ISSN:0975-9646

Barkha Kasab et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4350-4352

Enhancing the Privacy and Implementation of Access Control Mechanism by Anonymization

Barkha Kasab[#1], Prof. Vinayak Pottigar[*2], Prof. Swapnaja Ubale[#3]

[#]*Computer Science And Engineering Department, Solapur University,Solapur,India*

[*]*SKN Sinhgad College of Engineering,Korti,Pandharpur,India*

***Abstract*— For research purposes, it is observed to analyze the data that provide better services to users for further processing. There are various privacy measures like k-anonymity, l-diversity, t-closeness to protect the individual's data. So anonymization concept is introduced that implement access control mechanism. However, privacy is achieved at the cost of precision of authorized information. Access control mechanism for efficient anonymization of micro-data to preserve the privacy is done. The access control policies are given to roles while the privacy requirement is used to satisfy the k-anonymity or l-diversity that defines selection predicates.**

***Keywords*— anonymization, privacy preservation, generalization, suppression**

I. INTRODUCTION

Huge amount of sensitive information has been collected by governments, corporations, and individuals. Typically, that information is stored in a table, and each record is related to particular individual. Each record has a number of attributes. Organizations, such as hospitals, need to release micro-data for different purposes. However, sensitive personal information may be revealed in this process, which is very risky. So use privacy preserving techniques which reduce possibility of identifying sensitive information about individual.

In RBAC, the account access is restricted to only authorized users. For example, in a hospital system, the role of doctor is to check the medical condition such as prescribing, to perform certain tests, and maintain a record, about that person. As like that the role of researcher includes analysing and interpreting the information for study purpose about patients which is anonymous.

Anonymity Definition [1]:

Let us consider the definitions based on anonymity which gives role-based access control and preserve the privacy.

There are different attributes such as:

Identifier: Attributes, e.g., name, address, mobile no. that can uniquely identify an individual. These attribute are removed from publishing records before releasing it.

Quasi-identifier (QI): The set of attributes, e.g., age, gender, zip code, birth date, linked with external attributes to reidentify. To satisfy the anonymity requirements generalization method is applied on QI attributes.

Sensitive attribute: Attributes, e.g., disease or salary that is assumed to be not revealed which is associated to unique individual and hence protected.

II. LITERATURE REVIEW

A model is designed for static access control and relational data [1]. The access control policies define selection predicates defined to different roles. Hence the privacy requirement is to satisfy the k-anonymity or l-diversity. An additional constraint that satisfied by the PPM is the imprecision bound for each selection predicate.

B. Fung et.al implement the privacy preserving model[5]. That method transforms the original data into some anonymous form to prevent from unauthorized users This survey describes three types of linkage - record linkage, attribute linkage, and table linkage.

A. Rask et.al produces an approach that enables a SQL Server 2005 database to support row- and cell-level security based on an arbitrary security label scheme [3]. Access restriction on rows and cells is enforced inside the database by using intrinsic structures.

The privacy requirement in terms of k-anonymity has been shown by Li et.al [4] that after sampling.From an access control user perspective, the permissions based on selection predicates have different accuracy requirements that need to be satisfied by the privacy protection mechanism. The

proposed privacy-aware access control framework allows the access control mechanism which satisfies the imprecision constraints.

Space filling curves for k-anonymity and l-diversity are implemented [6]. Ghinita et al. also introduce the problem of accuracy- constrained anonymization for information loss for each equivalence class [7]. Similarly, Xiao et.al [8] proposes to add noise to queries based on size of the queries in a given workload to meet differential privacy. Here, query imprecision bounds for are not considered. To minimize the imprecision for a given set of queries the workload-aware anonymization technique is proposed.

III. METHODOLOGY

There are two types of privacy preserving methods – identity disclosure and attribute disclosure. To protect the identity disclosure the k-anonymity property is introduced. But attribute disclosure attack is not satisfied by k-anonymity alone. So to overcome the limitations of k-anonymity, l-diversity method is introduced. To overcome both identity disclosure and attribute disclosure attack a new technique called as t-closeness is introduced.

Also the techniques like, generalization and suppression are used to preserve the privacy in a new way.

Generalization: Generalization basically related to group of people or thing which is converted to less specific value. For ex, in Table I, which contains original micro-data, have SEX values as "Male" and "Female" which are generalized to "Any". Generalization techniques can be applied to either attribute or cell.

Suppression: In Suppression we simply remove the sensitive data to preserve the privacy. Suppression technique can be applied at the level of single cell, entire tuple, or entire column. This method allows to reduce the amount of generalization to satisfy k-anonymity property.

Let us consider the relation T= A1, A2, ..An , where Ai is an attribute, T* is the anonymized version of the relation T. Initially, apply Suppression techniques to selected quasi identifier Qi and perform generalization. After that the anonymized table T is generated. After applying suppression technique, the records in the table T are sorted and arranged in n groups G1, G2, G3…Gn. Then each group is ordered by suppressed value of Quasi identifier attribute Bi (i=1, 2, … m). Select the Quasi identifiers Qi, from dataset with more distinct values. From group Gi, calculate the next nearest least integer value Li, and next nearest most integer value Mi. Then the value of attribute is now read as range value Li <= Mi. Repeat this process until all the Qi values in each group Gi are suppressed.

Modules Involved:

A. Access Control for relational data

In Role-based Access Control (RBAC) permissions are defined on objects according to their roles in any organization. As per the assignment of roles to the user, it executes a query. The tuples satisfying the query predicate and the permission are returned. A view is presented that allows a predefined query to a user or application as like a table. Also, users are allowed to gain the access for the view. The implementation of Cell level access control for relational data is done by symmetric key encryption mechanism which replaces the unauthorized cell values by NULL values.

B. *Anonymization*

All personal information is stored in the original table get transformed so that it is hidden from unauthorized users to determine the identity of the individuals in that table. The identifier attributes are encrypted by using symmetric key encryption. Also use suppression and generalization methods to satisfy k-anonymity.

C. Permission and imprecision Bound

The Access control administrator defines the permissions along with the imprecision bound for each permission/query, user-to-role assignments, and role-to-permission assignments. The specification of the imprecision bound ensures that the authorized data has the desired level of accuracy.

D. Privacy protection mechanism

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 k-anonymity, l-diversity and variance

diversity. Investigation of privacy-preservation from the anonymity aspect is here done. The heuristics proposed for accuracy-constrained privacy-preserving access control.

IV. RESULT AND DISCUSSION

Table I shows the original micro-data before anonymization. After applying the actual generation and suppression method the k-anonymous table II is generated. Consider Age, Zip, sex, phone are quasi identifiers and disease as sensitive attribute. The selection of quasi-identifiers and sensitive attributes are depend on the organization according to their rules and regulations.

TABLE I
ORIGINAL MICRO-DATA BEFORE ANONYMIZATION

P_Id	P_age	P_zip	P_sex	P_phone	P_disease
1	21	413123	M	8856321244	fever
2	24	413147	F	8856785633	Heart attack
3	46	411452	M	7816390111	jaundice
4	38	412789	M	9830905634	cough

Here apply Suppression techniques to selected quasi identifier Qi= {AGE, ZIP, SEX, PHONE} and perform generalization.

TABLE II
TABLE AFTER ANONYMIZATION

P_Id	P_age	P_zip	P_sex	P_phone	P_disease
*	20-50	413***	ANY	885*******	fever
*	20-50	413***	ANY	885*******	Heart attack
*	20-50	411***	ANY	781*******	jaundice
*	20-50	412***	ANY	983*******	cough

V. CONCLUSIONS

Access control mechanism to relational data for privacy preservation using role based approach has been proposed which show empirically that the proposed approach satisfies imprecision bounds for more permission. Only authorized user now allowed to use the sensitive information. The privacy preserving module anonymizes the data to meet privacy requirements. Anonymization is done such a way that the researchers can discover useful knowledge from data without breaching the privacy of the individuals. So anonymization takes place depending on the category of attribute values.

ACKNOWLEDGMENT

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.

REFERENCES

[1] Zahid Pervaiz et.al , ‘‘Accuracy Constrained Privacy-Preserving Access Control Mechanism for Relational Data,” IEEE Transactions on Knowledge and Data Engineering, VOL.26, No.4 APRIL 2014.

[2] P. Samarati, ‘‘Protecting Respondents Identities in Micro-data Release,” IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6 pp. 1010-1027, Nov. 2001.

[3] 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.

[4] N. Li, W. Qardaji, and D. Su, Provably private data anonymization: Or,k-anonymity meets di_erential privacy, Arxiv preprint arXiv:1101.2604, 2011.

[5] B. Fung et al, ‘‘Privacy-Preserving Data Publishing: A Survey of Recent Developments, ACM Computing Surveys,” vol. 42, no. 4, article 14, 2010.

[6] G. Ghinita et.al and N. Mamoulis, ‘‘Fast data anonymization with low information loss, in Proceedings of the 33rd international conference on Very large data bases,” pp. 758769, VLDB Endowment, 2007.

[7] G. Ghinita et.al , ‘‘A framework for efficient data anonymization under privacy and accuracy constraints,” ACM Transactions on Database Systems (TODS), vol. 34, no. 2, p. 9, 2009.

[8] X. Xiao et.al, Ireduct: ‘‘Differential privacy with reduced relative errors, in Proceedings of the ACM SIGMOD International Conference on Management of Data, “2011.

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

[10] S. Rizvi et.al, ‘‘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