International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 11 | Nov 2021 www.irjet.net p-ISSN: 2395-0072

A System for Detection and Prevention of Data Leak

Aishwarya Jadhav[1], Prof. Pramila M. Chawan[2]

[1]M. Tech Student, Department of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Department of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *Technology is growing exponentially in the recent years and most of the organizations store their data in digital format. With the rapid growth in technology, there is a need for maintaining security of data. It is extremely essential as data loss might have huge effect on the organization. Preventing data leak has become one of the biggest challenge to the organizations. For the security purposes, the organizations have implemented several methods like implementation of policies, Firewalls, VPNs, etc. However, with the enhancement of data theft methods, these security measures are not reliable anymore. Hence there was a need for a system that can prevent data leak. Also, as employees have access to sensitive information of the company, they could leak the information either by negligence or on purpose. Hence, securing the data has become a big challenge for the organizations. In this article, we propose a system that will achieve the information security goals of the organization, and will be capable of detecting data leak at any states. The proposed system mainly focuses on preventing data leak.*

Key Words: **Data Leak Prevention, Sensitive Data, Data Leak**

1.INTRODUCTION

Security has become an important factor in our life. Security is required in all sectors of industry. An attacker has various methods to access the confidential information of any organization. Hence, preventing such attackers from accessing the information is the main aim of information security. We need to implement various strategies to secure the information.

Data leak occurs when the unauthorized users can access the confidential or sensitive data to. Data leak can happen intentionally through employees of the organization or malicious attackers. It can also be unintentional leak by employees. In any case, the data is transferred outside the organization. Data leak usually occurs through email. It can also occur through data storage devices such as laptops.

Data is one of the most precious asset for any organization. Therefore, the prevention of data leak is the most important task for any organization. Even with security measures like firewalls, data leak still occurs.

In any organization, the employees have access to sensitive data, Hence, there is a chance that the data leak occurs through employees rather than through malicious attackers.

1.1 Types of Data

Any organization must deal with three types of data to prevent data leak:

1. **Data in motion**

   It refers to data that is moving from the network to the outside world through the internet.

2. **Data at rest**

   It refers to data that is stored in the file systems, databases and other storage methods

3. **Data at the endpoint**

   It refers to data present at the endpoints in the network

Most of the organizations scan the emails that have been received from outside the organization for any malicious malwares. But, they do not check the emails sent outside the organization, thereby allowing the sensitive information to be sent outside the organization.

Most common causes of Data Breach are:

1. Hacking
2. Malware
3. Unintended Disclosure
4. Virus
5. Worms
6. Insider leak
7. Data loss

1.2 Causes of Data Loss

The following are common causes of data leak:

1. **Data loss due to natural disaster**

   The natural disasters like floods can destroy the hard disks, which leads to loss of data. In this case, data retrieval is possible through backups.

2. **Data loss due to improper handling**

   A disk can be damaged accidentally due to improper handling of the disk. This will lead to data loss. In this case, data retrieval is possible through backups.

3. **Data loss due to accidental drive format**

   In many cases, people accidentally format their drives and this leads to loss of data. In this case, data recovery is still possible.

4. **Data loss due to accidental deletion of data**

   In many cases, people can accidentally delete the data from the hard disk. Here, the data gets deleted unintentionally. In order to prevent this, the users should think carefully before they delete the data.

5. **Data loss due to intentional deletion of data**

   Sometimes, the users might delete a data intentionally from the hard drive and later on they want the data. The data can still be recovered from the recycle bin. If the recycle bin has been cleared, software to recover deleted recycle bin files can be used.

6. **Data loss due to corrupted system**

   Corrupted file system or database will inevitably lead to data loss. In this case, recovery of data is possible.

7. **Data loss due to power failure**

   If there is an occurrence of power failure and the user has not saved their file, there will be data loss. In order to prevent this, the users should keep saving as they work.

8. **Data loss due to software failure**

   In many cases, the software suddenly crashes or freezes while working. As a result, the program closes and all unsaved work is lost.

1.3 Causes of Data Leak

1. **Virus Attack**

   If a machine is infected by viruses and worms, spyware, adware, etc. this might result in corruption and loss of data. In order to prevent this, anti-virus should be used.

2. **Malicious Attack**

   Ill-intentioned and malicious attackers can hack into the system and steal, modify or delete valuable information. This will cause data leak.

Data leak prevention (DLP) is the practice for detection and prevention of data breaches and destruction of sensitive data. It is a set of tools and processes used to ensure that the unauthorized users do not access, delete or modify the sensitive data.

DLP software classifies data into different categories and checks for violations of policies defined by organizations. Once these policies are violated, DLP issues alerts, and other protective measures to prevent end users from accidentally or maliciously sharing confidential information.

2. LITRATURE REVIEW

In this section, summarization of existing research work is done. A new System for Detection and Prevention of Data Leak will be created based on the existing work with additional features.

In [1], the proposed system uses various techniques like Rule-based Regular Expressions, Database Fingerprinting, Exact File Matching, Partial Document Matching, Statistical Analysis, Conceptual/Lexicon which are used to secure the Sensitive data of an Organization.

In [2], the proposed system decides whether a particular chunk of data is permitted to be accessed or not. It uses an algorithm for data leak prevention with time stamp. Time stamp is used for giving permission to access a particular data, because in a particular period of time, the data is confidential and after the time stamp, the same data can become non - confidential.

In [3], the proposed system is based on a file system minifilter driver that will block unwanted file system operations, such as copying a confidential file to a removable storage device. The system allows to intercept and block I/O requests that originate from CopyFile or ReadFile APIs. It also blocks the external devices such as SD cards, USB drives or external hard drives.

In [4], a data leak prevention model is presented for classifying the data based on semantics. It uses data statistical analysis to detect evolved confidential data. It uses the information retrieval function Term Frequency-Inverse Document Frequency (TF-IDF) to classify documents under different topics. A Singular Value Decomposition (SVD) matrix was also used to visualize the classification results.

In [5], A data leak prevention method is proposed to check for data leaks in email communications of the organization. The method uses a combination of context and content analysis, which provides a better understanding of flow of confidential data in the organization. Contextual analysis is used to measure an RAI. It is achieved by calculating five context components in an email communication. Confidential data leaks to high RAI users are detected using Content semantics analysis. The results show the

possibilities of confidential data leaks among selected test subjects.

3. PROPOSED METHODOLOGY

3.1 Problem Statement

To develop a system for Detection and Prevention of Data Leak.

3.2 Problem Elaboration

In this system, the data will be secured by using DLP. DLP is a method to prevent the users from sending sensitive data outside the organization. The system is used to check for any activities of data transfer that might lead to data leak.

The system will control and monitor endpoint activities. It will also monitor the data in the cloud to protect data at rest, in motion, and in use. It will also generate reports and identify weakness in the system to enhance the security.

3.3 Proposed Methodology

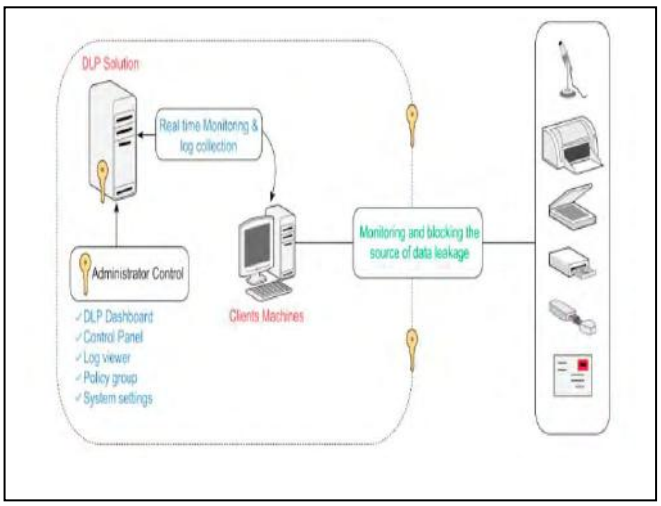


Fig -1: Data Leak Prevention model

The data leak can be prevented with the help of DLP. It can be implemented using the following measures:

1. Implement a centralized DLP which takes care of all states of data.
2. The involvement of skilled people with appropriate organization model.
3. Classification of data into two types: sensitive and non-sensitive
4. Implementation process has to be in phased manner
5. The DLP implementation should not affect the workflow of the business process
6. Reports must be more effective
7. Enough security measures must be taken before implementing the DLP

4. CONCLUSION

Data leak is a major issue for many organizations. Data leak can have a disastrous effect on any organization. Hence, preventing data leak is very important. In this paper, a system is proposed for detection and prevention of data leak, which will achieve the security goals of an organization. The proposed method is easy to implement and can be useful for many organizations.

REFERENCES

[1] S. Czerwinski, R. Fromm, and T. Hodes, "Digital Music Distribution and Audio Watermarking,"http://www.scientificcommons.org/430256 58, 2007. Available at: www.researchpublications.org NCAICN-2013, PRMITR,Badnera 399

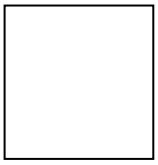
[2] Y. Li, V. Swarup, and S. Jajodia, "Fingerprinting Relational Databases: Schemes and Specialties," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 34-45, Jan.-Mar. 2015.

[3] Y. Cui and J. Widom, "Lineage Tracing for General Data Warehouse Transformations," The VLDB J., vol. 12, pp. 41-58, 2014.

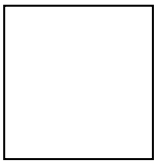
[4] Panagiotis Papadimitriou and Hector Garcia-Molina, "Data Leakage Detection," IEEE Trans, Knowledge and Data Engineering, vol. 23, no. 1, January 2013.

[5] P. Bonatti, S.D.C. di Vimercati, and P. Samarati, "An Algebra for Composing Access Control Policies," ACM Trans. Information and System Security, vol. 5, no. 1, pp.1-35, 2011.

BIOGRAPHIES



Aishwarya Jadhav is currently pursuing M. Tech from VJTI COE, Mumbai. She has done her B.E. (Computer Engineering) from Atharva College of Engineering.



Prof. Pramila M. Chawan, is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B.E. (Computer Engineering).