International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey on the Various Techniques of Data Leakage Detection

Saranya S. Devan[1], Minu Lalitha Madhavu[2]

M.Tech Scholar, Dept. of Computer Science & Engineering,
Sree Buddha College of Engineering, Alappuzha, Kerala, India

Asst. Professsor, Dept. of Computer Science & Engineering,
Sree Buddha College of Engineering, Alappuzha, Kerala, India

Abstract: ***Now-a-days trusted parties will be given sensitive data by the data distributor. So, this data must be protected and should not be found in an unauthorized place. This paper mainly focuses on a survey of the leakage of the data by the various agents within the organization and the different techniques to avoid this leakage. In several fields of research, data is highly confidential. Since, the data is shared by a large number of people; it is vulnerable to alteration and leakage. So, this paper mainly focuses on a survey of various data leakage detection techniques and the approach proposed mainly focuses on delegated access control with the tracing of MAC address so that the leaker can be detected and the data can be blocked to the unauthorized world.***

Keywords: Data distributor, agents, data leakage, MAC , delegated access control.

1. Introduction (*Heading 1*)

The data either private or sensitive that is distributed not intentionally or accidentally to an unauthorized entity is called as data leakage. Some of the sensitive data mainly include research information, account information, personal data, patient information, financial information etc. This sensitive information can be shared among a large number of people and thereby, it is vulnerable to various leakage and alterations. This mainly includes the data being collected by the unauthorized hands within or outside the organization. The data leakage can be direct or indirect loss. The damage that is tangible so that it is easy to measure and estimate quantitatively is called as direct loss. The damage that is much harder and difficult to quantify in terms of cost, time and place is called as indirect loss.

Traditionally, the leakage of the data can be detected by means of watermarking. The copy that is distributed will be embedded with unique codes. Whenever this copy is found in the hands of any unauthorized entity leaker can be easily detected. But, watermarking generally needs modification and also, the watermark can be smashed if the intended receiver is malicious. Also, perturbation can be useful to detect the leakage of data. In perturbation, the data is modified and the sensitivity is made less so that the sensitive data cannot be easily identified. For example, the exact values of certain data can be replaced by ranges, or provide an approximate value of the original data. But, perturbation also modifies the data which is not allowed by the data distributor.

The job of identifying the insider who leaks the sensitive data to the outside world is very difficult and faces many challenges. But, in many organizations there is a need to share the data among other organizations as a need for partnership. So, there arises the need to avoid data leakage. For example, the owner of an organization has leaked the data and it was seen in some website or in the laptop of another organization. So, the main goal of this paper is to detect when the sensitive data of the organization has been leaked by the agent or the leaker and also, the agent who leaked the data. And also find a solution so that the data will be blocked to the outside world. This paper is a survey on the various data leakage detection techniques and prevention strategies.

2. Related Works

The main goal of this paper is to detect when the sensitive data of the organization has been leaked by the agent or the leaker and also, the agent who leaked the data. And also find a solution so that the data will be blocked to the outside world. Some of the existing data leakage detection schemes include:

a) Perturbation Approach
In this technique, original sensitive data is made less sensitive [2]. The data is mainly approximated to the nearby value of the original data i.e the original data is modified. For example the exact value of the sensitive data can be replaced by ranges, rounded off to the nearest integer.

b) Watermarking Approach
In some applications the modification of the data is not allowed i.e the original sensitive data cannot be made less sensitive. The data provided by the data distributor cannot be altered [1]. So, the leakage of the data can be handled by embedding small codes into the original sensitive data when distributed. If any modification is found in this copy, the leaker can be easily detected.

Watermark is used as a proof of ownership. It is a signal that is embedded securely and robustly into the sensitive original data like image, audio, video etc and thereby producing a watermarked signal. It provides protection for the relational data. Some bit positions of some attributes of some tuple consist of specific values. These values are mainly determined only using the private key known to the owner. The person who has access to the private key will be able to modify the watermark.

Some small errors are inserted into the object being watermarked. The errors that are made intentionally are called marks. The collection of all the marks is called watermark. Even though watermark is found to be useful, it also requires the modification of the original data and once the intended receiver is malicious, the watermark is smashed.

c) Data Allocation Approach

- Explicit Data Request Allocation

  In this type of allocation [3], the request will be send by the agent with an appropriate condition. The input generated by the agent consists of the data and condition for the request. After processing, fake objects will be added to the data in an encrypted format.

  In the type of explicit requests not allowing the fake objects, the distributor is not allowed to add fake objects to the sensitive data. The data request mainly defines the data allocation. But, in the case of explicit requests allowing fake objects, the request cannot be modified or removed by the data distributor from the agent.

  In this type of request, the input will be a set of requests r1,r2…..rn from n agents along with the conditions for the request. The optimal algorithm mainly helps to determine the agents that are able to receive the fake objects. Then in each iteration one fake object is created and allocating the fake objects to the agent being selected. Each term is being minimized by adding the maximum number of fake objects to every set of requests to yield optimal solution.

- Sample Data Request Allocation

  In the sample data requests [3], there is no condition within the agent request. The request will be sending with no condition as per his query and will get the data. The leaked data mainly came from one or more agents. The distributor allots objects and keeps the chance of detection of guilt agent constant. The guilt probability mainly depends on the agents who received the objects being leaked. It does not depend on the identity of leaked objects.

3. Proposed Approach: Delegated Access Control with Data Leakage Detection

Access control is the mechanism of selective access to the sensitive data. If access control is provided to the data then it will provide restriction of access to the sensitive data stored. And therefore the data leakage can be avoided to a larger extent. This method mainly restricts the access of the various agents within the organization or outside the organization. Traditionally, the control of access mainly lies in the hands of the owner. But it results in high communication and computation overhead.

Delegated access control means the control of access is distributed over multiple parties. And thereby, the data leakage can be detected and also, the overhead can be reduced. Along with delegated access control, MAC address of the system is being traced and thereby, whenever the agent pass on the sensitive data to the outside world it can easily be detected as the data is being running on an untrusted network and also, the application or the data will be blocked to the unauthorized entity. The access control provided by various owners is shown in figure 1 and figure 2.

Figure 1: Owner Access Control

Figure 2: Cloud Access Control

The MAC registration done by the agent within the organization is shown in figure 3.

Figure 3: MAC Registration

4. Conclusion

In the real world, the organizations are facing the problem of data leakage. The data may be seen in other laptops or websites. This paper mainly presented a survey on the various data leakage detection techniques. In the proposed approach, the MAC address is traced with delegated access control so that the agent who leaked with data can be detected and data is blocked to the outside untrusted network.

References

[1] Hartung and Kutter," Watermarking technique for multimedia data", 2003.

[2] Priyanka Barge, Pratibha Dhawale and Namrata Kolashetti,"A Novel Data Leakage Detection" in International Journal of Modern Engineering Research, 2013.

[3] B. Sruthi Patil and M. L. Prasanthi," Modern Approaches for Data Leakage Detection Problems", in International Journal of Engineering and Computer Science,2013.