© May 2016 | IJIRT | Volume 2 Issue 12 | ISSN: 2349-6002

Privacy Preserving for Inference Control with Improved ND Algorithm

Maulik Joshi[1] , Chetna Chand[2]

[1]*M.E. Student, Department of Computer Engineering, Kalol Institute of Technology & Research Center*

[2]*Assistant Professor, Department Of Computer Engineering, Kalol Institute of Technology & Research Center*

***Abstract*- Online analytical processing (OLAP) is providing functionality of analysis of multidimensional data cube and it is support decision making and knowledge discovery technique. OLAP operation on such databases may reveal the information which is private to an individual. Privacy preserving is important in OLAP because private information can be shown through user query. So adversarial inference of private information its main issue in OLAP. Privacy preserving OLAP had focused on single aggregate function but which eliminates from consideration an important class of privacy breaches where partial information, but not exact values, of private data is disclosed .In this paper proposed technique provide protection for exact and partial disclosure in OLAP with more than one aggregate function with reducing processing time of OLAP cube.**

***Index Terms*- Privacy, preserving, OLAP, inference problem**

I. INTRODUCTION

Data warehouse is a collection of decision support technology such as OLAP and Data Mining that enables the knowledge worker to make better and faster decisions. OLAP is providing knowledge discovery and decision support technique in business intelligence system. The design of data warehouse and OLAP system by its very nature conflicts with security. At one side, the Goal of data warehouse is to make all data available to all users; especially the OLAP adhoc queries need open nature. Imposing security may hinder the analysis process. On the other side the data warehouse contains an organization’s valuable and summarized data that should be protected from all kinds of malicious access. In traditional databases, only base tables are considered for security model authorization and not the summarized data, while OLAP security should not only consider authorization on aggregated data but also on the detailed data.

We can apply query on multidimensional data cube. Among various ways of data analysis, OLAP is one of the most popular techniques. OLAP helps analysts to extract useful knowledge from large amount of data. Similar to any technology, OLAP is also double edged sword. Without sufficient security components, an OLAP system may become a powerful tool in the hands of malicious users in threating the privacy of individuals. Protection of private information is main issue in online analytic processing system; adversarial inference of private information from OLAP query answers is major privacy problem [1]. Privacy preserving OLAP had focused on single aggregate function but it is not consider important class of privacy breaches and partial information has been generated. In this proposed approach privacy protection in front of both exact and partial disclosure in OLAP systems using more than one aggregation function. Propose approach we consider SUM like function and MIN Like function [4]:

- MIN-like functions: MIN and MAX
- SUM-like functions: SUM, AVG, COUNT, MEDIAN, and STANDARD DEVIATION

It is provide guarantees that the privacy disclosure can not to exceed thresholds predetermined by the data owners. We will implement base paper algorithm with modification to make it faster and also reduce size & processing time of OLAP cubes. Our approach will efficient and can be implemented in existing OLAP systems with little modification. It is satisfies the simulate able auditing model and leaks no private information through query rejections.

II. PROBLEM STATEMENT

The data warehouse server stores private data and server may answer on the multidimensional aggregates of private data by users OLAP queries .However, it is a challenge to enable OLAP on private data without privacy breach of data owners.

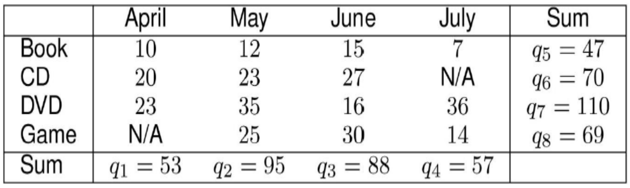
User may not access all individual data in data warehouse but privacy breach occurs when user will get certain information about private data point from OLAP queries but user don’t have right to access that private data. So here using query answer user can infer certain information of private data point. This privacy breach become as the inference problem.

Now, example of inference problem as per base paper [1]:

In this example attribute Y is sensitive cell in collection 1 so user can not access information of this private data point.

User asks two queries:

1. What is the total no. of items in collection1?
2. What is the value of attribute X in collection 1?



	April	May	June	July	Sum
Book	10	12	15	7	$q_5 = 47$
CD	20	23	27	N/A	$q_6 = 70$
DVD	23	35	16	36	$q_7 = 110$
Game	N/A	25	30	14	$q_8 = 69$
Sum	$q_1 = 53$	$q_2 = 95$	$q_3 = 88$	$q_4 = 57$	

So first query answer is 47 and second query answer is 7.

Using this two query we identified our private data information such as in collection have 47 total number of items and attribute X contain 7 items out of 47 so we easily to get attribute Y information from this two query.

III. RELATED WORK

OLAP privacy obtained from data perturbation [5] and show that our perturbation provides guarantees against privacy breaches. Here develop algorithms for reconstructing counts of sub cubes over perturbed data. It is also identify the tradeoff between privacy guarantees and reconstruction accuracy and show the practically of our approach the perturbation algorithm is publicly known; the actual random numbers used in the perturbation, however, are hidden. To allow clients to operate independently, we use local perturbations so that the perturbed value of a data element depends only on its initial value and not on those of the other data elements. Different columns of a row are perturbed independently. We use retention replacement schemes where an element is decided to be retained with probability p or replaced with an element selected from a probability distribution function (p.d.f.) on the domain of elements.. The perturbation algorithm is everyone known; the actual random numbers used for hide sensitive data in the perturbation Technique. To allow clients to operate independently, it is use local perturbations so that the perturbed value of a data element depend only on its initial value and not on those of the other data elements. It is also proposed reconstruction algorithms both analytically and empirically

Another data perturbation technique called uniformly adjusted distortion [2], in this technique initially distorts one cell and then uniformly distributes this distortion in the whole data cube. This also provides accuracy with range sum queries and high availability. It presented a simple but effective distortion technique for privacy preservation in data cube. Distortion technique would not affect the response time of OLAP system queries as all calculations will be done at source side before the interaction of the users. Also it is applicable without divide a data cube into blocks as oppose to the state of the art technique. This work proposes a simple but effective distortion technique, which replaces the original value (Ii of a cell with addition of noise $\varepsilon_{i.}$ like $\beta_i = \alpha_i + \varepsilon_{i.}$ (Where β_i represent the distorted value and i is the position of a cell). Our approach is based on Relative distortion like 50% rather than Absolute distortion in order to avoid the problem of scalability. The motivation for selection of perturbation based approach is that, if a malicious user infers the sensitive data and discloses it, this disclosure will be partial and still he will not be able to get the exact value. Also, this method allows answering all queries e.g. range sum and other queries without any restrictions.

In this paper propose an innovative framework based on flexible sampling-based data cube compression techniques[15] for computing privacy preserving OLAP aggregations on data cubes while allowing approximate answers to be efficiently evaluated over such aggregations. In this proposal, this scenario is accomplished by means of the so-called accuracy/privacy contract, which determines how OLAP aggregations must be accessed throughout balancing accuracy of approximate answers and privacy of sensitive ranges of multidimensional data.

In This Paper [4] address issues related to the protection of private information in Online Analytical Processing (OLAP) systems, where a major privacy concern is the adversarial inference of private information from OLAP query answers. Most previous work on privacy preserving OLAP focuses on a single aggregate function or addresses only exact disclosure, which eliminates from consideration an important class of privacy breaches where partial information, but not exact values of private data is disclosed. We address privacy protection against both exact and partial disclosure in OLAP systems with mixed aggregate functions. In particular, it is propose an information-theoretic inference control approach that supports a combination of common aggregate functions and guarantees the level of privacy disclosure not to exceed thresholds predetermined by the data owners. It is demonstrate approach is efficient and can be implemented in existing OLAP systems with little modification. It also satisfies the simulatable auditing model and leaks no private information through query rejections. Through performance analysis, It is show that compared with previous approaches, This approach provides more effective privacy protection while maintaining a higher level of query-answer availability.

In Base Paper N-D algorithm [1] provided protection against Inference problem. This paper implement N-D algorithm For SUM LIKE Queries and given protection from privacy breaches. It is also demonstrate SUM Like Query in result. . If any query leads to inference problem then rejection of query or precisely give answer. In this paper takes assumption for algorithm is for n-dimensions. It is provide protection against both partial and exact disclosure. Here propose an information-theoretic inference control approach that supports a combination of common according to its communication aggregate functions (e.g. COUNT, SUM, MIN, MAX, and MEDIAN) and guarantees the level of privacy disclosure not to exceed thresholds predetermined by the data owners.

IV. COMPARISON WITH OTHER OLAP PRIVACY PRESERVING TECHNIQUES

Basic we have two types of technique for privacy preserving in OLAP:

A. Inference control
B. Data perturbation

Using two techniques we prevented from privacy breach so how to privacy preserving using methods describe below:

A. Inference Control

Inference control [1], [4] approach, which is based on three tier architecture. It given layer between user query and cube and in this layer contains predefined inference free aggregation function. If any query leads to inference problem then rejection of query or precisely give answer.

B. Data perturbation

Data perturbation [5] is one more approach for privacy preserving in OLAP. In this technique include noise with input source and when query issues for information, we will get answer with some estimation rather than exact value. So it is disadvantage of this technique.

Distortion technique [2] is part of data perturbation technique .This technique through replace to original value of cell with noise. In distortion technique, perform relative distortion (means 50%) rather whole cube distortion for avoid scalability.

V. PROPOSED WORK

Proposed solution is based on inference control approach, it is providing protection in both case exact and partial disclosure of data from data cube. Due to privacy concerns, the owner of a data cube may not want a user to have access to all the information stored in it. Here we perform our approach with both MIN like and SUM like aggregates. In this approach if query leads to generate private information then query will be rejected. And query does not lead to inference problem after that it gives precisely answer to the user according query. Using this approach we improved performance of n-d algorithm with modification in algorithm. This approach provides privacy preservation of OLAP cube from inference problem. It will not give any inference value to malicious user because using inference value, user able to get private information.

A. Steps of proposed approach of inference Control Algorithm

```
1. Get query
2. Check for inference factors
    a. Call procedure ND-Random
        i. Select number at position (row x column) from the set.
        ii. (To exclude row and column from retrieval list) add it to a vector
        iii. Go to i.
    b. Check for type of query
    c. If q == MIN-Like then
        i. Go from first record to n-h record
        ii. Retrieve records after calculating security level for each record
        iii. Calculate values base on equation in vector
    d. Find set of all possible values of inference
        i. match appropriate labels with random values
        ii. find results with MIN query store in vector
        iii. set Max to get max value that satisfied inference results from group of MIN results
    e. End for
    f. End if
    g. If inference value > set level value
        i. Return invalid query
        End if
    h. Else
    i. The query is SUM-Like
        i. Find target score with match values
        ii. Set upper bound and lower bound
        iii. Set standard deviation measure = 0 and check for infinity by random value match
        iv. Store possible values in vector
        v. Match all vector values in with min values
        vi. If inference value > set level value
        vii. Return invalid query
        viii. End if
        j.  End if
        k.  If vector != null
            i.  Store vector values in db
        l.  End if
        m.  Return null
        n.  Exit
3.  Exit
```

B. Flow chart of proposed system

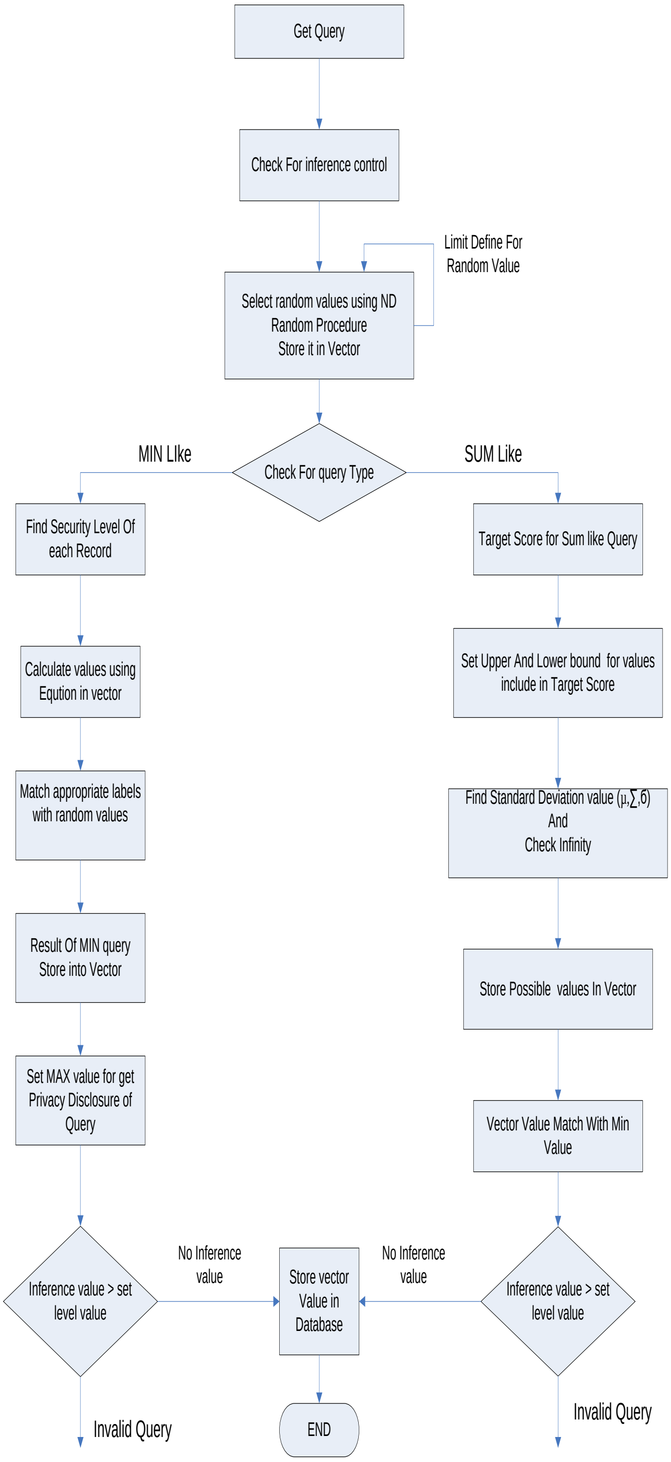

Get Query
Check For inference control
Limit Define For Random Value
Select random values using ND Random Procedure Store it in Vector
MIN LIke
Check For query Type
SUM Like
Find Security Level Of each Record
Calculate values using Eqution in vector
Match appropriate labels with random values
Result Of MIN query Store into Vector
Set MAX value for get Privacy Disclosure of Query
Target Score for Sum like Query
Set Upper And Lower bound for values include in Target Score
Find Standard Deviation value (μ,∑,6) And Check Infinity
Store Possible values In Vector
Vector Value Match With Min Value
Inference value > set level value
No Inference value
Store vector Value in Database
No Inference value
Inference value > set level value
Invalid Query
END
Invalid Query

Fig: - Flowchart of Proposed work

Using improved N-D algorithm we get better performance than previous approach. User apply any query on cube and that will go for check inference. After that it will be decided about query type means query is SUM like or MIN like. According to type of query further execution will be performed. If inference value is more than set level value then query will be rejected otherwise values store into database for showing values to the user.

VI. CONCLUSION

This paper proposed techniques for privacy preserving in OLAP from inference problem. For inference problem we have different techniques with their advantage and disadvantage. Through each technique we got control on inference problem but we should make more efficient method, for that purpose we proposed new approach we used randomization for minimize checking of every single value & reduce processing time of OLAP cube with modification of base paper algorithm. This approach performed with SUM like and MIN like aggregates. In both case we get efficiently privacy preserving of data in cube. It gives protection in exact or partial disclosure of data.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their precious comments and suggestions that contributed to the expansion of the work.

REFERENCES

[1] Rohit Goel, Mahesh Kumar,” Implementation of Privacy Preservation of N-D Algorithms for Online Analytical Processing”, IJIRCCE, Vol. 2, Issue 6, June 2014.

[2] Sara Mumtaz, Azhar Rauf, Shah Khusro,” A Distortion Based Technique for Preserving Privacy in OLAP Data Cube”,IEEE,2011.

[3] Gunwanti R. Bawane, Prof. Prarthana Deshkar,” Integration of OLAP and Association rule mining”,IEEE,2015.

[4] Nan Zhang, Member, Wei Zhao, Fellow,” Privacy-Preserving OLAP: An Information-Theoretic Approach”,IEEE, VOL. 23,2011.

[5] Rakesh Agrawal, Ramakrishnan Srikant, Dilys Thomas,” Privacy Preserving OLAP”, ACM,2005.

[6] Han Jiawei, Kamber M. Data Mining: Concepts and Techniques. San Francisco California: Morgan Kaufmann Publishers, 2001.

[7] Https:// www.wikipedia .org.

[8] Chaudhuri, S. and U. Dayal, "An overview of data warehousing and OLAP technology". ACM Sigmod record, 1997. 26(1): p. 65-74.

[9] Y. Li, H. Lu, and R.H. Deng, “Practical Inference Control for Data Cubes,” Proc. IEEE Symp. Security and Privacy, Extended Abstract, pp. 115-120, 2006.

[10] L. Wang, S. Jajodia, and D. Wijesekera, “Securing OLAP Data Cubes Against Privacy Breaches,” Proc. 25th IEEE Symp. Security and Privacy, pp. 161-175, 2004.

[11] Sung, S.Y., et aI., "Privacy preservation for data cubes",Knowledge and Information Systems, 2006. 9(1): p. 38-61.

[12] Hua, M., et aI., "FMC: An approach for privacy preserving OLAP.Data Warehousing and Knowledge Discovery", 2005: p. 408-417.

[13] Wang, L., D. Wijesekera, and S. Jajodia, "Cardinality-based inference control in data cubes". Journal of Computer Security, 2004. 12(5): p. 655-692.

[14] Wang, L., S. Jajodia, and D. Wijesekera, "Preserving privacy in on-line analytical processing data cubes". Secure Data Management in Decentralized Systems, 2007: p. 355-380.

[15] Alfredo Cuzzocrea, Domenico Saccà “Balancing Accuracy and Privacy of OLAP Aggregations on Data Cubes”, ACM,2010