International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 07, July-2015

Security and Accuracy Constrained Task-Role based Access Control and Privacy Preserving Mechanism for Relational Data

Pratik Bhingardeve
Computer Department
Smt.Kashibai Navale College of Engineering,
Vadgaon (BK), Pune 411041, India

Prof. D. H. Kulkarni
Computer Department
Smt.Kashibai Navale College of Engineering,
Vadgaon (BK), Pune 411041, India

***Abstract*— During a time where the moment subtle elements of our life are recorded and put away in databases, a reasonable Catch is developing between the need to protect the security of people and the need to utilize these gathered information for examination, open strategy definition, and other. Security saving miniaturized scale information distributed right now does not have a strong hypothetical establishment. Existing cryptography-based work for security safeguarding information mining is still too ease back to be in any way successful for extensive scale information sets to face today's huge information challenge. A PPM can utilize concealment and speculation of social information to anonymize and fulfill security prerequisites, e.g., k-namelessness and l-assorted qualities, against character and property exposure. Notwithstanding, protection is accomplished at the expense of accuracy of approved data. So I have come with an idea of aggregating these two techniques i.e. PPM and ACM with task-role based access to provide high security and privacy for our relational data.**

Keywords—Task Role Based Access Control,Encryption, Privacy Preserving, k-anonymit,l-diversity.

I. INTRODUCTION

In associate in nursing age where the minute details of our life recorded and confine databases, a clear contradiction is rising between the necessities to preserve the privacy of individuals and so the need to use these collected data for analysis, public policy formulation, and other. Privacy-preserving tiny data publication presently lacks a solid theoretical foundation. Existing cryptography-based work for privacy-preserving processing continues to be too slow to be effective huge for giant for large} scale data sets to face today's big data challenge. A PPM can use suppression and generalization of relative data to anonymize and satisfy privacy requirements, e.g., k-anonymity and L-diversity, against identity and attribute revealing. However, privacy is achieved at the worth of accuracy of authorized information [3].

Before going additional discussion concerning our work we have a tendency to should clear our basic ideas regarding this paper i.e. Privacy protecting and its varied algorithms and Access management ways that. The conception of privacy-preservation for sensitive data can want the group action of privacy policies or the protection against identity revealing by satisfying some privacy requirements. we've an inclination to analyze privacy-preservation from the obscurity facet. Anonymization algorithms use suppression and generalization of records to satisfy privacy requirements with tinyest distortion of small data. The obscurity techniques is also used with associate in nursing access management mechanism to verify every security and privacy of the sensitive data. The privacy is achieved at the worth of accuracy associate in nursing inexactitude is introduced at intervals the authorized data below AN access management policy [1]. Access management mechanisms unit of measure accustomed make sure that alone approved data is getable to users. However, sensitive data will still be exploited by approved users to compromise the privacy of shoppers. Databases within the planet unit of measure usually giant and sophisticated. The challenge of querying such infuse really terribly timely fashion has been studied by the information, process and knowledge retrieval communities, however seldom studied within the protection and privacy domain. We've got AN inclination to possess associate in nursing interest within the disadvantage of protective access privacy for users once querying giant databases of the gigabytes of data.

We initial analyze k-anonymization strategies and show however they fail to produce ample protection against re-identification, that it had been designed to guard. Access management mechanisms defend sensitive data from unauthorized users. However, once sensitive data is shared and a Privacy Protection Mechanism (PPM) isn't in situation, a licensed user will still compromise the privacy of someone resulting in identity revelation. A PPM will use suppression and generalization of relative knowledge to anonymize and satisfy privacy necessities, e.g., k-anonymity and l-diversity, against identity and attribute revelation [1]. However, privacy is achieved at the value of preciseness of approved data. K-anonymity could be a property possessed by bound anonymized knowledge. The construct of k-anonymity was developed as an endeavor to unravel the problem: "Given person-specific field-structured knowledge, turn out a unleash of the information with scientific guarantees that the people World Health Organization are the topics of the information can't be re-identified whereas the information stay much helpful." A unleash of knowledge is claimed to possess the k-anonymity property if the knowledge for every person contained within the unleash can't be distinguished from a minimum of k-1 people whose information conjointly seem within the unleash. Two main strategies for this are Suppression: during this methodology, bound values of the attributes are replaced by associate asterisk '*'. All or some

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

values of a column is also replaced by '*'. Generalization: during this methodology, individual values of attributes are replaced by with a broader class.

L-diversity could be a style of cluster primarily based anonymization that's accustomed preserve privacy in knowledge sets by reducing the roughness of a knowledge illustration [3]. The l-diversity model is associate extension of the k-anonymity model that reduces the roughness of information illustration mistreatment techniques as well as generalization and suppression specified any given record maps onto a minimum of k alternative records within the data. The l-diversity model handles a number of the weaknesses within the k-anonymity model wherever protected identities to the extent of k-individuals isn't like protective the corresponding sensitive values that were generalized or suppressed, particularly once the sensitive values among a gaggle exhibit homogeneity. The l-diversity model adds the promotion of intra-group diversity for sensitive values within the anonymization mechanism.

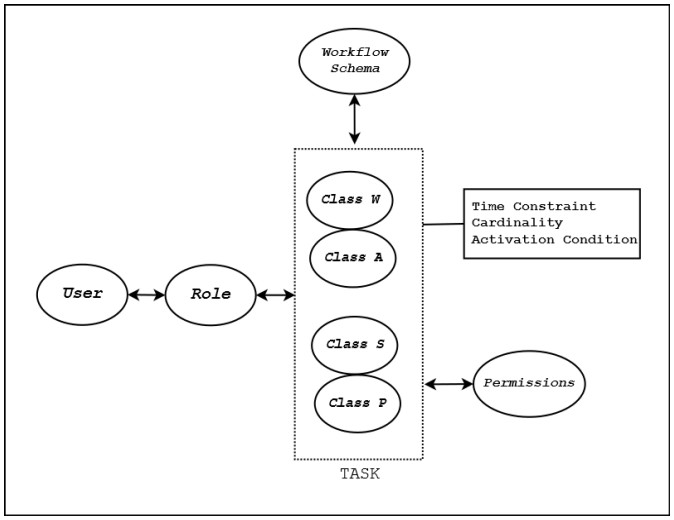


Fig 1: The Architecture of TRBAC Model.

Task-Role Based Access Control (TRBAC) model provides a lot of modeling power. In TRBAC, roles cluster tasks and users along, and permissions to business objects are certain with tasks [4]. TRBAC prevents users from access business objects once not death penalty the corresponding tasks, and therefore satisfies the confidentiality necessities of contemporary enterprises [2]. Business processes are operated supported not solely roles however conjointly tasks. With each as core ideas, our model provides a lot of modeling power. Roles cluster tasks and users along, and permissions to business objects are certain with tasks.

II. Related Work

For understanding the fundamental ideas we have a tendency to has referred varied reference papers for understanding ideas like PPM, K-anonymity, encoding at varied levels etc. In [1], they have used the conception of impreciseness certain for every permission to outline a threshold on the number of exactitude which will be tolerated. Existing work aware anonymization techniques minimize the impreciseness combination for all queries and also the impreciseness another to every permission within the anonymized small knowledge isn't notable. creating the privacy demand a lot of demanding (e.g., increasing the worth of k or l) ends up in extra impreciseness 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 are relevant within the context of workload-aware anonymization. The anonymization for continuous knowledge commercial enterprise has been studied in literature. during this paper the main focus is on a static relative table that's anonymized just one occasion. To exemplify our approach, role-based access management is assumed. However, the conception of accuracy constraints for permissions is applied to any privacy-preserving security policy, e.g., discretionary access management.

A. Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data[1].

Authors: Zahid Pervaiz, Walid G. Aref, Arif Ghafoor and Nagabhushana Prabhu.

Abstract and Conclusion— Access management mechanisms shield sensitive data from unauthorized users. However, once sensitive data is shared and a Privacy Protection Mechanism (PPM) isn't in situ, a licensed user will still compromise the privacy of an individual resulting in identity revealing. A PPM will use suppression and generalization of relative information to anonymize and satisfy privacy necessities, e.g., k-anonymity and l-diversity, against identity and attribute revealing. during this formulation of the said drawback, we tend to propose heuristics for anonymization algorithms and show through empirical observation that the projected approach satisfies impreciseness bounds for additional permissions and has lower total impreciseness than this The frame-work could be a combination of access management and privacy protection mechanisms. The access management mechanism permits solely licensed question predicates on sensitive information. The privacy-preserving module anonymizes the information to satisfy privacy necessities and impreciseness constraints on predicates set by the access management mechanism.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

B. A Delegation Framework for Task-Role Based Access Control in WFMS[2].

Authors: Hwai-Jung Hsu and Feng-Jian Wang.

Abstract and conclusion: Access management is very important for shielding data integrity in work flow management system (WFMS). Compared to traditional access management technology like discretionary, mandatory, and role based mostly access management models, task-role-based access management (TRBAC) model, AN access management model supported each tasks and roles, meets additional needs for contemporary enterprise environments. However, few discussions on delegation mechanisms for TRBAC area unit created. Within the framework, the methodology for delegations requested from each users and WFMS is mentioned. The constraints for delegate choice like delegation loop and separation of duty (SOD) area unit self-addressed. With the framework, a sequence of algorithms for delegation and revocation of tasks area unit created bit by bit.

C. Methods for Access Control: Advances and Limitations[3].

Author: Ryan Ausanka-Crues.

Abstract and conclusion: In this paper author focused on various types of access control mechanisms such as DAC, MAC and RBAC there way of working and there advantages and disadvantages. Access management models have return quite ways in which since the initial implementations of mackintosh and DAC. Researchers have learned volumes regarding the complexities of maintaining security policies through model applications and with RBAC have return terribly near seamlessly integrating and confidentiality. Work still must be done on translating policies into verifiable model implementations and in efficient and correct management of those implementations.

D. L-Diversity: Privacy beyond k-Anonymity[7]

Authors: Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer.

Abstract and conclusion: Publishing data about individuals without revealing sensitive information about them is an important problem. In recent years, a new definition of privacy called k-anonymity has gained popularity. In a k-anonymized dataset, each record is indistinguishable from at least k−1 other records with respect to certain "identifying" attributes. In this paper we show with two simple attacks that a k-anonymized dataset has some subtle, but severe privacy problems. First, we show that an attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes. Second, attackers often have background knowledge, and we show that k-anonymity does not guarantee privacy against attackers using background knowledge. We give a detailed analysis of these two attacks and we propose a novel and powerful privacy definition called ℓ-diversity. In addition to building a formal foundation for ℓ-diversity, we show in an experimental evaluation that ℓ-diversity is practical and can be implemented efficiently.

Following table focuses on various techniques and methods for access control mechanisms.

Characteristics	Simple RBAC [4]	Improved RBAC [5]	TAC [8]	TRBAC[5]
Delegation of Permissions	Grant	Grant & Transfer	No	Grant & Transfer
Delegation of Tasks	No	No	Transfer	Transfer
Delegation of Task Instances	No	No	No	Transfer
Time Constraints	No	No	No	Yes
Automatic Delegation	No	No	No	Yes

Table 1: Comparison between various ACM methods.

III. PROPOSED SYSTEM

Authentication and Security problem in existing system:

In existing system, we tend to user's doesn't have economical privacy and correct constraints and System powerless to retrieve information in made-to-order approach. Another issue is system doesn't give security for information to be retrieved or loaded. Our projected system tries to resolve these problems and our System provides additional security by adding cryptography to information. Information will be retrieved in a very made-to-order approach which will create users to access in a very additional versatile approach. Access management concentrates on anomaly users to avoid privacy problems which offer benefits like we tend to area unit able to formulate the accuracy and privacy constraints and construct of accuracy-constrained privacy-preserving access management for relative information can give security and privacy to users. The heuristics projected during this paper for accuracy strained privacy-preserving access management are relevant within the context of workload-aware anonymization. The framework may be a combination of access management and privacy protection mechanisms. The access management mechanism permits solely licensed question predicates on sensitive information. The privacy conserving module anonymizes the info to fulfill privacy needs and inexactitude constraints on predicates set by the access management mechanism.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

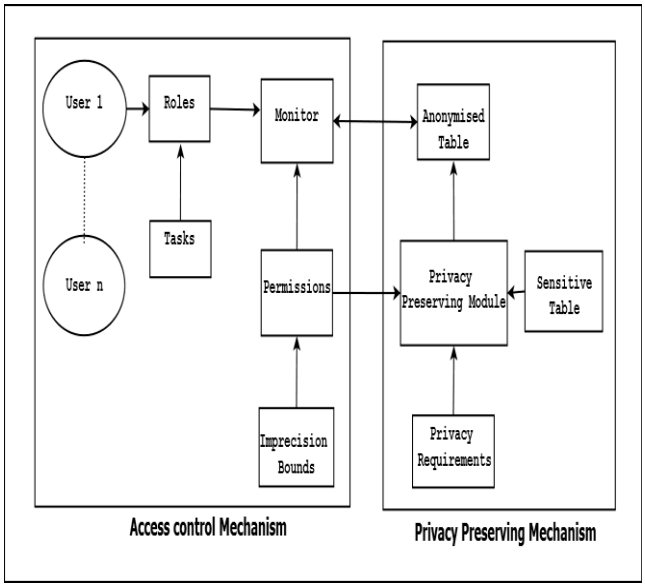


Fig 2. Proposed Architecture

IV. Mathematical Terms And Algorithm

The exact tuple values in a relation are replaced by the generalized values after the anonymization. In this case, access control enforcement over the generalized data needs to be defined. In this section, we discuss the Relaxed and Strict access control enforcement mechanisms over anonymized data. The access control enforcement by reference monitor can be of the following two types:

1. Relaxed. Use overlap semantics to allow access to all partitions that are overlapping the permission.

2. Strict. Use enclosed semantics to allow access to only those partitions that are fully enclosed by the permission.

Both schemes have their own pros and cons. Relaxed enforcement violates the authorization predicate by giving access to extra tuples but is beneficial for applications where low cost of a false alarm is tolerable as compared to the risk associated with a missed event. We further assume that under relaxed enforcement if the imprecision bound is violated for a permission then that permission is not assigned to any role.

In this section, the relaxed enforcement of access control is analyzed probabilistically. The access control policy administrator sets the imprecision bound BQi for each query, and requires that the imprecision bound for the least number of queries be violated by PPM. The policy administrator might revise the imprecision bounds for queries and further relax the access control policy if it is known with a high probability that a large number of queries will violate the bounds and access requests for roles will be denied.

Given:- n tuples, (tuples are uniformly distributed)
To Find:-

1. Expected imprecision for a randomly selected query.
2. Expected number of partitions overlapping the query.
3. We need to find the expected partition size $|P_{e}|$ and expected length of intervals l_{i}^{Qj} .

Proof:-

We use the domain length of each attribute in the domain space and then divide this length of the first QI attribute by 2. The length of interval is updated and the new partition will now contain tuples. For the next division, another QI attribute is selected and the process is repeated until the expected partition size is

Terms and Formula:-

- Identifier: Attributes, e.g., name and social security that can uniquely identify an individual. These attributes are completely removed from the anonymized relation.[1]

- 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[1]

- Sensitive attribute: Attributes, e.g., disease or salary, that if associated to a unique individual will cause privacy breach[1]

- Query Imprecision (impQi):- 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 impQi

$$(\text{impQi}) = |Qi(T^*)| - |Qi(T)| \quad (1)$$

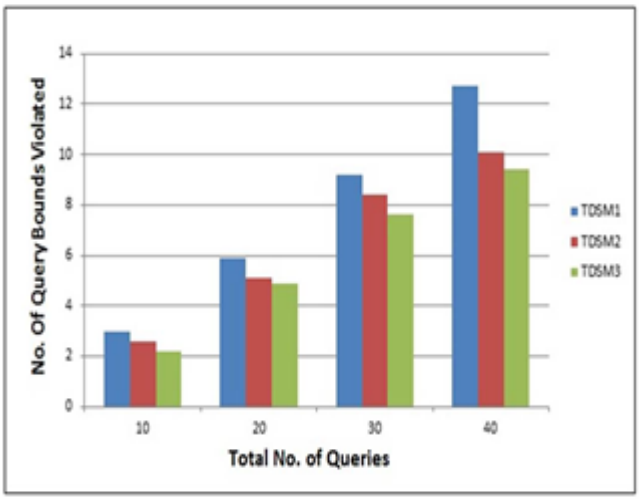
- Query Imprecision Bound (BQi): The query imprecision bound, denoted by BQi , is the total imprecision acceptable for a query predicate Qi and is preset by the access control administrator.

- Query Imprecision Slack (SQi): The query imprecision slack, denoted by SQi for a Query, say Qi, is defined as the difference between the query imprecision bound and the actual query imprecision.[1]

In the Top-Down Heuristic algorithm 3 (TDH3, for short), we modify TDH2 so that the time complexity of $O(d|Q|n\log n)$ can be achieved at the cost of reduced precision in the query results.

Given a partition, TDH3 checks the query cuts only for the query having the lowest imprecision bound. Also, the second constraint is that the query cuts are feasible only in the case when the size ratio of the resulting partitions is not highly skewed. We use a skew ratio of 1:99 for TDH3 as a threshold. If a query cut results in one partition having a size greater than hundred times the other, then that cut is ignored. TDH3 algorithm is listed in This Algorithm 3. In Line 4 of Algorithm 3, we use only one query for the candidate cut. In Line 6, the partition size ratio condition needs to be satisfied for a feasible cut. If a feasible query cut is not found, then the partition is split along the median as in Line 11. The time complexity of the TDH2 algorithm is $O(d|Q^2|n2)$, which is not scalable for large data sets (greater than 10 million tuples) hence TDH3 is better than both TDH1 and TDH2.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)



TDH3 Algorithm

Input: T, k, Q and BQi

Output: P

1. Initialize Candidate part.(CP←T)
2. For (CPi ε CP) do
   //DFT or preorder traversal
3. Find set of queries QO that overlap CPi such that ic>0;
4. Select query from QO with smallest BQi.;
5. Create query cut in each dimension;
6. Reject cuts in each dimensions;
7. select dimension and cut having least imprecision;
8. if (feasible cut found) then
9. Create new partition and add to CP
10. Else
11. Split CPi recursively along median till anonymity req. is satisfied.
12. Compact new partition and add to P
13. Update BQi
14. Return(P)

V. Implementation and Results

The proposed system is developed to improve security. The results of dissertation work are explained in this section. Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

There are mainly two modules in this project

1. User/Doctor.
2. Admin.
   1. User Module:-

This module consists of various sub modules which carries out following functions:-

i. Sensitive Data Retrieval:- 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.

ii. Encryption: - Encryption is the conversion of data into a form, called a cipher text, that cannot be easily understood by unauthorized people. Decryption is the process of converting encrypted data back into its original form, so it can be understood.

iii. Request for Permission Key:-User want to convert the encrypted data to original format then user enter permission key that key send by admin via SMS.

iv. Retrieve from Encrypted Database:- The process of decoding data that has been encrypted into a secret format. Decryption requires displaying the user data.

v. Anonymous Data Retrieval:-Anonymity is prone to homogeneity attacks when the sensitive value for all the tuples in an equivalence class is the same Overlap. Include all tuples in all partitions that overlap the query region. This option will add false positives to the original query result.

Doctor Data!!!

Doctor Id	Doctor Name	Location	Permission
7	mrt	Pune	Granted
8	abhi	Pune	Granted
9	jay	Pune	Granted
10	nikam	Mumbai	Granted
11	anil	Pune	Granted
12	sanjay	Pune	Granted
13	salim	Pune	Granted
14	amin	Pune	Granted
15	Dr. Deshmukh	Pune	Granted
16	p	Pune	Granted
17	dipti	Mumbai	Granted
18	nitin	Pune	Granted

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

2. Admin Module:-This module consists of various sub modules which carries out following functions:-

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

ii. Query Cut: - A query cut is dined 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 and the end of the query interval are considered to split a partition along the jth dimension [5].

iii. Median Cut: - The median cut generates a balanced tree with height lgn and the work done at each level is n. The partitions created by TDSM have dimensions along the median of the parent partition. A compaction procedure has been proposed in where the created partitions are replaced by minimum bounding boxes. This step improves the precision of the anonymized table for any given query workload by reducing the overlapping partitions[7].

VI. Conclusion

The planned additive approach of access management and privacy protection mechanisms in our system provides a lot of security by adding encryption to information and information is retrieved during a custom-made approach which will build users to access during as lot of versatile approach. 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. And by using TRBAC, the permissions to business objects are bound with tasks & time constraint is also added which is helpful for security constraint.

VII. Acknowledgment

I am very grateful to my guide Prof. D. H. Kulkarni and staff members for their generous help for their many thoughtful and insightful comments that have greatly influenced this paper. We are also thankful to my colleagues and parents for their invaluable support.

References

[1] Zahid Pervaiz, Walid G. Aref, Arif Ghafoor, and Nagabhushana Prabhu "Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data" IEEE Trans. On Knowledge And Data Engineering, Vol. 26, No. 4, April 2014.

[2] Hwai-Jung Hsu And Feng-Jian Wang, "A Delegation Framework For Task-Role Based Access Control In Wfms", Journal Of Information Science And Engineering 27, 1011-1028 (2011)

[3] Ryan Ausanka , Crues Harvey, " Methods for Access Control: Advances and Limitations" California university,2012

[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] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu, "Approximation Algorithms for k-Anonymity," J. Privacy Technology, vol. 2005112001, pp. 1-18, 2005.

[6] 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.

[7] 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.

[8] Sejong Oh And Seog Park "An Improved Administration Method on Role-Based Access Control in the Enterprise Environment" Journal Of Information Science And Engineering 17, 921-944 (2001)

(This work is licensed under a Creative Commons Attribution 4.0 International License.)