Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

A PRIVACY BASED PRECISION CONTROLLED STUDY ON LOG BASED ANALYSIS IN RELATIONAL DATA

P. KRANTHI PRIYA #1, K.V. SAMBASIVARAO #2
#1M.Tech Student, Dept. of CSE, NRI Institute of Technology, Agiripalli,A.P,India
#2 Dean, Dept of CSE, NRI Institute of Technology, Agiripalli, A.P, India

ABSTRACT:

Recently the amount of log and dimension data in the websites that needs to be processed and stored has exploded as the usage of site is increased. Therefore, the issue is suggested about the necessity of Enterprise Security Management (ESM) that is for integrated management of network system such as firewall, IPS, VPN, and etc. The precision control approaches characterize choice predicates precision to parts while the secrecy stabilization is to support the k-anonymity or l-diversity. A SSPPM it will fulfill the admittance control and local monitoring of data. Then again, security is accomplished at the premium of exactness of approved data. But, in our plan of the previously stated issue we didn't have key management of data, but in this we propose efficient results with authorized user and another hand original data sets will not be present for servers also. And best of our insight, the issue of fulfilling the exactness and requires the data maintains for various parts has not been considered some time recently. The procedures for workload-mindful anonymization for determination predicates have been examined inthe writing. Notwithstanding, when delicate data is shared and a Secrecy Stabilization Picket Picket Mechanism (SSPPM) is not set up, an approved client can at present trade off the security of a man prompting with accurate data. In this paper, we propose a precision controlled security safeguarding admittance control structure. That Admittance control components shield delicate data from unapproved clients. This type of approaches are used for data manage on mining with efficient manner. These kind of results produce key for authorized once only.

Keywords:precision controlled data, secrecy stabilization

Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

I.INTRODUCTION:

Several organizations and agencies publish their sensitive micro data, e.g., medical data, customer data or census data for research and other public sensitive information. In an age where the sensitive micro data of each individual are recorded and stored their secrecy details, an inconsistency arises between the necessity to protect the picket of and also to use these data for medical research, trend analysis andimprovement. Hence, the private information of an individual should not be revealed from the micro data. Organizations collect and analyse consumer data to improve their services. Admittance Control Mechanism (ACM) is used to ensure that only authorized information is available to the users. However, sensitive information can still be misused by authorized users compromising the picket of consumers. The concept of picket preservation for sensitive data can require the enforcement of picket policies or the picket against identity disclosure by satisfying some picket requirements. The anonymity techniques can be used with an admittance control mechanism to ensure both security and picket of the sensitive information. The picket is achieved at the cost of accuracy and imprecision is introduced in the authorized information under an admittance control policy. An integrated framework of achieving both picket and security is proposed through the integration of Admittance Control Mechanism with Picket Preservation [1] Technique to prevent the authorized user from misusing the sensitive information. The enforcement of picket policies or the picket against identity disclosure satisfying some picket requirements are the prerequisites for picket preservation of sensitive data. Even after removal of identifying attributes, the sensitive information issusceptible to liking attacks by the authorized users. So the present investigation is proposed to study the area of micro data publishing and picket definitions such as k-anonymity[2], l-diversity[3] and variance diversity procedures can be utilized with an entrance control component to guarantee both security and picket of the delicate data. The security is accomplished at the premium of exactness and imprecision is presented in the approved data under an entrance control strategy. In existing framework the heuristics proposed in this paper for exactness obliged picket safeguarding

admittance control are likewise significant in the connection of workload-mindful anonymization. The system is a mix of admittance control and picket assurance instruments. The entrance control system permits just approved inquiry predicates on touchy information. The picket safeguarding module anonymizes the information to meet security necessities and imprecision imperatives on predicates set by the entrance control system. Yet, it has a few impediments, for example, users doesn't have proficient picket and precise requirements. Framework not ready to recover information in altered way. Framework doesn't give security to information which propelled me to chip away at this. An exactness compelled picket protecting admittance control component, showed inFig.[1](Arrows speak to the course of data stream), is proposed. The picket insurance instrument guarantees that the security and precision objectives are met before the touchy information is admittanceible to the entrance control component. The consents in the entrance control arrangement are in view of choice predicates on the QI properties. The arrangement manager characterizes the consents alongside the imprecision destined for every consent/question, client to-part assignments, and part to authorization assignments. The imprecision bound data is not imparted to the clients in light of the fact that knowing the imprecision bound can bring about damaging the picket prerequisite. The picket security system is obliged to meet the security prerequisite alongside the imprecision headed for every authorization. While admittanceing data from database, the idea of imprecision bound is presented in every entrance from database to take care of the issue of where insignificant level of resilience is characterized for every entrance question. Present workload mindful anonymization strategies minimize the imprecision total for all question/consent. The idea of fulfilling the precision limitation for individual authorizations in an approach or workload has not been mulled over some time recently. Exactness compelled picket protectingadmittance control component significant in the workload–aware anonymization. The idea of nonstopinformation distributed has been additionally examined. Numerous entrance control components are there to manage social database. Part based Admittance Control that permits characterizing

authorization on item in view of parts in an association.

II.RELATED WORK :

Admittance control instruments for databases permit inquiries just on the approved piece of the database. Predicate based fine-grained admittance control has further been proposed, where client approval is constrained to predefined predicates. Implementation of admittance control and picket strategies has been considered. Notwithstanding, considering the communication between the entrance control systems and the security assurance components has been missing Related work deals with the previous work related to this paper. The existing methods only deals with either admittance control mechanism, or picket picket mechanism. There was no such a study related to the hybrid of both admittance control mechanism for relational data. Here it deals with the various methods used for the admittance control mechanism and picket picket mechanism. In the case of picket picket, the main method is k-anonymity method; k-anonymity has recentlybeen investigated as an interesting approach to protect sensitive data undergoing public or semi-public release from linking attacks. To protect respondents' identity when releasing micro data, 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 n to re-identify respondents and to infer information that was not intended for release. One of the emerging concepts in micro data picket is k-anonymity, which has been recently proposed as a property that captures the picket of a micro data 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. The disadvantages of the existing systems are:

1)There is no picket for users

2)There is a chance of the linking attacks even after the removal of identifying attributes from the sensitive data.

Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

III.METHODOLOGY:

There are lots of methods for providing the picket for the sensitive information stored in the database and there are different admittance control methods for admittancing the secured information stored in a database. In my project it deals with the introduction of both the admittance control mechanism and the picket picket mechanism together for protecting the sensitive information. Here it uses the anonymity method and thefragmentation method for the picket picket and the imprecision bound for both the admittance control and the picket picket method. The proposed system uses secure reversible Accuracy Constrained Picket-Preserving Admittance Control for relational database. The proposed method provides data publication in a picket preserved method. The framework of the proposed method is a combination of admittance control and picket picket mechanisms. The admittance control mechanism allows only authorized queries predicates on sensitive data. The picket preserving module anonymized the data to meet picket requirements and imprecision constraints on predicates set by the admittance control mechanism. Admittance Control Mechanisms (ACM) is used to ensure that only authorized information is available to users. However, there is chance of sensitiveinformation can still be misused by authorized users for their use. The confidential data can also be misused. The concept of picket-preservation for sensitive data requires the enforcement of picket of the secured sensitive data and picket policies or the picket against identity disclosure by satisfying some picket requirements. In the proposed method, it investigate picket-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 picket picket mechanism. An Anonymization algorithms use suppression and generalization of records to satisfy picket requirements with minimal distortion of micro data. The fragmentation technique and anonymity technique can be used with an admittance control mechanism to ensure both security and picket of the sensitive information. The picket is achieved at the cost of accuracy and imprecision is introduced in theauthorized information

Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

under an admittance 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 allqueries. Then the imprecision added to each permission/query in the anonymized micro data is not known. Making the picket 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 admittance control mechanism. However, the concept of accuracy constraints for permissions can be applied to any picket-preserving security policy. In the picket picket 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.

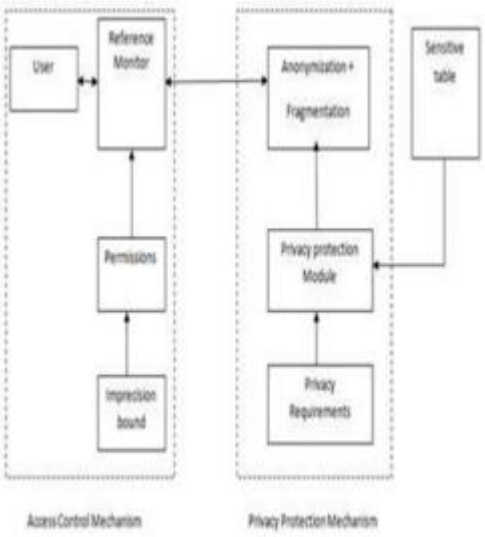


Fig. 1: Privacy Preserved Access Control for Relational Data

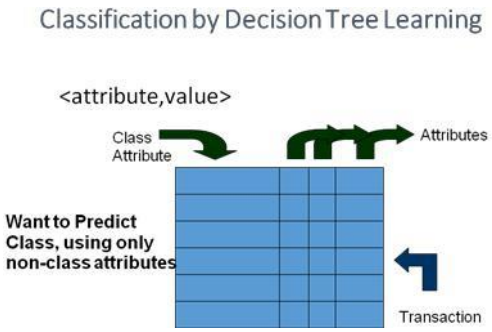
IV.PROPOSED METHOD:

Classification by Decision Tree Learning:

Step1: Rooted tree with nodes/edges

Step2: Internal Nodes => Attributes

Step3: Edges leaving nodes =>Possible values



Step4: Leaves => Expected Class for transaction

– Traverse tree using known attributes

– Predict class given leaf node's value

Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

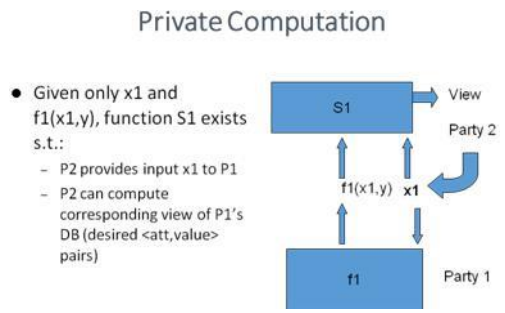
Step5: Top-down

Step6: At each level – find attribute that "best" classifies transactions => gives least overhead

– Best => Attribute that minimizes entropy (maximizes information gain)

– Entropy = -xlnx

– Entropy of class = 0



Algorithm:

Step1: Each party computes ID3 – decision tree learning – (O(# attributes)).

Step2:Combine results using cryptographic protocols like oblivious evaluation - (O(log(#transactions))).

Step3: Result - Each party gains results of data-mining without learning more than necessary

Step4: Each party computes their "share" of entropy

(a)For each attribute, combine values from each party

(b)Results in private computation of Entropy (-xlnx)

Step5: Choose attribute that minimizes entropy

(a)Provides maximum information gain

(b)Ensures most efficient tree with least overhead.

(c)Use oblivious Evaluation.

Step6:Efficient:

Large Databases accommodated: Algorithm relies on number of possible values for attributes, not number of transactions in database.

Step7: Private: Each step depends on local computation and private protocol Uses techniques like oblivious transfer/ evaluation to exchange information Paper proves individual steps are private, AND can predict control flow between steps ONLY based on input/output – so also private.

V.EXPERIMENTAL RESULTS:

The experiments have been carried out on two data sets for the empirical evaluation of the proposed heuristics. The first data set is the Adult data set from the UC Irvine Machine Learning Repository having 45,222 tuples and is the de facto benchmark for k-anonymity research. The attributes in the Adult data set are: Age, Work class, Education,Maritalstatus,Gender Occupation. The second data set is the Census data set from IPUMS. This data set is extracted for Year 2001 using attributes: Age, Gender, Marital status, Race, Birth place, Language, Occupation, and Income. The size of the

data set is about 1.2 million tuples. For the k-anonymity experiments, we use the first eight attributes as the QI attributes.

For the l-diversity experiments, we use Attribute occupation as the sensitive attribute and the first seven attributes as the QI attributes. For the l-diversity experiments, all the tuples having the occupation value as Not Applicable(0 in the data set) are removed, which leaves about 700k tuples. In the case of the variance diversity experiments, Attribute income is used the sensitive attribute and all the tuples having the income value as Not Applicable (9,999,999 in the data set) are removed, which leaves about 950k tuples.

We use 200 and 500 queries generated randomly as the workload/permissions for the Adult data set and Census data set, respectively. The experiments have been con- ducted for two types of query workloads. To avoid yielding too many empty queries, the queries are generated randomly using the approach by Iwuchukwu and Naughton. In this approach, two tuples are selected randomly from the tuple space and a query is formed by making a bounding box of these two tuples. To simulate the permissions for an access control policy, the query selectivity for both the data sets is set to range from 0.5 to 5 percent. For the first workload, if the query output is between 500 to 5,500 tuples for the Adult data set and 1,000 to 50,000 for the Census data set, the query is added to the workload. For the second workload (we will refer to this workload as the uniform query workload) this range (1,000 to 50,000 for Census data set) is divided into ten equal intervals and we add only 50 queries from each interval to the workload. Similarly, for the Adultdataset, 20 queries are added from each size interval. The first workload is used for the l-diversity and variance diversity experiments. The average query size for the Adult data set is 3,000 and for the Census data set is 25,000 for the uniform query workload. The imprecision bounds for all queries are set based on the query size for the current experiment. Otherwise, bounds for queries can be set according to the precision required by the access control administrator. The intuition behind setting bounds as a factor of the query size is that imprecision added to the query is proportional to the query size. Further, as no real relational policy data is available, we believe this approach can allow researchers to reproduce our workload and compare

Volume:15 Issue:1 May – June: 2017
INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS [IJCMI] ISSN: 0974-8652

their results with the approaches presented in this paper. For the k-anonymity experiments, we fix the value of k and change the query imprecision bounds from 5 to 30 percent with increments of 5. Then, we find the number of queries whose bounds have not been satisfied by each algorithm for the uniform query workload. The results for k-anonymity are given in Fig. 6 for the Adult data set for k values of 3, 5, 7 and 9. Heuristic TDH2 has the least number of query bound violations.

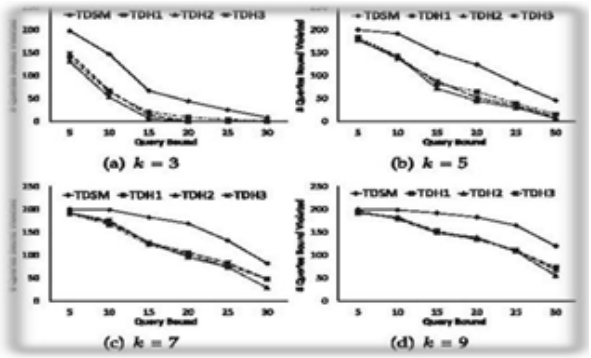


Fig. Adult dataset

In both cases, TDH2 has the lowest number of queries violating the imprecision bounds. The sum of imprecision for all queries is given in Fig., where TDH2 also has the lowest total imprecision for all values of k. In Fig., the total number of violated queries is given that. The number of queries for k-Anonymity which the imprecision bound is violated is given in Fig., for the Census data set using the uniform query workload of 500 queries. The results have the same behavior as that for the Adult data set.

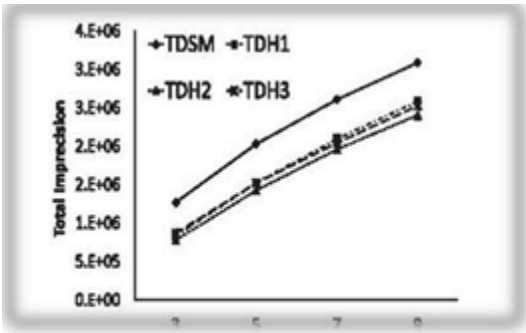


Fig. census data set

The reason for using the uniform query workload (50 randomly selected queries from each size range having cardinality between 0.5 to 5 percent of the data set) is that it helpsobserve the behavior of the queries violating the bounds for each algorithm output partitions to satisfy the imprecision bounds of queries that violate the bound by a less than 10 percent margin. Intuitively, there is more chance of violating the imprecision bounds for a query having a smaller imprecision bound. In Fig. 11, the number of queries violated for each size range (10 size intervals in 1k-50k) are plotted. Thus, for TDH1, TDH2 and TDH3 favor queries with smaller bounds initially. The behavior of TDSM follows the intuition as more queries in the smaller size range are

violated. For TDH1, the heuristic always favors the queries with smaller bounds when being considered for a partition split.

VI.CONCLUSION

An accuracy constrained picket p reserving admittance control framework for relational data has been proposed. The planned additive approach of admittance management and picket picket mechanisms in our system provides a lot of security and information is retrieved during a custom-made approach which will build users to admittance during as lot of versatile approach. Any admittance management concentrate on anomaly users to avoid picket problems security. The ACM allows solely licensed user predicates on sensitive information and PPM anonymizesthe information to satisfy picket necessities and inexactness constraints on predicates set by the admittance management mechanism. The framework is a combination of admittance control and picket picket mechanisms. The admittance control mechanism allows only authorized query predicates on sensitive data. The picket preserving module anonymizes the data to meet picket requirements and imprecision constraints on predicates set by the admittance control mechanism. This interaction is formulated as the problem of k-anonymous Partitioning with Imprecision Bounds (k -PIB). Hardness results are given for the k - PIB problem and the heuristics for partitioning the data are presented to satisfy the picket constraints and the imprecision bounds. In the current work, static admittance control and relational data model has been assumed. The proposed picket-preserving admittance is extended to control incremental data and cell level admittance control.

VII.REFERENCES:

[1]Bertino E. and Sandhu (2005),“DatabaseSecurity-Concepts Approaches, and allenges,”IEEE Trans.Dependable and SecureComputing, vol. 2, no. 1, pp. 2-19.

[2]Fung B. et al (2010), “Picket-Preserving Data Publishing: A Survey of Recent Developments,” ACM Computing Surveys, vol.42, no. 4, article 14, 2010.

[3] Ghinita G. et al (2009),“A Framework forEfficient Data Anonymization Under Picket andAccuracyConstraints,”ACMTrans.DataabaseSystems, vol. 34, no. 2, article 9.

[4] Li N. et al (2011), "Provably Private DataAnonymization: Or, k -AnonymityMeetsDifferential Picket," Arxiv preprint arXiv:1101.2604.

[5] LeFevre K. et al (2008), "Workload Aware Anonymization Techniques for Large-ScaleDatasets," ACM Trans. Database Systems, vol.33, no. 3, pp. 1-47.

[6] Rizvi S. et al (2004), "Extending Query Rewriting Techniques forFine-Grained Admittance Control," Proc. ACMSIGMOD Int'l Conf. Management of Data, pp.551-562.

[7] ZahidPervaiz and Walid G. Aref(2014), "Accuracy - Constrained Picket-Preserving Admittance Control Mechanism for Relational Data" IEEE Transactions OnKnowledge And Data Engineering, Vol. 26, No. 4.

[8] 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.