International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 06, June 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Data Leakage Detection Using Algorithms

Mrs. Seema G. Bavachkar, Mrs. Swati N. Pawar

Assistant professor, ATS's, SBGI, Miraj

bavachkarsg@sbgimiraj.org

bavachkarsg@sbgimiraj.org

ABSTRACT

This paper contains concept of data leakage, its causes of leakage and different techniques to protect and detect the data leakage. The value of the data is incredible, so it should not be leaked or altered. In the field of IT, huge database is being used. This database is shared with multiple people at a time. But during this sharing of the data, there are so many chances of data vulnerability, leakage or alteration. So, to prevent these problems, a data leakage detection system has been proposed.

KEYWORDS —watermarking guilty agent; explicit data, data leakage detection.

INTRODUCTION

Data leakage is the unauthorized transmission of sensitive data or information from within an organization to an external destination or recipient. Sensitive data of companies and organization that includes: -

1. Financial information,
2. personal credit card data
3. And other information depending upon the industry

The term can be used to describe data that is transferred electronically or physically. Data leakage threats usually occur via the web and email, but can also occur via mobile data storage devices [3] such as optical media, USB keys, and laptops.

Reasons of data leakage: -

1. Weak passwords.
2. Theft of a company Item from employee.
3. Accidentally emailing sensitive information or Publishing It Online.
4. Malicious attacks that Result in Data Leakage.
5. Phishing.
6. Loss of Paperwork.

I. EXISTING SYSTEM

Traditionally, leakage detection is handled by watermarking, e.g., a unique code is embedded [2] in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified.

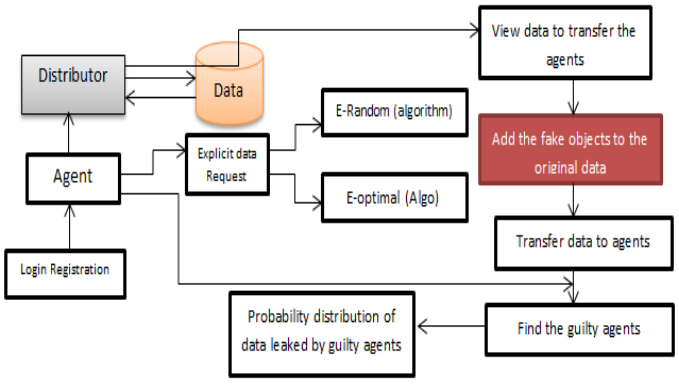
Disadvantages

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. E.g. A hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. We call the owner of the data the distributor and the supposedly trusted third parties the agents.

II. PROPOSED SYSTEM

- 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 develop unobtrusive techniques for detecting leakage of a set of objects or records.
- We 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 act as a type of watermark for the entire set, without modifying any invidual 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.

III. SYSTEM ARCHITECTURE



Algorithm

Allocation for Explicit Data Requests (EF)

Input: R1; . . .Rn, cond1, ..., condn, b1; . . . bn, B

Output: R1; . . .; Rn, F1; . . . ; Fn

1: R_ 0 Agents that can receive fake objects
2: for i_1, . . . ; n do
3: if bi > 0 then
4: R_ R U{i}
5: Fi_0 ;
6: while B > 0 do
7: i_ SELECTAGENT(R,R1; . . .;Rn)
8: f _CREATEFAKEOBJECT(Ri; Fi; condi)
9: Ri _Ri U{f}
10: Fi_ Fi U{f}
11: bi_ bi-1
12: if bi= 0 then
13: R_R\{Ri}
14: B _B – 1

The Algorithm finds agents that are eligible to receiving fake objects . Then, in the main loop is the algorithm creates one fake object in every iteration and allocates it to random agent. The algorithm minimizes every term of the objective summation by adding the maximum number of fake objects to every set,yielding the optimal solution, the algorithm just selects at random the agents that are provided with fake objects.

B . Evaluation of Sample Data Request Algorithm

With sample data requests agents are not interested in particular objects. Hence, object sharing is not explicitly defined by their requests. 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 more data objects the agents request in total, the more recipients on average an object has; and the more objects are shared among different agents, the more difficult it is to detect a guilty agent.

IV. MODULES

1. Data Allocation Module:

The main goal of experiment is the data allocation problem as how can the distributor "intelligently" give data to the gents 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.

2. 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.

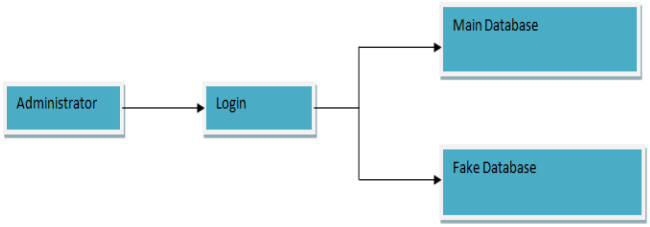


Fig. Fake object Module

3. 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.

4. Data Distributor Module:

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 Admin can able to view the which file is leaking and fake user's details also.

5.Guilty agent: -

. To detect when the distributor's sensitive data has been leaked by agents, and if possible to identify the agent hat leaked the data. Perturbation is a very useful technique where the data is modified and made "less sensitive" before being handed to agents. An unobtrusive technique is developed for detecting leakage of a set of objects or Records.

V. CONCLUSION

From the study of the data leakage, we can detect and prevent the data from the leak by using some algorithms and techniques. 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, in a perfect world we could watermark each object so that we could trace its origins with absolute certainty.

VI. REFERNCES

[1] Dr.A.R. Pon Periyasamy, E.Thenmozhi, "Data leakage detection and data prevention using algorithm" PG and Research department Nehru memorial collage (autonomous) Puthanmpatti, tamilnadu, india, April 2017.

[2] Rajesh kumar, "Data Leakage Detection" Double blind peer reviewed international Research Journal, Global journals Inc(USA) 2018.

[3] Shaj .v, K.P.KALIYAMURTHIE,"A Review on Data Leakage Detection"Department of information technology,Bharath University,India , April 2014.

[4] Panagiotis papadimtriou and Hector Garcia-Molina,"Data Leakage Detection",IEEE Trans, knowledge and Data Engineering, vol.23 no.1,january 2013.

[5] Data Leakage prevention: A news letter for IT Professional Issue 5 P.P(1-3).