Data Leakage threats and measures to overcome it

M.Jayaprabha, P.Felcy Judith
Assistant Professor, Associate Professor
T.John college,Bangalore

Abstract - Data is an asset to every organization and all measures are taken to protect it from unauthorized access and loss. Data loss can be detected with crucial data breach, data transmissions and prevents them by monitoring, detecting and blocking sensitive data while in-use or in-motion or at-rest. In modern business world, almost all organizations use technologies to manage their sensitive information. There is a requirement to protect such a key component of the organization cannot be overlooked. Data Leakage Prevention has been one of the most effective ways of prevent Data Loss. DLP detects and prevents unlicensed access to copy or send sensitive data, both intentionally and unintentionally. This generally happens by people who are authorized to access the sensitive information. DLP is designed to detect critical data breach in real time. This can be achieved by data monitoring. Data Loss Prevention is designed to fit with the organizational structure. It helps the organization protect both structured data and unstructured data. The data may be Personnel's personal information, financial information, business secrets, merger and acquisitions, health records etc. DLP technology is not only for large organizations but can also be used by other sector processing sensitive information like banking, Health care, aviation, government etc.

Introduction

Data Leakage Prevention (DLP) is a new technology that focuses on big problem like data leakage. Many of the problems that emerge are by unauthorized accesses. Such incidents can be handled by law suites due to data breach rights. There are many products available in the market to prevent data leakage. But all of them should be reviewed and tested several times before it could be used in the companies.

This paper covers• the unintentional data leakage • Possibilities to prevent data leakage • check for vulnerabilities in the software • A set of tests are to prevent data loss and • examine which forms of data leakage can be avoided by implementing a DLP suite. Some of the tools used may be encryption, or embedment of unclear information or keywords using different data types. There are two extra-ordinary DLP solutions like new McAfee Host Data Loss Prevention and the Websense Data Security Suite which is one of the leading products in the market. But both solutions contain several flaws which make them fail in few areas. McAfee did not delete confidential files and sensitive data while copying to external devices. The encryption of Data Security Suite is not sufficient in Web sense.

Few examples to show the different kinds of Data Leakage across the world

1. Loss of trust in an organization: In 2008, a camera was sold on Ebay that contained pictures of terror suspects and internal classified documents of MI6 of the British intelligence service. This kind of unintentional leakage is a loss of trust of an organization.
2. A British organization lost a CD containing data of more than 11,000 teachers. The CD was sent to another once through courier, but did never arrive at the destination. Fortunately, all information was encrypted so that nobody can use the lost data. Data becomes information when it is interpreted and transported.
3. Another biggest leakage happened at Germany with T-Mobile in 2008 where 17 million customer data were stolen due to security vulnerabilities in different systems and databases.

To overcome these problems the www consortium gave website privacy rights to protect the data flowing across the websites. It provides a database of all the leakage incidents.

How does data leakage happen and its effects:

Types of data leaked

1. Confidential information 2. Intellectual property 3. Customer data 4. Health records

Intentional / Internal data leakage

This kind of leakage happens via - Remote Access, Instant Messaging, Email, Web Mail; Peer-to-Peer, and even by File Transfer Protocol.

E-mail is a personal document may be sent to an unauthorized individual as an attachment. They may also choose to compress and / or encrypt the file, or embed it within other files. Steganography – a process of concealing a sensitive data within a non-secret data may also be utilized for this purpose.

Web Logs / Wikis →Web logs could be used to release confidential information into the internet, simply by entering the information in their blog. Wikipedia site is a common website which can be directly edited by anyone with access to it", such as wikipedia.org. These sites are mostly available to all internet users around the world, and possibly add confidential information into a wiki page.

FTP is a File Transfer Protocol that may be added through the firewall which is an intentional leakage. Uploading a file to an FTP server can be made only by a person who is aware of the process and therefore cannot be done by an average user on a daily basis

Removable Media / Storage are USB, hard drives, digital cameras, and even musical devices such as an Apple iPod
Others may take a hard copy of the document in briefcase and share it with other unauthenticated person or take a picture of the doc and send it through mobile phone. Files and folders may not be provided with authentications, allow sensitive data to be leaked or write inadequate queries in database.

External Threats

Phishers may start developing fraudulent employment web sites, and attempt to attract users to send their resumes directly to them on job site users. Hackers not only grab resumes but also e-mail ids and credit card numbers of job seekers by disguising as a trustworthy entity in websites.
SQL injection is embedding malicious code in DB query which is made by person who has good knowledge on Database.
Dumpster diving Sometimes organizations may not be destroying the hard copy information securely which may run the risk of confidential information falling into unauthorized hands. An attacker may decide to scrutinize the company's dumps and discover important information. The information may be stored in external devices like CDs or DVDs or printed document.

Safeguarding confidentiality

The loss of confidentiality is a common problem when engaging with information security. There are lots of approaches which follow a variety of ways to safeguard the confidentiality of data.

Hippocratic Databases

Hippocratic Databases are aimed that more granular access controls exist to ensure that only the owner should be authorized to access the database system. This can be achieved by attaching fake attributes to all stored information. These attributes allow fine grained access control. Another crucial requirement is the absence of side channels i.e. the executed queries should not provide additional information like statistical data which is based on a small number of data sets.

Email Leak Prevention

Sending and receiving Emails has become one of the most important communication mediums. Consequence of this is arising threat of email leakage. Emails containing confidential data may be sent to wrong recipients – e.g. due to misspelling or wrong use of the auto completion, a feature of modern mail agents which completes email addresses after the first letters.

Current DLP Approaches

There are three main capabilities of DLP solutions: • Identify • Monitor • React
Each of these steps in leakage prevention has to handle data at rest, in motion, and in use.
Identify: How to Find Valuable Content
First identify the valuable data. A central management should induce protocols and policies to be consistent and manageable. Rule-Based Regular Expressions are the most common technique for defining data via an abstract pattern. This approach produces a high rate of false positives due to the limited scope and missing context awareness. For example the word confidential or important can be with various confidential contexts. While processing, they can be used as a filter to reduce the amount of data for further processing.
Database Fingerprinting- is a biometric identification (ID) methodology. It uses digital imaging technology for storing and analyzing fingerprint data. Finger prints of authorized people should be stored in dB already. Again when the person is trying to access the data then the finger print matching should be done.
Cyclic hashing – is a process to scan large data in an efficient way. The first hash value indexes the first N characters in the document, the next hash value covers the next part which includes an overlapping. Thereby, it is important that the resulting index contains an overlapping map of the document. If suspicious documents should get examined, the same algorithm can be used to determine whether there is sensible data included. But it produces a high CPU load due to excessive hashing.

Data can be monitored at use, in rest and in motion

In motion - Data in motion can be an email or a FTP file transmission. The complete pathway or network must be monitored while transmitting the data. There are several monitoring points to intercept traffic like the web proxies, mail servers to monitor the complete data flow.
At Rest- An organization can recognize where its sensitive data is distributed by scanning the network sharing methods. Sometimes a software agent is used to assess the endpoint systems or application servers.
In Use- it is not possible to remotely monitor the data in use. To control every action, the user may take an endpoint agent to assess. This agent hooks up into the operating system functions to recognize all actions.

Handling Policy Breaches

There are several ways to handle the data loss situation. It is necessary to define appropriate policies to prevent data leakage rather than protecting it.

i) It would be sensible to delete documents in the public file to avoid any leakage. The file should be moved to a secure place leaving a message where it has moved.
ii) Encryption is another way to secure discovered data.
iii) The data can even undergo the initial phase of reconnaissance to build and improve a basic rule set.
iv) E-mail channels can be protected by proxies and gateways. This can be implemented using a plug-in by providing a dedicated service which acts as a proxy. Another way is to terminate the connection by TCP reset packets using sniffed sequence number.

General Problems

Some general problems include lack of awareness, no sufficient technical knowledge and man power in companies where security is needed to operate networks and systems. Investment to operate security solutions like administration, development and research is high. These circumstances will lead to loss of data. To avoid these problems DLP has deployed software, plug-in, and endpoint agents in every network to block all confidential data. One drawback to use DLP solution is its complexity. So to use any technology in a safe way it is necessary to understand the solution. But there are lots of notifications that remind the user that he is accessing valuable data. Another problem may be when a user prints a sensible document and by mistake throws it to the trash; it may probably reach an outsider. Even if a DLP solution could help by showing up a warning message, it is questionable whether awareness courses would not have been more efficient. Few studies show that 50% of all information leakage caused by insider threats are conducted by administrators as they have access to almost every data since they have to administer it. Also they know how a deployed DLP solution works and are able to disable or bypass it. Therefore the DLP solution loses lots of its benefit. Regardless on the latest idea of data breaches there are few classical problems like software vulnerabilities, inappropriate configuration of systems and weak passwords.

Alternative Solution Statements

There are many approaches for the improvement of confidentiality. To understand alternative solutions to data leakage, it is necessary to understand the different types of leakage. Data leakage can be classified broadly into two- intentional and unintentional. If suppose an Email is sent to a wrong recipient due to a typo error or use of the auto completion - new feature in modern mail, there is no way to get the email back. An error message can appear after every operation to confirm the recipient and mail-id if there are 2 similar names. An additional check is also done to check whether the email should really be with confidential data. This repeated error messages can reduce the data leakage. A loss to the reputation of an organization can happen even if a CD is entirely encrypted – when sending data. But its loss would result in data loss. A risk analysis should be performed at regular intervals to check whether the existing systems and networks are secure. Multilevel Security system or access control lists can help to restrict access on data. Another possibility would be the use of server based computing ie use thin clients which connect to terminal servers which reduces vulnerable data loss.

Conclusion

However data breach can be prevented by

1. Implementing referential integrity and concurrency constraints, data access policies, data masking and encryption techniques for changing original data values, and checksums for integrity checks on changed data at Data Warehouse
2. Using attestation process, peer review technique, Cloud Safety Net (CSN), a light weight approach for monitoring data propagation in PaaS clouds to discover data leakage between tenants, Traffic monitoring, Log/ Identity management in Cloud.
3. Confidentiality, Data Integrity, Availability, Replay Protection and storing data packet and replay it at later stage in IoT
4. Make protocols and policies, only authenticated person can make decisions in Data Mining
5. Regular Expressions are used to detect patterns of digits or characters. For example a 16 digit sequence could represent a credit card or debit card number.
6. Clustering is a technique which focuses on groups of documents which are similar, by correlating words, word counts, and patterns across the group of documents

References

[1] DATA LEAKAGE DETECTION , IJCSMC, Vol. 2, Issue. 5, May 2013, pg.283 – 288, ISSN 2320–088X, Ms. N. Bangar Anjali1, Ms. P. Rokade Geetanjali2, Ms. Patil Shivlila3, Ms. R. Shetkar Swati4, Prof. N B Kadu5

[2] Data Leakage Detection and Data Prevention Using Algorithm, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 4, April 2017, ISSN: 2277 128X

[3] Data Leakage Detection, International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2012, Sandip A. Kale1, Prof. S.V.Kulkarni2 Department Of CSE, MIT College of Engg, Aurangabad, Dr.B.A.M.University, Aurangabad (M.S), India1,2.

[4] https://www.veracode.com/security/guide-data-loss-prevention

[5] https://digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention

[6] http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Data-Leak-Prevention.aspx

[7] https://ieeexplore.ieee.org/abstract/document/5487521

[8] Data Leakage Detection, International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, p-ISSN: 2395-0072, Ghagare Mahesh1, Yadav Sujit2, Kamble Snehal3, Nangare Jairaj4, Shewale Ramchandra5