ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

The Method of Dynamic Access Control against Data Leakage and Data Misuse

Mayuri M. Gawade, Prof. P.S. Desai

M.E Student, Dept. of Computer Science, Pune University, Pune, India

Professor, Dept. of Computer Science, Pune University, Pune, India

ABSTRACT: Insiders may be the employees, business partners, service providers in an association eligible to access database of organization to fulfill tasks provided to them. There may be a possibility of mistreat of exposed data by the insiders. Around 26% of attacks are originated by insiders is surveyed by Cyber Security Watch (CSW). Associations are much anxious about misuse of crucial information from insiders and trying to mitigate misuse of vital data from insiders. This paper mainly focuses on how to mitigate misuse of crucial information in stored relational database of organization from insiders. This paper proposes concept, M-score which calculate extent of harm or loss to the organization if specific information goes in the incorrect hand and design applications by using M-score to reduce misuse of data from insiders. This paper proposes application of M-score, M-score based Dynamic Access Control (MDAC) which is real time dynamic access control system for insiders in the organization.

KEYWORDS: Data misuse, M-score, Misuse-ability weight, RDBMS.

I. Introduction

Data available in the association is the main plus point for the organization. Data is accessed by the insiders i.e. Employees, contractors, business partners in the association to accomplish tasks provided to them. At the same time there may be the possibility of misuse of data from insiders. If Insider may not have correct behavior he may disclose sensitive data to competitor organizations. Important data may contain data or information about clients, customers, and statistical data of the organization which may be misuse and leak by insiders which indirectly leads to loss or significant damage in business of the organization.

As per the surveys and studies major percentage of information is misused by insiders of the association. The 26% of attacks are originated by insiders is surveyed by Cyber Security Watch (CSW) [1]. From 26% of the attacks, 16 percent were caused by theft of sensitive data and 15 percent by disclosure of confidential data.

By providing limited access to the data or information to insiders is not good solution for association as due to this if legitimate insider cannot access data from the database then he cannot gives his best for organization. Main objective of this paper is to how to stop or minimize misuse of the data stored in database (Tabular format) from insiders in the organization. Many schemes have been proposed till date to address data misused from insiders. Many traditional systems based on user profiles that describe behavior of user and alerts whenever user’s behavior significantly diverges from the normal profile. Mainly there are two schemes for data against misuse from insider, first one is data centric and second one is syntax centric. According to data centric approach, data accessed by user is analyzed to detect or prevent misuse of data. In syntax centric approach data requested by user i.e. queries submitted by user to the user are analyzed to avoid or detect misuse of the data. Previous system does not address the extent of harm to the firm if particular information giving to the legitimate user is misused. This paper presents novel concept called as Misuse-ability weight of information or M-score i.e. possible degree of loss in business of firm if information given to insider is misused. We can apply M-score of information to prevent the misuse of the data from insiders. Paper proposes application of M-score name M-Score Based dynamic access control, which is access control system to information in the firm based on idea of Misuse-ability weight of the information

DOI: 10.15680/IJIRCCE.2016. 0401135

II. Related work

Many approaches have been developed to reduce data leakage and data misuse in database systems. Traditional methods can be classified into the data centric and syntax centric. According to the scheme called data centric, data accessed by user is evaluated to detect or prevent misuse of data. As per the syntax centric approach data requested by user i.e. queries submitted by user to the user are analyzed to mitigate or detect misuse of the data.

First method is “Detecting Anomalous Access Patterns in RDB”, proposed which is syntax centric approach. In this method, SQL-expression representation of queries is used to construct user profiles. When user submits query to database, features are mined from query and stored in the vectors and these vectors are then utilized to detect anomalous patterns.

Second is Data- centric approach to corporate executive Attack Detection [3] in info Systems presents a information-centric approach and considers a query’s expression syntax as root for discerning user intent; solely the resulting data matters. For each access to info, a statistical vector (S-Vector) is prepared holding numerous statistical details on the result-set information, like minimum, maximum, and average for numeric attributes, or counts of the various values for text attributes.

In “Insider Threat Prediction Tool Evaluating the probability of IT Misuse” G.B. Magklaras used taxonomy of corporate executive threats to compute the Evaluated Potential Threat (EPT) [4] live. This measure tries to estimate whether or not a user’s action is correlate with an area of the taxonomy so as to\'s labeled as malicious. The EPT is calculated by considering features describing the user, the context of the action and also the action itself. Additionally, the tool uses a group of malicious actions that were antecedently discovered.

Celikel et al. in “A Risk Management Approach to RBAC,” [5] introduce a Risk and decision Analysis gift a model for risk management in distributed info systems. The model is employed to determine the danger poses by a user so as to forestall insiders from misusing or abusing insider’s role privileges. Within the model, a Risk Priority number (RPN) is evaluated for each user, that is the product of the incidence Rating (OR) that reflects the quantity of times a similar question was issued with relevancy different the opposite users within the same role; the Severity Rating (SR) that measures the danger by pertaining to the standard of the information the user would probably get from the queries insider issued; and therefore the Detection Rating (DR) indicates however shut the behavior of the user is to the behavior of users in other roles. To forestall insiders from misusing their privileges, Bishop and Gates [6] recommend the Group-Based Access control (GBAC) mechanism that is a generalization of RBAC method. This method uses, additionally to the user’s basic description (role), the user characteristics and activity attributes like the time insider usually involves work or the customers with whom insider sometimes interacts.

Above all proposed methods described have not been considered the sensitivity level of the data to which the user may be exposed. This factor can greatly affect the outcome when trying to figure out the potential damage to the association if the data is leaked or misused. We are using data centric approach in which M-score of the result set of submitted query is analyzed to mitigate misuse of the information.

III. Proposed algorithm

To prevent misuse of crucial statistical data stored in relational database of an organization, we compute extent of damage (M-score) to association if particular data is misused by specific insider in specific context and to apply calculated M-score to design dynamic access control system for mitigating misuse of data.

IV. Proposed Work

(A) Misuse-ability Weight of the Data:

Data stored in an organization’s computers is very crucial and embodies the core of the organization’s power. A corporation doubtless desire to conserve and retains this power. At other side, this information is vital for daily work processes. Users among the organization’s perimeter such as employees, subcontractors can execute a variety of actions on this information such as query, report, and search, and will be exposed to sensitive info embodied among the info they access. In an attempt to see the extent of damage to associate organization that a user will cause using the data

DOI: 10.15680/IJIRCCE.2016. 0401135

insider has obtained, we initiate the thought of Misuse-ability Weight. By assignment a score that shows the sensitivity level of the info that a user is exposed to, the misuse-ability weight will conclude the extent of damage to the organization if the info is misused. Utilizing this info, the organization will then take accurate steps to forestall or reduce the damage

(B) Evaluate M-score of the data:

To assigning M-score to data based on the way data is stored and offered i.e. whether the data is in the tabular format or text format. This is one of the reasons why it is not possible to design generic function to estimate of M-score of given data. Our work focuses on data in tabular format i.e. Relational Database. This section describes steps to calculate M-score of data. Before starting actual steps we will some important keywords.

Table T [i, j]: T is the table of the data with dimensions no. of records[R] i and no. of attributes [A] j. As per our context mainly there are three types of the attributes namely Quasi-identifier attributes, Sensitive attributes, other attributes Quasi-identifier (Q) attributes. These attributes can be linked, possibly using an outer data source, to protect a particular entity that the particular information is about. For example First name, last name, mobile number etc. in the record is the possible fields to identify the entity in the table. Sensitive attributes (S) are the attributes that are used to assess the risk resulting from exposing the data to the legitimate user. Sensitive attributes are mutually excluded from quasi-identifier attributes. Consider example of the hospital database, then name of the patient, address are the quasi-identifier attributes but type of the disease, bill amount are the sensitive attributes. Other attributes are the attributes which are not helpful to calculate approximately misuse-ability of the data.

(C) Sensitivity Score Function:

The sensitivity score function f: C × S [0, 1] allocate a sensitivity score to each possible value x of S, according to the specific context C in which the table was exposed to particular legitimate user. Sensitivity score function is defined by assistance of the organization expert or domain expert such as manager or executive of the firm who knows correct sensitivity of the type data stored in the organization. This function is mainly domain dependent because firm to firm stored data differ and if firms are of same domain then also sensitivity of data may change as per organization. Context factor used in the Sensitivity function based on context in which data is exposed. For example context factor depends on when the data is requested whether it is working hour, break time, holidays.

1) Calculate Raw Record Score:

At this stage record is elected from table and sensitivity function is applied to the each sensitivity attribute [x] in the record and then sum of all sensitivity weight s performed then least amount of sum or 1 is taken as raw record score.

$RRS_i = \min(1, \sum_{Sj\ \varepsilon\ T} f(c, Sj[x_i]))$

2) Calculate Record Distinguishing Factor (D):

By utilizing the characteristic issue, the M-score integrates the individuality of the quasi-identifier's value within the table when weighting its misuse-ability. The DF measures to what degree a quasi-identifier reveals the precise entity it represents (e.g., a customer). It assigns a score vary between the [0,1], once the lower the score is, the tougher it\'s to differentiate one entity from another, given this quasi-identifier. In other words, the DF of record i specify the trouble a user can got to invest so as to seek out the precise entity insider is probing for .Usually, the DF is not easily acquired, and therefore we use the record distinguishing factor (Di) as an approximation.

3) *Calculate Final Record Score (FRS):*

In this step, uses RRS and D of record calculated in the above two step.
Given a table with r records, RS is calculated as follows:

$$RS= \max_{0\le i\le r} RS_i = \max_{0\le i\le r}\left(\frac{RRS_i}{D_i}\right)_i$$

Where RRS is Raw Record Score, Di Disinguishing Factor. Fraction of RRS and D of each record in the table taken andmaximun of them taken as the Final Record score.

4) *Calculate M-score:*

At this stage we consider quantity of the data to exposed, that is no of records in the table. M-score of the table is calculated as follows. Given a table with r records, the table’s M-score is calculated as follow

$$MSCORE = r^{\frac{1}{x}}\times RS = r^{\frac{1}{x}} \times \max_{0\le i\le r}\left(\frac{RRS_i}{D_i}\right)_i$$

Where r is the number of records in the table, x is a given parameter and RS. We apply a settable parameter x. This parameter sets the importance of the quantity factor within the table’s final M-score. The higher we set x, the lower the effect of the quantity factor on the final M-score.

V. Application: M-score Based Dynamic Access control System (MDAC)

In this we proposed new scheme called misuse-ability weight or M-score. Here we discuss the application of M-score name M-score based Dynamic Access control system. In this novel access control system we utilize M-score factor as base for access control to the data for insiders in the organization.

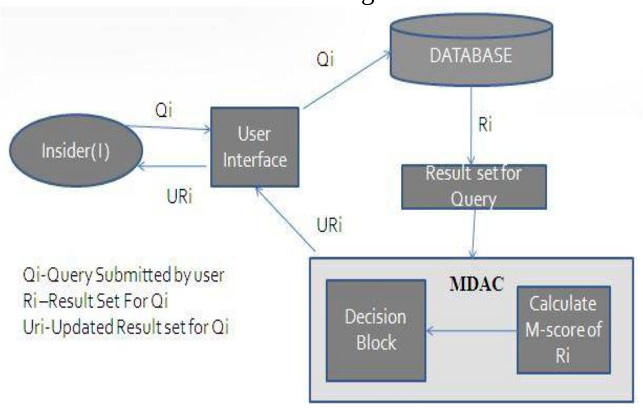


Fig 1 System Architecture MDAC

(A) *Working of MDAC*:

1) Enroll Insider in the system: At this step insider who frequently access data stored in the firm database is enrolled to the system and threshold M-score is allocated to the insider, which is the maximum M-score value of data he can access. Username and password is provided to the each user.
2) When user required to access data first and foremost thing for user is to login. Then user submits query to database normal result set is produced with respect to the query.
3) M-score of the result set is calculated as per context as described in the above section.

4) Decision Block checks m-score of the result set calculated in the step 3 and threshold of the user taken from step 2 when user logins.
5) If Threshold of user is greater or equal to the M-score of the result set then access is approved to the user. If Threshold is less than the M-score of the Result set, then mode in which system is working is checked Binary or Subset Closure.
6) If the system is working in the Binary mode then the access is starved against the requested data.
7)If system is working in the subset closure mode then ,most sensitive attribute column in the table is chosen and detached from result set to prevent the M-score of the result set and new result set is generated. And steps are repeated from step no 4 till M-score of result set is less or equal than threshold of the user. Above procedure is shown in algorithm as follows.

VI. Details of Dissertation Work

A) *Mathematical Model:*

1) Get result set from database
Input: Query (Q) submitted by user of the system (S)

Process: Get result set (RS) from database (DB) for valid query (Q).

Output: RS for query (Q) from database (DB)

2) MDAC
Input: Result set (RS) for query (Q) and Threshold M-score (TM) value of User (U) who submitted query (Q)

Process: For RS
Calculate Raw Record score (RRS) of each Record (i) in RS

RRS_i = min (1, $\sum_{Sj\ \varepsilon\ T}$ f(c, Sj[x$_i$]))

Where S is sensitivity score, x is attributing in the RS, C is the context factor

Calculate Final Record score (FRS) for RS using
RRS as follows

RS= $\max_{0\le i\le r} RS_i$ = $\max_{0\le i\le r}\left(\frac{RRS_i}{D_i}\right)_i$

Where RRSi is a raw record score of row in the resultset (RS) , Di is distingusing factor
Calculate M-score of the result set as follows:

Mscore = $r^{\frac{1}{x}}\times$RS

Where r is the number of records in the result set (rs) and x is mscore parameter.

If Mscore of RS is less than TM then give RS to U,
Else check mode of the MDAC IF MODE is binary
then message “Not allowed” to U IF MODE is subset closure mode then attrribute with h ighest mode are removed untill M-score of RS <= TM, and Updated result set (URS) is given to U.

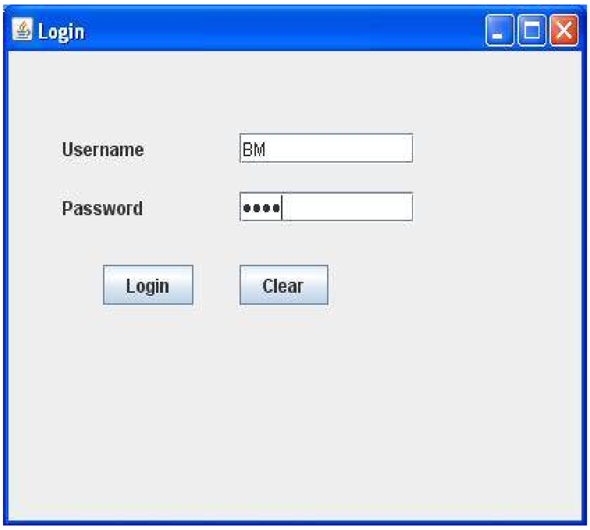
Output: RS or URS or Message”Access denied”

B) Algorithm of MDAC :

```
Input :M-score of RS Table And Threshold M-score Value of  the User
    If  (M-score of Rs table >= Threshold)
        then
        {
            if (Mode : Binary)
               then
                   {
                     Display Message ”Access denied”
                     }
            else
                  {
                     Remove Most Sensitive Data Till
                     M-score of RS Table<=Threshold
                     Show Subset Of RS(i.e. Appropriate RS)
                    }
     }
     Else
         {
          Show Result Set To the user
                       }
```

VII. IMPLEMENTATION DETAILS

1. Login Window:

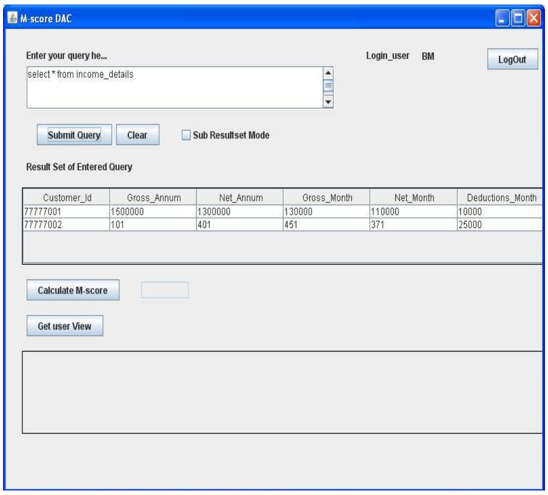


Username BM

Password ••••

Figure2: Login Window.

2. M-Score Calculation:



select * from income_details

Login_user BM

Customer_Id	Gross_Annum	Net_Annum	Gross_Month	Net_Month	Deductions_Month
77777001	1500000	1300000	130000	110000	10000
77777002	101	401	451	371	25000

Figure 3: M-Score Calculation.

VIII. EXPERIMENT AND RESULT

In our research, we developed MDAC for housing finance organization. Sensitivity score function is developed by utilizing the organization expert. After complete implementation of MDAC for housing finance we test results given by MDAC from organization expert. To test MDAC, we login to housing finance database as specific user (u) and queries (q) db for Result Set (Rs). If users (u) m-score threshold value is greater than or equal to the m-score of Result Set (Rs) then user will get access to Result set and if M-score is more than threshold then user will not grant access to result set in binary mode. In experiment we use machine with processor Intel dual core, RAM of 1 GB, operating system used is windows XP. We used MySQL 5.1 database as back end for housing finance application, SQLyog, and Swing as front end with J.D.K. 1.7. We test 20 queries per user (There are 4 different users with different user) and Result Set given by MDAC to that specific user from Expert of housing finance. MDAC is successful if result set given by MDAC matches with what result set expected by organization expert. Result of the experiment is shown in table 1. Total result sets of 80 queries are tested from organization expert. 72 out of 80 result set of queries are matched with expert expectation. It means developed application MDAC is 90% accurate.

Table 1: Result of Performed Experiment

User	No. of Results matches with expert expectation.	No. of Results matches with expert expectation.
Sanction Officer	18	2
Disbursement officer	19	1
First Identification agent	16	4
Property valuation agent	19	1

IX. CONCLUSION

In this paper we mostly direct attention on the misuse of sensitive data from insiders in the association. We studied traditional methods to prevent misuse. In this paper we have proposed one new scheme called misuse-ability weight of data, which is helpful to mitigate misuse of data from insiders in the firm. Then we proposed novel and efficient application namely M-score based on Dynamic Access Control System, which is in turn depends upon the misuse-ability weight concept.

X. FUTURE WORK

We proposed M-score function to evaluate misuse-ability weight which is association dependent, so in future there is possibility of upgrading the M-score function to construct it organization independent.

Until now there is no such model designed for M-score function for particular organization, so in future there is possibility to extend generic model for developing M-score function.

REFERENCES

[1] Cyber Security Watch Survey, http://www.cert.org/archive/pdf/ecrimesummary10.pdf, 2013.
[2] A. Kamra, E. Terzi, and E. Bertino, “Detecting Anomalous Access Patterns in Relational Databases,” Int’l J. Very Large Databases, vol. 17, no. 5, pp. 1063-1077, 2008.
[3] S. Mathew, M. Petropoulos, H.Q. Ngo, and S. Upadhyaya, “Data- Centric Approach to Insider Attack Detection in Database Systems,” Proc. 13th Conf. Recent advances in Intrusion Detection, 2010.
[4] G.B. Magklaras and S.M. Furnell, “Insider Threat Prediction Tool: Evaluating the Probability of IT Misuse,” Computers and Security, vol. 21, no. 1, pp. 62-73, 2002.
[5] E. Celikel et al., “A Risk Management Approach to RBAC,” Risk and Decision Analysis, vol. 1, no. 2, pp. 21-33, 2009.
[6] M. Bishop and C. Gates, “Defining the Insider Threat,” Proc. Ann. Workshop Cyber Security and Information Intelligence Research, pp. 1- 3, 2008.