Data Leakage Detection

Prof. R. A. Gulhane
Assistant Professor
Deptt.of Computer Sci. and Engg.
PRMIT & R, Badnera

Mr. Akshay R. Bijawe
Final year Bachelor of Engineering Student
Deptt.of Computer Sci. and Engg.
Sipna's C.O.E.T,Amt.

Ms. Nupur R. Jaiswal
Final year Bachelor of Engineering Student
Deptt.of Computer Sci. and Engg.
Sipna's C.O.E.T,Amt.

Abstract

As the watermarking techniques have been successfully utilized for copyright protection of multimedia data, yet the research of database water-marking technique is still facing a lot of challenges due to the differences between the relational database and multi-media data. *In this paper, we are focusing on detecting when agents have leaked the distributor's sensitive data, and if possible to identify the agent that leaked the data. We present a model for calculating "guilt" probabilities in cases of data Leakage. We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker. The robustness of the method is confirmed by detecting the data leakage by using watermarking technique the results that display the immunity of the embedded watermark to several kinds of data, such as compression, filtering, scaling, cropping, and rotation. Based on this technique we can easily detect the data leakage.*

1. Introduction

The robustness of the method is confirmed by detecting the data leakage by using watermarking technique the results that display the immunity of the embedded watermark to several kinds of data, such as compression, filtering, scaling, cropping, and rotation. Based on this technique we can easily detect the data leakage.

Data Leakage can occur through a variety of methods - some are simple, some complex. As such, there is no single "silver bullet" to control Data Leakage. Data leakage detection is an increasingly important part of any organization's ability to manage and protect critical and confidential information. Examples of critical and confidential data that applications can access include: Intellectual Property, Corporate Data, and Customer Data.

Sometimes data is leaked and found in unauthorized places. Nowadays, more and more data are sold and transmitted on the internet. Databases are being used widely in many important fields, such as, banking and so on. With the fast growth of database business on the net, the data may be unsafe after passing through the unsecure network. The datafor the following suspicion.

- First, the data receiver may suspect that the data is tampered by an unauthorized person.
- Second, they may suspect the data received are not produced and provided by the authorized suppliers.
- Third, the suppliers and purchasers actually with different interest should have different roles of rights in the database management or using. So how to protect and verify the data becomes very important here.

The goal is to detect when the distributor's sensitive data has been leaked by agents, and show the probability for identifying the agent that leaked the data. We study unobtrusive techniques for detecting leakage of a set of objects or records. Specifically, we study the following scenario: After giving a set of objects to agents, the distributor discovers some of those same objects in an illegitimate 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.

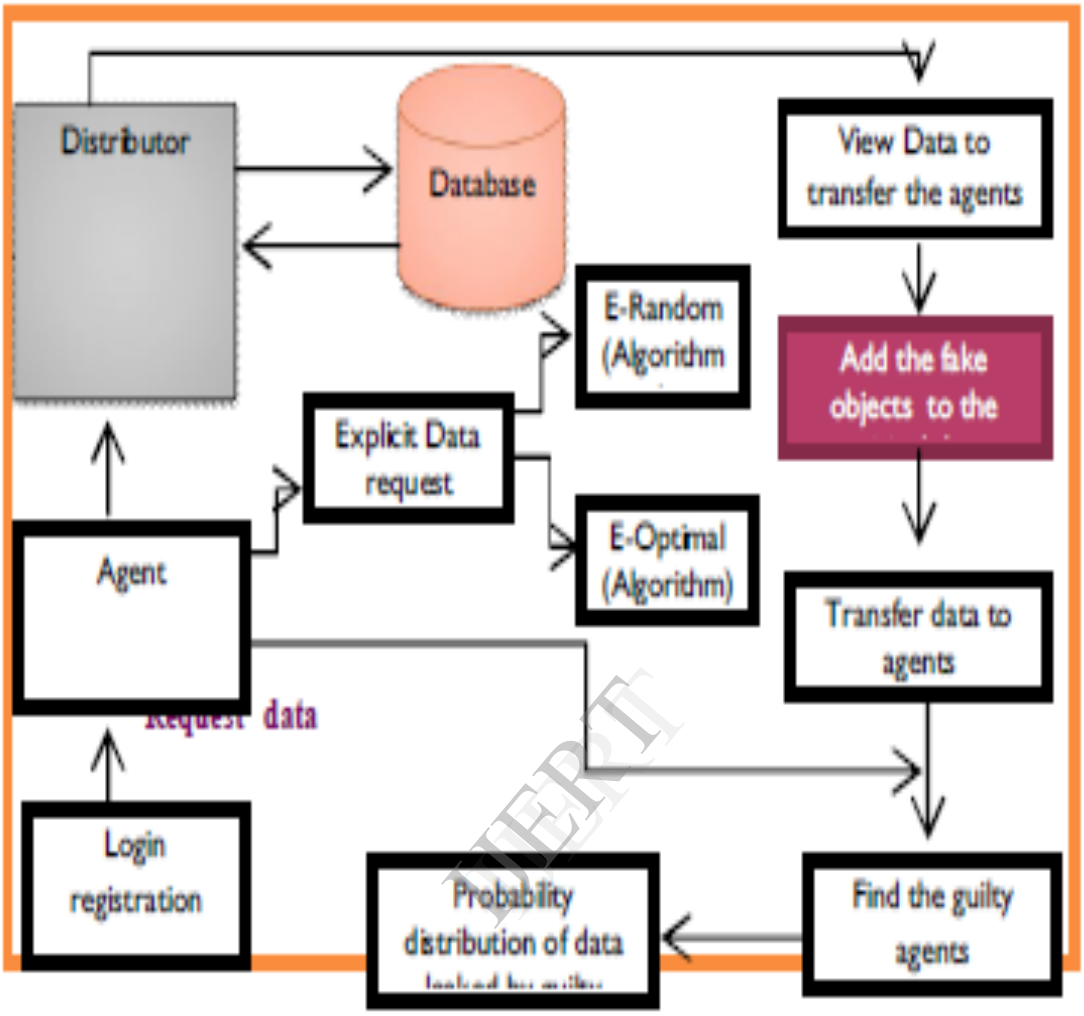

Fig: System Architecture

2. Methodology

Our goal is to detect, when the distributor's sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data. Perturbation is a very useful technique where the data is modified and made "less sensitive" before being handed to agents.

We propose to develop unobtrusive techniques for detecting leakage of a set of objects or records. In this we propose to develop a model for assessing the "guilt" of agents. We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker.

Finally, we also consider the option of adding "fake" objects to the distributed set.

- ✓ 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.

3. Modules

- **Data Allocation Module:**

The main focus of our project is the data allocation problem as how can the distributor "intelligently" give data to agents in order to improve the chances of detecting a guilty agent, Admin can send the files to the authenticated user, users can edit their account details etc. Agent views the secret key details through mail. In order to increase the chances of detecting agents that leak data.

- **Fake Object Module**:

The distributor creates and adds fake objects to the data that he distributes to agents. Fake objects are objects generated by the distributor in order to increase the chances of detecting agents that leak data. The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. Our use of fake objects is inspired by the use of "trace" records in mailing lists. In case we give the wrong secret key to download the file, the duplicate file is opened, and that fake details also send the mail. Ex: The fake object details will display.

- **Optimization Module**:

The Optimization Module is the distributor's data allocation to agents has one constraint and one objective. The agent's constraint is to satisfy distributor's requests, by providing them with the number of objects they request or with all available objects that satisfy their conditions. His objective is to be able to detect an agent who leaks any portion of his data. User can able to lock and unlock the files for secure.

- **Data Distributor Module**:

A data distributor has given sensitive data to a set of supposedly trusted agent (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 Admin can able to view the which file is leaking and fake user's details also.

- **Agent Guilt Module**:

To compute this, we need an estimate for the probability that values in S can be "guessed" by the target. For instance, say some of the objects in T are emails of individuals. We can conduct an experiment and ask a person with approximately the expertise and resources of the target to find the email of say 100 individuals. If this person can find say 90 emails, then we can reasonably guess that the probability of finding one email is 0.9. On the other hand, if the objects in question are bank account numbers, the person may only discover say 20, leading to an estimate of 0.2. We call this estimate point the probability that object t can be guessed by the target. To simplify the formulas that we present in the rest of the paper, we assume that all T objects have the same point, which we call p. Our equations can be easily generalized to diverse points though they become cumbersome to display. Next, we make two assumptions regarding the relationship among the various leakage events. The first assumption simply states that an agent's decision to leak an object is not related to other objects.

4. Application

Systems like reservations, stock exchange or banking where data is of a very large size and critical After registering the agents the system will supply necessary data to the agents to solve customer queries about say balance in the bank, complaint about a transaction or enquiry about reservation. Detect any leakage of data from the agent using hidden information

4.1 Techniques used for leakage detection

- Assign an auto generated internal unique identity key to each agent.
- Add these keys to the data at the time of supply.

- Add few fake objects to the data those can uniquely identify agents.
- Track the keys if any leakage is found and detect the exact agent responsible.

5. Conclusion

We can conclude that using this application we can deal with outsourcing of critical data that needs a leakage detection system. So we can detect the agents who are responsible for data leakage thus helping the business owner protect their important data.

References

[1] A. Shabtai et al., "A Survey of Data Leakage Detection and Prevention Solutions" Springer Briefs in Computer Science, 2012.
[2] Panayiotis Papadimitriou, Hector Garcia-Molina, " Data Leakage Detection" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2010.
[3] Polisetty Sevani Kumari, Kavidi Venkata Mutyalu. "Development of Data leakage Detection Using Data Allocation Strategies" International Journal of Computer Trends and Technology- volume 3, Issue 4, 2012 ISSN: 2231-2803, pp. 548.
[4] Sandip A. Kale , Prof.S.V. Kulkarni, "Data Leakage Detection" IOSR Journal of Computer Engineering (IOSRJCE) ISSN : 2278-0661, volume 1, Issue 6 (July-Aug2012), pp. 32-35.
[5] Mr.V.Malsoru, Naresh Bollam, "REVIEW ON DATA LEAKAGE DETECTION" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 1, Issue 3, pp.1088-1091.