Indian Journal of Science and Technology, Vol 9(22), DOI: 10.17485/ijst/2016/v9i22/90927, June 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Multi User Profile Orient Access Control based Integrity Management for Security Management in Data Warehouse

G. Thangaraju[1]* and X. Agnise Kala Rani[2]

[1]Department of Computer Science, Karpagam University, Coimbatore - 641021, Tamil Nadu, India; gtrmsu@gmail.com
[2]Department of Computer Applications, Karpagam University, Coimbatore - 641021, Tamil Nadu, India; agneskala72@gmail.com

Abstract

Background/Objectives: The Aim of this research work to demonstrate that the security enhancement of data warehousing, the methods face major challenges in integrity management and have the responsibility to restrict the unauthorized user. **Methods/Statistical Analysis:** There are many access control methods discussed earlier for the problem of integrity management, and some of them have been discussed using the user profiles. Still they suffer from the problem of efficiency in integrity management. **Findings:** To overcome such issues, in this paper a multi-user profile orient access depths measure based integrity management is proposed. The method maintains a set of Meta data which keep track of data objects in a hierarchical manner according to their importance and sensitivity of the data. The method first identifies a set of objects being specified from the input query and the sensitive tree; the method computes the access depthness measure. The access depthness is computed based on the level of objects being called and the access level the user has been given and the number of objects the user has access. **Application/Improvement:** Based on computed access depths measure user query is being processed, and the method improves the performance of integrity management.

Keywords: Access Depth Measure, Access Control, Data warehouse Security, Integrity Management, Multi User Profile

1. Introduction

Integrity management is the process of maintaining such security to the data or information from the data warehouse. This research paper focuses on improving the efficiency of integrity management and in what way the safety can be enforced in an efficient manner.

Access control in the data warehousing environment can be enforced in many ways. The simple password-based approach is not enough for the management of large warehouse where to exist a large number of users. Different users in the environment have different access rights, and the system has to restrict them according to their rights given. The user profile is the Meta data about the access rights of different users and using such Meta data the system can identify whether the user has the rights to read the data object. Here the difference in the query depends on the request as read or write. In some situations, there will be rights for the user to read the data base and in some other there will be both read and write. On the other side, some of the data tables_have no rights to read or write. Still the user of the organization can be restricted in attribute level so that the user or client will be allowed to access a few number of attributes but not some others. Such operation can be performed based on the user profile.

*Author for correspondence

For any given request from the user, the method can identify a set of data tables needs which are to be accessed. Using the data profile and the list of tables needed to access, we can compute the multi-attribute access depths measure. Using computed measure the user can be restricted from accessing the data warehouse in an efficient manner.

2. Related Works

In[1], presents a method of sequence enrichment based on an encryption method which conserve the data type of the plaintext source. This technique be mainly responsive for multifaceted data warehouse environment. The first processing steps involve replacing each plaintext character in the string by an integer that symbolizes its position, or index, within the selected alphabet. This amount is along with zero and one fewer than the total number of characters in the alphabet. If a plaintext character is not in the legal alphabet, it is copied to the output and detached from the string to be encrypted.

In[2], presented an approach for corpus-based text or character classification based on WHIRL, a DBS (Database System) that augments traditional RBT (Relational Database Technology) with textual similarity operations developed in the IRC (Information Retrieval Community). Not only does the approach carry out competitively when compared to modern text categorization methods, but we also show that it enables the assimilation of a range of up till now unexploitable sources of info into the organization process in a justlyforceful and in general fashion.

In[3], proposed a data mask technique for protecting perceptive business data in DWs that balance security strength with database performance, using the method i.e., modular mathematical operator. It also empowers a fake data inoculation technique for confusing attacker & cumulative the overall sanctuary strength. This thechnique can be easilyexecuted in any of the DataBase Administration System (DBAMS) and patently used, lackingchange to submission source program. Investigational estimations using a real time DW and TPC-H decision fundingstandardapplied in leading commercial DBMS Oracle 11g and Microsoft SQL Server 2008 demonstrate its overall effectiveness. At the final conlusion the implementation cost is high the infromation should be in secure.

In[4], presented the concept, originate the Enterprise Data warehouse market increasingly aggressive, as illustrated by tighter clustering of top Companies. Oracle,Teradata,SAPthen IBM lead by contribution high-performance, in all aspects. Teradata delivers the greatestfeatures based on today's Enterpise Data warehouse. The Oracle consumes new databse mechanism named Exadata it meats the EDW requirements. The product Sybase, lately acquired by SAP, remains to enhance IQ's massively parallel columnar data for real-time analytics. IBM has ramped up its Bigdata and EDW solution focus and now sets the step on petabyte-scale Hadoop addition. SAP is fast evolving and joining BW and BWA into a high-performance EDW through an in-memory, columnar organization enhanced for real-time analytics. EMC Greenplum proves solid execution and sustained invention. The IBM product Netezza has integrated Microsoft has hurledmoney-spinning EDW appliance for midmarket and large initiative, and Strong Performer Vertica Schemes continue to enhance its high-performance all-columnar EDW architecture.

In[5], Best Practices, offers best performs for with Oracle Forward-thinkingSecurity TDE (Transparent Data Encryption). It runs the ability to encrypt complex application data on storage medium entirely translucent to the submission itself. TDE addresses encryption requirements connected with communal and private confidentiality and protection mandates such as PCI and California SB1386. Oracle Progressive Security TDE support encryption was announced in Oracle Database 10g Release 2, it enabling the major applications.

In[6], Demonstrate the DWH security implemented by the benchmark with the tool of data generation, in this tool takes the bulk queries as input and produces the result with the database instances. This bench mark also provides the testing tools like an application testing and database security. It also provides the evaluation results of TPC-Work loads.

In[7], proposes an log based security with the implementation of examination based on logs. To maintain data privacy, various answers have been future and proven effective in their security purpose. However, they familiarize significant overheads manufacture them unfeasible for the data warehouse. Therefore in the direction of avoid these outlays and to increase data sanctuary, data masking method have been proposed. The solution accomplishes the haphazardness of cloaked values which increases the overall security strength.

In[8], deliberates that the last numerous ages should be proprietary by means of global businesses made up of massive folders contain processer users search

investigations and sites visited; administration agencies accruing subtle data and inferring information from unspecified data with little inducement to deliver countries habits of correcting false data; and persons who can easily syndicateopenly available data to derive info that – in former times – was not so willingly accessible. Security in data warehouses develops more important as dependable, and appropriate safety mechanisms are obligatory to attain the desired level of confidentiality defense.

In[9], proposed a data security based on masking technologies for protecting sensitive business data in the data warehouse environment that balances security strength with database performance, using a formula based on the modular operator for mathematical worker. Our solution can perceptible arbitrariness and transmission of the masked values while announcing small storage space and query achievement time overheads. It also empowers a false data injection method for confusing attackers and increasing the overall security forte. It can be smoothly realized in any DBMS and clearly used, deprived of variations to request programe code.

In[10], proposed a novel approach for modeling Structured Query Langues statements to apply machine-learning techniques, such as outlier detection or clustering, to detect malevolent behavior at the catalog transaction level. The method incorporates the describe tree structureof Structured Query Language queries as representative e.g., for linking SQL queries with applications and individual benevolent and malevolent questions. The practicality of our approach onreal-world datais established.

In[11], proposed a method with the making use of cryptographic algorithm provided by Shamir's for security enhancement, which has some advantages including security of the admin, the security based on client-side aggregation. It privileges that security is upheld even after N or more waitpersons colud. We provide that much exploration has been done to ensure the safety of the single mist and cloud server while multi-clouds have customary less consideration in the area of security. We insist the affecting to multi-clouds due to its ability to diminution security risks that affect the cloud adding a user. The key conclusion is that wished-for work provides discretion, data integrity, improved availability, and magnitude to handle multiple requirements at a time.

In[12], presented a concept which is amixture of information hiding with cryptography, it is used to control main cloud server. It ensures data honesty and service obtainability. The self-possession of STORM is that it requirements code implementation in their server systems, and it does not contract with numerous version of data.

In[13], Redundant Array of Cloud Storage (RACS) is a method for Intercloud storeduring the day of 2010. This system is alike to RAID and usually used by recordings and file schemes, and imitation offers better fault forbearance. But the problem is incapable to work together with vendor shut in and financial failure. Presented a plan for Intercloud storage space named ICStore in 2010. ICstore is client centric dispersed protocol which can handle data integrity issue but has poor presentation in case of data interruption and service ease of use. Same thing happened with encrypted cloud VPN.

In[14], presents security mechanisms to access the PHR based on ontology, the mechanisms incorporate with the HSP and CSP, the security methodology used in both service providers is ABE. His contribution in security of PHR based on with the following properties, 1. Object Properties 2. Data Properties he provides the OWL Structure for data and object properties. Finally he concluded, in future OAC for PHR might be extended by inference engine. The conclusion of endorsement based upon logics and conditional data which are defined in the regulation. It is necessary that the strategy itself to be defined as ontology constructed with certain regulations. The conditional data might be accessed by using SPARQL query language in the PHR systems.

In[15], presents analysis reports of data mining tools and techniques for information retrieval, the tools taken for analysis are Rapid Miner, Weka, and R Programming, and also the techniques used for this analysis are classification, clustering, neural networks and genetic algorithms. The planned method implement a normalization concept for analysis and new technique to analysis the tools and techniques. The planned technique puts support an extent to retrieve data in much competent way as compared to others. This proves to be valuable in the cases where brilliant information retrieval is required. Enlargement of technique which provides privileged accuracy in the accessible ones is sought afterwards.

In[16], proposed a model for successful management of customer relationships with the context of Business Intelligence. The three main hypotheses were confirmed. They are IT (Information Technology), knowledge organization and supervisory context based on BI (Business Intelligence), were identified as three self-determining variables in victory of CRM (Customer Relations Management), conclusion can be drawn that BI has a

direct and optimistic effect on success of CRM, and realization of BI in organizations is essential requirement for their more and more achievement and propensity toward customer-orientation.

In[17], proposed a system introduces a new plan in which all the mobility functionalities and the HA functionalities are incorporated as a sole unit as DAP in the access level itself. Security Analysis is done to the paper with the Confidentiality and integrity, Authentication and prevention methods are used as follows Man-in-the-middle Attack Prevention, Replay assault Prevention, and fake BU Attack Prevention. The performance investigation of the proposed and existing scheme is based upon the full signal cost.

All the methods discussed above have the problem of ensuring security and restricting access in the data warehouse and produce higher false classification.

3. Methodology

3.1 User Profile Orient Access Control based Integrity Management

The multi-user profile based integrity management approach reads the input query and identifies a set of relational objects or data objects being required to complete the query. The method maintains the Meta data in the form of a tree structure, and it maintains different data table in different level. Based on identified data tables and tree the method computes the access depths measure to restrict the user from unauthorized access. The entire process can be split into different stages namely Query preprocessing, Access Tree Generation, multi-attribute access depths measure computation and Validation. This section discusses each of the stages in detail.

3.1.1 Query Preprocessing

The method takes the input query and performs parsing of the input query. From the parsed result, the method identifies the set of data tables mentioned in the query. Also, the method identifies the list of functions being mentioned in the query. Also, the method identifies the list of attributes being accessed by the function or the query. All these information are passed to the multi-user profile based multi-attribute access depths measure computation. Based on the result from the MAADM value, the method decides the access result of the user query.

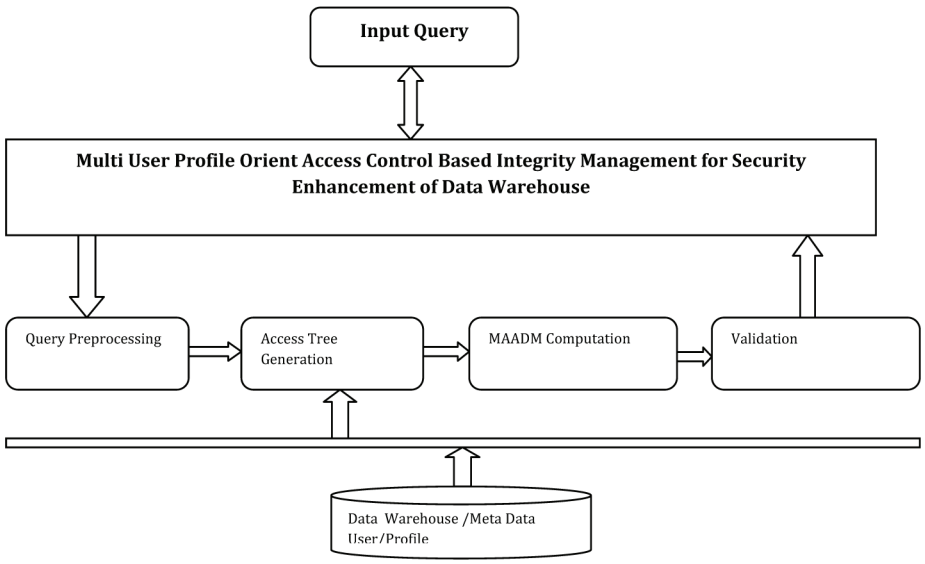

Figure 1. Architecture of the planned method and the functional components.

Algorithm

Input: Data Warehouse Metadata Md, Query Q

Output: Result Rs.

Start

Read metadata Md.

Read input query Q.

Initialize Table set Tas.

Initialize Functions Set Fs.

Initialize Attribute set As.

Key set Ts = $Ts = \int Parse(Q,'')$

For each term Ti from Ts

Identify the data table name.

If $\int_{i=1}^{size(md)} Md(i).name == Ti$ then

Add to table set Tas = $\sum Tables \in (Rs) + Ti$

End

If $\int_{i=1}^{size(functions)} Function.name == Ti$ Then

Add to function set Fs = $\sum Functions(Md) + Ti$

End.

Identify list of attributes accessed.

Attr set As= $\int_{i=1}^{size(Attr(Tas(i)))} \sum AttributesRequired$

End

MAADM set Ms = Compute Multi-Attribute Access Depthness Measure (Tas, Fs, As).

For each level l

If Ms(i)>As then

Process the query and send result

Else

Drop the query.

End

Stop.

The algorithm displayed above performs preprocessing of input query and identifies the set of data tables required and set of functions accessed and then identifies a list of attributes. Using all the above the method computed the multi-attribute access depths measure and based on the value of MAADM value the method restrict the user from accessing the system.

3.1.2 Access Tree Generation

In this stage, the method generates the access tree from the Meta information available. The method maintains a different level of access modes by placing the attributes at different levels. First the method creates the root node, and the first level contains leaf of attribute nodes which has open access. According to the number of attributes which has public access, the method creates some nodes and each assigned with different attributes of different data tables. Similarly, according to the Metadata information, the method creates some levels according to the importance of attributes. If the method classifies the attributes as threelevels, then there will be three levels in the tree. The number of levels can be extended up to any level.

Pseudo Code of Access Tree Generation:

Input: Meta Data Md

Output: Access Tree At.

Start

Create Tree T.

Create Root node Rn.

Add Rn to T.

`For each data table Di from Dm

Identify all attributes.

As = $\int_{i=1}^{size(Dm)} \sum Attr \in Dm(i)$

For each attribute Ai from As

Identify level l.

L = $\int_{k=1}^{size(level)} Md(k) \in Ai$

Create Node N.

Initialize N with Ai.

Add to tree T.

T = $\int \sum (Nodes \in Md(k)) \cup N$

End

End

Stop

The above-discussed algorithm generates the access tree which will be used to compute the MAADM value.

3.2.3 MAADM Computation

The method reads the access tree and the user profile with the term set, function set and attribute set. Using all this, for each level the method computes the access depths measure based on some attributes at any level has been required, and some attributes in any level are given. The lookup is performed from the user profile and computes the access depthness measure for each level. The computed value will be given for query processing.

Pseudo Code forMAADM Computation:

Input: User Profile Up, Access Tree At, Attribute set As.

Output: MAADM set Ms.

Start

For each level l from tree AT

Extract total Attributes.

$$LA = \int_{(i=1)}^{(size(nodes))} \Sigma Node.Attr \in At(i)$$

Compute a number of attributes.

NA = size(LA).

Identify number of attributes required.

$$Nr = \int_{i=1}^{NA} \sum Attr(As) \in LA$$

ComputeMulti Attribute Access depthness measure MAADM.

$$MAADM = \frac{Nr}{NA}$$

Add to Ms.

End

Stop

The algorithm discussed above computes the multi-attribute access depthness measure which will be used to restrict the user from unauthorized access.

3.2.3 Validation

At this stage, the method performs all the operations mentioned above by using the procedures. First, the method performs query preprocessing and then generates Access tree using the Meta data. Second, the method MAADM computes depths, and based on the value returned, the method executes the query for the user and returns the results to the user. The results to the user are fully based on the values of function sets, attribute sets, obtained on the query submitted which are computed depending up on the customer profile and depthness of the client query.

4. Results and Discussion

The planned method has been designed and implemented using the SQL data base which has a number of relational databases. The warehouse has been created with thousands of relational database and has been evaluated from the user query in lacks. The details of evaluation has been listed below:

Table 1 shows the details of simulation parameters being used to perform an evaluation of the proposed approach.

Figure 2, shows the efficiency of integrity management produced by different methods and it shows obviously that the proposed technique has produced efficient integrity management than other technique.

Figure 3, shows the comparative analysis of query processing produced by various techniques and it shows visibly that the MAADM method has produced more effectiveness with the other methods.

Table 1. Details of simulation parameters

Parameter	Name
Data Warehouse Tool	SQL
Number of databases	7000
Number of queries	110000
Number of users	650

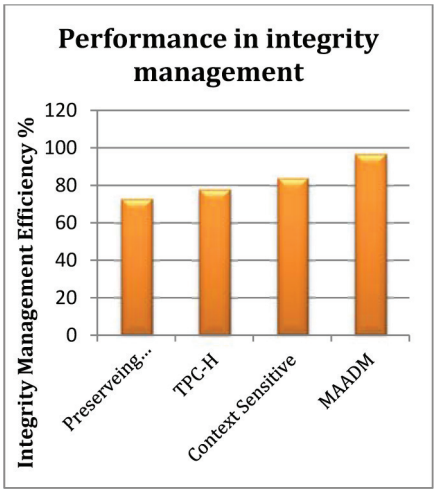

Figure 2. Comparison of integrity management efficiency.

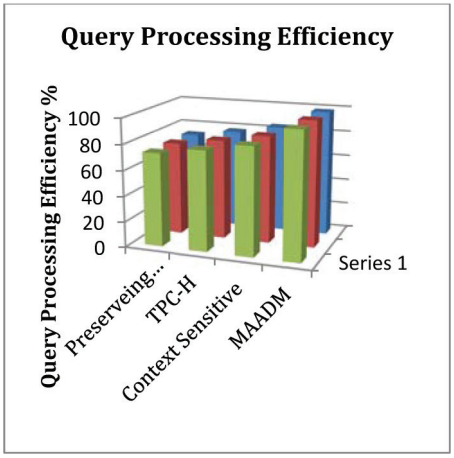

Figure 3. Comparison of query processing efficiency.

5. Conclusion

An intelligent user profile orient access control method is proposed for the integrity management of data warehouse systems. The proposed method improves the performance of the query processing and improves efficiency also.

6. References

1. Reddy MS, Reddy MR, Viswanath R, Chalam GV, Laxmi R, Rizwan MA. A schematic technique using data type preserving encryption to boost data warehouse security. IJCSI International Journal of Computer Science Issues. 2011 Jan; 8(1).
2. Hirsh H. Integrating mulitple sources of information in text classification using whril. Snowbird Learning Conference; 2000 Apr.
3. Barbosa M, Farshim P. Randomness reuse: Extensions and improvements. 12th Institute of Mathematics and its Applications (IMA) Int Conference on Cryptography and Coding; 2009.
4. Kobielus J. The forrester wave: Enterprise data warehousing platforms. Forrester Research Report, Q1; 2009.
5. Oracle Corporation. Oracle advanced security transparent data encryption best practices. Oracle White Paper; 2010 Jul.
6. Lo E, Cheng N, Hon W. Generating databases for query work loads. Int Conf on Very Large Data Bases (VLDB); 2010.
7. Amritpal S, Umesh N. Implementing log based security in data warehouse. International Journal of Advanced Computer; 2013.
8. Weippl ER. Security in data warehouses. IGI Global. Data Ware Housing Design and Advanced Engineering Applications. Ch: 015; 2010.
9. Santos RJ, Bernardino J, Vieria. Balancing security and performance for enhancing data privacy in data warehouses. International Joint Conference of IEEE Trust Com-11/IEEE ICESS- 11/FCST -11; 2011.
10. Bockermann C, Apel M, Meier M. Learning SQL for database intrusion detection using context sensitive modeling. Int Conference on Knowledge Discovery and Machine Learning (KDML); 2009.
11. Mirajkar S, Biradar SK. Using secret sharing algorithm for improving security in cloud computing. IJARCST. 2014; 2(2):395–8.
12. Bowers KD, Juels A, Oprea A. HAIL: A high availability and integrity layer for cloud storage. CCS'09: Proc 16th ACM Conf on Computer and communications security; 2009. p. 187–98.
13. Goodson GR, Wylie JJ, Ganger GR, Reiter MK. Efficient by zantine-tolerant erasure-coded storage. DSN'04: Proc Intl Conf on Dependable Systems and Networks; 2004. p. 1–22.
14. Mohan K, Aramudhan M. Ontology based access control model for healthcare system in cloud computing. Indian Journal of Science and Technology. 2015 May; 8(S9):213–7. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.
15. Verma A, Kaur I, Singh I. Comparative analysis of data mining tools and techniques for information retrieval. Indian Journal of Science and Technology. 2016 Mar; 9(11). DOI: 10.17485/ijst/2016/v9i11/81658. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.
16. Shahraki A, Dezhkam A, Dejkam R. Developed model of management of successful customer relationship in the context of business intelligence. Indian Journal of Science and Technology. 2015 Dec; 8(35). DOI: 10.17485/ijst/2015/v8i35/68800. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.
17. Mathi S, Lavanya M, Priyanka R. Integrating dynamic architecture with distributed mobility management to optimize route in next generation internet protocol mobility. Indian Journal of Science and Technology. 2015 May; 8(10):963–74. DOI: 10.17485/ijst/2015/v8i10/58213. ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.