Data Leak Prevention System

Prasad Jadhav[1], P. M. Chawan[2]

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

---***---

Abstract - *In recent years, many companies and institutions tend to keep most of their data in digital format. There is a large specter of information stored by such entities, like medical records, contracts, internal procedures, etc. that can be considered as being confidential, thus protecting them is a great concern. Since employees have access to such information, either by negligence or bad intention, they could leak the information. Hence, information security has become a big concern for the organizations.*

In this article we propose a software architecture known as Data Leak Prevention System that can achieve the goals of information security of the organizations. This architecture is designed to highly regulate access to private data, and furthermore, to identify which parts of the system can be subjected to external hacking or inside attacks.

The proposed architecture focuses mainly on preventing massive data leaks. The architecture guarantees that any access to sensitive data is logged into an external system which cannot be affected by the attackers.

Key Words: DLP - Data Leak Prevention

1. INTRODUCTION

This Currently, security has become an essential factor in our day to day life. Security is required in Industrial sector as well as government sector. A malicious attacker can use various methods to access the private information. To avoid this is one of the goals of the information security. As we know that we require security in our daily life, similarly we need to implement various strategies to secure the information.

Data leak is the unauthorized exchange of data between an organization and an external destination or recipient. A data breach or data leak is the release of confidential or sensitive information to the unauthorized users. Data leak can happen because of a programmer assault, intentional leak by employees of the organization, or unintentional loss or exposure of data. It implies that the data is transferred electronically or physically. Data leak usually occurs via the web and email. It can also occur via mobile data storage devices such as optical media, USB keys, and laptops.

Data leak is also known as data theft and is a huge problem for data security. Regardless of size or industry, it can cause serious damage to any organization. Any organization will want to protect themselves from threats like declining revenue, tarnished reputation, massive financial penalties or lawsuits. There are many different types of data leak like Accidental Breach, Disgruntled or Ill-Intentioned Employee, Electronic Communications with Malicious Intent, etc. and it is important to understand that the problem can be initiated via an internal or external source. The following are common causes of data loss.

1. **Natural Disaster:** Your hard drive can be damaged due to fire, flood or some other unforeseen disasters. However, the data can still be retrieved in such situations.

2. **Accidental Damage**: If a drive or disk is mishandled or accidentally dropped, this may cause trauma and loss of data. Data recovery is also possible in this case.

3. **Accidental drive format**: Sometimes users accidentally format their drives and this results in instant loss of data. However, it is possible to recover your data in a situation like this. Users can get help from the experts.

4. **Accidental Deletion of Data**: There are times when you accidentally delete a file or a program from your hard drive. This is an unintentional deletion which may go unnoticed for a long time. Administrative errors also fall under this category. The best way is to think carefully before you delete any data or program.

5. **Intentional Deletion of Data**: You may have deleted a file intentionally from your system and later decided you wanted the file back. You can still recover your data from the recycle bin. If you have emptied your recycle bin, you can use software recover deleted recycle bin files.

6. **Corrupted Data**: If your file system or database is corrupted, then it will inevitably lead to loss data. At the same time, it is possible to recover data from a corrupt file system using an appropriate software tool.

7. **Power Failure**: If you experience power failure before you have the opportunity to save your work, you may lose valuable data. It is better to keep saving as your work.

8. **Software Failure**: When the application software suddenly crashes or freezes while working, this may

result in severe damage to the hard drive. This causes the program close suddenly and all unsaved work is lost.

9. **Virus Attack**: If a machine is deeply infected by viruses and worms, spyware, adware and some deadly computer parasites, this can lead to total corruption and loss of data. Installing a very good anti-virus program will reduce the possibility of having a fatal virus attack.

10. **Malicious Attack**: Professional hackers or competitors can invade into the system and destroy it. This will obviously lead to loss of data.

Data Leak Prevention (DLP) is the practice of detecting and preventing data breaches, exfiltration, or unwanted destruction of sensitive data. Organizations use DLP to protect their data and comply with regulations and policies. The term DLP refers to defending organizations against both data loss and data leak prevention. The term data leak refers to an event in which important information to the organization is leaked to the unauthorized environment. Data leak prevention focuses on preventing illicit transfer of data outside the institutional boundaries.

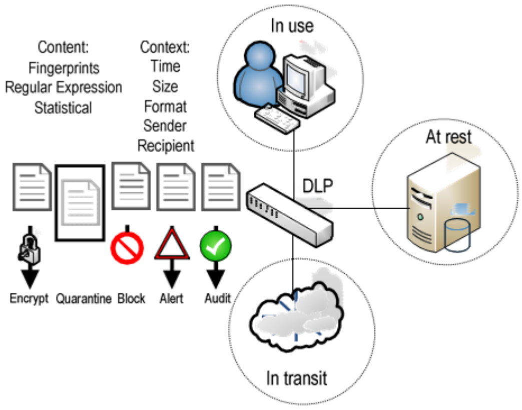

Figure 1. DLP: Data Leak Prevention Model

Data Leak Prevention (DLP) systems are increasingly being implemented by various organizations. Unlike the standard security mechanisms such as firewalls and intrusion detection systems (IDS), the DLP systems are designated systems which are used to protect in use, in transit and at rest data. DLP system analytically uses the content and surrounding context of confidential data to detect and prevent unauthorized access to confidential data.

DLP system that use content analysis techniques are largely dependent upon regular expressions, data fingerprinting, and statistical analysis to detect data leaks.

2. PROPOSED SYSTEM

The proposed system suggests a Data Leak Prevention System which will continuously monitor the activities of the employees and will notify the admin if any malicious activity is detected.

2.1 SYSTEM PARAMETERS:

The system uses a set of parameters to identify Data Leak which are as follows:

- ➢ Time Restriction
- ➢ Extension Restriction
- ➢ Keywords Restriction

1) Time Restriction

Time Restriction restricts the employees through time constraints. Here, the admin can decide a particular time frame for data transmission. The transmission is permitted only in this time frame and will be restricted outside this time frame.

2) Extension Restriction

The Extension Restriction technique is used for restricting the files with such extensions which cannot be read by the system. Transmission of such files pose a threat to the confidentiality of the information of an organization. The system can only read files in the following format:

- ➢ Excel (.xls)
- ➢ Word (.doc)
- ➢ PDF (.pdf)
- ➢ Text File (.txt)

Hence, the Extension Restriction feature allows the admin to block all other types like Jpg, jpeg, png, mp3, mp4, etc. which cannot be read by the DLP system.

3) Keywords Restriction

Keywords Restriction feature enables the admin to block the transmission of files which includes any confidential data. The admin can set the keywords to block the transmission of such files.

2.1 SYSTEM ARCHITECTURE:

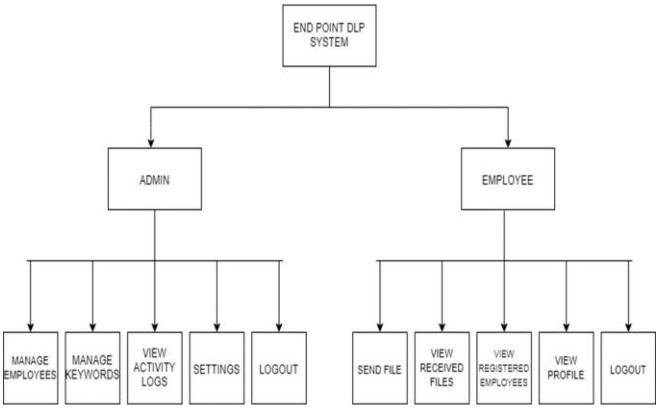

Figure 2. System Architecture

The system comprises of 2 major modules with their sub-modules as follows:

- ADMIN
- EMPLOYEE

Sub - Modules of Admin:

- **Login:** Admin can login using credentials.
- **Manage Employees:** Admin will be adding, updating and deleting employee details. Once a user is registered login credentials will be sent via mail to the respective employee.
- **Manage Keywords & File Extensions:** Admin can add, update and delete keywords and file Extensions.
- **View Activity Logs:** Admin can see the activity logs.
- **Settings:** Admin will Set file sharing timings as well as Email id to receive incident email.

Sub - Modules of Employee:

- **Login:** Employee can login using received credentials via mail
- **Send File:** Employee can send file to anyone inside or outside the organization. If any suspicious activity found, an incident mail will be sent to admin.
- **View Received Files:** Employee can download the received files.
- **View Registered Employees:** Employee can also see other employees.
- **View My Profile:** Employees can view their profile.

3. WORKING

Admin monitors the activities of employees to check whether there is any suspicious activity or not. The employee can send file to anyone inside or outside organization.

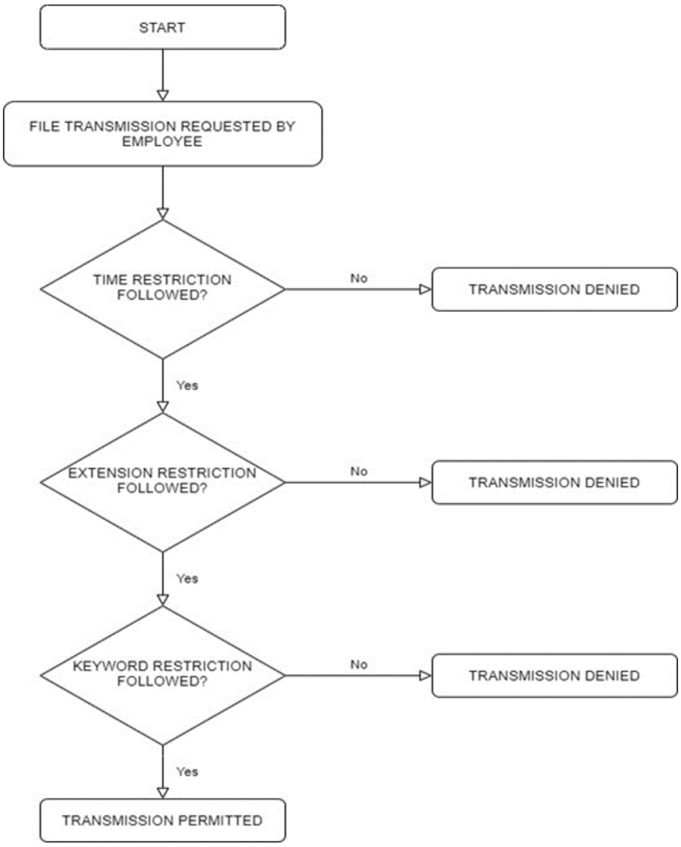
START
FILE TRANSMISSION REQUESTED BY EMPLOYEE
TIME RESTRICTION FOLLOWED?
No
TRANSMISSION DENIED
Yes
EXTENSION RESTRICTION FOLLOWED?
No
TRANSMISSION DENIED
Yes
KEYWORD RESTRICTION FOLLOWED?
No
TRANSMISSION DENIED
Yes
TRANSMISSION PERMITTED

Figure 3. Operational Flowchart

Whenever there is a mail/file to be sent from employee to outside world, the DLP system will check all the conditions/restrictions set by admin and will either block the mail or allow the mail.

The system will identify data leak based on the three parameters: Time Restriction, Extension Restriction and Keywords Restriction.

The system will also generate mail to admin and create activity log.

4. CONCLUSION

In this paper, a Data Leak Prevention System is proposed to detect and prevent the data leak, thereby achieving the security goals of an organization. The proposed technique has been tested against different scenarios in which the DLP system dealt with various types of data. This technique is simple, 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

Prasad Jadhav is currently persuing M. Tech from VJTI COE, Mumbai. He has done his B.E.(IT) from Sardar Patel Institute of Technology

Pramila M. Chawan is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B. E.(Computer Engg.) and M.E (Computer Engineering) from VJTI COE, Mumbai University. She has 27 years of teaching experience and has guided 75+ M. Tech. projects and 100+ B. Tech. projects. She has published 99 papers in the International Journals, 21 papers in the National/International conferences/symposiums. She has worked as an Organizing Committee member for 13 International Conferences, one National Conference and 4 AICTE workshops. She has worked as NBA coordinator of Computer Engineering Department of VJTI for 5 years. She had written proposal for VJTI under TEQIP-I in June 2004 for creating Central Computing Facility at VJTI. Rs. Eight Crore (Rs. 8,00,00,000/-) were sanctioned by the World Bank on this proposal.