International Academic Institute
for Science and Technology

International Academic Journal of Science and Engineering
Vol. 2, No. 8, 2015, pp. 36-48.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

A new method to protect privacy in data cubes

Fariba Hakim nezhad[a] , Mansoor Amini lari[b]

[a] *A student of MAof software group, Fars researches sciences, Islamic Azad Uiknversity, Marvdasht, Iran.*

[b] *Professor assistant of computer group, Marvdasht, Islamic Azad University, Marvdasht, Iran.*

Abstract

Toprotect privacy while providing analysis on information is one of the most important issues in online analytic processing systems. One of the challenges is to prevent derivation of sensitive cases through non sensitive accumulated data. This article presents and proves a new algorithm to prevent deriving due to hiding extra data and also data itself which these extra data are necessary and efficient. based on this, this approach can provide the most information to the user and protect security. This strategy doed not effect on online running in online analytic processing system. Conducted analysises and emperical comparison indicate efficiency and possibility of forbidden minimal covering.

Keywords: data mining, online analytical processing, privacy protection, data cube

Introduction:

Today privacy protection in databases and online systems discussion is an important and essential one. Because these information systems maintain thousand people. To enter into this concept it is necessary to have more knowledge on these systems. Analytical information systems are systems against operational systems that make it possible to analyze mass data resulted from operational systems for all levels of users. Operational systems in big organizations such as banks that daily doing much information processing and produce various information are used. These organizations information banksare faced to a lot of data obtained from financial transaction, official, accounting and Accure and precise analysis and processing of operational information Can help produce statistical results to making effective managerial macro decisions and to managers to make optimal decisions for their organization success. To analysis and processing this information and facilitate and accelerate reporting operation and different inquiries rather than data direct analysis from within the operational systems, the analytical databases and systems are used that are outside the operational systems area and are of high speed (Cuzzocrea, 2012).

Analytical database are prepared from different data resources of an organization and or even several associated organization. This database provides a suitable bed to summarize data recorded into operational systems and independent and integrated and are made available to the managers to extract information properly(Giannotti et al., 2013). Online analytic processing (OLAP) is one of the most popular decision supporting techniques in commercial intelligence systems. However, this is a big challenge to analyze privacy data without violating privacy of the data owners. Online analytic processing and data mining with successful extraction of information provides required knowledge to use on various areas sucg as marketing, meteorology, medical analysises and national security, but still there is no guarantee by which can evaluate special data,., without violating to that information owner privacy (Agrawal and Sirkant, 2013). For example, in a medical system, how to do analytical processing on the patients privacy information without disclosuring that information is one of the issued faced with. Some organizations such as health securities organization and examining health status in United States and data management organization and analytical systems in European union with regard to the occurred sensivities on this have created a set of compelling rules on data management and systems analysis.

These kind of concerns increase parallel to developing using data analysis systems. This article presents a new algorithm to prevent deriving bu hiding extra data and the sensitive data itself and proves that this extra information is both necessary and efficient. Based on this, this approach can provide the most probable information to the user and protect privacy. Online analytic processing is an important infrastructure to advanced data analysis and knowledge discovery. While most of the previous researches on Online analytic processing have focused on models of this type of processing, making data cube and datawares, maintaining and density methods and efficient methods of responding to research. However, among these issues, the important issue is that the privacy protection problem to responding to Online analytic processing inquiry is examined (Ioannidis et al., 2014). To clarify this matter and to understand the matter importance an example is presented below.

Example 1- consider table 1 that indicates patients information in some hospitals

Table1- patients tables

Hospital	disease	number of patients
Forest	lung cancer	16
Forest	Diabetes	63
Memorial	Diabetes	87
Memorial	Heart attack	32

Figure1 also expresses data cubes derived from table 1.

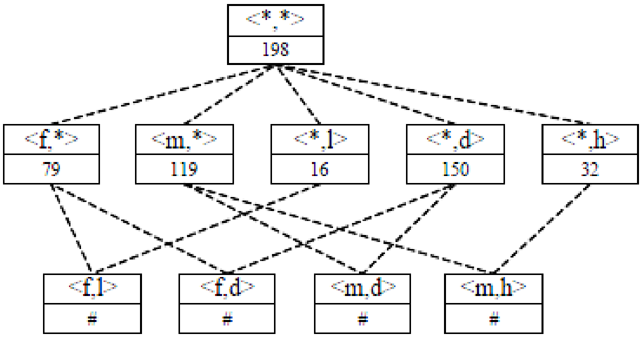

Figure1- data cubes based on table 1

Supposed that the hospitals don’t want to announce individual diseases population generally but they agreed to share total number of all cases in a hospital or total number of a special disease in all hospitals. This is the case that in data cubes based on table1, value of cells <f,1>, <f,d>, <m,d>, <m,h> should be hidden and not exposed to the users. A simple and direct policy is to decline all accesses to sensitive cells. However such a policy is unsufficient to protect privacy. Since just some parts of evaluation is hidden, cube structure can be discovered outside of other reality table columns and accordingly, sensitive values can be detected from other unprotected cells. For example,value <f,d> is exactly equal to <I,*> because <I,*> makes dense this record. . in addition to subtract <f,1> from <*,f> , makes <f,d> clear. Now matter is that wheter wecan create a better security policy that protect privacy strongly? In addition, we need such a policy that hide less information as possible. We call this privacy preserving OLAP problem. We introduce in this article a method to soledthis problem by hiding minimal set of unprotected cells that contribute to sensitive and confidential cells and accordingly information leakage prerequisite no longer will be created.

For example if we hide <I,*> and <h,*> in figure1, the sensitive cells value<m,d>, <f,1>, <m,l> and<f,d> never will be obtained by accessing to other unprotected cells. Here, there are some newchallenges to preserve online analytic processing problem privacy compared to difficulties in controlling privacy in statistical databases and data mining and we have some works on this.

1-Sensitive data items can be distributed in various granularity levels in online analytic processing.we suggest a general model and a solution that can manage this.

2-To online analytic processing it is important to provide more information for the users while maintaining sensitive data. We prove that our algorithm just hide necessary data.

3-online analytic processing operation usually need short response time. We avoid derivation before the user interacts to the system and accordingly the algorithm doesn’t effect on running online analytic processing online.

Proposed method framework

A data cube consists of a set of dimensions and measures with aggregate functions that defined on it. We have concentrated on SUM function in this article.

Every node of data cube is called cuboid and each tuple in cubiod is a Cell as well. Two cuboids C1 and C2 follow a partial order on each dimension, if and only if each of them shares an identical property or C2 is of higher level in dimension hierarchy. We say in this case that C2 is an Ancestor of C1 andC1 is Descendant of C2. If C1$\leq C2$ and here there is no cuboid C such as C1$\leq C$ and C$\leq C2$ we say that C2 isfather of C1 and correspondingly C1 is son of C2.

These definitions are applying to cells as well. In example 1, cuboids <hospital, disease>were <hospital,*> and cells <f,i> are smaller or equal to <*,f>.
Access control in online analytic processing systems decided based on multi dimensional data model is in cuboids and cells. We define confidential and sensitive information as a forbidden set {c1,…,cm} where ci is data cube cell. We suppose that the forbidden set comprises all confidential and sensitive cell and their descendants because a confidential and sensitive cell can be calculated by all descendants simple sum.all cells that are not in the forbidden set constitue Available set that is available to the users for example the available set in figure 1 comprises all cells except <f,l>, <m,d>, <f,d> and <m,h>. however we have indicated that some confidential information can be obtained by incorporating cells into the available set. We define the available set and information taken from that as Available set closure.

Definition no. 1(Available set closure)
Considering Available set as A the Available set closure(A) defined as below:

1- if cell $c \in A$ then cell is $c \in C(A)$
2- if $c \in C(A)$ then $k*c \in C(A)$ that is a real number.
3- If cells $c2 \in C(A)$ then $c1+c2 \in C(A)$.

When limitation of the available set and forbidden set has commonality Inference occurs. In this case we say the forbidden set compromised. Cells in the available set caused inference is known as Source of the inference.

Definition no.2 (compromisability)
Having data cube L and forbidden set information in L , information endangers when $C(L- F) \cap F \neq \emptyset$
To prevent compromisability we hide some cells into the source that the sensitive cells cant be calculated through uncomplete source. Whereas the hidden cells can be inferred by cells with higher grain and based on this more cells should be hidden to protect them. Finally, we can find a set of cells near the forbidden set and no cells there can cause to infer internal cells.
Elements of a Paper.

Definition no.3: Minimal cover MC
Considering data cube L and a forbidden set information in L a set systemis defined as minimal cover information as MF© with following conditions

$S \subseteq L-F$
$C(L-F-S) \cap (F+S) = \emptyset$
$\forall\ S' \subset S,\ C(L-F-S')\ ) \cap (F+ S') \neq \emptyset$

The first condition represents that the minimal cover isa subset of the available set. The second condition represents that after hiding the minimal cover, residual cells don’t cause to infer the minimal set and the forbidden set. The third condition is that any subset of the minimal set cant satisfy the second condition that guarantees that all existing cells are necessary to remove inference and are indispensable. The main goal of method presented in this article by considering data cube as L and thr forbidden set as FI online analytic processing privacy preservation problem is minimal cover for F, (MCF) that avoids compromisimg information and in this state bans accessing to the least information as possible.

Overview on the online analytic processing privacy preservation
From definitions above it is clear that the minimal cover should be free of inference for the forbidden set and itself. Otherwise one can disclosures information initially by inferring the minimal cover values and then accessing to the forbidden set. A subset of the minimal cover that is just free from inference for the

forbidden set is called the minimal partial cover. We take two steps to find the forbidden set minimal partial cover and then expand it to the minimal cover to preserve absolute security.
Step 1- to find the minimal partial cover for the forbidden set. Wefind the minimal partial cover for the forbidden set by Linear system Theory within which the minimal partial cover hiding removes all inferences directed to the forbidden set and here hiding each one of the minimal partial cover is not feasible.
Step 2-expanding minimal partial cover to minimal cover: we then consider the minimal partial cover found in step 1 as the forbidden set and repeat the minimal partial cover for the newly found cells until there is no longer need to hide another cell.

To find minimal partial cover
We examine in this part that how we can find a minimal partial cover for a forbidden set. First we present the vector code for each cell in cuboid as following.

Definition4 (vector code) : considering a cuboid C, vector code $\bar{c}$ for cell c in C or C father cuboids is defined in form of (a1,…,an), where analytic is the number of cells in C. In this condition, ai is equal to 1 if c∈ C and otherwise ai is equal to 0, in other word ai is equal to 1 if c aggregates ith cell in C.
c∈ Father(C) and otherwise is equal to 0. Above problem has been shown below.

$$a_i = \begin{cases} \text{` } c \text{ is the } i_{th} \text{ cell in } C_{(c \in C)} \\ \cdot \text{ otherwise} \end{cases} \text{ or } a_i = \begin{cases} \text{` } \text{if } c \text{ aggregates the } i_{th} \text{ ce in } C_{(c \in Father(C))} \\ \cdot \text{ otherwise} \end{cases}$$

For example in cuboid < hospital , disease> in figure 1, cell vector code <*,f> is (1,1,0,0) and the vector for <f,l> is (1,0,0,0). Cell corresponded to $\bar{c}$ can be inferred by c1,…,cn if the vector codes $\bar{c}n1\bar{c}$ can incorporate linearly into vector code $\bar{c}$. To determine whether this happens or not we examine three solutions of equation 1(x1,…,x are real number).

$$X_1 \times \bar{c}1 + \ldots + X_n \times \bar{c}n = [\bar{c}1, \ldots, \bar{c}n] \times [X_1, \ldots, X_n]^T = \bar{c}$$

Equation 1 has no solution. Cell c related to $\bar{c}$ can't be calculated with any other cells and accordingly there isno need to hide another information.
Equation 1 only has a non- zero solution. $\bar{c}$ can be calculated from certain combining with $\bar{c}n1\bar{c}$,…,. If xi,xj,…are non-zero components of the solution, then corresponded cells $\bar{c}i, \ldots, \bar{c}j$ are avoidable for inferring $\bar{c}$. Based on this only hiding one of the $\bar{c}i, \ldots, \bar{c}j$ can avoid inference.
Equation 1 has more than one non-zero solution. To avoid all inferences it is necessary that we hide a cell which its cell corresponded to solution X always be non-zero. If there is no such cells we need to find a set of cells at least in one of the cases used in each solution. We have created method based on linear system to avoid inferring certain cells. This example has been indicated in example 2.
Example 2: we seek to find minimal partial cover for cell <f,d> in example 17 and here security necessities are as before. Suppose following conditions.
c_1=<f,*>, c_2=<m,*>, c3=<*,l>, < c4=<*,d>, c5=<*,h>

Corresponded vectors here are 5,…,$\bar{c}1\bar{c}$.

1-we create equation based on making b=$\bar{c}$, A=[$\bar{c}1, \ldots, \bar{c}5$].

$$AX=\begin{bmatrix}1&0&1&0&0\\1&0&0&1&0\\0&1&0&1&0\\0&1&0&0&1\end{bmatrix}\times X=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}$$

2- equation 3 solution is X=X0+k*X1 that here X0=[1,0,-1,0,0]T, X1=[-1,-1,1,1,1]T and k is also a real number. If ithcomponent of X is non-zero then $\bar{c}$ *is used to calculate* <f,d>. as an example if we consider k=0 then X=[1,0,-1,0,0]T that is exactly the same case shown in the figure1.

3- We seek to find a component of X that is always non zero or find a set of components that at least one of them in each X is non-zero.

If X=X0,:k=0 , the first and third components is non-zero.

If k$\neq 0$: by cautiously choosing a value for k the first or third component can be zero but other components never will be zero. So, a cell exists in {c1, c3] and another one in {c2, c4, c5} from minimal partial cover <f,d>.

Introduced algorithm

Now we generalize the minimal partial cover presented below:

Input: forbidden set information and given cuboid of C

Output : a minimal partial cover (MPC) from F

Method:

1- Creating matrix of coefficient A={$\bar{c}$1,…, $\bar{c}$n}

2- For each cell c in F

3- If Ax=$\bar{c}$ there is some solutions

4- Find X Solutions for Ax=$\bar{c}$.

5- Find set of components Mc that at least one of

6- of Ax=0 and X0 is a certain solution to Ax=c. here independent components r exist in X that take 0 in x0 and 1 in each (xi, i=1,…,r) respectively. As an example in picture 18 last them in each X be non-zero.

7- Return MPC= UcC∈ information MC.

Based on a forbidden set information in cuboid C first create coefficient matrix A using unprotected cells in C or C fathers. Then for each cell c in information if Ax=c has some solutions find set of components into the solution such that at least one of them in each X is non-zero. Here we use linear system theory to find such cells. Solution of Ax=0 can be presented as x=x0+[x1,…,xr]*[k1,…,kr] where x1,…x is basic solution three components are independent. Suppose that X2[i] and X0[i] is non zeroin all ith components of X0 to X3 and X2[j] is an independent componentcomprising 1 in Xq2. So each one of X[i] or X[j] isused and corresponded cells are minimal partial cover.

$$\overset{X}{\begin{bmatrix} \# \\ X[i] \\ \# \\ \# \\ X[j] \\ \# \end{bmatrix}} = \overset{X_0}{\begin{bmatrix} \# \\ X_0[i] \\ \# \\ 0 \\ 0 \\ 0 \end{bmatrix}} + \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \times \begin{array}{c} \begin{array}{ccc} X_1 & X_2 & X_3 \end{array} \\ \begin{bmatrix} \# & \# & \# \\ 0 & X_2[i] & 0 \\ \# & \# & \# \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{مستقل} \qquad \text{independent component}$$

Figure 2- an example of minimal partial cover

is given. (r+1)th from nth component of X is an independent component. If X0[i]$\neq$ 0 and only if Xd1[i],…,Xdj[i] from X1[i],…,Xn-r[i] is non-zero then :

1- At least one of the components X[i], X[r+d1],…,X[r+dj] will be non zero in X.

2- Each subset of components X[i], X[r+d1],…,X[r+dj] can ba all zeroin X.

Lemma 2: algorithm 1 returns minimal partial cover from forbidden set FS.

Now,we create a surface framework for expanding minimal partial cover to minimal cover for each cuboid with two optimization strategy as following.

Remove single children presumption:

A cell if and only if has a desecendent in its son cuboid it is called single gensis cell.single gensis fathers of the forbidden set are certainly sensitive. In example 1 if we hide two single gensis cells <1,*> and <h,*> all inferences are removed. Based on this we add to our algorithm all single gensis fathers of sensitive cells to minimal cover. This can cause to elaminatemajor part of inference and reduces cells that we should examine for inference.

To find candidate domain: we examine all unprotected fathers and ancestors of the forbidden cells for inference in algorithm 1. However all of them are not dangerous.

Example 3: a 2 dimensional cube has been shown in figure 3. Cell <a2,b1> that marked with sterick in cuboid <A,B> is sensitive. We create coefficient matrix A for cuboid <A,B>. Two vectors of column A are related with 8 father cells and 5 unprotected cells in cuboid <A,B>. however just the column vector A[10], A[1], A[2], A[5], A[6] and A[9] is of inference probability <a2,b1> since other cases have zeros in corresponded components. We call matrix formed by A[10], A[1], A[2], A[5], A[6] and A[9] and their non zero components candid domain of the forbidden set. Candid domain can be found by putting it into father cells of the forbidden set and then adding repeatedly into cells are shared with candid domain.

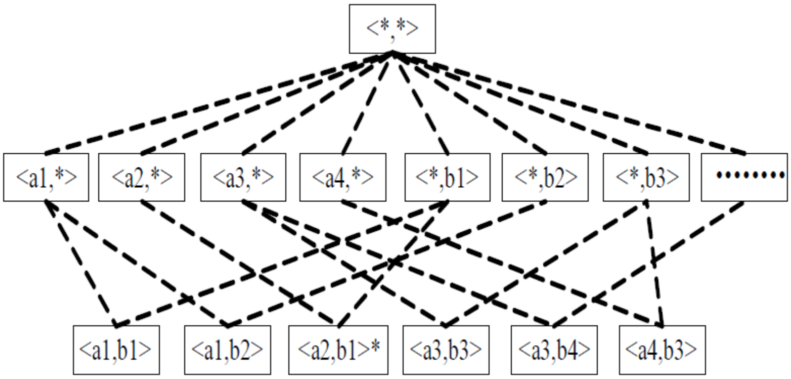

Figure 3- 2 dimensional data cube

$$A=\begin{array}{r}<a_1,b_1>\\<a_1,b_2>\\<a_2,b_1>^*\\<a_3,b_3>\\<a_3,b_4>\\<a_4,b_3>\end{array}\begin{bmatrix}1&0&0&0&1&0&0&0&1&0&0&0&0\\1&0&0&0&0&1&0&0&0&1&0&0&0\\0&1&0&0&1&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&1&0&0&0&1&0&0\\0&0&1&0&0&0&0&1&0&0&0&1&0\\0&0&0&1&0&0&1&0&0&0&0&0&1\end{bmatrix}$$

Figure 4- coefficient matrix for <A,B>

Introduced algorithm is in following. We have used a surface framework to expand minimal partial cover to minimal cover.

Input: forbidden set FS

Output : MC of FS

Method :

For each given cuboid c* in cuboid until : FS∩C*≠ Ø

Add single gensis father toMC

Find CR for FS

m)m=FMPC(FS,CR) is minimal partial cover of FS that returned by FMPC.

Return MC.

Algorithm 2- FMC: a surface algorithm to find a minimal cover

As indicated in algorithm 2 we initially have ranked cuboids in cube in terms of being granularity ascendant order. Then we have used two optimalization strategies for each cuboid and cited to algorithm 1 to find minimal partial cover of forbidden set in this cuboid. The minimal partial cover should be further examined. This process should be repeated until there is no minimal partial cover in the current cuboid. Algorithm 2 returns a minimal cover of the forbidden set.

Results:

Method proposed in this article expressed in the previous part in detail. To evaluate the proposed method, it has been done and tested.

All the experiments have been conducted on a PC Pentium4 2.84 GHz with main memory 512Gb and windows 7. The algorithm has been applied using Borland C++ Builder 6 with Microsoft SQLServer 2000. We have used a set of compound data and real data set with standard TPC-H for our experiments. In compound datasets we created data from a Zipfian distribution where z on 0,1,2,3 was variable. Data sets sizes varied from 20000 to 80000 cells with 3 dimensions and 4 granularity levels in one dimension. In this article some comparison occur on various parameter of Zipf. We applied forbidden minimal cover for benchmark TPC-Hand datasets that their parameters were z=0, 1, 2, 3. We chose randomly 2 cuboid as as forbidden set and compared extra cells hidden by the introduced method FMC and SeCube. Figure 5 indicates these results. when z=0 and data are distributed uniformly, less extra cells need to hide in compared to data along with Skewed Data. Since some dimension values reveal less in Skewed Data these scatter data are the main source of reference.

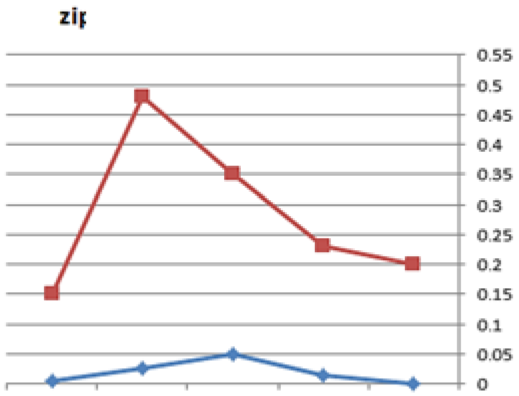

Figure 5- protected cells size diagram in cube in term of forbidden set

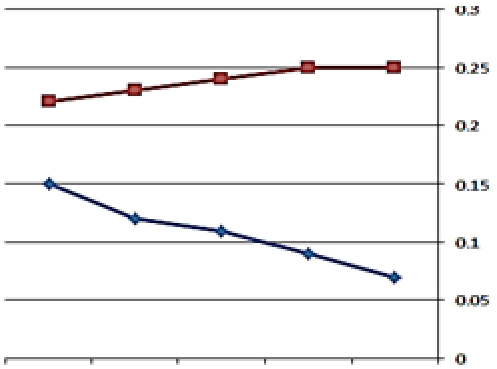

Figure 6- hidden extra cells size in term of forbidden set

We have conducted some experiments to state the proposed method quality in terms of zipf to evaluate different factors.

Where parameter zipf put as z=1 and changed the set size. Figure 6 shows the extra cells size hidden by secube and the forbidden minimal cover that here forbidden minimal cover in compared to secube hides less cells in all cases.

This picture examines protected cells size in cube in terms of forbidden various set size against secube method. As clear from the figure the introduced method is of suitable quality. In addition to this another comparison is necessary to evaluate candid size in terms of various percent of the forbidden set size. Figure 7 shows candid domain on different size of the forbidden set.

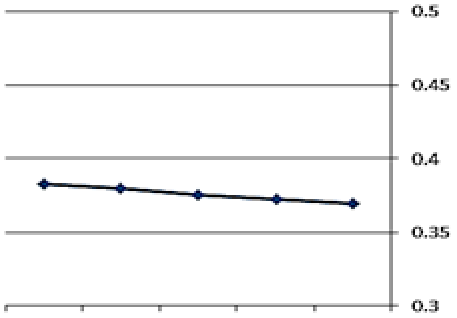

Figure 7- candid domain size diagram in terms of forbidden set.

As it is apparent from diagram presented in the figure the candid domain size remains below 40% in all cases and this means that we need to just 40% of all cube for inference. We also have examined running time of the introduced method for different size of the forbidden set. Diagram presented in figure 8 represents this comparison.

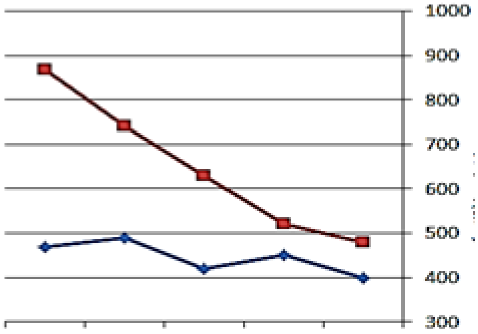

Figure 8- diagram of running time in terms of forbidden set percentage

Conclusion

In this article we have created an effective and efficient algorithm to refer to online analytic processing privacy preservation problem. Main idea is to hide a part of data that caused to inference and accordingly, other sensitive information can't be calculated. We can guarantee that all hidden information are necessary and based on this the most possible information can be provided for users. All of these tasks are conducted before user interacts to the system and so, this will not be affected on running online analytic processing system online. Our algorithm is partially based on linear system theory and based on this the accurate problem can be proved. Exprimental results also can show the algorithm efficiency. Works related to future are on applying this method for other aggregating functions and improving algorithm efficiency. We also have planned to develop this work to inference problem solving that created by comprising two aggregating functions in a cube.

References:

A. Cuzzocrea, Overcoming Limitations of Approximate Query Answering in OLAP,2012.

B. Ding, M. Winslett, J. Han, and Z. Li. Differentially private data cubes: optimizing noise sources and consistency. In SIGMOD, 2011.

B. Gedik and L. Liu, Location Privacy in Mobile Systems: A Personalized Anonymization Model, Proc. IEEE Int. Cof. Distributed Computing Systems, pp. 620-629, 2009.

Bertino, E., Fovino, I. N., and Provenza, L. P." A framework for evaluating privacy preserving data mining algorithms". Data Mining and Knowledge Discovery, 11(2), 2005, 121-154.

C. Tai, P. S. Yu, and M. Chen. k-support anonymity based on pseudo taxonomy for outsourcing of frequent itemset mining. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010, pages 473--482, 2010.

C. Y. Chow, M. F. Mokbel and X. Liu. A, Peer-to- Peer Spatial Cloaking Algorithm for Anonymous Location-based Services, Proc. ACM Int. Symp. On Advances in Geographic Information Systems, pp. 171–178, 2010.

D. Boneh, B. Waters, Conjunctive, subset, and range queries on encrypted data, Proceedings of the 4th conference on Theory of cryptography, February 21-24, 2014.

F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and W. H. Wang. Privacy-preserving mining of association rules from outsourced transaction databases. IEEE Systems Journal, 7(3):385--395, 2013.

F. Liu, K. Hua, and Y. Cai, Query l-diversity in location-based services, Proc. the 1st. Int. Workshop on Mobile Urban Sensing, 2010.

G. Ghinita, P. Kalnis, and S. Skiadopoulos, PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems, Proc. World Wide Web, pp. 237- 246, 2009.

Ge, X., L., Zhu, J., and Shi, W. "Privacy-preserving distributed association rule mining based on the secret sharing technique. In Software". Engineering and Data Mining (SEDM), 2010 2nd International Conference on 2010, June, pp. 345-350. IEEE.

I. Ioannidis, A. Grama, M. Atallah, A Secure Protocol for Computing Dot-Products in Clustered and Distributed Environments, Proceedings of the 2002 International Conference on Parallel Processing, p.379, August 18-21, 2011.

J. J. Gardner, L. Xiong, F. Wang, A. Post, J. H. Saltz, and T. Grandison. An evaluation of feature sets and sampling techniques for deidentification of medical records. In IHI, 2010.

J. Rizvi, R. Haritsa, Maintaining data privacy in association rule mining, Proceedings of the 28th international conference on Very Large Data Bases, p.682-693, August 20-23, 2014, Hong Kong, China

K. Chaudhuri, C. Monteleoni, D. Sarwate, Differentially Private Empirical Risk Minimization, The Journal of Machine Learning Research, 12, p. 1069-1109, 2011.

M. F. Mokbel, C. Chow and W. Aref, The New Casper: Query Processing for Location Services without Compromising Privacy, Proc. Int. Conf. onVery Large Data Bases, pp. 763–774, 2011.

M. Hardt, N. Rothblum, A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis, Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, p. 61-70, October 23-26, 2010

N.Matatov, L.Rokach, , and O. Maimon, "Privacy-preserving data mining: A feature set partitioning approach". Information Sciences, 180(14), 2010, 2696-2720.

R. Agrawal, R. Srikant, Privacy-preserving data mining, Proceedings of the 2000 ACM SIGMOD international conference on Management of data, p.439-450, May 15-18, 2013.

V.Ciriani, C., Foresti, S., and P.Samarati. "k-anonymity". In Secure Data Management in Decentralized Systems 2007, pp. 323-353. Springer US.

Y. Chang, M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data, Proceedings of the Third international conference on Applied Cryptography and Network Security, p.442-455, June 07-10, 2014.