An Overview of Data Leakage Prevention System

[1]Vaishnavi Kature, [2]Ankita Washimkar
[1] KDK College Of Engineering, Nagpur ,[2] KDK College Of Engineering, Nagpur.
[1]307vaish@gmail.com,[2]ankitawashimkar96@gmail.com

Abstract— An enterprise data leak causes a big loss of organization. As the organization progresses into the more technological environment, the amount of the digitally stored data increases dramatically, but keeping the track on data used in any organization is no longer as easy as before .Security practitioners have always had to deal with data leakage issues that arises from various ways like email and other internet channels. Data leakage is the unauthorized transmission of data or information within an organization or from an organization to the external destination. The data stored in any device can be leaked in two ways; if the system is hacked or if the internal resources intentionally or unintentionally make the data public. Hacking can be prevented by carefully configuring your Firewalls and other security devices but if the internal resources make data public then we must have some preventing solutions. DLP is used to identify sensitive content by using deep content analysis to per inside files and with the use if network communications. DLP is mainly designed to protect information assets in minimal interference in business processes.

Index Terms— sensitive data, data leakage, internal attack, external attack, data leakage prevention

I. INTRODUCTION

Data Leakage is an incident when the confidentiality of information has been compromised. It refers to an unauthorized transmission of data from within an organization to an external destination. The data that is leaked out can either be private in nature and are deemed confidential whereas Data Loss is loss of data due to deletion, system crash etc. Totally both the term can be referred as data breach, has been one of the biggest fears that organization face today.

Data Leakage Prevention (DLP) is a computer security term which is used to identify, monitor, and protect data in use, data in motion, and data at rest [1]. DLP is used to identify sensitive content by using deep content analysis to per inside files and with the use if network communications. DLP is mainly designed to protect information assets in minimal interference in business processes. It also enforces protective controls to prevent unwanted incidents. DLP can also be used to reduce risk, and to improve data management practices and even lower compliance cost.

As organization was facing issues with data loss, the objective of our paper is to analyze the evaluation of how well DLP fills security gap in comparison with previously used technology in a motive to solve data leakage problem. This is a very important need for the capability to exchange confidential information securely and easily as the organization is dealing with sensitive payroll data. This is done by doing a detailed study and a case research on Data Leakage Prevention technology in organization.

II. LITERATURE REVIEW

The issue of data loss or data breach has been one of the biggest fears that most of the organizations face today. In some organizations, there is a wide hole in controlled and in secure environment which was created to protect electronic resources. This hole is the way where the business and individuals communicate with each other over the Internet [2].

Data loss prevention (DLP) is interested in identifying sensitive data and also is one of the most critical issues facing CIOs, CSOs and CISOs. DLP is now today's strict regulatory and ultra competitive environment. In creating and implementing a DLP strategy, the task can seem to be intimidating. For this the effective solutions are available. This paper presents best practices for preventing leaks, enforcing compliance, protecting company's brand value and reputation in organization [3].The data loss issue is being exposed from confidential information about a customer to dozens of company's product files and documents being sent to a competitor. This can be caused in many ways either accidental or deliberate, or even with insiders in realizing sensitive data about customer's personal information, intellectual property, or other confidential information in violation of company policies and regulatory requirements. Here considering few of high-profile examples:

- AOL posts search engine data contains personal information about its members.
- DuPont employee leaks $400 million in intellectual property.
- Former Ceridian employee who accidentally posts ID and also bank account data for 150 employees of an advertising firm on a website Like the above there are many more data loss problem occurred and the list goes on [4].

In organization, today's employees with available access to electronically expose sensitive data, the scope of sensitive data loss problem is greater than outsider's threat protection. In order to cover all the loss bearings, an organization has the potential to encounter:

- Data in motion – Any data that is moving through the network to the outside via the Internet.
- Data at rest – Data that resides in files systems, databases and other storage methods.
- Data at the endpoint – Data at the endpoints of the network (e.g. data on devices such as USB, external drives, laptops, mobile devices, etc) [3, 4].

Now the world is connected using electronic communication where we are electrically connected in numerous ways. It doesn't matter where we are around the globe. Accessing the electronic data has become more crucial in day-to-day business.

For instance, many companies have development offices at offshore level, and /or international level in which all were exponentially increase the opportunity for data loss. Confidential information can travel even to the far corners of the earth using simple email communication.

Over the years, organizations have spent large amount of resources in a motive to protect their information. In their effort, majority was focused on preventing outsiders from hacking into the organization. Unintentional information loss from employees and partners are the results of majority of all leaks in leading firms. Research conducted on data loss prevention indicates that more than half of security breaches are caused by insiders. In an organization, employees can cause a sudden damage for their company even with the simple click of a mouse [3], [5].

II. EXISTING SYSTEM

There is a large security gap between the existing systems which are used to prevent the data leakage and the real life scenario. Gap is sometimes called, the space between where we are and where we want to be. The gap analysis is undertaken as means of bridging that space. It is a technique for determining the steps that are need to be taken in moving form a current state to desired future state. It begins with questionnaire “what is” and proceeds to “what should be” and finally highlights the gaps“ that exist and need to be filled”. Here comes what is security gap? Security gaps are nothing but the vulnerabilities or weakness in the organization which is a threat and can be exploited to make an attack.

There are two ways of attacks such as External and Internal. External Attacks are those attacks which are done by hackers and other people from the outside of an organization network. It is done by finding the vulnerability and exploiting that to make an attack. Malware infection, DDOS attack, Man in the middle are few types of attack which are done to gain monetary benefits or to harm the organization assets. Internal Attack is performed from the internal perimeter of the organization by a disgruntled employee, contractors or vendors either for monetary benefits or to take away some confidential, sensitive data out of the organization. Software code, PCI DSS information, financial reports, NER report are few examples of inside attack which are performed from inside of the network.

Why the gap is a problem? The gap becomes a problem when there is a false feel of information security is created as this false feeling does not protect against threats. This might due to the causes such as organization may not be aware of information security risk to their operations, by default acceptance of unknown level of risk, unconscious deciding on risk level, relaying on ineffective controls, not able to justify the spending of security, etc.

Though organization has many security frameworks and techniques that are available today but the overall security structure or measurements is far from acceptance. The false feel of security has various causes such as interests, language, education, uncertainty, knowledge, view on process control, and methods to handle information (in) security. All these can be looked at in various ways [6].

PROPOSED SYSTEM

Data Leakage Prevention (DLP) is a computer security term which is used to identify, monitor, and protect data in use, data in motion, and data at rest. DLP is used to identify sensitive content by using deep content analysis to per inside files and with the use if network communications. DLP is mainly designed to protect information assets in minimal interference in business processes. It also enforces protective controls to prevent unwanted incidents. DLP can also be used to reduce risk, and to improve data management practices and even lower compliance cost.

Data Loss Prevention is found to be the data leakage control mechanism that fits naturally with the organizational structure of businesses.DLP is considered as preventive control which when applied helps organization prevents data leakage of sensitive information (Personal identifiable information, financial information, trade secrets etc.)

A. Usage of DLP Technology in contrast with other existing technologies in organization

There are various technologies being used in the organization to prevent data loss. Though these technologies are very powerful but can help majorly an outside attack on data, whereas the current DLP technology deployed is mainly focused on inside attacks. Below are the currently used technologies in the organization for preventing security breaches. Here we concentrate on how these technologies are addressing the security issues in comparison with DLP.

1) Anti-Malware: Anti-malware is software used to protect malware attacks on computers, this software get into the operating system's core or kernel functions in the same way as malware, which attempt to operate from there. Each time the operating system does some job, the anti-malware software checks that the OS is doing an approved task. Though this anti-malware software works in real time environment very effectively but it only looks for threats from outside, by scanning and signature validation it ensures that malware infection be removed. This anti-malware software helps in data loss prevention from external threats but for internal threats it doesn't have any mechanism.

2) Firewall: Firewall is a software or hardware that helps in keeping network secure. Its objective is to control the incoming and outgoing traffic of networks by analyzing the data packets and determining whether it should be allowed through or not. A network's firewall frames a brigade between an internal network (Secure), and an external network, i.e. Internet (Insecure).

There are different types of Firewalls used in the organization, and it is one of the best security features to be implemented. But the major problem here is Firewall works on Access Controlled List often know as ACL"s. These ACL"s either allow or deny completely. For example if a rule is set to deny any outgoing traffic with certain set of data, then it will block all such traffic and it will not even allow the legitimate traffic to flow.

3) IDS/IPS: Intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities. It identifies a potential security breach, and logs the information and gives an alert by signaling.

Intrusion prevention systems monitor network, system activities for malicious activity. It mainly identifies malicious activity, log information, attempt to block/stop activity, and report activity. Though both IDS/IPS and Firewall relate to network security, an intrusion detection system (IDS) differs from a firewall in that a firewall looks outwardly for intrusions in order to stop them from happening. Firewalls limit access between networks to prevent intrusion and do not signal an attack from inside the network. 4) SIEM: Security Information Event Management (SIEM) is a tool used on enterprise data networks to centralize the storage of logs which was generated by the software running on the network. This has various features such as gathering information, analyzing the information and also presenting the information from network and security devices; identity and access management applications; vulnerability management and policy compliance tools; database logs; application logs; external threat data and OS. It monitors and helps manage user and service privileges, and directory services; as well as providing log auditing and review and incident response. Though this technology can collect events or logs and store for certain period of time but it doesn't have the capabilities of preventing/protecting data loss

All the above technologies are used to prevent external attacks and act very minimal for preventing Insider attacks/threats.

In contrast to the above technologies used for Loss Protection/Prevention, DLP provides a policy based approach to secure data. It enables customers to classify their sensitive data, discover data across the enterprise, enforce controls, and generate reports to ensure compliance with established policies.

III. PROPOSED PLAN OF WORK

To prevent data loss in the organization by using Data Loss Prevention technology, the organization is in need for the capability to exchange confidential information securely and easily.

1. Confidential Data
   - Credit Card / Client Information
   - Customer privileged data
   - Employee personal data
   - Business Confidential data
2. Secure data from
   - Employee Error, Employee Theft

     Data stored in any storage device can be leaked in two different ways.

     1. If the system is hacked.
     2. If internal resource intentionally/ un-intentionally makes the data public.

Hacking can be prevented by carefully configuring your Firewalls and other security devices. We will be discussing the second scenario i.e. if an internal resource makes the sensitive data public. Consider the possibility of an employee leaking the sensitive data. Now there are various ways in which data can leave the organization via internet, Email, webmail, FTP etc. Consider the possibility that an employee needs to forward the confidential data through Email or and uploading those files on to a server which can be accessed by outside world. Before reaching that confidential data to the unauthorized person we need to enforce some policies in order to avoid the violation of the organization health.

To achieve the primary requirement is to scan the whole outbound traffic. We will maintain the DLP (data link prevention) server, which would scan the complete attachment to match the patterns. In case the patter matches, the attachment will be corrupted with the User designed message and an automated response E-mail will be sent out.

The following figure shows the exact scenario.

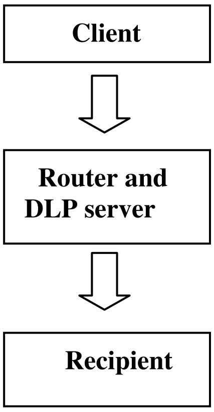

Figure 1: Transmission of an Email between client and DLP server.

For the execution of the above DLP solutions we will require the following modules:

1. Email System
2. DLP Server
3. DLP Algorithms for pattern matching
4. Router Programming
5. File corruption system
6. Integration module to integrate all modules

IV. CONCLUSION AND FUTURE WORK

Data discovery and classification is a prerequisite to a successful deployment of a Data Loss Prevention solution. Understanding the data flows and classifying information enables organizations to protect sensitive information while avoiding relatively benign information like family photos or grocery shopping lists.

DLP has helped the organization in providing a quick, practical framework to:

- Discovering sensitive information
- Protecting this information.
- Evaluating and refining DLP policies and rules once the knowledge is obtained about the nature of the organization's internal and external information flows.

After implementation the results were clearly seen in the organization as how the security gaps are filled on all the three different modules for data loss prevention

REFERENCES

[1]RichardE.Mackey,Available:http://viewer.media.bitpipe.com/1240246133_118/1258558418_168/sComplian ce_sSecurity_Data-Protection_final.pdf [2] Bradley R. Hunter, Available:
http://www.ironport.com/pdf/ironport_dlp_bookle

t.pdf

[3]Webspy,Available:http://www.webspy.com/resources/whitepapers/2008%20WebSpy%20Ltd%20%20Inform ation%20Security%20and%20Data%20Loss%20Prevention.pdf [4] Data loss problems, Available: http://www.webspy.com/reso urces/whitepapers/2009WebSpy Ltd Information Security and Data Loss Prevention.pdf [5] Report, the Office of the U.S. Trade Representative, Available: http://www.ustr.gov/about -us/press-office/reports-and-publications/archive [6] Lubich, H.P; "The changing roel of IT security in an Internet world, a business perspective"; Available: http://www.terena.nl/conference/archieve/tnc2000/proceedings/2A/2a2.html [7]Sithirasenan, E. ; Muthukkumarasamy, V., "Word N-Gram Based Classification for Data Leakage Prevention", Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on 16-18 July 2013, 578 – 585, Melbourne, VIC, 13971211, 10.1109/TrustCom.2013.71. [8]Pham, D.V. "Threat analysis of portable hack tools from USB storage devices and protection solutions," IEEE ISBN: 978-1-4244-8001-2 [9] http://en.wikipedia.org/wiki/Data_loss_prevention_software

[10]http://www.cisco.com/c/en/us/solutions/enterprise-networks/data-loss-prevention/index .html

[11] Bai Xiaoping; Wei Yuanfeng; , "Study on the signal detection and simulation of universal serial bus 2.0 IP core circuit system, "SoutheastCon, 2007. Proceedings. IEEE , vol., no., pp.59-62, 22-25 March 2007 [12] S. Jithesh and U. Naveen, "Improved key management methodology for enhanced media security in IMS networks", New York, US: Institute of Electrical and Electronics Engineers Inc., 2007, pp. 903-907.

[13] AK. Gupta, U. Chandrashekhar, S.V. Sabnis and F.A, "Building secure products and solutions", Bell Labs Technical Journal, Hoboken, US: John Wiley and Sons Inc., 2007.3, pp. 21-38 [14] R.A. Shaikh, S. Rajput, S.M.H. Zaidi and K. Sharif, "Comparative analysis and design philosophy of next generation unified enterprise application security", Piscataway, US: Institute of Electrical and Electronics Engineers Computer Society, 2005, pp. 517-524. [15] Data Leakage Prevention A newsletter for IT Professionals Issue 5. [16] Data Leakage Detection SandipA.Kale1, Prof. S.V.Kulkarni2 Department Of CSE, MIT College of Engg, Aurangabad, Dr.B.A.M.University, Aurangabad (M.S), India1, [17] Journal Of Information, Knowledge And Research In Computer Engineering Issn: 0975 – 6760| Nov 12 To Oct 13 | Volume – 02, Issue – 02| Page 534 Data Leakage Detection Nikhil Chaware 1,Prachi Bapat 2, Rituja Kad 3, Archana Jadhav 4, Prof.S.M.Sangve.