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

A Framework for Privacy-Preserving Access Control Mechanism in Relational Data

Nemalipuri Saikiran
M.Tech Student
Department of CSE,
Sri Venkateswara College of Engineering and Technology.

RVLS N Sastry, M.Tech
Associate Professor
Department of CSE,
Sri Venkateswara College of Engineering and Technology.

Abstract:

Access control Mechanisms are security features that control how users and systems communicate and interact with other systems and resources. Access control mechanisms give organization the ability to control, restrict, monitor, and protect resource availability, integrity and confidentiality. Privacy Protection Mechanism (PPM) uses suppression and generalization of relational data to anonymize and satisfy privacy needs.

The access control policies define selection predicates available to roles while the privacy requirement is to satisfy the k-anonymity or l-diversity. Imprecision bound constraint is assigned for each selection predicate. Essentially analysis within the data processing or information mining with sub space of information security is loosely classified into access management analysis and data privacy analysis. Each plays vital role in information security thence we've integrated these methodologies to boost our security on relative information. Mistreatment privacy protective mechanism we will generalize and suppress our relative information to anonymize and satisfy privacy needs against identity and attribute speech act. In this paper we implement a Privacy Protection Mechanism for protecting sensitive information from unauthorized users.

Keywords: ***Privacy, k-anonymity, l-diversity, Imprecision bounds, Access control.***

Introduction:

Data privacy problems are getting progressively vital for our society. This can be proved by the very fact that the accountable management of sensitive knowledge is expressly being mandated through laws like the Sarbanes-Oaxley Act and therefore the insurance movability and answerability Act (HIPAA). Protective individual privacy is a crucial downside. Access management mechanisms area unit accustomed make sure that solely approved data is obtainable to users. However, sensitive data will still be illused by approved users to compromise the privacy of shoppers. Databases within the globe area unit typically massive and sophisticated. The challenge of querying such infuse in a very timely fashion has been studied by the database, data processing and knowledge retrieval communities, however seldom studied within the security and privacy domain. We have a tendency to have an interest within the downside of protective access privacy for users once querying massive databases of many lots of or thousands of gigabytes of knowledge. This can be a more durable downside than in alternative domains as a result of the matter contents of queries area unit themselves protected against the info server.

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

The concept of privacy-preservation for sensitive data can require the enforcement of privacy policiesor the protection against identity disclosure by satisfying some privacy requirements. We investigate privacy-preservation from the anonymity aspect. 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.

Existing System:

Organizations collect and analyze consumer data to improve their services. Access Control Mechanisms (ACM) is 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.

Disadvantages of Existing System:

- Minimize the imprecision aggregate for all queries.
- The imprecision added to each permission/query in the anonymized micro data is not known.
- Not satisfying accuracy constraints for individual permissions in a policy/workload.

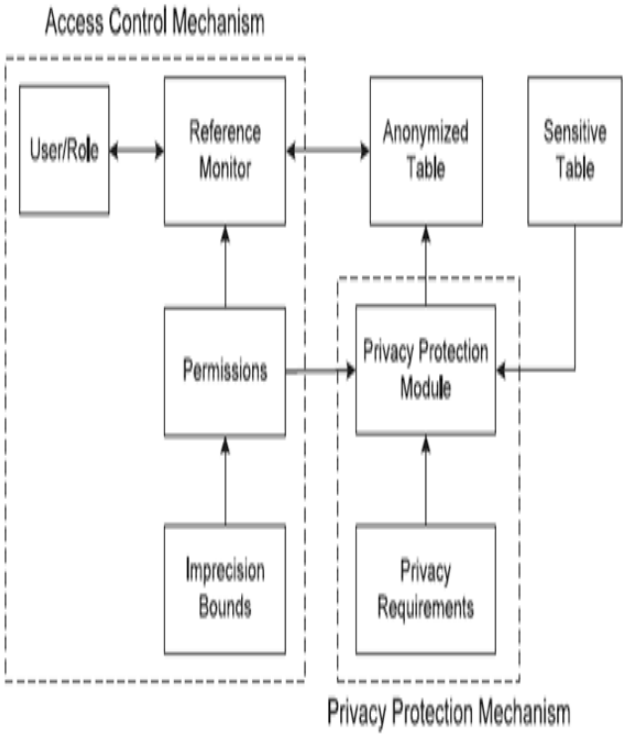
Proposed System:

The heuristics proposed in this paper for accuracy-constrained privacy-preserving access control are also relevant in the context of workload-aware anonymization. The anonymization for continuous data publishing has been studied in literature. In this paper the focus is on a static relational table that is anonymized only once. 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.

Advantages of Proposed System:

- Formulate the accuracy and privacy constraints.
- Concept of accuracy-constrained privacy-preserving access control for relational data.
- Approximate the solution of the k-PIB problem and conduct empirical evaluation.

System Architecture:



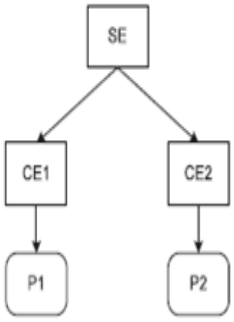
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Implementation Modules:

1. Access control policy
2. Anonymity
3. Anonymization with impresision 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

	QI_1	QI_2	S_1
ID	Age	Zip	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 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. a 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. b is a 2-anonymous and 2-diverse version of table in Fig. a. 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

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

code and age attributes, there are at least two tuples in each equivalence class.

Anonymization with impresision Bounds:

We formulate the problem of k-anonymous Partitioning with Imprecision Bounds and present an accuracy-constrained privacy-preserving access control framework. 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. 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 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 because knowing the imprecision bound can result in violating the Privacy requirement. The privacy protection mechanism is required to meet the privacy requirement along with the 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_Qj
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_CPi^QOj > 0
    Sort queries QO in increasing order of B_Qj
    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)
```

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

Conclusion:

The planned additive approach of access management and privacy protection mechanisms in our system provides a lot of security by adding cryptography to information and information is retrieved during a custom-made approach which will build users to access during as lot of versatile approach. Any access management concentrates on anomaly users to avoid privacy problems security .The ACM allows solely licensed user predicates on sensitive information and PPM anonymizes the information to satisfy privacy necessities and inexactness constraints on predicates set by the access management mechanism.

References:

[1] Zahid Pervaiz, Walid G. Aref, Senior Member, Arif Ghafoor, Fellow, and Nagabhushana Prabhu,"Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data", IEEE TRANSACTIONS, VOL. 26, NO. 4, APRIL 2014.

[2] 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.

[3] S. Chaudhuri, R. Kaushik, and R. Ramamurthy, "Database Access Control & Privacy: Is There a Common Ground?" Proc. Fifth Biennial Conf. Innovative Data Systems Research (CIDR), pp. 96-103, 2011.

[4] N. Li, W. Qardaji, and D. Su, "Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy," Arxiv preprint arXiv:1101.2604, 2011.

[5] Femi Olumofin and Ian Goldberg "Preserving Access Privacy Over Large Databases", University of Waterloo Waterloo, Ontario, Canada N2L 3G1, 2012

[6] Yung-Wang Lin, Li-Cheng Yang, Luon-Chang Lin, and Yeong-Chin Chen "Preserving Privacy in Outsourced Database", International Journal of Computer and Communication Engineering, Vol. 3, No. 5, September 2014.

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

[8] JaideepVaidya, BasitShafiq, Wei Fan,DanishMehmood, and David Lorenzi "A Random Decision Tree Framework for Privacy-Preserving Data Mining",IEEE transactions on dependable and secure computing, vol. 11, no. 5, september/october 2014

[9] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "A Framework for Efficient Data Anonymization Under Privacy and Accuracy Constraints," ACM Trans. Database Systems, vol. 34, no. 2, article 9, 2009.

[10] X. Xiao, G. Bender, M. Hay, and J. Gehrke, "Ireduct: Differential Privacy with Reduced Relative Errors," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2011.