DIFFERENCE OF DLP SYSTEM METHODS

[1]Irgasheva D.Y., [2]Turobova G.O.

[1] Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, professor, [2]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, teacher

https://doi.org/10.5281/zenodo.7857154

Abstract. *Sensitive and confidential data are a requisite for most companies, so protection for this data takes great attention by company's top management, administrators and IT managers. Data leakage causes negative impact on companies. The traditional security approaches, such as firewalls, can't protect data from leakage. Data leakage/loss prevention (DLP) systems are solutions that protect sensitive data from being in non-trusted hands. This paper is an attemptto survey and study DLP systems that will be conducted as wellas a comparison with other security and data protection approaches.*

Keywords: *Data Leakage Prevention (DLP), confidential data, firewall, unauthorized user, cloud computing.*

Data in each company is one of the most important assets; therefore the protection of this data must take the first priority. Although the companies have security measurements and technical parries such as firewalls, still thedata leakage occurs.

The data leakage happens when sensitive data is revealed to unauthorized parties whether it's intentionally or not. The data leaked may cause serious threats to a company. The lossof confidential or sensitive data can severely impact a company's reputation, customers and employee confidence, competitive advantage and in some cases lead to the closure of the company, or political crises such as WikiLeaks leaks.

Data leakage problem must be solved using the Data Leakage/Loss Prevention System (DLP). DLP solutions helpidentifying, monitoring, protecting and reducing the risks of sensitive-data leakage. It is used to detect and prevent unauthorized user from getting sensitive data, and even to protect confidential data that can be accidentally shared. In this paper we will first talk about the existing security approaches used in data protection and in the second section we will talk about data leakage prevention systems, finally we will compare between them.

Response box: inputs information from the previouscomponents and forms an appropriate response. IDS/IPS approach: IDS/IPS techniques can be dividedinto two approaches: A signature-based or pattern matching: IDS/IPS dependson a database of known attacks. These known attacks are loaded into the system as signatures [1].The biggest disadvantage of signature-based systems is that it can trigger only on signatures that have been loaded. Data Leakage Prevention is a solution designed to detectpotential data breach incidents in timely manner and prevent used on enterprise data network to centralize the storage oflogs which was generated by the software running on the network, as well as gathering information, analyzing the identifiable, credit cards, financial and legal information. Exposing this data outside a company's securityperimeter leads to consequences detrimental to the company. DLP solutions help to protect data from going outside company [2].

DLP use Deep Content Inspection (DCI) that considered the evolution of Deep Packet Inspection with the ability to look at what the actual content contains instead of focusing on

individual or multiple packets. Deep Content Inspection allows services to keep track of content across multiple packets so that the signatures they may be searching for can cross packet boundaries and yet they will still be found.

Content Matching: works for structured and unstructured data, using keywords, pattern matching, regular expression, file types, file size, file properties, sender, recipient, and network protocol information to detect data loss incident [3]. Content matching, apply action, use Match-Join algorithms is shown Algorithm1. For each two tuples t1 ∋ T1 and t2 ∋ T2 that match, the match-join table J contains there concatenated tuple. Two tuples match if and only if for every common categorical attribute A: t1.A and t2.A are on the same generalization path in the taxonomy tree for A.

Learning Method (LM):

DLP uses this approach; the basic idea from it is to use machine learning techniques such as Vector space model (SVM) [4], to determine the "confidentiality level" of the scanned email message. Method is the vector space model. Vectors represent documents, and vector features represent terms and their frequency of appearance. The vectors are used as learning sets to build a probabilistic model, on the basis of which decisions are made whether or not documents are confidential. This method is efficient for detecting unstructured content in cases where a deterministic technique is difficult to implement and statistical metrics are the best approach.

Comparison between DLP methods

In the first method content atching is effective in case identifying all keywords and regular expression, but not effective in case changing document format, otherwise the second method needs to enter huge sample of documents until to increase accuracy issues and reduce the rate of false positive and false negative .The use of any of the previous methods based on its existing policies where these methods do not address the encrypted data and does not address the hidden data within the images, audio and video. I suggest enhancements on this algorithm to deal with encrypted and hidden data as future work.

Comparison between DLP (DCI) method and other existing methods (DPI)

Through my study of the existing systems and DLP system, each

system has advantages and disadvantages, the existing systems are providing protection for networks from the outside and provide periodic reports on the security status of the network and systems and send alarm in case attack occurs, existing systems work with ad hoc approach which don't support centralized approach and without having the ability to content aware so without having the ability to prevent data leakage. DLP system works in conjunction with security tools that companies may already have deployed both on endpoint computers (for example, laptops and desktops) and on the network. These may include network and personal firewalls IDS/IPS, antivirus, antispam, encryption and digital rights management tools. The main difference between a DLP system and existing technologies is that DLP systems are content-aware; they are designed to give visibility into where the company's most sensitive data is stored, who has access to it, and where and by whom it is sent outside the company's network. Existing security applications cannot perform this level of monitoring. Additionally, DLP systems must provide comprehensive functionality to prevent this sensitive data from being sent outside the organization through an endpoint computer or available.

Disadvantage of the DLP system: DLP system cannot read the encrypted data and the data hidden within images, audio and video, and in implementation stages; because depending on collected data from all business departments may cause weakness in accuracy issues. Also the Deep Packet Inspection has the ability to inspect the network packet headers. This ability allows firewalls to implement allow/block network access policies since the intent of the packages can be determined where they come from, where they are going, and what ports they are passing through, but inability to comprehend information in that packets. [5].

In recent years, the challenge of dealing with the malicious insider has been acknowledged, and several methods have been proposed for solving this problem. Data leakage is one of the main goals of a malicious insider, and therefore most of the methods proposed for insider threat detection are also applicable for detecting and preventing data leakage.

Initially, Maybury et al [6] presented the results of a collaborative study involving a characterization and analysis of the methods used to counter malicious insiders in the U.S. intelligence community. The study proposes a generic model of malicious insider behavior, distinguishing motives, actions, and associated observables. Several prototype techniques were developed for providing early warning of malicious insider activity, including the use of honey tokens, network traffic profiling, and knowledge-based algorithms for structured analysis and data fusion. Hong et al [7] surveyed proposed methods for detecting insider attack in the research literature, including host-based user profiling based on features such as database and file system accesses, system calls, and OS commands; network based detection; and use of honey pots. Franqueira at all [8] distinguished between internal insiders and external insiders. Mun at all [9] proposed the use of an intrusion detection through the network security levels to documents and monitoring user access to documents.

I suggest working on solving the problems of the existing system, such as work to develop classifiers algorithms or integrate current products that are able to read the encrypted data and the data hidden within images, audio and video, and dealing with content that is not a plain text, such as binary files. Algorithm suggested provides file-unzipping capabilities to interpret a file when the content is obscured several levels down; for example, when an Excel spreadsheet is embedded in a zipped Word file. And policies enhanced in collecting data from business department. using DLP to protect data in cloud computing (Private and Public), this needs more enhancements in the design, policies and procedures to control all the data and application as required. Mobile devices and particularly smart phones are expected to become the main computerized devices that members of the organization use and will be used in the future. Because smart phones are used to access the organization's confidential data such as emails and documents, it is expected that they will be used to leak information accidentally and intentionally. There have been several attempts to extend the organization's security perimeter into smart phones. There is a need for future research to find new approaches to give members of the organizations the access to confidential information through their smart phones, and on the other hand to prevent this information from leaking through the smart phone intentionally or accidentally.

Conclusions

This paper describes the importance of the information regards to companies and the seriousness of corporate data leakage. This paper studied the current systems used to protect data and the DLP system in terms of their components and methods used within them, and the

differences between them. Showing the difference between existing systems and DLP system, as well as the importance of covering the shortage existed in current DLP systems such as of developing policies, integration with other systems such as encryption, audio, video and images, so it is recommended to provide a scientific solutions to the problem of data leakage and mitigation.

REFERENCES

1. Dhiren & Maulik, Hardik Joshi, Bhadresh K,Patel, "Towards Application Classification with vulnerability signature for IDS/IPS",SecurIT'12,Augest,17-19,2019,Kollam,Kerala,India.
2. Preeti Raman, Hilmi Güneş Kayacık, and Anil Somayaji
3. System for detecting insider attackers. The proposed system is based on assigning grades and privilege levels to users and August 6, 2017.
4. Salem, B.M., Heshkop, S., and Stolfo, S.J. 2020. A survey of insider attack detection eesearch. Insider Attack and Cyber Security- Beyond the Hacker, Springer.
5. wedgenetworks http://www.wedgenetworks.com/resources/ technology/deep-content-inspection-with-wedgeos.html.
6. Erez Shmuelia,c, Tamir Tassab,c, Raz Wassersteina, Bracha Shapiraa, Lior Rokach, Limiting Disclosure of Sensitive Data in Sequential Releases of Databases,2017s
7. Mun, H., Han, K., Yeun, C.Y., and Kim, K. 2018. Yet another intrusion detection system against Insider Attacks. Proceedings, Symposium on Cryptography and Information Security .
8. Mohammad A. Faysel , and Syed S. Haque Towards Cyber Defense: Research in Intrusion Detection and Intrusion Prevention Systems ,July 2018
9. SearchSecurity, http://searchsecurity.techtarget.com/definition/ security-information-and-event-management-SIEM.