International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

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

Madhuri S. Lambe[1], Vrunda K. Bhusari[2]

[1]P.G. Student, Savitribai Phule Pune University, Department of Computer Engineering, BSIOTR, Wagholi, Pune, Maharashtra, India

[2]Assistant Professor, Savitribai Phule Pune University, Department of Computer Engineering, BSIOTR, Wagholi, Pune, Maharashtra, India

Abstract: ***To prevent the misuse of sensitive data by the authorized users and provide both privacy and security of the sensitive data. New approach has investigated privacy-preservation from the anonymity aspect. The access control mechanisms and privacy preservation mechanisms protect the data from unauthorized or third party user. When there is a lack in privacy preserving mechanism (PPM) and data is shared with others, the authorized user may need to compromise the privacy of data or Information. The privacy preservation can be achieved through anonymization techniques like generalization or suppression. Along with privacy the precision of the authorized data is important. The aim of the work is to provide better security and minimum level of precision to the retrieved data, for that in this paper an accuracy constrained privacy preserving access control mechanism is implemented with additional constraint on each selection predicate called imprecision bounds. New approach plan to extend the proposed privacy-preserving cell level access control .Todays fast growing world, the malicious intent or hacking purpose also increasing. So there is a need to provide a better security to our system.***

Keywords: Access Control, Anonymization, Privacy preservation, Query evaluation.

1. Introduction

Several organizations and agencies publish microdata, e.g., medical data, customer data or census data for research and other public benefit purposes. In an age where the microdata of each individual are recorded and stored, an inconsistency arises between the necessity to protect the privacy of individuals and also to use these data for medical research, trend analysis and societal improvement. Hence, the private information of an individual should not be revealed from the microdata.

Data owners constantly seek to make better use of data they possess and utilize to extract useful information.Human Data mining and knowledge discovery in databases are two areas that extract the previously unknown patterns from large amounts of data. Recent search in data collection, data distribution and related technologies have initiated where existing data mining algorithms should be reconsidered of privacy preservation. Privacy preserving is a research direction in data mining, where algorithms are used for the side-effects they suffer in data privacy. Twofold is the main consideration in privacy preserving data mining. First, sensitive data like identifiers, names, addresses and the like should be modified from the database, the recipient of the data not to be able to compromise another persons privacy. Second, Sensitive knowledge which can be mined from a database because knowledge can compromise data privacy. The main objective in privacy preserving data mining is to develop algorithms for modifying the original data in some way, so that the private data and private knowledge remain confidential so authorized information only get to that user. When confidential information can be access from released data by unauthorized users is also commonly called the database inference problem. The privacy-preservation for sensitive data can require the privacy policies or the protection against identity disclosure by satisfying some privacy/accuracy requirements [1]. It allows queries only on the authorized part of the database. Predicate based access control has proposed, where user authorization is limited to pre-defined predicates. Privacy preservation and access control mechanism are important concept in every information sharing system. New approach are going to implement cell level security access control mechanism this helps to normal user to protect personal information as per they wants to do not share with others.

There are two levels Access Control:

A. Table level Access Control Mechanism

1) Tuple Level: When evaluating user queries, assume a model called Truman. In this model, a user query is modified by the access control mechanism and only the authorized tuples are returned.
2) Column level: It allows queries to execute on the authorized column of the relational data only
3) Cell level: It is implemented by replacing the unauthorized cell values by NULL values [2].

B. Role-based Access Control

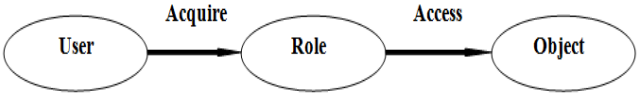


Figure 1: Role Based Access Control

The permissions on objects based on roles in an organization.
1) U = user1, user2, user3 where U is a set of Users.
2) R = role1, role2, role3 where R is a set of Roles.
3) P = permission1, permission2, permission3 where P is a set of Permission.

2. Preliminaries

Based on anonymity, role-based access control and privacy definitions based on anonymity are over-viewed. Query evaluation semantics, imprecision, and the Selection Mondrian algorithm are explained. A relation T = {A1;A2; : : : ;An} where A_i is an attribute, T* is the anonymized version of the relation T. Consider T is a static relational table.

2.1 Attributes

1) Identifier: Attributes which is uniquely identify an individual. These attributes are completely removed from the anonymized relation.
2) Quasi-identifier(QI): Attributes, that can potentially identify an individual based on other information available to an adversary. QI attributes are generalized to satisfy the anonymity requirements.
3) Sensitive attribute: Attributes, that if associated to a unique individual will cause a privacy breach.

2.2 Anoymity Definitions

1) Equivalence Class (EC): It is a set of tuples having the same QI attribute values.
2) k-anonymity Property: A anonymized table satisfies the k-anonymity property if each equivalence class has k or more tuples.
3) Query Imprecision: Difference between the number of tuples returned by a query evaluated on an anonymized relation and the number of tuples for the same query on the original relation.
4) Query Imprecision Bound: It is the total imprecision acceptable for a query predicate and is preset by the access control administrator.

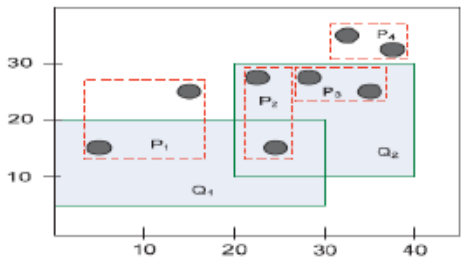


Figure 2: Anonymization satisfying imprecision bounds

5) Query Cut: The splitting of a partition along the query interval values. For a query cut using Query, both the start of the query interval and the end of the query interval are considered to split a partition along the dimension.

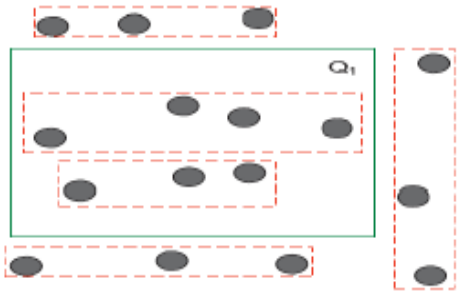


Figure 3: Query Cut

3. Literature Review

The most relevant concepts underlying the notion of databsase security and summarize the most well-known techniques. E. Bertino focus on access control systems, on which a large amount of research has been devoted, and describe the key access control models. B.Fung[3] Work on Attacks Model and Privacy Models. A fine-grained access control model depend on authorization views that allows authorization transparent querying; that is, user queries in terms of the database relations, and are valid if information contained in authorization views[4]. They extend on authorization-transparent querying by introducing a new notion of validity, conditional validity. They give a powerful set of inference rules to check for query validity. They demonstrate techniques by describing how an existing query optimizer can be extended to perform access control checks by incorporating inference rules. Enforcement of access control and privacy policies have studied [5]. The relation between the access control and the privacy protection mechanisms has been missing. Newly, Chaudhuri et al. have considered access control with privacy mechanisms [6].The privacy requirement in terms of k-anonymity shown by Li et al. [7] that k-anonymity offers similar privacy guarantees.

4. System Implementation

4.1 System Architecture

An accuracy-constrained privacy-preserving cell level access control mechanism is a combination of access control and privacy protection mechanisms.

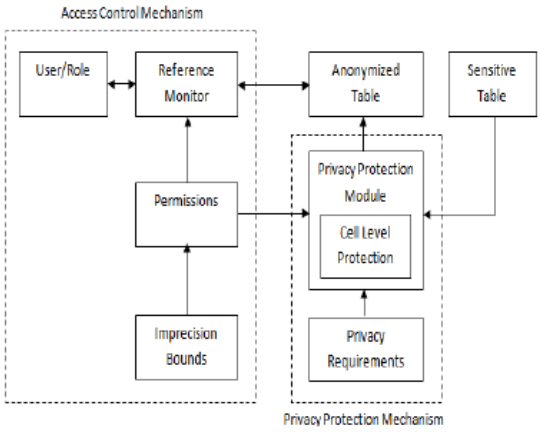


Figure 4: Accuracy Constrained Privacy Preserving Cell Level Access Control

Individual sensitive is easily inferred by an attacker. Protecting data privacy is an important problem,in micro-data distribution.Anonymization is to protect individual privacy, with minimal impact on the quality of the resulting data. The system produce an anonymous view based on a target class of workloads, consisting of one or more data mining tasks and selection predicates. The advantage of anonymity is resisting the attacker's inference attacks.

Access control mechanism for relational data is constructed with the privacy preservation based model. Role Based Access Control (RBAC) scheme protects the sensitive data

with minimum imprecision values. K-Anonymity model is integrated with minimum imprecision based data access control mechanism. Privacy preserved data access control mechanism is improved with incremental mining model and cell level access control. The proposed system reduces the imprecision rate in query processing. Access control mechanism is adapted for incremental mining model. Time complexity is reduced in the proposed system. The proposed system provides the dynamic policy management mechanism.

A.Access Control Mechanism

The Access control mechanism allows only authorized query predicates on sensitive data.

1) User/Role: It allows defining permissions on objects based on roles in an organization. An RBAC is composed of a set of Users, a set of Roles, and a set of Permissions. Assume that the selection predicates on the QI attributes define permission [8]. UA is a user-to-role assignment relation and PA is a role to permission assignment relation. When a user assigned to a role executes a query, the tuples that are used satisfying the conjunction of the query predicate and the permission are returned.
2) Permissions: It based on selection predicates on the QI attributes. The imprecision bound for each query, user-to-role assignments, and role-to permission assignments.
3) Imprecision Bound: It ensures that the authorized data has the desired level of accuracy. The imprecision bound can be used to meet the privacy requirement. The privacy protection mechanism is required to meet the privacy requirement according to the imprecision bound for each permission[9].

B. Privacy Protection Mechanism

PPM ensures that the privacy and accuracy goals are met before the sensitive data is available to the access control mechanism.

1) Privacy Protection Module: It anonymizes the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism.
2) Sensitive Table.

Table 1: Sensitive Table.

	QI1	QI2	S1
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

The Table 1 does not satisfy k-anonymity because knowing the age and zip code of a person i.e quasi identifier allows associating a disease to that person.

3) Anonymous Table.

Table 2: Anonymous Table

	QI1	QI2	S1
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

The proposed system get information in a anonymous version of sensitive table. 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.

4.2 Algorithm

Two algorithm are going to implemented in proposed system.

1) Top-Down Heuristic 1 (TDH1)
Input: T, K, Query and BQj.
Output: Partition (P).
1. Initialize set of candidate partitions.
2. for ($CP_i \epsilon$ CP) do
3. Search the set of queries that overlap candidate partitions.
4. Arrange queries in increasing order of BQ_j.
5. while (feasible cut is not found) do
6. Select query from QO.
7. Generate query cuts according dimension.
8. Select dimension and cut having least overall imprecision.
9. Check feasible cut found or not if feasible cut then
10. Generate new partitions and add to CP.
11. otherwise
12. Split candidate partitions recursively along median upto anonymity requirement is satisfied.
13. Shrink new partitions and add to P
14. return (P)

2) Top-Down Heuristic 3 (TDH3)
Input: T; K; Query and BQj
Output: Queryset and Resultset.
1. Initialize candidate partitions.
2 .for each cp_i in cp {
Find the queries which gets the overlap Results and add it into Queryset.
Select Query from Queryset with small Resultset.
Create querycut with each dimension.
Select Imprecision for all queries into queryset.
If (feasiblecut found)
if(check with usersecured attributes)
add it into CP
Else
do recursively until anonymity requirement is not satisfly.
Add into Resultset.

Update BQ_j.
3. Return (Resultset)

4.3 Flow of System

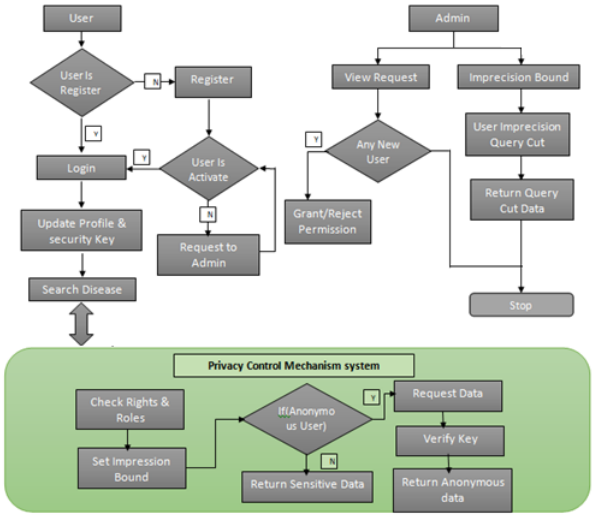


Figure 5: Flow Diagram

4.4 Mathematical Model

$S=\{A, D_o, A_d, A_r, O_p, S, F\}$
Ai=Each Patient have only one account and contains user id.
Do=Doctor login to system.
Ad=Admin login to system.
Ar,=Authenticate request to admin.
Op= Operations.
S=Success {connection establish, sending or retrieving data successfully }
F=Failure {not a valid user, incorrect password}

1. To add new account :
n
Σ Ai=1 if 1=true , 0=false
i=1
Patient alredy exists
Else
Add new account in it.

2.To login in Account :
n n n
Σ Ai =1 then Σ Ai = Σ Ud =1
i=1 i=1 i=1
(verifies password in user Details according to the patient authority ID present in Account)
account information which contain patient id and password are match then Login Successfully.
Else
login failed.

3.Operations:
n
ΣOpi =1 then Operation Success
i=1
Else
Operation Failed

Operations are patient Authority Validation, Storing and Viewing patient details , doctor Authority Validation, Viewing patient details.

5. Results

Consider one data sets for the empirical evaluation of the proposed heuristics. The data set is the Medical data set from the UC Irvine Machine Learning Repository having 450 tuples and is the defacto benchmark for k-anonymity research. The attributes in the Medical data set are: User Id, Name, Password, Blood Group, Email Id, Mobile no., Location, Date of Birth, age, Address, Gender, Disease name, Pincode.

Registration	User_Id	User_Name	User_Password	Blood_Group
3	111	maruti	maruti	O+
4	454	sanket	sanket	O+
5	3435	abhi	abhi	O+
7	656	kantilal	kanti	O+
8	7634	MRT	mrt	O+
9	4355	ashvini	aa	B+
10	kunti	kunti	kunti	o+ve
11	madhuri	madhuri	madhuri	b+ve
12	mayur	mayur	mayur	a
13	az73646	zadeabhi	123456	A+
14	1234	Mahavir	mahi	B+
15	11	madhuri11	madhuri11	0
16	AGLambe123	AshishL	Ayush123@	0-
17	ML	mayurl	mayurl	A
18	55	madhuri55	madhuri55	0
19	11	11	11	o
20	mm	mm	mm	0

We use 20 and 50 queries generated randomly as the workload/permissions for the Medical data set. In this query set we remove all queries which have empty value.

Generate query load: We set query impression (k) value according to that we generate quires and store it into Query table.

Median Query: Select all set of queries which satisfy all attribute values and store it into Median Query set.

Query Cut: select randomly queries and store it into Query Cut set and divide it into an equal intervals Which is called Uniform Query set.

Retrive From Encrypted Database		
Age	PinCode	Disease
24	8484	Fever
24	54212	fever
25	22	fever
24	2234	fever
28	411028	fever
27	413304	fever
27	440028	fever
26	411057	Fever
27	411024	fever

Retrive From Encrypted Database		
Age	PinCode	Disease
1-30	60-90	Fever
1-30	60-90	fever
1-30	1-30	fever
1-30	60-90	fever
1-30	60-90	fever
1-30	60-90	fever
1-30	60-90	fever
1-30	60-90	Fever
1-30	60-90	fever

We are going to implement 2 algorithms
1) TDH1
2) TDH3

For the k-anonymity experiments, we fix the value of k and change the query imprecision bounds from 5 to 20 value with increments of 5. Then, we find the number of queries whose bounds have not been satisfied by each algorithm for the uniform query workload. The results for k-anonymity are given for the Medical data set for k values of 5,10,15, 20.

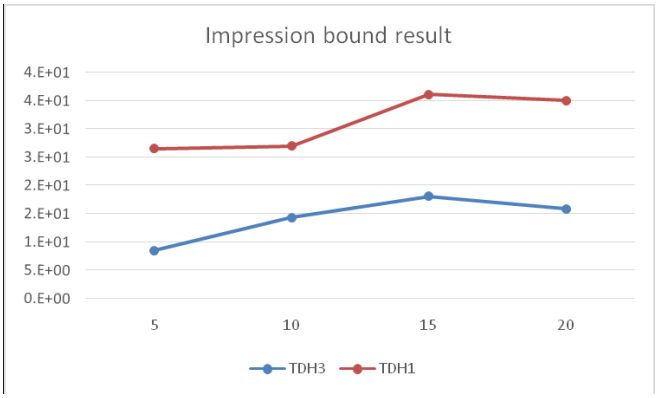
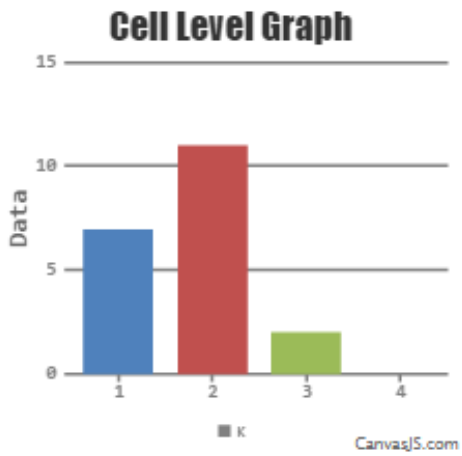


Figure 6: Imprecision Bound Result



6. Conclusion

The system architecture is a combination of access control mechanism and privacy protection Mechanism(Cell Level). New approach provide data access control to user it means we prevent anonymous user to access confidential data. If any user wants to access data which is secure then he/she wants to raise the request to the responsible user then this user send you the key which is used to access his/her data. It strictly prevent anonymous user to access unauthorized data. The proposed system implemented table level, row level, and cell level access control mechanism. Query set is to access the data and this sets assigned to user for roles and permission purpose. The access control mechanism allows only authorized query predicates on sensitive data. The privacy preserving module anonymizes the data to get privacy requirements. Imprecision bound estimation is optimized. The system proposed privacy-preserving access control to incremental data and dynamic access control. The system reduces the time complexity. The advantage of anonymity is resisting the attacker's inference attacks.

7. Acknowledgement

I would like to express my sincere gratitude to my guide Prof. Vrunda K. Bhusari for her continuous support, patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this paper.

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] K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, and D. DeWitt, Limiting Disclosure in Hippocratic Databases, Proc. 30th Intl Conf. Very Large Data Bases, pp. 108-119, 2004.
[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] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy, Extending Query Rewriting Techniques for Fine-Grained Access Control, Proc. ACM SIGMOD Intl Conf. Management of Data, pp. 551-562,2004.
[5] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi, Extending Relational Database Systems to Automatically Enforce Privacy Policies, Proc. 21st Intl Conf. Data Eng., pp. 1013-1022, 2005.
[6] 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.
[7] N. Li, W. Qardaji, and D. Su, Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy, Arxiv preprintarXiv:1101.2604, 2011.

[8] S. Chaudhuri, T. Dutta, and S. Sudarshan, Fine Grained Authorization through Predicated Grants, Proc. IEEE 23rd Intl Conf. Data Eng., pp. 1174-1183, 2007.
[9] Zahid Pervaiz, Walid G. Aref, Senior Member, IEEE, Arif Ghafoor, Fellow, IEEE, and Nagabhushana Prabhu, Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data, IEEE Tran.knowledge and Data Eng.,vol.26,n0.4,April 2014.

References

Ms. Madhuri S. Lambe. Received the Bachelors degree (B.E.) Computer Engineering in 2010 from College of Engineering and Technology, Akola. She is now pursuing Masters degree (Computer Engineering), from BSIOTR, Wagholi, Pune, Maharashtra.

Prof. Vrunda K. Bhusari. received her M.Tech(Computer Engineering) from Bharati Vidyapeeth, Pune and now she is working as Assistant professor, Department Of Computer Engineering, Bhivarabai Sawant Institute of Technology & Research, Wagholi, Pune, Maharashtra.. Her research areas include Network Security.