International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 01 | Jan -2017 www.irjet.net p-ISSN: 2395-0072

Prompt Detection of Transformed Data Brench

N.KEERTHANA1,P.SIVAKAMASUNDARI[2]

[1]M.E(Computer Science),Adhiparasakthi Engineering College,Melmaruvathur,Tamil nadu, India
[2]Professor(Dept of Computer science),Adhiparasakthi Engineering College,Melmaruvathur, Tamil nadu,India

---

Abstract-*Statistics from security firms, research institutions and government organizations show that the number of data leaks instances has grown rapidly in recent years. Among various data-leaks cases, human mistakes are one of the main causes of data loss. According to a report from Risk Based Security(RBS) the number of leaked sensitive data records has increased dramatically during the last few years., from 412 million in 2012 to 822 million in 2013.Deliberately planned attacks, inadvertent leaks(forwarding confidential emails to unclassified email accounts)and human mistakes(assigning the wrong privilege)lead most of the data-leaks incidents. Detecting and preventing data leaks requires a set of complementary solutions, which may include data leak detection data confinement stealthy malware detection and policy enforcement. Network data-leak detection(DLD) typically perform deep packet inspection(DPI) and searches for any occurrences of sensitive data patterns.*

Index Terms-Risk based security(RBS),Data leak detection(DLD),Deep packet inspection(DPI).

1. INTRODUCTION

Data leak detection is a software that are designed to detect potential data breaches, data ex filtration transmissions and prevent them by monitoring, detecting and blocking sensitive data while in-use (endpoint actions), in-motion (network traffic), and at-rest (data storage). In data leakage incidents sensitive data is disclosed to unauthorized personnel either by malicious intent or inadvertent mistake[1]. Such sensitive data can come in the form of private or company information, intellectual properties (ip), financial or patient information, credit-card data, and other information depending on the business and the industry.

The terms data prompt and data leak are closely related and are often used interchangeably, though they are somewhat different]data loss incidents turn into data leak incidents in cases where media containing sensitive information is lost and subsequently acquired by unauthorized party[2]. Data leak detection server detects and prevent unauthorized attempts to copy or send sensitive data, intentionally or unintentionally, without authorization, mainly by personnel who are authorized to access the sensitive information. In order to classify certain information as sensitive, these solutions use mechanisms, such as exact data matching, structured data fingerprinting, statistical methods, rule and regular expression matching.

2. EXISTING SYSTEM

DLD solutions include a number of techniques for identifying confidential or sensitive information. Sometimes confused with discovery, data detection is a process by which organizations use a DLD technology to determine what to look for (in motion, at rest, or in use).Data is classified as structured or unstructured. Structured data resides in fixed fields within a file such as a spreadsheet, while unstructured data refers to free-form text as in text documents or PDF files.. The leak of sensitive data on computer system is a serious threat in organizational security. Sometimes trusted 3rd parties may act as a point of data leakage. Data leakage mainly happens due to human errors. Detecting and preventing data leaks requires a set of complementary solutions, which may include data-leak detection data confinement stealthy malware detection and policy enforcement.

2.1 Inadvertent data leak:

The sensitive data is accidentally leaked in the outbound traffic by a legitimate user. Inadvertent data leak may be due to human errors such as forgetting to use encryption, carelessly forwarding an internal email and attachments to outsiders.

2.2 Malicious data leak:

A rogue insider or a piece of stealthy software may steal sensitive personal or organizational data from a host.

2.3 Data traffic and time consumption:

Data traffic on proxy server and mail server impacts the performance of data leakage detection technique and time delay of common legitimate users.

2.4 Static filtering of authorized users: Static approaches of authorized user filtering technique affect the efficient of data leakage detection.

3. SYSTEM DESIGN

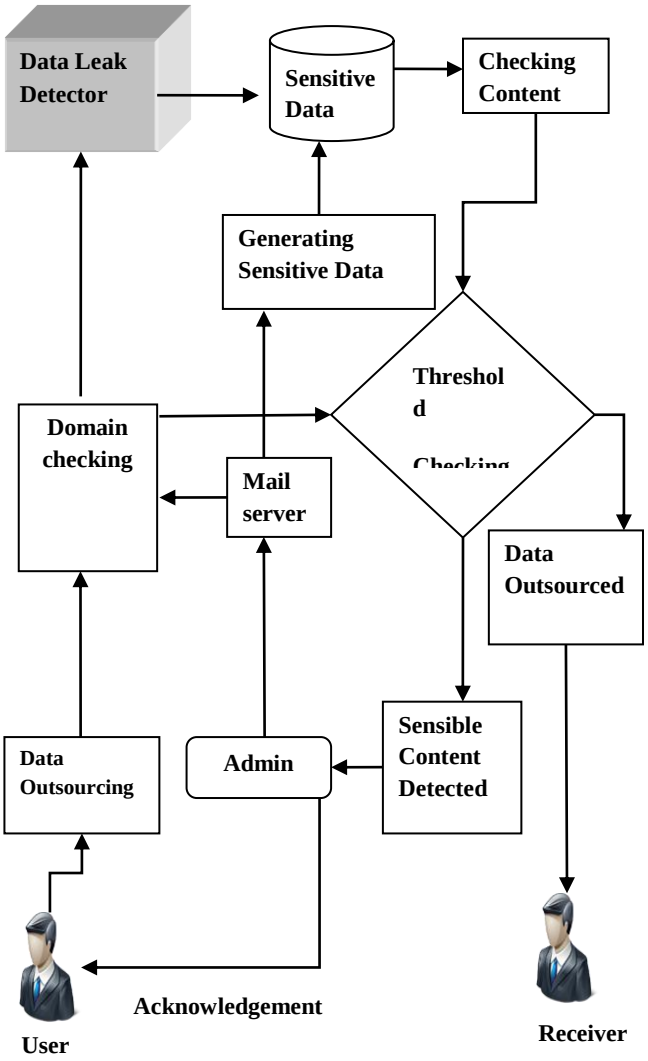

Fig 1-System Architecture

4. IMPLEMENTATION

4.1 Content outsourcing without DLD:

User registers in mail server with their name, authorized job position and their authorized e-mail domain. And the users can transfer their file using without any restriction of sensitive content checking. There is no content checking and domain filtering on their transformed sensitive data. Sensitive content is outsourcing from one organization to another organization performed by user[3]. The content can be of any file (text, document). Outsourcing will not reach DLD and directly reach its destination or organization. Here outsourcing mechanism of transferred data is offending over the protocol.

4.2 Build data leakage detection framework:

Mail server data owner generates a sensitive data and stored in the cloud and create the directory for Lucene search framework and other data leakage detectors. Data owner's cloud contains much sensitive information about their authorized customer's details, information technology source, and database and server details. This sensitive information is maintained by Data Leak Detector. Using this DLD referenced directory perform data leak detection mechanism. The DLD consist of Lucene search engine framework, Levenshtein distance algorithm and own shuffled checking algorithm. The DLD directly configured with cloud and can refer every data transformation outsourcing from authorized user transformation.

4.3 Content outsourcing with DLD checker:

DLD is the one will check all the outsourcing content before it transmit to the other organization. All the outsourced contents are check with sensitive data. All the sensitive data are maintaining in index file. Using this index file DLD identify the sensitive data concurrently with domain filtering and threshold assigning based on their email domain. DLD will check every line of the sending data with the sensitive file. DLD will not allow any sensitive data will leak to any of the other organization.

In proxy mail server the every occurrence of transformed contents are filter by users email domain. All users details are retrieved from the cloud using their email. Then threshold assigned for the users based on their authorized job position and the transferred content has been tested by Lucene framework search engine, Levenshtein distance checking and shuffling algorithm.

4.4 Sensitive Data Detection:

Once the DLD framework checks the outsourced content, if any data leak is identified means DLD will detect

the sensitive data. Here DLD will check not only the sensitive data and also it will check some access condition. Every data owner maintain common access condition every file. For example, all the contents are encrypted before they outsourced. If DLD identified any sensitive information outsourcing means they will detect the sensible content in between of the file outsourcing.

For the purpose of false alert, we maintain threshold of every domain and users position. If the sensible content percentage of transferred file exceeds the threshold percentage which trigger alert mail to Admin of the proxy mail server. Alert mail consists of entire details about the users even what are the sensible contents are pings from the transferred content by the DLD framework.

3. CONCLUSION:

Data leak detection framework is developed to avoid exposure of sensitive data and also provide privacy preserving to sensitive data. For future work, the static implementation of web service used to maintain the users and sensitive content instead of database.

REFERENCES

[1] X. Shu, J. Zhang, D. Yao, and W.-C. Feng, "Rapid and parallel content screening for detecting transformed data exposure," in Proc. 3rd Int. Workshop Secure. Privacy Big Data (Big Security), Apr./May 2015, pp. 191–196.

[2] L.De Carli, R.Sommer, and S. Jha, "Beyond pattern matching: A concurrency model for Stateful deep packet inspection," in Proc. ACM SIGSAC Conf. Computer. Commun. Secure. 2014, pp. 1378–1390.

[3] X. Shu, D. Yao, and E. Bertino, "Privacy-preserving detection of sensitive data exposure," IEEE Trans. Inf. Forensics Security, vol. 10,no. 5, pp. 1092–1103, May 2015

[4] (Feb. 2015). Data Breach QuickView: 2014 Data Breach Trends. YEDataBreachQuickView.pdf, accessed Feb. 2015.