ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology

National Conference on Innovative Applications and Research in Computer Science and Engineering (NCIARCSE-2017)

AGTI's Dr. Daulatrao Aher College Engineering, Vidyanagar Extension, Karad

Vol. 4, Special Issue 4, January 2017

Data Leakage Detection

Mr. Sagar Ravindra Dalvi[1], Ms. Shamika Rajendra Khatu[2]

Dept of Computer Engineering, Gharda Institute of Technology, Khed[1,2]

Abstract: A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the data is leaked and found in an unauthorized place (e.g., on the web or somebody's laptop). The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. Data allocation strategies (across the agents) that improve the probability of identifying leakages has been proposed. These methods do not rely on alterations of the released data (e.g., watermarks). In some cases distributor can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty party.

Keywords: Guilty agent, data distributor, fake object, data leakage.

I. INTRODUCTION

Data leakage is defined as the accidental or unintentional distribution of private or sensitive data to unauthorized entity. Sensitive data of companies and organizations includes intellectual property (IP), financial information, patient information, personal credit-card data, and other information depending on the business and the industry.
Furthermore, in many cases, sensitive data is shared among various stakeholders such as employees working from outside the organizational premises (e.g., on laptops), business partners and customers. This increases the risk of confidential information falling into unauthorized hands. Whether caused by malicious intent, or an inadvertent mistake, by an insider or outsider, exposed sensitive information can seriously hurt an organization.
The potential damage and adverse consequences of a data leak incident can be classified into the following two categories: direct and indirect loss. Direct loss refers to tangible damage that is easy to measure and estimate quantitatively. Indirect loss, on the other hand, is much harder to quantify and has a much broader impact in terms of cost, place and time. Direct loss includes violating regulations (such as those protecting customer privacy) resulting in fine/settlement/customer compensation fees; litigation of lawsuits; loss of future sales; costs of investigation and remedial/restoration fees. Indirect loss includes reduced share-price as a result of the negative publicity; damage to company's goodwill and reputation; customer abandonment; and exposure of Intellectual Property (business plans, code, financial reports, and meeting agendas) to competitors.

II. EXISTING SYSTEM

Perturbation

Application where the original sensitive data cannot be perturbed has been considered. Perturbation is a very useful technique where the data is modified and made "less sensitive" before being handed to agents. For example, one can add random noise to certain attributes, or one can replace exact values by ranges. However, in some cases it is important not to alter the original distributor's data. For example, if an outsourcer is doing our payroll, he must have the exact salary and customer bank account numbers. If medical researchers will be treating patients (as opposed to simply computing statistics), they may need accurate data for the patients.

Watermarking

Traditionally, leakage detection is handled by watermarking, e.g., a unique code is embedded in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified. Watermarks can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious.

III. PROPOSED SYSTEM

Unobtrusive techniques for detecting leakage of a set of objects or records have been studied. After giving a set of objects to agents, the distributor discovers some of those same objects in an unauthorized place. (For example, the data may be found on a web site, or may be obtained through a legal discovery process.)At this point the distributor can assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. Using an analogy with cookies stolen from a cookie jar, if Freddie with a single cookie has been cached, he can argue that a friend gave him the cookie. But if Freddie with 5 cookies has been cached, it will be much harder for him to argue that his hands were not in the cookie jar. If the distributor sees "enough evidence" that an agent leaked data, he may stop doing business with him, or may initiate legal proceedings. A model for assessing the "guilt" of agents has been developed. An algorithm for distributing objects to agents, in a way that improves our chances of identifying a leaker

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
National Conference on Innovative Applications and Research in Computer Science and Engineering (NCIARCSE-2017)
AGTI's Dr. Daulatrao Aher College Engineering, Vidyanagar Extension, Karad
Vol. 4, Special Issue 4, January 2017

has been proposed. The option of adding "fake" objects to the distributed set also been considered. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of watermark for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty.

Figure 1welcome page

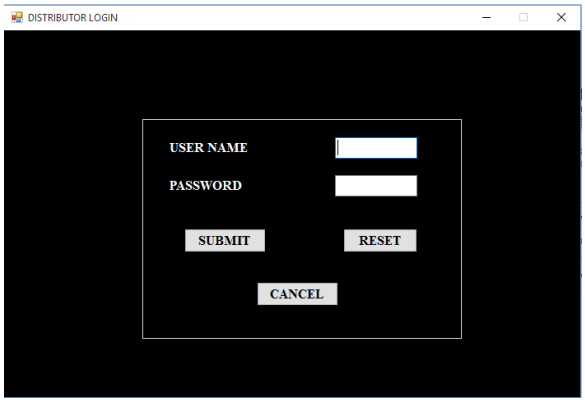

Figure 2Distributor Login

IV. METHODOLOGY

ALGORITHM STEPS:

Step1: Distributor select agent to send data.
The distributor selects two agents and gives requested data R1, R2 to both agents.

Step2: Distributor creates fake object and allocates it to the agent.
The distributor can create one fake object (B = 1) and both agents can receive one fake object (b1 = b2 = 1). If the distributor is able to create more fake objects, he could further improve the objective.

Step3: Check number of agents, who have already received data.
Distributor checks the number of agents, who have already received data.

Step4: Check for remaining agents.
Distributor chooses the remaining agents to send the data. Distributors can increase the number of possible allocations by adding fake object.

Step5: Select fake object again to allocate for remaining agents.
Distributor chooses the random fake object to allocate for the remaining agents.

Step6: Estimate the probability value for guilt agent.
To compute this probability, we need an estimate for the probability that values can be guessed by the target.

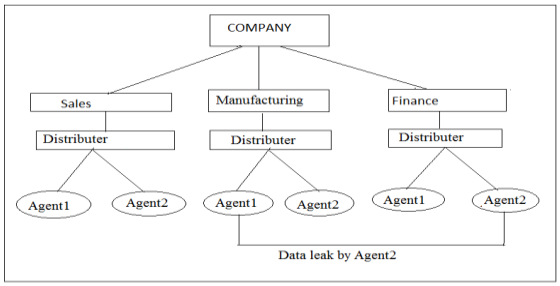

Figure 3System Design

V. ADVANTAGE

- If the distributor sees "enough evidence" that an agent leaked data, he may stop doing business with him, or may initiate legal proceedings.
- We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker.
- Consider the option of adding "fake" objects to the distributed set. Such objects do not correspond to real entities but appear.
- If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty.

VI. MODULE DESCRIPTION

• LOGIN / REGISTRATION:
This is a module mainly designed to provide the authority to a user in order to access the other modules of the project. Here a user can have the accessibility authority after the registration.

• DATA TRANSFER:
This module is mainly designed to transfer data from distributor to agents. The same module can also be used for illegal data transfer from authorized to agents to other agents.

• GUILT MODEL ANALYSIS:
This module is designed using the agent – guilt model. Here a count value(also called as fake objects) are incremented for any transfer of data occurrence when agent transfers data. Fake objects are stored in database.

• AGENT-GUILT MODEL:
This module is mainly designed for determining fake agents. This module uses fake objects (which is stored in database from guilt model module) and determines the

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

IARJSET

International Advanced Research Journal in Science, Engineering and Technology

National Conference on Innovative Applications and Research in Computer Science and Engineering (NCIARCSE-2017)

AGTI's Dr. Daulatrao Aher College Engineering, Vidyanagar Extension, Karad

Vol. 4, Special Issue 4, January 2017

guilt agent along with the probability. A graph is used to plot the probability distribution of data which is leaked by fake agent.

VII. CONCLUSION

In a perfect world there would be no need to hand over sensitive data to agents that may unknowingly or maliciously leak it. And even if we had to hand over sensitive data, These methods do not rely on alterations of the released data (e.g., watermarks). In some cases we can also inject realistic but fake data records to further improve our chances of detecting leakage and identifying the guilty party. We propose data allocation strategies (across the agents) that improve the probability of identifying leakages.

ACKNOWLEDGEMENT

For all the efforts behind this paper work, we first would like to express our sincere thanks to the staff of Dept. of computer Engg., for their extended help and suggestions at every stage. It is with a great sense of gratitude that I acknowledge the support, time to time suggestions to my guide **Prof. R.B. Pawar.**

REFERENCES

[1] Priyanka Barge, Pratibha Dhawale, Namrata Kolashetti Ass. Prof., Department of Computer Engineering, NIRMALA CHOUHAN .A Novel Data Leakage Detection .International Journal of Modern Engineering Research (IJMER).

[2] DATA LEAKAGE DETECTION USING DATA ALLOCATION STRATEGIES Jaymala Chavan, Priyanka Desai Thakur College of Engg. Tech, Mumbai, MH, India. International Journal of Advances in Engineering Technology, Nov. 2013.