Journals: International Journal of Software & Hardware Research in Engineering
ISSN-2347-4890
Volume 4 Issue 2 February, 2016

Privacy-Preserving And Access Control Mechanism For Relational Data Using PACT

Miss Suvarna Shingare[1];Prof.Srinu Dharawat[2]
ME Student,GSMCOE[1];Research Scholar, GSMCOE[2]
sdshingare18@gmail.com[1];sreevasmtech@gmail.com[2]

ABSTRACT

Access management mechanisms addresses the problem of accessing information from unauthorized user's to precise information in individual data record. However, once sensitive information is shared and a Privacy Protection Mechanism (PPM) isn't in situation, a licensed user will still compromise the privacy of someone leading to identity revelation. A PPM will use suppression and generalization of relative knowledge to anonymize and satisfy privacy requirements, e.g., k-anonymity and l-diversity, against identity and attribute revelation. However, privacy is achieved at the value of precision of approved info. During this paper, we tend to propose associate degree accuracy constrained privacy-preserving access management framework. The access management policies outline choice predicates out there to roles whereas the privacy demand is to satisfy the k anonymity or l-diversity. an extra constraint that must be glad by the PPM is that the inexactitude certain for every choice predicate. The techniques for workload-aware anonymization for choice predicates are mentioned within the literature. However, to the simplest of our data, the matter of satisfying the accuracy constraints for multiple roles has not been studied before. In our formulation of the said downside, we tend to propose heuristics for anonymization algorithms and show through empirical observation that the planned approach satisfies inexactitude bounds for additional permissions and has lower total inexactitude than the present state of the art.

Index Terms— Access control, privacy, k-anonymity, query evaluation

1. INTRODUCTION

The construct of privacy-preservation for sensitive information will need the social control of privacy policies or the protection against identity speech act by satisfying some privacy needs [1].Access management Mechanisms area unit accustomed make sure that solely approved in fois on the market to users. In this project, we tend to area unit planned privacy-preservation. The sensitive info, even when the removal of distinctive attributes, remains at risk of linking attacks by the approved users. We use the construct of in exactitude sure for every permission to outline a threshold on the quantity of inexactitude that may be tolerated. the problem of satisfying accuracy constraints for individual permissions during a policy or workload has not been studied before. The heuristics planned during this project for accuracy-constrained privacy-preserving access management also are relevant within the context of workload-aware anonymization. During this project the most focus is on a static relative table that's anonymized one time. To exemplify our approach, role-based access management is assumed. However, the construct of accuracy constraints for permissions will be applied to any privacy-preserving security policy, e.g., discretionary access management the contributions of the project area unit as follows. First, we tend to formulate the accuracy and privacy constraints because the drawback of k-anonymous Partitioning with in exactitude Bounds (k-PIB) and provides hardness results. Second, we tend to introduce the construct of accuracy-constrained privacy-preserving access management for relative information. Third, we tend to propose heuristics to approximate the answer of thek- PIB drawback and conduct empirical analysis.

2. LITERATURE REVIEW

Previously in literature survey we are going to discuss all recent methods over the Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data

E. Bertino and R. Sandhu[1] they proposed a system to alice owns a k-anonymous database and needs to determine whether her database, when inserted with a tuple owned by Bob, is still k-anonymous. Also, suppose that access to the database is strictly controlled, because for example data are used for certain experiments that need to be maintained confidential. Clearly, allowing Alice to directly read the contents of the tuple breaks the privacy of Bob (e.g., a patient's medical record); on the other hand, the confidentiality of the database managed by Alice is violated once Bob has access to the contents of the database. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting Alice and Bob know the contents of the tuple and the database, respectively. In this paper, we propose two protocols solving this problem on suppression-based and generalization-based k-anonymous and confidential databases. The protocols rely on well-

known cryptographic assumptions, and we provide theoretical analyses to proof their soundness and experimental results to illustrate their efficiency.

P. Samarati[2] they proposed a system to an Today's globally networked society places nice demand on the dissemination and sharing of knowledge. Whereas within the past discharged data was largely in tabular and applied mathematics type, several things decision these days for the discharge of specific information (micro data). so as to shield the namelessness of the entities (called respondents) to that data refers, information holders usually take away or inscribe specific identifiers like names, addresses, and phone numbers. De identifying information, however, provides no guarantee of namelessness. Discharged data usually contains alternative information, like race, birth date, sex, and code, which will be connected to in public obtainable data to re identify respondents and inferring data that wasn't supposed for revelation. During this project we have a tendency to address the matter of cathartic micro data whereas safeguarding the namelessness of the respondents to that the info refer. The approach relies on the definition of k-anonymity. A table provides k-anonymity if makes an attempt to link expressly characteristic data to its content map the data to a minimum of k entities. We have a tendency to illustrate however k-anonymity will be provided while not compromising the integrity (or truthfulness) of the data discharged by victimization generalization and suppression techniques. We have a tendency to introduce the conception of least generalization that captures the property of the discharge method to not distort the info over required to attain k-anonymity, and gift an rule for the computation of such a generalization. we have a tendency to additionally discuss potential preference policies to decide on among completely different least generalizations.

B. Fung, K. Wang, R. Chen, and P. Yu[3] they proposed a system to The collection of digital information by governments, corporations, and individuals has created tremendous opportunities for knowledge- and information base decision making. Driven by mutual benefits or by regulations that require certain data to be published, there is a demand for the exchange and publication of data among various parties. Data in its original form, however, typically contains sensitive information about individuals, and publishing such data will violate individual privacy. The current practice in data publishing relies mainly on policies and guidelines as to what types of data can be published and on agreements on the use of published data. This approach alone may lead to excessive data distortion or insufficient protection. Privacy-preserving data publishing provides methods and tools for publishing useful information while preserving data privacy. Recently, PPDP has received considerable attention in research communities and many approaches have been proposed for different data publishing scenarios. In this survey, we will systematically summarize and evaluate different approaches to PPDP, study the challenges in practical data publishing, clarify the differences and requirements that distinguish PPDP from other related problems and propose future research directions.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam [4] they proposed a system to the social network data to be published in public. With the rapid growth of web application development the need for private data to be published has grown multifold. Representing the individual sensitive labels in a graphical structure for the ease of access and quick re-identification is an important issue to be addressed. In this paper, we have made a detailed surveyed about the existing techniques that preserve the sensitive data in social network data. It is pragmatic that preserving the graph structure and label re-identification by adding some noise nodes to the graph makes difference in degree is inferred from literature. Several techniques dealt in literature to deal with abnormal graph structure are studied in detail. The various possible attacks on social network data and techniques to prevent and handle the same are studied in detail.

K. LeFevre, D. DeWitt, and R. Ramakrishnan[5] they proposed a system to find data for quality prediction. Early in the life cycle, projects may lack the data needed to build such predictors. Prior work assumed that relevant training data was found nearest to the local project. To provide defect data-set owners with an effective means of privatizing their data prior to release, the MORPH understands how to maintain class boundaries in a data-set. MORPH is a data mutator that moves the data a random distance, taking care not to cross class boundaries. The value of training on this MORPHed data is tested via a 10-way within learning study and a Cross learning study using Random Forests, Naive Bayes, and Logistic Regression for ten object-oriented defect data-sets from the ROMISE data repository. Measured in terms of exposure of sensitive attributes, the MORPHed data was four times more private than the unMORPHed data. Also, in terms of the f-measures, there was little difference between the MORPHed and unMORPHed data (original data and data privatized by data-swapping) for both the cross and within study. We conclude that at least for the kinds of OO defect data studied in this project, data can be privatized without concerns for inference efficacy.

3. PROPOSED SYSTEM

3.1 System Architecture:

The architecture of PACT is shown in Figure 1. PACT has two phases: the offline phase the initial processing that takes place before any query is processed; and the online phase, which shows how an inter-organization query is processed in runtime. PACT is a middleware system that requires very few changes to be done on the legacy systems of any organizations involved. The offline procedure of PACT is to (1) translate the (syntactic) access control policy of each organization to a semantic access control policy against the organization's ontology, and (2) prepare the other metadata used by the mediator.

To illustrate the online aspect of PACT, suppose an employee of Organization A needs some information from organization B. In Step 1, since the user does not know B's data schema, the users SQL query is written against the user's user ontology. In this way, the actual column and table names used in the query will be obfuscated by as ontology.Then the obfuscated query will be encrypted. In Step 2, a SQL parser is used to "decompose" the query into several column-level or table-level access requests. However, at this stage these requests are expressed with A's ontology and role lattice, and they cannot be directly processed by Organization B. Hence, in Step 3, the mediator translates these requests into several semantic accesses requests expressed with B's ontology and role lattice via an algorithm called semantic request mediation. This algorithm uses encrypted mappings between terms in A's ontology and B's ontology and the mapping between roles in A's role lattice and B's role lattice. In Step 4, these requests are checked against B's semantic access control policy. In Step 5, the filtered yet authorized semantic requests will be decrypted and translated into some syntactic access requests against B's schema, In Step 6, the SQL query is processed by B's DBMS. The DBMS may forgo the security checking since it has already been done. However, the query results cannot be directly returned to A because they are not expressed against A's ontology and the user can be confused about the meaning of the results. In Step 7, the responder translates the data and sends it back to the user.

We present the set of core techniques used by the PACT system and demonstrate their uniqueness and merits. First, we discuss the offline operations of PACT. Second, we discuss the runtime operations of PACT. Although for clarity we only address the scenario with two organizations, PACT can easily handle multiple organizations with one or more mediators.

3.2 Algorithm Used:

3.2.1.*Top-Down Heuristic 1 (TDH1):*

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. 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.

3.2.2.Top-Down Heuristic 2 (TDH1)

In the Top-Down Heuristic 2 algorithm (TDH2, for short), the query bounds are updated as the partitions are added to the output.

3.2.3 Top-Down Heuristic 3 (TDH1)

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.

3.3 Mathematical Model**:**

Set Theory:

1. Query Imprecision is defined as ,

$$impQ_i = |Q_i(T^*)| - |Q_i(T)|, where$$

$$|Q_i(T^*)| = \sum_{Ec\ overlaps\ Q_i} |EC| \qquad (1)$$

The queryQi is evaluated over T # by including all the tuples in the equivalence classes that overlap the query region.

2.The query imprecision slack, denoted bys Qi for a Query, say Qi, is defined as,

$$S_{Q_i} = \begin{cases} B_{Q_i} - impQ_i, & if\ impQ_i \leq B_{Q_i} \\ 0 & otherwise \end{cases} \qquad (2)$$

3. The partition imprecision cost is a vectorfic

Q1 Pi ;...;ic Qn Pi g, where ic Qj Pi is the imprecision cost of a Partition Pi 2Pwith respect to a Query Qj.

$$ic_{p_i}^{Q_j} = |P_i - Q_j|,$$

$$impQ_i = \sum_{P_i \in P} ic_{P_i}^{Q_j} \qquad (3)$$

where minus sign denote the set difference.

4. LetIQj be a non-negative random variable that denotes the query imprecision. Then, the expected imprecision for a queryQj is,

$$E(I_{Q_j}) \leq \left[\left(\prod_{i=1}^{d}\left\lceil\frac{l_i^{Q_j}+l_i^{Pe}}{l_i^{Pe}}\right\rceil\right) * |Pe|\right] - |Q_j|. \quad (4)$$

5. Let I Qi be a non-negative random variable that denotes the query imprecision. LetX1;...;Xn be an independent Poisson trial, where Xi is a random variable that is equal to 1 if a query, say Qi,

$$E[X] = \sum_{i=1}^{n} pi \leq \sum_{i=1}^{n} \frac{E(I_{Q_i})}{(B_{Q_i}+1)} \quad (5)$$

3.3.4 Proposed Algorithm:

Real-time Query Processing:

Step 1: Schema Obfuscation and Query Encryption.

When a query is issued by an employee of the requestor organization, the query is written against the employee's user ontology instead of the requestor's schema. In this way, the requestor's schema is kept private. We refer to this process as schema obfuscation. Conceptually, schema obfuscation replaces a schema term in the query with a randomly chosen synonym in the user ontology.

Step 2: SQL Query Parsing.

When an encrypted query arrives at the mediator, it is parsed and all the table and column names are extracted. The mediator expands wildcards (like select *) by replacing the wildcard character using the encrypted attributes associated with the table in the query.

Step 3: Encrypted Query Rewriting

Step 4: Semantic Access Control

After a query has been rewritten, PACT checks to verify if the translated role has the permissions to access the tables and columns in the rewritten queries. A novelty of PACT is that it does semantic access control at the mediator. The advantage is that the queries that are rejected are not sent to the responder.

Step 5: Semantic to Syntactic Query Translation

At the responder, a semantic query is translated to a syntactic query by replacing ontology terms in the query with their equivalent terms that appear in the responder's database.

Step 6: Query Evaluation.

The query is then evaluated at the database and the results returned to the requestor.

Step 7: Returning the Results. The results of the query are sent back to the requestor after the data is translated using the split-merge attributes rules and the conversion functions. The data is sent back directly from the responder to the requestor bypassing the mediator

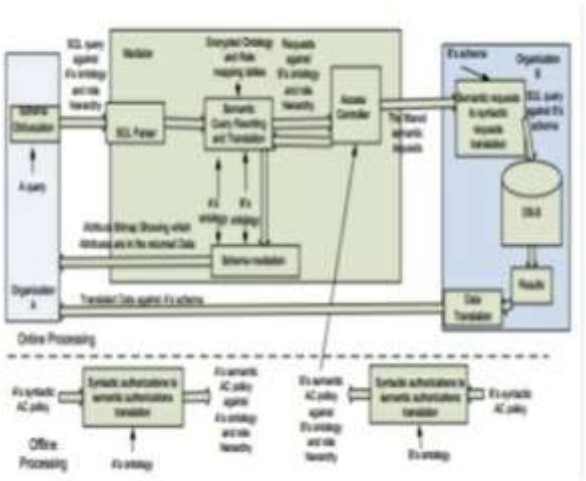

Fig 1: Proposed System Diagram

Results of Practical Work:-

The results compared here are time graph between existing and proposed system is shown in below figure.

Graph of Time comparison

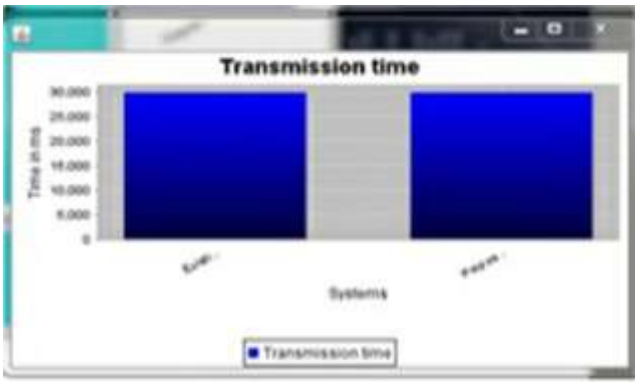

Fig.2 Time Comparison between proposed and existing system

Fig.3 Performance for a different query workload for the Census

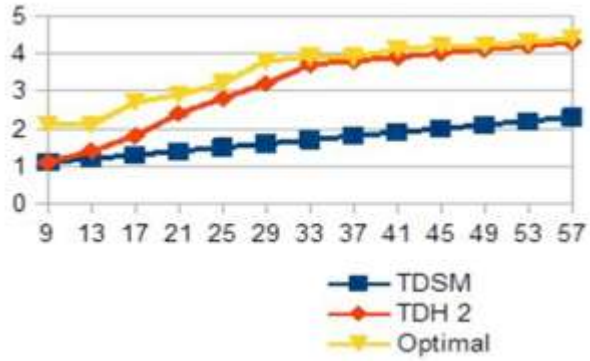

Fig 4.Comparison with optimal solution.

5.CONCLUSION AND FUTURE WORK

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 problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB).

6.ACKNOWLEDGEMENT

The specially thank to my guide "Prof. Srinu Dharawat" and department of computer Engg. Of GSM COE, Balewadi giving his contribution in writing paper.

7.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,

[6] T. Iwuchukwu and J. Naughton, "K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc. 33rd Intl 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 Technical White 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 Intl 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 Intl 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 Intl Conf. Very Large Data Bases, pp. 108-119, 2004.

[13] D. Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn, and R. Chandramouli, "Proposed NIST Standard for Role-Based Access Control," ACM Trans. Information and System Security, vol. 4, no. 3, pp. 224- 274, 2001

[14] 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.

[15] E. Otoo, D. Rotem, and S. Seshadri, "Optimal Chunking of Large Multidimensional Arrays for Data

Warehousing," Proc. ACM 10th Int'l Workshop on Data Warehousing and OLAP, pp. 25-32, 2007.

[16] W. Hoeffding, "On the Distribution of the Number of Successes in Independent Trials," The Annals of Math. Statistics, vol. 27, no. 3, pp. 713-721, 1956.

[17] R. Agrawal, P. Bird, T. Grandison, J. Kiernan, S. Logan, and W. Rjaibi, "Extending Relational Database Systems to Automatically Enforce Privacy Policies," Proc. 21st Int'l Conf. Data Eng., pp. 1013-1022, 2005.

[18] 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.

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

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

