INTERNATIONAL JOURNAL OF ADVANCE RESEARCH AND DEVELOPMENT

IJARnD

(Volume2, Issue4)
Available online at www.ijarnd.com

Data leakage Detection

[1]Prof. Nilima Nikam, [2]Mayuri Form are, [3]Priyanka Palve, [4]Smita Halde
[1234]Yadavrao Tasgaonkar Institute of Engineering and Technology
Department Of Computer Engineering
mayurifokmare@gmail.com, smitahalde123@gmail.com, priyankapalve94@gmail.com

ABSTRACT
This paper explores the issue of data leakage detection. The unauthorized person transfers some important data to unknown person or outside the world from organization record. We developed data leakage detection software to overcome the important data leakage from the organization. Data leakage detection is a strategy for making sure that users do not send sensitive data or information to outside the world. In the organization, sometimes sensitive data must be handed over to trusted third parties from a user (guilty user).

The distributor can easily accessible all information about the guilty user for example: which file shares to a fake agent from organization data records and their details etc. The main goal of the software identifies data leakages using some data allocation strategies and find out the fake agent and users who leak the important data. Data leakage prevention software also provides the essential security to data.

Keywords: Distribution Model, Information Security, Data security.

INTRODUCTION

In the course of doing business, sometimes sensitive data must be handed over to supposedly trusted third parties. For example, a hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that required 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. Our goal is to detect when the distributor's sensitive Data have been leaked by agents, and if possible to identify the agent that leaked the data.

METHODOLOGY

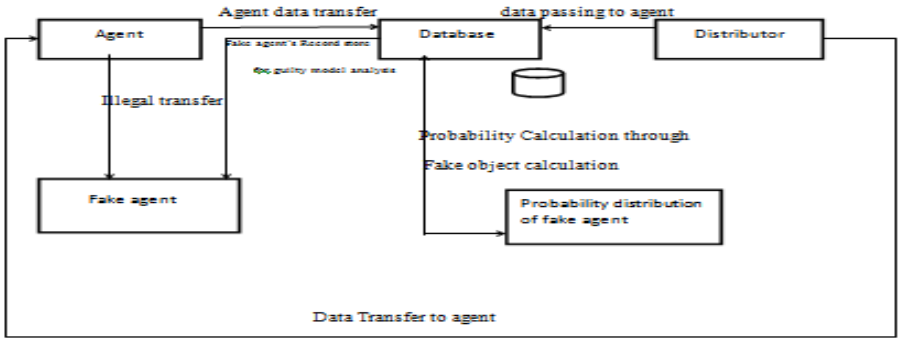


Fig. 4.1 Block diagram of system architecture

A distributor can insert original as well as fake records in the Database. A new agent can be registered by entering personal details. A registered agent can Login and make a request to the distributor for data. The request can be of two types-Sample or Explicit.

The system then extracts the requested data from the main database and performs the addition of fake records to the set of original records. It then provides this data to the agent. The agent may pass on this data to an unauthorized party.

Algorithm

Allocation for Explicit Data Requests In this request the agent will send the request with appropriate condition. Allocation for Sample Data Requests In this request agent request does not have the condition. The agent sends the request without condition as per his query he will get the data.

1: a ← 0|T| a[k]: number of agents who have received object tk

2: R1 ← ∅, . . . ,Rn←∅

3: remaining ←

4: while remaining > 0 do

5: for all i = 1, . . . , n : |Ri| < mi do

6: k ← SELECT OBJECT (i, Ri) May also use additional Parameters

7: Ri ← Ri∪ {tk}

8: a[k] ← a[k] + 1

9: remaining ← remaining – 1

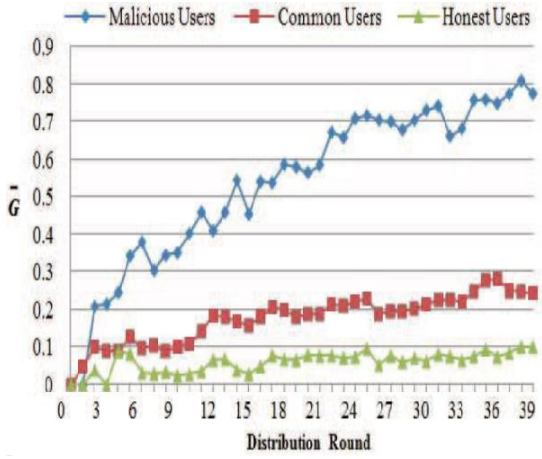
Proposed System

In the proposed system 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 unauthorized place. (For example, the data may be found on a website 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.

In the proposed approach, 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 individual members. If it turns out that an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty. In the Proposed System the hackers can be traced with a good amount of evidence.

Experiment Results

To verify effects of the guilt probability on improving the probability of marking leakage source, in this paper we select two simulation distributing files experiments, one taking the guilt probability into account and the other not, then we analyze and compare two experiment results. These two experiments are set the same scene that means in the course of distributing files all the users act same - the specific user requires, receives or leaks a specific file at the specific moment



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. However, in many cases, we must indeed work with agents that may not be 100 % trusted, and we may not be certain if a leaked object came from an agent or from some other source, since certain data cannot admit watermarks.

In spite of these difficulties, we have shown that it is possible to assess the likelihood that an agent is responsible for a leak, based on the overlap of his data with the leaked data and the data of other agents, and based on the probability that objects can be "guessed" by other means. Our model is relatively simple, but we believe that it captures the essential trade-offs. The algorithms we have presented implement a variety of data distribution strategies that can improve the distributor's chances of identifying a leaker.

We have shown that distributing objects judiciously can make a significant difference in identifying guilty agents, especially in cases where there is a large overlap in the data that agents must receive. It is difficult to directly detect the leakage behavior of users in practical applications the guilt probability can be used to infer the leaker when the file is found leakage. If the user's guilt probability is very high, then we can conclude that the user may be a leaker. Based on the above analysis, the distribution strategy taking account of the guilt probability can increase the gap between users of different types, helping more accurately determine a malicious user and then take measures for data leakage prevention.

CONCLUSION

A distributor can insert original as well as fake records in the Database. A new agent can be registered by entering personal details. A registered agent can Login and make a request to the distributor for data. The request can be of two types-Sample or Explicit.

The system then extracts the requested data from the main database and performs the addition of fake records to the set of original records. It then provides this data to the agent. The agent may pass on this data to an unauthorized party.

REFERENCE

[1] Sandip A. Kale, Prof. Kulkarni S.V. (Department Of Computer Sci. &Engg,MIT College of Engg, Dr.B.A.M.University, Aurangabad(M.S), India, Data Leakage Detection: A Survey, (IOSR Journal of Computer Engineering (IOSRJCE)ISSN : 2278-0661 Volume 1, Issue 6 (July-Aug 2012), PP 32-35 www.iosrjournals.org

[2] IEEE Transactions On Knowledge And Data Engineering, Vol. 22, No. 3, March 2011 Data Leakage Detection Panagiotis Papadimitriou, Member, IEEE, Hector Garcia-Molina, Member, IEEE P.P (2,4-5)

Allocation Strategies

[1] Chun-Shien Lu, Member, IEEE, and Hong-Yuan Mark Liao, Member, IEEE Multipurpose Watermarking for Image Authentication and Protection

[2] A. Shabtai, a. Gershman, M. Kopeetsky, y. Elovici Deutsche Telekom Laboratories at Ben-Gurion University, Israel. Technical Report TR-BGU-2409-2010 24 Sept. 2010 1 a Survey of Data Leakage Detection and Prevention Solutions P.P (1-5, 24-25)

