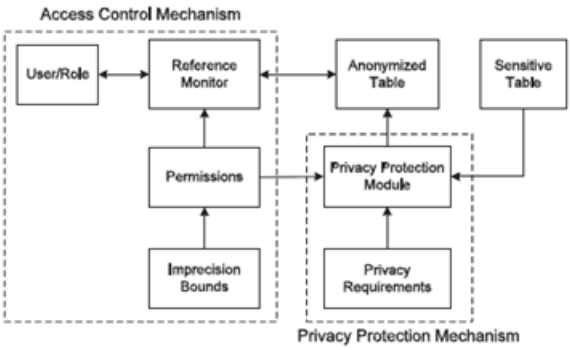


Fig 1.System model for application anonymization

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.

ADVANTAGES:

1.accuracy constrained privacy preserving access.
2.It's maintain data's in secure manner.

k- anonymity:

k-anonymityis a property possessed by certain anonymized data. In the context of k-anonymization problems, a database is a table with n rows and m columns. Each row of the table represents a record relating to a specific member of a population and the entries in the various rows need not be unique. The values in the various columns are the values of attributes associated with the members of the population.

Name	Age	Gender	State of domicile	Religion	Disease
Ramsha	29	Female	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	Kerala	Hindu	Viral infection
Salima	28	Female	Tamil Nadu	Muslim	TB
Kaker	27	Male	Karnataka	Parsi	No illness
Joan	24	Female	Kerala	Christian	Heart-related
Bahuksana	23	Male	Karnataka	Buddhist	TB
Rambha	19	Male	Kerala	Hindu	Cancer
Kishor	29	Male	Karnataka	Hindu	Heart-related
John	17	Male	Kerala	Christian	Heart-related
John	19	Male	Kerala	Christian	Viral infection

Fig 2. Before k-anonymity

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 all 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 with a broader category. For example, the value '19' of the attribute 'Age' may be replaced by ' ≤ 20', the value '23' by '$20 < \text{Age} \leq 30$' , etc.

Name	Age	Gender	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	Tamil Nadu	*	TB
*	20 < Age ≤ 30	Male	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	Karnataka	*	TB
*	Age ≤ 20	Male	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	Karnataka	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Heart-related
*	Age ≤ 20	Male	Kerala	*	Viral infection

Fig 3. 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.

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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

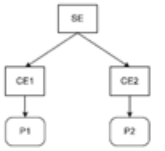
Here, planned system solely considers conjunctive queries (The adversative queries will be expressed as a union of conjunctive queries), wherever every question will be expressed as a d-dimensional hyper-rectangle. The linguistics for question analysis on associate degree anonymized table Ta must be outlined. once the equivalence category partition (Each equivalence category will be delineate as a d-dimensional hyper-rectangle) is totally b within the question region, all tuples within the equivalence category ar a part of the question result.

V.IMPLEMENTATION MODULES:

1.Access control policy
2.Anonymity
3.Anonymization with impression Bounds
4. Accuracy-Constrained Privacy-Preserving Access Control
5.Top-Down Heuristic

Access control policy:

Role	Designation
SE	State Epidemiologist
CE1	County 1 Epidemiologist
CE2	County 2 Epidemiologist



Permission	Authorized Query Predicate
P1	Location = County 1 ∧ Age = 15-65 ∧ Syndrome = Influenza
P2	Location = County 2 ∧ Age = 15-65 ∧ Syndrome = Influenza

Syndromic surveillance systems are used at the state and federal levels to detect and monitor threats to public health. The department of health in a state collects the emergency department data (age, gender, location, time of arrival, symptoms, etc.) from county hospitals daily. Generally, each daily update consists of a static instance that is classified into syndrome categories by the department of health. Then, the surveillance data is anonymized and shared with departments of health at each county.An access control policy is given in Fig. 1 that allows the roles to access the tuples under the authorized predicate, e.g., Role CE1 can access tuples under PermissionP1. The epidemiologists at the state and county level suggest community containment measures ,e.g., isolation or quarantine according to the number of persons infected in case of a flu outbreak. According to the population density in a county, an epidemiologist can advise isolation if the number of persons reported with influenza are greater than 1,000 and quarantine if that number is greater than 3,000 in a single day.

The anonymization adds imprecision to the query results and the imprecision bound for each query ensures that the results are within the tolerance required. If the imprecision bounds are not satisfied then unnecessary false alarms are generated due to the high rate of false positives.

Anonymity:

	QI_1	QI_2	S_1
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

(a) Sensitive table

ID	QI_1 Age	QI_2 Zip	S_1 Disease
1	0-20	10-30	Flu
2	0-20	10-30	Fever
3	20-30	10-30	Diarrhea
4	20-30	10-30	Fever
5	20-30	10-30	Flu
6	30-40	20-40	Fever
7	30-40	20-40	Flu
8	30-40	20-40	Diarrhea

(b) 2-anonymous Table

anonymity is prone to homogeneity attacks when the sensitive value for all the tuples in an equivalence class is the same. To counter this shortcoming, l-diversity has been proposed and requires that each equivalence Fig. 1. Access control policy. class of T_ contain at least l distinct values of the sensitive attribute. For sensitive numeric attributes, an l-diverse equivalence class can still leak information if the numeric values are close to each other. For such cases, variance diversity has been proposed that requires the variance of each equivalence class to be greater than a given variance diversity parameter. The table in Fig. 2a does not satisfy k-anonymity because knowing the age and zip code of a person allows associating a disease to that person. The table in Fig. 2b is a 2-anonymous and 2-diverse version of table in Fig. 2a. The ID attribute is removed in the anonymized table and is shown only for identification of tuples. Here, for any combination of selection predicates on the zip code and age attributes, there are at least two tuples in each equivalence class.

Anonymization with impresision Bounds:

we formulate the problem of k-anonymousPartitioning with Imprecision Bounds and presentan accuracy-constrained privacy-preserving access controlframework. Imprecise data means that some data are known only to the extent that the true values lie within prescribed bounds while other data are known only in terms of ordinal relations. Imprecise data envelopment analysis (IDEA) has been developed to measure the relative efficiency of decision-making units (DMUs) whose input and/or output data are imprecise. In this paper, we show two distinct strategies to arrive at an upper and lower bound of efficiency that the evaluated DMU can have within the given imprecise data.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The optimistic strategy pursues the best score among various possible scores of efficiency and the conservative strategy seeks the worst score. In doing so, we do not limit our attention to the treatment of special forms of imprecise data only, as done in some of the studies associated with IDEA. We target how to deal with imprecise data in a more general form and, under this circumstance, we make it possible to grasp an upper and lower bound of efficiency.

Accuracy-Constrained Privacy-Preserving Access Control:

An accuracy-constrained privacy-preserving access control mechanism. (arrows represent thedirection of information flow), is proposed. The privacyprotection mechanism ensures that the privacy and accuracygoals are met before the sensitive data is available tothe access control mechanism. The permissions in theaccess control policy are based on selection predicates onthe QI attributes. The policy administrator defines thepermissions along with the imprecision bound for eachpermission/query, user-to-role assignments, and role-topermissionassignments . The specification of theimprecision bound ensures that the authorized data hasthe desired level of accuracy. The imprecision boundinformation is not shared with the users because knowing the imprecision bound can result in violating thePrivacy requirement. The privacy protection mechanismis required to meet the privacy requirement along withthe imprecision bound for each permission.

Top-Down Heuristic:

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 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 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.

Algorithm 1: TDH1

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

```
Initialize Set of Candidate Partitions(CP ← T)
for (CP_i ∈ CP) do
    Find the set of queries QO that overlap CP_i
    such that ic^{QO_j}_{CP_i} > 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
return (P)
```

VI.RESULT AND CONCLUSIONS:

Figure 4.

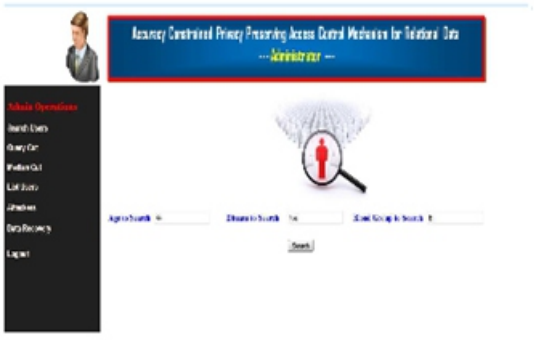


Figure 5.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

CONCLUSION:

An accuracy-constrained privacy-preserving access control framework for relational data has been proposed. The framework is a combination of access control and privacy protection mechanisms. The access control mechanism allows only authorized query predicates on sensitive data. The privacy preserving module anonymizes the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism. We formulate this interaction as the problemof k-anonymous Partitioning with Imprecision Bounds (k-PIB). We give hardness results for the k-PIB problem and present heuristics for partitioning the data to the satisfy the privacy constraints and the imprecision bounds. In the current work, static access control andrelational data model has been assumed. For future work, we plan to extend the proposed privacy-preserving access control to incremental 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.