Open access Journal International journal of Emerging Trends in Science and Technology

Malicious Agent Detection In Multi Party Data Access Structure

Authors

Prof.DevendraShamkuwar[1].Miss Narode Arti V [2].Mr. Shendage Rahul B[3] Gaikwad Swapnil R.[4]

Dept of Information Technology of Pune University.

Sharadchandra Pawar College of Engineering, Otur (Pune)

Email:*devendra8319@gmail.com, rahul.sh7190@gmail.com,*

Abstract — A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some amount of data is leaked and found in an unauthorized machines (e.g., on the internet or computer). The distributor has to findout from whom that data get leaked. We propose data allocation strategies (across the agents) that increase the chances of detecting leakages. These techniques do not need to modify data of the released data (e.g., watermarks). In some cases we can also embed real but fake data records to further increase our chances of detecting leakage and identifying the guilty agent. Our goal is to detect when the distributors sensitive data has been leaked by agents, and detect the agent that leaked the data. Perturbation is a very useful method where the data is modified and made less sensitive before being handed to agents. we develop secure techniques for detecting leakage of a set of objects or records.

Index Terms – Data leakage, Detection Guilty Agent, Distributor.

I. Introduction

In today's technically empowered data rich environment, it is a challenge for data holders to prevent the leakage of data. Loss of large volumes of confidential information has become regular headline event, which force the companies to re-issue cards, inform customers and mitigate loss of goodwill from negative publicity. While considering the protection of company's electronic assets from outsider threats from intrusion prevention systems to firewalls to vulnerability management organizations now turn their attention to an equally dangerous situation: the problem of data disclosure by the insiders. Whether its email, instant messaging, webmail, a form of website, electronic communications exiting the company still go largely uncontrolled and unmonitored on their way to their destinations with the ever present potential for confidential information to fall into wrong hands. The importance of digital watermarking for digital assets such as relational databases to preserve their copyrights is becoming more and more important as time goes by. In the last few years, a large number of techniques have been implemented for hiding copyright marks on relational databases.

In this paper, we present an watermarking technique for relational data that is secure from different attacks . While previous techniques have been mainly concerned with introducing errors into the actual data, our concept is new data that are not original and we says them "fake" data, to the relation as watermarks. We will says that our concepts leads to an effective technique that is secure against various forms of non-maliciousattacks .

Watermarking is a form of information hiding with a goal of preserving the copyright of the digital asset such as multimedia, software, and database. There are many discussions on what the characteristics of a robust digital watermarking are. For instance, Petit colas et al have proposed the following characteristics for a robust watermarking Marks should not degrade the perceived quality of the work, if multiple marks are inserted in a single relational database, then they should not interfere with each other, if various copies of an object are distributed with various marks, then different users cannot do anything.their copies in order to makes a new copy that identifies none of them, and the mark should survive all attacks that do not degrade the work's perceived quality. Despite the differences on characterizing robust watermarking, they all share a common idea. The key idea behind any sort of digital watermarking is to introduce imperceptible (so that the attacker can not detect them) and tolerable (to ensure that the value of data is not greatly depreciated) errors to the object.

II. EXISTING SYSTEM

Generally,data leakage detection is handled by watermarking, e.g., a unique code is added in each distributed data. If that data is later got to an unauthorized person , the leaker can be detected. Watermarks concept generally used in images, video and audio data. The main techniques of Watermarking to identify a owner of image or audio or data and, hence, is subject toattacks where a pirate claims ownership of the data or weakens a merchants claims.

In the watermarking, watermark on the data ,image ,audio video can be easily removed , so identification of the owner get destroyed .

And data is distributed to agent for analysis , processing purpose but if that data copy is leaked then detection of that data and leaker of data is very difficult because they were probably detected .so distributor is not sure about agent and data leakageidentifies.

III. PROPESED SYSTEM

Our goal is to detect when the distributors sensitive data has beenleaked by agents, and to identify the agent that leaked the data . When malicious agent leak distributors data to unauthorized party for benefit then it is difficult to identify that leaker or agent of that data .So we are implementing this system .in this system we using allocation strategies for distributing data to agent and in this allocation we are adding some fake object as real in data. This will increase our chances of detecting malicious agent.

In this system we are downloading data which is on domain , after that we using extraction process for extracting downloaded that data and by using DOM tree we get original data at leaf node of tree. Then that data is compared with original database if data is leak then we get leaked data and If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be sure
that agent was guilty. so detecting of data leakage easily findout.

IV. SYSTEM IMPLEMENTATION

In this paper we are using Data Allocation Strategies for distributing data to agents . for identifying data leakage we are using extraction process in proposed system. In extraction process we are using there steps .

1)Data Allocation Strategies :

In this strategies Data can be allocated or distributed
To agents in two ways as follows.

Explicit Data Request:

In the explicit request, agent is interested in some objects. So data is distributed according to agent or user request instead of randomly distribute, For e.g Employee salary in between 50000 to 200000. In this as per agent request data is send .

Sample Data Request:

In sample data requests agents are not interested in particular objects. In this distributor distribute data to agent randomly.The distributor is "forced" to allocate certain objects to multiple agents only if the number of requested objects exceeds the number of objects in set T. The lots of data objects the agents request in total, the more agent on average an object has got ; and the more objects are distributed among various agents, the more difficult it is to identify a guilty agent.

Adding of fake object:

The distributor able to add fake objects to data .to increase the chances in detecting guilty agents. By adding fake object in data, we will see what agent can do with that data. The idea of perturbing data to detect leakage is not new. In many times data perturbation is not possible .Because if if we are take data of Employee salary then perturbation of data is not possible.

2) Detection Module :

For leakage detection we are using here three techniques are as follows

Web Data Extraction System :

Web data extraction is a system which automatically extract data from web pages with modifying content and sent or pass it to database or other application. In this step data downloaded from domain and that downloaded data is in HTML format .so the data is in unreadable format so that data is parse into tokens for parsing that data we are using HTML parser.

Document object Model :

The Document Object Model (DOM) is a cross-platform for representing and interacting with objects in HTML documents. By using DOM , tree is made that is called DOM Tree. The nodes of every document are organized in a tree structure, called the DOM tree, whose topmost node is the *Document* object. When an HTML page is rendered in a browser, the browser downloads the HTML into local memory and parses it, using the DOM to construct the internal data structures employed to display the page in the browser window. The DOM is also the way in which JavaScript sees the state of the browser and the current HTMLpage.

Depth –First Search:

Depth-First Search is an algorithm which is used for traversing or searching tree. In it we are traversing from first root node to its child node then again we are going to its deeper and deeper node until we get desired or leaf node.

In this system we are using DFS algorithm on DOM tree for finding original data which always at leaf node of the tree.In DOM tree topmost node is Document object which we are considering root node for DFS .At the leaf node of DOM tree we get actual data.
Then at the end we are compare extracted data with original data in database . if in that data we get some fake object then distributor is sure about data is leak .then distributor can check in database , who have given that fake object and according to that malicious agent identified.

3) Modules :

Distributor Module .
Add Customer
Manage Customer
Release Database.

ii) Agent Module:
Data Analysis
Leak Data

Detection Module :
Download Data
Generate Tree

Distributor Module :

Add Customer :
In this distributor can add their customer details .for their organization purpose .

ManageCustomer :
In this if distributor wants to modify , update , delete customer information then that can b e done .

Release Database :
In this Data is distributed to the agent for processing or analysis on that data .the data can be distributed to the agents in two ways . one is Explicit Request Another is Sample Request
.

Agent Module :
In this there are two types of agents . Malicious AgentAnd Non malicious Agent.

Analysis & processing Data:
Non-Malicious agents are those who canprocessing on data or analysis on data .

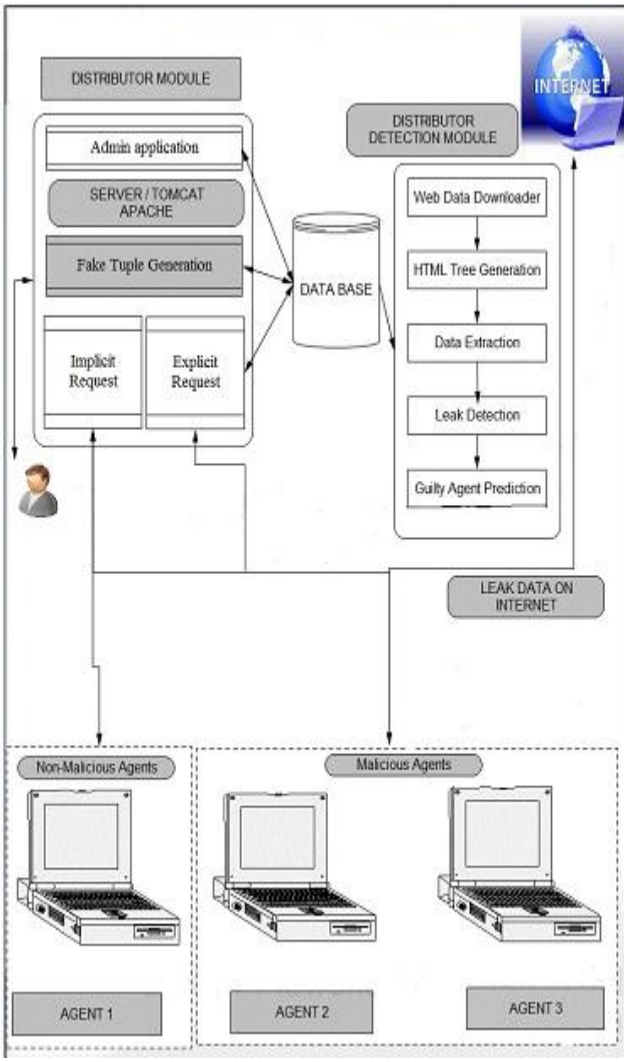
Leak data :
Malicious agents are those who canleak that data by unauthorized person .

Detection Module :

Download Data :
Distributor can download data from domain For leakage detection .

Generate tree :
Distributor can generate tree of that data the compare that data with its original database .if data is leaked then match downloaed data and data leakage found .

Architecture Diagram



Mathematical Model

Let S = { D, FT , FGT() , IR(), ER(), WDE() , D', DE()}

Where,
S= Sis the set of all functions and Entities which comes in system.
D= D is the original database that contains set of tables T.
FT= FT is a set of false tuples.
FGT()= Pseudo Random Function.
IR()= IR is Implicit Request
ER()=Explicit Request.
WDE()= Web data extraction function
D'= Database which is put be test.

FT = FT is set of falsetupelos.
(ft1, ft2,……ftn) € FT
R = FGT(time speed)

Where ,
R is set of Random indices generated by the Pseudo Random Function.

Ds = IR(D1)

Implicit Request

Ds=ER(D)

Explicit Request

D' is Database to be test

D'= WDE()

DFS Function
DFS(G, V)

Where (G, V) graph *G* and a vertex *v*
WDE()= Web Data Extraction

DE (D, FT, D')
{
Return true ;
Agent Guilty
}
{
Return False;
Agentnot Guilty

}

N Sets so the Time Complexity is **O(N).**

This Gives full output so this system is P-Complete problem

V. FUTURE ENHANCEMENT

In future we can add functionfor Audio,Video data, E-mail filtering .in this system.

VI. CONCLUSION:

The data distributor Strategies improve the distributor's chances of identifying a leaker . It has been shown that distributing objects can make significant difference in identifying guilty agents , especially in cases where there is large overlap in the data that agents must receive. In Some cases real but fake data records are added to increase the chancesof detecting leakage and identifying the guilty agents. And here we download the data which leak and compare thatdat a with original database so leaker of data can be identify easily.

VI. References

1)A .Santhi Lakshmi , Asso.Prof.A.Bhaskar , PadmavatiVanka
"To Predict Guilty Agent using Fake Object Injection ", at International Journal of Engineering Resarch And Technology (IJERT)ISSN:2278-0181, VOL.1 Issue 5, July -2012.

2)Panagiotis Papadimitriou , Member IEEE , Hector G Garcia-Molina Member IEEE; "Data leakage detection ", at IEEE Transaction On Knowledge and Data Engineering , VOL. 22. 2010.

3)Kanimozhi.C-1, Christobel Diana.S-2, "Detection of Data Leakageand Guilty Agents by Injecting Fake Records", International Conference on Computing and Control Engineering (ICCCE 2012), 12 AND13 April, 2012.\

4)R. Agrawal and J. Kiernan, ―Watermarking Relational Databases,||Proc. 28th Int'l Conf. Very Large Data Bases (VLDB '02), VLDB Endowment, pp. 155-166, 2002.

5)Bing Liu, Robert Grossman, YanhongZhai-"Mining Data Recordsin Web Pages."

6)Robert Baumgartner, Wolfgang Gatterbauer, Georg Gottlob - "Web Data Extraction System"