Palestine Polytechnic University
Deanship of Graduate Studies and Scientific Research

Data Leakage/Loss Prevention (DLP) Systems Analysis and Solutions

Submitted by

Yousef Saleh

In partial fulfillment of requirements for the degree of Master in Informatics

August, 2015

The undersigned hereby certify that they have read and recommend to the Deanship of Graduate Studies and Scientific Research at Palestine Polytechnic University the acceptance of a thesis entitled *"Data Leakage/Loss Prevention (DLP) Systems Analysis and Solutions"* submitted by Yousef Saleh in partial fulfillment of the requirements for the degree of Master in Informatics.

Graduate Advisory Committee

Dr. Radwan Tahboub (Supervisor)

Signature:_____ Date:_____

Dr. Liana Tamimi (Internal Examiner)

Palestine Polytechnic University, Palestine

Signature:_____ Date:_____

Dr. Ahmad Asadeh (External Examiner)

Birzeit University, Palestine

Signature:_____ Date:_____

Thesis approved by

Dr. Murad Abusubaih

Dean of Graduate Studies and Scientific Research

Palestine Polytechnic University

Signature:_____ Date:_____

Declaration

I declare that the master thesis entitled *"Data Leakage/Loss Prevention (DLP) Systems Analysis and Solutions"* is my own original work, and hereby certify that unless stated, all work contained within this thesis is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgement is made in the text.

Yousef Saleh

Signature:_____ Date:_____

Statement of Permission to Use

In presenting this thesis in partial fulfillment of the requirements for the master degree in Informatics at Palestine Polytechnic University, I agree that the library shall make it available to borrowers under rules of the library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made.

Permission for extensive quotation from, reproduction, or publication of this thesis may be granted by my main supervisor, or in his absence, by the Dean of Graduate Studies and Scientific Research when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Yousef Saleh

Signature:_____ Date:_____

Dedication

I dedicate this dissertation to many friends who have supported me throughout the process. I will always appreciate all they have done, especially Anas for the many hours of proofreading.

Acknowledgement

We would like to thank Almighty for giving me an opportunity to study in Palestine Polytechnic University (PPU) and to work on this Master Thesis about Data leakage Prevention.

A very big 'thank you' to my supervisor Dr.Radwan Tahboub for guidance, support and the help given to me to complete this required work.

To all friends Mahmoud, Ahmad, Anas for very useful comments and suggestions to fine tune thesis. To my family who support me throughout my course duration.

I would like to thank companies that help me in doing our test.

I would like to thank my instructors, Doctors and all people in Palestine Polytechnic University (PPU) who taught me and helped me to complete my study program.

الملخص

نظم منع تسرب أو سرقة البيانات هي عبار عن أنظمة حماية للبيانات الحساسة والهامة بحيث توفر هذة الأنظمة الحماية لتلك البيانات من اولئك الذي لديهم صلاحيات الاطلاع على هذه البيانات من سرقة أو تسريب لتلك البيانات، لما يمثل تسريب هذه البيانات الحساسة من اضرار جسيمة للشركات والمؤسسات على جميع المستويات. هذه الأنظمة هي عبارة عن أنظمة متطورة وحديثة وتفوق الأنظمة الموجوده حاليا في حماية البيانات مثل أنظمة الجدر النارية ومضادات الفيروسات وغيرها من الأنظمة الموجودة بحيث تم عمل مقارنة بينهما. وعلى الرغم من أن أنظمة منع تسرب البيانات أنظمة متطورة وحديثة إلا أنها تعاني من بعض المشاكل والضعف في منع تسريب أو سرقة هذه البيانات. حيث تم اختبار خمس من أنظمة منع تسريب البيانات الاكثر شيوعا حيث كشفت النتائج عن وجود ضعف ومشاكل في هذه الأنظمة كما تم اقتراح لحل هذه المشاكل بتقديم نموذج وقائي لمعالجة مشكلة البيانات المخفية في الصور بحيث يضمن تدمير لهذه البيانات المخفية ومنع تسريب البيانات من خلالها وذلك من خلال تقديم اقتراحين لمعالجة هذه المشكلة وهما استبدال البت الاخيره من البكسل الى صفر والطريقة الثانية استبدال البت الاخيرة من البكسل الى واحد وبنائهما واختيار أحد هذين الخيارين لبنائه في نظم منع تسريب البيانات.

Abstract

Sensitive and confidential data is a requisite for most of the companies, so protection of this data takes great attention by company 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 the solutions that protect sensitive data from being in non-trusted hands. This thesis is an attempt to study DLP systems that will be conducted as well as a comparison with traditional security approach and discover the weakness of DLP approach.

In this thesis we are testing and evaluating five different DLP systems implemented in the IT market and showing the weaknesses in these systems and propose a model to enhance these weaknesses.

A simple approach has been developed to solve some of these weaknesses which enhances the effectiveness of DLP systems. Furthermore to these evaluations and tests, a complementing proof of concept solution may be developed so that it could be built-in with DLP solutions.

The experimental results showed that DLP systems suffer from weaknesses such as, all DLP not work in safe mode, cannot discover encryption and compressed file also cannot discover data hidden when using steganography technique and still need improvement and enhancement.

The idea of our model is using a precaution method to destroy hidden

data before being sent by email, USB flash memory or uploaded outside the network.

The importance of our proposed model is to protect data and ensure that no hidden data might be leaked. The results show after executing proposed model on samples of hidden data in the images, the hidden data has been destroyed minimal effect on resolution and quality on these images.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BIOS	Basic Input/Output System
BMH	Boyer-Moore-Horspool
DAR	Data-At-Rest
DIM	Data-In-Motion
DIU	Data-In-Use
DLP	Data Leakage/Loss Prevention
DSM	Data Security Management
DPI	Deep Packet Inspection
FGDLP	FortiGate Data Leakage Prevention
FLSBO	Flipping the Least Significant Bit (LSB) to One
FLSBZ	Flipping the Least Significant Bit (LSB) to Zero
FTP	File Transfer Protocol
HIDS/HIPS	Host-based Intrusion Detection/Prevention System
HIPAA	Health Insurance Portability and Accountability Act
HTTP	Hypertext Transfer Protocol
ID	Identity Number
IDS	Intrusion Detection System
ILDP	Information Leak Detection and Prevention
IPS	Intrusion Prevention Systems
LSB	Least Significant Bit
MCDLP	McAfee Data Leakage Prevention
MSE	Mean Square Error
MyDLP	MyDLP Data Leakage Prevention
NIDS/NIPS	Network-based Intrusion Detection/Prevention Systems
NSA	National Security Agency
PCI-DSS	Payment Card Industry Data Security Standard
PSNR	Peak Signal to Noise Ratio
RSA	Rivest-Shamir-Adleman
SDLP	Symantec Data Leakage Prevention
SIEM	Security Information Event Management
SSN	Social Security Number
TDLP	Trustwave Data Leakage Prevention
VSM	Vector Space Model
WSDLP	Websense Data Leakage Prevention

Chapter 1

Introduction

Data in every company is one of the most important assets; therefore the protection of this data must take the first priority. Although the companies have security measures and technical parries such as firewalls that mean the data leakage still occurs [9]. The data leakage happens when sensitive data such as credit card information and social security number an revealed to unauthorized parties whether it's deliberately or not [9]. Data that's leaked could cause severe threats to a company [9]. The loss of confidential or sensitive data can severely impact an organization's reputation, client and employees confidence, competing advantage (some case lead to organization closure), or even political crises such as WikiLeaks' leaks [10].

To avoid this type of loss/leakage, systems focused on identifying and protecting sensitive data were introduced in 2006 and became known as Data Leakage or Loss Prevention (DLP) [9]. DLP solutions help identify, monitor, protect and lower the risks of confidential-data leakage [11]. It is used to detect and prevent unauthorized user from getting sensitive data out, and even to protect confidential data that can be accidentally shared [12]. There are currently many vendors provide various DLP solutions for various operating systems such as Symantec [13], Websence [14], OpenDLP [15], MacAfee [16], Trustwave [17], MyDLP [18], RSA [19] and FortiNet [20].

Criminal attacks are particularly likely to happen from the inside; a recent study

estimated that 90 % of criminal computer crimes were committed by employees of the company [21]. An example of the inside threats is Edward Snowden - an American computer professional - who leaked sensitive-classified information from the National Security Agency (NSA) to the mainstream media [22]. Because he had a high level of rights as an administrator/manager, nothing prevented him from obtain data that he shouldn't obtain, without warning anyone. This example displays how a threat exists and the results can damage the targeted organization.

Data Loss Prevention is necessary for organizations, to keep the sensitive data from being leaked outside the organization from employees who have the access to this data. Any organization has the right to keep this data because it represents its characteristic and any loss for these data can destroy the company. So it is mandatory for the IT vendors to keep improving DLP Solutions.

Also organization should discuss the reasons why their employees will leak the data. Employee's sometimes behave aggressively against their company because of many reasons such as work pressure, efforts unfair rewards, political issues. [2].

There are many products that provide solutions to data leakage, such as, websence, openDLP, McAfee, trustwave, symantec, myDLP, trendMicro, RSA and Fortinet. Most of these does not provide sufficient protection against intentional attacks from insider. Protection from intentional attacks can be partially achieved using regular expressions, key phrase matching and file type (e.g. .doc, .zip, .pdf). The defined key phrase matching is applied to all data which needs to be read or written from removable devices and met the specification in the policy rule. So improving the methods for detecting sensitive data can increase the effectiveness of DLP solutions which leads to enhance DLP security.

1.1 Problem Definition and Contributions

One of the primary goals of an attack with an organization network is to get sensitive data, including the credit card numbers of its customers, plans of the organization, source

codes.

Today's networks are very large that manual examination of data leakage is not practical and a costly option. To be able to handle the security problems, organizations invest in intrusion detection and prevention systems (IDS/IPS), firewalls and virus scanners. Nevertheless, only 57 % of all the data breaches are the result of external hacking attacks, based on to Open Security Foundation DataLossDBm survey, 43% of most recorded incidents are from insiders [23].

Information leakage/loss problem became a current problem in several companies. This even became a major problem for business owner. This kind of issue appears in the context where the confidential data and organization files are safe by a protection model. In business, the main loss/leakage of data is primarily taking place through insider attack. Although the protection systems such as Firewall, IDS, IPS and other systems are very strong, but they can help only against the outside attack on data (Not always). Because these systems don't support majorly towards insider attacks, which causes an information loss/leakage problem.

The goal of this thesis is to obtain knowledge about the existing DLP systems and understand how they work and what weaknesses are existing in these systems. Also, how can we improve these weaknesses. This is done by testing different five DLP systems and examining what leakages are in those systems.

According to the outcomes of our tests in chapter 3, we propose a model to prevent data from being leaked. We expect to contribute in the following:-

1. Survey about DLP systems.
2. Implementing some DLP systems test based.
3. Test and compare between these DLP systems.
4. Provide evaluate suggestions to decrease the weaknesses and improve the selected systems.

1.2 Thesis Outline

The remaining chapters of this thesis are organized as following:

Chapter 2: shows the technologies used in Data Protection. Chapter 3: covers the investigation and evaluation of prevention systems. Chapter 4: presents the results and discussion. Chapter 5: presents the suggested model about problems. Chapter 6: concludes our work and presents some future lines of research about improving this project.

Chapter 2

Data Protection Technologies

This chapter presentes the existing technology used in data protection such as IDS/IPS, antimalware, antivirus and firewall. Also, it presents the definition of data leakage prevention systems, and show the differences between existing technology and DLP systems, after that about the DLP products used in market such as Websence, Opendlp, Symantec, Trustwave, MacAfee, MyDLP, RSA and Fortinet and some problems faced by these systems such as steganography. There are several technologies used to protect data, many focus on protecting data from the outside while the DLP system concentrates on data protection from inside [24].

2.1 Existing Technologies Used in Data Protection

2.1.1 IDS / IPS

Intrusion Detection System (IDS) is usually a device or software application that monitors networks or system activities for malicious. It primarily determines malicious, log information, try to block or stop and report activities [25].

Intrusion Prevention Systems (IPS) are expansion of intrusion detection systems as they both monitor system activities for malicious, network traffic, but IPS can stop or

block intrusions discovered. This could also execute actions for example indicating a warning, or leaving the malicious packets [25].

IDS/IPS Components: IDS/IPS could be put into two classes: Network-based Intrusion Detection/Prevention Systems (NIDS/NIPS) and Host-based Intrusion Detection/Prevention Systems (HIDS/HIPS).

NIDS/NIPS inspect packets on the network and looks at the data attempting to recognize an attack. (HIDS/HIPS) monitor traffic on one specific host. HIDS/HIPS excels at detecting/preventing unauthorized access and activity. HIDS/HIPS looks at the state of a system and verify that the behaviour of the packet appears as expected [1].

Both NIDS/NIPS and HIDS/HIPS can often scan for attack, track hackers' movements, and alert an administrator about ongoing attacks. Most IDS/IPS consists of more than one application or hardware device. IDS/IPS consist from the following components [25]:

- Network Sensors: Discover and send data to the systems.
- Central Monitoring System: Processes and examines data sent from sensors.
- Report Analysis: Provides information about how to counteract a particular event.
- Database: Keep information such as IP about the attacker.
- Response Box: Inputs information from the previous elements and forms a suitable reaction.

IDS/IPS Approach: could be divided into two techniques:

Signature-based or Pattern Matching: IDS/IPS depends upon a database of identified attacks.

These identified attacks are loaded into the system as signatures. The largest problem with signature-based systems is it can trigger only on signatures that have been loaded [1, 24]. The latest attack may go undetected. Snort [24] is a good example of signature-based IDS/IPS as shown in Figure 2.1.

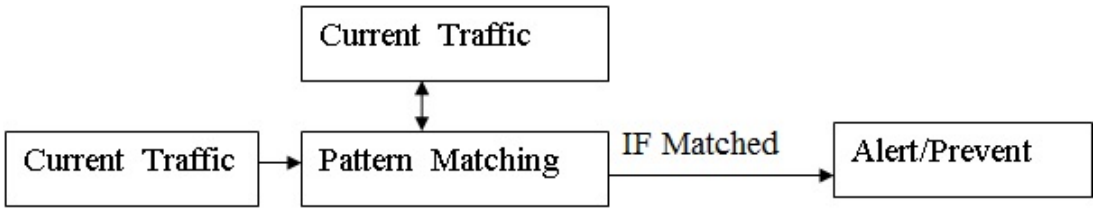

Figure 2.1: Signature-Based IDS/IPS [1] [2]

Anomaly-Detection Systems: require administrator to implement profile of authorizes activates or place the IDS/IPS into learning mode so it can learn what constitutes normal activity. As shown in Figure 2.2.

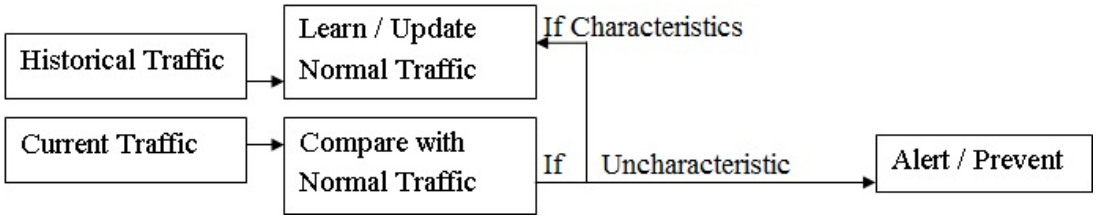

Figure 2.2: Anomaly -Based IDS/IPS [2] [3]

Anomaly detection can be a helpful one at monitoring behaviours that's greatly different from normal activities [3].

2.1.2 Anti-Malware

Malware is software designed to damage computer operations, collect sensitive data, and gain unauthorized access to computer system.
Malware has two categories [23]:

- Malware which modifies the recourse such as memory, BIOS Code, PCI devices expansions EEPROMS [26].
- Malware that does not modify any of those resources, but only the resources which are dynamic by nature, such as by modifying some function pointers in some kernel data structures, so that the attackers code gets executed instead of the original system or application.

Types of Malware:

- Virus: Replicates by attaching its program instructions to an ordinary host software or file, so the virus instructions are run when the host program is run. For example micro virus, file virus, boot sector virus, and e-mail virus [26].

- Network Worm: Self-propagating program that spreads on the network, usually Internet. Not like viruses, may not depend on other programs or victim actions (such as opening an infected email attachment or clicking on a web link for a malware Web site) for replication, dissemination, or execution. Such as swarm worm [27].

- Trojan Horse (or, simply, Trojan): A destructive program that masquerades as a harmless software. Stealth ware such as key loggers, spyware, root kits, trapdoors, and certain adware represents a subset of Trojans that is intentionally designed to be hard-to detect or undetectable. Trojan horse software installs itself on the victim's computer when the victim opens an email attachment or computer file containing the Trojan, or clicks on a web link that directs the victim's browser to a web site from which the Trojan is automatically downloaded, Such as Backdoor Trojan [28].

- Spyware (non-Trojan): Non-Trojan stealth-ware that has the same objectives and performs the same types of actions as spyware Trojans. A number of bots have spyware capabilities, and are referred to as spy-bots such as Adware [23].

- Embedded Malicious Code: Any type of malicious logic embedded in a valid executable program by its developer, integrator, distributor, or installer. Most frequently a logic Trojan, time bomb or bomb [27].

Anti-Malware is any software that provides a protection for computers and systems from malware, viruses, spyware and other harmful programs. 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 will be removed [28].

2.1.3 Firewalls

Firewalls are devices or software that permit or deny network transmissions based on a set of rules (access rule) and is used to protect networks from unauthorized access while permitting legal communication to pass [27].

Firewall is 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 packet and determining whether it should be allowed through or not [27].

2.1.4 Vender Patches

Updating the systems and applications, and installation of new hotfixes that recommended by vendors [24].

2.1.5 SIEM

Security Information Event Management (SIEM) is a tool used on enterprise data network to centralize the storage of logs which was generated by the software running on the network, as well as gathering information, analyzing the information and also presenting the information from network and security devices [29].

2.2 Deep Packet Inspection (DPI) Method

All the previous techniques are using Deep Packet Inspection (DPI) method. DPI looks at the packet, finds anomalies in the traffic and alerts the administrator or prevents the traffic. DPI classifies passed traffic depending on policies, these policies include information about layer 3 and layer 4 content of the packet as well as the information that describe the content of the packet payload [30].

The following steps summarize how DPI Architecture works [31]:

- Pattern Definition Language Interpreter uses signatures that can be written to detect and prevent against known and unknown protocols.
- The Deep Packet Inspection reassembles TCP packets arriving out-of-order.
- Deep Packet Inspection engine preprocessing involves normalization of the packet payload. For example, an HTTP request may be URL encoded and thus the request is URL decoded in order to perform correct pattern matching on the payload.

Deep Packet Inspection engine postprocessors perform actions which may either simply pass the packet without change, or can fall a packet or could possibly reset a TCP connection [4]. As shown in Figure 2.3.

All the above technologies use DPI method.

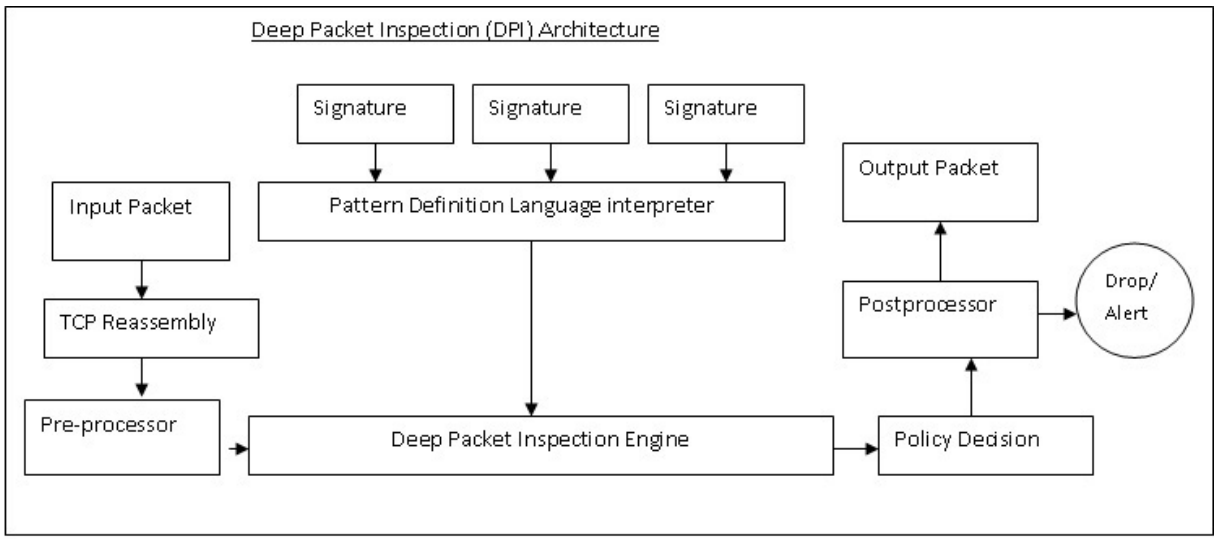

Figure 2.3: Deep Packet Inspection Architecture. [4]

2.3 Data Loss

The data loss problem is uncovering sensitive information for a customer or outsider of organization such as sensitive documents and files being delivered to arrival. This happens in several ways: unintentional or deliberate, or even with insiders knowing sensitive data about clients such private information, intellectual property, or any other sensitive information [32]. Employees could deliberately or accidentally send information by using e-mail. Or they can print and then get copies of sensitive documents. They can also download data to a laptop that's subsequently lost or stolen, or copy it onto an external storage device such as a USB drive. When sensitive data is lost, businesses usually should pay for damage mitigation and victim remediation. They may also have regulatory fines and the requirement to conduct additional audits. If a data leak receives the public attention, it could damage an organization's reputation, causing current and potential customers to lose trust [9]. There are also several accesses in which sensitive data can leave a company through the internet: this includes E-mail, Webmail, HTTP (message boards, blogs and other websites), Peer-to-peer sites and sessions and FTP. the figure 2.4 show the tree of technologies used in data protection.

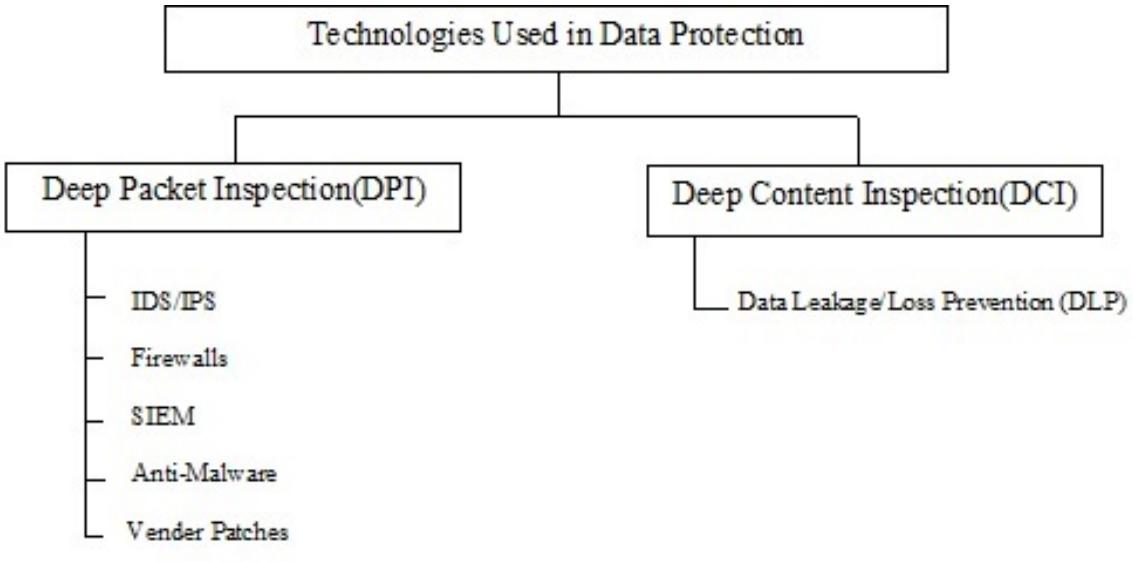

Figure 2.4: Technologies Used in Data Protection

2.4 Sensitive Data

Sensitive data is any data if leaked may cause damage to person or business. Sensitive data may have private information or information about a business. Sensitive data examples my include Social Security Number (SSN), credit card number, banking information, private information for patients, financial data for companies, private information for students, research data inside college and sources code for software [9].

In 2014, US State of Cybercrime Survey discovered that almost one-third of participants said insider crimes are more costly or damaging than incidents perpetrated by outsiders. Because internal threat actors hold the advantage since they are more likely to know where valuable data is stored and what processes and technologies are in place to protect this information and prevent theft [33].

Ponemon Institute (2013) executes Security Survey and clarifies very well the confidential data leakage problem with the different statistical numbers which may be very helpful to us to be aware of the size of this trouble properly [34].

2.5 Data Leakage Prevention (DLP)

Data Leakage Prevention is a solution designed to detect potential data breach incidents in timely manner and prevent them by monitoring data while in-use (endpoint actions) or in-motion (network traffic) or at-rest (data storage). DLP is considered as a paradigm shift in information security. It addresses risks; these risks are focused around handling certain types of data with information security risk, such as personally identifiable, credit cards, financial and legal information. Exposing this data outside a company security perimeter leads to consequences detrimental to the company. DLP solutions help to protect data from going outside company [9] [35].

Due to the creation and expansion of data, it has became increasingly critical for businesses to put measures in place to secure this data. DLP solutions help to protect this

data while allowing the organization to function with minimal obstruction to operations. Sometimes Data Leak Prevention (DLP) known as Information Leak Prevention (ILP) or Information Leak Detection and Prevention (ILDP) [11].

2.5.1 The Importance of the DLP Systems.

Data security breaches pose serious to company. The loss/leakage of sensitive or confidential data can severely impact an organizations reputation, customers and employee confidence, and competitive advantage. Failure to comply with regulatory standards can cost companies huge monetary fines in addition to the loss of current and future customers and employees. DLP is essential for the following causes [36]:

- Help company focus on securing business process and the company assets.
- Empower employees to make correct decision on sensitive data handling and help employees understand when their actions may expose sensitive data to risks.
- Preserve clients and employee commitment.
- Reduces likelihood and costs of data breaches.

When confidential data exist in particular business leakages deliberately or unintentionally to the incorrect people or companies, who don't have an authorized permission to see them, considerable damage might be created to somebody or a company. In addition for legal expenses, companies have to deal with the high concrete costs of recovery; for example losing money, time, resources, business, and reputation [37].

2.5.2 DLP Stakeholders

Effective DLP is a corporate-wide initiative to prevent the loss of the organizations most sensitive data. Deploying DLP involves people, processes, and technology at all levels within company. While it is the responsibility of all employees to protect the company

sensitive data, it is important that stakeholders work together to define the success metrics and fully understand the impact on the business with respect to organization culture, functional changes and employee behaviour [38]. The stakeholders often include representative from the business data owners, compliance and privacy, information security, legal and auditing.

2.6 How DLP Work

Each DLP system today varies in an way that used in data leaks problem, however, many important areas stay the same. Many DLP systems have predetermined key word content for example “Password”, “Policy”, ”ID person”. They also include dictionaries with bank card number patterns to the various manufactures(template) for example Visa, Master Card, American express [9].

The DLP system scans the documents being transmitted and in case there are numbers matching any of these templates or key words based on the policy set for this type of event, the DLP system will do the defined action for example block, alert, and reroute [11].

Also DLP offers the ability to do some specific actions to some file location, sender, receiver and types. For example: preventing the file extension. If the latter is performed, it might always be able when the file extension is modified. In case a file is modified from password.txt to password.jpg, the system might still identify this file as a text[48].

2.7 Data Leakage/Loss Prevention Structure

The goal of DLP systems is to protect data in different phases the general DLP structure is shown in Figure 2.5 [5].

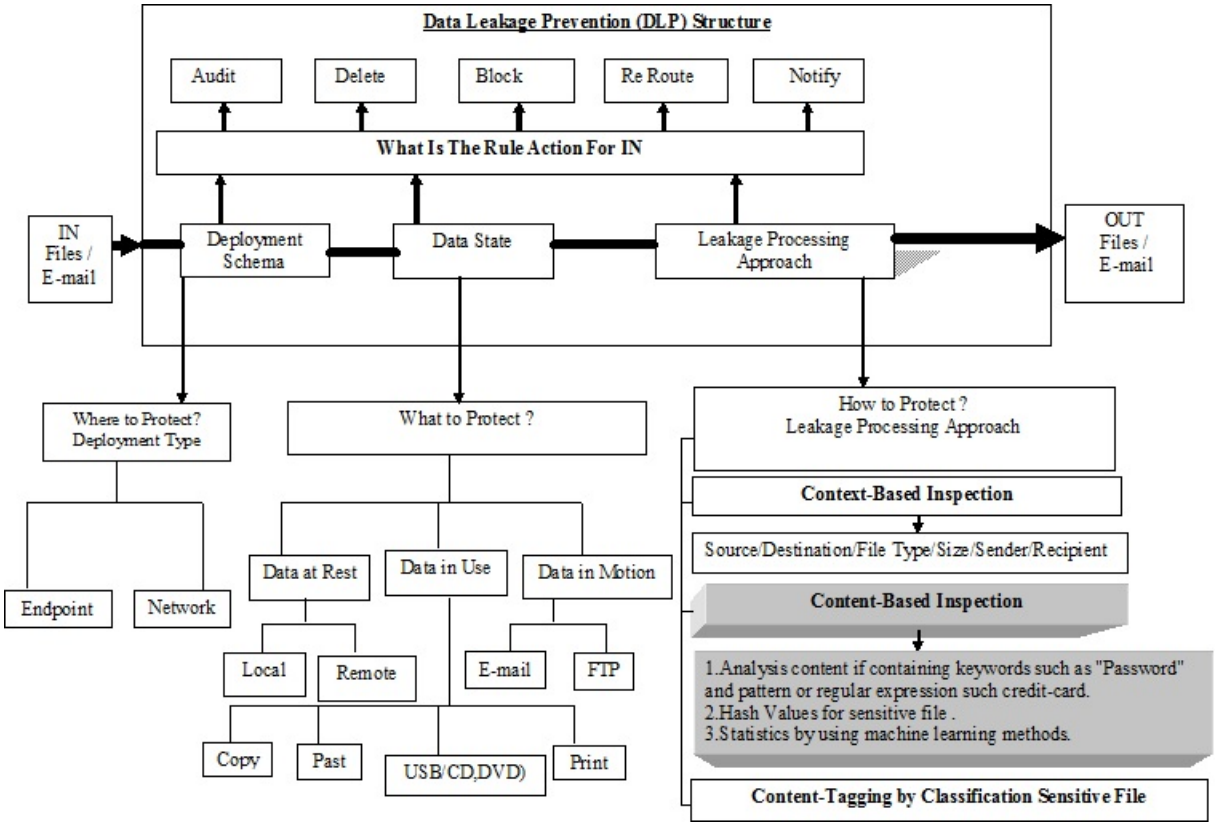

Figure 2.5: DLP Structure [5] [6]

2.7.1 Phases of Data in DLP System

DLP systems separate between three levels of data during their lifecycle: data-at-rest (DAR), data-in-motion (DIM) and data-in-use (DIU) [9].

Data-At-Rest: are described as all data in computer storage. To maintain data-at-rest from becoming accessed, stolen, or modified by unauthorized people, safety measures - for example data encryption and access control - are generally used [38]. A prerequisite of these security measures is content discovery, which serves to discover where all the data are stored. The best way to achieve this is applying the content discovery features of DLP products [2]. Such as, a policy may need that customer credit card numbers be saved only on approved servers [7]. If data is detected with an unauthorized server, it may be encrypted or removed, or an alert could be sent to the data owner as shown in figure 2.6.

Data-In-Use: are any data that a user is communicating. Endpoint-related systems are used to protect data-in-use and to monitor data as the user interacts with them [39].

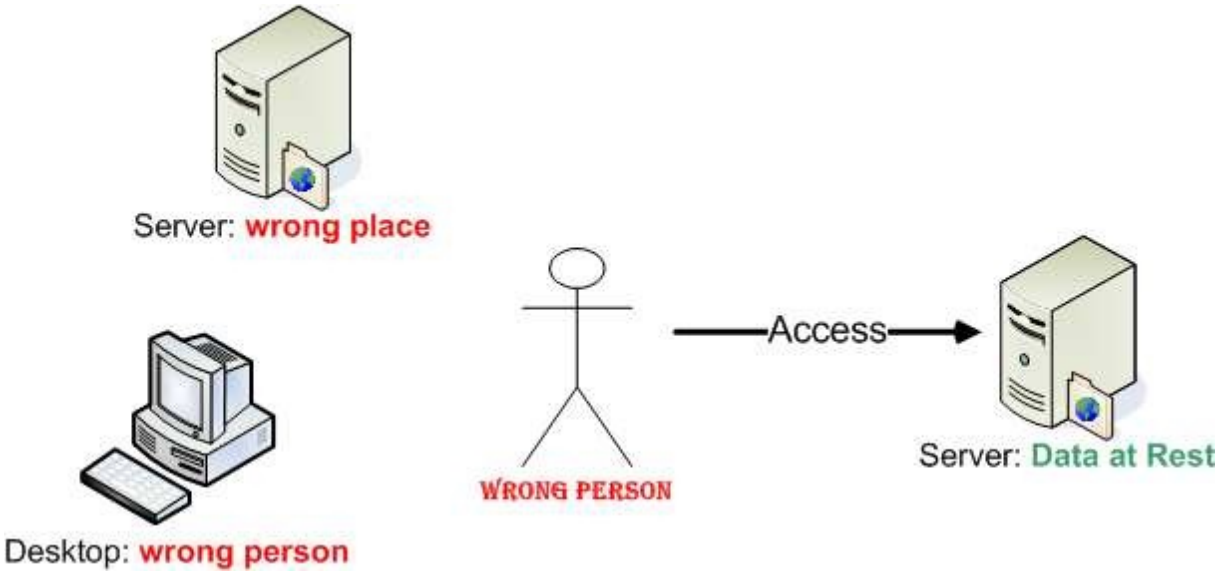

Figure 2.6: Data-at-Rest [7]

Usually, an agent is used to monitor the data while they're used or moved from an endpoint device or client through different output channels to peripheral devices. The actual concept is when an attempt is made to send sensitive data, the possible leakage might be quickly detected and resolved (e.g., blocked) before the data can be sent. Data-in-use tools may monitor the following activities [39]:

- Copy, paste and screen-capture operations involving confidential data.
- Transfer of sensitive data from one device to another such as USB drives, CD/DVDs or smart phones as shown in figure 2.6.
- Printing confidential content.

Data-In-Motion: are data being sent via a network. These data could be sent inside the internal network of the organization or may go over into an external network [40]. DLP solutions are used to discover and check data that are being sent throughout communication channels over a network using known protocols, including SMTP (simple mail transfer protocol), HTTP (hypertext transfer protocol) or FTP (file transfer protocol)as shown in figure 2.7 and figure 2.8.

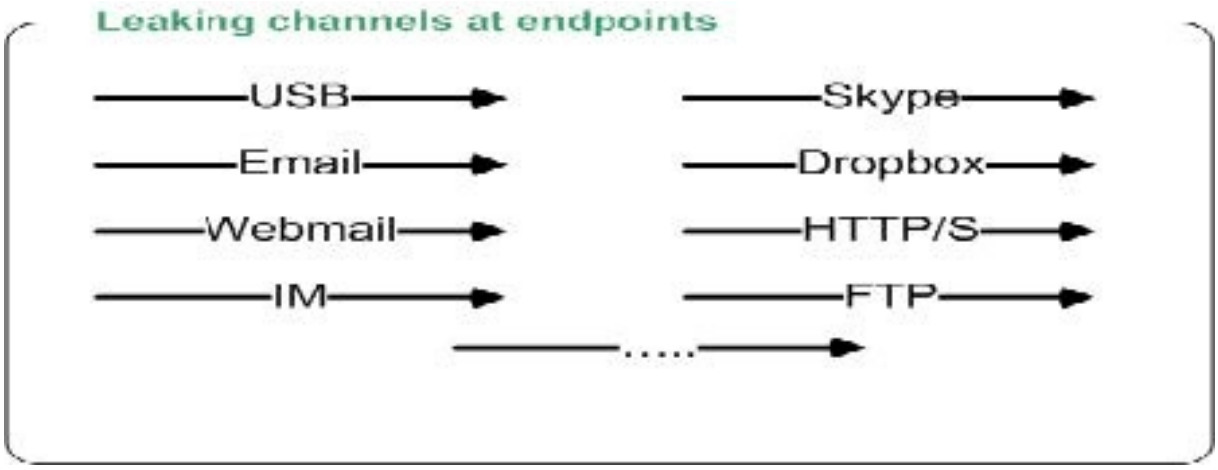

Figure 2.7: Data in Use [7]

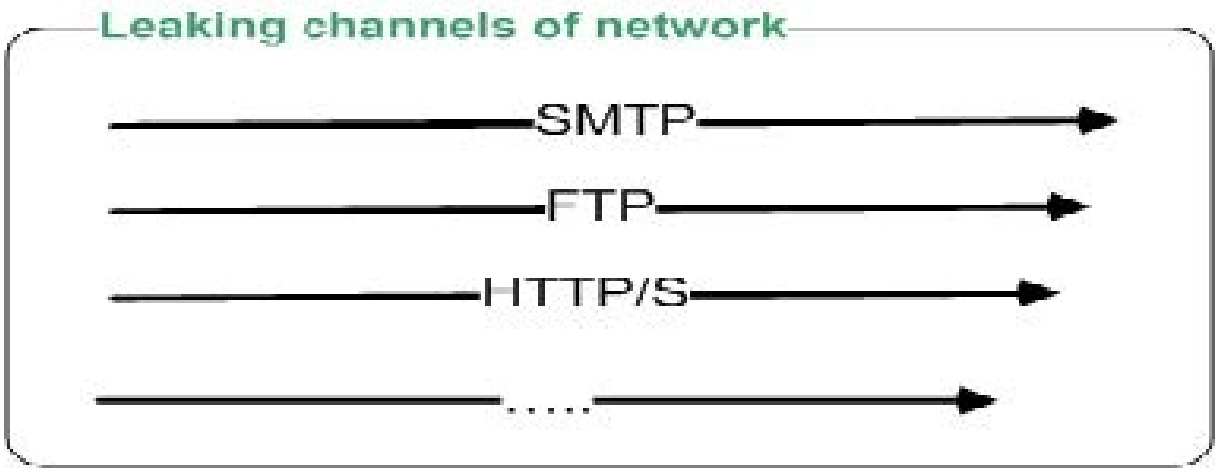

Figure 2.8: Data in Motion [7]

2.7.2 Data Leakage/Loss Prevention Capabilities

Data Leakage/loss Prevention system is to deliver a unified solution to discover, monitor and protect confidential or sensitive data wherever are stored or used, across endpoint, network, and storage systems [39]. DLP solutions are used to address the business problem of protecting confidential data [40]. The solution components are designed as following and shown in Figure 2.9 [9].

- Discover where confidential or sensitive data is stored.
- Monitor how data is being used.
- Protect data to prevent its loss.

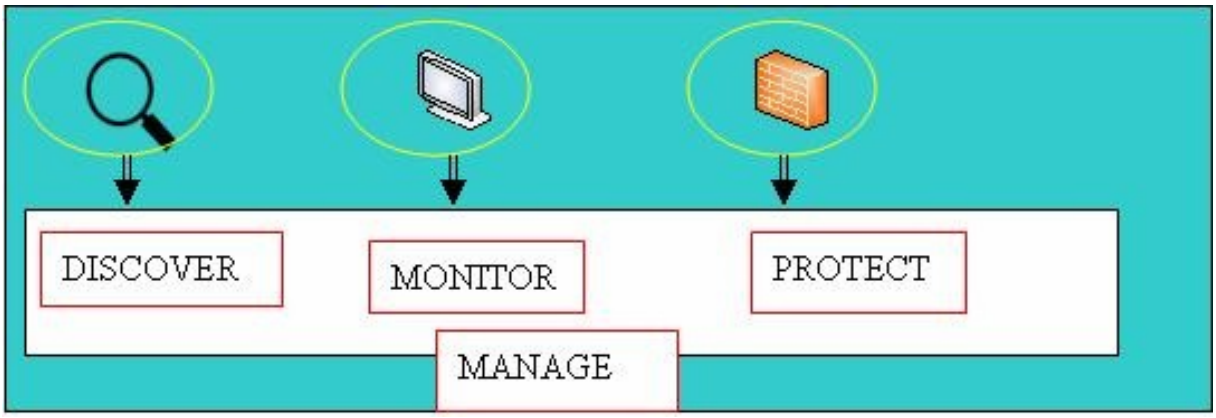

Figure 2.9: DLP Loss Modes [5]

2.7.3 Main Deployment Options

Two primary options are used when setting up DLP products:

Endpoint: DLP software deployed on endpoint devices or clients. They typically protect data as they are being used by the user (data-in-use) and detect sensitive data stored on the endpoint (data-at-rest) [7].

Network: A DLP solution could be implemented on the network level. By examining

network traffic, and controlled by a predefined policy, events could be terminated and the suspect transmissions could be blocked [7].

2.8 Leakage Handling Approach

Leakage occurrences can be treated by using detective/discover/investigate approach. In this approach, the system will try to discover leakage incidents and will make suitable remedial action to handle any determined leakage incident. Such as, if a local DLP agent discovers a file including confidential data on an un-authorized server, it could move the file to the secure place, such as a secured repository [7]. Which includes:

- Context-based inspection.
- Content-based inspection.
- Content tagging.

2.8.1 Context-Based Inspection:

This model controls the numerous security systems such as firewalls, proxies, intrusion detection / prevention systems (IDS/IPS), and spam filtering. The idea of context means contextual information taken from the monitored data, for example source, destination, size, recipients, sender, header/metadata information, time stamps, file type, location, format, application, and queries or transactions. A sample of a context inspection-based technique is the packet-filter firewall, which determines whether a network packet might be permitted to go through on. The basis of specific filtering conditions is for example source/destination IP address, source/ destination port, and also other packet attributes. A context-based DLP solution can prevent Java files from being sent of an organization, block all encrypted files, or prevent copy-and-paste operations from specific applications [7].

2.8.2 Content-Based Inspection:

This method discovers data leakage by examining content by using a number of techniques such as:

- Combining Vocabulary: containing keywords for example "Policy", "Password", "Salary" and patterns or standard phrase matching (e.g., a 16-digit pattern for a credit card number).

  Many systems include common dictionaries that address regulations and laws for example PCI-DSS(Payment Card Industry Data Security Standard), HIPAA(Health Insurance Portability and Accountability Act), GLBA(Gramm–Leach–Bliley Act), and SSN (Social Security Number). This kind of detection method is the simplest and fastest to configure [7].

- Fingerprinting: an approach that takes "fingerprints" from confidential files or database entries and seeks actual fingerprints to detect leakage. A fingerprint is an ideally unique hash value associate with a set of data. Hash values for confidential files are kept in databases or locally on the device under inspection. The system compares these hash values with parts of the inspected data. Fingerprints can be created on database records [7].

- Learning Method (LM): DLP uses this approach the basic idea from it is to use machine learning techniques such as Vector Space Model (VSM), to determine the "confidentiality level" of the scanned email message [7]. Approach used 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 documents are confidential or not.

One of the most used algorithms in content inspection is the Boyer-Moore-Horspool algorithm or Horspool's algorithm is an algorithm to find substrings in strings. It had been

created by Nigel Horspool in 1980 [41]. It's a simplification of the Boyer-Moore string searches algorithm that is related to Knuth-Morris-Pratt algorithm. The main element of the algorithm is summarized in appendix C:

2.8.3 Content Tagging:

In this method, a label is associated with a file including confidential data, as well as a policy is forced based on the given label [41]. Content will remain tagged even if processed by other applications. For example, a word file that was labeled as confidential will stay labeled as confidential even if it is encoded or zipped. Labels could be given to files in a different way: manually through the creator of the confidential data; automatically using content- or context-based analysis; automatically for all files kept in a specific area; or automatically to all files made by specific applications or users [7].

2.9 Other Security Systems

Through my study of the existing systems and DLP system, each system has advantages and disadvantages.

1) 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.

2) 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 (such as, laptop computers and desktops) and also on the network. These may include network and personal firewalls IDS/IPS, antivirus, antispam, encryption and digital rights management tools.

3) 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 most

sensitive information is stored, who has access to it, and where and by whom it is sent outside the company network. Existing security applications cannot perform this level of monitoring.

4) DLP systems must offer comprehensive functionality to prevent this sensitive data from being sent outside the organization through an endpoint computer or through the network.

2.10 DLP Products and Technologies

To be able to obtain an excellent idea of how existing DLP system work, both open source and commercial products from DLP systems. A check of the eight following DLP systems has been conducted:

1. Websence Data Loss/Leakage Prevention(WSDLP).
2. Open Data Loss/Leakage Prevention(OpenDLP).
3. Symantec Data Loss/Leakage Prevention(SDLP).
4. Trustwave Data Loss/Leakage Prevention(TDLP).
5. MyDLP Data Loss/Leakage Prevention(MyDLP).
6. RSA Data Loss/Leakage Prevention(RDLP).
7. MacAfe Data Loss/Leakage Prevention(MCDLP).
8. Furtinet Data Loss/Leakage Prevention(FGDLP).

1) WSDLP is one of the new DLP solutions which arose in 2007 was the WSDLP suite. It is currently available in version 5.8 and exists both as a full DLP suite, including network monitoring and data discovery, and as an endpoint monitoring solution. The central management is realized via a plug-in which could be integrated into the Websense management console Data Security Management (DSM) [42].

This management console is the central point for administering all kinds of Websense security and software including: the Data Security Protector Components and Endpoint Agents. So the central maintenance of DLP policies can be easily integrated into the workflows of administration process [42].

The solution basically provides the monitoring of removable storage media like USB flash or CD/DVD discs. All content which should be written to these media is monitored and analysed based on the central policies. Websense Data Security protects company from data loss by:

- Discovering the location of sensitive data inside the network
- Monitoring data as it travels inside or outside the organization
- Protecting data while it is being manipulated in office applications, with policy based controls that align with business processes.

The Websense DLP solution consists of several models:

- Data Security Management (DSM).
- Data Security Protector (DSP)
- Data Security Endpoints (DSE)

2) OpenDLP is a free and open source, agent and agentless based, centrally-managed, suitable for windows, unix, MySQL databases. OpenDLP can concurrently discover sensitive data at servers, laptops and workstations. OpenDLP has two components [15] [43]:

- central management for managing agents and push policies for all endpoint agents.
- microsoft windows agent used to execute scans on different systems. It is currently available in version 0.5.1 released 2012.08.27.

3) The symantec approach for data loss prevention is to deliver a unified solution to discover, monitor, and protect confidential data wherever stored or used across endpoint,

network, and storage systems. Symantec provides software solution to address the business problem of protecting confidential data. The technology components are designed to address the following issues [13] [44]:

- Discover where confidential data is stored.
- Monitor how data is being used.
- Proactively protect data to prevent its loss.

Symantec's group of product works together in conjunction with a centralized management platform to report and reduce data loss risk [45].

4) TDLP assists companies discover, monitor and secure data at rest, in motion, and in use to prevent leak this data [17].

TDLP is focused on the three support beams of data security: monitor, protect, discover. The solution allows companies to achieve complete visibility into all external attacks and insider risk.TDLP administrator handles all the TDLP products from a central console [46].

5) MyDLP is free and open source data loss/leakage prevention system. It may be installed on both network servers and workstations. MyDLP is a DLP solution available in two versions; MyDLP Enterprise Edition and MyDLP Community Edition. For these versions, we evaluated the freely available Enterprise Edition. MyDLP can be used to provide solutions for data leakage prevention in network and endpoint protection and discover sensitive data[64]. Supported data inspection channels include web, mail, and file transfer to removable storage devices and printers. The management console is the central point for administering and managing policies for all endpoint agents. It is easy to deploy any policies that we need [18].

6) RSA DLP discover and monitor the location and flow of sensitive data for example client credit card data and salary slip [19]. Teach clients and enforce controls to stop loss of sensitive data through email, web, PCs, laptops, and more [47] [48].

7) MCDLP among the DLP products which happened in 2008 was the McAfee Host Data Loss Prevention package. It is currently included in version 4.8 and is available both as a full DLP package, such as network monitoring and information discovery, and also as an endpoint monitoring solution. The central management is recognized by using a plug-in which can be built-into the MacAfee management console called ePolicy Orchestrator (ePO) [49] [50].

This administration console is a vital point for providing all types of MacAfee security software including the anti virus parts and endpoint agents. So the central maintenance of DLP policies can be easily integrated into the workflows of administration process [16] [51].

MCDLP Administrator handles all the MacAfee DLP products from a central console, then shows incidents and events discovered by them on its dashboards. In the unified policy design, policies could be configured to discover incidents and violations everywhere on intranet i.e. in network traffic, in repositories and on endpoints. Actions can be added to any kind of rule to address any issue when discovered.

8) The FGDLP product enables to stop confidential data from leaving the network. When specify confidential data patterns, data matching these patterns will be blocked, or logged allowed while moving through the FortiGate unit. DLP system filters based on file type, file size, or a regular expression.

Data leak prevention inspects network traffic for data patterns you specify. You define whatever patterns you need the FortiGate unit to find in network traffic [20].

2.11 Steganography in DLP Systems

Problem Description

Steganography is the art and science of concealing data by embedding information and files in image. steganography works by changing amount of useless or unused data in regular computer files (such as graphics, sound, text, or HTML). This concealed data could be plain text, cipher text, or even image [52]. steganography and encryption are

both used to ensure data confidentiality. However, the primary difference in between them is that with encryption, anyone can see that both sides are communicating secretly. steganography hides the existence of secret message and in best case; nobody can see that both parties are communicating secretly. For example we have two pictures [Figure 2.9 and Figure 2.10].

Figure 2.10: Original Image (without embedded data) [8]

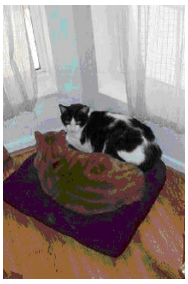

Figure 2.11: Steganography Image (contains embedded data) [8]

DLP systems have weakness in detection steganography leaks, so any hidden sensitive data could be sent to other parties.

How Data is Being Hidden

An electronic picture contains a matrix of colour and depth values. In the normal gray scale picture, 8 bits/pixel are used. In that typical full-colour image, we will find 24 bits/pixel, 8 bits given to each colour components [53].

Least Significant Bit (LSB) insertion is a very common and simple way of embedding information inside an picture. The least significant bit basically, the 8th bit of some or all of the bytes inside an image is modified to a bit from the secret message. If you use a 24-bit image, a bit of each of the red, green and blue colour components can be used, since they are each represented by a byte. In other words, you can store 3 bits in each pixel. An 800 600 pixel image, can thus store a total amount of 1,440,000 bits or 180,000 bytes of embedded data [53]. Such as a grid for 3 pixels of a 24-bit image can be as the following:

(00101101 00011100 11011100)

(10100110 11000100 00001100)

(11010010 10101101 01100011)

If the number 200, which binary representation is 11001000, is inserted to the least significant bits of this part of the image, the resulting grid can be as follows:

(00101101 00011101 11011100)

(10100110 11000101 00001100)

(11010010 10101100 01100011)

While the number was being inserted to the first 8 bytes of the grid, just the 8 bolded bits (LSB) were changed based on the embedded message [53]. On average, only 1/2 of the bits in an image need to be changed to conceal a secret message while using maximum cover size. Because there are 256 possible intensities of every primary colour, modifying the LSB of a pixel leads to small changes in the intensity of the colours. These modifications can't be observed through the human eye, thus the message is successfully hidden. The benefit of LSB embedding is its simplicity and many techniques use these techniques [53].

2.12 Literature Review of DLP Systems

Lately, the challenge of handling the malicious insider became acknowledged, and man of strategies are already suggested for handling this problem. Data leakage is one of malicious insider attacks, and therefore, most of the approaches were suggested for insider threat detection and prevention [27].
At first, Maybury et al [54] offered due his collaborative research which including a characterization and analysis of the useful approaches to counter malicious insider in the U.S. Intelligence Community. The research suggests a generic model of malicious insider behavior, differentive purposes, actions, and associated observables. Numerous prototype methods were produced for providing early warning of malicious insider activity.

Gaonjur and Bokhoree [55] focused on the risks of insider threats in IT outsourcing environments and suggested minimization of these risks using non-deceptive techniques such as intrusion-detection systems, and also deceptive techniques such as honey pots. A properly secured network with suitable intrusion-detection systems and honey pots configuration will assure the overall security and integrity of user's sensitive information.
Salem et al [27] and lately Hong et al [28] surveyed some suggested techniques for discovering insider attack in the study literature, including host-based user profiling depending on features such as database, file system access, system calls, and OS commands, and also network based detection.

Franqueira et al [56]. distinguished between internal insiders and external insiders. He surveyed numerous detection approaches, including auditing, updating access rights, behavior monitoring, security policies, and security assurance. He also suggested the variation of keys when using these techniques for discovering internal and external insiders. The author mentioned that external insiders aren't susceptible to the same security measures that are used for detecting malicious internal insiders. So, mitigating the external

insider threats must begin with evaluating the existent of the potential insider by mapping and determining the data being shared, the business partners and etc.

Spitzner et al [57] mentioned the following primary benefits of honey pots. First, honey pots gather data only if somebody or something malicious communicates with them. This makes the data gathered through the honey pots very brief, accurate, simple to handle, and easy to analyze. Second, honey pots can determine and catch new attacks. Because any action using the honeypot is anomalous by definition, new or hidden attacks are detectable and result in a reduced false negative rate.

Belasco et al [58] prevents e-mail leakage intentionally or inadvertently and handles the email with very confidential information to non-desired users. Using the learning machine, he additionally provides a solution to this problem. He were able to discover leakage in 82% of the test cases with this solution.

Luft [59] raised the problem: "can data loss prevention stop data leakage?". After evaluating two DLP systems, the indications were that DLP system even today has problems and weaknesses in preventing data from leakage. Furthermore, he discovered that the system did not correctly secure the connection between the DLP server and DLP agent. There's a probability for interception and eavesdroping for the traffic that could include confidential data such as accident reports. Additionally as the author described, there are many of available DLP systems that need to be evaluated in similar way.

Sara Ghorbanifar and Glenn Fryklund [60] propose enhancing DLP systems' security by evaluating four different existing kinds of DLP systems available in today's market, and revealing their weaknesses and suggested methods for improving their security. They concluded that the current DLP systems are still in need of improvement. No one has tested all the DLP solutions and fully covered. The possible leakage points that could

exist in the business world, there's a need for continued evaluation of DLP systems.

Through our studies in the previous research, most of it does not provide sufficient protection against intentional attacks from insider. Protection from intentional attacks can be partially achieved using a standard Right Management System (RMS), access control and authentication, anomaly detection (aimed at detecting data), and honey pots. In our opinion, improving the methods for detecting sensitive data can increase the effectiveness of DLP solutions which leads to enhance DLP security.

Chapter 3

DLP Systems Evaluation

This chapter presents the model and methodology that will be used in evaluating DLP systems. Also presents the implementation, analysis and testing of the five DLP systems along with found problems.

3.1 DLP Systems Test Model

The DLP test was implemented to evaluate whether there's any information leakage occurs or not in these systems, using different operating systems such as Windows Server 2008, Linux with different databases such as SQL, Oracle. In order to check if the DLP systems or the product is secured or not, if DLP systems is vulnerable to attacks, and if someone hack the system, the test model was built based on several processes to determine if a DLP systems protect data and maintain functionality as intended as described below:

- Built Test Environment.

Due to the limited use of resources and setup, the intended test environment consisted of a virtual environment with computers running on several operating systems:

1. Windows 7 running service pack.

2. Windows 2008 server R2.

3. Ubuntu Linux.

4. DataBase Server (SQL server, Oracle).

5. Two Laptops - HP 250

6. Two Servers - Dell poweredge 2950.

7. Dell Optiplex with 7010 (16 GRAM CPU intel core i7)

The benefit of using a virtual environment is the ease to create clones of whole virtual environments that will decrease the time it takes to build all environments [61]. In addition, it allows easily managing and jumping between environments when needed.

- Built Test Cases.

The real examination of the application comes from the capabilities. The examination of endpoint agents requires a subset of all test cases. The endpoint agent for removable media is examined, the Facebook/Dropbox/Skype for upload files is examined, the web mail for attached files is examined, the printer for sensitive data is examined, and so Table 3.1 lists the necessary evaluation items.

Table 3.1: Test Cases For DLP Systems

Discover	Are sensitive key-word matched and is data blocked?
React	Is sensitive data blocked? Are incidents reported properly?
security	Can any weaknesses found?

- Results.

To check the DLP products specific tests performed to determine whether the solution can satisfy the requirements placed in Table 3.1. To examine the capabilities in the system, an easy policy was created: Every file which contains the string Password/Policy/Salary/Credit-card should be blocked from being written for any removable storage devices.

The first step is to create policy which determines the key-word by defining the sensitive data from content classifier as shown in figure 3.1. Any file that contains any of these key-words should be blocked.

The file should be blocked when copied to removable media or when uploaded to Facebook/Skype/Drop box and when file sent through e-mail or printer.

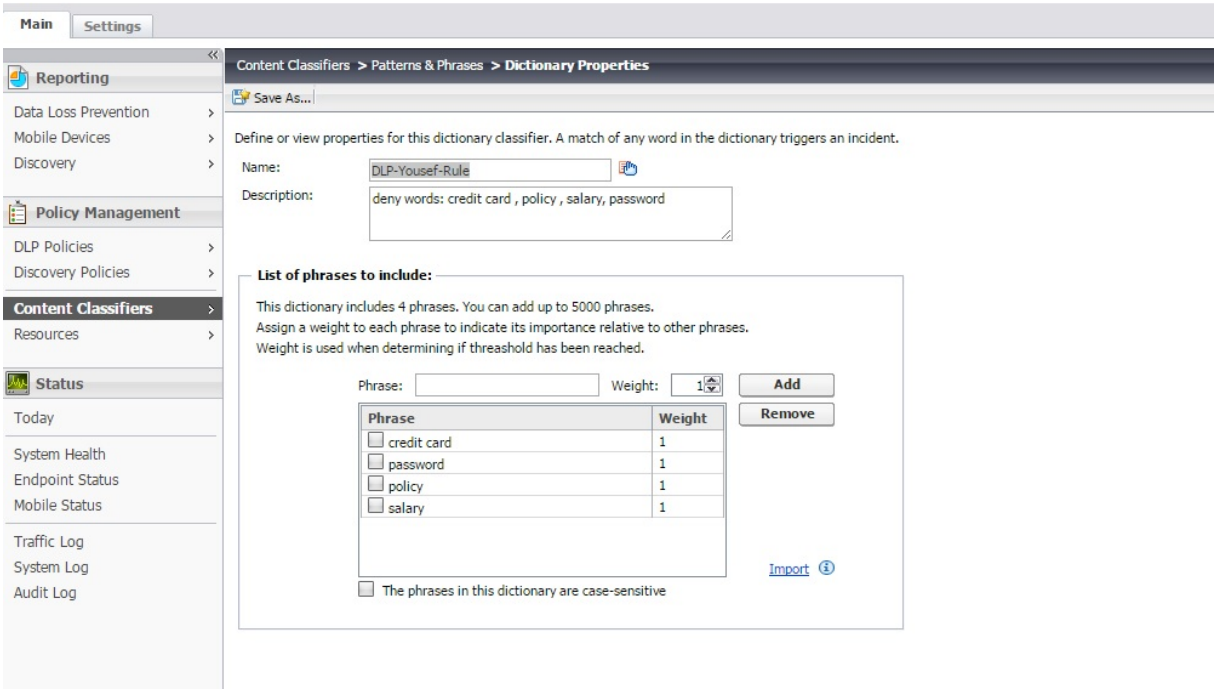

Figure 3.1: Key-Words Matching Definition

This policy could be applied using the text pattern shown in Figure 3.1. The reaction rule should be blocked.

3.2 DLP System Test Methodology

Because the DLP sector still is new and growing, we should evaluate certain products before relying on them [11].

The five DLP system was mentioned in section 2.10, has become installed and completely tested, once a time on these computers. In order to cover the leakage point with external hard disks and printers, a USB flash drive and a network printer has been used.

The tests also look at when a user has different privileges on the local machine. The tests were to leak sensitive data using webmail clients (hotmail), and cloud syncing software (Dropbox), social media (Facebook), instant messaging software (Skype), printers and USB devices. The data attempted to leak was:

1. A file containing the words considered secret: "Password", "Policy", "Credit card" and "Salary". These have then been put in a custom content dictionary to determine if the DLP systems can detect files that contains these words.

2. The same file sent through USB memory, Email, Facebook, Dropbox and Skype.

3. The same file compress by use different format such as (.zip, .7z and .rar.) and sent through, USB memory, Email, Facebook, Dropbox and Skype.

4. The same file is encrypted and sent through USB memory, Email, Facebook, Dropbox and Skype. (it was stored within a Kruptos tool).

5. The same file is encrypted, compressed and sent through USB memory, Email, Facebook, Dropbox and Skype.

6. The same file is renamed, put in video file format, and sent through USB memory, Email, Facebook, Dropbox and Skype.

7. The same file is renamed, put in picture file format and sent through USB memory, Email, Facebook, Dropbox and Skype.

8. The same file is hidden in picture file by using any hiding software and sent through USB memory, Email, Facebook, Dropbox and Skype.

9. Put secret words in different file types such as .PDF, .XLS, and .PPT, and sent through USB memory, Email, Facebook, Dropbox and Skype.

10. The same file is opened and sent through network/local printer.

For each of the experiments, every file that contains any sensitive word was set to be blocked. The tests were run through two various phases for every DLP solution. During the 1st phase, the attempt would be to build the DLP solution based on the recommendations and instructions available in order to prevent leakage of the data. Trials to move/leak the data were created using the applications and software mentioned before, and the result was Passed (P)/ Failed (F) for each experiment to show whether it was able to properly prevent the data from being leaked or not. For the tests that system didn't cover, a mark of not available (NA) is used. This stage also included how the solution reports prevention notification to both administrators and users. The second phase included trials to disable the DLP solution by different ways and then showing how well it was performed in each state. The tests were run again and graded the solution in the same way depending on how well it performed. To obtain attempts of attacking the DLP solution we tried:

1. Booting the computer in a safe made and check if the DLP agent works or not.
2. Stopping/killing the DLP agent to prevent it from running.
3. Disable/Uninstall for DLP agent.

the third phase will check if Communication Secure between central DLP server and the endpoint, and can be able to prevent communication between them.

These test cases produce concrete technical tests as summarized in Table 3.2.

These tests were executed on all suggest DLP systems with same configuration as listed in Table 3.3.

Table 3.2: Summary of Test Cases

Description	**File type**	**Email**	**Facebook**	**Printer**	**USB**	**Dropbox**	**Skype**
Phase1							
File is Attach/send							
File with sensitive data attach/send							
File is compressed with sensitive data attach/send	.ZIP, .RAR, .7Z						
File encrypted with sensitive data attach/send	Krup-tos						
File encrypted and compressed with sensitive data attach/send							
File is put in video file format attach/send	.Mp3						
File is put in picture file format attach/send	.Jpg, .png						
File Hiding sensitive data attach/send	Stegano graphy tech-niques						
Put sensitive data in different file types and attach/send.	.doc , .xls, .pdf, .ppt						
The same file is opened and sent through printer							
Phase 2							
The DLP agent is working safe mode							
Can stop DLP agent							
Phase 3							
Secure communication between Central DLP server and endpoint							

Table 3.3: DLP Systems Configuration

Phase 1	**File type**	**Configuration**
Test case		
File is compressed and sent through email, Facebook, web-mail,....	.ZIP, .RAR,.7Z	Policy set to block any file contain sensitive data.(Password/Policy/Salary/ Credit Card)
File is encrypted and sent through email, Facebook, web-mail,....	Kruptos	
File encryption and compression		
File is put in video file format	Mp3	
Put in picture format	Jpg, png	
Hidden secret data	Steganography techniques.	
Put secret words in different file types	.doc , .xls, .pdf, .ppt	
The same file is opened and sent through local/network printer		
Phase 2		
The DLP agent is working even in safe mode.		
Can stopping DLP agent		
Phase 3		
Secure communication between central DLP server and endpoint and the ability to prevent communication between them		

3.3 DLP Products Evaluation

This section presents the examination of the five DLP products based on the approach mentioned in previous section (3.1 and 3.2) as well as discovered problems/weakness and suggested improvements.

3.3.1 Product1 Data Loss/Leakage Prevention (P1DLP)

Product1 Data Security protects company from both information leaks and data loss at the perimeter and inside the company, as mentioned in section 2.10.

3.3.1.1 P1DLP Test Environment

Figure 3.2 shows the basic connection of the individual components. The central tool for administration is the Policy Server using Data Security Management (DSM). The Data Security Management Server is a Windows 2008-based machine on where you install the product1 Data Security software [42]. This machine provides the core information loss technology, capturing fingerprints, applying policies, and keeping incident forensics [61]. one can install multiple data security servers, sharing the analysis load, but one must be the primary as the management server [62].

The protector is a Linux-based machine that intercepts and analyzes traffic on any type of channels, such as email, HTTP, FTP, and chat [62]. It is an essential component of product1 data security providing monitoring and blocking capabilities to prevent data loss and leaks of sensitive information [63]. The protector can exactly monitor sensitive information in transit on any port [61]. Endpoints that are deployed on user's computers (PCs, laptops, and servers.) enable administrators to analyze content within a user's working environment and block or monitor policy breaches as defined by the endpoint profiles. DLP product1 using a Microsoft SQL Server created during the installation. These components are all separated services; it is possible to install them on various machines as shown in Figure 3.3. Dependent on the number of clients, this allows weighting

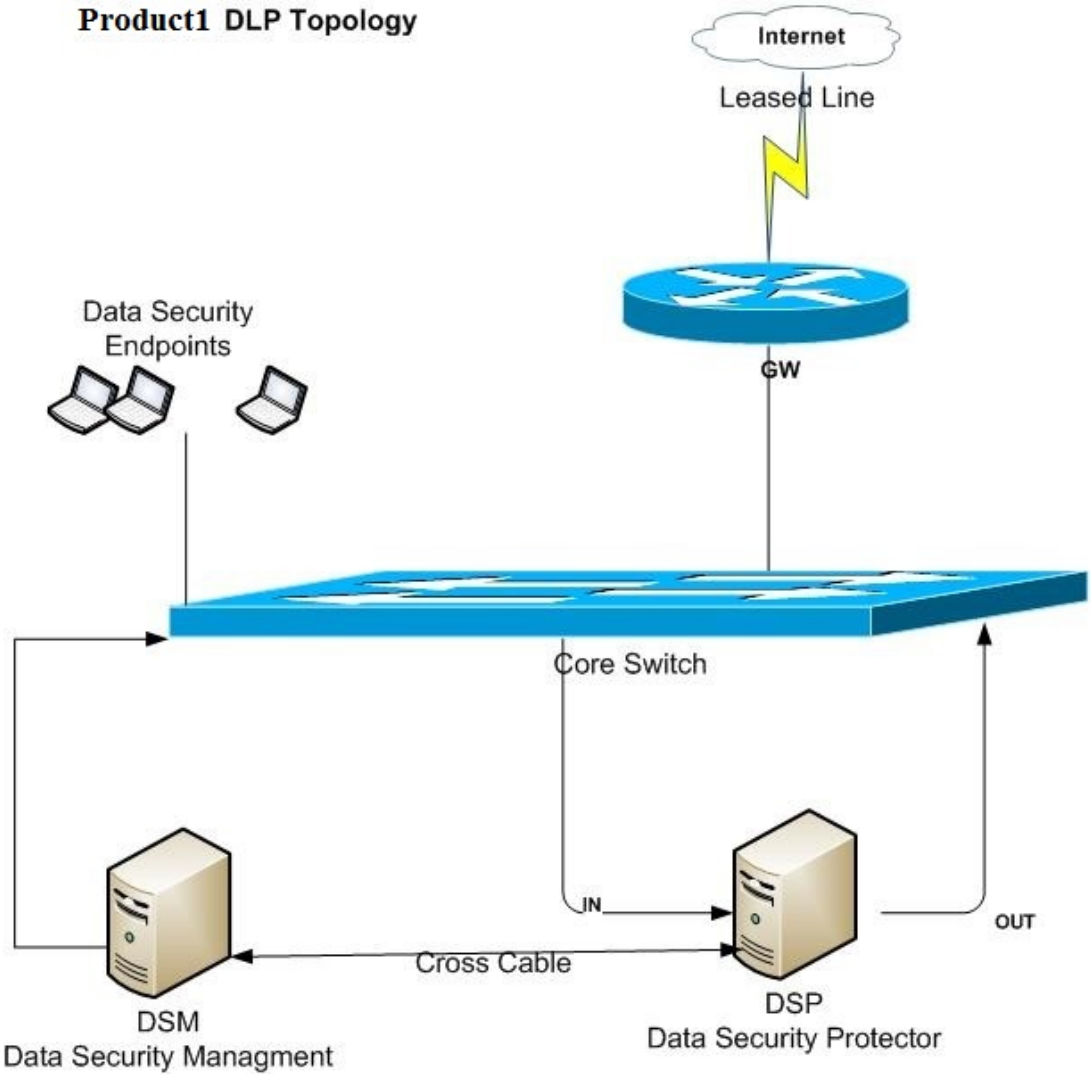

Figure 3.2: P1DLP Test Environment.

the performance according to the number of clients. In the built test environment, all services, including the database and the domain controller, ran on a single virtual machine as depicted in Figure 3.2. By applying this environment, we'll be able to monitor every removable storage device which is linked to each of the two client systems. There are many ways to determine sensitive data: Regular expressions, keyword matching and file type such as doc, zip, pdf. The defined keyword matching are applied to all data which should be read or written from removable devices and met the specification of the policy rule. This rule specification can include constraints to specific keyword matching types

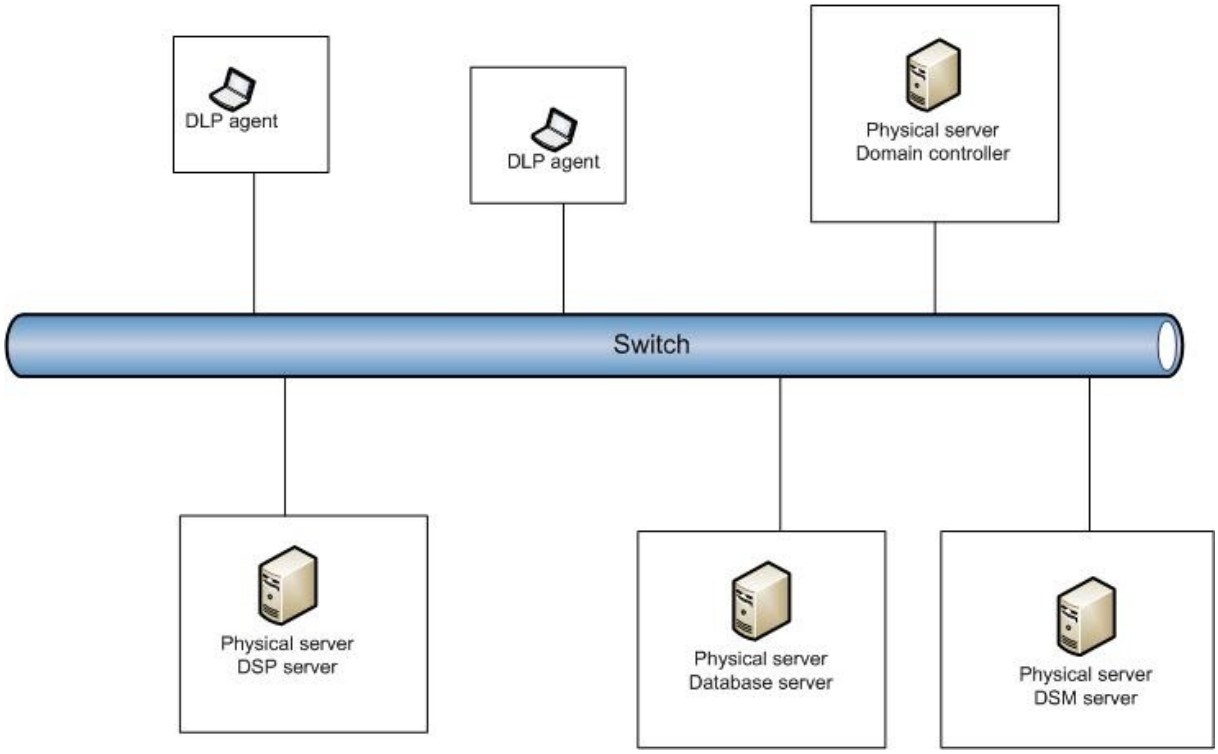

Figure 3.3: P1DLP Topology in Physical Server

or file type. Additionally, we will be able to mark files manually and thus mark special files which didn't meet any category. In case a breach of policy is detected, a notification is shown to the user, the data is blocked and an alert message is sent to administrator or manager.

The DSM Server is the main management and analysis component. Additionally it gives you log files, statistics and saved incidents [61]. Figure 3.2 shows the basic interaction

between the single components. The product1 solution installs its own database server by default. It may be possible to determine an existing database server, but since there is no need for a dedicated installation, it's not placed in Figure 3.2. The DSM Server is divided into three tools [42]:

1. Incident and Reports.
2. Policy Management.
3. Status and Logs.

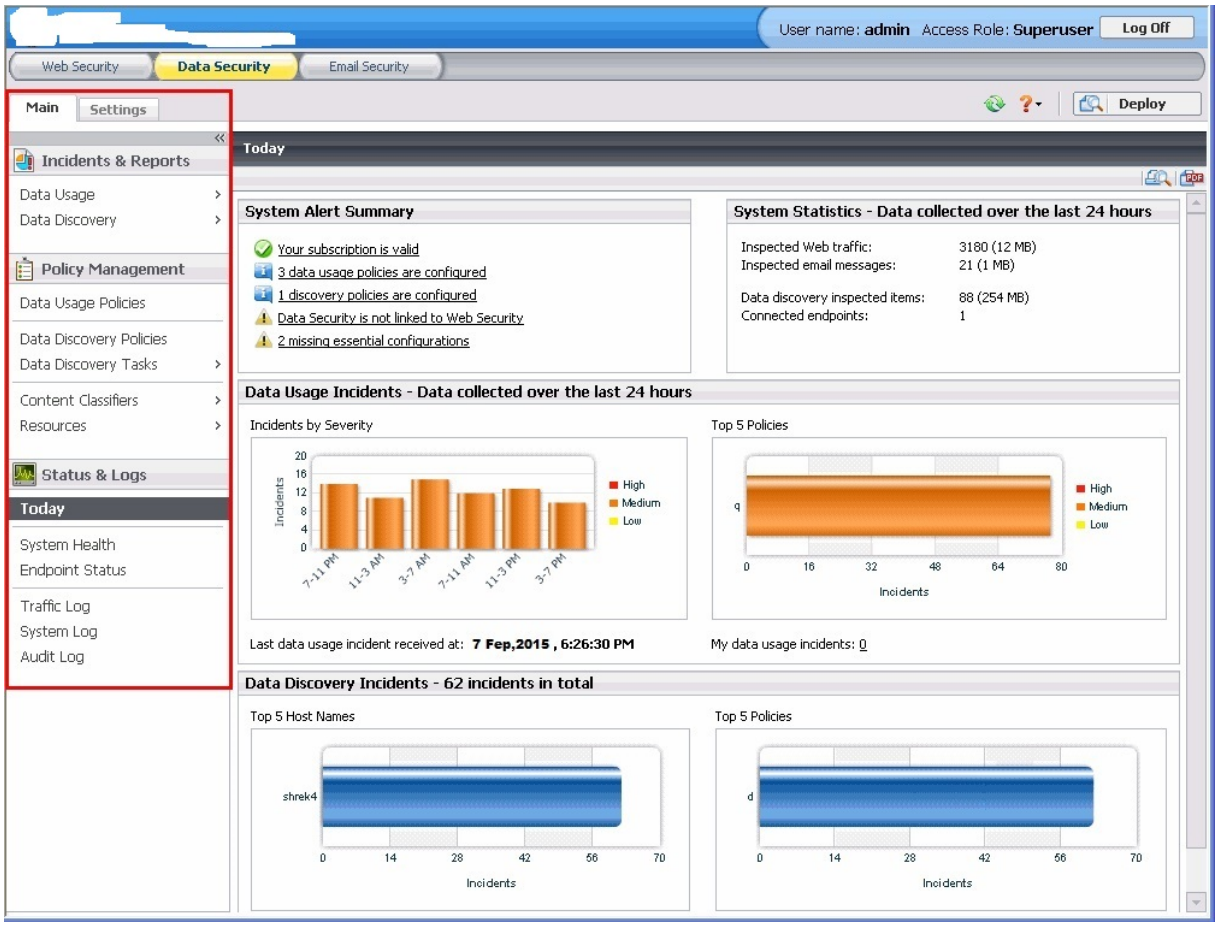

Figure 3.4: DSM Server Dashboard

DSM views and manages data usage incidents relevant to the active administrator, and creates or manages a network policy or endpoint policy. We can create policies from scratch or by using a predefined regulatory template. Also define the source and destination of the data you want to protect, the endpoint device or application that may be in

use, and the remediation or action taken when a violation is discovered (such as block or notify), as shown in figure 3.4.

3.3.1.2 P1DLP Test Cases

The same test cases are applied to all DLP product as mention in section 3.2. the DLP solutions can address the same leakage data and use similar protection methods as mentioned in section 3.1.

3.3.1.3 P1DLP Results

This section presents the results of the tests that performed on different test cases mentioned in the previous section (3.2). The first test was copying of a text file containing the text PASSWORD to an USB flash. As wasn't to become expected otherwise, the text file was blocked (Figures 3.5 and 3.6) and stored at database from the DSM server (Figure 3.6). After this test with correct functionality in DLP system, further tests about the key-word matching were performed. This policy could be applied using the text pattern shown in Figure 3.1. The reaction rule should be blocked.

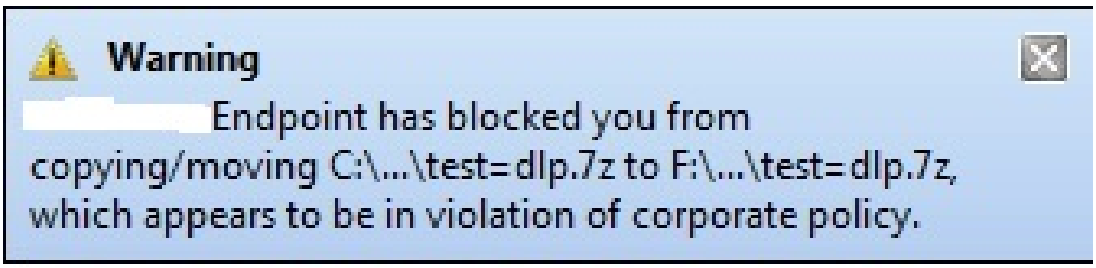

Figure 3.5: Password File Sent to a USB Flash by P1DLP.

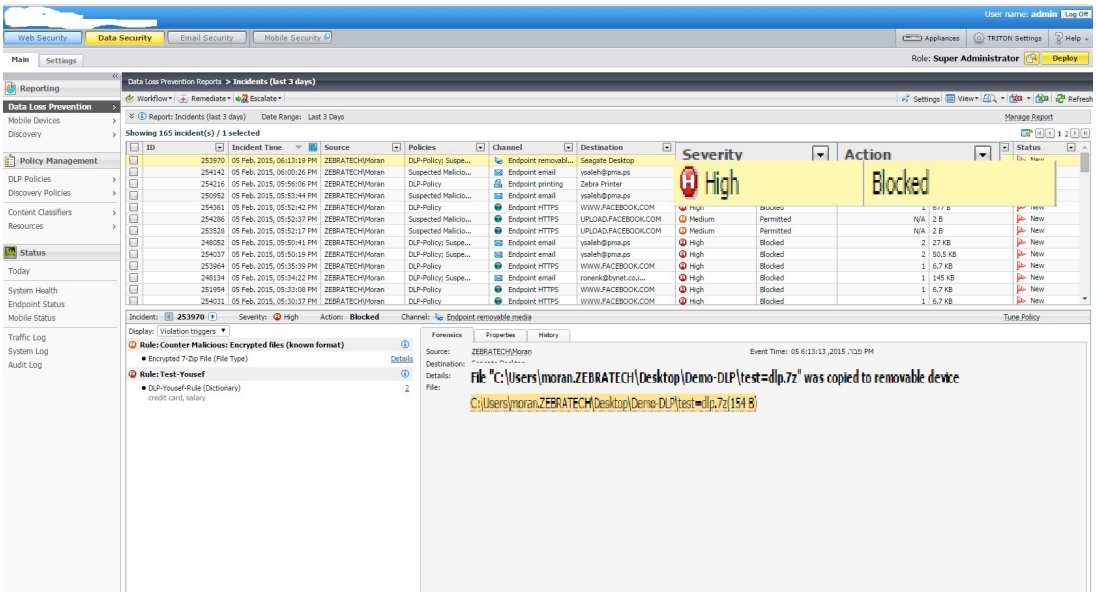

Figure 3.6: The Password File is Correctly Blocked by P1DLP.

Each DLP solution can handle many file types that may be understood and parsed. Therefore the following mixtures of more complex file types were generated:

1. A file including the any phrase: and sent through, USB memory, Email, Facebook, Dropbox and Skype.

2. A file including the phrase: “Password”, “ Policy”, “credit card”, and “salary” and sent through, USB memory, email, Facebook, Dropbox and Skype, as shown in figure 3.6.

3. The same file is compressed into Formats: .zip, .7z and .rar, and sent through USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.6.

4. The same file is renamed as video file format and sent through, USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.7.

5. The same file is renamed as picture file format and sent through USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.10.

6. Put secret words in different file types such as .PDF, .XLS, and .PPT as shown in figure 3.7.

7. The same file is opened and sent through network printer as shown in figure 3.9.

Since all these file types are supported, all files were detected to become valuable content and therefore are correctly blocked by P1DLP.

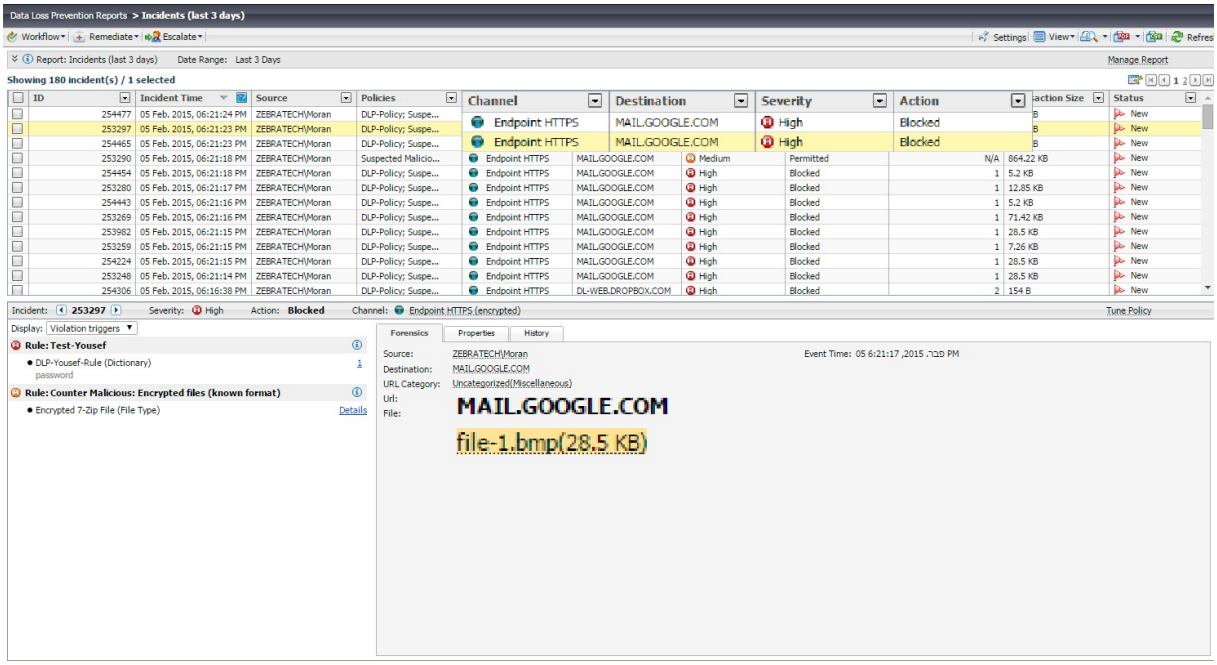

Figure 3.7: Password File in zip Format Sent Through Email by P1DLP.

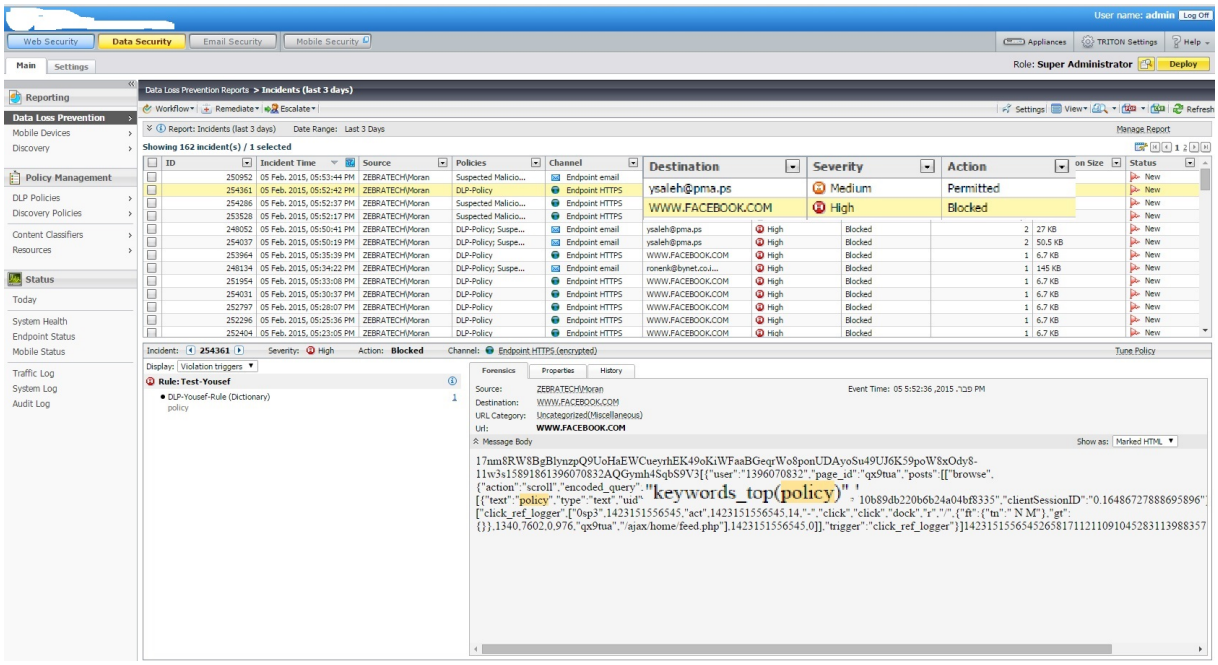

Figure 3.8: Password File in MP3 Format Sent Through Email by P1DLP.

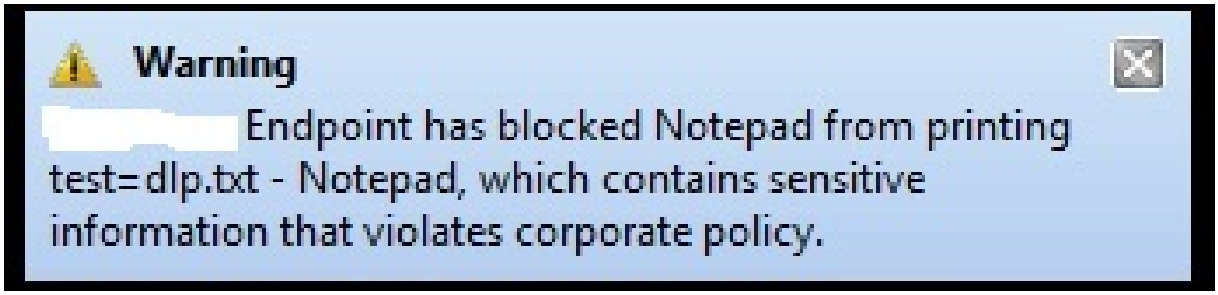

Figure 3.9: Password File is Printed on Printer by P1DLP.

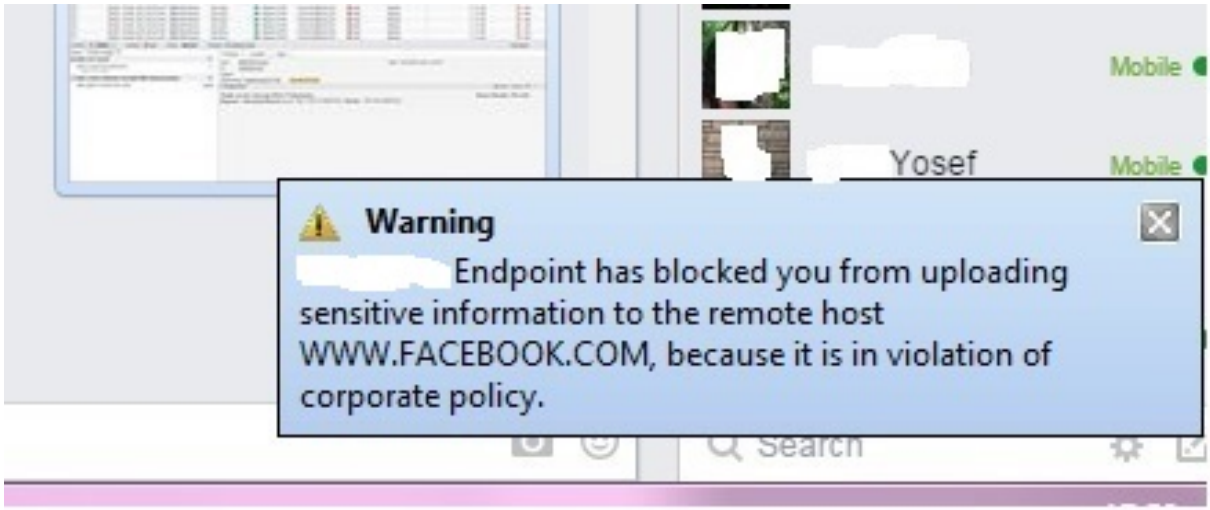

Figure 3.10: Password File Sent Through Facebook by P1DLP.

Disabled Agent

The user cannot disable the DLP agent until the super-admin creates a verification code on DSM server and sends it to the user to put this code on end point to disable the agent on the client host as shown in figure 3.11.

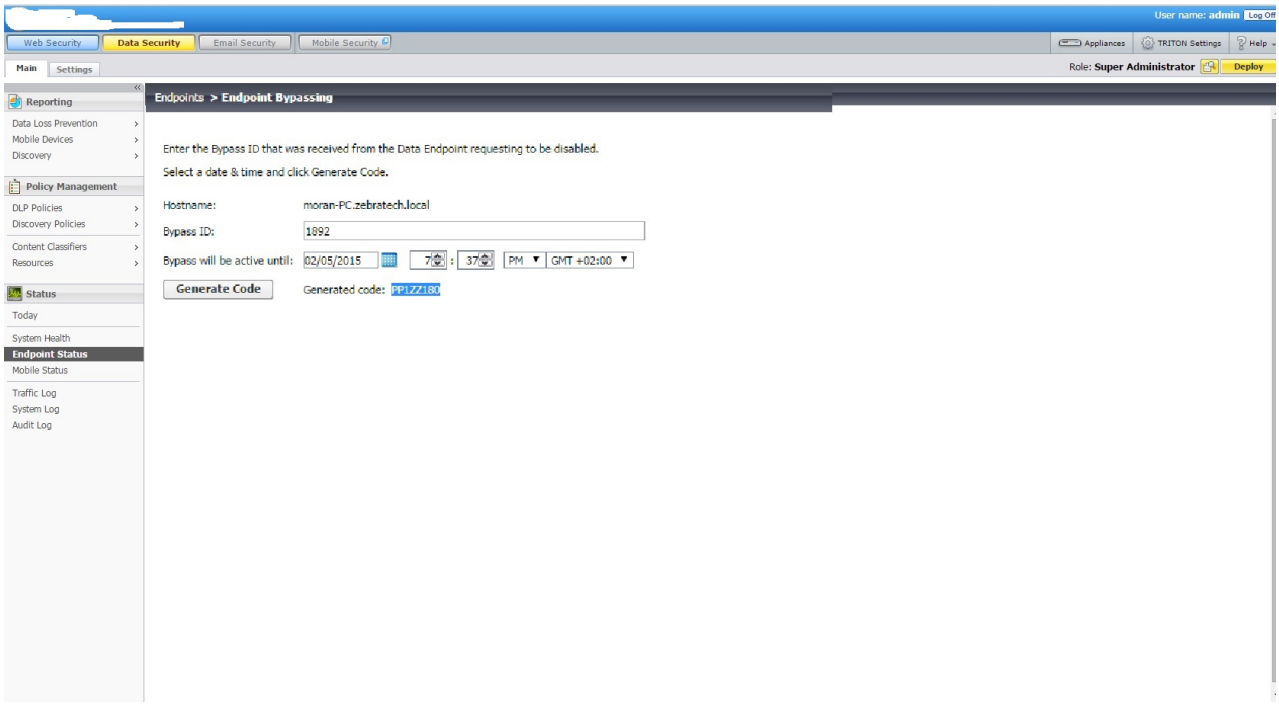

Figure 3.11: Agent Disable in P1DLP.

When disconnect the DLP agent from DLP server and send a file that is blocked by a policy, the file is also blocked (Offline mode works). When the agent reconnects to DLP Server, the logs are sent to the server when the connection has been re-established.

Secure communication between central DLP and endpoint.

The endpoint agent communicates with the server over HTTPS. This implicitly means that every communication is encrypted. These messages include the exchange of policy updates and event logging.

3.3.1.4 P1DLP Leakages Cases.

To check whether a real deep content analysis is performed, the test was repeated using the same word file by hide this file in an image by using stenography technique. The

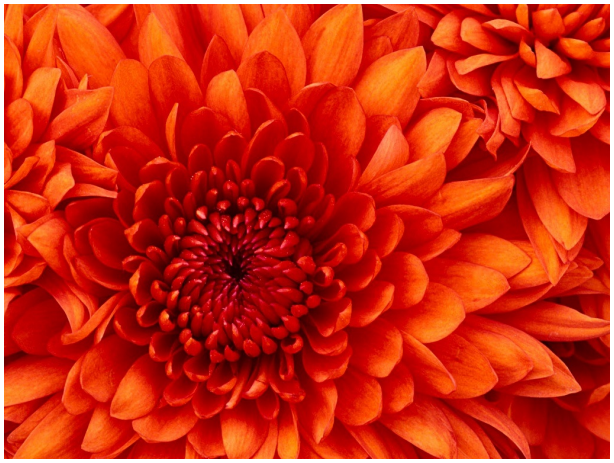

Figure 3.12: Steganography Hidden Secret Data.

sensitive words is put in an image by hiding these data in that image.We use OurSecert tool to hide data in image as show in figure 3.12. But the file was not detected through the discover process. In general, this means that the hidden data types are not discovered in product1 DLP as shown in figure 3.13. The image that contains a sensitive data was successfully sent to flash and e-mail.

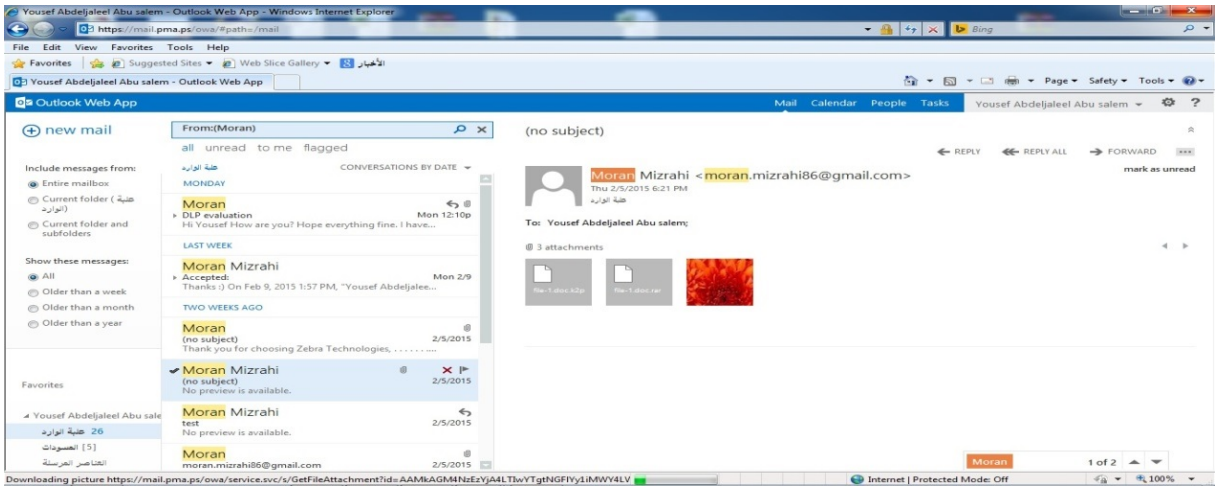

Figure 3.13: Send Hidden-In-Image Data Through Email in P1DLP

- In case when file contains sensitive encrypted data and sent through, USB memory, email, Facebook, web-mail and skype. (It was stored within a Kruptos file format).
- The same file is encrypted, compressed and sent through, USB memory, email, Facebook, web-mail and Skype.

- Product1 Agent does not work in safe mode.

Since all these cases are not supported, all these files were not detected and so not blocked in DLP Product1.

Table 3.4 shows the summarized result of P1DLP tests.

Table 3.4: Summarized Result of P1DLP Tests.

Description	**File Type**	**P1DLP**
Phase 1		
File is send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P
File with sensitive data attached/send.	.TXT,.Doc	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	P
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	F
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	P
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	P
File Hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .pdf, .ppt	P
The same file is opened and sent through printer.	Local/network	P
Phase 2		
The DLP agent is working safe mode.		No
Can stop DLP agent		No
Phase 3		
Secure communication between Central DLP server and endpoint		Yes

3.3.2 Product2 Data Loss/Leakage Prevention(P2DLP)

The P2DLP approach for data loss prevention is to deliver a unified solution to discover, monitor, and protect confidential data wherever stored or used across endpoint, network, and storage systems.

Product2 Data Leakage Prevention Architecture

The P2DLP solution consists of many products designed to work together to monitor and protect sensitive data wherever the data is stored and however it is sent out of your organization network [18].

The architecture is made up of the management platform and a combination of one or more detection servers that detect and protect data in the following three areas as figure 3.14 shows:-

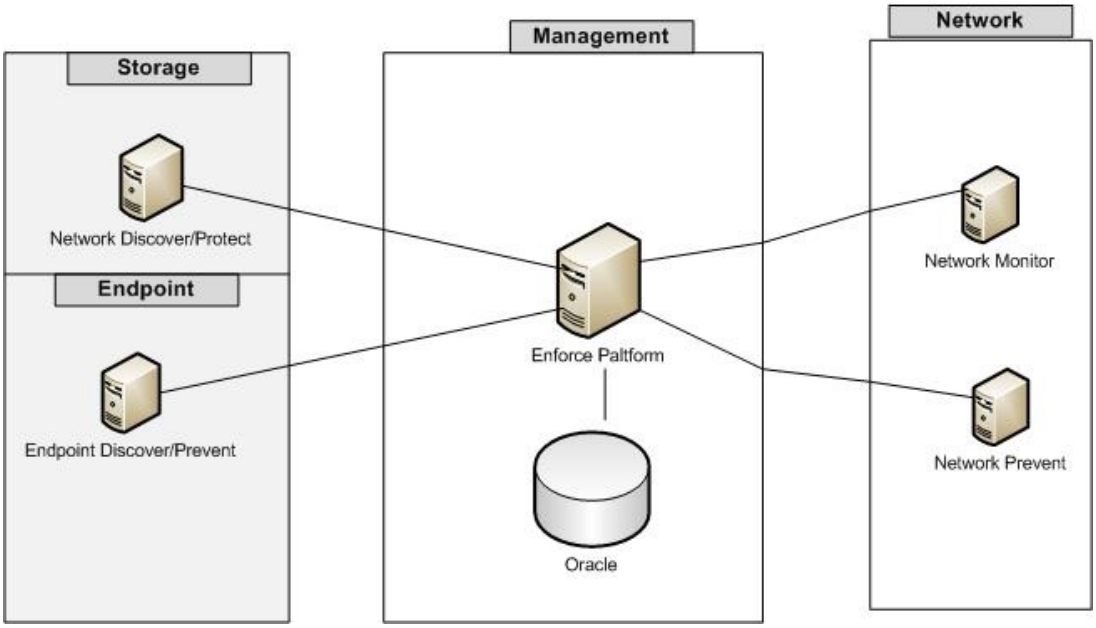

Figure 3.14: Overview of P2DLP Architecture

Network

- Network coverage is commonly referred to as data in motion
- Network coverage is monitoring and protecting data that is being transmitted over network to Internet. For example including common business applications, including e-mail as Simple Mail Transport Protocol (SMTP), web mail as Hyper Text Transport Protocol (HTTP), and File Transfer Protocol (FTP) [44].

Storage

- Storage coverage is commonly referred to as data at rest.
- Storage coverage scans and protects data that is stored in enterprise data repositories. For Example including public and private shares, and databases [64].

Endpoint

- Endpoint refers to devices such as laptops, desktops, or workstations.
- Endpoint coverage monitors and protects data as it moved to or off the endpoint machines. For examples including data downloading to hard drive and copying data to removable media (USB) [45].

The Product2 DLP Products

- Management Platform

Enforce platform is the central management platform, consisting of enforce server and oracle 10g database server. All users log into the enforce console. The user interface manages all DLP pollicises, workflow, reporting, users, roles, system management, and security. All detection servers are managed by the enforce platform. Enforce pushes appropriate policies, detection, and server configurations to the detection servers [44].

The oracle 10g standard edition database is delivered with the enforce platform. This database holds all incident snapshot details, reporting data, enforce platform configurations, and system information [44].

- Detection Servers

Detection servers inspect data according to policy rule criteria and if a match is found, detection servers create an incident that is transmitted securely to the enforce server for reporting and remediation. The six P2DLP detection servers are [45]:

Network

- P2DLP Network monitor (Network monitor) monitors network communications and searches for violations of data security policy.
- P2DLP Network Prevent (Network Prevent) proactively stops or redirects violating network compunctions [45].

Storage

- P2DLP Network Discover (Network Discover) scans storage location to expose confidential data.
- P2DLP Network Protect (Network Protect) works with network discover to protect stored data with automatic quarantine or copy of files.

Endpoint

- P2DLP Endpoint Discover (Endpoint Discover) includes agent-based scans of endpoints including laptops, desktops, and workstations at remote offices to expose confidential data.
- P2DLP Endpoint Prevent (Endpoint Prevent) includes agent-based monitoring and prevention of endpoint actions originating on laptops, desktops, and workstations at remote offices to expose confidential data. These actions may include monitoring files downloaded to local drives or blocking files from being copied to removable media devices, burned to CD/DVDs, copied or pasted to a drive, and printed [44].

3.3.2.1 P2DLP Test Environment

The test environment for P2DLP solution consists of several components [44] [13] [45]:-

- Active Directory
- Network Monitor
- Network Prevent
- Mail Getaway
- Web Getaway
- DLP enforce Management
- DLP for Client.

Figure 3.15 shows the basic connection of the individual components. The central tool for administration is the policy server using enforce management. The enforce management server is a windows 2008-based machine where you install the enforce management software. This machine provides the core information loss technology, capturing fingerprints, applying policies, and keeping incident forensics. In the built test environment use seven components: oracle database 10g, active directory, network monitor (as physical server), network prevent, web gateway, mail gateway and enforce management as a virtual machine by using VMware workstation. The component running different operating systems.

- Windows server 2008.
- Windows 7 running service packs 1.
- Oracle 10g.

By using this environment, we can monitor every removable storage device that is connected to each of the client systems. There are many methods to determine sensitive

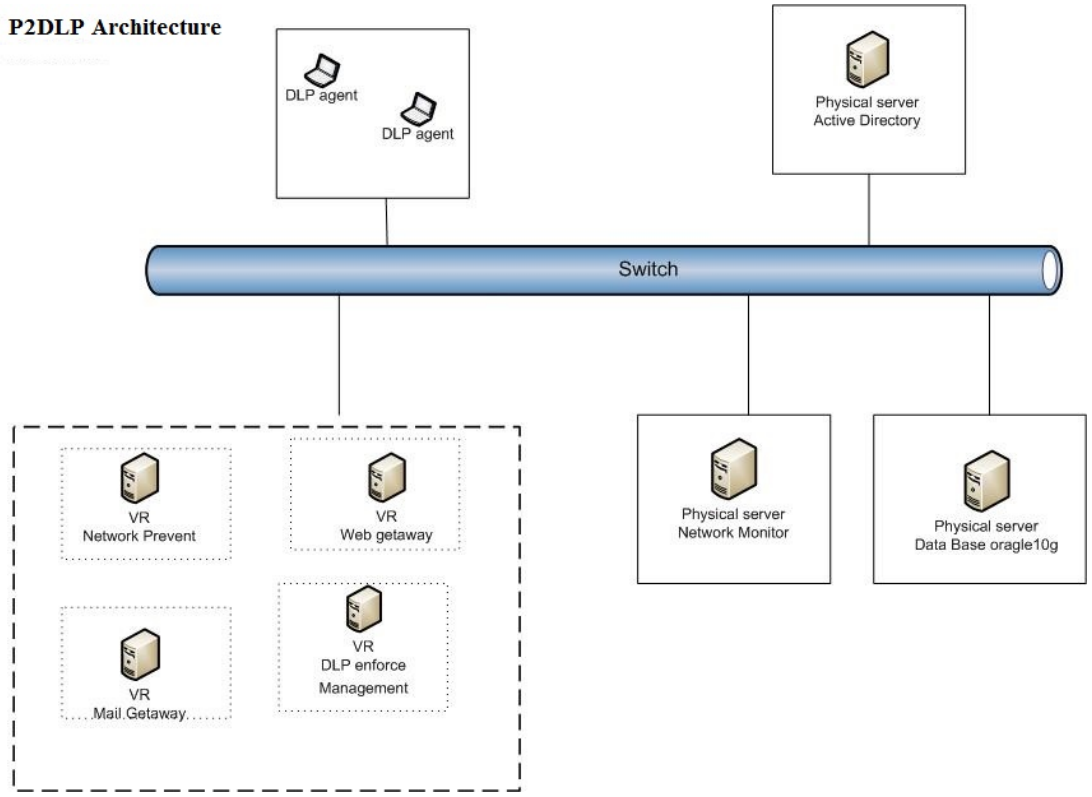

Figure 3.15: Test Environment For P2DLP

data: regular expressions, key phrase matching and file type (e.g. .doc, .zip, .pdf). The defined key phrase matching are applied to all data which needs to be read or written from removable devices and met the specification in the policy rule.

DLP enforce management is graphical user interface that manages as show in figure 3.16. GUI offers the ability to manage DLP policies, work-flow, reporting, users, roles, system management, and security. All detection servers are managed through the enforce platform. Enforce forces suitable policies, detection, and server configurations on the detection servers.

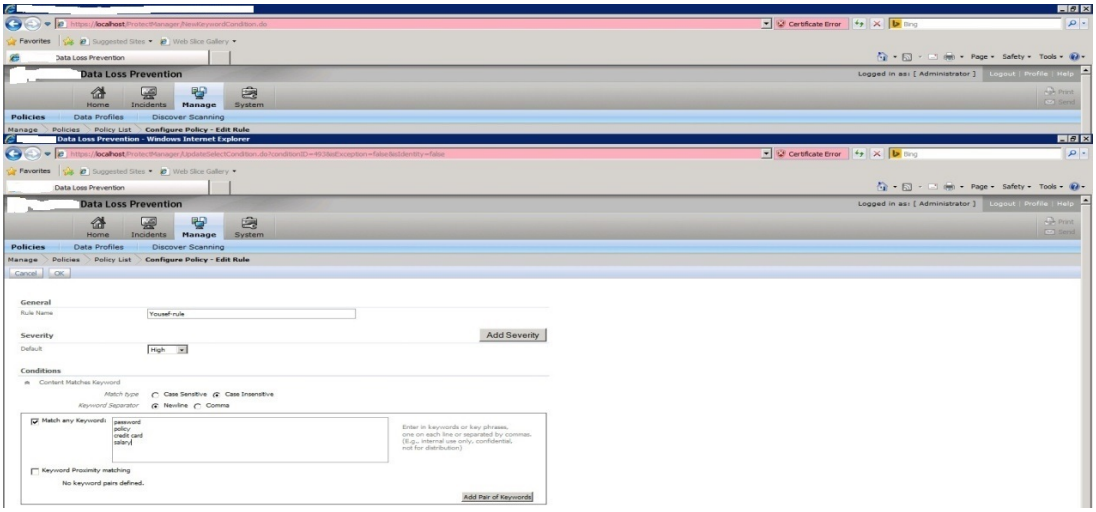

Figure 3.16: P2DLP Enforces Management

3.3.2.2 P2DLP Test Cases

The same test cases are applied to all DLP product as mention in section 3.2. The DLP solutions can address the same leakage data and use similar protection methods as mentioned in section 3.1.

3.3.2.3 P2DLP Results

The initial test was copying of the text file including the text PASSWORD to a USB flash. As expected, the text file was blocked (Figure 3.17) and stored in database on the enforce management server. After this test with correct functionality of the system, additional tests about the key-word matching were performed.

The initial test was copying of the text file including the text PASSWORD to a USB flash. As expected, the text file was blocked (Figure 3.17) and stored in database on the enforce management server. After this test with correct functionality of the system, additional tests about the key-word matching were performed.

Every DLP solution is designed for many file types that could be understood and parsed. and so the following mixes of more complicated file types were produced:

- A file including the any phrase: and sent through, USB memory, Email, Facebook, Dropbox and Skype.

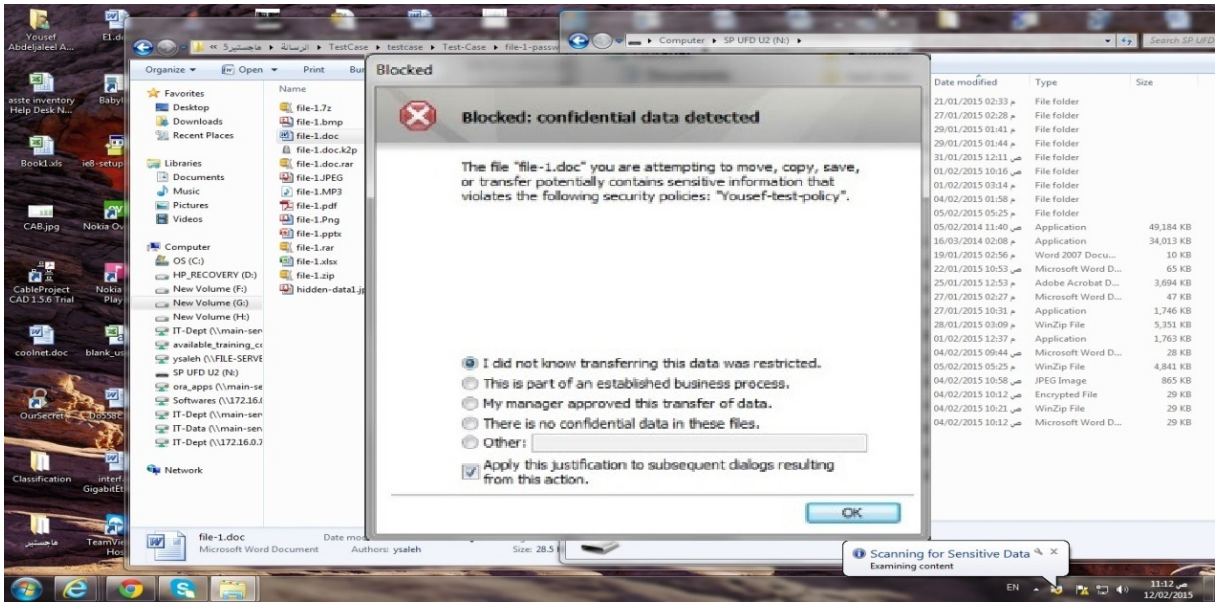

Figure 3.17: Block When Copying the Password File to a USB Flash in P2DLP.

- A file including the phrase: “Password”, “Policy”, “credit card”, and “salary” and sent through, USB memory, Email, Facebook, Dropbox, and Skype as shown in figure 3.17.

- The same file is compressed into Formats: .zip, .7z and .rar, and sent through USB memory, email, Facebook, Dropbox and Skype, as shown in figure 3.18.

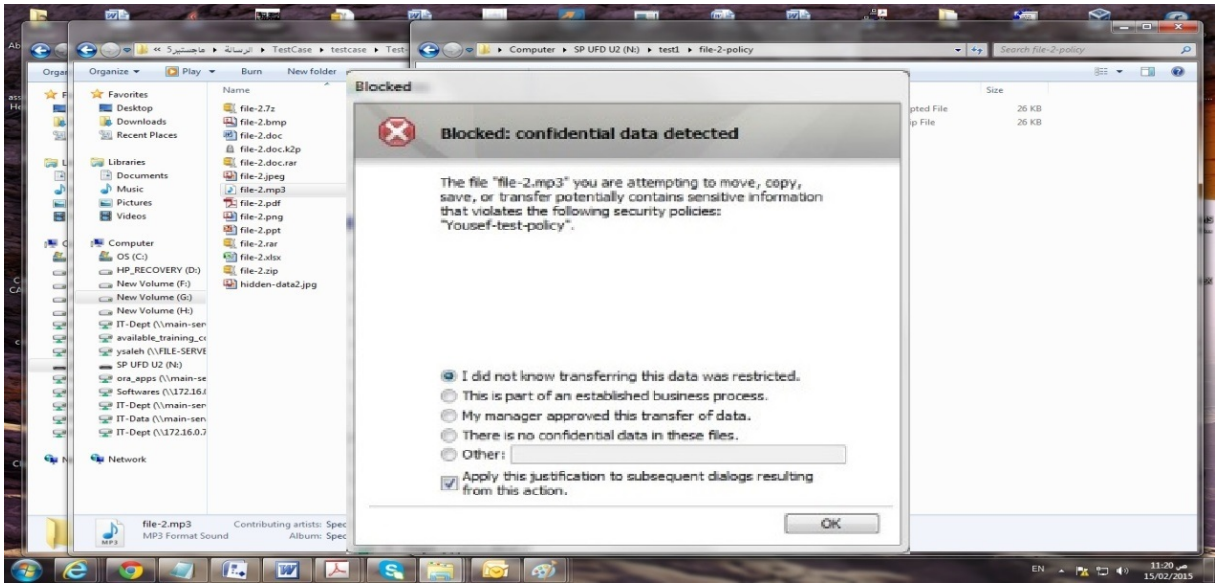

Figure 3.18: Block File When Sent Compressed File in P2DLP.

• The same file is renamed as video file format and sent through, USB memory, Email, Facebook, web-mail and Skype, as shown in figure 3.19.

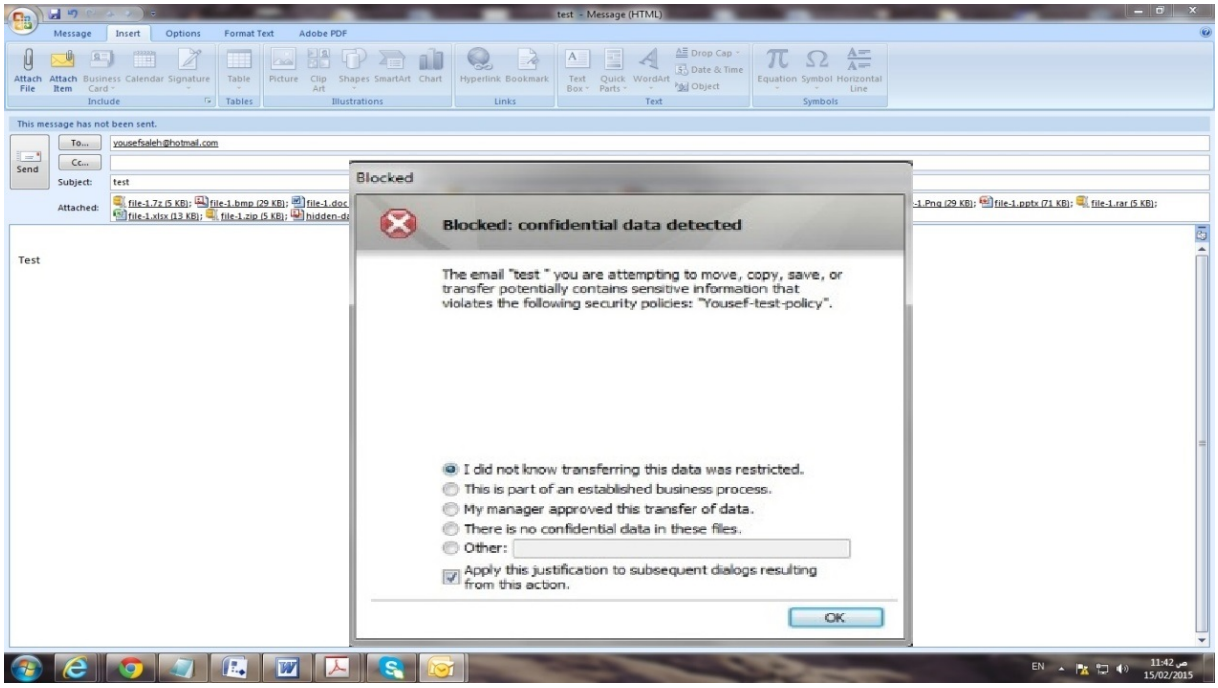

Figure 3.19: Block File When Sent in .mp3 Format in P2DLP

• The same file is renamed as picture file format and sent through USB memory, Email, Facebook, web-mail and Skype, as shown in figure 3.20.

• Put secret words in different file types such as .PDF, .XLS, and .PPT and send to USB as shown in figure 3.21.

• The same file is opened and sent through network printer as shown in figure 3.22.

Because these file types are supported, all these files were discovered as a valuable content and so are blocked.

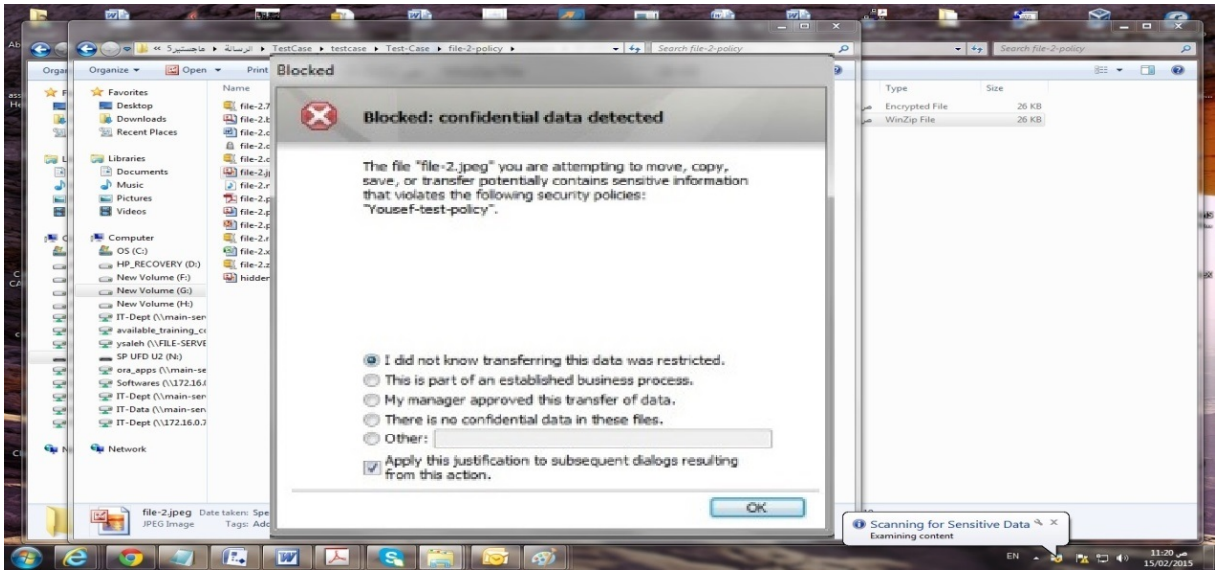

Figure 3.20: Block File When Copy In .jpg Format in P2DLP

Figure 3.21: Block File When Copy In .ppt/.pdf/.xls Formats in P2DLP

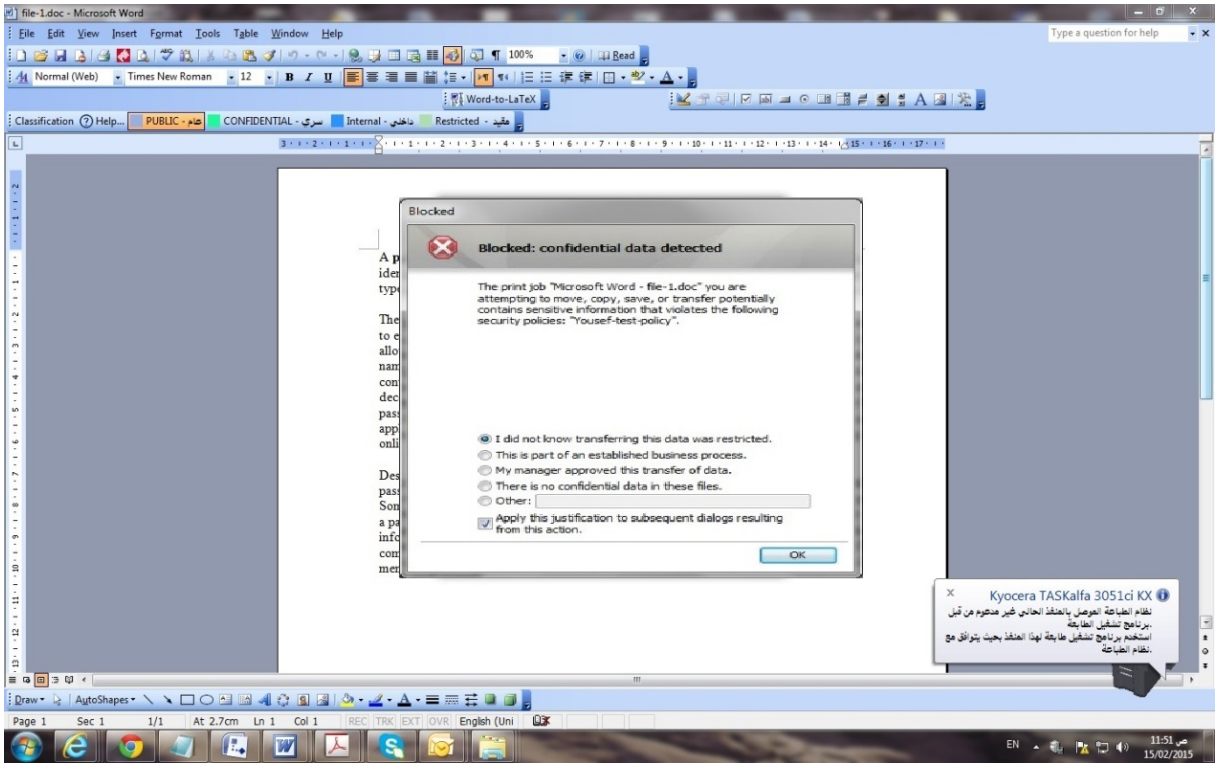

Figure 3.22: Block File When Print to Printer in P2DLP

Disabled/Turn-off P2DLP Agent

Disabling DLP agent does not delete the agent from endpoint server. Disabling an agent disable all data loss prevention monitoring on that endpoint computer. The associated events are still visible on the endpoint server. Unlike deleted agents, disabled agents can re-enabled. In the disabled state, agents do not receive policy or configuration update. Restarts the agent to retrieve the latest policy and configuration. The client cannot disable DLP agent. The administrator only can do it in advanced agent settings as shown in figure 3.23.

Secure communication between central DLP and endpoint

The endpoint agent communicates with the enforce management server over HTTPS. The P2DLP installation program asks if you want to install network prevent in hosted environment. If you choose to install a network prevent detection server in a hosted environment, you must use the SSL key tool utility. This ensures secure communication from the enforce server to the hosted network prevent server, and to all other detection

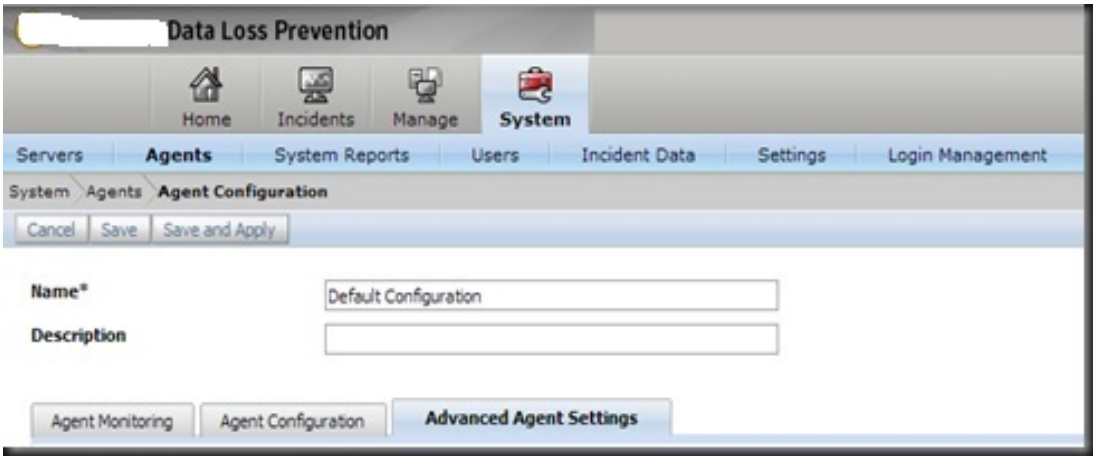

Figure 3.23: Disable Agent in P2DLP

servers that you install.

3.3.2.4 P2DLP Leakage Cases.

To test whether an actual deep content analysis is conducted or not, the test was repeated with similar word file but hiding the sensitive data within the image by using stenography approach. When copying the hidden data to USB flash memory, the DLP system cannot discover this leak as shown in figure 3.24. The sensitive words are embedded in an image by hiding these data in that image. We used OurSecert tool to hide this data, this tool allows a user to embed hidden data inside a carrier file, such as an image or video, and later extract that data, but the file wasn't detected over discovering process. Generally, that means the hidden data types aren't discovered in P2DLP.

- In case when file contains sensitive encrypted data and sent through, USB memory, email, Facebook, Dropbox and Skype. (stored within a Kruptos file format), the DLP system cannot discover this leak.

- The same file which is encrypted, compressed and sent through, USB memory, email, Facebook, web-mail and Skype, as shown in figure 3.25, the DLP system cannot discover this leak.

- P2DLP Agent does not work in safe mode.

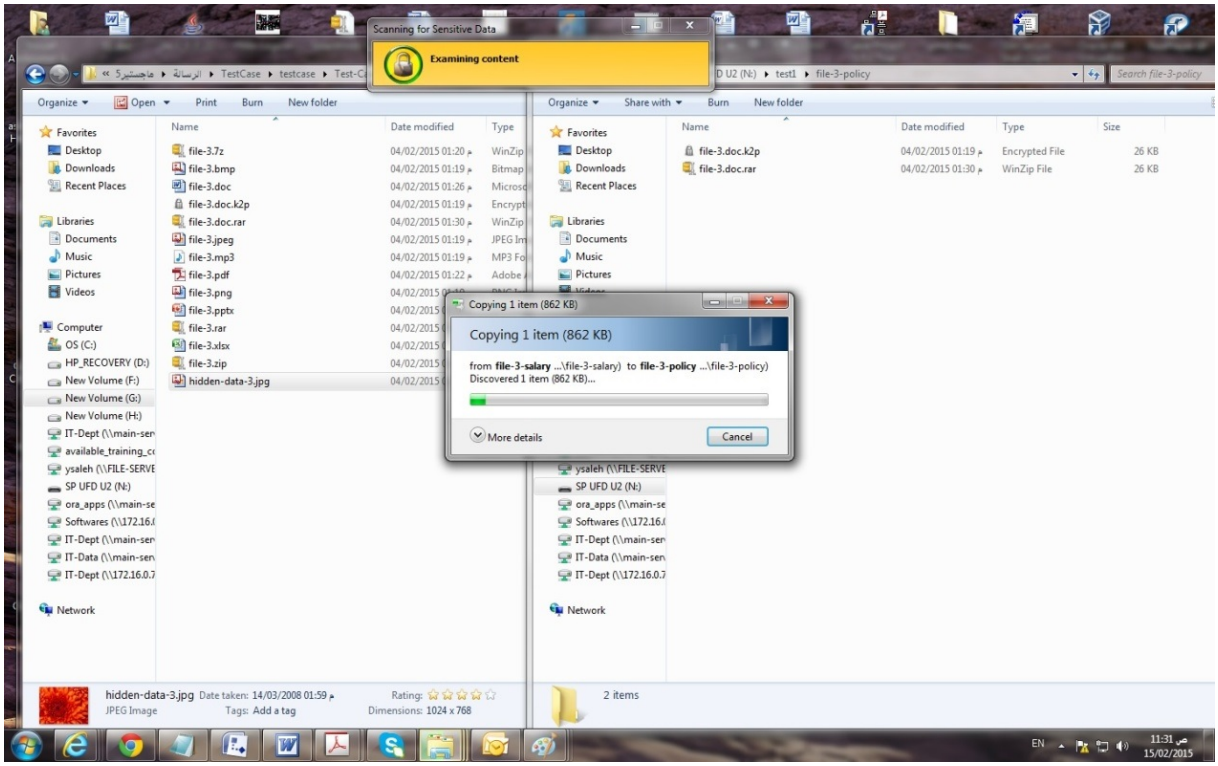

Figure 3.24: Hidden Secret Data Copy to USB Flash Memory in P2DLP.

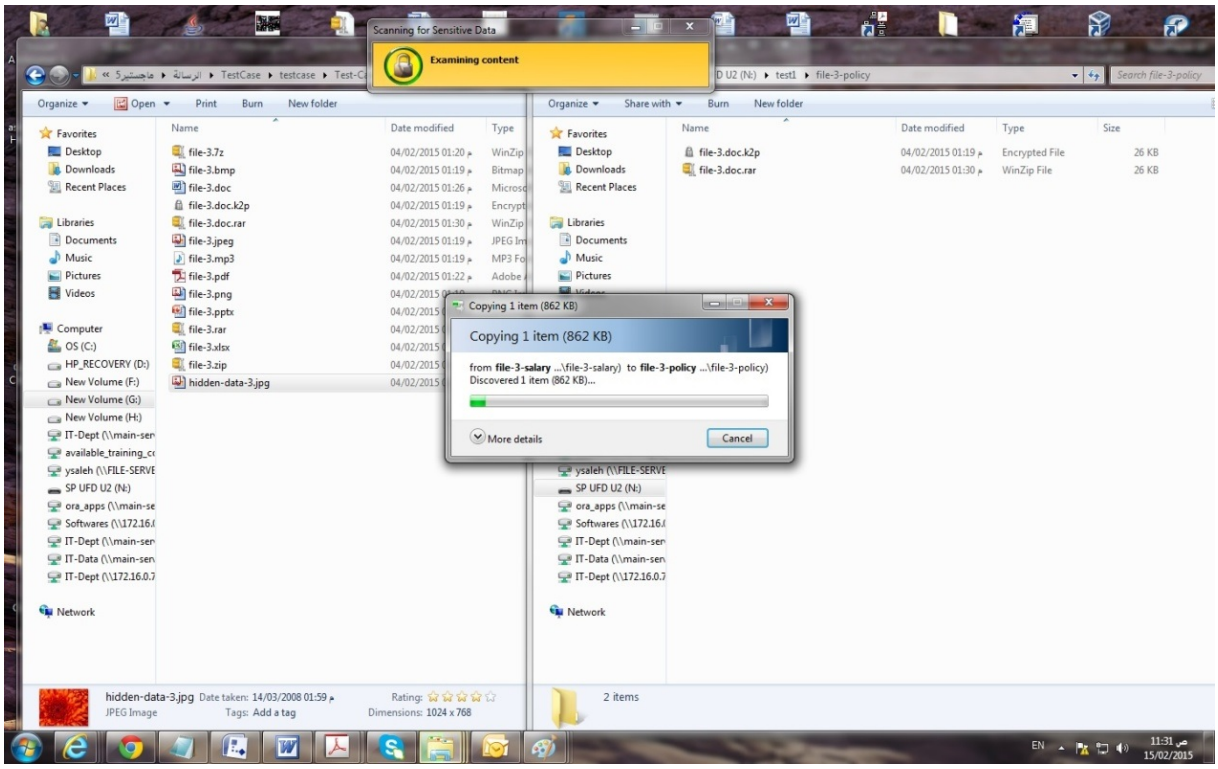

Figure 3.25: File Encrypted and Compressed is Copied (Didn't Blocked)in P2DLP

Because these cases aren't supported, and cannot block by P2DLP, these considered as weaknesses. Table 3.5 shows the summarized result of P2DLP tests.

Table 3.5: Summarized Result of P2DLP Tests

Description	**File Type**	**P2DLP**
Phase 1		
File is send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P
File with sensitive data attached/send.	.TXT,.Doc	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	P
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	F
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	P
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	P
File Hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .pdf , .ppt	P
The same file is opened and sent through printer.	Local/network	P
Phase 2		
The DLP agent is working safe mode.		No
Can stop DLP agent		No
Phase 3		
Secure communication between Central DLP server and endpoint		Yes

3.3.3 Product3 Data Loss/Leakage Prevention(P3DLP)

P3DLP is free and open source data loss/leakage prevention system. It may be installed on both network servers and workstations.

3.3.3.1 P3DLP Test Environment

We installed P3DLP server virtual appliance i386 version 12.04 on VMware machine. The server OS is Ubuntu 12.04 LTS. P3DLP can monitor, handle data motion and save data in organization. It is possible to pass, log, archive and quarantine moving data, encrypt removable devices and delete discovered files on storages using policy actions [18]. The two main components of P3DLP are P3DLP Network Server and P3DLP Endpoint. These two components work together to protect sensitive information in organization.

The figure 3.26 shows the architecture of P3DLP data leakage prevention system.

Apache HTTP Server was running to simulate the content of the web application. Although port 80 was open, only port 443 was used for accessing the administration web application. P3DLP used MySQL database created during the installation, used to store all events and logs for incidents. The squid proxy with ICAP (Internet Content Adaption Protocol) module is used for content inspection in network discover. Therefore network monitoring is done on the same server as the DLP administrative functions. The P3DLP beacon makes an auto discovery for sensitive data on server side [18].

These components are separated services, and it is possible to install them on various machines. In the built test environment, all services, including the database, ran on a single virtual machine.

By using this environment, we be able to monitor every removable storage device linked to each of the two client systems. There are many ways to catch sensitive data: regular expressions, key-word matching and file type matching such as doc, zip, pdf.

Management Console is a web based management user interface that enables users to configure policies, review brief history about incidents and monitor user activity.To

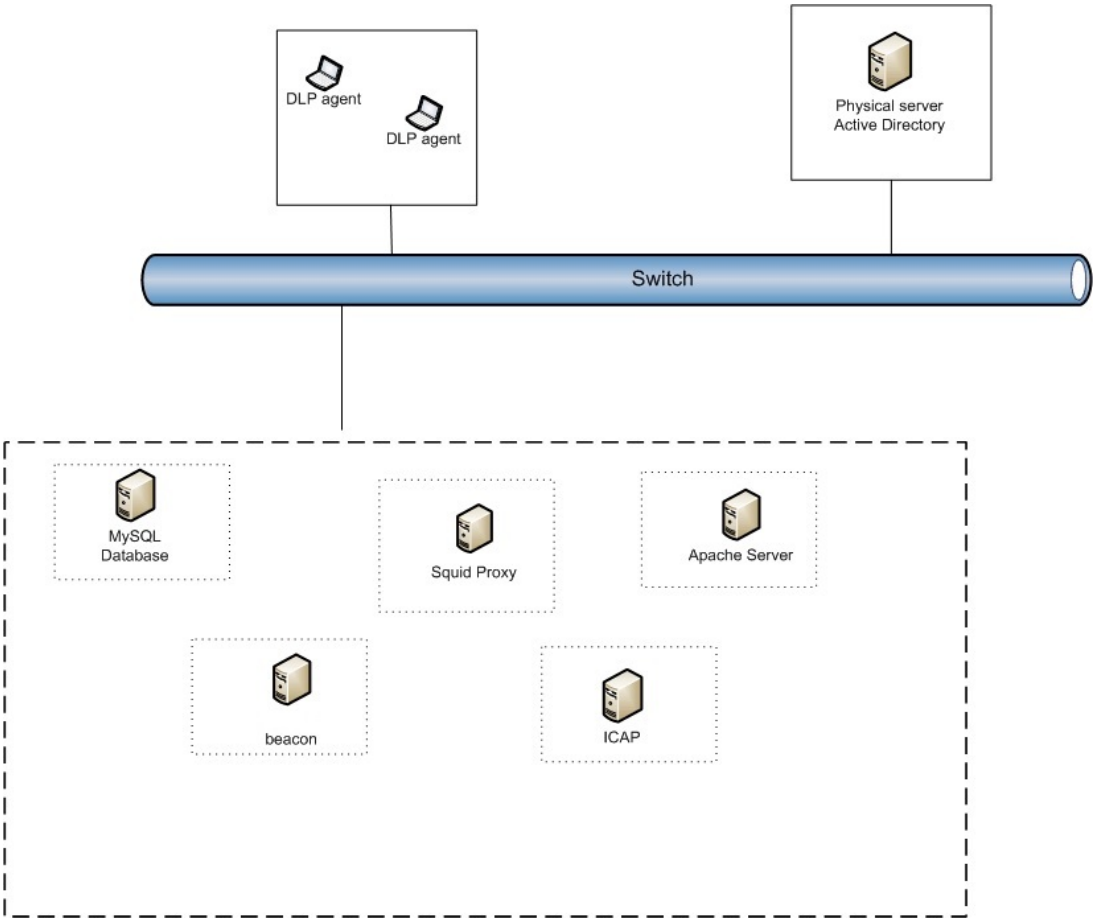

Figure 3.26: Test Environment for P3DLP Solution.

connect to management interface, we have to use web browser as shown in figure 3.27.

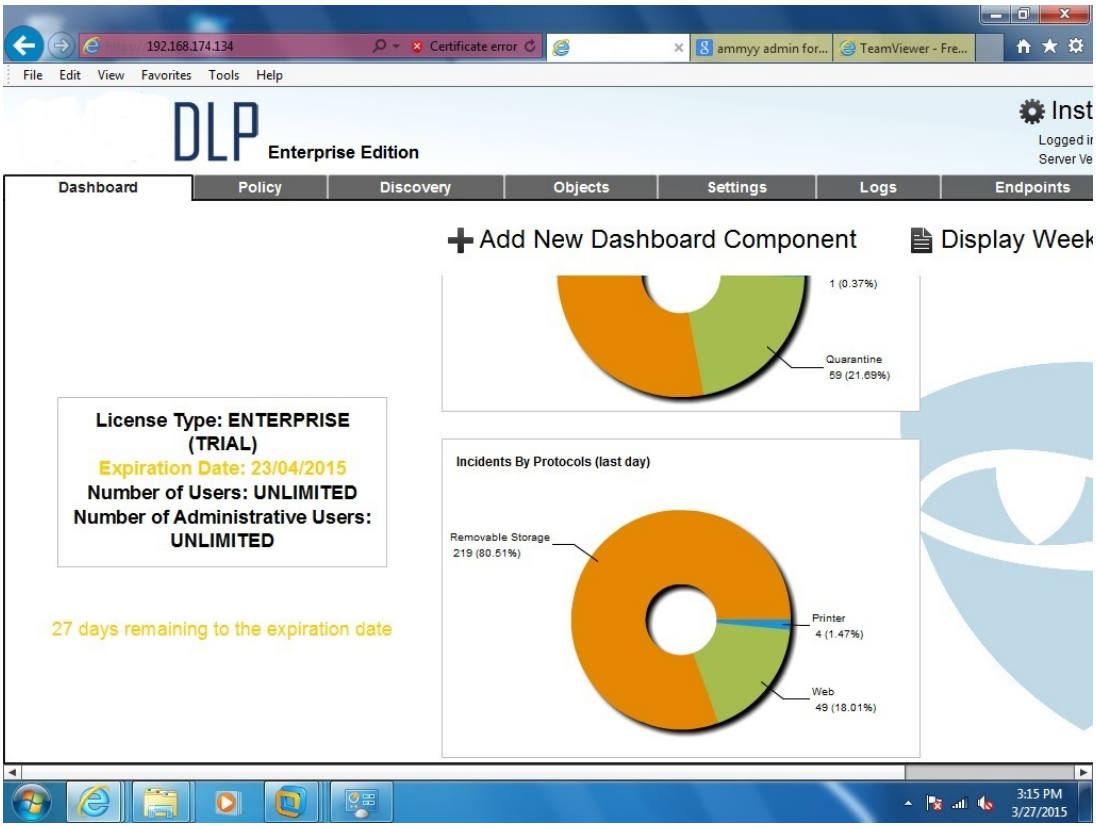

Figure 3.27: P3DLP Server Dashboard.

P3DLP Endpoint protection helps you to detect and to prevent any data moved to removable devices such as USB flash from workstations or laptops at your company. Also it covers any document printed using network or local printers connected to PC's and enables you to enforce policy on stored data on PC's in network.

3.3.3.2 P3DLP Test Cases

The same test cases are applied to all DLP product as mention in section 3.2. The DLP solutions can address the same leakage data and use similar protection methods as mentioned in section 3.1.

3.3.3.3 P3DLP Results

This section presents the results of the tests that performed on different test cases mentioned in the previous section (3.2). At first, we created a policy which contains a key-word to define the sensitive data through user-defined a policy as shown in figure 3.28. Any file contains any of these key-words should be blocked. The file should be blocked when copying to removable media, when uploaded through Facebook/Skype/Drop box or when sent through e-mail or printer.

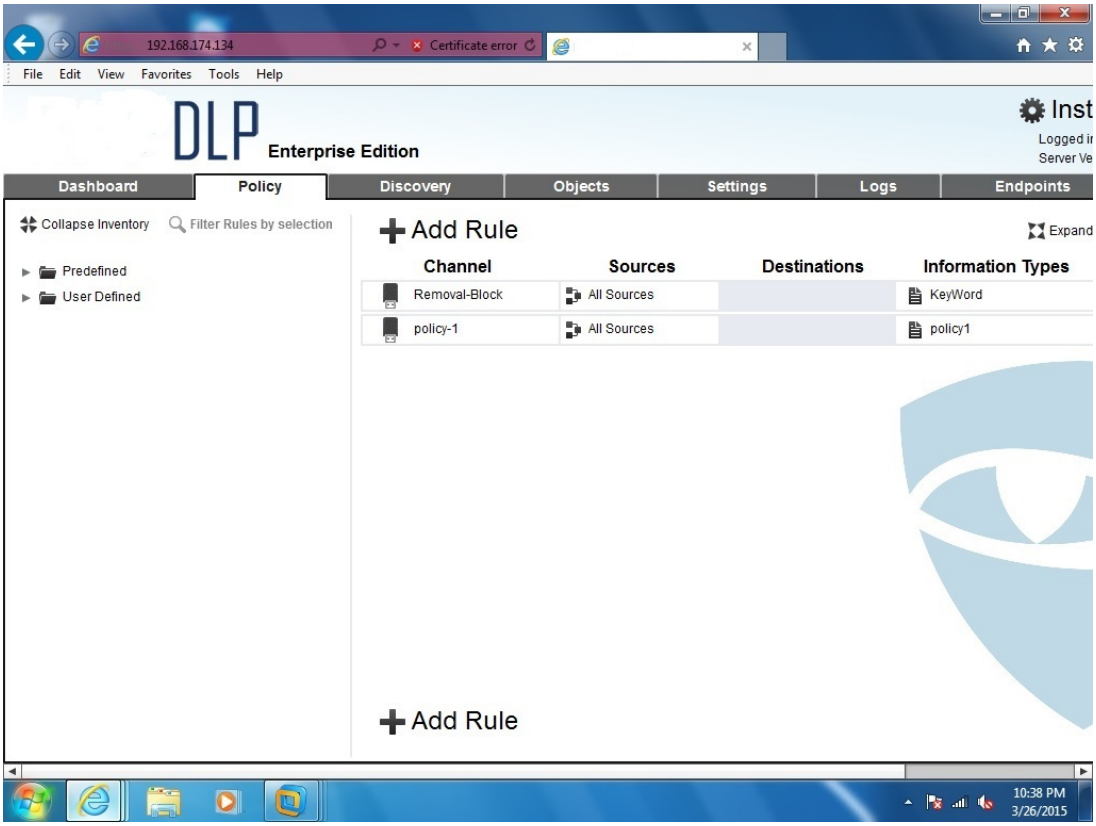

Figure 3.28: Define the Key-Word in P3DLP.

This policy could be configured using the text pattern shown in Figure 3.28 and the rule action should be “block”. The initial test was copying of the text file including the text POLICY to a USB flash. As expected, the text file was blocked (Figure 3.29) and stored in database on the enforce management server. After this test with correct functionality of the system, additional tests about the key-word matching were performed.

P3DLP system handles the test cases that contain many file types that may understand

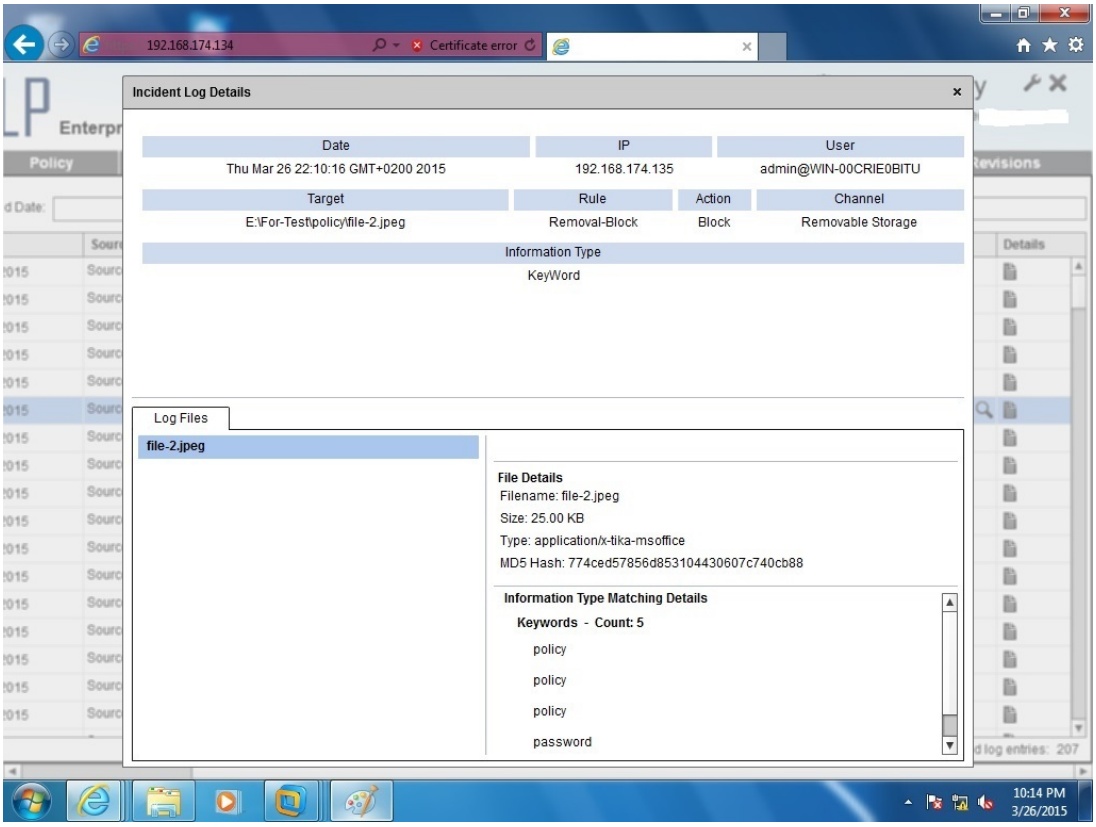

Figure 3.29: When Copying Password File to a USB Flash in P3DLP.

and parse. Therefore the following different file types were generated:

- A file including the any phrase: and sent through, USB memory, Email, Facebook, Dropbox and Skype.
- A file including the phrase: “Password”, “Policy”, “credit-card”, and “Salary” and sent through, USB memory, Email, Facebook and Skype, as shown in figure 3.30 and figure 31.
- The same file is compressed into Formats: .zip, .7z and .rar, and sent through USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.31.
- The same file is renamed as video file format and sent through, USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.32.
- The same file is renamed as picture file format and sent through USB memory, Email, Facebook and Skype, as shown in figure 3.33.

- Put secret words in different file types such as .PDF, .XLS, and .PPT as shown in figure 3.34.

- The same file is opened and sent through network printer as shown in figure 3.35.

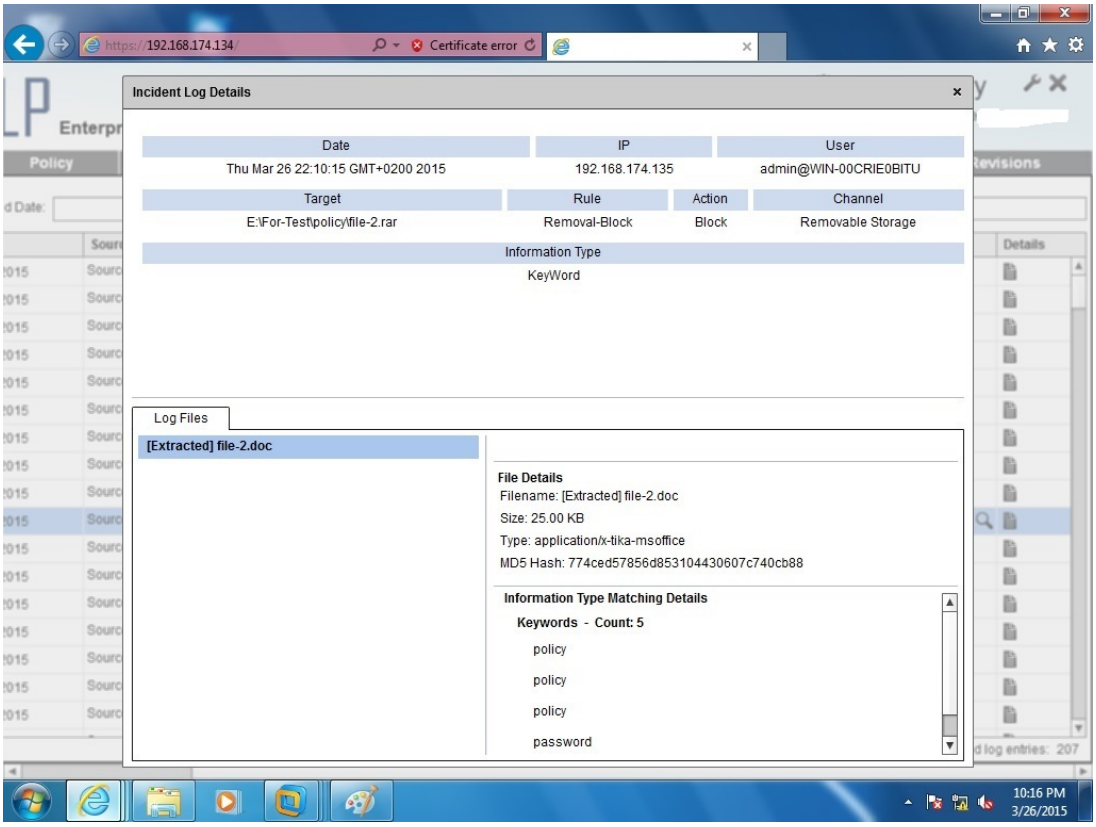

Figure 3.30: The Password File in .rar Format Sent to USB Flashes in P3DLP

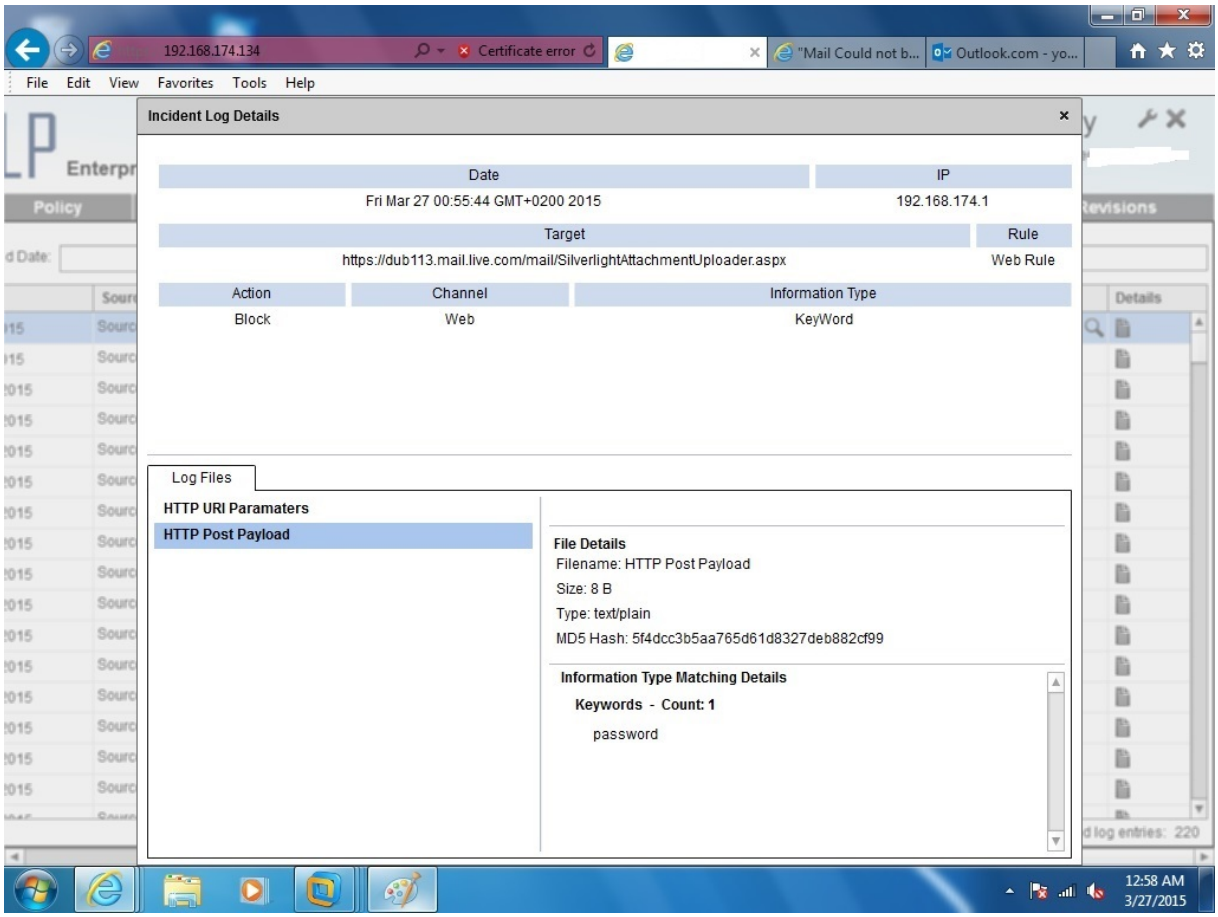

Figure 3.31: The Password File in .zip Format Sent Through Email in P3DLP

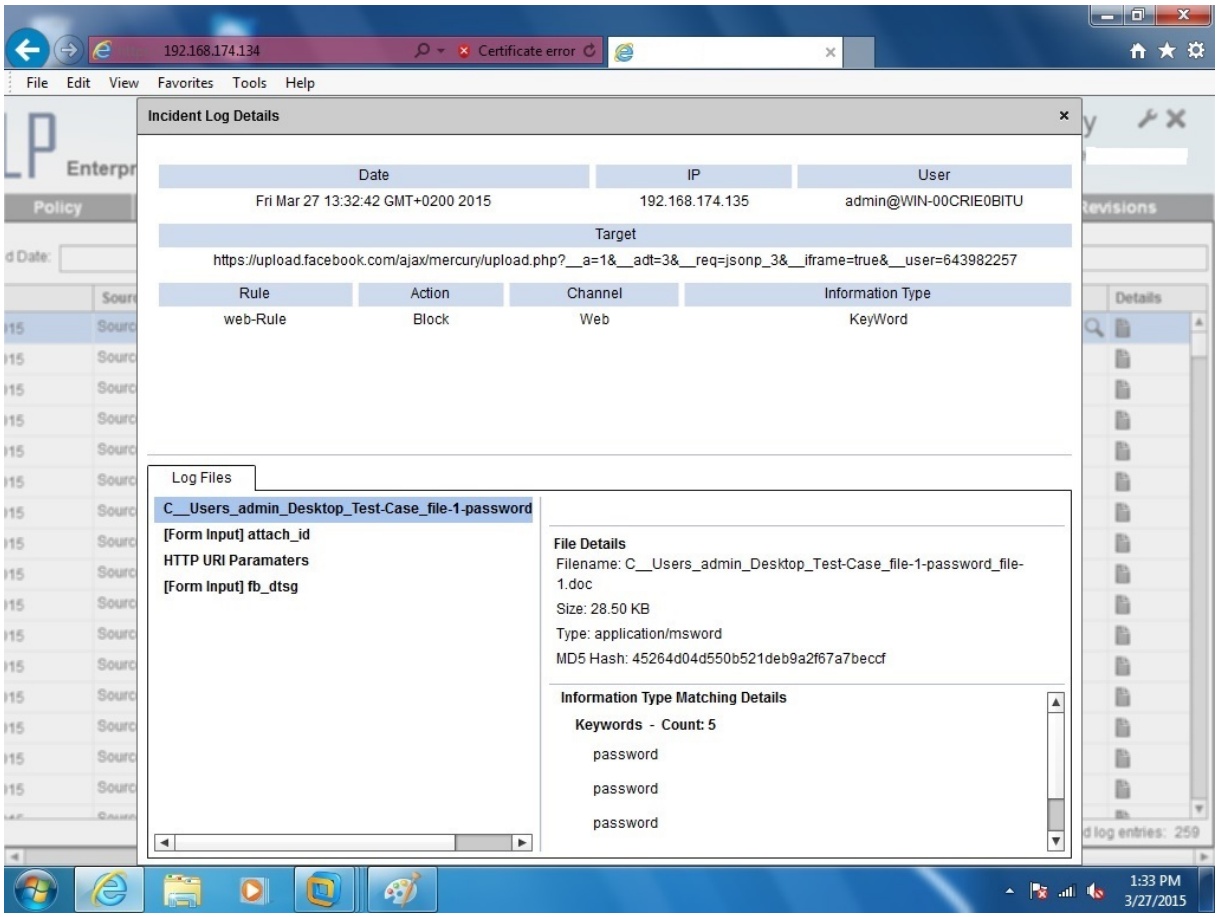

Figure 3.32: The Password File Sent Through Facebook in P3DLP

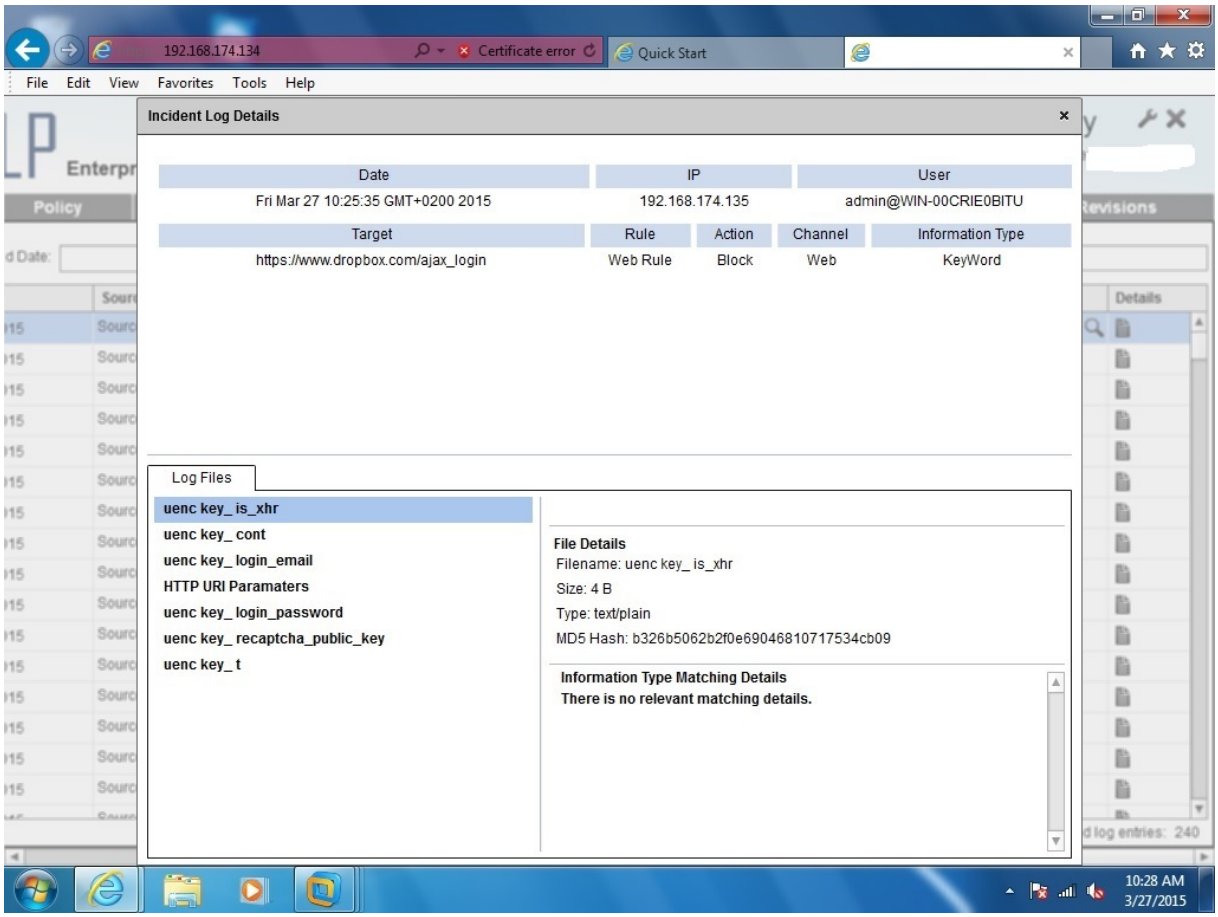

Figure 3.33: The Password File Sent Through Dropbox in P3DLP

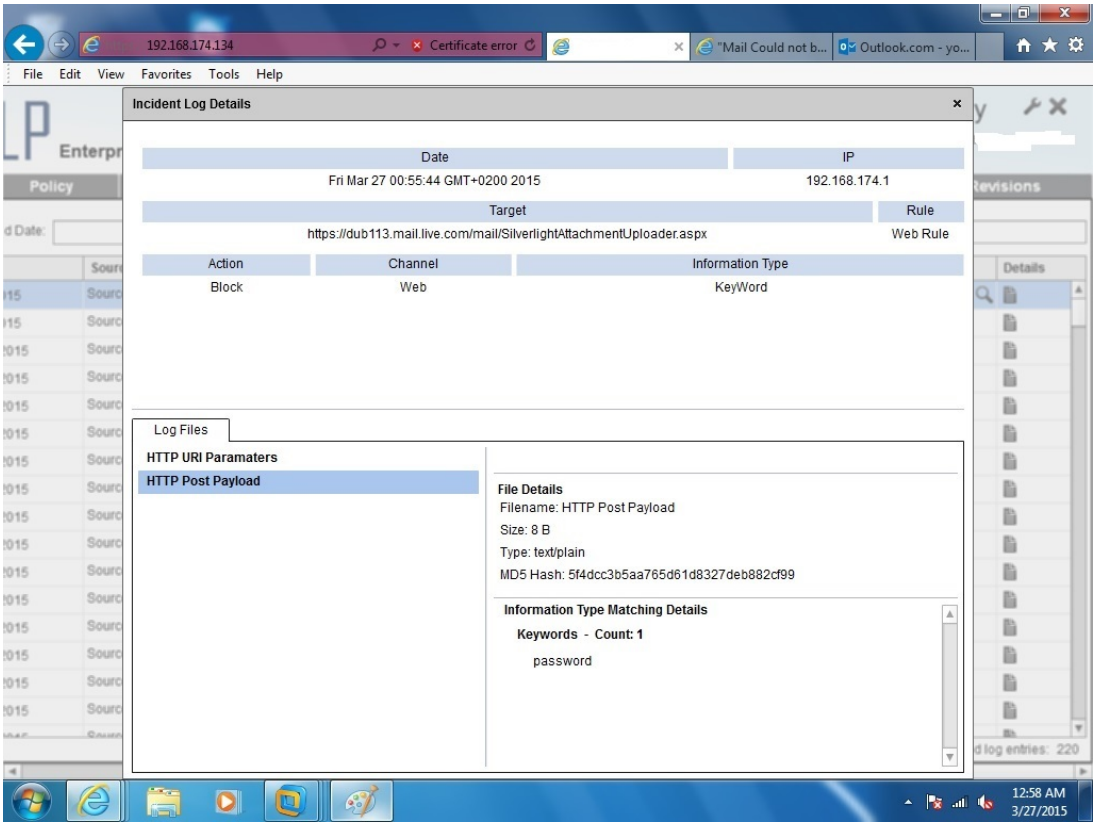

Figure 3.34: The Password File Sent Through Email in Different Formats in P3DLP

P3DLP Disable Agent

The user can uninstall the P3DLP agent from workstation as shown in figure 3.36.

3.3.3.4 P3DLP Leakage Cases

To check whether P3DLP system can find and analyse the content data in case hidden. The test was done by using the same word file by hiding this file into image by using stenography technique. P3DLP didn't detect these hidden data and sent it to any destination as shown in figure 3.37.

- The file is encrypted, compressed and sent through, USB memory, email, Facebook, Email and Skype. These files are not blocked in P3DLP system and successfully sent.

- P3DLP Agent does not work in safe mode as shown in figure 3.38.

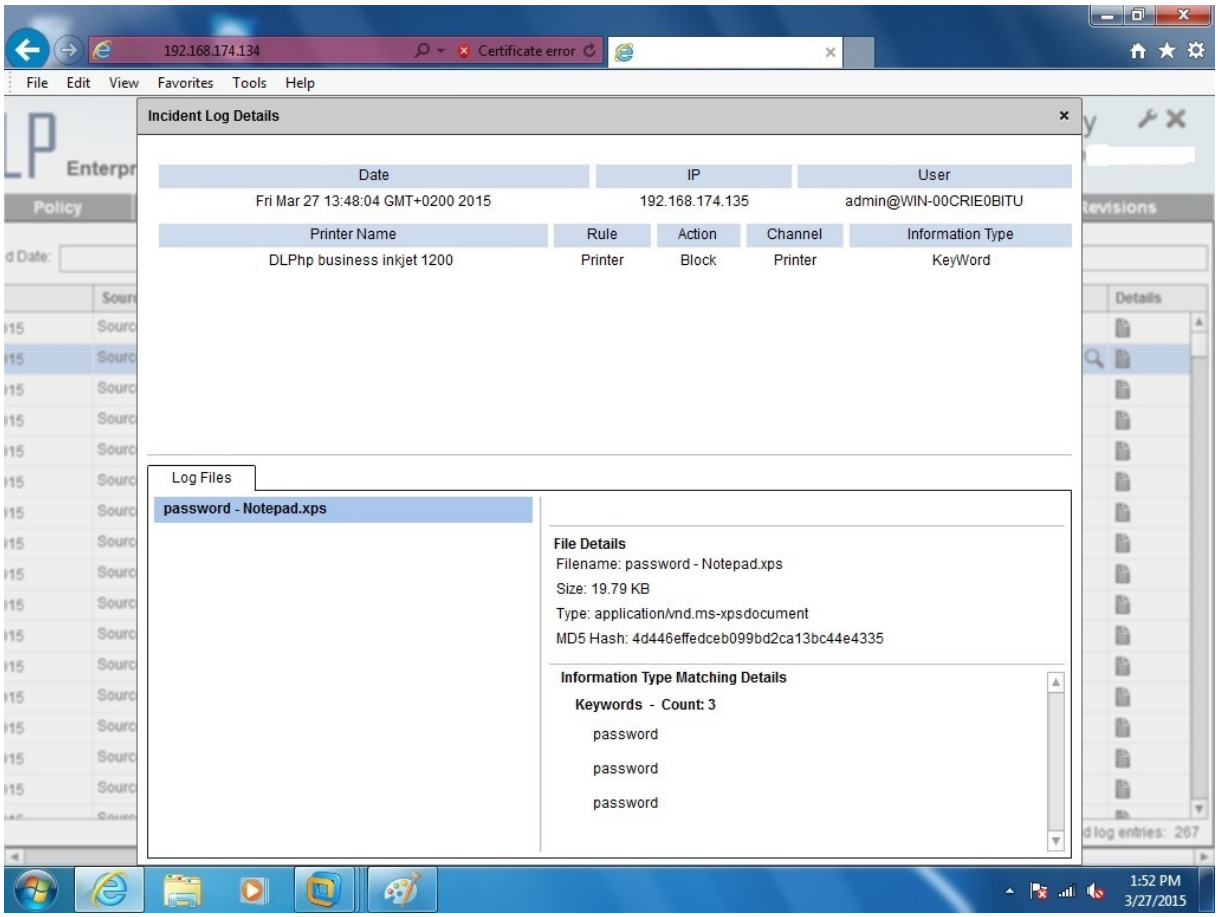

Figure 3.35: The Password File Sent to Printer in P3DLP

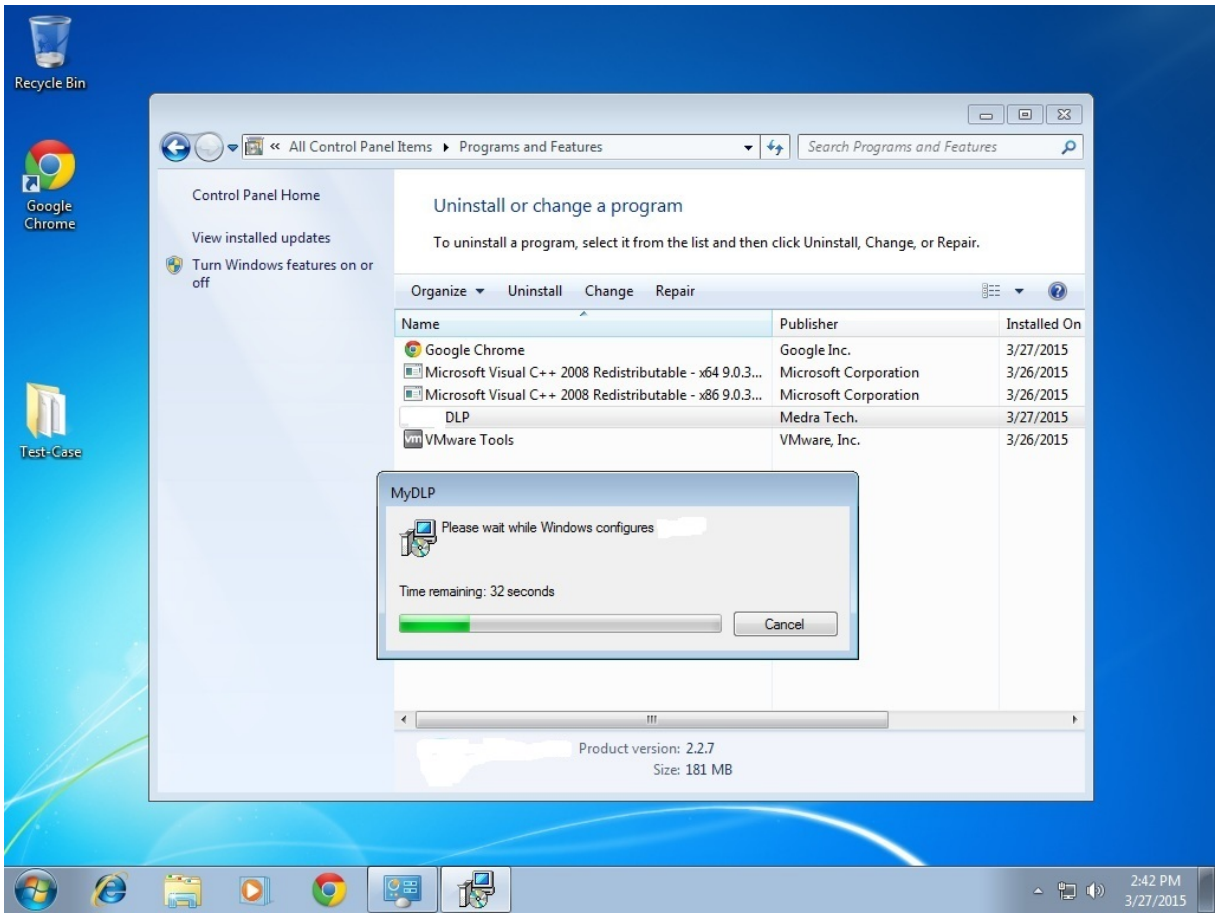

Figure 3.36: Uninstall P3DLP Agent

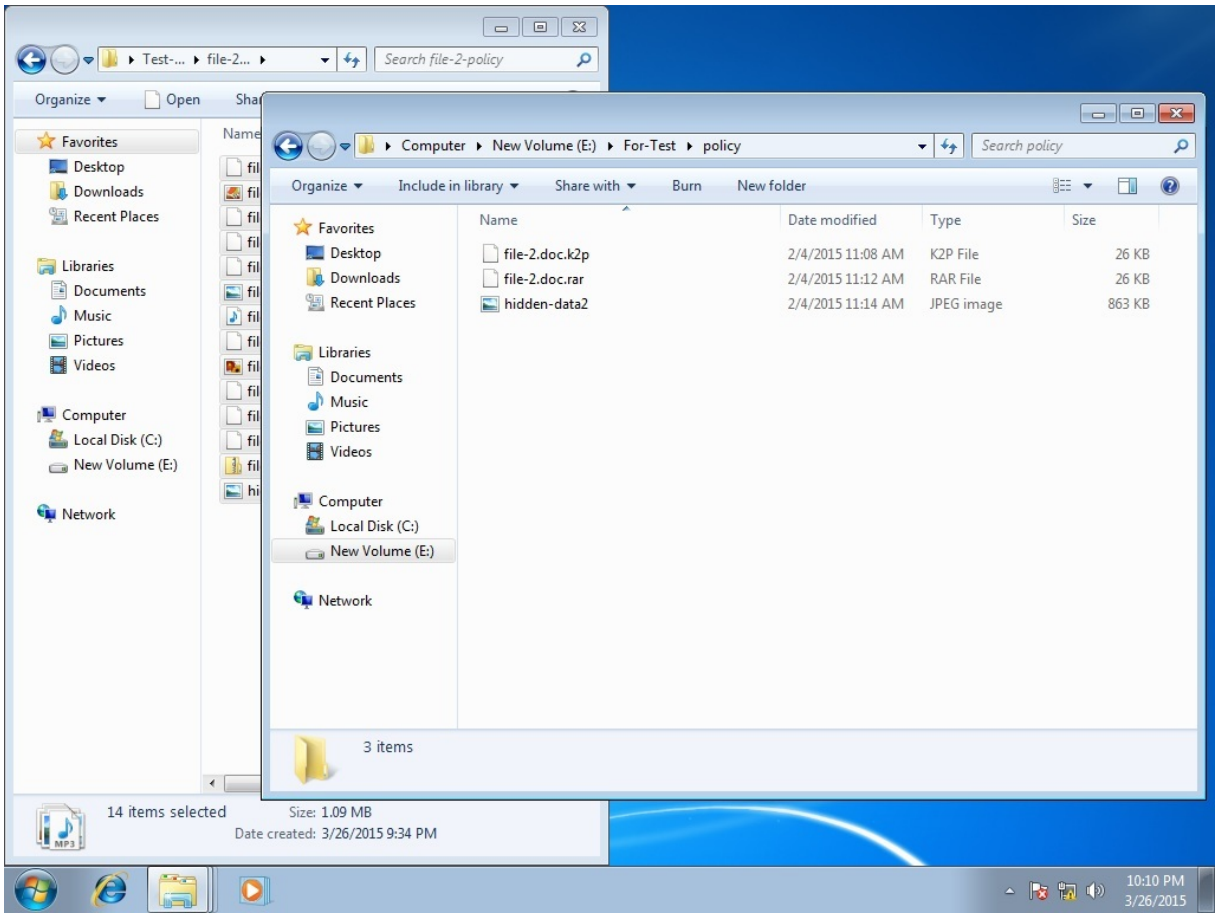

Figure 3.37: Hidden Data was Sent to USB Flash successfully in P3DLP

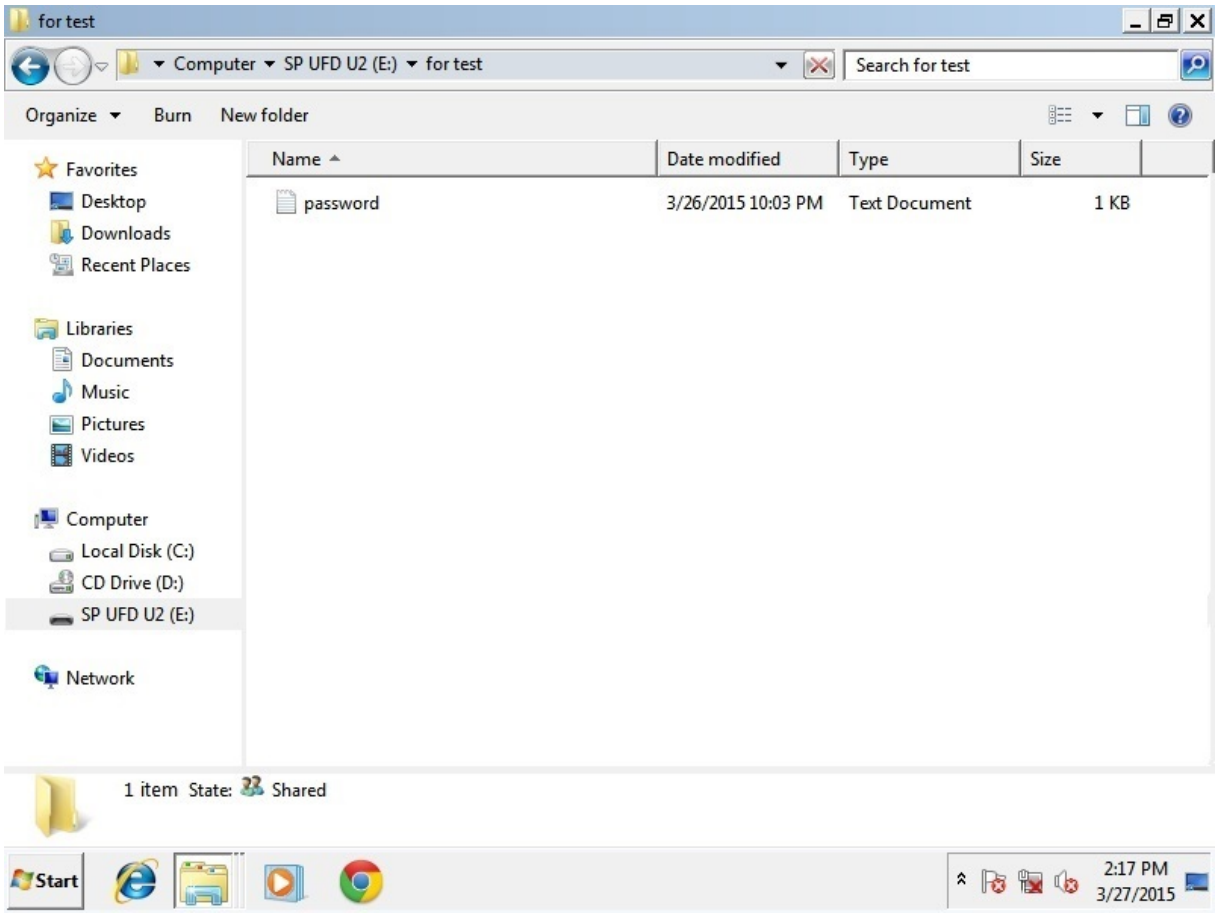

Figure 3.38: P3DLP Doesn't Work in Safe Mode

The following Table 3.6 shows the summarized result of P3DLP tests.

Table 3.6: Summarized Result of P3DLP Tests

Description	**File Type**	**P3DLP**
Phase 1		
File is send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P
File with sensitive data attached/send.	.TXT,.Doc	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	P
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	P
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	P
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	P
File Hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .pdf, .ppt	P
The same file is opened and sent through printer.	Local/network	P
Phase 2		
The DLP agent is working safe mode.		No
Can stop DLP agent		Yes
Phase 3		
Secure communication between Central DLP server and endpoint		Yes

3.3.4 Product4 Host Data Loss/Leakage Prevention(P4DLP)

Among the DLP products which happened in 2008 was the product4 Host Data Loss Prevention package. It is currently included in version 4.8 and is available both as a full DLP package, such as network monitoring and information discovery, and also as a endpoint monitoring solution.

3.3.4.1 P4DLP Test Environment

The P4DLP solutions are built to address data in motion, data at rest, and data in use.

The P4DLP solution consists of several models [16]:

1. The P4DLP Monitor and analyzes all content over a network.
2. P4DLP Prevent monitors all email and applies actions to resolve any problems.
3. P4DLP Discover and monitors file systems and servers, finds significant data, and reports data that is in violation of policy.
4. P4DLP Endpoint finds important events occurring at endpoints and reports any policy violations.

Figure 3.39 displays the fundamental connection of the individual components. The main tool for management is the Policy Server. Using ePO 4.8, the DLP administration plug-in combines into the existing anti-virus console [51]. This plug-in enables in determining DLP policies, reaction, global configurations to the endpoints agents and setting these policies for the managed systems. For saving the settings, a Microsoft SQL Server installation is required. A DLP database is created through the installation. the policy, report and notification events are stored in the database [16].

Endpoints that are deployed on user's computers (PCs, laptop, servers) enable administrators to analyze content within a user's working environment and block or monitor policy breaches as defined by the endpoint profiles [49] [51].

These components are all separated services, it is possible to install them on various machines. Dependent on the number of clients, this allows to weight the performance according to the number of clients. In the built test environment, all services, including the database and the domain controller, ran on a single virtual machine as depicted in Figure 3.39.

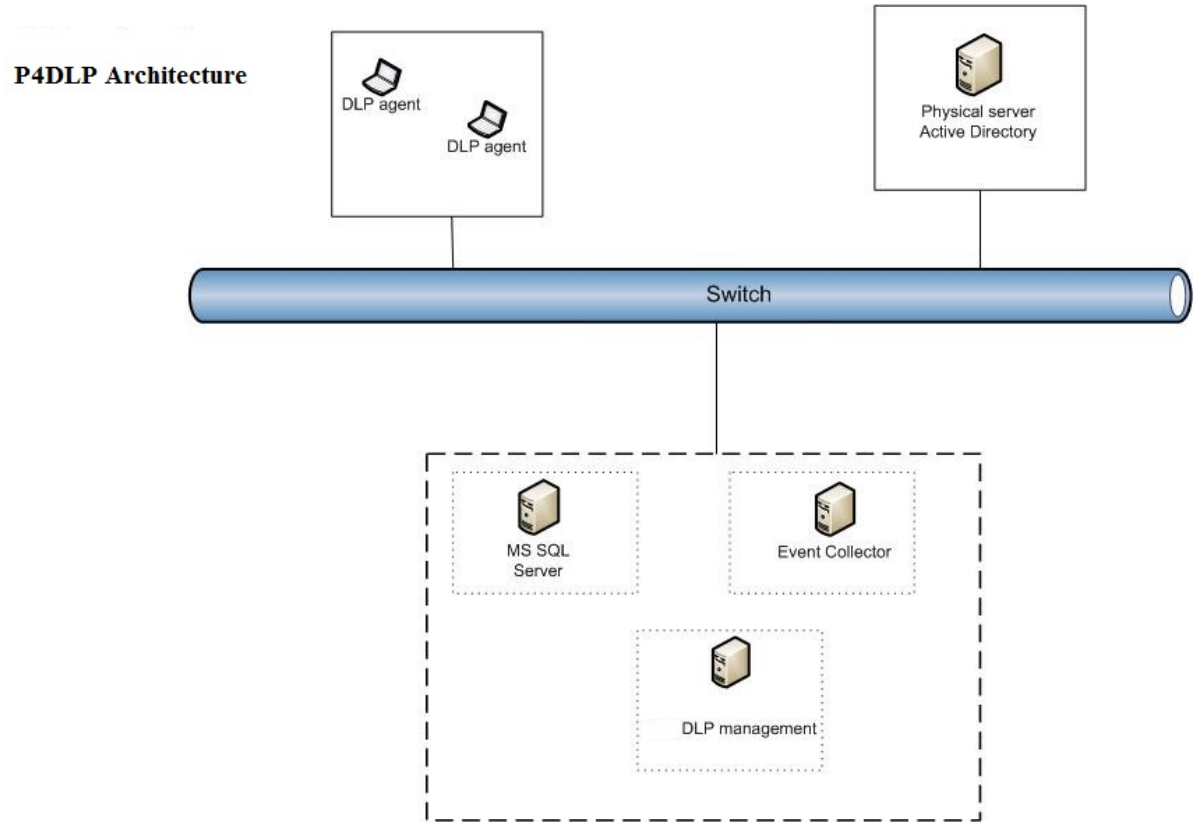

Figure 3.39: Test Environment For P4DLP

By applying this environment, we will be able to monitor every removable storage device which is linked to each of the two client systems. There are many ways to determine sensitive data: regular expressions, key-word matching and file type such as doc, zip, pdf. The defined key-word matching are applied to all data which should be read or written from removable devices and met the specification of the policy rule.

This rule specification can include constraints to specific key-word matching types or file type. Additionally, we'll be able to mark files manually and thus mark special files

which didn't meet any category. In case a breach of policy is detected, a notification is shown to the user, the data is blocked and an alert messages is sent to administrator or manager.

The ePO Server is the main management and analysis component [50]. Additionally it gives you log files, statistics and saved incidents. Figure 3.39 shows the basic interaction between the single components. The P4DLP solution installs its own database server by default. It may be possible to determine an existing database server, but since there is no need for a dedicated installation. ePO views and manages data usage incidents relevant to the active administrator, and creates or manages a network policy or endpoint policy. We can create policies from scratch or by using a predefined regulatory template. Also define the source and destination of the data you want to protect, the endpoint device or application that may be in use, and the remediation or action taken when a violation is discovered (such as block or notify). As show in figure 3.40.

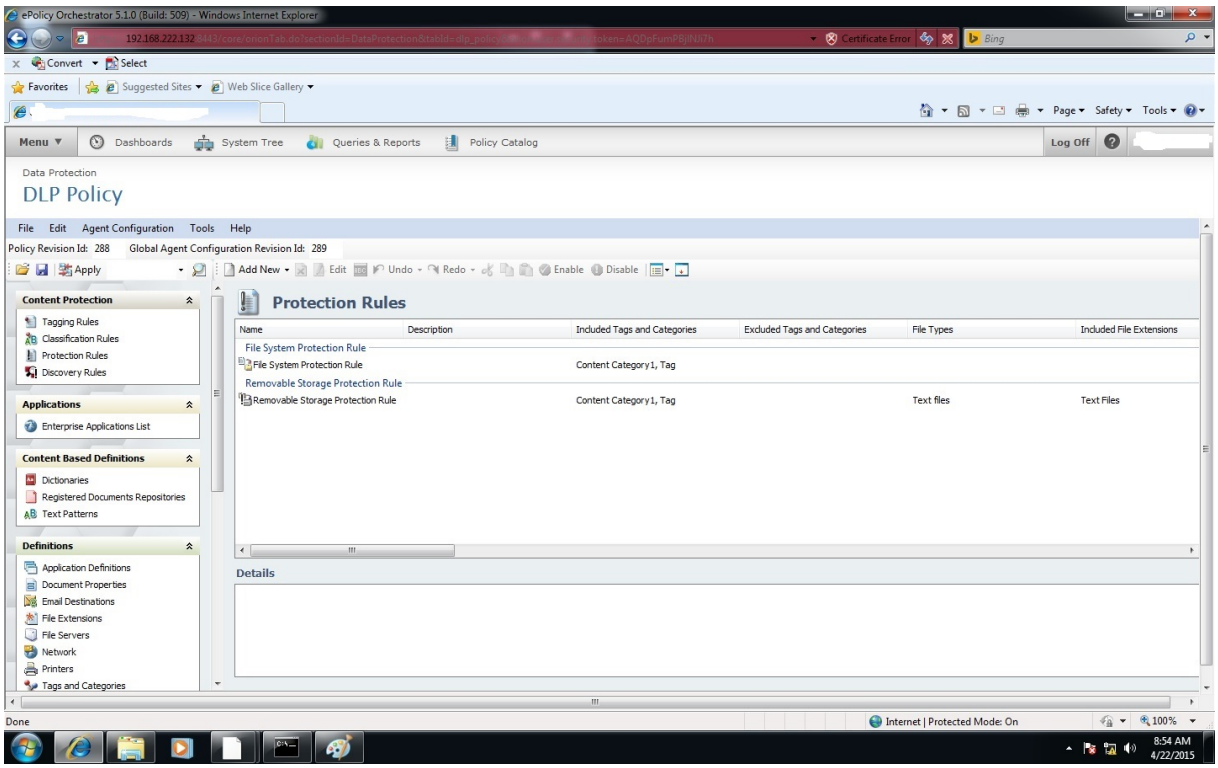

Figure 3.40: ePO Dashboard

3.3.4.2 P4DLP Test Cases

The same test cases are applied to all DLP product as mention in section 3.2. The DLP solutions can address the same leakage data and use similar protection methods as mentioned in section 3.1.

3.3.4.3 P4DLP Results

This section will present the results of the tests that performed on different test cases mentioned in the previous section (3.2). The first step is to create policy which determines the key-word by defining the sensitive data from content classifier as shown in figure 3.42. Any file that contains any of these key-words should be blocked.

The file should be blocked when copied to removable media or when uploaded to Facebook/Skype/Dropbox and when file sent through e-mail or printer.

This policy could be applied using the text pattern shown in Figure 3.41. The reaction rule should be blocked.

The first test was copying of a text file containing the text SALARY to an USB flash. As wasn't to become expected otherwise, the text file was blocked (Figures 3.42 and 3.43) and stored at database from the P5DLP (Figure 3.43). After this test with correct functionality in DLP system, further tests about the key-word matching were performed.

Each DLP solution can handle many file types that may be understood and parsed. Therefore the following mixtures of more complex file types were generated:

- A file containing any word and sent through, USB memory, Facebook, Dropbox and Skype.
- A file containing the sensitive words: " Password" , " Policy", "credit card" , and "salary" and sent through, USB memory, Facebook, Email and Skype, as shown in figure 3.42.
- The same file is compressed in different Formats such as .zip, .7z and .rar. and sent through, USB memory, Facebook, Email, Dropbox and Skype as shown in figure

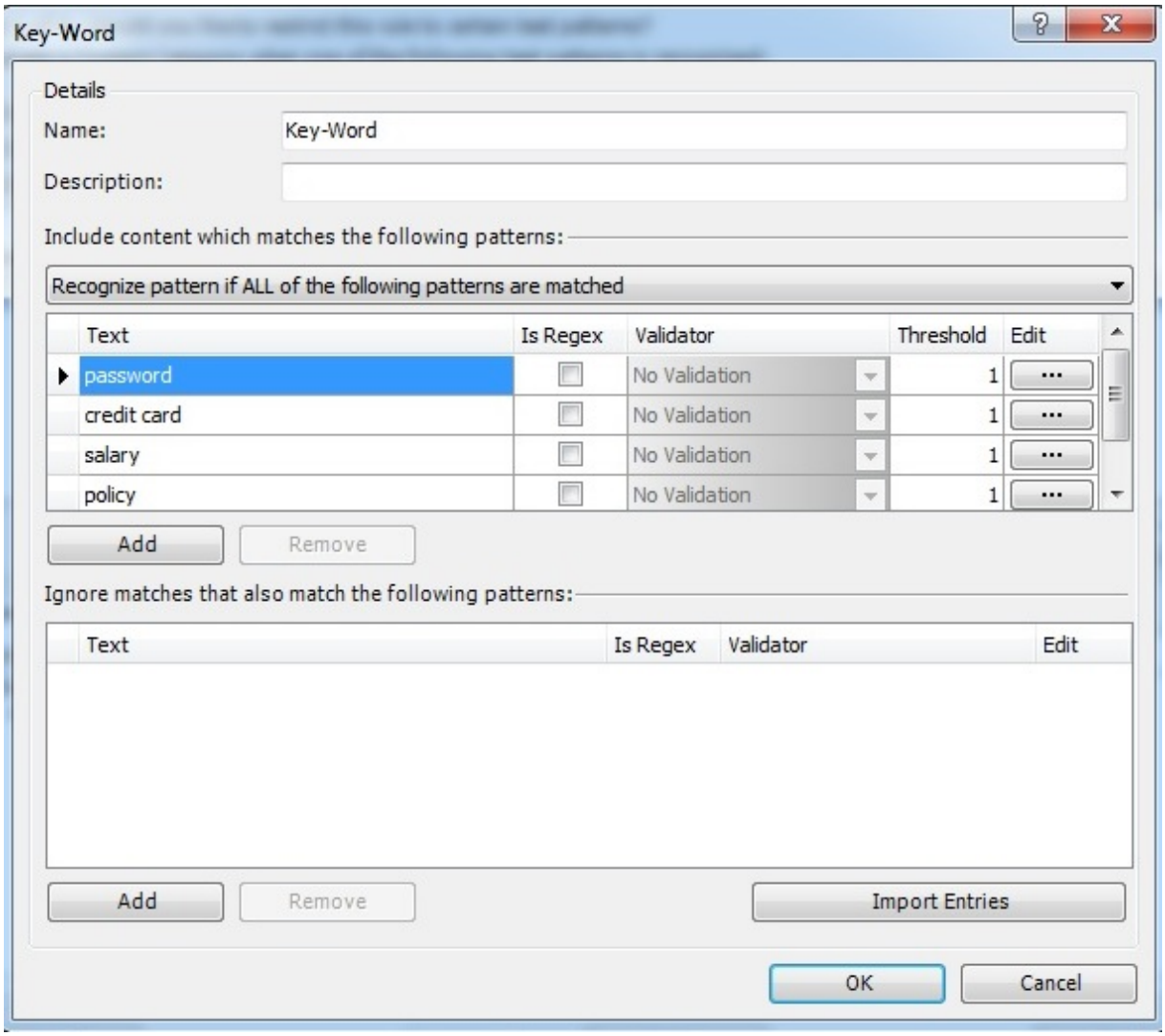

Figure 3.41: Define the Key-Words Matching in P4DLP.

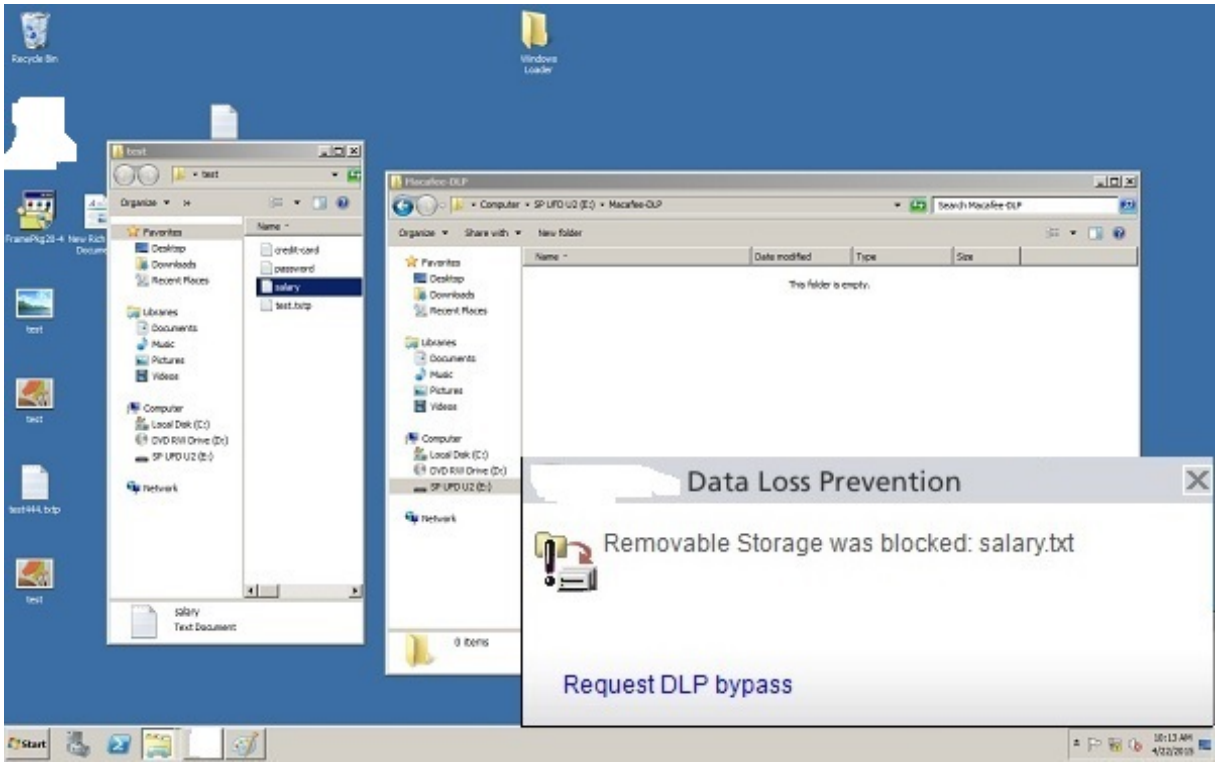

Figure 3.42: The Salary File Sent to a USB Flash in P4DLP.

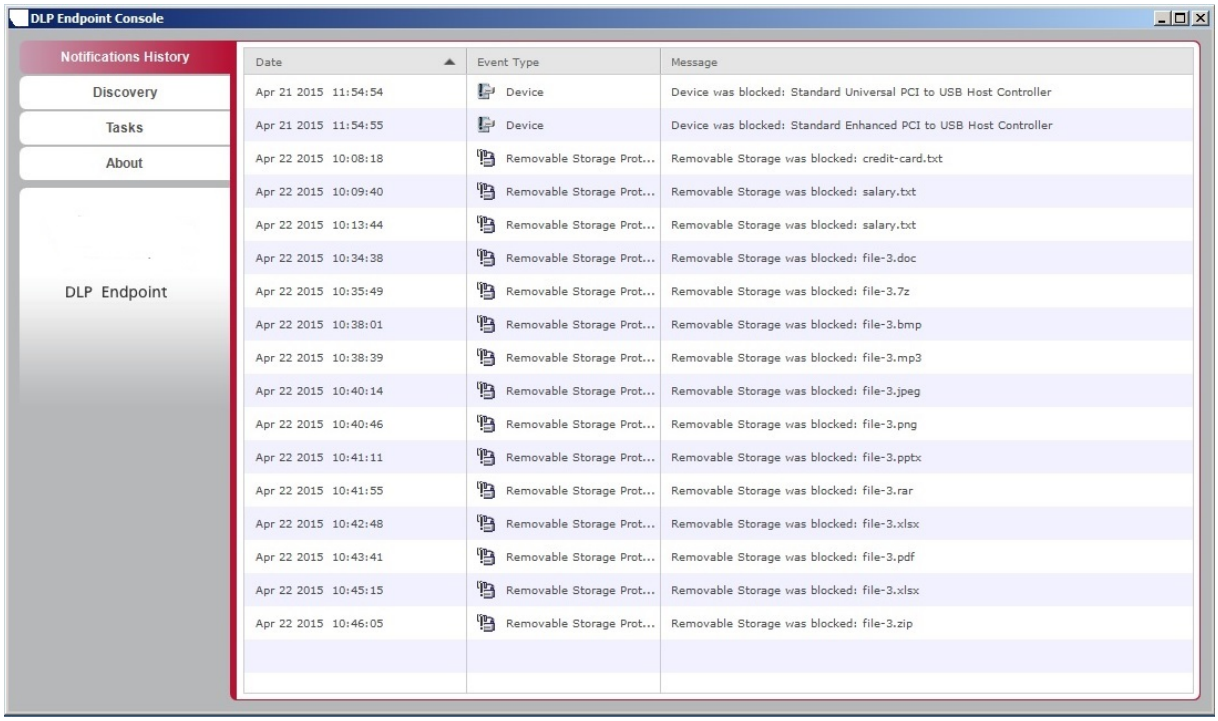

Figure 3.43: The Salary File is Correctly Blocked in P4DLP.

3.44.

- The same file is opened and sent through network printer as shown in figure 3.45.
- The same file is renamed as video file format and sent through, USB memory, Dropbox, Facebook, Email and Skype, as shown in figure 3.46.
- The same file is renamed as picture file format and sent through, USB memory, Email, Facebook, Dropbox and Skype, as shown in figure 3.47.
- Put secret words in different file types such as .PDF, .XLS, and .PPT and sent through, USB memory, Email, Facebook and Skype, as shown in figure 3.48.

Since all these file types are supported, all files were detected to become valuable content and therefore are blocked.

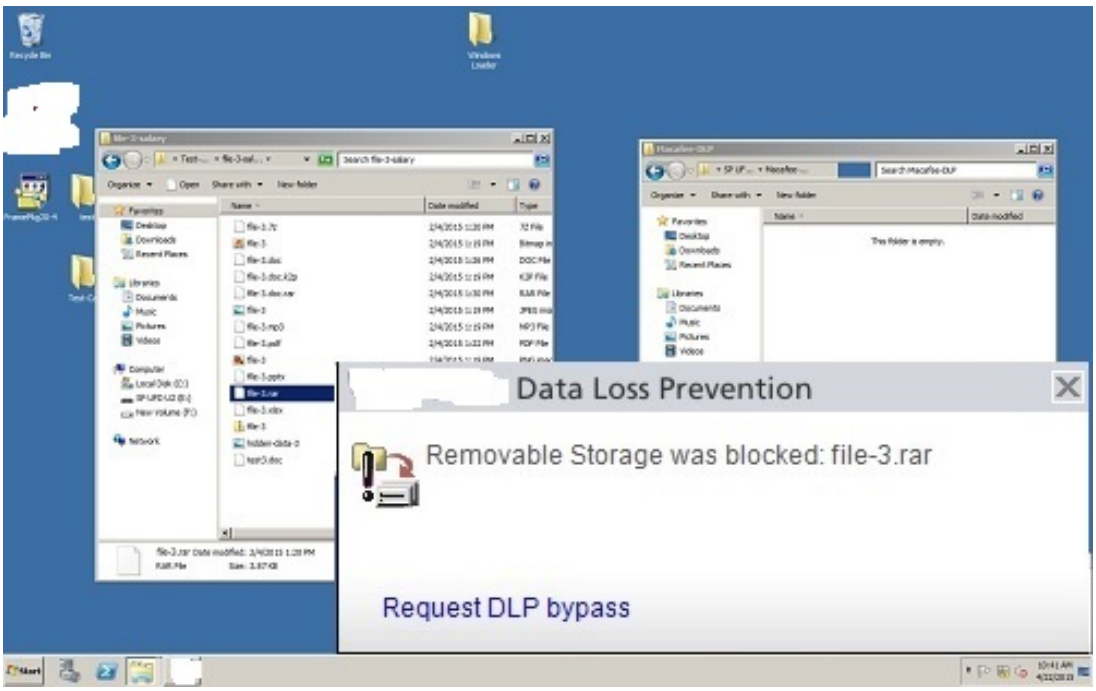

Figure 3.44: The Salary File in .rar Format Sent to USB Flash in P4DLP

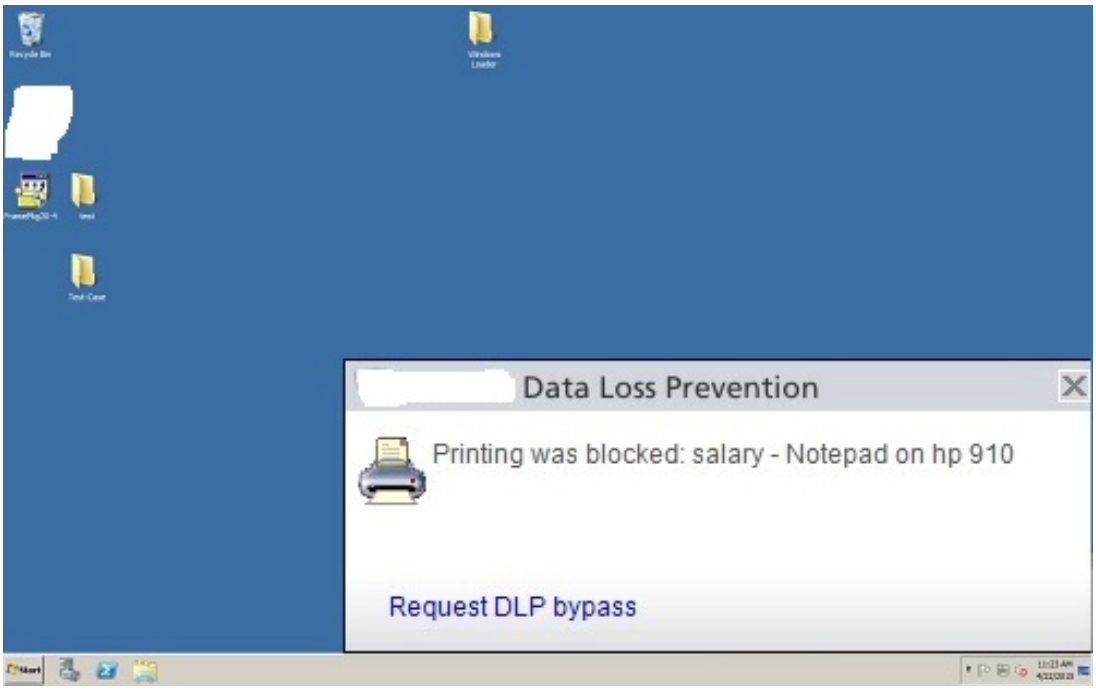

Figure 3.45: The Salary File Sent to Printer in P4DLP

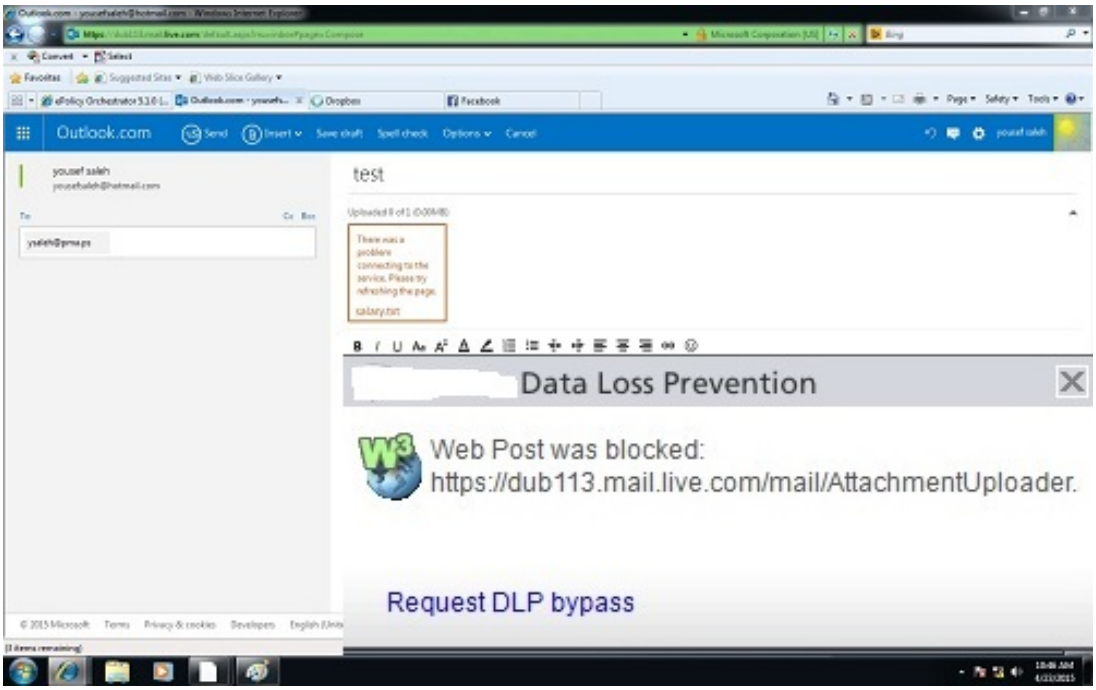

Figure 3.46: The Salary File Sent to Email in P4DLP

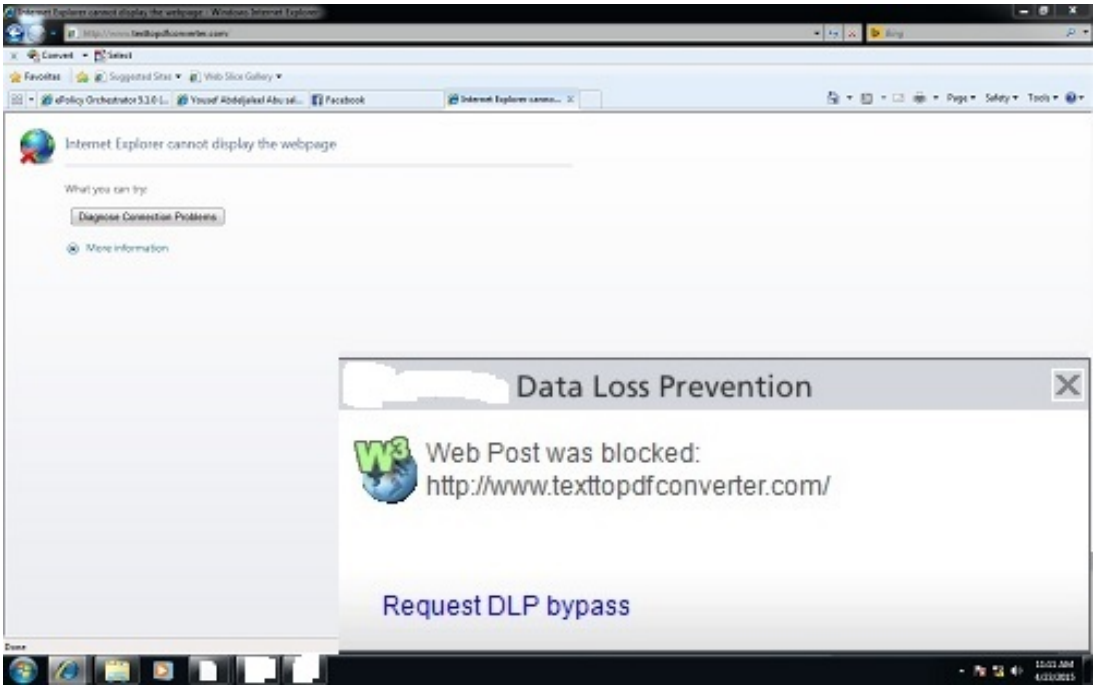

Figure 3.47: The Salary File Uploaded to Internet in P4DLP

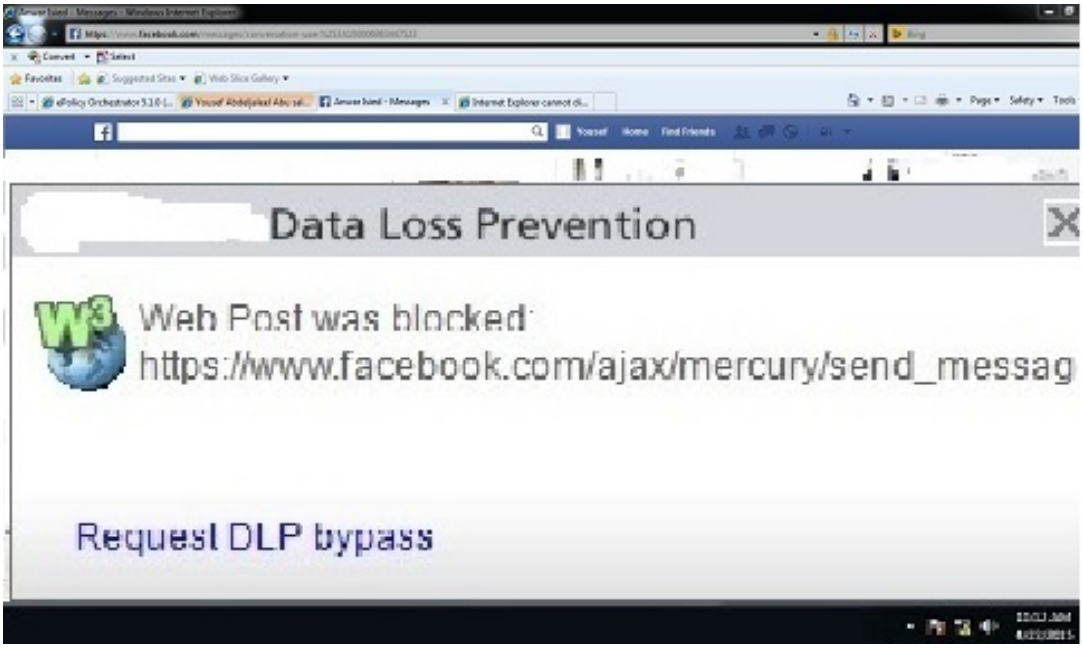

Figure 3.48: The Salary File Uploaded to Facebook in P4DLP

Disabled Agent

The user cannot disable the DLP agent until the super-admin creates a verification code on ePO server and sends it to the user to put this code on end point to disable the agent on the client host as shown in figure 3.49.

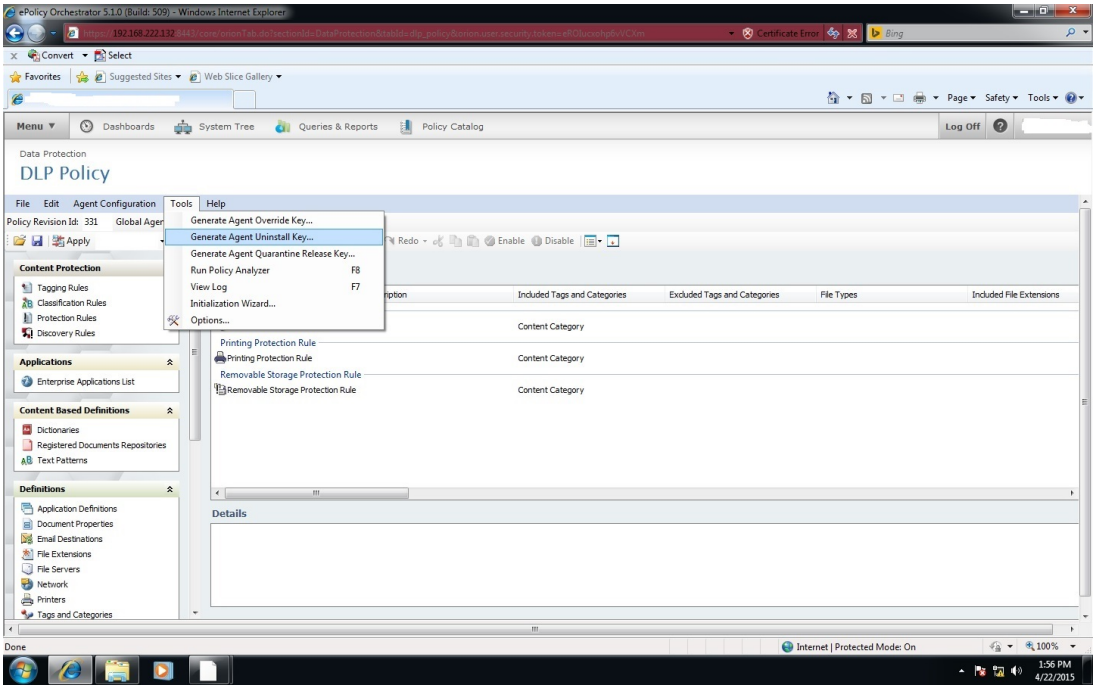

Figure 3.49: Agent Disable in P4DLP

Secure communication between central DLP and endpoint.

The endpoint agent communicates with the server over HTTPS. This means that every communication is encrypted.

3.3.4.4 P4DLP Leakage Cases

To check whether a real deep content analysis is performed, the test was repeated using the same word file by hide this file in an image by using stenography technique.

The sensitive words is put in an image by hiding these data in that image. We use OurSecert tool to hide data in image as show in figure 3.50. But the file was not detected through the discover process. In general, this means that the hidden data types are not

Figure 3.50: Steganography Hidden Secret Data in Image

discovered in P4DLP as shown in figure 3.51. The image that contains a sensitive data was successfully sent to flash and e-mail.

- In case when file contains sensitive encrypted data and sent through, USB memory, email, Facebook and Skype. (It was stored within a Kruptos file format).
- The same file is encrypted, compressed and sent through, USB memory, email, Facebook, Dropbox and Skype.
- P4DLP Agent does not work in safe mode as shown in figure 3.52.

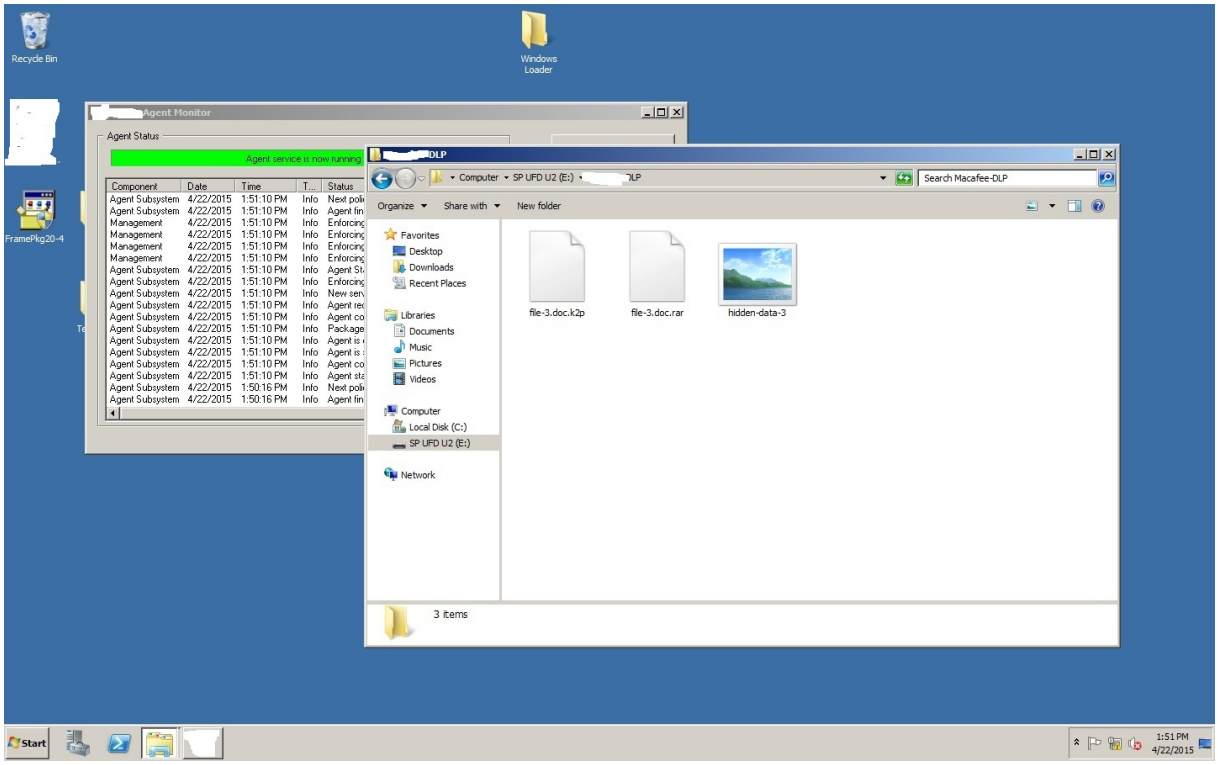

Figure 3.51: Send Hidden-in-Image Data Through USB

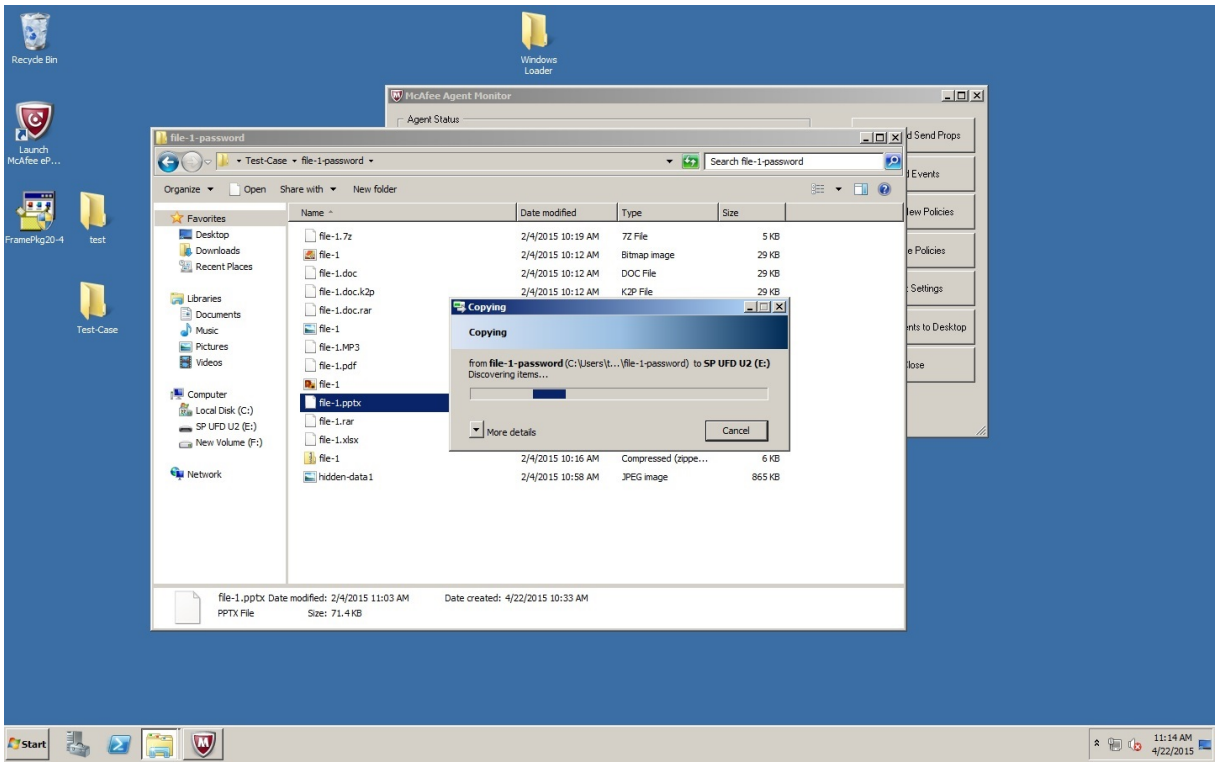

Figure 3.52: P4DLP Agent in Safe Mode

Since all these cases are not supported, all these files were not detected and so not blocked in P4DLP.

The following Table 3.7 shows the summarized result of P4DLP tests.

Table 3.7: Summarized Result of P4DLP Tests

Description	**File Type**	**P4DLP**
Phase 1		
File is send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P
File with sensitive data attached/send.	.TXT,.Doc	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	P
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	F
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	P
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	P
File Hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .pdf , .ppt	P
The same file is opened and sent through printer.	Local/network	P
Phase 2		
The DLP agent is working safe mode.		No
Can stop DLP agent		No
Phase 3		
Secure communication between Central DLP server and endpoint		Yes

3.3.5 Product5 Data Loss/Leakage Prevention (P5DLP)

The P5DLP product enables to stop confidential data from leaving the network.

3.3.5.1 P5DLP Test Environment

P5DLP inspects network traffic for data patterns you identify. Administrator specifies whatever patterns the P5DLP appliances to look for in network traffic. The DLP feature is divided in to a number of parts.

1. DLP Sensor : A DLP sensor is a package of filters. Filters can analyze traffic for known files using DLP fingerprints, for files of a specific type such as (.txt .pdf .doc .xls .ppt .docx .pptx) or name, files bigger than a particular size, data matching a specified regular expression [20].

Figure 3.53 displays the connection of the individual components. The main tool for management is the P5DLP appliances. This appliance is hardware component enables in determining DLP policies, reaction.

This solution inspects network traffic when moving through the P5DLP appliances only and not provide endpoint solutions. This solution consider agent-less and didn't detect any leakage occurs in workstation, Laptops and workstations in user and server side such as when copying to USB/CD/DVD, Printer.

3.3.5.2 P5DLP Test Cases

The same test cases are applied to all DLP product as mention in section 3.2. The DLP solutions can address the same leakage data and use similar protection methods as mentioned in section 3.1.

3.3.5.3 P5DLP Results

This section will present the results of the tests that performed on different test cases mentioned in the previous section (3.2).

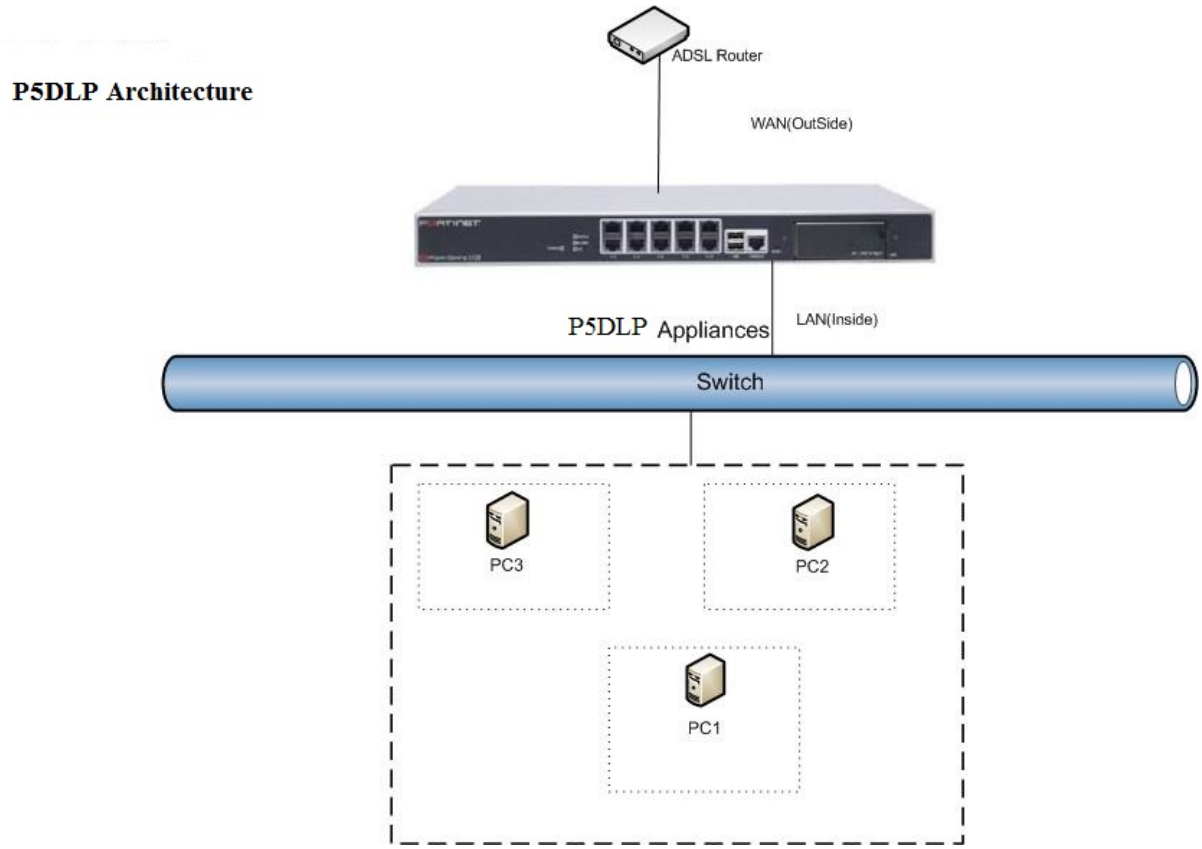

Figure 3.53: Test Environment For P5DLP

The first step is to create policy which determines the key-word by defining the sensitive data from DLP Leak Prevention as shown in figure 3.54. Any file that contains any of these key-words should be blocked.

The file should be blocked when copied to removable media or when uploaded to Facebook/Skype/Dropbox and when file sent through e-mail or printer.

This policy could be applied using the text pattern shown in Figure 3.55. The reaction rule should be blocked.

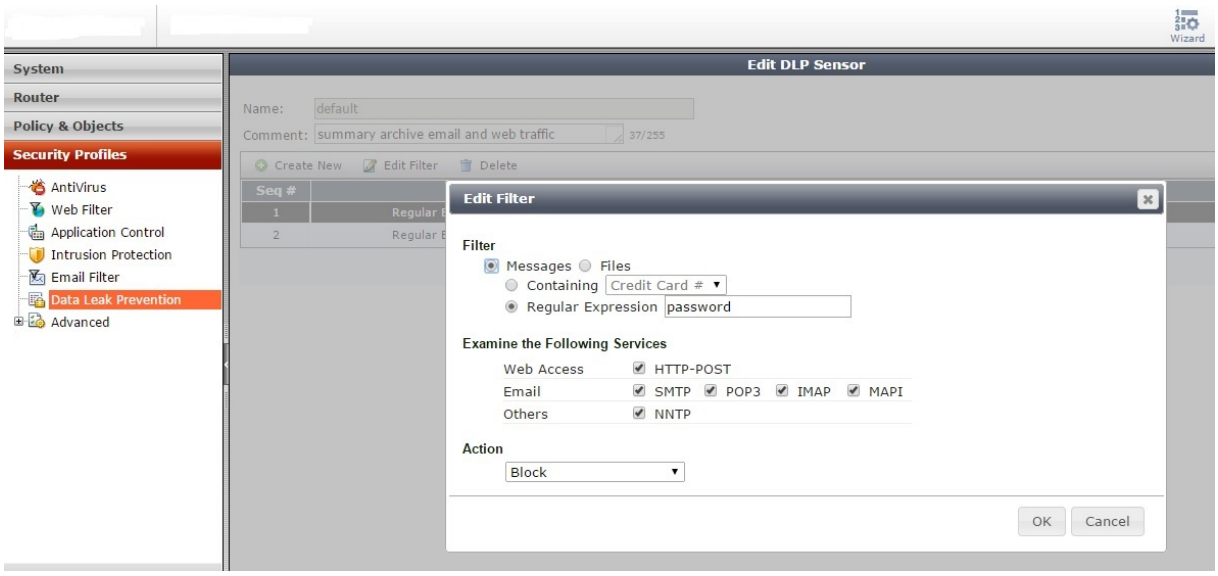

Figure 3.54: P5DLP Data Leak Prevention Dashboard

The first test was copying of a text file containing the text Password to an USB flash. As wasn't to become expected otherwise, the text file was sent successfully.

Each DLP solution can handle many file types that may be understood and parsed. Therefore the following different of more complex file types were generated:

- A file containing any word and sent through, USB memory, Facebook, Email and Skype.
- A file containing the sensitive word: " Password" , and sent through, Facebook, Email, Dropbox and Skype. as shown in figure 3.55.

Since all these file types are supported, all files were detected to become valuable content and therefore are blocked.

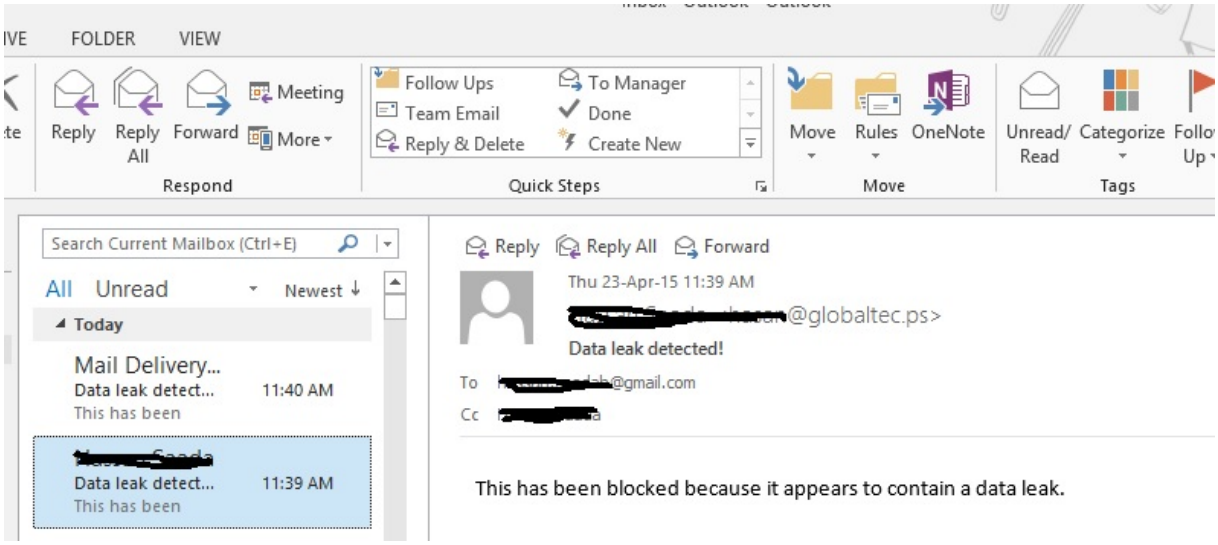

Figure 3.55: The Password File Sent to Email in P5DLP

Secure communication between central DLP and endpoint

The communicates with the P5DLP appliances over HTTPS. Which means that every communication is encrypted.

3.3.5.4 P5DLP Leakage Cases

To check whether a real deep content analysis is performed, the test was repeated using the same word file by hide this file in an image by using stenography technique. The sensitive words is put in an image by hiding these data in that image. We use OurSecert tool to hide data in image. But the file was not detected through the discover process. In general, this means that the hidden data types are not discovered in P5DLP appliances DLP. The image that contains a sensitive data was successfully sent to Email.

- In case when file contains sensitive encrypted data and sent through, USB memory, Email, Facebook, Dropbox and Skype. (It was stored within a Kruptos file container).

- The same file is encrypted, compressed and sent through, USB memory, Email, Facebook, Dropbox and Skype.

- The same file is compressed in Formats: .zip, .7z and .rar. and sent through,

Facebook, Email, Dropbox and Skype.

- The same file is renamed as video file format and sent through, Email, Facebook, Dropbox and Skype.
- The same file is renamed as picture file format and sent through, Email, Facebook, Dropbox and Skype.
- Put secret words in different file types such as .PDF, .XLS, and .PPT and sent through, Email, Facebook, Dropbox and Skype.

Since all these cases are not supported, all these files were not detected and so not blocked in P5DLP appliances DLP. The following Table 3.8 shows the summarized result of P5DLP tests.

Table 3.8: Summarized Result of P5DLP Tests

Description	**File Type**	**P5DLP**
Phase 1		
File is send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P
File with sensitive data attached/send.	.TXT,.Doc	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	F
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	F
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	F
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	F
File Hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .pdf, .ppt	F
The same file is opened and sent through printer.	Local/network	F
Phase 2		
The DLP agent is working safe mode.		No
Can stop DLP agent		No
Phase 3		
Secure communication between Central DLP server and endpoint		Yes

Chapter 4

Results and Discussions

This chapter presents the results of tests that performed to five DLP system. Also presents the weaknesses discovered, and suggest to solve some of weaknesses.

4.1 Results

As discovered, in chapters 4, it is shown that DLP solutions don't yet fully prevent data loss/leakage. Web sense, OpenDLP, Symantec, Trustwave, MacAfee, RSA, MyDLP and FortiNet systems are some of the popular solutions in DLP industry.

There were clearly several possibilities for loss /leakage (e.g. copying sensitive data to USB flash memory in safe mode) in Product1, Product2, Product3, Product4 and Product5 DLP systems. It would be risky to depend on these systems in security concept.

These results are summarized in Table 4.1 to provide a quick review around the abilities of the mentioned systems. When a solution passed a test, a check mark passed (P) is used; otherwise a failed (F) mark is used. A not available (NA) mark when the system doesn't support.

Table 4.1: Completed Test Results for P1DLP, P2DLP, P3DLP, P4DLP and P5DLP

Description	**File Type**	**P1**	**P2**	**P3**	**P4**	**P5**
Phase 1						
File is send/Attached through						
USB, Facebook, E-mail, Dropbox, Skype.	.TXT,.Doc	P	P	P	P	P
File with sensitive data attached/send.	.TXT,.Doc	P	P	P	P	P
File is compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.ZIP, .RAR,.7Z	P	P	P	P	F
File encrypted with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Kruptos	F	F	P	F	F
File encrypted and compressed with sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.		F	F	F	F	F
File is put in video file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.mp3	P	P	P	P	F
File is put in picture file format send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Jpg,png,.bmp	P	P	P	P	F
File hiding sensitive data send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	Steganography techniques.	F	F	F	F	F
Put sensitive data in different file types send/Attached through USB, Facebook, E-mail, Dropbox, Skype.	.doc , .xls, .df, .ppt	P	P	P	P	F
The same file is opened and sent through printer.	Local / network	P	P	P	P	F
Phase 2						
The DLP agent is working safe mode.		No	No	No	No	No
Can stop DLP agent		No	No	Yes	No	No
Phase 3						
Secure communication between Central DLP server and endpoint		Yes	Yes	Yes	Yes	Yes

4.2 Discussion

The results show that DLP systems will not be able to completely secure confidential data. As long as there are ways of hiding data, DLP systems will not discover it.

The outcomes of our tests results enable us to get general image of different conclusions. First one, we are able to abstract weaknesses and issues in DLP solutions that have common problems of DLP approach that will be described in Section 4.2.1.

Then the second step that we will suggest solutions to these problems was described in chapter 5 and finally, we will develop a proposed solution and do a comparison between the proposed solutions based on processing time and image size.

4.2.1 Common Problems

DLP architecture can be useful target to attackers. A lot of valuable information is saved on incident database. Descriptions of most secret patterns and confidential data could be found on DLP server. Therefore, all of the DLP systems need to proper protect the mass of sensitive data.

Keeping database for common patterns located on sensitive files and documents, and in addition to hashes related to sensitive files, is dangerous. Sometimes copies of sensitive data are stored in the DLP management server.

DLP endpoints are managed through the central server through polices and rules updates. When an attacker accesses one management server, he may obtain information of how to effectively identify sensitive files and its locations, and so, if he succeeded, he can access directly to all sensitive data and control of all security procedure in the company or organization network.

In addition to that, attacker can change policy definitions, administer endpoints and so deleting critical system files, or learn how these sensitive files being transferred to the server or remote locations. By controlling the software update procedure, attacker possibly could hack the endpoint software update source to malware source such as key

loggers, Trojans, remote shell.

DLP system infrastructure should be hard guarded to face these threats. If poor passwords are used, the passwords could be brute forced hacked and an attacker could log in to remove all events. To reduce these risks, strong and complex password with suitable authentication should be used to assure that no passwords could be predicted. Super-admin password should be divided between two persons or more for security reasons to reduce risk.

We observed many common issues during testing these DLP systems. The tests were unsuccessful in following:

- Blocking hidden data; we will propose suggestions to solve this problem in chapter 5.

- Work in safe mode : No system can protect data against user operations in safe mode of the main operating system. It is possible for the user to make any leak operation during these process such as copy, move and upload.

- Preventing a compressed encrypted file : The systems failed when the file is compressed and encrypted. We can solve this issue by preventing any encrypted data to get out the computer unless a permission key provided by administrator is obtained. Also a justification should be provided by the user to his manager and system administrator. Except that, the encrypted file should be prevented.

Chapter 5

Suggested DLP Solution Model

As mentioned in previous chapter, the data hidden is very dangerous and any user can leak sensitive data by using these stenography techniques. To avoid data leakage outgoing from users by uploading or sending files such as email, dropbox, and etc., we are suggesting precaution model. The importance of our proposed model is to protect data and ensure that no hidden data might be leaked.

5.1 Precaution Model for Data Leakage Prevention

The main descriptions of our model are as the following:

Input: The files that will be uploaded.

Model reaction: Check if the uploaded files are images.

Output:

- If DLP system discovers any breach of a pre-defined policy then apply the action (block, send notification to manager and reroute) or set allow action (send email, copy or upload) and send notification to system logs.
- If DLP system cannot discover any breach of policy, then execute precaution model to ensure destroy data hidden and send notification to system logs and manger.

Precaution Model Description

1. The user wants to upload file through an email, USB or dropbox that contains an attachment, from his official account.

2. DLP agent scans if the attachment is allowed or not based on a predefined policy based on file type, location and context. Checked if he actually has access rights as defined in corresponding policy. If the operation is incompatible with the predefined policy, the DLP agent takes action such as block or sends notification to system logs and manger.

3. If the user is allowed with predefined policy, then send the file to precaution model.

4. Another operation occurred in precaution model. During this operation the system ensures that the file is pure from any hidden data. Then it will be forwarded to the appropriate target such as USB, email and so on. Also a notification will be sent to system logs repository and manger.

Figure 5.1 shows the work-flow diagram for suggested precaution model .

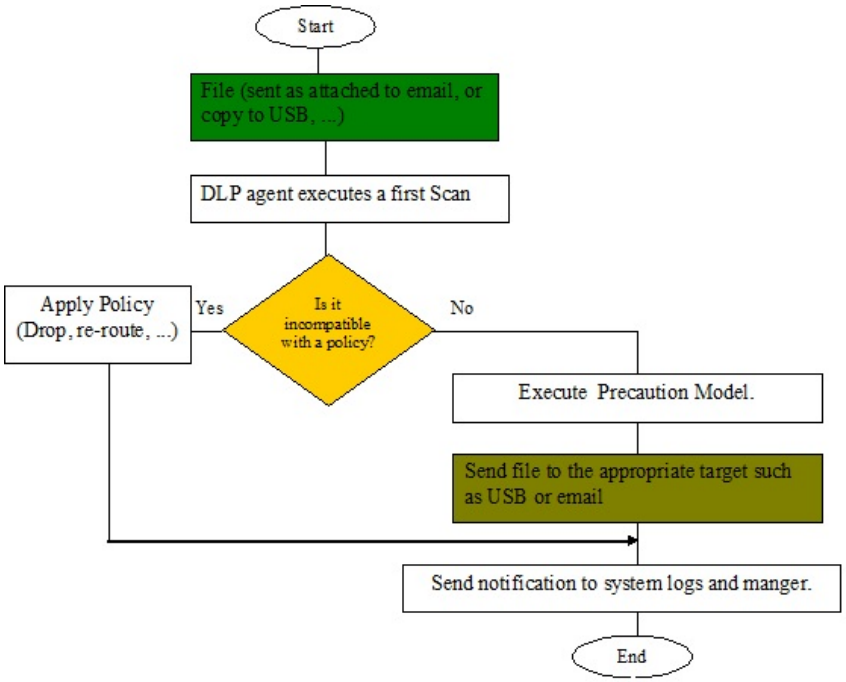

Figure 5.1: Features Flow Chart Diagram of the Precaution Model

5.2 Precaution Model Implementation

In implementation process we want to implement two methods to destroy hidden data that it has been mentioned in section 2.11 and a comparison between these two methods then choose the best method of them based on processing time and image size. The goal of our method is to destroy hidden data without any effect on the quality of the image.

5.2.1 Flipping the Least Significant Bit (LSB) to One's(FLSBO)

We proposed a method to destroy the hidden data by flipping the least significant bit to one. The proposed method will perform the (OR) bitwise operation of each pixel (24bit) with this bit sentence (000000000000000000000001), Consequently.

This will guarantee that the LSB will turn to one. Any hidden data will be destroyed. For example, if we have an image X including hidden data, applying our suggested method will occur as following:

Image X as Table 5.1

Table 5.1: Image X Including Hidden Data.

111111101010011001111010	111111101010011001111010	...	...	.....
.				
.				
111100111010011001111010	111111101010011001111010	...	...	

OR with (000000000000000000000001)

Result will be an image like as shown in Table 5.2:

Table 5.2: Result of OR Operation with Image X.

111111101010011001111011	111111101010011001111011	...	...	.....
.				
.				
111100111010011001111011	111111101010011001111011	...	...	

Figure 5.2: Image(X) has Hidden Data

Figure 5.3: Image (X) After Removing Hidden Data Using OR Operation.

The implantation of this proposed method is done using C# as shown in appendix (A). The pseudo code for this method is shown in the following Table 5.3:

Table 5.3: The Pseudo Code for (FLSBO).

00:	ImageMatrix = getImget ();
01:	for x=0 to imageMatrix.width
02:	for y=0 imageMatrix.hight
03:	Begin
04:	pixel= imageMatrix[x, y]
05:	pixel=pixel OR (000000000000000000000001)
06:	imageMatrix[x,y] = pixel
07:	End
08:	SaveImageFile (imageMatrix)

5.2.2 Flipping the Least Significant Bit (LSB) to Zero(FLSBZ).

We proposed a method to destroy the hidden data by flipping the least significant bit to zero. The proposed method will perform the (AND) bitwise operation of each pixel (24bit) with this bit sentence (111111111111111111111110), Consequently.

This will guarantee that the LSB will turn to zero. Any hidden data will be destroyed. For example, if we have an image Y including hidden data, applying our suggested method will occur as following:-

Image Y table 5.4

Table 5.4: Image Y Including Hidden Data.

111111101010011001111 0011	111111101010011001111 0011	...	...	.....
.				
.				
111100111010011001111 0011	111111101010011001111 0011	...	...	

AND with (111111111111111111111110)

Result will be an image like the following as table 5.5: The implantation of this proposed

Table 5.5: Result After Applying AND Operation

111111101010011001111 0010	111111101010011001111 0010	...	...	.....
.				
.				
111100111010011001111 0010	111111101010011001111 0010	...	...	

Figure 5.4: Image(Y) with Hidden Data

method is done by using C# as shown in appendix (B). The pseudo code for this method as shown in the following Table 5.6:

Figure 5.5: Image (Y) After Removing Hidden Data Using AND Operation

Table 5.6: The Pseudo Code for (FLSBTZ)

00:	ImageMatrix = getImget ();
01:	for x=0 to imageMatrix.width
02:	for y=0 imageMatrix.hight
03:	Begin
04:	pixel= imageMatrix[x, y]
05:	pixel=pixel AND 11111111111111111111110
06:	imageMatrix[x,y] = pixel
07:	End
08:	SaveImageFile (imageMatrix)

After executing these two methods on different samples of images containing hidden data, the results were show in Table 5.7 and Table 5.8:

After comparing the results of the tests, we found that Flipping the Least Significant Bit (LSB) to Zero method is the best way to destroy hidden data because the time required for destroying hidden data is the least, the size of image after destroying hidden data is decreased and image resolution remains the same with the image before destroying the hidden data.

Reasons of using the proposed flipping least significant bit to zero (FLSBZ) include the following:-

1) The time required for destroying hidden data is less than LSBTO method, this appear when applying the two methods on the same samples of images with same environments.

2) The size of image after destroying hidden data is decreased, when using images from type of JPEG, the JPEG type is using loss compared with digital images, so when applying FLSBZ and FLZBO on this type of images, similar bit will cause a decrease in the size of the image.

Table 5.7: Results After Executing Flipping the Least Significant Bit(LSB) to Zero

Image Format	**Size-KB**	**Size-KB**	**Resolution-DPI**	**Time ([μs])**	**Size-KB**	**Resolution-DPI**
.JPG	859	865	96	0.869	122	96
.JPG	827	832	96	0.866	123	96
.JPG	582	587	96	0.965	91	96
.JPG	9	14	96	0.059	11	96
.JPG	758	763	96	0.899	64	96
.JPG	763	768	96	0.869	147	96
.JPG	549	554	96	0.868	94	96
.JPG	760	766	96	0.889	109	96
.JPG	607	612	96	0.892	88	96
.JPG	80	86	96	0.657	86	96
.JPG	70	75	96	0.394	68	96
.JPG	37	42	96	0.476	23	96
.JPG	207	212	96	0.527	48	96
.JPG	9	14	96	0.086	8	96
.JPG	109	114	96	0.101	77	96
.JPG	16	21	96	0.086	15	96
.JPG	13	18	96	0.085	16	96
.JPG	55	60	96	0.202	30	96
.JPG	6	12	96	0.055	8	96
.JPG	8	14	96	0.053	10	96
.JPG	15	21	96	0.055	19	96
.JPG	113	119	96	0.252	66	96
.JPG	11	17	96	0.055	15	96
.JPG	7	12	96	0.055	8	96
.JPG	17	22	96	0.056	20	96
.JPG	42	47	96	0.100	29	96
.JPG	91	97	96	0.736	80	96
.JPG	2506	2512	96	6.705	411	96
.JPG	12	18	96	0.086	15	96
.JPG	14	20	96	0.085	13	96
.JPG	84	89	96	0.388	46	96
.JPG	10	15	96	0.085	12	96
.JPG	174	179	96	0.013	122	96
.JPG	110	116	96	0.485	88	96
.JPG	163	169	96	0.777	164	96
.JPG	602	607	96	0.831	154	96
.JPG	111	116	96	0.481	113	96
.JPG	71	77	96	0.309	72	96
.JPG	77	82	96	0.535	78	96
.bmp	134	140	96	0.775	48	96
.bmp	148	154	96	0.0569	197	96
.bmp	43	49	96	0.0235	57	96
.bmp	82	87	96	0.0352	109	96
.bmp	36	42	96	0.0115	48	96
.bmp	149	155	96	0.0624	198	96

3) Image resolution remains the same with the image before destroying the hidden data when applying FLSBZ and FLZBO methods.

Table 5.8: Results After Executing Flipping the Least Significant Bit(LSB) to One

Image Format	Size-KB	Size-KB	Resolution-DPI	Time ([μs])	Size-KB	Resolution-DPI
.JPG	859	865	96	0.918	122	96
.JPG	827	832	96	0.923	123	96
.JPG	582	587	96	0.986	91	96
.JPG	9	14	96	0.097	11	96
.JPG	758	763	96	0.914	64	96
.JPG	763	768	96	0.922	147	96
.JPG	549	554	96	0.946	94	96
.JPG	760	766	96	0.909	109	96
.JPG	607	612	96	0. 933	88	96
.JPG	80	86	96	0.633	86	96
.JPG	70	75	96	0.288	68	96
.JPG	37	42	96	0.515	23	96
.JPG	207	212	96	0.540	48	96
.JPG	9	14	96	0.121	8	96
.JPG	109	114	96	0.1025	77	96
.JPG	16	21	96	0.106	15	96
.JPG	13	18	96	0.087	16	96
.JPG	55	60	96	0.221	30	96
.JPG	6	12	96	0.108	8	96
.JPG	8	14	96	0.123	10	96
.JPG	15	21	96	0.128	19	96
.JPG	113	119	96	0.282	66	96
.JPG	11	17	96	0.081	15	96
.JPG	7	12	96	0. 071	8	96
.JPG	17	22	96	0.087	20	96
.JPG	42	47	96	0.164	29	96
.JPG	91	97	96	0.769	79	96
.JPG	2506	2512	96	6.80	411	96
.JPG	12	18	96	0.114	15	96
.JPG	14	20	96	0.106	13	96
.JPG	84	89	96	0. 413	46	96
.JPG	10	15	96	0.099	12	96
.JPG	174	179	96	1.343	122	96
.JPG	110	116	96	0.527	88	96
.JPG	163	169	96	0.806	164	96
.JPG	602	607	96	0.893	154	96
.JPG	111	116	96	0.537	113	96
.JPG	71	77	96	0.354	72	96
.JPG	77	82	96	0.564	78	96
.JPG	134	140	96	0.809	48	96
.bmp	148	154	96	0.0582	197	96
.bmp	43	49	96	0.0264	57	96
.bmp	82	87	96	0.0375	109	96
.bmp	36	42	96	0.0203	48	96
.bmp	149	155	96	0.0704	198	96

Image Quality Measurement

Image quality is a characteristic of an image that calculates the recognized image destruction. Imaging systems may present some levels of deformation or artifacts in the signal,

so the quality assessment is an essential problem.

These two methods used in measuring image quality, the first method called MSE (Mean Square Error) in statistics, the mean squared error (MSE) of the estimator calculates the average of the squares of the "errors", that is, the difference between the estimator and what is estimated [65].

The second method which called PSNR (Peak Signal to Noise Ratio) is most often useful to measure the quality of reconstruction of loss compression codecs (e.g., for image compression). The signal in such cases is the original data, and the noise is the error introduced by compression [66].

When we made a comparison in quality between original images with images after destroying hidden data, we noticed that after executing the test on MATALAB program, there is no a significant change on quality when using MSE (Mean Square Error) [67] and PSNR (Peak Signal to Noise Ratio) [68] methods.

The Table 5.9 shows the result of the quality comparison test.

Table 5.9: Quality Comparison

Quality Comparison by Using MSE and PSNR				
Method	**(FLSBTZ)**		**(FLSBTO)**	
	(MSE)	**(PSNR)**	**(MSE)**	**(PSNR)**
1.JPG	13.5487	36.8118	13.5487	36.8118
2.JPG	2.8878	43.5252	2.8878	43.5252
3.JPG	2.0239	45.0690	2.0239	45.0690
4.JPG	1.6089	46.0768	1.6089	46.0768
5.JPG	2.8503	43.5819	2.8534	43.5772
6.JPG	4.7089	41.4016	4.6879	41.4210
7.JPG	5.8689	40.4453	5.9240	40.4047
8.JPG	5.4356	40.7783	5.4350	40.7788
9.JPG	1.3958	46.6827	1.4019	46.6635
10.JPG	2.0724	44.9661	2.689	44.9733
11.JPG	7.6044	39.3201	7.6129	39.3153
12.JPG	23.4558	34.4283	23.5047	34.4193
13.JPG	4.3076	41.7885	4.3170	41.7790
14.JPG	1.6908	45.8499	1.6902	45.8513
15.JPG	6.1642	40.2321	6.1659	40.2308
16.JPG	27.3429	33.7624	27.3607	33.7595
17.JPG	17.1257	35.7943	17.1146	35.7971
18.JPG	4.8617	41.2629	4.8620	41.2627
19.JPG	11.8822	37.3818	11.8787	37.3831
20.JPG	2.1513	44.8038	2.1414	44.8238
21.JPG	23.0830	34.4979	23.0830	34.4979
22.JPG	4.9021	41.2270	4.9271	41.2049
23.JPG	6.2504	40.1717	6.2499	40.1721
24.JPG	40.4556	32.0610	40.5058	32.0556
25.JPG	0.3711	52.4358	0.3530	52.6525
26.JPG	3.4737	42.7228	3.4718	42.7252
27.JPG	0.9131	48.5256	0.8871	48.6511
28.JPG	0.4473	51.6244	0.4791	51.3263
29.JPG	0.0177	65.6587	0.0161	66.0609
30.JPG	4.2126	41.8853	4.2108	41.8872
31.BMP	0.0060	70.3231	0.0124	67.1926
32.BMP	0.0076	69.3437	0.0047	71.3692
33.BMP	0.0064	70.0583	0.0087	68.7423
34.BMP	0.0056	70.6874	0.0144	66.5531
35.BMP	0.0073	69.5013	0.0104	67.9770

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This thesis describes the importance of information regarding to companies and the seriousness of corporate data leakage. This thesis studied the traditional systems used to protect data, DLP systems and the differences between them.

In this thesis we've examined five current DLP systems and evaluated their problems and weaknesses. After reviewing our test results, we sketched the conclusion that no DLP solution is 100 % secure and robust because all solutions showed potential weaknesses and insufficient protection ways when identifying and protecting data.

After executing the tests on five DLP products we noticed from the results of tests, that DLP products still suffer from some weaknesses and share in the same problems such as hidden data, encryption and safe mode.

We showed in our evaluation that the data protection will remain after disconnecting two DLP nodes. If a disconnection between DLP agent and DLP server occurs, the DLP still protect data. As well as our testing analysis, we've discovered some vulnerabilities and suggested improvements to systems.

We have also developed a model by our own, for further investigation to the possible weaknesses. It is a simple proof of concept solution that could be combined with the

existing DLP solutions. The idea of our model is using a precaution method to destroy hidden data before being sent by e-mail, USB flash memory or uploaded outside the network. This model could be provided by other DLP solutions. This model is built and developed using C #.

Our results indicate that there are several threats against DLP systems. Many steps should be taken to ensure the DLP safety, such as restricting user privileges, restricting entering to safe mode of the OS, or preventing any data encryption.

It is also necessary to keep testing DLP solutions in depth more and more to reveal the software vulnerabilities. Even if the security of the software is not related to the approach of DLP, it is necessary to assure that the software doesn't contain any vulnerability. In other words, it isn't possible to prevent leakage if the software management is already hacked and sensitive data could be revealed.

Because this thesis could only cover five DLP solutions, much crucial efforts should be done to check more DLP products. This kind of examination process should continue to test other DLP solutions to have an overall image of how robust is the security for these types of products.

The tests outcomes showed that DLP systems suffer from weaknesses for example, all DLP not operate in safe mode, can't find out file encryption and compressed file also can't find data hidden when using steganography technique but still need improvement and enhancement in discovering data.

An effective method has been developed to resolve some of these weaknesses which enhance the effectiveness of DLP systems, since the time required to destroy the hidden data is very small. Also the size of picture after destroying hidden data is decreased; image resolution not affected after destroying the hidden data, and no major change in image quality.

6.2 Future Work

In our tests, we used sample of the widely used DLP solutions. Resuming our experiments and extending them with other similar applications may expose more information related to DLP systems and their weaknesses. Additional examininations should be made for some other OS's or devices such as smart phone.

Smart phones are predicted to become the primary computerized devices in the companies and could be used mainly in the future as the Bring Your Own Device (BYOD) comes in the field. Since smart phones widely-used to access the companies sensitive data for example emails and files, it's predicted to be familiar with leak information mistakely or deliberately. There's a requirement for future research to discover methods on how to permit employees of such an organizations to access sensitive information through their smart phones without leaking these information outside the company or organization via malicious applications that the user downloaded from the mobile application markets.

Continuing the development for creation of our DLP model and integrating it with other DLP solutions could greatly improve the effectiveness of current DLP systems.

Finally, future research on how DLP systems used in Cloud computing to prevent data loss/leakage as the Cloud computing is one of the present IT subjects that is new to business channels.

Bibliography

[1] M. A. Faysel and S. S. Haque, "Towards cyber defese: Research in intrusion detection and intrusion prevention systems," *IJCSNS*, pp. 316–323, 2010.

[2] R. Tahboub and Y. Saleh, "Data leakage/loss prevention systems (dlp)," *International Journal of Information Systems*, pp. 13–19, 2014.

[3] T. Aulakh, "Intrusion detection and prevention system: Cgi attacks," *San JosA State University*, pp. 7–8, 2009.

[4] S. Dharmapurikar, P. Krishnamurthy, T. Sproull, and J. Lockwood, "Deep packet inspection using parallel bloom filters," *IEEE*, pp. 65–78, 2004.

[5] P. Liwei Ren, "Overview of data loss prevention (dlp) technology," *TsinghuaUniversity, Beijing, China*, pp. 65–78, 2012.

[6] A. Shabtai, Y. Elovici, and L. Rokach, "A survey of data leakage detection and prevention solutions," *Springer Publishing Company*, 2012.

[7] P. Liwei Ren, "Dlp systems: Models, architecture and algorithms," *UCSC, Santa Cruz*, pp. 19–21, 2013.

[8] K. Westphal, "Steganography revealed," *Indiana University of Pennsylvania,IUP*, pp. 16–20, 2013.

[9] P. Raman, H. G. KayacAk, and A. Somayaji, "Understanding data leak prevention," *Annual Symposium on Information Assurance (ASIA) Journal*, pp. 7–8, 2011.

[10] P. Karhula, ""what is the effect of wikileaks for freedom of information?"," *International Federation of Library Associations and Institutions*, 2012.

[11] P. Kanagasingham, "Data loss prevention," *SANS Institute InfoSec*, pp. 16–20, 2008.

[12] T. H. Tomoyoshi Takebayshi, Hiroshi Tsuda, "Data loss prevention technologies," *FUJITSU SCI*, pp. 170–179, 2011.

[13] Symantec, "Symantec data loss prevention 11.5: Administration," *Symantec*, 2010.

[14] Websense, "Triton - data security help," *Websense Data Security*, 2010.

[15] OpenDLP, "Rreadme-0.5.1," *Opendlp Resources*, 2015.

[16] McAfee, "Mcafee complete data protection—advanced," *McAfee*, 2014.

[17] Trustwave, "Data loss prevention for government," *Trustwave Resources*, 2013.

[18] M. T. Ltd, "Mydlp administration guide v2," *Medra Teknoloji Ltd*, 2013.

[19] EMC, "Rsa data loss prevention endpoint," *EMC Resources*, 2012.

[20] Fortinet, "Fortios handbook security profiles for fortios 5.0," *Fortinet*, 2014.

[21] C. Waxer, "The top 5 internal security threats," *IT Security*, 2007.

[22] R. Carroll, "Snowden used simple technology to mine nsa computer networks," *The Guardian*, 2014.

[23] S. A. Williamson, P. Varakantham, D. Gao, and O. C. Hui, "Active malware analysis using stochastic games," *AAMAS2012*, pp. 65–78, 2012.

[24] D. Maulik, H. Joshi, and B. K, "Towards application classification with vulnerability signature for ids/ips," *SecurIT12.*, pp. 17–19, 2012.

[25] P. A. Hilmi A. Lahoud and P. Xin Tang, "Information security labs in ids/ips for distance education," *SIGITE06*, pp. 19–21, 2006.

[26] J. Rutkowska, "Introducing stealth malware taxonomy," *COSEINC advanced malware lab*, pp. 19–21, 2006.

[27] Salem, B.M, and Stolfo, "A survey of insider attack detection research.insider attack and cyber security- beyond the hacker," *Springer.*, pp. 23 – 27, 2008.

[28] J. Hong, J. Kim, , and Cho, "The trend of the security research for the insider cyber threat," *International Journal of Future Generation Communication and Networking*, pp. 65–78, 2010.

[29] J. M. Butler, "Benchmarking security information event management (siem)," *SANS Institute InfoSec*, 2009.

[30] P. S. G. Asmaa Shaker Ashoor, "Intrusion detection system (ids) intrusion prevention system (ips):case study," *International Journal of Scientific Engineering Research Volume 2*, 2012.

[31] H. Mun, K. Han, C. Yeun, and K. Kim, "Yet another intrusion detection system against insider attacks proceedings," *SCIS*, pp. 22–25, 2008.

[32] C. G. Networks, "Financial services data loss assessment," *Code Green Networks, Available http://www.codegreennetworks om/resources/downloads /assessment_financial.pdf*, pp. 19–21, 2009.

[33] C. leadership team, "Key findings from the 2014 us state of cybercrime survey," *pwc*, pp. 22–25, 2014.

[34] Rajesh and Parthasarathy, "Data breach statistics," *MENTIS, Available: http://www.mentisoftware .comrescs_data_breach_statistics.html*, pp. 65–78, 2013.

[35] S. Al-Fedaghi, "A conceptual foundation for data loss prevention," *International Journal of Digital Content Technology and its Applications. Volume 5, Number 3*, pp. 293 – 303, 2011.

[36] E. Tittel, "Understanding data loss prevention," *Ed's Tom's IT Pro*, pp. 19–21, 2013.

[37] IBM, "Cost of data breach study," *Ponemon Institute Research Report*, pp. 17–19, 2014.

[38] A. S. Security, "Data loss prevention: Keeping sensitive data out of the wrong hands," *pwc*, pp. 22–25, 2008.

[39] S. Liu and R. Kuhn, "Data loss prevention," *IEEE*, pp. 16–20, 2010.

[40] N. Kumaresan, "Key considerations in protecting sensitive data leakage using data loss prevention tools," *ISACA Journal Volume 1*, pp. 23–27, 2014.

[41] P. Liwei Ren, "Dlp systems: Models, architecture and algorithms," *UCSC, Santa Cruz*, pp. 19–21, 2013.

[42] Websense, "Deployment and installation guide," *Websense Data Security*, 2012.

[43] OpenDLP, "Readme-vm-0.5.1," *Opendlp Resources*, 2015.

[44] J. Jesse, "Symantec dlp overview," *it.northwestern.edu*, 2014.

[45] R. Mogull, "Best practices for endpoint data loss prevention," *Symantec*, 2012.

[46] Trustwave, "Best practices in email, web and social media security," *Trustwave Resources*, 2014.

[47] EMC, "Rsa data loss prevention datcenter," *EMC Resources*, 2013.

[48] EMC, "Rsa data loss prevention network," *EMC Resources*, 2013.

[49] McAfee, "installation guide mcafee data loss prevention," *McAfee*, 2012.

[50] McAfee, "Mcafee data loss prevention endpoint 9.3.200," *McAfee*, 2014.

[51] McAfee, "Installation guide revision b mcafee epolicy orchestrator 5.1.0 software," *McAfee*, 2014.

[52] E. E. and A. bed Elgabar, "Comparison of lsb steganography in bmp and jpeg images, information technology," *King Abdul Aziz University*, pp. 91–95, 2013.

[53] A. b. E. Eltyeb E, "Comparison of lsb steganography in bmp and jpeg images, information technology," *King Abdul Aziz University*, pp. 91–95, 2013.

[54] B. C. Maybury, Penny Chase, "Analysis and detection of malicious insiders," *International Conference on Intelligence Analysis,McLean*, 2005.

[55] Gaonjur and Bokhoree, "Risk of insider threats in information technology outsourcing:can deceptive techniques be applied?," *Journal of Security and Management*, pp. 65–78, 2006.

[56] Franqueira, Cleeff-A, Eck-P, and R. Wieringa, "External insider threat: a real security challenge in enterprise value webs," *5th International Conference on Availability, Reliability and Security*, pp. 16–20, 2010.

[57] L. Spitzner, "Honeypots: catching the insider threat," *19th Annual Computer Security Applications Conference (ACSAC03)*, pp. 170–179, 2003.

[58] J. Blasco, J. C. Hernandez-Castro, J. E. T. and, and A. Ribagorda, "Bypassing information leakage protection with trusted applications.computers & security," *Computers & Security Volume 31, Issue 4*, pp. 557–568, 2012.

[59] M. Luft, "Can data leakage prevention prevent data leakage?," *Master's thesis, University of Mannheim*, pp. 23 –27, 2009.

[60] S. Ghorbanian and G. Fryklund, "Master's thesis, improving dlp system security," *Faculty of Computing Blekinge Institute of Technology*, 2014.

[61] VMware, "virtualization-basics," *VMware, Inc.*, date access 2015- 2/1/2015.

[62] Websense, "Unified data loss prevention for gateways,endpoints and discovery," *Websense Data Security*, 2013.

[63] Websense, "Web, email and data loss protection (dlp) within a unified hybrid security architecture," *Websense Data Security*, 2013.

[64] Symantec, "Symantec data loss prevention data insight," *Symantec*, 2010.

[65] G. Lehmann, E. L. Casella, "Theory of point estimation (2nd ed)," *New York: Springer. ISBN 0-387-98502-6. MR 1639875*, 1998.

[66] S. T. Welstead, "Fractal and wavelet image compression techniques," *SPIE Publication*, pp. 155–156, 1999.

[67] MathWorks, "Mean-squared error," *http:// www.mathworks.com/help/images/ref/ immse.html ?searchHighlight=MSE*, date access 2015- 5/8/2015.

[68] MathWorks, "Compute peak signal-to-noise ratio (psnr) between images," *MathWorks,Available:http:// www.mathworks.com/ help/vision/ ref/ psnr.html*, date access 2015- 5/8/2015.

Appendix

(A) Source Code For Flipping the Last Significant Bit (LSB) to one

```
Public void OrWithZero ()
    {
      Bitmap temp = (Bitmap)_current Bitmap;
      Bitmap bmap = (Bitmap) temp.Clone ();

      Stopwatch stopwatch = new Stopwatch ();

      Stopwatch. Start ();
      Color c;
      For (into i = 0; i < bmap.Width; i++)
      {
        For (int j = 0; j < bmap. Height; j++)
        {
          c = bmap.GetPixel (i, j);
          int Argb = c.ToArgb();
          int NewRes = (Argb | 1);
          c = Color.FromArgb(NewRes);
          int cR = c.R;
          int cG = c.G;
          int cB = c.B; ;

          if (cR < 0) cR = 1;
          if (cR > 255) cR = 255;

          if (cG < 0) cG = 1;
          if (cG > 255) cG = 255;

          if (cB < 0) cB = 1;
          if (cB > 255) cB = 255;

          bmap.SetPixel (i, j, Color.FromArgb ((byte)cR, (byte)cG, (byte)cB));
        }
      }
      _current Bitmap = (Bitmap) bmap.Clone ();
      Stopwatch. Stop ();
      MessageBox.Show ("The Time Process " + stopwatch.Elapsed.ToString());
```

Figure 6.1: Source Code For FLSBO

(B) Source Code For Flipping the Last Significant Bit (LSB) to Zero

```
{
    Bitmap temp = (Bitmap)_currentBitmap;
    Bitmap bmap = (Bitmap) temp.Clone();

    Color c;
    for (int i = 0; i < bmap.Width; i++)
    {
        for (int j = 0; j < bmap. Height; j++)
        {
            c = bmap.GetPixel (i, j);
          int Argb = c.ToArgb();
          int NewRes = (Argb & 16777214);
                    if(i==j)
          {
            if (NewRes == 111) { }
          }
          c = Color.FromArgb (NewRes);
          int cR = c.R;
          int cG = c.G ;
          int cB = c.B; ;
          if (cR < 0) cR = 1;
          if (cR > 255) cR = 255;
          if (cG < 0) cG = 1;
          if (cG > 255) cG = 255;
          if (cB < 0) cB = 1;
          if (cB > 255) cB = 255;
          bmap.SetPixel (i, j, Color.FromArgb ((byte) cR, (byte)cG, (byte)cB));
        }
    }
    _currentBitmap = (Bitmap) bmap.Clone ();
  }

} }
```

Figure 6.2: Source Code For FLSBZ

(C) Boyer-Moore-Horspool (BMH) Algorithm [41]

Character comparison can be made from right to left, starting from the end of the pattern. Ending Character Heuristics Consider that we are pointing to character $R[i]$ and try to compare it with the ending character of P

Bad character **if** $R[i] \neq P[m]$ *and* $R[i]$ *is not included in P*ś *alphabet* **then**
    it is safe for the pointer to skip m positions arriving at $R[i+m]$
end
if $R[i] \neq P[m]$, $R[i]$ *is included in P*ś *alphabet and* $R[i]$ś *last occurrence within* P *has distance* q *from the end of* P **then**
    it is safe for the pointer to skip q positions arriving at $R[i+q]$
end

Good character **if** $R[i] = P[m]$, P *is not matched , and* $R[i]$ *has no other occurrences within* P **then**
    it is safe for the pointer to skip m positions arriving at $R[i+m]$
end
if $R[i] = P[m]$, P *is not matched and* $R[i]$*'s last occurrence other than* $P[m]$ *has distance* q *from the end of* P **then**
    it is safe for the pointer to skip q positions arriving at $R[i+q]$
end
Matched instance **if** $R[i] = P[m]$ *and* P *is matched* **then**
    save the instance.
end
It is almost safe to move the pointer to skip m positions arriving at $R[i+m]$. Character comparison can be made from right to left, starting from the end of the pattern.

Ending Character Heuristics

Consider that we are pointing to character $R[i]$ and try to compare it with the ending character of P

Bad character

if $R[i] == P[m]$ *and* $R[i]$ *is not included in P*ś *alphabet* **then**
    it is safe for the pointer to skip m positions arriving at $R[i+m]$
end

Algorithm 1: Boyer-Moore-Horspool (BMH) Algorithm [41]