Privacy-Preserving Access Control Mechanism for Relational Data

Shijith T P
PG Student
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology, Palakkad, Kerala

Sreelakshmi R
Assistant Professor
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology, Palakkad, Kerala

AmbikadeviAmma T
Professor
Department of Computer Science & Engineering
Jawaharlal College of Engineering and Technology, Palakkad, Kerala

Abstract

The Access control mechanism avoids the unauthorized access of sensitive information. It protects the user information from the unauthorized access. The privacy protection mechanism is a much important concern in the case of sharing the sensitive information. The privacy protection mechanism provides better privacy for the sensitive information which is to be shared. The generally used privacy protection mechanism uses the generalization and suppression of the sensitive data. It prevents the privacy disclosure of the sensitive data. The privacy protection mechanism avoids the identity and attributes disclosure. The privacy is achieved by the high accuracy and consistency of the user information, ie., the precision of the user information. In this paper, it proposes a privacy persevered access control mechanism for relational data. It uses the accuracy constrained privacy preserving access control mechanism for relational data framework here. It uses the concept of imprecision bound for the privacy preserving access control mechanism. The imprecision bound is set for all queries. For the privacy protection mechanism it uses the combination of both the k-anonymity method and fragmentation method. The k-anonymity method uses the suppression and generalization. Here it uses the clustering method for fragmentation. It also proposes the heuristic method for anonymization process.

Keywords: Access control, Privacy, k-anonymity, Imprecision Bound

I. INTRODUCTION

Data mining is the process of extracting the useful information from the database. It is possible to efficiently extract or mine knowledge from large amounts of vertically partitioned data within quantifiable security restrictions. In other words the data mining is the process of discovering the interesting knowledge from large amounts of data stored either in databases, data warehouses or other information repositories. Knowledge Discovery in Databases (KDD) is the process of extracting knowledge from large quantities of data. The KDD process assumes that all the data is easily accessible through centralized access mechanisms such as federated databases and virtual warehouses. Moreover, advances in information technology and the ubiquity of networked computers have made personal information much more available. Privacy advocates have been challenging attempts to bring more and more information into integrated collections. Database security is the important requirements of the database. Database security is a very broad area that addresses many issues, like legal and ethical issues regarding the right to access certain information. Some information may be stored to be private and cannot be accessed legally by unauthorized persons. The sensitive data is accessible to authorized users only. The database security provides the security for the sensitive information from the unauthorized access. The database security is based on the access control mechanism and the privacy protection mechanism.

The Access Control Mechanisms (ACM) is used to ensure that only authorized information is available to users The sensitive information which is the user is not authorized to access will not accessed by the user. The authorized user can only access the authorized data. In a multiuser database system, the Database Management System (DBMS) must provide techniques to enable certain users or user groups to access selected portions of a database without gaining access to the rest of the database. Its importance comes in a large organization where numerous workers working. There must be some important data which are not published to all the workers. There uses the access controls mechanism for providing the access to the secured data to the particular authorized user only. For example, sensitive information such as employee salaries or performance reviews should be kept confidential from most of the database system's users. A DBMS typically includes a database security and authorization subsystem that is responsible for ensuring the security of portions of a database against unauthorized access. Privacy is the one of the most important concern of human life. It gives more importance to protect the privacy of the personal life. In the case of

database, there will be huge amount of data to be kept privately. These data may contain sensitive information about the persons, confidential information about some organizations and so on. These data has to be protected by using some methods. It is the privacy protection mechanism (PPM). The general method is to transform the original data into some anonymous form to prevent from accessing its record owners' sensitive information. There are numerous methods to provide the privacy for the sensitive data. The anonymization method is one of the important privacy protection mechanisms. The anonymization process will transforms the sensitive information to some anonymized form. K-anonymity, l-diversity, etc., are some of the anonymization method. For a given query from unauthorized user, it will provide the anonymized data through the privacy preserving techniques.

In this paper it deals about the privacy protected access control mechanism. It will provide the security for the sensitive information. For an example, in the case of hospital management system there should be a number of patients. Some of the patients may have the disease which has to be isolated and so on. While publishing the patients' data to the state medical board for disease surveillance system, they should anonymize the personal data of the patient. For this purpose it can use the proposed method for the secured access control and privacy protection mechanism.

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. METHODOLOGY

There are lots of methods for providing the privacy for the sensitive information stored in the database and there are different access control methods for accessing the secured information stored in a database. In my project it deals with the introduction of both the access control mechanism and the privacy protection mechanism together for protecting the sensitive information. Here it uses the anonymity method and fragmentation method for the privacy protection and the imprecision bound for both the access control and the privacy protection method.

The proposed system uses secure reversible Accuracy-Constrained Privacy-Preserving Access Control for relational database. The proposed method provides data publication in a privacy preserved method. The framework of the proposed method is a combination of access control and privacy protection mechanisms. The access control mechanism allows only authorized queries predicates on sensitive data. The privacy preserving module anonymized the data to meet privacy requirements and imprecision constraints on predicates set by the access control mechanism. Access Control Mechanisms (ACM) is used to ensure that only authorized information is available to users. However, there is chance of sensitive information can still be misused by authorized users for their use. The confidential data can also be misused. The concept of privacy-preservation for sensitive data requires the enforcement of privacy of the secured sensitive data and privacy policies or the protection against identity disclosure by satisfying some privacy requirements. In the proposed method, it investigate privacy-preservation from the anonymity aspect. The sensitive information, even after the removal of identifying attributes, is still in danger to linking attacks by the authorized users. Here it uses the data fragmentation and the anonymization method for the purpose of the privacy protection mechanism. Anonymization algorithms use suppression and generalization of records to satisfy privacy requirements with minimal distortion of micro data. The fragmentation technique and anonymity technique can be used with an access control mechanism to ensure both security and privacy of the sensitive information. The privacy is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an access control policy. Here use the concept of imprecision bound. The imprecision bound is a threshold value which determines the amount of imprecision that can be tolerated for each query. Existing anonymization techniques minimize the imprecision aggregate for all queries. Then the imprecision added to each permission/query in the anonymized micro data is not known. Making the privacy requirement more stringent results in additional imprecision for queries. Here proposed a heuristic algorithm for the partitioning process. The partitioning of data occurs according to the query cut. The proposed method is mainly focus on the static relational table which can anonymize only once. To represent this, assume the role-based access control mechanism. However, the concept of accuracy constraints for permissions can be applied to any privacy-preserving security policy.In the privacy protection mechanism it uses the concepts of both data fragmentation and encryption. In this proposed method it uses the k-anonymity method as the encryption method and clustering for the fragmentation process.

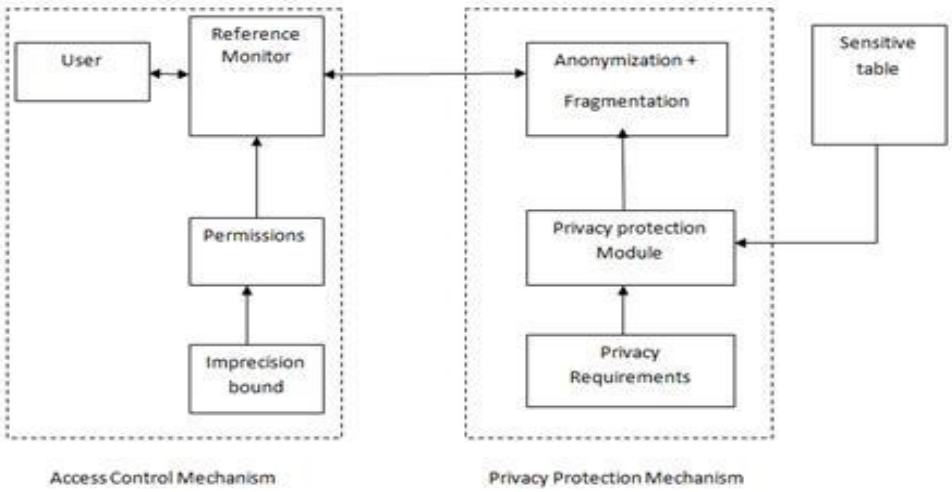

Fig. 1: Privacy Preserved Access Control for Relational Data

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.

In this proposed method accuracy constrained access control method is proposed. The permissions in the access control policy are based on the Quasi-Identifier (QI) attributes. QI attributes are attributes that can potentially identify an individual based on other information available to an adversary. QI attributes are generalized to satisfy the anonymity requirements. Gender, zip code, date of birth etc are quasi identifiers. 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 s 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.

After the anonymization process the exact table value will get replaced by the anonymized values. The values are the generalized values. The access control enforcement over the generalized data contains the relaxed enforcement and strict enforcement. Relaxed enforcement uses overlap semantics to allow access to all partitions that are overlapping the permission and strict enforcement uses enclosed semantics to allow access to only those partitions that are fully enclosed by the permission. Both approaches have advantages and disadvantages. Relaxed enforcement violates the authorization predicate by giving access to extra tuples but is beneficial for applications where low cost of a false alarm is tolerable as compared to the risk associated with a missed event. On the other hand, strict enforcement is suitable for applications where a high risk is associated with a false alarm as compared to the cost of a missed event. In this proposed method it focuses on relaxed enforcement. It assume that under relaxed enforcement if the imprecision bound is violated for a permission then that permission is not assigned to any role.

In this section, the access control enforcement method is analyzed probabilistically. Here it uses the relaxed enforcement method, so the analyzing occurs on the relaxed enforcement method. Here the main process is that the access control policy administrator will set the imprecision bound for each query. The access control policy administrator requires that the imprecision bound for the least number of queries be violated by Privacy Protection Mechanism. The policy administrator has to revise the imprecision bound for the queries. It will deny the queries which highly violate the bounds.

Here proposed top-down heuristics for multi- dimensional partitioning to satisfy imprecision bounds. The top down heuristic algorithm is introduced for getting better result than the existing top down selection Mondrian algorithm. 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 TDSM, the partitions are splits along the median, so it uses the median cut.. In the heuristic method, the splitting of the partition is performed along the query cut. By using the query cut method it will choose the dimension where, the imprecision should be minimum for all the queries. If there present multiple queries that overlaps a partition, then it has to select the query for the query cut. For this purpose the queries having the imprecision greater than zero will taken and sorted. The queries having the imprecision bound is small will taken because the queries with less imprecision bound. If the queries doesn't allows the query cut then it will uses the median cut. In the heuristic method the term feasible cut is very important. The feasible cut is the partition resulting from the split should satisfies the privacy requirements.

Top down Heuristic Algorithm

1) Step 1. Initialize the set of candidate partition.
2) Step 2. Sort the queries overlapping the candidate partition with imprecision greater than zero.
3) Step 3. Select the queries with least imprecision bound.
4) Step 4. Checks for the feasible split of the partition along the query interval.
5) Step 5. If a feasible cut is found, then the resulting partitions are added to the candidate partition.
6) Step 6. If feasible cut is not found, then the candidate partition is checked for the median cut.

The heuristic algorithm will helps to provide the secured access control mechanism. The imprecision bound is set by the administrator. The imprecision bound is not known to the user. So it provides the secured access control method.

IV. RESULT AND DISCUSSION

The proposed system that combines the idea of secured access control mechanism and privacy protection mechanism for the relational data. The heuristic algorithm used here will improve the efficiency of the access control mechanism. The combination of the anonymization and fragmentation used here it has improved the privacy of the sensitive information in the relational data.

V. 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.

REFERENCES

[1] ZahidPervaiz, Walid G. Aref, ArifGhafoor, and NagabhushanaPrabhu, Accuracy-Constrained Privacy-Preserving Access Control Mechanismfor Relational Data", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, NO. 4, April 2014.
[2] Ms. S.Kokila, Dr. T. SenthilPrakash, and Ms. P.Maheswari, " Privacy and Security Ensured Database Rights Management Scheme", International Journal On engineering Technology and Sciences – IJETS™ ISSN (P): 2349-3968, ISSN (O): 2349-3976 Volume 1 Issue 6, October 2014.
[3] Yung-Wang Lin, Li-Cheng Yang, Luon-Chang Lin, and Yeong-Chin Chen, "Preserving Privacy in Outsourced Database", International Journal of Computer and Communication Engineering, Vol. 3, No. 5, September 2014.
[4] T.Sujitha, V.Saravanakumar, C.Saravanabhavan, "An Efficient Cryptographic approach For Preserving Privacy In Data Mining", International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013.
[5] ZahidPervaiz, ArifGhafoor, and Walid G. Aref , "Precision bounded access control for privacy preserving data stream", CERIAS Tech Report 2013-7.
[6] Alaa H Al-Hamami, and Suhad Abu Shehab, "An Approach for Preserving Privacy and Knowledge In Data Mining Applications", Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, No. 1 Jan 2013.
[7] N.Punitha, R.Amsaveni, "Methods and Techniques to Protect the Privacy Information in Privacy Preservation Data Mining" IJCTA | NOV-DEC 2011.
[8] 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.
[9] Gabriel Ghinita, PanosKalnis and Yufei Tao," Anonymous Publication of Sensitive Transactional Data", IEEE Transactions on Knowledge and Data Engineering, vol. 23, Issue.2,pp.161-174,2011.
[10] N. Li, W. Qardaji, and D. Su, "Provably Private Data Anonymization: Or, k-Anonymity Meets Differential Privacy," Arxiv preprint arXiv:1101.2604, 2011.
[11] Shahidul Islam Khan, Dr. A. S. M. LatifulHoque, "A New Technique for Database Fragmentation in Distributed Systems", International Journal of Computer Applications (0975 – 8887) Volume 5– No.9, August 2010.
[12] Xin Jin, Nan Zhang, Gautam Das, "Algorithm-Safe Privacy-Preserving Data Publishing" EDBT 2010, March 22–26, 2010.
[13] 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.
[14] Gabriel Ghinita, PanagiotisKarras, And PanosKalnis," A Framework for Efficient Data Anonymization under Privacy and Accuracy Constraints", ACM Transactions on Database Systems, Vol. 34, No. 2, Article 9, Publication date: June 2009.
[15] LalanthikaVasudevan , S.E. DeepaSukanya, and N. Aarthi, "Privacy Preserving Data Mining Using Cryptographic Role Based Access", Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 VolI IMECS 2008, 19-21 March, 2008.