Detection of Malicious Agent Responsible for Data Leakage

Bhagyashri Hapase
Department of Computer Engineering
AISSMS College Of Engineering
Pune, Maharashtra, India

Mona Bhutada
Department of Computer Engineering
AISSMS College Of Engineering
Pune, Maharashtra, India

Poonam Dhotre
Department of Computer Engineering
AISSMS College Of Engineering
Pune, Maharashtra, India

Chaitali Chavan
Department of Computer Engineering
AISSMS College Of Engineering
Pune, Maharashtra, India

N. R. Talhar
Department of Computer Engineering
AISSMS College Of Engineering
Pune, Maharashtra, India

Abstract

Sometimes distributor needs to provide sensitive data to supposedly trusted agents in the organization for performing some operations on data. But many times this data is found in unauthorized place. Such transmission of data to unauthorized place is known as data leakage. Due to the leakage of data, organization's secured data is at risk. This data can be sold for profit which destroys customers trust, spoils brand name, damages advantages and reputation of that organization. To overcome this problem, we develop a model which detects the agent who has leaked the data. Such agents can be called as guilty agents. This model describes how distributor can smartly provide data to the agents. By providing unique fake tuples to the distributed data, exact guilty agent can be detected.

Keywords— distributor, data leakage, guilty agents, fake tuples, distributed data.

I. INTRODUCTION

In today's era for doing business, sensitive data is handed over to the agents. The distributor owns the data of an organization. Agents in an organization represent different stakeholders who have to achieve their own objective which sometimes conflicts interest of other stakeholders. The agents in companies have their own decision making capability and can directly interact with their distributor.

For example, hospital can provide the patient's database to researchers who will advice better treatment for the patients. In business world companies may have partnership with other companies which share there data for processing.

The objective is to identify the guilty agent when distributor's sensitive data have been leaked by some agents. Perturbation and watermarking are the two techniques which can be helpful for such situation. In perturbation technique the data is made less sensitive by modifying the original data and this modified data is handed over to agents. In some cases original data need not to be changed like customers bank account number, PAN card number. However, to overcome such situations we can use addition of fake tuples technique. Using this technique distributor can detect exact guilty agent.

II. OBJECTIVE

Data leakage is an unauthorized release of secure information to an untrusted environment. The goal is to estimate the leakage of the data and to find out one of the agent in particular have leaked the data. Data allocation is intelligently done by the distributor by addition of fake tuples while distributing the data to the agents. We implement and analyse the model which detects the guilty agent without modifying original data.

III. EXISTING SYSTEM

Data transfer or data distribution plays an important role in real world. This data transfer is carried out through distributor who is the owner of the data and the agent makes use of the data further, the data transfer and detection of guilty agents in case of data leakage is carried out by using the technique named Perturbation. In this the data is discriminated into sensitive as well as non sensitive data. Distributor must satisfy the requested constraints. Hence when the data is leaked to any parties by one or more agents automatically sent to the distributor and sequentially the leaked data had made into unreadable format. Original data is thus modified in perturbation technique.

Another technique used for data leakage detection is watermarking. In this technique unique code is embedded in each distributed copy of data. If this distributed copy is found in some unauthorized place, the leaker can be detected. Watermarks were used in data whose digital representation includes considerable redundancy such as images, videos and audio data. Watermarking is used to identify the data owner.

IV. PROPOSED SYSTEM

It is possible to predict that an agent is responsible for the leak. This can be identified by comparing the data leaked by the agent and original data. Using data distribution strategies, the distributor's chances of identifying the leaker can be improved. In this module, technique for accessing the guilt of agents is developed. For implementing this model, the distributor intelligently allocated data by adding fake objects to the distributed data set. But these fake objects appear realistic to the agents but they do not correspond to real entities. These fake objects are unique and they do not replace or modify original objects. Using the fake tuple addition technique, the distributor is more confident that a particular agent was a leaker.

V. PROBLEM DEFINATION

Distributor needs to satisfy all the constraints requested by the agents. Agents may have implicit or explicit request. The data must be distributed among agents according the request they have made. Distributors constraint is to satisfy and fulfil all condition made by their agents. After this data allocation, objective of the distributor is to detect the leakage of data distributed over agents and also to detect the exact guilty agent who have leaked this distributed data. The constraint that needs to be satisfied by the distributor is strict. Distributor can not provide the different perturbed version of objects requested by agent or he can not just deny serving the data requested by agents. Hence some fake objects are added into the distributed data and provided to agents. The main objective is to maximize the chances of detecting the exact agent in the organization who is leaking the data provided to him.

VI. MODULE DESCRIPTION

- Database Maintenance
- Agent Maintenance
- Fake object addition
- Data distribution
- Detection of guilty agent

A. Database Maintenance

In this module, distributor maintains agent registration details and the sensitive data that must be provided to these agents. Here the database designing is also done.

B. Agent Maintenance

The detail information about agents is maintained and it keeps the record of the data given to them. As the history of agents is maintained it helps to detect the exact guilty agent.

C. Fake object addition

The distributor will allocate the data satisfying the constraints requested by agents. Fake tuples are added by the distributor while distributing this data. This fake tuples are unique and look real enough to every agent. The original data is not at all affected or modified. This technique improves the effectiveness in detecting exact guilty agent.

D. Data distribution

The distributor handles two types of request: implicit and explicit. These two types of requests are considered, as the agents may request the data based on some constraints which must be satisfied by distributor. While distributing the data, the distributor adds unique fake tuples along with requested data to every agent. For example in banking system, accountant may request for 100 records of customer. This is an implicit request. In the same example if accountant requests for 100 records of customers having saving account then it is explicit request.

E. Detection of guilty agent

Suppose the dataset X has been leaked by particular agent A. This dataset X is found by the distributor on particular domain for which the read write privileges are provided. Distributor will compare leaked dataset X with original distributed data. The distributor comes to know that agent A have leaked the data. But Agent A might claim that he is innocent and might blame other agents in the organization. Since fake tuples provided to every agent are unique, the distributor is confident about the exact guilty agent.

VII. SYSTEM IMPLEMENTATION

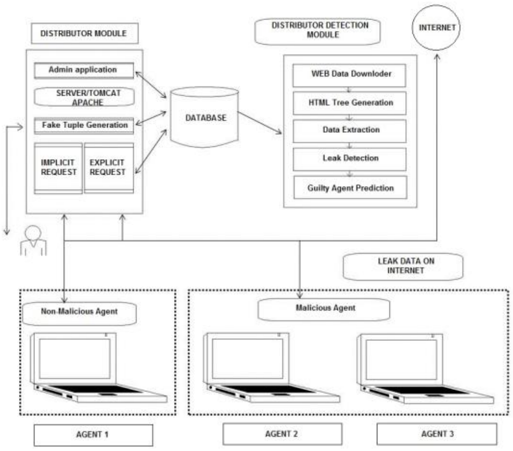

Figure 1. Flow Diagram

Initially, the agent will login into the system. Only authenticated agents are allowed to access the system. Distributor is the owner of the data who maintains the database. When the agent requests for the data, the distributor adds unique fake tuples and distributes the data. It may happen that one of the supposedly trusted agent leak the sensitive data of an organization to an unauthorized place. Due to data leakage company is at serious risk. This destroys customer's trust, spoils brand name and reduces reputation of that organization. So it is necessary to identify who has exactly leaked the data.

After the data distribution malicious agent will release this data on particular domain of which read write privileges has been provided to him using FTP Uploader. Distributor will download the data from the domain using FTP downloader. Then he will parse downloaded data using HTML or XML Parser. Distributor will compare the parsed data with original database of an organization. If it matches distributor will come to know that organization's data has been leaked. Comparison of parsed data and distributed data is done. If parsed data tuples matches with the unique fake tuples of particular agent. Then distributor will definitely identify the exact guilty agent.

VIII. SOFTWARE DESCRIPTION

Technologies used:

- Java Servlets
- J2EE
- JDBC
- My SQL Database
- Net Beans
- FTP Uploader
- FTP Downloader
- HTML Parser
- XML Parser
- Nano XML Parser

IX. CONCLUSION

Ideally, sensitive data need not be given to the agents. But in real world, we have to provide organization's data to its agents for processing various tasks. These agents can knowingly or unknowingly leak this sensitive data. To trace the exact agents who have leaked the data, unique fake tuple addition technique is used in which realistic but fake tuples are injected while distributing the data. This technique helps to find exact guilty agent.

ACKNOWLEDGEMENT

This paper involves number of respected helping hands. We are grateful to Prof. N. R. Talhar for his dedication and valuable guidance. We would like to thank the Department of Computer Engineering, AISSMS COE, Pune for their uninterrupted help and support.

REFERENCES

[1] S. Umamaheswari, H.Arthi Geetha, authors of IEEE paper on *Detection of Guilty Agent*, Proceedings of the National Conference on Innovations in Emerging Technology-2011Kongu Engineering College, Perundurai, Erode, Tamilnadu, India.17 & 18 February, 2011.pp.23-26. February 2011

[2] A. Santhi Laxmi, Padmavati Vanka, A. Bhaskar, authors of International Journal On

To predict guilty agents using fake object injection, International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 5, July – 2012 ISSN: 2278-0181

[3] Rupesh Mishra, D. K. Chitre, authors of International Journal On *Data Leakage and Detection of Guilty Agent*, International Journal of Scientific & Engineering Research, Volume 3, Issue 6 and June-2012 ISSN 2229-5518

[4] S.Jenila, K.Sivasankari, R.Arudselvi, J.Maria Monica, B.Saranya, authors of International Journal on *Guilt Model Process for Identifying Data Leakage and Guilty Agent in Data transmission*, International Journal of Computer Applications (0975 – 8887) Volume 42– No.6, March 2012

[5] N. Bangar Anjali, P. Rokade Geetanjali, Patil Shivlila, R. Shetkar Swati, N B Kadu, authors of Research Article On *Data Leakage Detection*, International Journal of Computer Science and Mobile Computing IJCSMC, Vol. 2, Issue. 5, May 2013, pg.283 – 288, ISSN 2320–088X

[6] N. Sandhya, G. Haricharan Sharma, K. Bhima, authors of International Journal On *Exerting Modern Techniques for Data Leakage Problems Detect*, International Journal of Electronics Communication and Computer Engineering Volume 3, Issue (1) NCRTCST, ISSN 2249 –071X,2012

[7] Rudragouda G Patil, author of International Journal On *Development of Data leakage Detection Using Data Allocation Strategies*, International Journal of Computer Applications in Engineering Sciences, volume 1,Issue 2,June-2011,ISSN 2231-4946 **[**Vol I, Issue II, June 2011**]**

[8] What is Detection of Guilty Agent [Online] Available: https://docs.google.com/document/d/1qVSNpQft0Whs3X7Yv6EktWUUFl4mPGjbjSF2i35YwQ/preview?pli=1

[9] Implementation of data leakage detection [Online] Available: http://www.techgig.com/projects/Implementation-Of-Data-Leakage-Detection-In-Distributed-Network-39145

[10] Data Leakage Detection Project [Online] Available: http://www.faadooengineers.com/threads/10452-Data-Leakage-Detection-Project

[11] Data Leakage Detection [Online] Available: https://www.classle.net/projects/node/2999

[12]Detection-of-data-misuse-using-allocation-strategic-data-mining-techniques[Online]Available: http://www.upublish.info/article/detection-of-data-misuse-using-allocation-strategic-data-mining-techniques/409661