IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-10 E-ISSN: 2347-2693

Enabling Privacy Preservation Technique to Protect Sensitive Data with Access Control Mechanism Using Anonymity

Barkha Kasab[1*], Vinayak Pottigar[2] and Swapnaja Ubale[3]

Department of Computer Science and Engineering, Solapur University, Solapur, India
SKN Sinhgad College of Engineering, Korti, Pandharpur, India

www.ijcseonline.org

Received: Sep/28/2015 Revised: Oct/10/2015 Accepted: Oct /22/2015 Published: Oct /31/ 2015

***Abstract*—** Access control mechanisms shield sensitive data from unauthorized users. On the other hand, when sensitive information is released and a Privacy Protection Mechanism (PPM) is not in set up, an authorized user can still compromise the privacy of a person leading to identity exposure. A PPM can use concealment and speculation of social information to anonymize and fulfill protection prerequisites here some algorithm i.e. k-anonymity and l-diversity used against identity as well as attribute disclosure. However, security is accomplished at the expense of exactness of authorized data or information. Paper describes an accuracy-constrained privacy-preserving access control model. Role based access control policies define selection predicates available to roles and it should be satisfy the k-anonymity or l-diversity. An extra limitation that should be fulfilled by the PPM is the imprecision headed for every choice predicate. However, the problem of satisfying the accuracy constraints used for multiple roles has not been studied before. In our formulation ,technique used heuristics for anonymity algorithms and also done experiments to show proposed approach satisfies imprecision bounds for more permissions and find has lower total imprecision than the earlier methods.

***Keywords*—**Access control, privacy, k-anonymity, l-diversity

I. INTRODUCTION

The gathering of computerized data by governments, companies, and people has made gigantic open doors for learning and data based choice making. There is an interest for the trade and distribution of information among different gatherings. Associations, for example, hospitals need to release micro data for exploration and other open advantage purposes. In any case, sensitive personal information (e.g., medical condition of a specific person) may be uncovered in this procedure. Information in its unique structure, on the other hand, commonly contains touchy data about people, and distributed such information will damage singular security.

A substantial classification of protection assaults is to re-distinguish people by joining the distributed table with some outside tables demonstrating the foundation learning of clients. To fight this kind of assaults, the framework will propose.

The structure is a blend of access control and security assurance components. Access control mechanism shields delicate data from unapproved clients. The past work controls access to table yet doesn't give line level or cell-level security inside of table. We will give both Role-based and granular level access control component for social information [24]. The security saving module will anonymize the information utilizing encryption key component to meet protection necessities and imprecision requirements on predicates set by the entrance control system. The anonymization adds imprecision to the question results and the imprecision destined for every inquiry guarantees that the outcomes are inside of the resistance needed [25].

II. METHODOLOGY

A. *General Systen Architectur access Contorl model*

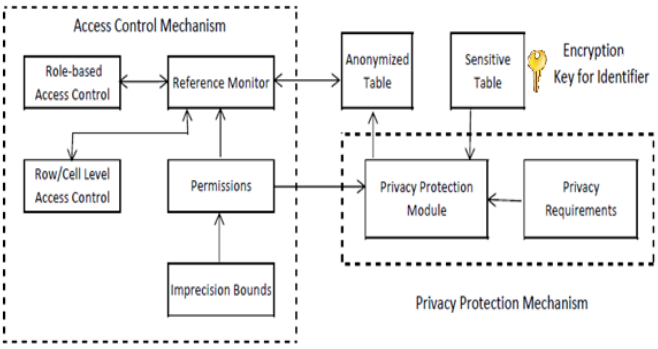

Figure: 1 Accuracy-constrained privacy-preserving access control model

Above figure 1 Illustrate accuracy-constrained privacy-preserving access control mechanism. Before the original data is available to the access control mechanism, the privacy protection mechanism ensures that goals like the privacy and accuracy are met. The permissions in the access control policy are based on selection predicates on the QI (Quasi Identifier) attributes. Permissions are defined by the policy administrator along with the imprecision bound to each permission or query, user-to-role and role-to permission assignments [18].

B. Access Control for relational data

Earlier Literature used Role-based Access Control. The proposed work provides both Role-based Access Control and granular level access control for relational data. Role-based Access Control (RBAC) defines permissions to the objects based on roles in an organization. When a user assigned to a particular role executes a query, the tuples satisfying the query predicate and the permission are returned. Users can be access to view information with grantee, but denied access to underlying tables. In proposed work, presenting both Role-based and granular level access control for relational data. Row level access control for relational data will allow tuple-level permissions [17]. Cell level access control for relational data will be implemented by replacing the unauthorized cells by NULL values using symmetric key encryption mechanism.

C. *Anonymization*

In Anonymization, an original table containing personal information will be transformed in encrypted form, so that it will difficult to an intruder for the determination of the identity of the individuals in that table. We will be encrypted the identifier attributes from the original table by using symmetric key encryption technique. Anonymization algorithms will provide suppression and generalization of records to satisfy privacy requirements with minimal distortion of data.

D. *Permission and imprecision bound*

The Access control administrator will define the permissions with the imprecision bound to each permission/query, user-to-role , and role-to-permission assignments. Imprecision bound tells that the authorized data has the desired level of accuracy. The difference between the number of tuples gained by a query executed on an anonymized relation (T*) and the number of tuples in same query on the original relation (T) is called as the Query Imprecision. The imprecision for query Qi is denoted by impQi, in this paper used to present query workload anonymization technique to minimize the imprecision for individual queries.

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

$$| Qi(T^*) | = \Sigma EC \text{ overlap } Qi \, | EC |$$

E. Privacy protection mechanism

Privacy protection model will provide k-anonymity, l-diversity and variance diversity. Paper used to propose the heuristic for partitioning the data to satisfy the privacy requirements.

III. LITERATURE SURVEY

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

Zahid Pervaiz, Walid G. Aref, Arif Ghafoor, Nagabhushana Prabhu [1] described approach is designed for static access control and relational data model. The access control policies define selection predicates available to roles while the privacy requirement is to complete the k-anonymity or l-diversity constraints. An additional constraint that satisfied by the PPM is the imprecision bound for each selection predicate.

B. Privacy-preserving data publishing: A survey of recent developments

B. Fung, K. Wang, R. Chen, and P. Yu [2] have suggested a promising approach to information sharing, while preserving individual privacy and protecting sensitive information. The general objective is to transform the original data into some anonymous form to prevent from concluding its record owners' sensitive information. Describes three types of linkage - record linkage, attribute linkage, and table linkage.

C. Implementing row and cell-level security in classified databases using sql server 2005

A.Rask, D. Rubin, and B. Neumann [5] explain SQL Server 2005 database to support row and cell-level security based on randomly security label scheme. Access restriction on rows and cells are enforced inside the database by using intrinsic structures like views and SQL Server data encryption.

E. Bertino and R. Sandhu survey the most relevant concepts underlying the notion of database security and summarize the most well-known techniques. Then focus on access control systems, on which a large body of research has been devoted, and explain the key access control models, namely, the discretionary and mandatory access control models, and the role-based access control (RBAC) model [4].

P. Samarati introduce the concept of minimal generalization that introduces the property of the release process which will not distort the data more than needed to achieve k-anonymity, and present an algorithm for the computation of such a generalization [6].

A. Machanavajjhala and et.al give a detailed analysis of two attacks, and propose a novel and powerful privacy criterion called ℓ-diversity that can defend against such attacks [7].

Access control mechanisms for databases allow queries to the authorized unit of the database [8], [10]. Predicate based fine-grained access control has further been proposed, where user authorization is limited for pre-defined predicates [11]. Enforcement of access control and privacy policies has been studied in [23]. However, the interaction between the access control mechanisms and the privacy protection mechanisms is not yet studied.

Recently, Chaudhuri et al. implemented access control with privacy mechanisms [12]. They use the definition of differential privacy where the random noise is added to original query results to satisfy privacy constraints. However, they have not thought about the accuracy constraints for permissions. We define the concept of privacy requirement with k-anonymity concept. It has been shown by Li et al. [13]

that after sampling, k-anonymity produces same privacy guarantees as those of differential privacy.

The authors illustrate by experiments that anonymized data which uses biased Rþ-tree related to the given query workload which is more accurate for those queries than to the unbiased algorithm.Ghinita et al. have proposed algorithms related to space filling curves for k-anonymity and l-diversity [14][18].

The existing literature on workload-aware anonymization has a focus for the minimization of overall imprecision for a given set of queries. The anonymization related to the imprecision constraints for individual queries has not been studied before. We follow the imprecision definition of LeFevre et al. [18] and introduce the constraint of imprecision bound for each query in a given query workload.

R. Sandhu and Q. Munawer proposes a novel policy administration mechanism, referred to as collaborative policy administration (CPA for short), to simplify the policy administration [22].

IV. PROBLEM FORMULATION

In this section, presenting the problem Statement, state our privacy goal, describing model and algorithm strategies used.

A. *Problem Statement*

An access control mechanism addresses the problem of developing accurate models about aggregated data without making any access to precise information of individual data record. A previously studied approach introduces random disruption to individual values to preserve privacy before data are published. Available previous solutions of this approach are limited in their tacit assumption on data miners. In this work, by using Access control mechanisms protect sensitive information from unauthorized users.

V. SOLUTION TO THE PROBLEM STATEMENT

A. *K-Anonymization Algorithm[3][16]*

In this algorithm we used two techniques one is Generalization and second is Suppression. In generalization, the data will be decrypted as per System Algorithm. The generalization applied on some of the data. In Suppression, the whole data will be decrypted. It put the star instead of Data. The primary reason of using Generalization and Suppression is that the details of data or information could not be visualized and no-one can extract the information record [9][15].

The K-Anonymzation Algorithm is as follows:-

Input: An integer k, and a [k, 2k -1]- cover $\gamma = \{S_1, \ldots, S_t\}$ of $D = \{R_1 \ldots\ldots R_t\}$.

Output: A [k,2k -1]-clustering,γ^0 , of D.

1: Set = $\gamma^{0=}$ γ.
2: **while** γ^0 has intersecting subset **do**
3: let $S_j, S_t \in \gamma^0$ be such that $S_j \cap S_t \neq \emptyset$ and let $R \in S_j \cap S_t$.
4: **if** $|S_j| > k$ **then**
5: Set $S_j = S_j \setminus \{R\}$.
6: **else if** $|S_t| > k$
7: $= S_j \setminus \{R\}$
8: **else** $\{|S_t| = |S_t| = k\}$
9: **Remove** S_t and S_j from γ^0 and replace them with $S_j \cup S_t$.
10: **end if**
11: **end while**

Input: Table D, integer k.

$C=F_{[k,2k-1]} = \{S \quad D : k < |S| < 2k - 1\}$

Output: Table g(D) that satisfies k-anonymity.

1: Invoke algorithm 1 with $C = F_{[k,2k-1]}$
2: Convert the resulting [k , 2k -1] – cover γ into a [k , 2k -1]-clustering, γ^0 , by invoking algorithm **A**
3: output the k-anonymization g (D) of D that responds to γ^0

B. *L-Diversity Algorithm Flow[3][16]*

In L-diversity, we also applied the rule of generalization and suppression but the most important fact in L-diversity is Quasi Identifier. Quasi Identifier means (QI) it will combine two attributes and applied the algorithm on new record. [15]

Input: A clustering $\gamma = \{C_1, \ldots\ldots..C_t\}$ of the records in the table D: a targeted diversity parameter l 1.

Output: A coarser clustering that respects l-diversity.

1: compute div(C_i) for all $C_i \in \gamma$.
2: let C_m be the cluster with minimal diversity in γ .
3: **if** div (C_m)$\geq$ l **then**
4: **output** γ and stop.
5: **end if**
6: Compute cost (C_i,C_m) for all $C_i \in \gamma \setminus \{C_m\}$.
7: Find the cluster $C_i \in \gamma \setminus \{C_m\}$ for which cost (C_i,C_m) is minimal.
8: Remove C_i and C_m from γ and add to γ the cluster $C_i \cup C_m$.
9: Go to step 2.

C. *The Response Time for K-anonymity and l-diversity Technique:*

Here used different experiment dataset for patients executed the Java program and calculated the response for both K-Anonymity and L-Diversity. Selected different random Dataset 10, 50, 100, 250, 500, 750, 1000, 2000, 5000 and 10000 records. The result of the time response with seconds in the following table and chart represented and measured.

Table 1: Response Time between two techniques

Number of Records	K-Anonymity	L-Diversity
~10	0.63 s	0.52 s
~50	0.81 s	0.74 s
~100	0.93 s	0.71 s
~250	1.75 s	1.6 s

~500	5.0 s	3.0 s
~750	7.0 s	7.0 s
~1000	9.0 s	10 s
~2000	35 s	40 s
~5000	80 s	85 s
~10000	168 s	177 s

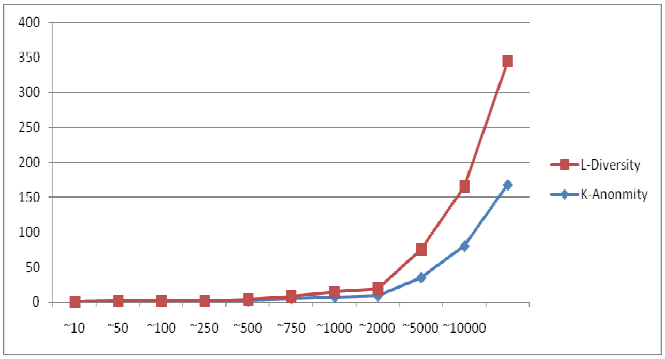

Figure: 2 K-Anonymity & L-Diversity Response Time Chart

D. Compare the efficiency of implementation for K-Anonymity & L-Diversity after modify the datasets:

Here represented and measured the number of dataset which are modified after implementing these K-Anonymity and L-Diversity techniques [19][21].

Table 2: Compare between numbers of modify datasets

Number of Records	K-Anonymity	L-Diversity
10	4	6
50	25	30
100	60	75
250	180	210
500	370	300
750	510	630
1000	760	860
2000	1672	1748
5000	3800	4590
1000	6400	7900

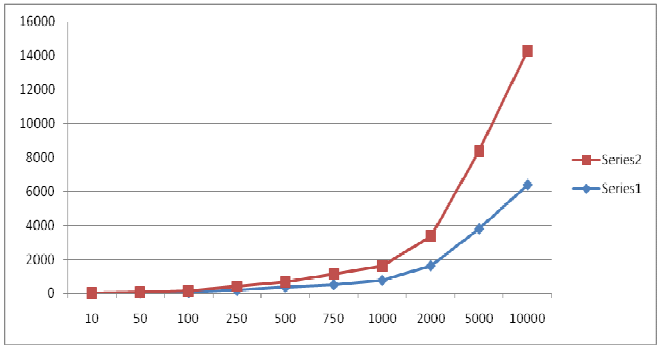

Figure: 3 K-Anonymity & L-Diversity modify datasets

VI. RESULT ANALYSIS

Table 3: Compare between expected and generated outcome

Sr.No	Expected Output	System Output	Percentage Output
Dataset1	100	88.75	88.75
Dataset2	93	77.4	86
Dataset3	98	85.75	87.5
Dataset4	86	68.58	79.75
Dataset5	83	77.19	93
Dataset6	97	87.59	90.3
Dataset7	85	70.97	83.5
Dataset8	84	68.63	81.7
Dataset9	87	73.34	84.3
Dataset10	89	86.51	97.2

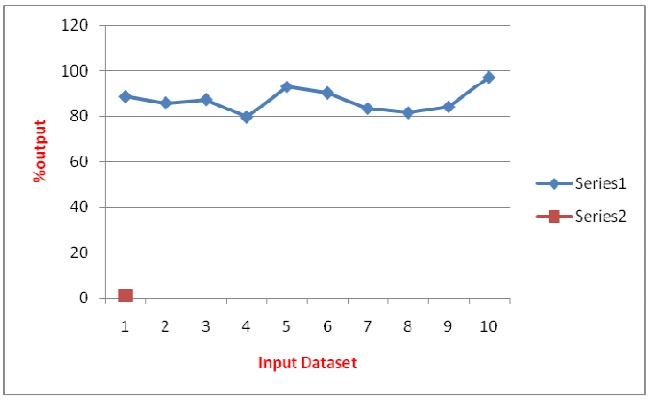

Figure 4: Output graph

ACKNOWEDGEMENT

I would like to thank my guide Prof. Vinayak Pottigar and co-guide Prof. Swapnaja Ubale for their valuable contribution in completing my work. I would also express thank to my family for moral support.

CONCLUSION

An accuracy-constrained privacy-preserving access control model for relational data has been proposed with role based access rights. The model is a combination of access control and privacy protection mechanisms. The access control mechanism permits only authorized query predicates on sensitive data and also provides generalization and suppression over the data. Anonymization offers more privacy options rather to other privacy preservation techniques (Randomization, Encryption, and Sanitization). The anonymization itself contains several techniques that require concluding best one and implement imprecision constraints on predicates rule by the access control mechanism. Here explained and compared between different types of Anonymization. After that, implemented two algorithms K-Anonymity and L-Diversity with .net

programs and used different datasets like patient's records. Earlier work, static access control and relational data model has been assumed but here actual relational data is used to experiment the result. We plan to extend the work privacy-preserving access control to incremental data and attribute based access control over the relational data.

REFERENCES

[1] Zahid Pervaiz, Walid G. Aref, Arif Ghafoor, Fellow, Nagabhushana Prabhu "Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data" ,IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, **APRIL 2014**

[2] 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.**

[3] Abdullah Abdulrhman AlShwaier, Dr,Ahmed Zayed Emam "A Novel Approach for DATA PRIVACY on E-HEALTH CARE SYSTEM ", International Journal of Engineering, Business and Enterprise Applications (IJEBEA)

[4] 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.**

[5] 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**.

[6] P. Samarati, "Protecting Respondents' Identities in Microdata Release" , IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov. **2001**.

[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] K. Browder and M. Davidson, "The Virtual Private Database in oracle9ir2" , Oracle TechnicalWhite Paper, vol. 500, **2002**.

[9] 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**.

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

[12] 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**.

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

[14] G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "Fast Data Anonymization with Low Information Loss" , Proc. 33rd Int'l Conf. Very Large Data Bases, pp. **758-769**, **2007**.

[15] 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**.

[16] Kenig B, and Tassa T, A Practical Approximation Algorithm for Optimal k-Anonymity, Division of Computer Science, The Open University, Raanana, Israel. Cited at: http://www.openu.ac.il/Personal_sites/tamirtassa/Download/Journals/optimal_k_anon.pdf

[17] 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**.

[18] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity", Proc. 22nd Int'l Conf. Data Eng., pp. **25-25**, **2006**.

[19] J. Friedman, J. Bentley, and R. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time", ACM Trans. Mathematical Software, vol. 3, no. 3, pp. **209-226**, **1977**.

[20] A. Meyerson and R. Williams, "On The Complexity of Optimal k-Anonymity", Proc. 23rd ACM SIGMOD-SIGACT-SIGART Symp.Principles of Database Systems, pp. **223-228**, **2004**.

[21] 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**.

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

[23] 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**.

[24] T.Monika**,** S.JayaPrakash, "Application specific Anonymization and Privacy – Preserving Access Control Mechanism for Relational data", National Conference on Research Advances in Communication, Computation, Electrical Science and Structures(NCRACCESS-2015) pp. 16-22, ISSN: 2348-8387,2015 at

[25] T.M.Arun Prabu, C.Anuradha, "Privacy preserving access control mechanism for electronic mail", International Journal of Computer Sciences and Engineering and scientific technology, March 2015.

AUTHORS PROFILE

Barkha Kasab is working toward the ME degree in the SKN Sinhgad College of Engineering, Korti-Pandharpur, Solapur University. Her research interests include access control, data security and preservation.