STRATEGY FOR DATA LEAKAGE DETECTION

Amrutha D V[1], Arka Jyoti Das[2], Pradeep Kumar H S[3]

[1] *Amrutha D V, Department of ISE, NIE, Mysore, Karnataka*
[2]*Arka Jyoti Das, Department of ISE, NIE, Mysore, Karnataka*
[3]*Pradeep Kumar H S , Department of ISE, NIE, Mysore, Karnataka*

ABSTRACT

This paper contains concept of data leakage detection, various causes of leakage and a strategy to detect the data leakage. The data is very important in current era, so it should not be leaked or modified. In any organization, huge database is used to store the data. This database is shared with multiple users simultaneously. During this sharing of the data, there will be chances of data vulnerability, leakage or modification. So, to overcome these problems, a data leakage detection strategy has been proposed. This paper outlines the data leakage detection strategy in a brief manner

Keyword : - *data leakage ,distributor, agent, and target etc….*

1. INTRODUCTION

Data leakage is a situation in which sensitive or private data is leaked intentionally or unintentionally to unauthorized entity. Sensitive data can be in any form depending upon the type of industry or organization based on the business they carry out. Examples are, financial information, credit card information, property information, patient database, etc. This sensitive information is shared among various stakeholders such as employees working from outside the organizational premises, business partners and clients. This might cause the risk of sensitive information falling into unauthorized hands, thus gaining access to private data. There will be a huge impact of data leakage on the organization leading to different kinds of loss. 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. Indirect loss includes reduced share-price as a result of the negative publicity; damage to company's goodwill and reputation, etc. Distributor is an entity wherein, he is distributes his sensitive information to trusted parties which are known as Agents. If the data has been leaked and found in unauthorized place, distributor should find out or assess from where and by whom the data has been leaked. Our goal is to propose a system that helps distributor in finding the guilty agents.

1.1 Existing System

The traditional method used for data leakage detection is watermarking. Watermarking is a technique where unique code is embedded in each distributed copy. If that copy is later found with unauthorized agent then guilty agent can be identified. Watermarks are very useful in the field of data leakage detection but the drawback of this technique is that, it involves the alterations or modification of the original data. And watermarks are not effective because it can 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 of sensitive customer information.

1.2 Proposed System

Our main goal is to detect when the distributor's sensitive data has been leaked by agents, and to identify the agent that leaked the data. We are using a hashing technique which is known as Secure Hashing Algorithm (SHA-1). SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function. SHA-1 produces a 160-bit (20-byte) hash value known as a message digest. In this method a randomly generated number is hashed and appended with the distributor's file. This file is then sent to agent. When one of the leaked file is recovered from an unauthorized source, it is then scanned for encrypted value. When the value is found, it is compared to the value stored in database. If it matches then the guilty agent is found and barred from further access to his account. This method of denying access to the guilty agent is called as Lock Down mechanism.

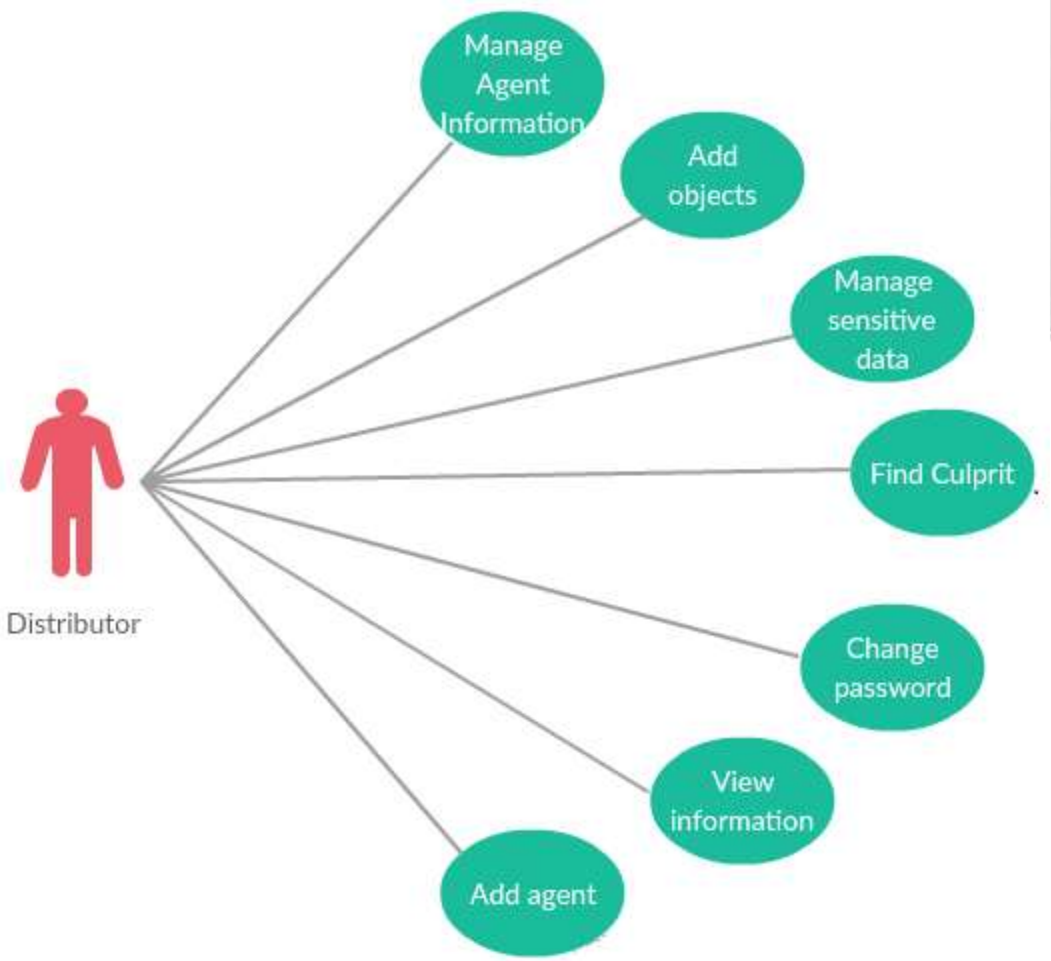

Fig.1.2.a : use case diagram for distributor

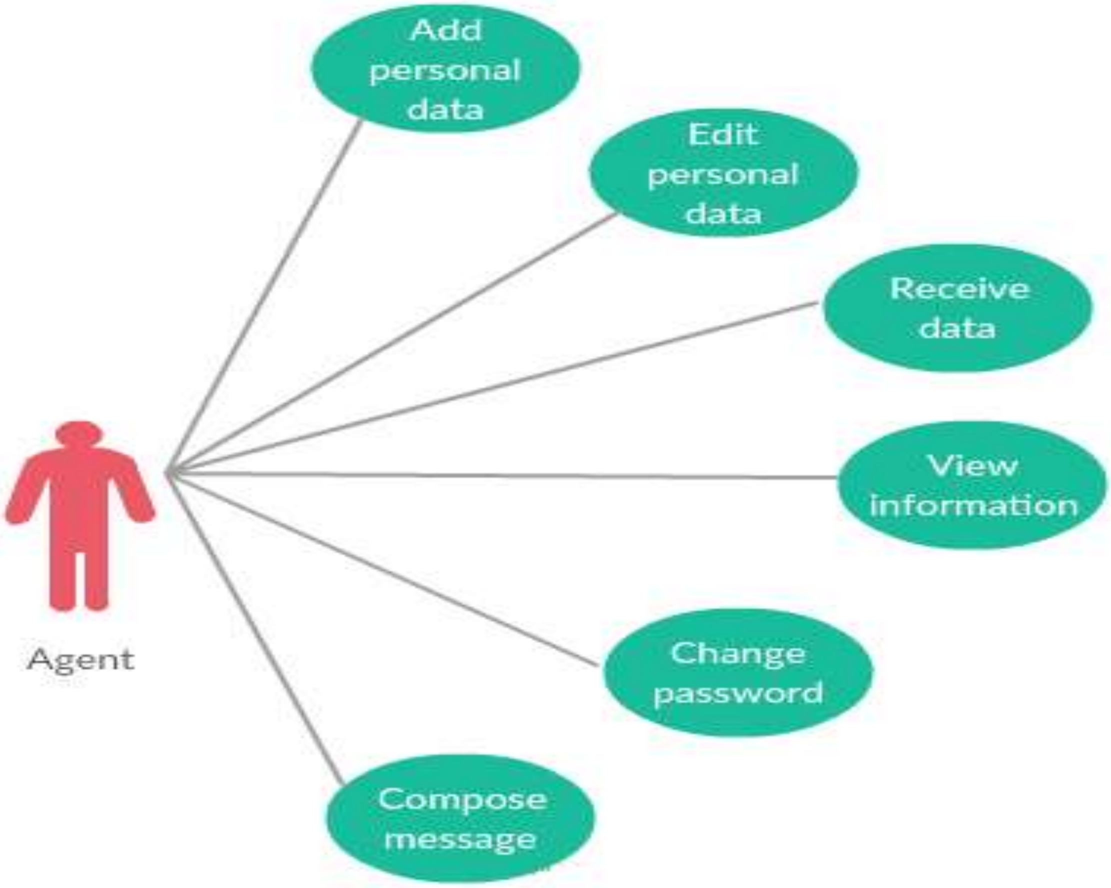

Fig.1.2.b : use case diagram for agent

1.3 Control flow diagrams

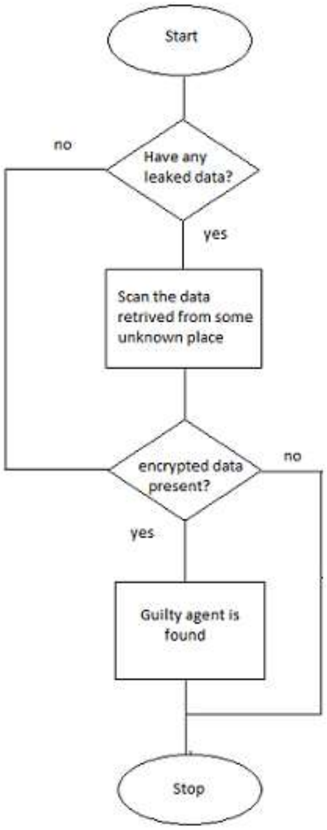

Fig.1.3.a Control flow diagram for detecting guilty agent

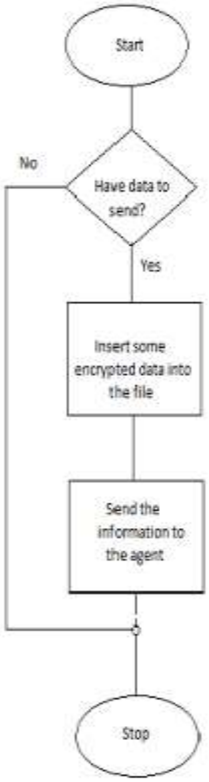
Start
No
Have data to send?
Yes
Insert some encrypted data into the file
Send the information to the agent
Stop

Fig.1.3.b Control flow diagram for inserting encrypted data into file and sending it to agent

2. LITERATURE SURVEY

Nowadays, most of the businesses, be it small-scale or large-scale all are making use of the network in some or the other form. It is seen that there is a excessive usage of cloud to store the data of the companies or organizations in the cloud storage which are more flexible i.e. they can be accessed remotely and can add storage value to it, then it is stored in physical drives, cloud is being used by most of the notable companies , marketplace and academic world. Examples for notable cloud services includes Amazon, Google, Microsoft, Yahoo, and Sales force. Also, top database vendors, like Oracle, as they are adding cloud support to their databases.

The providers are enjoying the opportunity to be able to build a market to sell cloud space or storage to the people who are willing to buy for a relatively low-cost as compared to physical storages, these low-cost and pay for use cloud can only work if the providers are able to provide the required amount of protection to the data. Quick provisioning, quick flexibility, everywhere network contact, hypervisor defense against network vulnerability, economical failure recovery and data storage solution, on-request security checks, synchronized detection of system altering and rapid re-construction of services.The cloud provides this compensation, until some of the risks are better understood. The basic concept of the cloud , based on the services they offer, form application service provisioning, grid and service computing, to Software as a Service. Despite of the specific architecture, the dominant concept of this computing model is that customers' data, which can be of individuals, organizations or enterprises, is processed remotely in unknown machines about which the user not aware. The ease and efficiency of this approach, however, comes with privacy and security risks. Confidentiality of data is the main hurdle in implementation of cloud services[1].

3. METHODOLOGY

In this paper we provide a description of all the steps that the data leakage prevention strategy will follow as to not all data to be accessed by unauthorized user outside the circle to which the administrator has determined.

In this paper we provide a description of all the steps that are followed by distributor in data leakage detection strategy in order to detect the guilty agent who leaks the sensitive data.

Steps to be followed are:

1. Distributor identifies trusted agents whom he has to send sensitive data.
2. A number is generated randomly, which is unique.
3. Generated number is then hashed using SHA-1 hashing technique.
4. Hashed value obtained as a result of hashing algorithm is appended to distributor's file. This Hashed value is also stored in database corresponding to each agent it is sent to.
5. File is then sent to trusted agents.
6. If any Distributor's file is the leaked by the agent, Distributor recovers this file from an unauthorized source.
8. File is scanned for hashed value, which was inserted before it was sent to all agents.
9. If a hashed value is found, it is then compared with the hashed value stored in distributor's database.
10. If a match is found, agent corresponding to that hashed value is considered as guilty agent.
11. That particular agent is denied of all kind of access and actions he could perform. This method is known as Lock Down mechanism.

4. CONCLUSION

In this paper, we conclude by saying that data leakage detection system is very useful as compared to watermarking model. We can provide security to our data during its distribution or transmission and even we can detect if the data gets leaked. Thus using this methodology security of data is ensured and detection technique is provided. This model is very helpful in various industries, where the data is distributed through any public or private channel and shared with third party. This system is relatively simple, but we believe that it captures the essential trade-offs.

5. REFERENCES

[1] https://en.wikipedia.org/wiki/Intrusion_detection_sys tem

[2] http://www.iosrjournals.org/iosr-jce/papers/vol1-issue3/S0132836.pdf

[3] https://www.paloaltonetworks.com/documentation/glossary/what-is-an-intrusion-prevention-system-ips

[4] https://vxheaven.org/lib/pdf/Introducing%20Stealth %20Malware%20Taxonomy.pdf

[5] https://digitalguardian.com/blog/cisos-guide-data-loss-prevention-dlp-strategy-tips-quick-wins-and-myths-avoid

[6] https://www.digitalocean.com/community/tutorials/w hat-is-a-firewall-and-how-does-it-work

[7] http://blogs.discovermagazine.com/sciencenotfiction/ 2010/11/15/information-converts-to-energy-at-28-percent/#.WLv2Z7M2vIU

[8] https://community.websense.com/blogs/websense-news-releases/archive/2015/03/23/research-penalties-punishment-amp-prison-for-serious-data-breaches-say-e-crime-congress-respondents.aspx

[9] http://www.webopedia.com/TERM/F/firewall.html