International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

ACM Access control using k-anonymity, query evaluate on outsourced database

Ghousia Begum[1]& M.Mohanrao[2]
[1]M-Tech Dept. of CSE Megha Institute of Engineering & Technology for Women
[2]Assistant Professor Dept. of CSE Megha Institute of Engineering & Technology for Women

Abstract
Micro data refers to series of records, each record with information on an individual unit like a patient or an organization. Access Control Mechanisms (ACM) safe the sensitive information from unauthorized users. Even sanctioned users may misuse the data to reveal the privacy of individuals to whom the data refers to. Privacy prosperous Mechanism (PPM) anonymize the relational data to avoid identity and attribute disclosure. It is achieved by generalization or suppression. Role-predicated access control gives users the sanctions to access the data predicated on their roles. The Access Control Mechanism define sallow predicates available to purposes while the privacy required is to fill the k-anonymity or l-diversity. Define bound constraint is assigned for each individual predicate. Top down Selection Mondrian (TDSM) algorithm is utilized for query workload-predicated anonymization algorithm is constructed utilizing acquisitive heuristics and kd-tree model. Query cuts are culled with minimum bounds in Top-Down Heuristic1 algorithm (TDH1). The query jumps are modified as the partitions are integrated to the output in Top-Down Heuristic 2 algorithm (TDH2). The price of decreasedaccurecy in the query results is utilized in Top-Down Heuristic 3 algorithm (TDH3). Repartitioning algorithm is utilized to reduce the total imprecision for the queries. The privacy preserved access control framework is enhanced to provide incremental mining features utilizing R+-tree. Data insert, expunge and update operations are associated with the partition management mechanism.

***Keywords**:* Access control; k-anonymity; query evaluation; outsourced database; Access Control Mechanisms (ACM); Role-predicated access control; Privacy Auspice Mechanism

1. Introduction

Current ecumenically networked society greatly demands the sharing and propagation of data. While data relinquished in the past was in tabular and calculate statistical form(macrodata),there is a necessity for the relinquishment for categorical data to perform statistical analysis on them(microdata). Microdatarefers to series of records, each with information on an individual unit like a person or an industrial unit.It sanctions the recipient to perform a whole incipient analysis on them as needed. In order to safety the identity of individuals to whomthe data refers to,when relinquishing microdata, data holders often abstract or encrypt expressed identifiers, this as names and gregarious security numbers.

Organizations accumulate and analyze consumer data to amend their accommodations. Access Control Mechanisms (ACM) iswont todetermine that only sanctioned information is usable to users. How, sensitive data can still be misused by sanctioned users to compromise the privacy of users. The conception of privacy-preservation for sensible data can accept the Social control of privacy policies or the safe against identity disclosure by gratifying some privacy requisites [1]. In this paper, we inquireprivacy-preservation from the anonymity view. The sensibleinfo, even after the

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

abstraction of describing attributes, is even capable to associating attacks by the sanctioned users [2]. This dilemma has been analyzed extensively in the place of micro data issuing [3] and privacy definitions, e.g., k-anonymity [2].

The other quasi-identifiers reveal the privacy. The "linking attack" [4] should be managed to secure privacy of individuals. These linking attacks can be managed by anonymizing data in tables. Anonymization is the process of abstracting the identity particulars by obnubilating or transforming the information. An innocent table is the one which is composed after transmuting the data that does not distinguish the individual characteristics. There are sundry anonymization methods that avail in holding privacy.

2. Related Work

2.1 Existing System:

ORGANIZATIONS amass and analyze consumer data to ameliorate their accommodations. Access Control Mechanisms (ACM) habituated to ascertain only sanctioned information is usable to users. Sensible info can even be abused by sanctioned users to compromise the secrecy of users. The privacy-preservation for sensitive data concept can need the social control of secrecy policies or the auspice against identity disclosure by gratifying some privacy requisites. Subsisting workload cognizant anonymization method understate the impreciseness aggregative for all questions and the impreciseness integrated to each sanction/query in the anonymized micro information is not kendo. Making the privacy requisite more stringent (e.g., incrementing the value of k or l) end results in unessential imprecision for queries.

2.2 Proposed System:

Access control mechanisms for databases sanction queries only on the sanctioned part of the database. Predicate predicated fine-grained access control, where utilize sanction is circumscribed to pre-defined predicates. Enforcement of access assures and secrecy policies have been studied. Still, studying the fundamental interaction between the access control mechanisms and the secrecy protection mechanisms has been losing. Lately, Chaudhuri et al. have analyzed access assure with secrecy policy. They utilize the definition of differential privacy whereby arbitrary noise is integrated to whole query results to satisfy privacy constraints. They have not considered the precision constraints for approves. The privacy needed in conditions of k-anonymity. It has been expressed by Li et al. That after sample, k-anonymity offers kindred privacy guarantees as those of differential privacy. The Precision-constrained privacy preserving access control framework sanctions the access control administrator to designate imprecision constraints that the privacy safe mechanism is needed to meet along with the secrecy requisites.

The privacy-cognizant access control challenges are inside to the dilemma of workload-cognizant anonymization. In our analysis of the within work, we settle on query-vigilant anonymization. For the concept of state the art in k-anonymity techniques and algorithms program, we concernthe reader to a recent study paper. Workload-cognizant anonymization is first studied by LeFevreetal.They have advised the Cull Mondrian algorithm [4], which is a modification to the avaricious multidimensional partitioning algorithm Mondrian. In their algorithm, predicated on the given query-workload, the avaricious splitting heuristic minimizes the sum of imprecision for all quetions. Iwuchukwu and Naughton have advised an Rþ-tree predicated anonymization

algorithmic view. The writersexemplify by experiments that anonymized data utilizing partial Rþ-tree predicated on the given query workload is more precise for those queries than for an equitable algorithm. Ghinita et al. have proposed algorithms predicated on space filling curve balltowards k-anonymity and l-diversity [10].

They additionally introduce the quandary of precision-constrained anonymization towards a given bound of acceptable data loss for each one parity class [8]. Similarly, Xiao et al. [9] propose to integrate noise to questions according to the size of the questions in a given workload to gratify differential secrecy. Jumps for query imprecision have not been considered. The subsisting literature on workloadaware anonymization has a concentrate to decrease the overall imprecision for a given fixedquestions. Anonymization with impreciseness constraints for individual questions has not been studied afore. Ye imprecision definition of LeFevre et al. And bring inyerestraint of impreciseness bound for each questions in a given questions workload.

3. Exactness-forced secrecy-preserving access control:

The privacy aegis mechanism ascertains that the privacy and precision goals are met afore the sensitive information is available to the check control policy. The sanctions in the access control policy are predicated on search predicates on the QI assignable. The policy decision maker defines the sanctions along with the imprecision bound for every sanction/query, exploiter Toronto duty assignments, and role-to sanction assignments [18]. The designation of the imprecision bound ascertains that the sanctioned data has the desired level of preciseness. The impreciseness bound data is not shared with the delegates because kenning the imprecision bound can result in breaching the privacy requisite. The privacy auspice mechanism is required to meet the privacy requisite along with the imprecision bound for each sanction.

(i) Access control enforcement:

The exact tuple values in a cognation are superseded by the generalized values after the anonymization. In this case, access control Mechanism over the generalized informationrequired to be defined. In this section, discussion about the Relaxed and Rigorous assure control social controlpolicies over anonymized information. The access checkMechanism by reference monitor can be of the following two types: 1. Relaxed - Utilization of overlap semantics to sanction access to all divisions that are lapping the sanction. 2. Rigorous- Utilization of enclosed semantics to sanction access to only those partitions that arecomprehensive enclosed by the sanction. Both schemes have their own advantages and disadvantages. Relaxed enforcement infringes the sanction predicate by giving access to extra calculating but is benign for applications where low cost of an erroneous alarm is tolerable as compared to the jeopardy associated with an escaped case. Examples let in epidemic surveillance and airport privately. On the other way, rigorous enforcement is felicitous for applications where a high danger is linked with a mendacious alarm as likened to the cost of a loosed event. An example is a mendacious apprehend in instance of shoplifting.

In this paper, the concentrate is on untaxed enforcement. Still the proposed methods for anonymization are withal valid for stringent enforcement because the proposed heuristics decrease the overlap among divisions and questions. Further surmise that under decompressed enforcement if the imprecision bounce is breached for a sanction then that sanction is not assigned to any role.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

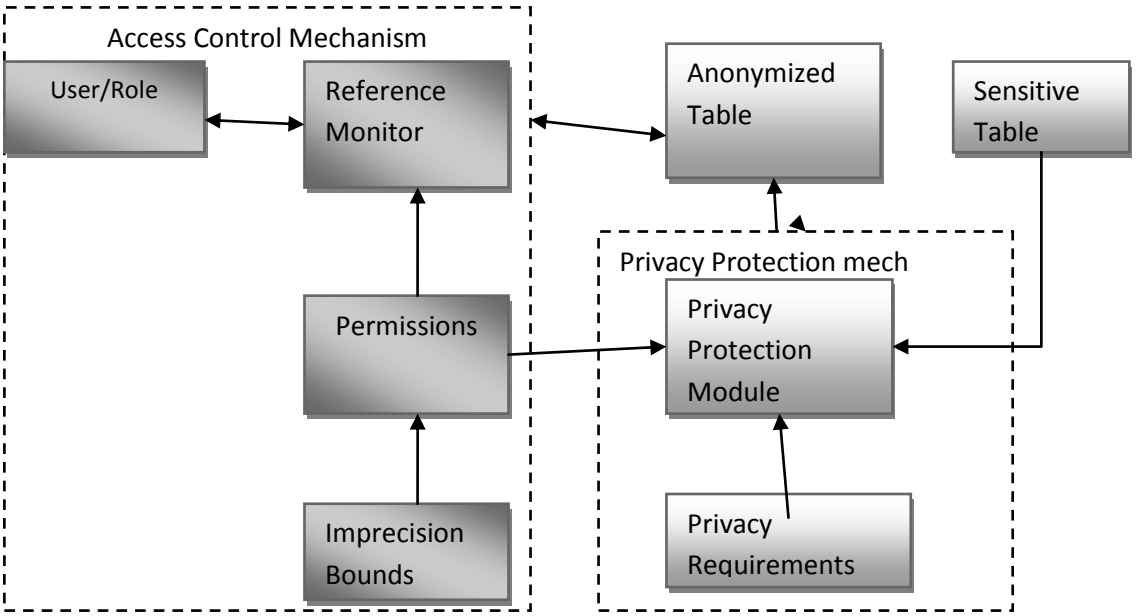


Fig 1: Privacy-preserving access control

A privacy-preserving access control framework is shown in Figure 1, where the privacy auspice mechanism ascertains that the privacy and precision goals are met afore the sensible information is available to the access control mechanism. The access control policies define sanctions for roles predicated on cull predicates. Privacy Bulwark Mechanisms (PPM) use suppression and generalization to anonymize and gratify privacy requisites. The procurement of the privacy goals is achieved at the cost of the precision of the data available to the sanctioned users. The access control mechanism needs to designate the caliber of imprecision that can be abode by the utilizer for each sanction. This designation of the imprecision bound ascertains that the sanctioned information has the desired level of precision. Then, the privacy auspice mechanism needs to meet the privacy requisite along with the imprecision bound for each sanction.

4. Data partitioning for privacy preservation:

In this, three algorithms predicated on avaricious heuristics are proposed. All three algorithms are predicated on kd-tree structure. Beginning with the completely tuple blank the nodes in the kd-tree are recursively divided till the division size is among k and 2k. The leaf nodes of the kd-tree are the output divisions that are represented to parity classes. Heuristic 1 and 2 have time involution of O(d 2 Q n 2). Heuristic 3 is a modification over Heuristic 2 to have O(d|Q|nlgn) involution, which is same as that of TDSM.

4.1 Top-Down Heuristic 1 (TDH1):

The TDH1 algorithm is named in Algorithm 1. in the first place, the whole tuple outer space is contributed to the set of candidate partitions. In the Lines 3-4, the query lapping the candidate division with to the lowest degree impreciseness bound and impreciseness greater than zero is picked out. The while loop in Lines 5-8 checks for a feasible burst of the division along query separations. If a feasible cut is found, then the leaving divisions are added to CP. other than, the candidate division is checked for average cut in Line 12. A workable cut thinks of that each division resulting from split should fulfill the privacy demand. The traversal of the kd-tree for partitions to consider in Set CP can be depth-first or breadth-first.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

However, the order of traversal for TDH1 does not matter.

Input: T,K,Q and BQ j

Output: P

1 Initialize set of candidate partitions(CP CP) do∈

2 for (CP i

3 Find the set of queries QO that overlap CP I such that ic j I QOCP > 0

4 sort queries QO in raising order of BQj

5 while (feasible cut is not found) do

6 Select query from QO

7 Create query cuts in each dimension

8 Select dimension and cut having least overall imprecision for all queries in Q

9 if (feasible cut found) then

10 Create new partitions and add to CP

11 else

12 Split CP i recursively along median till anonymity requirement is satisfied

13 Compact new partitions and add to P

14 return (P)

Algorithm 1: TDH1.

4.2 Top-Down heuristic rule 2 (TDH2):

In the Top-Down Heuristic 2 algorithm, ye query jumps are modified as the partitions are added to the output signal. This update is carried out by deducting the icQj Pi value from the imprecision bound BQj of each query, for a Partition, say Pi, that is being added to the output signal. For instance, if a devision of size k has impreciseness 5 and 10 for questions Q1 and Q2 with impreciseness bound 100 and 200, then ye bounds are altered to 95 and 190, severally. The best effects are reached if the kdtree traversal is depth-first (preorder). governtraversal for the kd-tree ensures that a given partition is recursively split till the leaf node is accomplished. Then, the query bounds are changed. Initially, this approach favors queries with littler bounces. As more divisions are added up to the outturn, all the queries are treated fairly. During the question bound modify, if the imprecision bound for any query gets violated, then that question is put on low precedence by replacement the question bound by the query sizing. The intuition behind this decision is that whatever future partition splits TDH2 builds, the question jump for this query cannot be satisfied. Therefore, the concentrate should be on the left queries.

Input: T,K,Q and BQj

Output: P

1 Initialize set of candidate partitions (CP CP) do //Depth first preorder traversal∈

2 for (CP i

3 Find the set of queries QO that overlap CP I such that icQOjCPi> 0

4 Sort queries QO in increasing order of BO j

5 While (feasible cut is not found) do

6 Select query from QO

7 Create query cut each dimension

8 select dimension and cut having least Overall imprecision for all queries in Q

9 if (Feasible cut found) then

10 Create new partitions and add to CP

11 else

12 Split CP i recursively along median till anonymity requirement is satisfied

13 Compact new partitions and add to P Q∈ Q j ∀

14 Update BQ j according to ic Qi pi ,

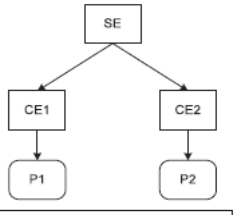
15 return (P)

Algorithm 2: TDH2.

5. Implementation

Access control policy:

Role	Designation
SE	State Epidemiologist
CE1	County 1 Epidemiologist
CE2	County 2 Epidemiologist



Permission	Authorized Query Predicate
P1	Location = County 1 ∧ Age = 15-65 ∧ Syndrome = Influenza
P2	Location = County 2 ∧ Age = 15-65 ∧ Syndrome = Influenza

Fig 2: Access control policy.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

The Syndromic surveillance systems are utilized at the country and union degrees to discover as well as monitor threats to public health. The section of wellness in a state collects the emergency section data (age, sexuality, localization, time of advent, indication, etc.) from county hospitals associating. In normally, each day by day modify consists of a stable case that is relegated into complex classes by the section of the health. Then, the surveillance information is anonym zed and shared with departments of health at from each county. The approach control mechanism is show in Figure. 2 that sanctions the characters to access the calculate below the approve proclaim, e.g., part CE1 can access calculating under permit P1. The epidemiologists at the state and county degree suggest community of interests containment evaluates, e.g., isolation or quarantine according to the number of persons infected in case of a flu outbreak. Granting to the universe tightness in a county, an epidemiologist can propose closing off if the number of persons reported with influenza are more preponderant than 1,000 and quarantine if that number is more preponderant than 3,000 in a single day.

A. Anonymity:

	QI_1	QI_2	S_1
ID	Age	Zip	Disease
1	5	15	Flu
2	15	25	Fever
3	28	28	Diarrhea
4	25	15	Fever
5	22	28	Flu
6	32	35	Fever
7	38	32	Flu
8	35	25	Diarrhea

(a) Sensitive table

	QI_1	QI_2	S_1
ID	Age	Zip	Disease
1	0-20	10-30	Flu
2	0-20	10-30	Fever
3	20-30	10-30	Diarrhea
4	20-30	10-30	Fever
5	20-30	10-30	Flu
6	30-40	20-40	Fever
7	30-40	20-40	Flu
8	30-40	20-40	Diarrhea

(b) 2-anonymous Table

Fig 3: Anonymity sensitive tables.

Anonymity is prostrate to homogeneousness approaches when the sensitive value for all the tuples in an equipollence class is equipollent. To contravene this defect, l-diversity has been placed and demands that each equipotent Fig. 2. Access control policy. Classes of T_ contain at least 1 discrete assesses of the sensibleassign. For sensitive numeric attributes, an l-diverse parity class can still leak data if the numeric values are proximate to each another. For such cases, variation variety has been introduced that requires the variance of each parity class to be more preponderant than a given variance variety parameter. The table in Fig. 3a does not slake k-anonymity because kenning the age and zip code of individual punishments associatingillness to that person. The table in Fig. 3b is a 2-innominate and 2-diverse interpretation of table in Fig. 3a. The ID attribute is swiped in the anonymized table and is shown only for recognition of holding. Here, for any heap of cull proclaims on the zip write in code and age attributes, there are at least two plus in each parity class.

B. Accuracy-Constrained Privacy-Preserving Access Control:

A precision-constrained privacy-preserving access assures mechanics. (Arrows represent the direction of information flow), is proposed. The assure bulwark mechanics determines that the privacy and precision goals are met afore the sensitive data is available to the access control mechanism. The sanctions in the access control policy are predicated on cull predicates on the QI attributes. The policy administrator defines the sanctions along with the imprecision bound for each sanction/query, utilizer-to-role assignments, and role-to sanction assignments. The designation of the imprecision bound ascertains that the sanctioned data has the desired level of precision. The impreciseness bound data is not shared with the users because

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 09, September 2015
Available at http://internationaljournalofresearch.org

kenning the imprecision bound can result in infringing the Privacy requisite. The privacy aegis mechanism is required to meet the privacy requisite along with the imprecision bound for each sanction.

C. Top-Down Heuristic:

In TDSM, the divisions are divide by the median value. Consider a division that inter sections a query. If the median withal comes down deep down the query then even after devidingye partition, ye imprecision for that query will not transmute as both the incipient divisions even overlap the query as illustrated. In such heuristic rule, we propose to split the division along the question cut and then optate the proportion along which the impreciseness is minimum for all questions. If multiple questions overlap a suppuration, then the query to be utilized for the cut needs to be picked. The questions having imprecision more preponderant than zero for the partition are sorted predicated on the imprecision bound and the query with minimum imprecision bound is picked. The suspicion abaft this decision is that the queries with more diminutive leaps have let down tolerance for error and such a division break ascertains the defragmentation in imprecision for the query with the most diminutive imprecision bound. If no feasible cut slaking the privacy requisite is found, then the next question in the sorted list is utilized to check for division split. If none of the questions sanction partition split, then that division is divide along the average and the leading divisions are incorporated to the end product after compaction.

6. Experimental results

Fig 4: Administrator home Page.

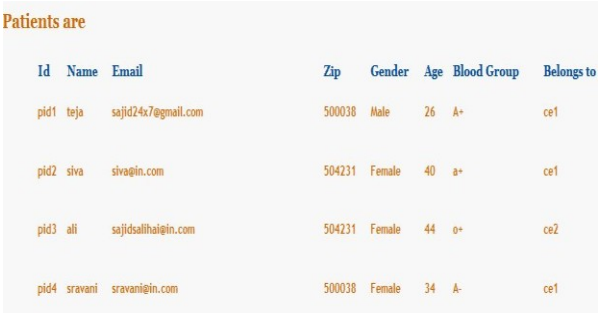


Fig 5: Patients sensitive data.

Fig 6: Sensitive Data.

Fig 7: Making Data Anonymization Page.

7. Conclusion

Access control mechanism for relational data is constructed with the privacy preservation predicated model. Role Predicated Access Control (RBAC) scheme provides security to the data by sanctioning access predicated on sanctions. K-Anonymity model is integrated with minimum imprecision predicated data access control mechanism. Partitioning utilizing R+-trees results in less number of overlapping partitions. Hence precision is ameliorated and time involution is reduced in the system. Privacy preserved data access control mechanism is ameliorated with incremental mining model. The system reduces the imprecision rate in query processing. Access control mechanism is acclimated for incremental mining model.

8. References

[1] S. Chaudhuri and Sudarshan, "Fine Grained Authorization through Predicated Grants," Proc. IEEE 23rd Int'l Conf. Data Eng.,, 2007.

[2]R.Agrawal,P.Bird,T.Grandison,J.Kiernan,S. Logan and W.Rjaibi,"Extending Relational Database Systems to automatically Enforce Privacy Policies,"*Proc.21st Int'l Conf. Data Eng.,*pp.1013-1022,2005.

[3] S. Chaudhuri, Kaushik and R. Ramamurthy, "Database Access Control & Privacy: Is There a Common Ground?" Proc. Fifth Biennial Conf. Innovative Data Systems Research, 2011.

[4] 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.

[5] N.Li,W.Qardaji, and D.Su,"Provably Private Data Anonymization:Or,k-Anonymity Meets Differential Privacy," *Arxio preprintarXiv:1101.2604*,2011.

[6] X. Xiao, G. Bender, M. Hay and J. Gehrke, "Ireduct: Differential Privacy with Reduced Relative Errors," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2011.

[7]ZahidPervaiz,WalidG.Aref,ArifGhafoor,NagabhushanaPrabhu,"Accuracy-constrained Privacy Preserving Access Control Mechanism for Relational Data,"IEEETrans.Knowledge and Data Engineering,vol.26,no.4,pp. 795-807,2014.

[8] 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.

Author Profile

Dr.Shaik Abdul Muzeer
Professor & Principal
Megha institute of Engineering & Technology for Women

Dr.S.A.Muzeer, at present working as a principal of Megha institute of engineering & Technology has completed his PG and P.HD in Electronics & Communication Engineering and published around 25 Papers in National & International Journals. His area of research is Digital signal processing and Bio-medical engineering