International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

Access Control Mechanism for Authorized Query Predicates on Sensitive data

D. Venkata Ashok [1]& Mr. D. Durga Prasad[2]
[1]PG Scholar,Dept of MCA ,MIC College of Technology, Krishna Dist, A.P. India
[2] Associate Professor,Dept of MCA, MIC College of Technology, Krishna Dist, A.P. India

Abstract:
Data privacy issues are increasingly becoming important for many applications. Protective individual privacy is a crucial downside. However, sensitive data will still be ill-used by approved users to compromise the privacy of shoppers. Traditionally, research in the database community in the area of data security can be broadly classified into access control research and data privacy research. Access Control Mechanisms (ACM) is used to ensure that only authorized information is available to users. Privacy Protection Mechanism (PPM) uses suppression and generalization of relational data to anonymize and satisfy privacy needs. Recent research studied the problem of publishing data in databases without revealing the sensitive information, moving to the privacy preserving paradigms of k-anonymity and L-diversity. K-anonymity protects against the identity of an individual's record. L-diversity, in addition to this, safeguards against the association of an individual with specific sensitive information. The aim of this paper is to provide better security and minimum level of precision to the obtained 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. The accuracy constraints are satisfied for multiple roles. We propose heuristics for anonymization algorithms to show empirically that the proposed approach satisfies imprecision bounds for more permissions and has lower total imprecision than the current state of the art.

Keywords: K-Anonymity; L-Diversity; Suppression; Generalization; Privacy

I.INTRODUCTION

COMPANIES gather and resolve user information to develope their quality services. Access Control Mechanisms (ACM) is used to protect the sensitive information from the unauthorized users. However, sensitive information can still be used improperly by correct users to compromise the privacy of users. The theme of privacy protection for important data can require the compulsion of privacy rules. In this paper, we inquire privacy preservation from the anonymity view. The important information, even after the deleting of sensitive attributes, is still infectible to linking attacks by the correct users [2]. This difficulty has been studied widely in the area of micro data publishing [3] and privacy definitions, e.g., k-anonymity [2], l-diversity [4], and variance diversity [5]. Anonymization algorithms use suppression and generalization methods to satisfy privacy necessity with minimal problem of micro data. The anonymity techniques can be used with an ACM to ensure both guarantee and protection of the important data. The protection is done at the cost of exactness and imprecision is introduced in the authorized data under an access control rule. We use the method of imprecision bound for each prohibition to define a threshold on the amount of imprecision that can be tolerated. Existing workload aware anonymization techniques [5], [6] reduce the imprecision aggregate for all questions and the imprecision added to each question in the anonymized micro data is not known.
Developing the protection needs more stringent returns in extra imprecision for questions. However, the solution of fulfill accuracy

constraints for individual prohibition in a workload has not been known before. The methods proposed in this paper for efficient access control mechanism are also related in the background of workload-aware anonymization. The anonymization for related data publishing has been studied in other papers [3]. In this paper the aim is on a fixed relational table that is anonymized one time only. To show our approach, role-based access control is considered. However, the concept of accuracy constraints for prohibition can be addressed to any security rules, e.g., discretionary access control. The impact of this paper is as follows. First, we develope the guaranteed and privacy restrictions as the difficulties of k-anonymous Partitioning with Imprecision Bounds (k-PIB) and give hardness outcome. Second, we provide introduction the theme of efficient access control mechanism for relational database.
Third, we propose methods to the solution of the k-PIB problem.

II. RELATED WORK

Related work deals with the previous work related to this paper. The existing methods only deals with either access control mechanism, or privacy protection mechanism. There was no such a study related to the hybrid of both access control mechanism for relational data. Here it deals with the various methods used for the access control mechanism and privacy protection mechanism. In the case of privacy protection, the main method is k-anonymity method; k-anonymity has recently been investigated as an interesting approach to protect sensitive data undergoing public or semi-public release from linking attacks. To protect respondents' identity when releasing microdata, data holders often remove or encrypt explicit identifiers, such as names and social security numbers. De-identifying data, however, provide no guarantee of anonymity. Released information often contains other data, such as race, birth date, sex, and ZIP code that can be linked to publicly available information to re-identify respondents and to infer information that was not intended for release. One of the emerging concepts in microdata protection is k-anonymity, which has been recently proposed as a property that captures the protection of a microdata table with respect to possible re-identification of the respondents to which the data refer. In the k-anonymity method there used two operations, suppression and generalization. The suppression technique the sensitive information is replaced by special characters like asterisk „*". The generalization method will replace the sensitive information with broader range.

Table-1: Sensitive Table

ID	Age	Zip	Disease
1	8	15	Fever
2	14	23	Flu
3	26	27	Flu
4	35	36	Cold
5	27	28	Fever
6	12	19	cold
7	22	30	Diarrhea
8	27	17	Cold

Table-2: Anonymized Table

Age	Zip	Disease
0-20	10-30	Fever
0-20	10-30	Flu
20-30	10-30	Flu
20-40	20-40	Cold
20-40	20-40	Fever
0-20	10-30	cold
20-30	20-40	Diarrhea
20-40	10-30	Cold

The other major method for anonymization is the l-diversity method. L-diversity method reduces the

granularity of representation of the data. In this section, it derives the principle of l-diversity in two ways. First, it will derive the data from the table and make sure that there will not occupies any privacy breach. Then it will re-derive the l-diversity principle from a more practical starting point and show that even under less than ideal circumstances, l-diversity can still defend against background knowledge that is unknown to the data publisher. The l-diversity method is an extension of the k-anonymity method. In the l-diversity method the first it uses the generalization or suppression method for the anonymization. The l-diversity model uses intra-group diversity for sensitive values in the anonymization process if the sensitive values show the homogeneity nature. The l-diversity is more efficient than the k-anonymity method. It avoids the attacks like background knowledge attack and others in k-anonymity method.

Top down Selection Mondrian (TDSM) algorithm is proposed by LeFevre et al. The TDSM algorithm is a greedy algorithm. The TDSM algorithm was developed to minimize the total imprecision for all queries. In this method the imprecision bound of the queries are not considered. The Top Down Selection Mondrian algorithm begins as the complete tuple as one partition and then all the partitions are recursively divided until the time new partitions meet the privacy requirement. Two decisions need to be made for the division of the partitions, i) Choosing a split value along each dimension, and ii) Choosing a dimension along which to split. The split value is chosen along the median and then the dimension is selected along which the sum of imprecision for all queries is minimum in the case of the TDSM algorithm

The disadvantages of the existing systems are:
1) There is no privacy for users
2) There is a chance of the linking attacks even after the removal of identifying attributes from the sensitive data.

III.Proposed System:

The algorithms which are used in this project are

1) Top-Down Heuristic 1 (TDH1)
2) Top-Down Heuristic 2 (TDH2)
3) Top-Down Heuristic 3 (TDH3)

3.1Top-Down Heuristic 1 (TDH1)

In TDSM, 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 as illustrated in Fig. 1. 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.

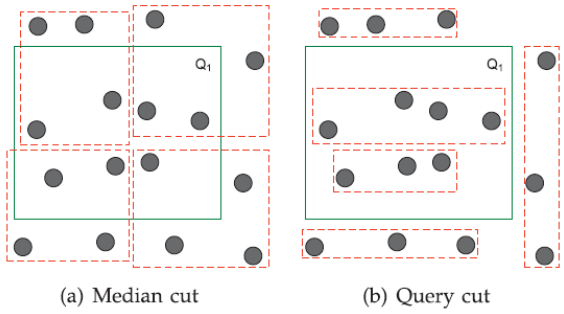

Fig.1 Comparison of median and query cut

3.2 Top-Down Heuristic 2 (TDH2)

In the Top-Down Heuristic 2 algorithm (TDH2, for short), the query bounds are updated as the partitions are added to the output. This update is carried out by subtracting the icQj Pi value from

the imprecision bound BQj of each query, for a Partition, say Pi, that is being added to the output. For example, if a partition of size k has imprecision 5 and 10 for Queries Q1 and Q2 with imprecision bound 100 and 200, then the bounds are changed to 95 and 190, respectively. The best results are achieved if the kd-tree traversal is depth-first (preorder). Preorder traversal for the kd-tree ensures that a given partition is recursively split till the leaf node is reached. Then, the query bounds are updated. Initially, this approach favors queries with smaller bounds. As more partitions are added to the output, all the queries are treated fairly. During the query bound update, if the imprecision bound for any query gets violated, then that query is put on low priority by replacing the query bound by the query size. The intuition behind this decision is that whatever future partition splits TDH2 makes, the query bound for this query cannot be satisfied. Hence, the focus should be on the remaining queries.

3.3 Top-Down Heuristic 3 (TDH3)

The time complexity of the TDH2 algorithm is OðdjQj2n2Þ, which is not scalable for large data sets (greater than 10 million tuples). In the Top-Down Heuristic 3 algorithm (TDH3, for short), we modify TDH2 so that the time complexity of Oðdj QjnlgnÞ 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 1:99for 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 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.

System Architecture

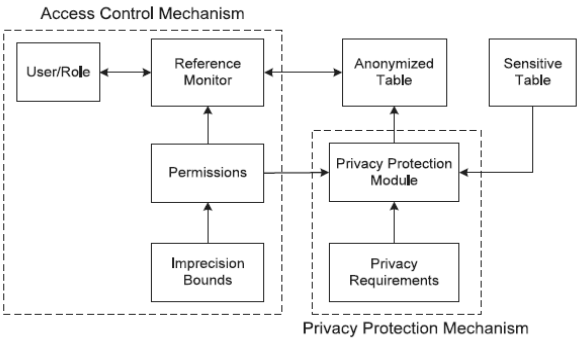

Fig .2 System Architecture

An accuracy-constrained privacy-preserving access control mechanism, illustrated in Fig.2 (arrows represent the direction of information flow), is proposed. The privacy protection mechanism ensures that the privacy and accuracy goals are met before the sensitive data is available to the access control mechanism. The permissions in the access control policy are based on selection predicates on the QI attributes. The policy 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. The imprecision bound information is not shared with the users because knowing the imprecision bound can result in violating the privacy requirement. The privacy protection mechanism is required to meet the privacy requirement along with the imprecision bound for each permission.

Anonymization is the process of making the data anonymized, i.e., the sensitive data is made privacy protected. In this proposed method it uses the k-anonymity method and the data fragmentation method for the privacy protection. The important term used here is the imprecision bound. By definition the query imprecision bound is the total imprecision acceptable for a query predicate and is preset by the access control administrator. The access control mechanism is based on the imprecision bound. The anonymization method

uses the combination of the k-anonymity method and the fragmentation method

The approaches for preserving privacy divide into two categories in this paper. The first one is data encryption, and the second one is data fragmentation. The typical approach for data encryption is encrypts entire databases on the client side before store to the non-trusted third party provider, external database. Here it uses the k-anonymity method for the encryption. Through the process of suppression and generalization it anonymized the sensitive information. The second approach for preserving privacy is data fragmentation. For the better result use the clustering methods for the purpose of the fragmentation. The fragmentation is done as horizontal fragmentation and vertical fragmentation. The horizontal fragmentation is done through the rows of the relation and vertical fragmentation is done through the columns of the relation. This provides better privacy for the relational data. The clustering method performs the fragmentation process efficiently.

IV.CONCLUSION

In secured relational data storage, it needs good access control mechanism and privacy preserving access control mechanism. In this paper a privacy-preserving access control framework for relational data has been proposed. The proposed framework is a combination of access control and privacy protection mechanisms. The access control mechanism allows only the authorized query predicates on sensitive data. The privacy preserving module anonymized and fragmented the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism. For the anonymization process proposed a k-anonymity method and for the fragmentation introduces the clustering analysis method. It formulates this interaction as the problem of k-anonymous Partitioning with Imprecision Bounds (k-PIB). It gives hardness results for the k-PIB problem. This paper presents a heuristics method for partitioning the data to satisfy the privacy constraints and the imprecision bounds. This proposed paper gives a secured access control mechanism and privacy protection mechanism for the relational data. In the current work, static access control and relational data model has been assumed. For future work, it plan to extend the proposed privacy-preserving access control to cell level access control and can use the l-diversity instead of k-anonymity method.

V. REFERENCES

[1]Zahid Pervaiz, Walid G. Aref, Arif Ghafoor and Nagabhushana Prabhu-IEEE transactions on knowledge and Data Engineering vol-26, No.4, April 2004.

[2]P.Samarati, “Protecting Respondents’ identitied in microdata release”, IEEE Trans. Knowledge and Data Eng.; vol 13, no 6,Nov. 2001.

[3]A.Machanavajjha,D.Kifer,J.Gehrke,andM.Venk itasubram aniam, “L-Diversity: Privacy Beyond k-anonymity,” ACM Trans.

[4]D.Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn and R. Chandramauli, “ Proposed NIST Standard forrole-base access control”, ACM Trans. Information and System Security.

[5]N. Li,W.Qadarji, and D.Su, “Provaby Private Data Anonymistaion: Or, k-anonymity meets Differential Privacy,” Arxiv preprint arXiv:1101.2064,2011.

[6]G.Ghinita, P.Karras, P.Kalnis, and N. Mamoulis, “Fast Data Anonymisation with Low Information Loss,” Proc. 33rd Int’l Conf. Very Large Databases,pp. 758-769,2007 .

[7]X.Xiao, G.Bener, M.Hay, and J. Gehrke, “Ireduct : Differential Privacy with Reduced Relative Errors,” Proc. ACM SIGMOD Int’l Conf.Management of Data,2011.

[8] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Workload-AwareAnonymization Techniques for Large-Scale Datasets," ACMTrans. Database Systems, vol. 33, no. 3, pp. 1-47, 2008.

[9]S.Chaudhari, T. Dutta and S. Sudarshan, " Fine Grained Authorisation through Predicted Grants," Proc. IEEE 23rd Int'l Conf Data Eng.,pp. 1174-1143,2007.

[10]T.Iwuchukwu nd J.Naughton, "K-Anonymisation as Spatial Indexing:Toward Scalable and Incremental Anonymisation", Proc. 33rdInt'l Conf.Very Large DataBases, pp.746-757,2007.

[11]G.Ghinita, P.Karras, P.Kalnis, and N. Mamoulis, A Framework for Efficient Data Anonymisation Under Privacy and Accuracy Constraints, "ACM Trans. Database Systems, vol. 34, no. 2, article 9, 2009