Cyber Guard

(Data Loss Prevention Solution)

By

Hamza Tahir

Dawood Aijaz

Wareesha Ahmed

Gulzar Ahmed Butt

Supervisor

Asst. Prof. Waleed Bin Shahid

Submitted to the faculty of Department of Computer Software Engineering,
Military College of Signals, National University of Sciences and Technology,
In partial fulfillment for the requirements of B.E Degree in
Computer Software Engineering

JULY 2020

CERTIFICATE OF CORRECTIONS & APPROVAL

Certified that work contained in this thesis titled“_Cyber Guard (Data Loss Prevention Solution) ” ,carried out by Hamza Tahir, Dawood Aijaz, Gulzar Ahmed Butt and Wareesha Ahmed under the supervision of Asst. Prof. Waleed Bin Shahid for partial fulfillment of Degree of Bachelors of Computer Software Engineering, in Military College of Signals, National University of Sciences and Technology, Islamabad during the academic year 2019-2020 is correct and approved. The content that has been written from other sources it has been rightly acknowledged. This work is original with 10 % plagiarism.

Approved by

Supervisor

AP Waleed Bin Shahid

Date: 11.07.2020

DECLARATION

No portion of work contained in this thesis has been submitted in the interest of another award or qualification in either this institute or somewhere else. We hereby declare that the work contained in this report and intellectual contents of this report are the product of our work. This thesis report has not been formerly published in any structure nor does it include any verbatim of the published resources which can be treated as violation of international copyright decree. We also affirm that we do recognize the terms ‘plagiarism’ and ‘copyright’ and that in case of any copyright infringement and plagiarism established in this thesis, we will be held fully accountable of the consequences of any such violation.

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Students

Hamza Tahir --- Reg # 198667

Hamza

Dawood Aijaz ---Reg # 136576

Dawood

Gulzar Ahmed --- Reg. # 174229

Wareesha Ahmed--- Reg. # 199869

Wareesha

Signature of Supervisor

AP Waleed Bin Shahid

Waleed

Acknowledgements

We are thankful to our God, the Almighty Allah, who helped us throughout this work, directed us every time when we were stuck and in dire need of His help and for every idea which You put in our brains to make it better. Certainly, we could not have achieved this without Your invaluable help and direction. Everyone who helped us during this time, whether our teachers or any other person was Your will, so indeed none be worthy of praise but You and You are the most merciful.

This thesis serves as a tribute to our advisor, Asst. Prof. Waleed Bin Shahid for the time, patience, and efforts they have spent on us. we are indebted for the vision, knowledge, and mentality I acquired from him, and we, the team members, feel privileged and proud to have benefited from his mentoring and guidance.

We would like to thank Asst. Prof. Mobeena Shehzad, for being our Final Year Project coordinator, for her guidance and support and helping us to complete this required work. We are also thankful to Dr. Muqeem Sheri for being on evaluation committee and for mentoring us throughout the project with valuable inputs and for his comments and suggestions that have shaped this work.

We would also like to thank our instructors, professors and all the people in Military college of Signals (NUST) who taught us and helped us to complete our study program.

We are very much thankful to our dear parents who cared for us when we were not able to even walk and helped us at every step of life.

Finally, we would like to express our gratitude to our friends and all those who people supported us and helped us during the course of four years.

Dedicated to our remarkable teachers and dear parents whose substantial support and assistance led us to this marvelous achievement.

Abstract

In this modern world, organizations use different security solutions to protect their sensitive and critical business data. This need of protection is a very important for organizations and need to be over emphasized. Data Loss Prevention (DLP) is one of the efficient ways of preventing this Data Loss. DLP identifies and stops the unauthorized actions of copying or sending sensitive data, both deliberately or/and accidentally. This solution is developed to detect data leak incidents timely and this is done by monitoring data. It is, without any doubt, appropriate for business activities of organization because it does not disrupt the work flow. Data loss prevention is not only suitable for protecting structured data but also deals with unstructured data. DLP helps organizations in preventing data loss of sensitive information like Personal identifiable, financial, trade secrets etc. and act as a preventive measure.

This thesis has main aim of presenting all the essential information and describing the key terms about data loss prevention, in a simple and understandable way. We have tried to solve sensitive data loss problem and tried to develop an effective and working DLP solution. This has been carried out by conducting a comprehensive study of research on Data Loss Prevention technology and by analyzing already developed solutions. A system architecture is presented, which is designed to implement the basic requirements of a DLP. After the careful implementation of design, we analyzed the system and evaluated with test case and propositions are formed to decrease the defects and make system more efficient. A recent technology i.e. Cloud DLP, which is very famous and used now a days, has been taken into account as future research in this thesis.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

1.1 Problem Statement

For some past years, everything is being digitized in organizations. Latest technologies are being used in workplaces by employees in their daily work life and everything is being shared digitally among co-workers and clients. So, the corporate dependency on digital data to achieve business goals has increased very much. The digital data related to any business surely elevate the organization's activities of people that are involved internally and externally in an organization. This evolution has given rise to a e major challenge, organizations are encountering nowadays, that is to protect sensitive data from theft or being leaked intentionally or unintentionally. A survey of 2014 found out that almost half of data leakages were related to business environment [1]. There are various ways through which data travels within and outside the organization network and there are many formats of this data like e-mails, word documents, images, zip files, databases and chats on social media etc. The organization's data consists of both sensitive and non-sensitive data, non-sensitive data is not harmful if it gets out of the organization, but sensitive data is valuable information which needs to be protected and prevent any leakage to unauthorized personnel. This protection is needed due to several reasons such as data can private in nature, or financially important for organization or contains business regulatory policies and objectives information, or it contains information about competitors.

Organizations must develop different policies and rules to prevent data loss. Moreover, the workers of an organization should also take part in enacting best solution for data leakage protection. No organization can accomplish full protection of sensitive data only by protecting end-points, organization networks and outside communication. Although many data protection solutions exist e.g. intrusion detection system/intrusion prevention system (IDS/IPS), firewalls, and virtual private network (VPN), but they are not very effective [2]. These solutions are also not consistent with each other, they are deployed at different levels of organization and each solution requires different type of management and skill set. Thus, they are not so much effective and sensitive data is still leaked from organizations. Such data leaks are very dangerous for organizations because it can cause bad reputation of organization, thus losing business partners or valuable customers trust, failure in competitive race of market and financial loss [3] For resolving such concerns new and better solutions are being which monitors and protects the sensitive data of organizations. These solutions differ in their functionalities and modes of protecting data, but all together they come under the umbrella of data loss prevention (DLP) [4].

Developing a Data Loss Prevention (DLP) solution can help mitigating the threats related with human factors. DLP is a technical security solution that implements policies for analyzing information and data classification, protecting and monitoring of sensitive information [3]. Moreover, DLP also helps the employees to get an understanding of sensitive data and its classification in company.

1.2. Objectives

The main objective is to look into one of the Cyber Security field related to enterprises i.e. data loss and the main focus is on data loss due to insider attacks (data thefts caused by employees of organizations) and develop a solution to solve this data loss prevention problem. This thesis will give an understanding which data is sensitive for an enterprise, what are the risks of its leakage in an organization, and how such thefts can be stopped and mitigated after studying it. Already developed solutions and researches on DLP will be analyzed and their limitations will be described. Requirements will be documented and then on the basis of those system will be designed. Functionalities and limitations of the system will be described in detail. Accuracy of the solution will be evaluated and the technology will be identified which will be taken into account as future research in this thesis. And this thesis also aims to answer the many un-answered questions in detail like what the DLP solution itself is? and how data theft happens in an organization? how policies are set according to data? what are the key identifiers of data and its classification? And these ambiguities will be cleared after studying it thoroughly and findings will be concluded at the end.

1.3. Methodology

The main motive is to solve sensitive data loss problem and show an effective and working DLP solution so this will be carried out as a result of combination of comprehensive research on the related workings of others, analysis of existing solutions and all details about system developed in this thesis. Explanation of the problem will be carried using help of research papers, internet, as well as interactions with people qualified in the field and the available developed systems. After survey requirements of a DLP solution will be identified and a peculiar architecture is suggested. The functionalities of DLP products available in market are very much different so the basic main features of Data loss prevention toolkit on which our system is developed are only focused on. Then implementation of the proposed design will be carried out. This implementation will be validated by test cases and propositions are formed to decrease the defects and make system more efficient. Then survey will be done to find some good technology that can enhance

DLP functionality to prevent data loss and it will be described how work will be performed on this research in future.

1.4 Thesis Outline

Rest of the chapters are organized as following:

Chapter 2: Literature and research articles related to data loss prevention are reviewed and described.

Chapter 3: describes what is Cyber Guard Data Loss Prevention Solution? This chapter also covers the functionalities of the proposed solution.

Chapter 4: specifies the requirements for the Cyber Guard.

Chapter 5: This chapter discusses the design of the suggested model. Architecture of every component of the system and data design is described in detail.

Chapter 6: Describes how Cyber Guard solution is working to prevent data loss.

Chapter 7: Discusses deployment of the proposed toolkit. Installation and Configuration of the solution comes in this chapter.

Chapter 8: Offers analysis and evaluation of the system with a few test cases of data leakage.

Chapter 9: Concludes our work and presents some future lines of research about improving this project.

CHAPTER 2: LITERATURE REVIEW

2.1. Definition of Data

Data can be defined as information computed or stored by a computer. The data can take many forms such depending on its use such as text documents, images, audio clips, source code, software programs, or other types of data. This data can be stored in many ways such as HDD, Discrete Drives, Cloud storages and databases [5].

In context of data loss, the data can be broadly classified in to 3 subcategories.

- **Data in use**: data in use can be defined as the data that resides on user’s endpoint systems. The user performs routinely actions on this data.
- **Data in motion**: data in motion can be defined as the data moving over the internet. Such as data moving over different internet protocols (HTTP, FTP, SMTP)
- **Data at rest**: data at rest is the static data stored in organizations storage servers such as database and file servers.

2.2. What is Data loss?

Data loss happens when valuable, sensitive or personal information of an organization is lost due to theft, human error, viruses, malware, insider attack

In context of DLP the data loss can be describes as leakage of valuable, sensitive or personal data from the organization. There are variety of ways through which the data leakage can occur. The data is a valuable commodity however in some cases data can become liability. If data is lost or leaked the organization reputation can be damages and may have to face different lawsuits rooting from different government regulation to protect user’s information in hands of different organizations

1. The payment card industry data security standards (PCI DSS)
2. The health insurance Portability and accountability act (HIPPA)
3. The federal information security manager act (FISMA)

It is approximated that the 50 % of data leakage are linked with data leaks from outside the organization such as hacks. If excluding hacks from outside this percentage drops and the insider attacks percentage increase from 40% to 60%. The most frequently effected data is names, social security number, address and medical data [6].

In 2014 a report was published by PPI on data leaks. In the report different source of data leaks were considered for 486 cases, outside in 245 cases, unknown in 88 cases. Incidents with data leaks from inside an organization are most common. The report shows that 60% of the data leakage the users use their correct identity and have valid access to the data. The report also shows that the 273 breaches, which is 33% of the total cases are unintentional. 214 data loss incidents were caused by insiders accidentally [6].

2.3. Data loss Prevention

A notable cybersecurity research and consulting company, Securosis that has been biggest admirer of data loss prevention solutions, defines Data Loss Prevention as:

“Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use, through deep content analysis.” [7]**.**

Several attempts and studies have been carried out in field of data leakage. Different names and terminology have been used to describe DLPS such as data loss/leak prevention, information loss/leak prevention

Data loss prevention (DLP) is typically describes as any software/tool that tracks confidential information, monitors that data as it moves through and out of the enterprise network and prevents unauthorized disclosure of data by creating and enforcing disclosure rules and policies [8].

Basically, a Data loss Prevention is a software the detects and monitor potential data ex-filtration, transmission and generates alert to notify person in charge. The high-level purpose of the DLP is to prevent data from getting in possession of unauthorized party. Other words linked with information leak detection (ILDP), information leak prevention (ILP), content monitoring and filtering (CMF) [9].

DLPs are dedicated analytical systems used to protect data from unauthorized accessed. DLPs have ability to analyze content. Secondly DLPs can be deployed inside an organization to monitor the data related activities. The third feature is to stop the various data leaks by taking various counter measures such as generating alerts, blocking the activity. The major plus point of the DLPs is that they provide proactive measures to stop data leakage.

A typical DLP works by monitoring sensitive data by using regular expression, data fingerprinting and machine learning.

An effective DLP should cover all the communication channels to protect the data. The communication channels can be various such as CD, Email, FAX, smart phone, Web, printer, virtual drives, USB. The channels can be categorized into 3 sets:

1. In Motion
2. In Use
3. At rest

A data loss solution must also provide a central management console through which the administrator can manage the activities of the DLP have insights and create rules and policies and enforce them. This management should have an easy to use interface and a short learning curve.

2.4. Importance of Data Loss Prevention

As mentioned earlier the data loss can be trouble-some for the organization and all other security solutions such as IPS/IDS are non-proactive solutions and does not protect against insider attacks this highlights the importance of the Data loss prevention solution. Data loss prevention (DLP) is described as a software or process that identifies confidential data, tracks that data as it moves through and out of the enterprise and prevents unauthorized disclosure of data by creating and enforcing disclosure policies.

A data loss is important for a business so that it can protect it from potential scrutiny, lawsuits and reputation loss. The importance of the DLP for an/ organization can be derived from following points:

1. Organization have a system from protecting from outsider attacks but does not have any protection against insider attacks.

2. Organization wants to have control over their confidential and valuable data.
3. Organization is required to obey national or organizational regulations.
4. Proactively stop wrong use of information through various channels
5. Gain competitive advantage in both brand value and reputation

The above-mentioned objectives can only be achieved by deploying a well thought DLP in the organization [10].

2.5. Weakness of Data loss prevention

In this passage we describe the short coming of the DLP. The DLP does not come without its limitations. In some cases, the DLP can be insufficient or even obsolete. The DLP systems can be affected by the false or improper data classification. If the sensitive data is not described properly the DLP may have tough time protecting it. The rules and policies should be well defined. The second major weakness of DLP is cryptography. If the sensitive data being leaked if encrypted using some encryption algorithm the DLP will not be able identify hence unable to stop the sensitive data transmission

The third weakness of the DLP is that there is a huge number of data communication channels through which data can be leaked. A DLP solution can only cover a limited number of channels. The unprotected channels can be exploited. The fourth weakness of the DLP is that they produce a large number of false positives. In some cases, the number of false positives can be so huge that the DLP itself can become a major problem. The false positives can divert the attention and resources from the real problems.

The fourth issue with the DLP is that they require a lot of human input to work properly. Constant attention is required by the administrator. The human input can also be a source of errors. The fifth issue with the DLP is that not many companies have the required infrastructure and resources to deploy full fledge DLP system. The DLP can bottleneck their network and consume all the computational resources [11].

2.6. Comparison with IDS/IPs and Firewalls

There are many security products available in the market. Each have its own purpose. Some examples are intrusion detection system (IDS), Intrusion Prevention System (IPS), firewall and lastly Data loss prevention software's

The IPS/Firewall system does not support network partitioning of network security however DLP does support this

IPS/Firewalls does not support SSL offloading however DLP does. The DLP can also validate encrypted session however the IPS/Firewall are insufficient to do so. The DLP can inspect encoded traffic while the IPS/Firewall are not design for this purpose

The best-selling point for the DLP is that it stops data theft and provides protection against these incidents. The DLP also provide TCP connection pooling but the IPS/Firewall does not. The DLP can also help with Request/Response logging and application access logging. DLP has a built-in authentication system so that authorized personals can carry out their task. However, lack of authentication in IPS/Firewall can create a lot of issues for valid actions. The DLP is a proactive solution which help preventing issues on time

The DLP has much more functionalities as compared to other security products. This give DLP an competitive edge. This also reduces the cost as the DLP covers a lot of functionalities in a single product [12].

2.7. Work done and Methods used to solve Data Loss in different researches

A lot of work has been done in field of DLP both by academia and the industry. Many research papers have been published describing different methodologies to build a DLP.

In a research published the DLP analyze data in 2 major categories

- **Context analysis**: this approach analyzes the meta data and the properties attached to the confidential data. The DLP keeps a check on the different attributes such as size of files, file extension, file name.
- **Content analysis**: In this the DLP focuses on the content of the files and directly analyzes it. DLP does this by comparing the data under observation with the predefined rules and policies. If the rule violation is detected an alert is generated.

A DLP solution can be either be preventative or detective. Depending on the needs of the organization they can select one of following methods:

Methods:

- **Data mining and text clustering**: This approach involves predicting when a data leakage will occur by learning about leakage information from the past. This is useful for the detection of unstructured documents such as source code files, long text files, excel files and binary large object (BLOBS)
- **Data identification:** This approach compares the data under observation such as data flowing through the organization network, data on the organization servers with the confidential data defined in the rules and policies. This is best for structured data such as different identification numbers, SSN, credit card numbers,

IBAN number and email. The data identification can also be based on the key word matching

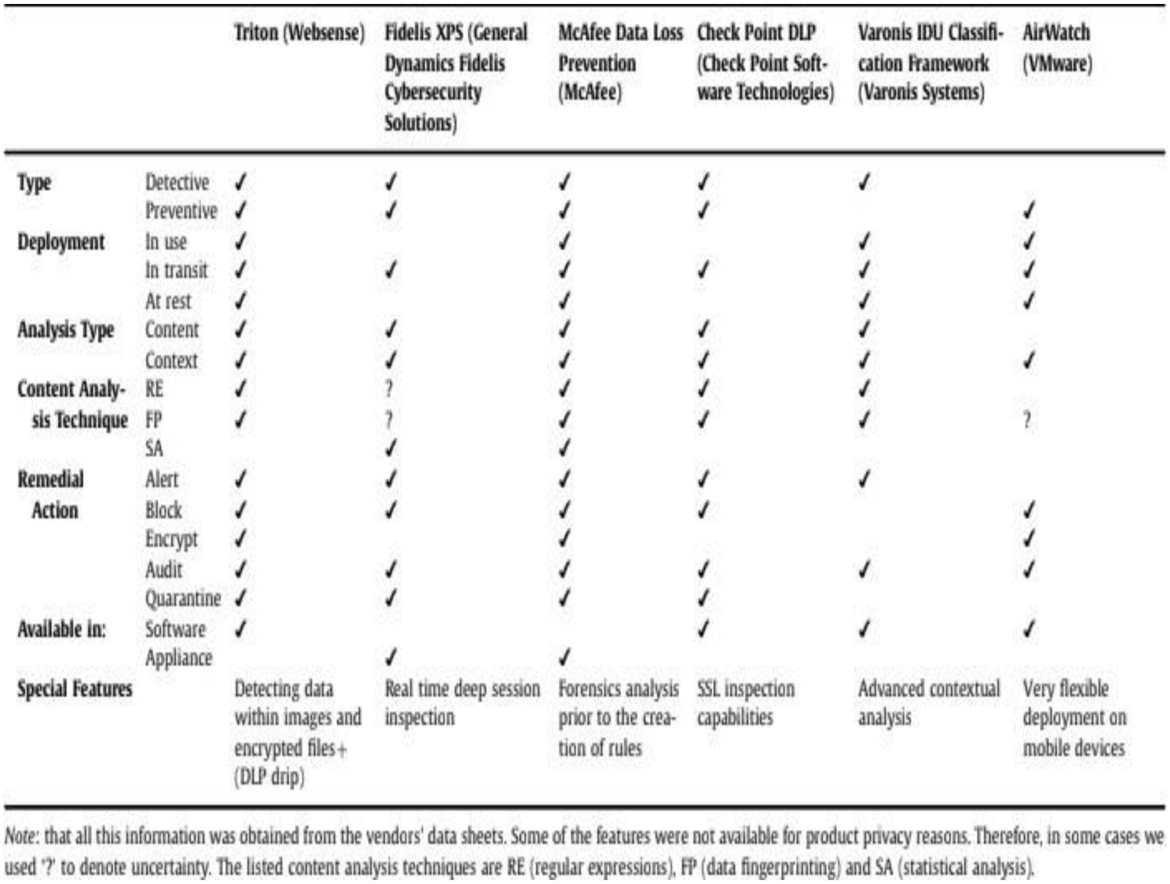
- **Quantifying and limiting:** The quantifying ways to monitor the communication channels through which the data can be leaked. This can be used for data in use or data in motion. Example the organizations network can be monitored to identify data leaks. All the internet protocols such as HTTPS, SMTP and FTP can be monitored closely, and the traffic can be analyzed
- **Social and Behavior analysis:** This approach closely monitors the user's activity inside an organization. If any **malicious** activity is detected appropriate response can be taken. This approach in most difficult of all approaches and can generate a lot of false positives
- **Regular Expression:** Regular Expression are one of the most famous approaches used in DLPs. The regular expression is used to define the structure of the confidential data. Mostly the regular expressions are used DLPs for exact of partial detection of data

All of the above-mentioned techniques have been proved to be beneficial. A combination of such method can be used to build a complete DLP [1].

2.8. Existing solutions:

DLP are offered by various cyber security vendors. The vendors are continuously improving their product to tackle new threats through an iterative approach. Each solution offers a wide range of functionality. In this passage only a few DLPs have been considered most of them were nominated for DLP of 2015 award.

As shown in the table 2.1. different DLPs from different vendors offers different combination of features However some of the DLPs have a special and unique feature such as Triton DLP has ability to detect sensitive data among images using OCR and computer vision. Some of the DLP offers coverage to mobile devices as well. These DLPs are offered at different prices depending of the size of organization, data states to cover and number of functionalities. Some of the vendors also offer a subset of functionalities or getting some specific functionalities this helps the client to cherry pick the features they want for their organization [1].



		Triton (Websense)	Fidelis XPS (General Dynamics Fidelis Cybersecurity Solutions)	McAfee Data Loss Prevention (McAfee)	Check Point DLP (Check Point Software Technologies)	Varonis IDU Classification Framework (Varonis Systems)	AirWatch (VMware)
Type	Detective	✓	✓	✓	✓	✓	
	Preventive	✓	✓	✓	✓		✓
Deployment	In use	✓		✓		✓	✓
	In transit	✓	✓	✓	✓	✓	✓
	At rest	✓		✓		✓	✓
Analysis Type	Content	✓	✓	✓	✓	✓	
	Context	✓	✓	✓	✓	✓	✓
Content Analysis Technique	RE	✓	?	✓	✓	✓	
	FP	✓	?	✓	✓	✓	?
	SA		✓	✓			
Remedial Action	Alert	✓	✓	✓	✓	✓	
	Block	✓	✓	✓	✓		✓
	Encrypt	✓		✓			✓
	Audit	✓	✓	✓	✓	✓	✓
	Quarantine	✓	✓	✓	✓		
Available in:	Software	✓			✓	✓	✓
	Appliance		✓	✓			
Special Features		Detecting data within images and encrypted files+ (DLP drip)	Real time deep session inspection	Forensics analysis prior to the creation of rules	SSL inspection capabilities	Advanced contextual analysis	Very flexible deployment on mobile devices

Note: that all this information was obtained from the vendors' data sheets. Some of the features were not available for product privacy reasons. Therefore, in some cases we used '?' to denote uncertainty. The listed content analysis techniques are RE (regular expressions), FP (data fingerprinting) and SA (statistical analysis).

Table 2.1. List of Nominated DLP Solutions in 2015

CHAPTER 3: CYBER GUARD DATA LOSS PREVENTION

3.1. Cyber Guard

Cyber Guard monitors and protects the pilferage and exfiltration of sensitive organization data, either at rest or motion, thereby minimizing risk and protecting organizations of all sorts. Cyber Guard is a client server architecture, different from IDS/IPS and firewalls. The server regulates and enforces all organization-specific policies. The client's agent would perform content analysis on user data and communications using the server's policy-driven engine. This would empower organizations to track all the network traffic, monitor actions on endpoints, find the sensitive data on different data storage devices like severs and laptops across the organization and perform nee measures on the basis of per-defined policies to prevent data from leakage.

The Cyber Guard Data Loss Prevention solution includes the following components:

- Data at Rest
- Data in Use
- Data in Motion

3.1.1. Data at Rest

Data at rest module of Cyber Guard helps an organization to find sensitive content spread across the organization's network e.g. data on laptops, desktops, and servers. It searches the company's entire network and checks file contents according to company's defined policies and in case of any irregularity it alerts the administrator and takes action.

This module main functionality is content discovery so it finds where the sensitive data is stored in database servers, storage network or file databases of endpoints. When these sensitive files are found DLP analyzes the contents of these files according to defined policies and take actions.

Functionalities of data at rest

1. Scan endpoint(windows) file system using an agent
2. Scan database (MySQL) no agent required

3.1.2. Data in Use

This Cyber Guard module also known as DLP Endpoint controls all the endpoints which are in mostly use of end users of organizations. It monitors data activity at endpoints for any inconsistent action with respect to policies, generates alert about that end user, and takes necessary action to stop and mitigate the data loss.

It helps an organization to find sensitive content spread across the organization's network and actions being performed on sensitive data like copy/paste, snipping tool, or changing sensitive file name. Organization define its policies about the data and DLP solution performs necessary counter measures against those unlawful actions to stop the leakage of data based on those policies so that it cannot be misused. All such actions are logged in the database and alerts are sent to the administrator.

Functionalities for data in use

1. Monitor and Block Copy and paste for USB and normal drives
2. Block Screenshot of sensitive data

3.1.3. Data in Motion

Data in motion identifies sensitive content when it is travelling in and out of the organization’s network like when traffic going on internet, and if any user send the content on network then according to company defined policies it is checked, if there is sensitive data necessary actions are taken and it is reported. This module of DLP continuously monitor network traffic.

Data which goes to network is mostly in the form of packets. Traffic on network is checked through proxy server, all the data in an organization goes in/out through proxy server which identifies and captures right packets which has sensitive content. Headers and contents of the HTTP/HTTPs request are identified. If sensitive data checked against pre-defined policies is detected flowing, it takes the action according to set rule and alert this violation.

The violations are logged in a database. The administrator can check reports to find which end user did that violation and logging of data provides necessary auditing in case of a security breach.

Functionalities for data in flow

1. Monitor data out (http/https)
2. Monitor attachments in web mail

CHAPTER 4: REQUIREMENTS SPECIFICATION

4.1. Functional Requirements

Major functional requirements of Cyber Guard DLP are

4.1.1. Real-Time Analytics

DLP is considered effective only if it generates real time alerts in some cases, notifications and analytics. So, this is very important requirement in the course of building DLP solution. These help security administrator to know about an attack/violation as early as possible and manual measures can also be taken if required, especially if the attack is related to very sensitive data.

4.1.2. Rules and Policy Management

This is the core feature of data loss prevention solution. Implementation of this requirement helps to create new rules and policies and enforce them on organization network. DLP has some general data related pre-defined rules and policies, however this helps DLP user to create new, modify old and enforce rules and policies according to own requirements. It should have a user-friendly graphical user interface so that even a lay man can use it easily.

It will help in defining which data is sensitive and how to protect it. The monitoring of end-point ands action in case of any violations should all be defined in policies and rules. The admins should be able to see through management server which policy and rules was violated in unauthorized action.

This is important because variety of data is generated daily in organizations. So, every data cannot be protected by same method, a new method is always required so this policy management will help in protecting this new data.

4.1.3. Admin Management

A central management server/ interface or enforcement platform is very essential for admins which are managing the security data through. DLP should have a user-friendly interface for dashboard because it is not always used by technical persons even the executives of company or business or finance department can also use it if required, so that even lay man (non- technical persons) can use it easily. It should be accessible only to admins.

4.1.4. Content Analysis

The feature of analyzing content is used to identify sensitive data and various techniques can be used to analyze content by DLP and then sensitive data can be protected according to measures defined by policies. In this requirement various techniques should be specified for content analysis so that data can be protected by DLP.

Some of the techniques which can be used to analyze content are Regular Expression Matching, Key Words Matching, Pre-defined data identifiers and File Content Matching. It is essential because any solution cannot protect data if it does not know about data. e.g. a person is sending a mail and there is an attachment, DLP cannot tell whether the attachment contains sensitive content or not until it analyses it.

4.1.5. End-Point Monitoring

The solution should be able to control all the endpoints of organizations. It should monitor data activity at endpoints for any inconsistent action with respect to policies, generates alert about that end user, and takes necessary action to stop and mitigate the data loss.

4.1.6. Data at Rest Monitoring

This requirement enables DLP to find sensitive content spread across the organization’s network e.g. data on laptops, desktops, and servers. It will search the company’s entire network and checks file contents according to company’s defined policies and in case of any irregularity it will alert the administrator and take action.

4.1.7. Data in Motion Monitoring

This will identify sensitive content when it is travelling in and out of the organization’s network in the form of http/https traffic on local and external network, and if any user send the content on network then according to company defined policies it will be checked, and if there is sensitive data necessary actions will be taken and reported. This feature of DLP will continuously monitor http/https network traffic.

4.1.8. Logging

DLP server will log all the incidents in which policy violation occurs. It will record the following information: Date and time of the violation, type of policy violated, end-user, the damage caused and the actions taken. These logs help admins to know about security gaps and will help in making policies better so that no data is leaked next time.

4.2. Non-Functional Requirements

Non-functional requirements of cyber guard are specified below:

4.2.1. Usability requirements:

The graphical user interface of the system should be friendly and interactive so that even a lay man can use it easily because sometimes executives of the company ask for management console to use or business, finance and other departments can also request to use it if required.

As the management console is web based so there should be minimum delay to access it because the delay access time is not very good thing for security-based solutions.

4.2.2. Performance Requirements

The communication link of end-point service(which monitors end-points), proxy server (which monitors network traffic) and database servers with central management server should be very good, because in case of any violations the admins should be immediately informed and if there is too much lag then system will not be an effective one. So, configuration of the system should also be good to achieve high performance.

4.2.3. Safety and Security Requirements

Management console should be password protected and the password must be changed regularly.

Only admins can access the management console with password for viewing, configuring and managing purposes.

The server on which management server runs should be in a secure location so no one un-authorized can access it.

Information transmission from end-point, proxy server and database servers with central management server should be secure and should not be tempered.

4.2.4 Software Quality Attributes

The data loss prevention system should be up and running 24/7 because it has the most important work in organization which is to protect the data. There should be proper measures to handle this like backup generators and servers, in case one server goes down or electricity goes off then system should not stop its working.

The implemented system should be very accurate and reliable in its workings because this is very sensitive job. If important data of organizations get leaked it can cause a lot of damage to them. So, it should be properly tested before deployment.

The system should be easy to use. It can be easily deployed in an organization. Its installation and configuration should be very simple.

CHAPTER 5: SYSTEM ARCHITECTURE

5.1. Overview

Cyber Guard DLP protects data through all the stages it moves, from its generation (first step) or record to its eventual registry and/or removal at the end of its functional life. As discussed in Chapter 3 Cyber Guard DLP includes three main components:

- Data at Rest
- Data in Motion
- Data in Use

5.2. Architectural Design

Following diagrams illustrates the System architecture of Cyber Guard DLP.

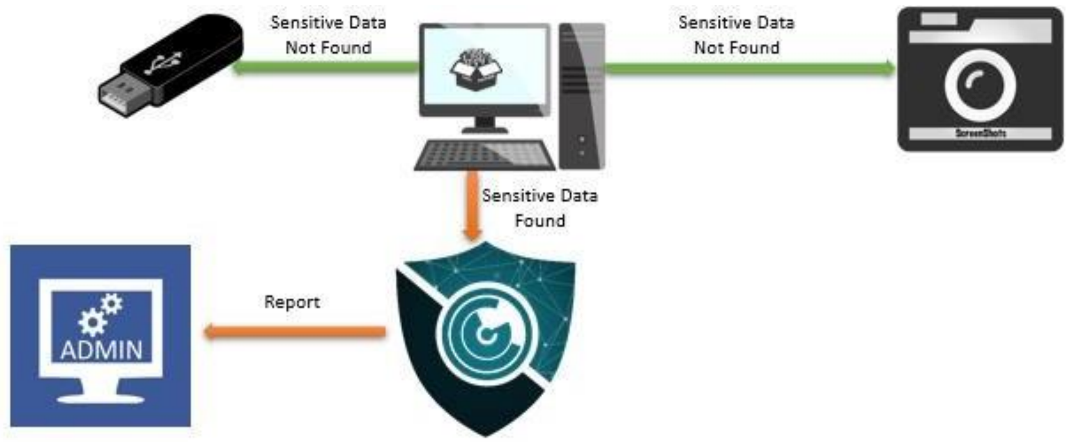


Figure 5.1. Architecture Diagram of Data in Use

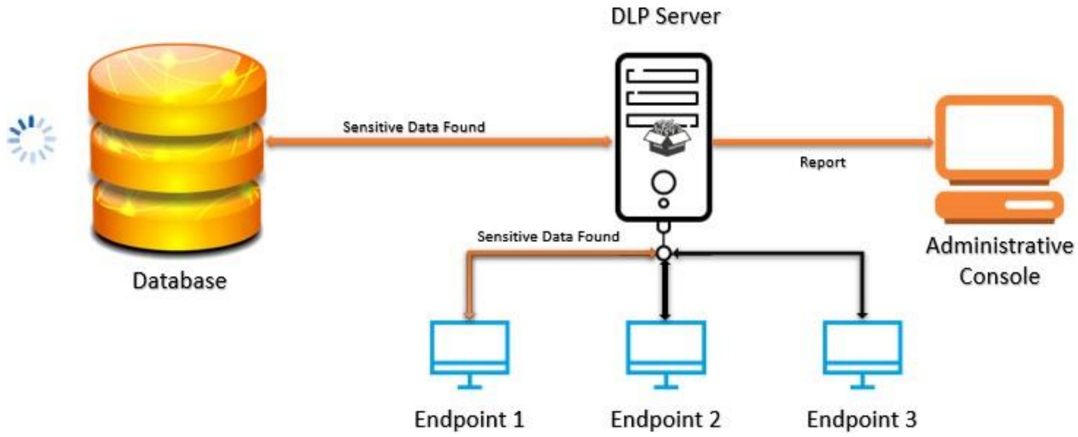


Figure 5.2. Architecture Diagram of Data at Rest

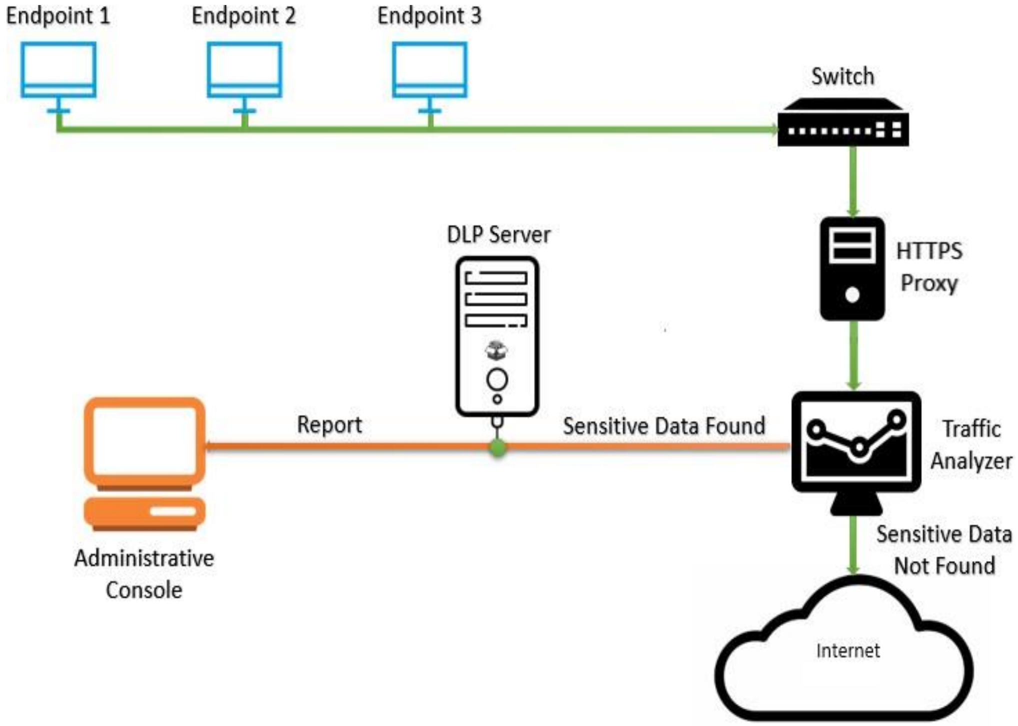


Figure 5.3. Architecture Diagram of Data in Motion

5.3. Decomposition Description

5.3.1. Data in Motion

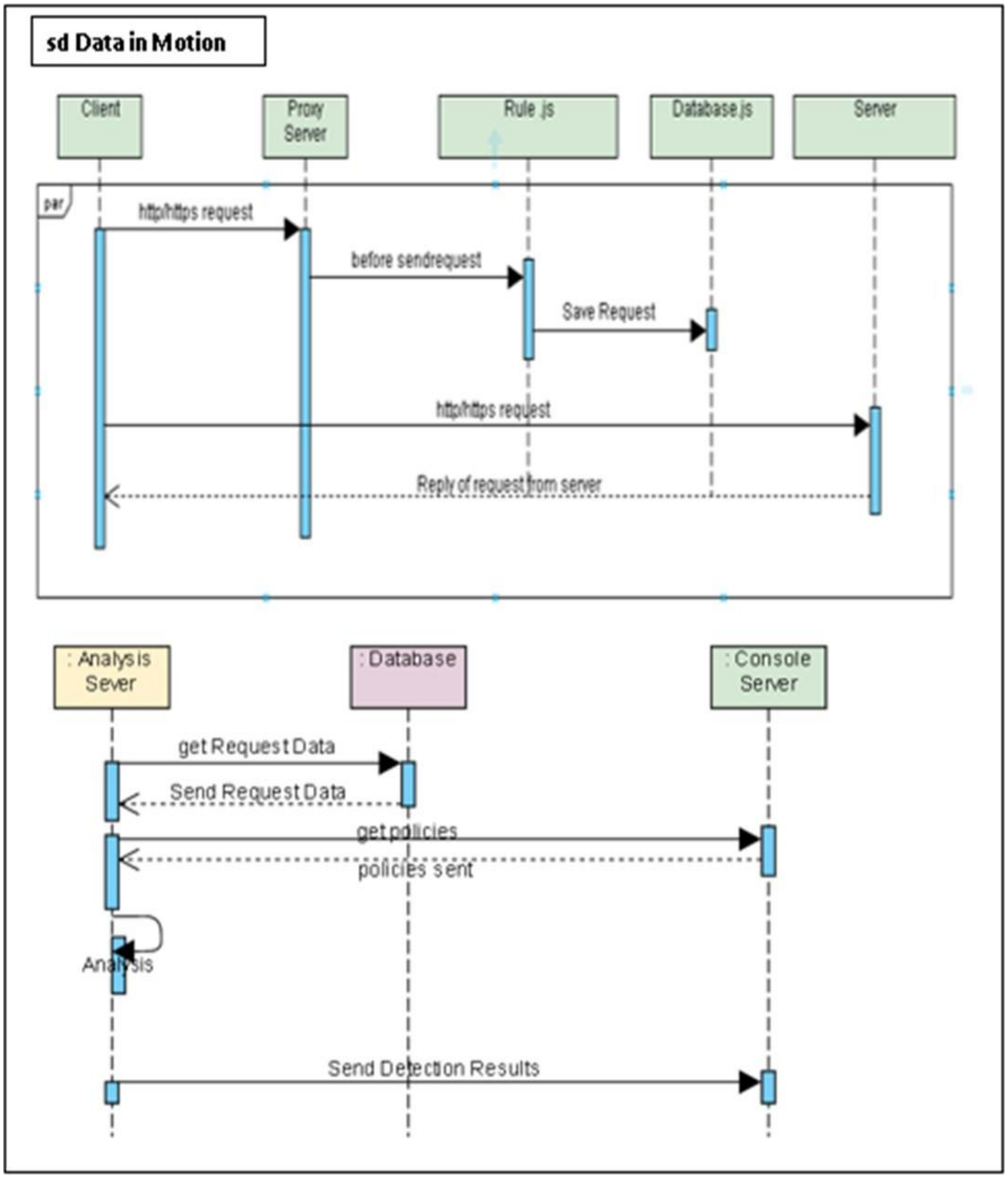


Figure 5.4. Sequence Diagram of Data in Motion

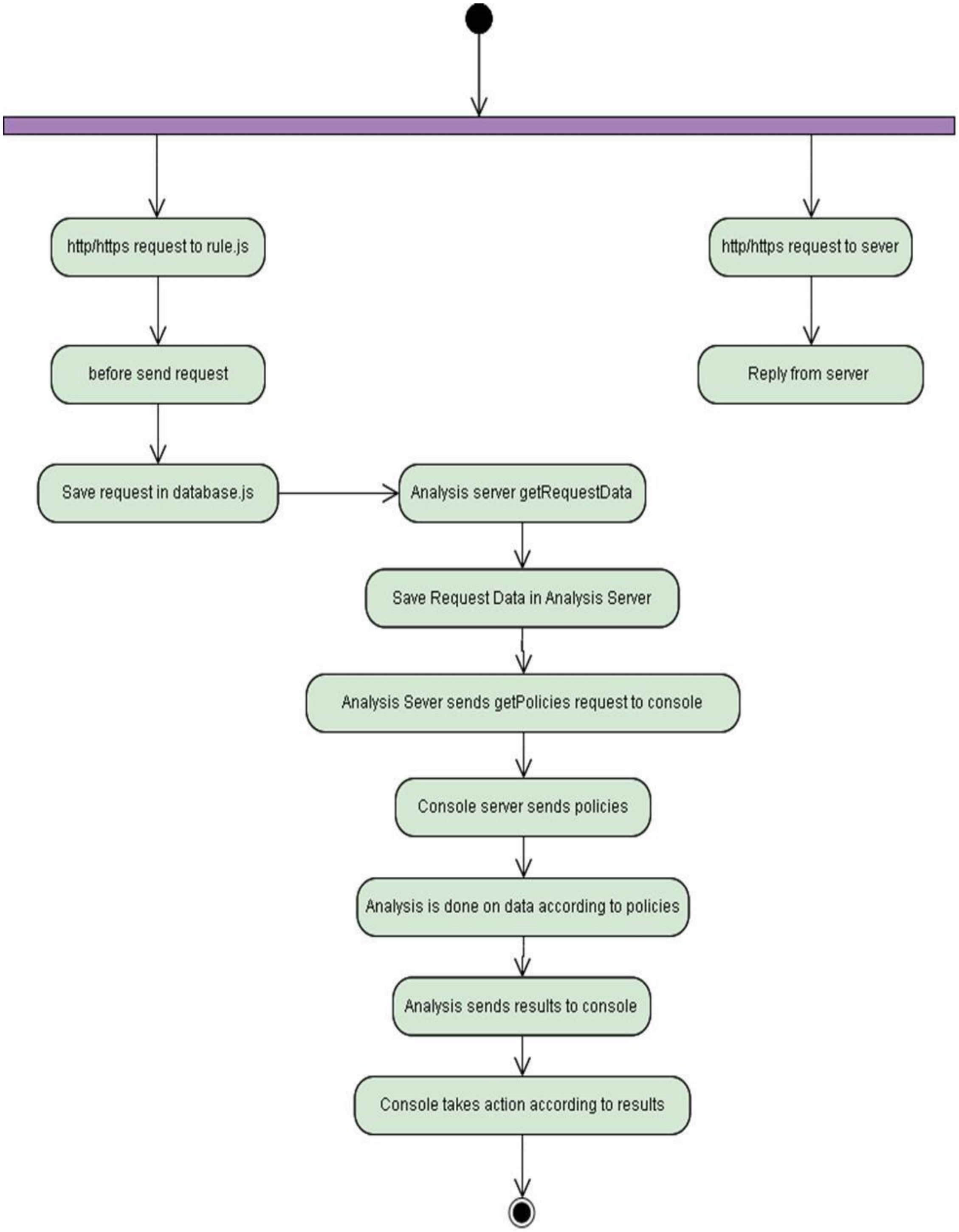

http/https request to rule.js
before send request
Save request in database.js
http/https request to sever
Reply from server
Analysis server getRequestData
Save Request Data in Analysis Server
Analysis Sever sends getPolicies request to console
Console server sends policies
Analysis is done on data according to policies
Analysis sends results to console
Console takes action according to results

Figure 5.5. Activity Diagram of Data in Motion

5.3.2. Data at Rest

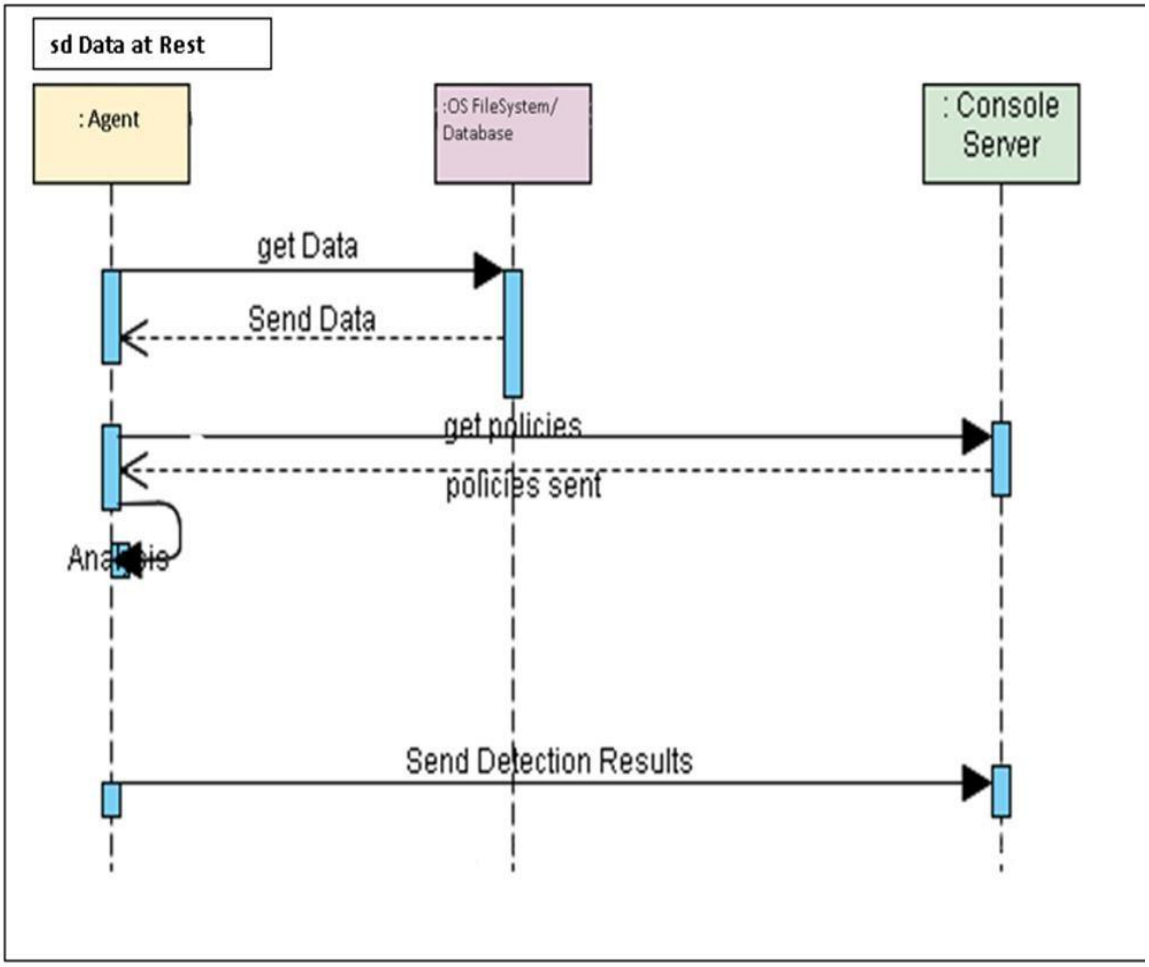


Figure 5.6. Sequence Diagram of Data at Rest

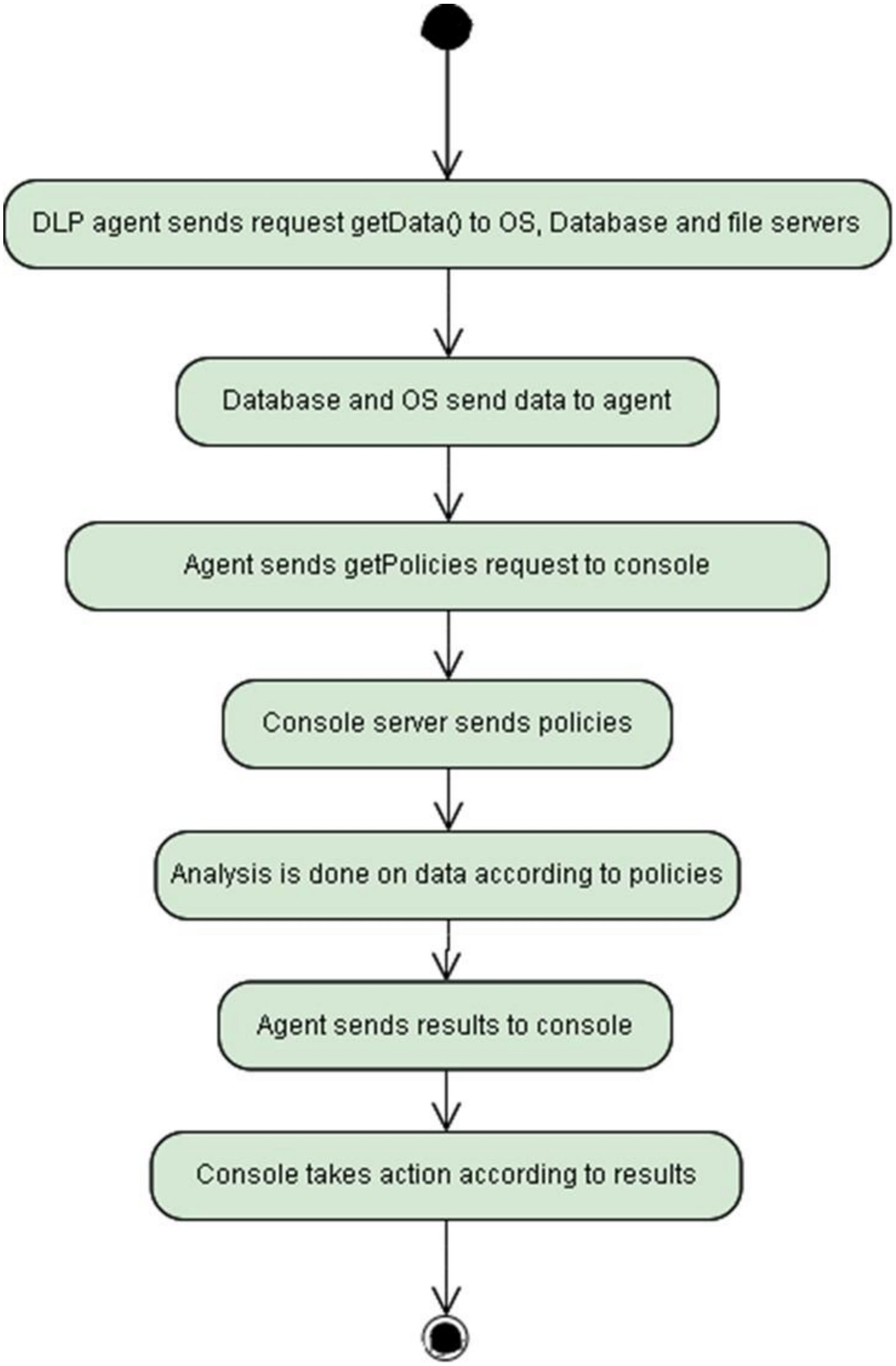

DLP agent sends request getData() to OS, Database and file servers
Database and OS send data to agent
Agent sends getPolicies request to console
Console server sends policies
Analysis is done on data according to policies
Agent sends results to console
Console takes action according to results

Figure 5.7. Activity Diagram of Data at Rest

5.3.3. Data in Use

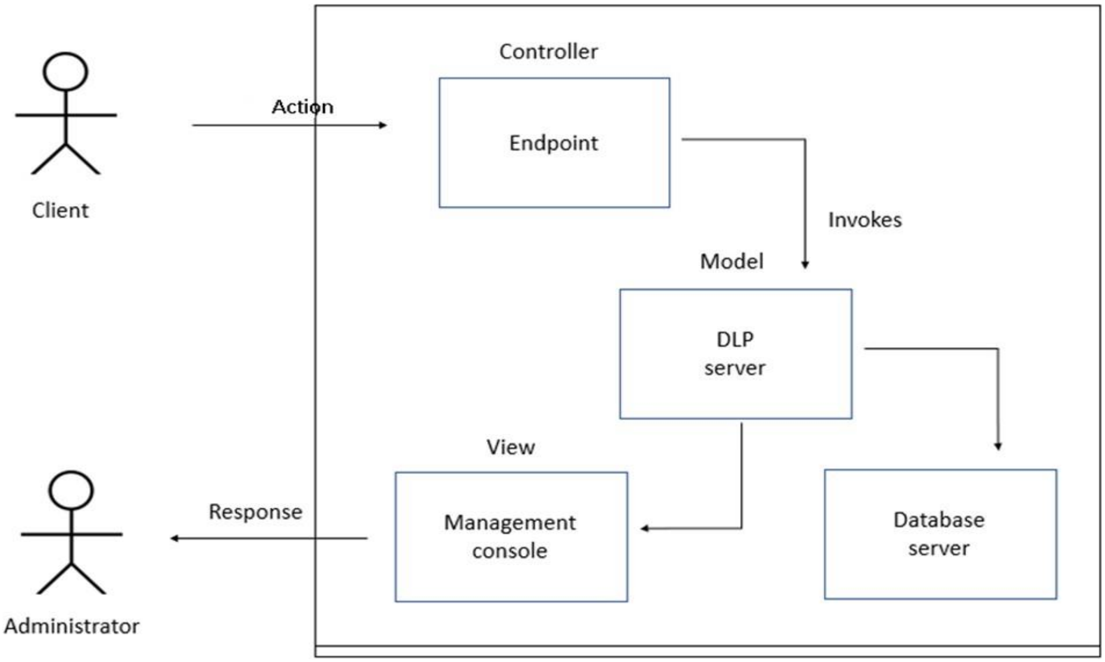


Figure 5.8. Block Diagram of Data in Use

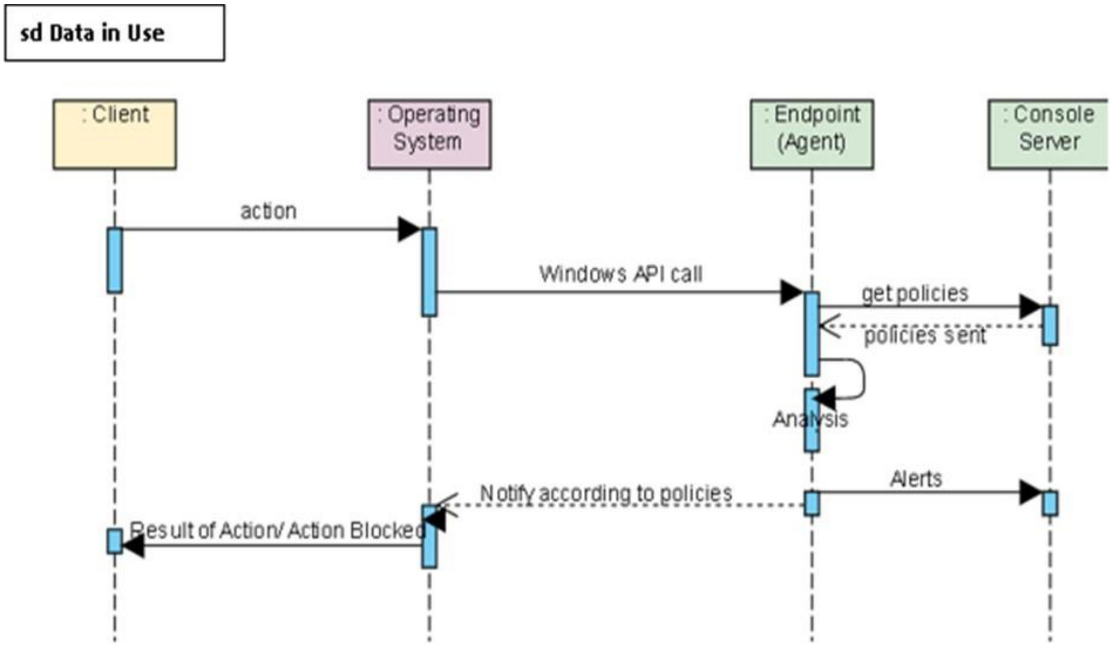


Figure 5.9. Sequence Diagram of Data in Use

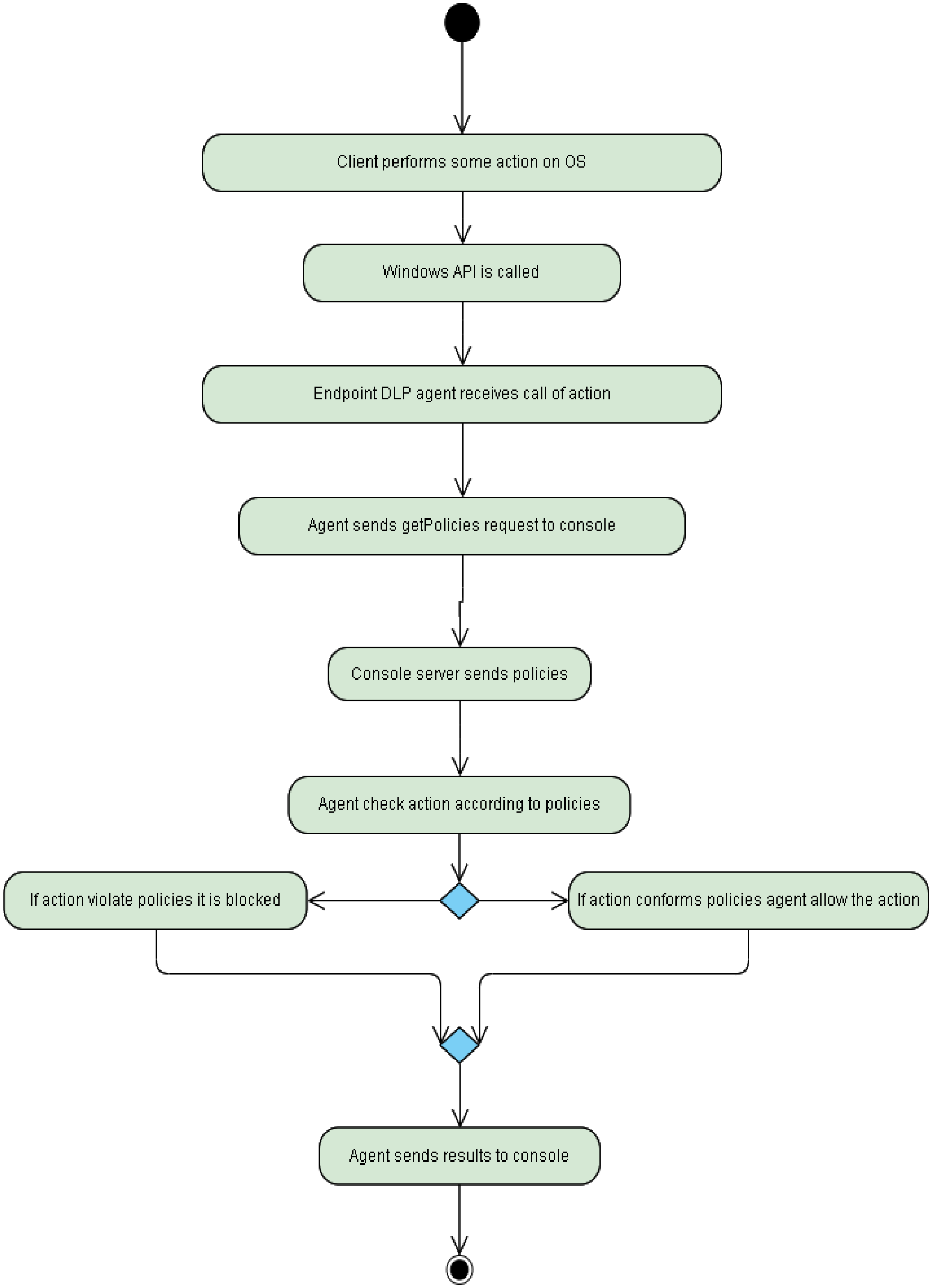

Client performs some action on OS
Windows API is called
Endpoint DLP agent receives call of action
Agent sends getPolicies request to console
Console server sends policies
Agent check action according to policies
If action violate policies it is blocked
If action conforms policies agent allow the action
Agent sends results to console

Figure 5.10. Activity Diagram of Data in Use

5.3.4. DLP Enforce Platform

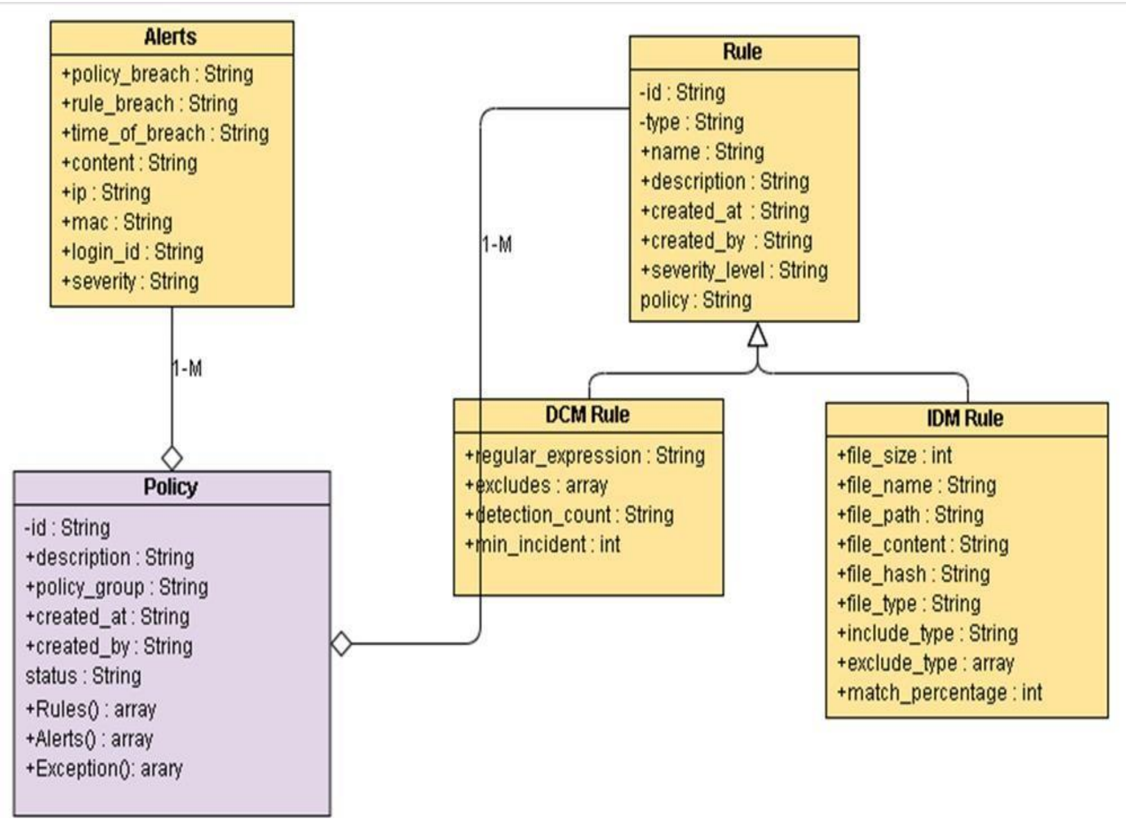


Figure 5.11. Class Diagram of Management Console

5.4. Data Design

5.4.1. Data Description

The database used is MongoDB. It is a cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with schema.

5.4.2. Data Dictionary

User Table

Field	**Type**	**Description**
UserID	Integer	Primary Key
IPaddress	Integer	IP address of computer
PCname	Text	Name of user device.
AccountName	Text	Logged on user

Table 5.1. Data Fields for End_user

Agent Table

Field	**Type**	**Description**
AgentID	Integer	Primary key
AgentVer	Integer	The version of software installed on user's device.
AgentStatus	Text	Whether the agent is running or not.

Table 5.2. Data Fields for Agent

Policy Table

Field	**Type**	**Description**
PolicyID	Integer	Primary key
PolName	Text	Name of policy
RuleType	Text	Whether its related to web traffic, data storage etc.
Action	Text	How to tackle policy violation.

Table 5.3. Data Fields for Policy

Administrator Table

Field	**Type**	**Description**
AdminID	Integer	Primary key
AdminName	Text	Name of administrator
PCname	Text	Name of device administrator uses to log in.

Table 5.4. Data Fields for Administrator

Storage Table

Field	Type	Description
Inf_type	Text	The kind of sensitive information (ID, password etc.)
File_type	Text	The kind of file (.doc, .xls) of information.

Table 5.5. Data Fields for Storage

Logs Table

Field	Type	Description
Date	Integer	Primary key
Time	Timestamp	Time of violation
Source	Text	Name of device and IP address.
Destination	text	Destination of violation

Table 5.6. Data Fields for Logs

FOR WEB TRAFFIC:

Request Detail:

Data	Type
_Id	String
Request_data	Object

timestamp	Time
URL	String
Request_options	Object
protocol	String

Table 5.7. Data Fields for Request Details of Web Traffic

Request options:

Data	**Type**
hostname	String
port	Int
path	String
method	String
headers	Object

Table 5.8. Data Fields for Request Options of Web Traffic

Headers:

Data	**Type**
accept	String
Accept-encoding	String
Accept-language	String
Cache-control	String
connection	String
User agent	String

Table 5.9. Data Fields for Headers of Web Traffic

Request data:

Data	**Type**
data	Array of bytes or int
type	String

Table 5.10. Data Fields for Request data of Web Traffic

CHAPTER 6: WORKING OF CYBER GUARD

6.1. Summary Screen

The display screen of **Summary** is a robust dashboard of Cyber Guard that presents management and analytical summaries of sensitive data related incidents and reports.

Data in an organization is very much distributed as there are hundreds of endpoints in a corporate and different storage devices and servers. So, it is very essential that policies are properly defined and enforces on these environments and properly checked. The major feature of DLP is its management console which is very unified, perform different types of analysis on data that help in forming best policies but major feature is that it is enforcement platform for policies on every device thus helps in preventing data loss. It helps you to monitor day to day activities like alerts, user audits, access log and policy violation and provide business intelligence.

6.1.1. Web Console Access

By entering the IP-address of server machine running the management server and port number in any browser web console can accessed. After entering login screen prompts which can be seen in figure below. Users can login by entering credentials of username and password and access the dashboard.

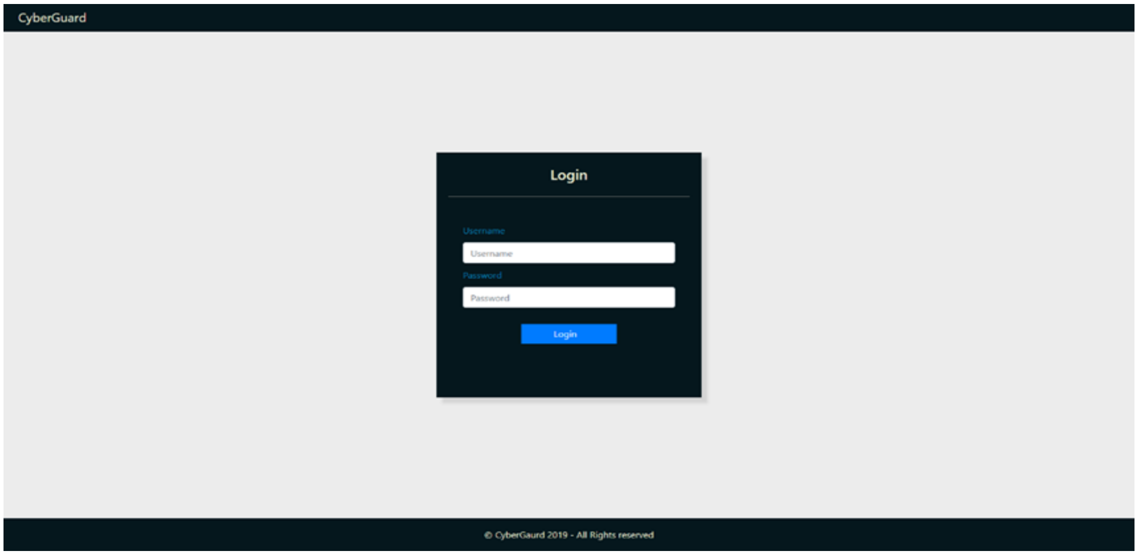


Figure 6.1. Login Screen

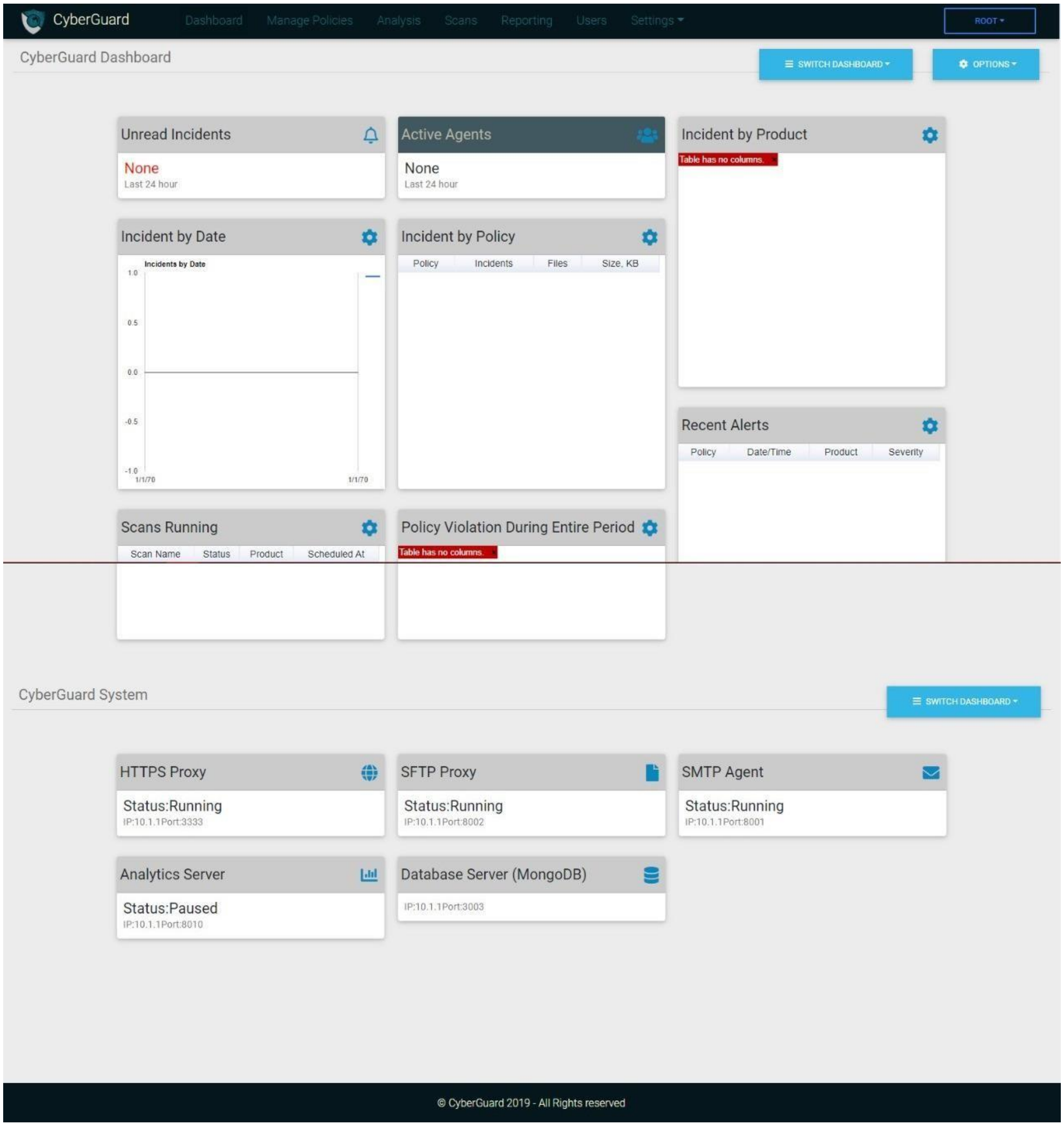
The user of this console are administrators and power users. These people are decided by a company and are given authorization and privileges according to their roles which they are handling in DLP.

The management console has very high privileges, so it is necessary for its users(administrators) to log out of the system before leaving the system

6.1.2. Web Console Navigation

Summary screen appears as default page after logging in to web console. This web-page offers a comprehensive view of all the important matters like recent information leaks and which actions are taken on them, which endpoints were violating policies recently and type of violations and quick access to different tabs and settings.

Administrators view any system’s health (incidents and sensitive data) and the alerts on violations from dashboard and take necessary actions on them easily through web console. Web Console helps administrators to see the incidents by hostname and policy category so that they can easily identify where are the most risks. The dashboard of Cyber Guard is shown in figure 6.2.



CyberGuard Dashboard

Unread Incidents
None
Last 24 hour

Active Agents
None
Last 24 hour

Incident by Product
Table has no columns.

Incident by Date

Incident by Policy

Policy	Incidents	Files	Size, KB

Recent Alerts

Policy	Date/Time	Product	Severity

Scans Running

Scan Name	Status	Product	Scheduled At

Policy Violation During Entire Period
Table has no columns.

CyberGuard System

HTTPS Proxy
Status:Running
IP:10.1.1Port:3333

SFTP Proxy
Status:Running
IP:10.1.1Port:8002

SMTP Agent
Status:Running
IP:10.1.1Port:8001

Analytics Server
Status:Paused
IP:10.1.1Port:8010

Database Server (MongoDB)
IP:10.1.1Port:3003

© CyberGuard 2019 - All Rights reserved

Figure 6.2. Dashboard Screen

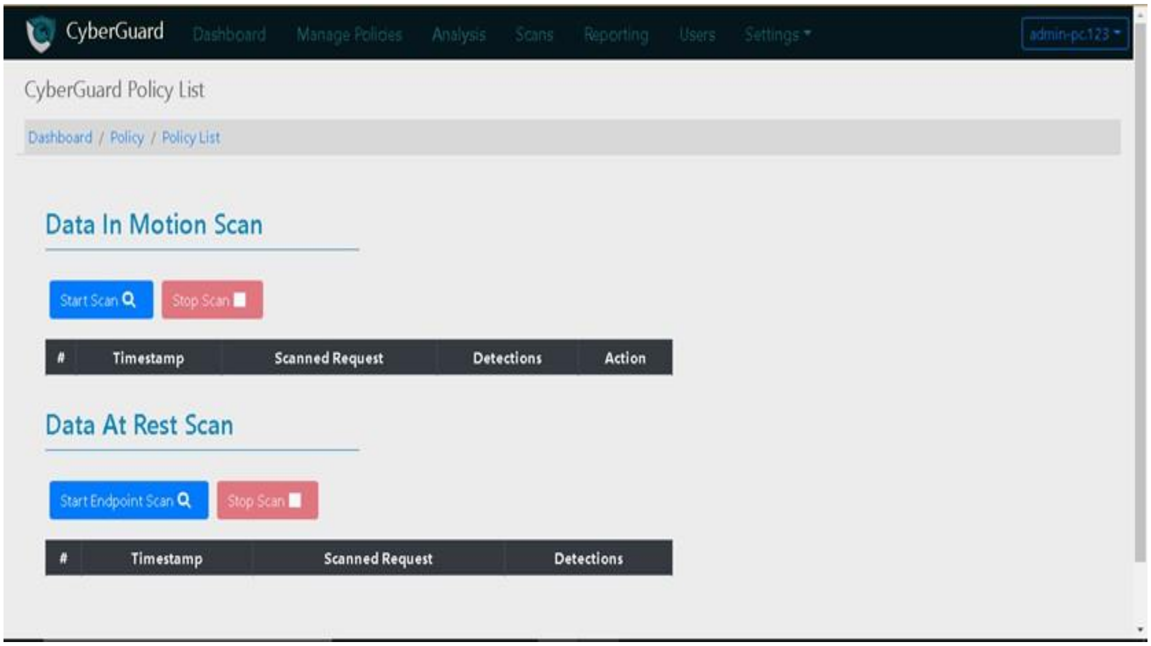
6.2. Functional Areas of Management Console

6.2.1. Analysis

There is an analytics engine in Cyber Guard DLP whose purpose is to find and prioritize risks incidents., Incidents which are of high priority are displayed on main dashboard. The incidents across all the endpoints due to defined policies. These incidents are processed to generate reports which tells about the data theft according to priority.

6.2.2. Scans

Policies define sensitive data identification rules, then the Cyber Guard searches for sensitive data during DLP scanning process and then take action when the content is found. Scanning in DLP is automatically done after some time, but anyone can manually scan the all endpoints or a specific endpoint by selecting the SCANS tab.



CyberGuard Policy List

Data In Motion Scan

#	Timestamp	Scanned Request	Detections	Action

Data At Rest Scan

#	Timestamp	Scanned Request	Detections

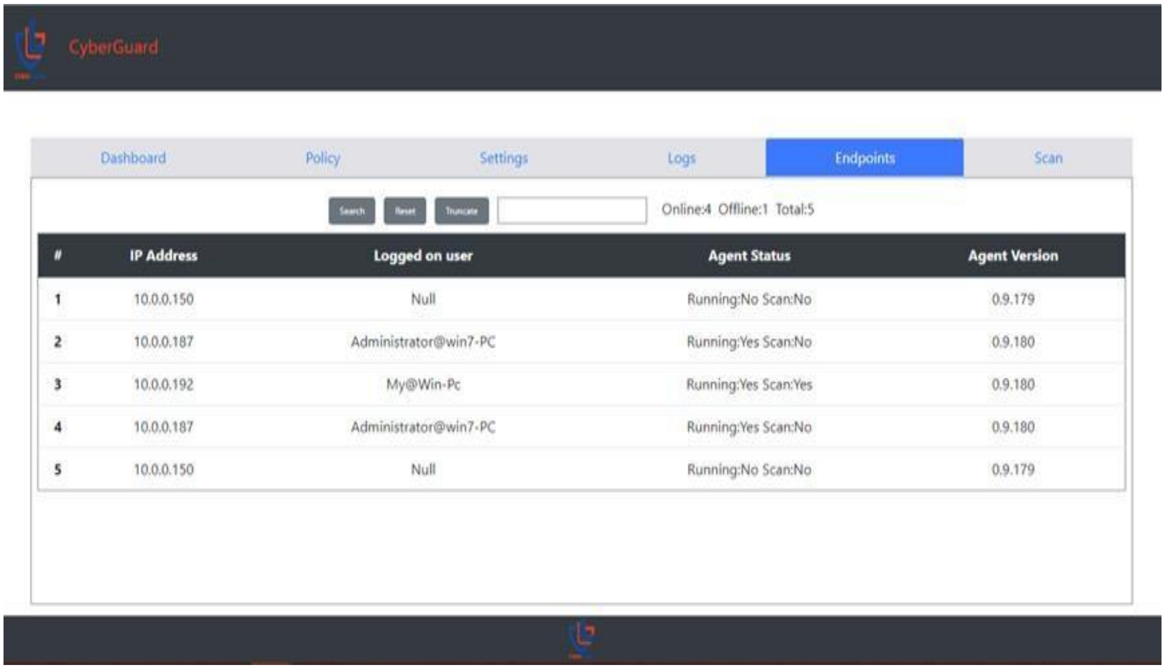
Figure 6.3. Scans Screen

6.2.3. Users and Endpoints

There are different endpoints in an organization, one of the main features of DLP is to monitor continuously the activities of users on these endpoints. DLP endpoint runs as a service on every endpoint to monitor according to predefined policies so that if user takes any step which is leading to data loss, action will be taken according to policies and reports to management console. And the second feature is that DLP scans endpoints for sensitive user data on automated basis and checks it against policies.

The Endpoint Status screen shows the results of the scans run on end points and also shows the very descriptive information for a specific endpoint.

Navigate screen > Endpoints screen appears.



Online:4 Offline:1 Total:5

#	IP Address	Logged on user	Agent Status	Agent Version
1	10.0.0.150	Null	Running:No Scan:No	0.9.179
2	10.0.0.187	Administrator@win7-PC	Running:Yes Scan:No	0.9.180
3	10.0.0.192	My@Win-Pc	Running:Yes Scan:Yes	0.9.180
4	10.0.0.187	Administrator@win7-PC	Running:Yes Scan:No	0.9.180
5	10.0.0.150	Null	Running:No Scan:No	0.9.179

Figure 6.4. Endpoints Screen

6.2.4. Administration and Settings

Use these tabs to specify Cyber Guard DLP specific settings. An administrator user can configure and/or manage the server configuration, agents, crawlers, data, endpoints, user accounts and roles according to his requirements.

6.2.5. Servers and Proxies Information

The real time visualization of the servers and proxies which are running.

- Servers
  - Analytical Server
  - Database Server
- Proxies

  HTTPS proxy

Cyber Guard DLP gives information that which servers and proxies are running on which IP address and port. It also gives information which is running and which is not and one can access their settings by clicking it. We can also scan database server for sensitive data anytime otherwise it is automated. Proxies deal with the sensitive data which is going through network.

6.2.6. Rules & Policies

There is so much diversity in data within only one organization and it is very much difficult to identify this sensitive data according to only one organization. Once one organization has defined almost all digital asset and its plan to stop, then different specific rules are defined for data detection like data identifiers, templates, policies etc.

- **Rules**

The process of automated data loss prevention and finding potential threats is done by rules defined by Cyber Guard which are key part to these activities. Rules help to detect un-authorized events in real time and sometimes take action against these events. Cyber Guard DLP has integrated Intelligent Rules Engine

- These rules in combine with Cyber Guard's every action monitoring feature, such as: files, and web traffic etc. to decide which action or content the rule should consider confidential.
- One can form different sets of rules for different level of clearance of people.
- The monitoring can be set to different rules like one can monitor an endpoint according to only one specific rule.
- Cyber Guard Agent enforces the rules on endpoints which are deployed on DLP management server and run as service on endpoint.

New rules can be easily created according to company's data. When one rule is violated an alert is generated on cyber guard management console.

A complete record of every violation event with its metadata is saved in cyber guard server. Reports are generated on schedule basis and these reports can be seen from the Alerts screen.

Data Detection Rules

1. Regular Expression

Cyber Guard recognizes sensitive digital assets according to a set of given expressions, which is given as input to DLP given the type of data of the organization where it is being used. Expression matching has more efficiency in case of structured data, such as ATM & credit card numbers, CNIC, or phone numbers.

2. Match Key-Word List

Keywords are unique words or phrases. For example, "Drugs", "Overdose", "Murder", and "steroids" are some keywords that someone can think of sensitive in medical field. Most common

CyberGuard also recognizes sensitive assets using a predefined list and user can also define its keyword list according to its organization data.

3. Predefined data identifiers

Cyber Guard Data Loss Prevention solution helps to recognize and verify sensitive data which is in a pattern, helping various system defined data identifiers. Cyber Guard Data Loss Prevention provides predefined data identifiers such as (Mastercard, Visa, IBAN, CNIC, SSN, Email, URL.) so that sensitive content can be identified according to them. It is very efficient and time saving technique because it is very easily implemented and the data is matched with respect to very short type of data.

These are basically algorithms that integrate content matching with key-words validators (luhn check, mod 10 algorithm, IBAN validator) to identify sensitive data. Predefined data identifiers are like RegExr but they are more effective because they have ability to match the contents more accurately and reduce false positives. Key word validators are measure to ensure the accuracy that checks how much it is detecting sensitive data and make sure that it works properly.

4. Match File Contents

Cyber Guard gives many conditions to match the content of messages. Many of these conditions need a reference data profile and index.

The file content matching rules only work for Data in Flow and Data at Rest. Files to be monitored and protected should be provided to the server (by creating a "Indexed Document Matching" rule) so that they are index and fingerprinted by the server. Every content passing through the network or at rest is monitored against these files specified in a specific rule.

Input of sensitive data is given to Cyber Guard servers and agents in the form of messages. Type of message is identified by Cyber Guard e.g. a text document. According to the type of message Cyber Guard either divide the message content into different components (header, subject, body, attachments) or just leave the message as it is. DLP checks the message or its components against match conditions. If a condition is fulfilled and it also meet the requirement of sensitive content then necessary action is taken.

• Policies

After installing and configuring DLP system settings, rules are defined properly which help in detecting sensitive data, after all this we are ready to form a policy.

Information loss prevention policies define parameters on which data can be shared within organization and outside the organization meanwhile protecting it. These policies help in making data useful to people whom it is concerned while preventing it from people who are not authorized to use it.

The policies screen can be used to integrate policies with rules templates to prevent sensitive data loss. It is the best practice that one first creates such policy that covers most of the strictest rules and it covers as many endpoints it can. Then policies should be created for left out or as exception from all users.

From the Dashboard main tab,

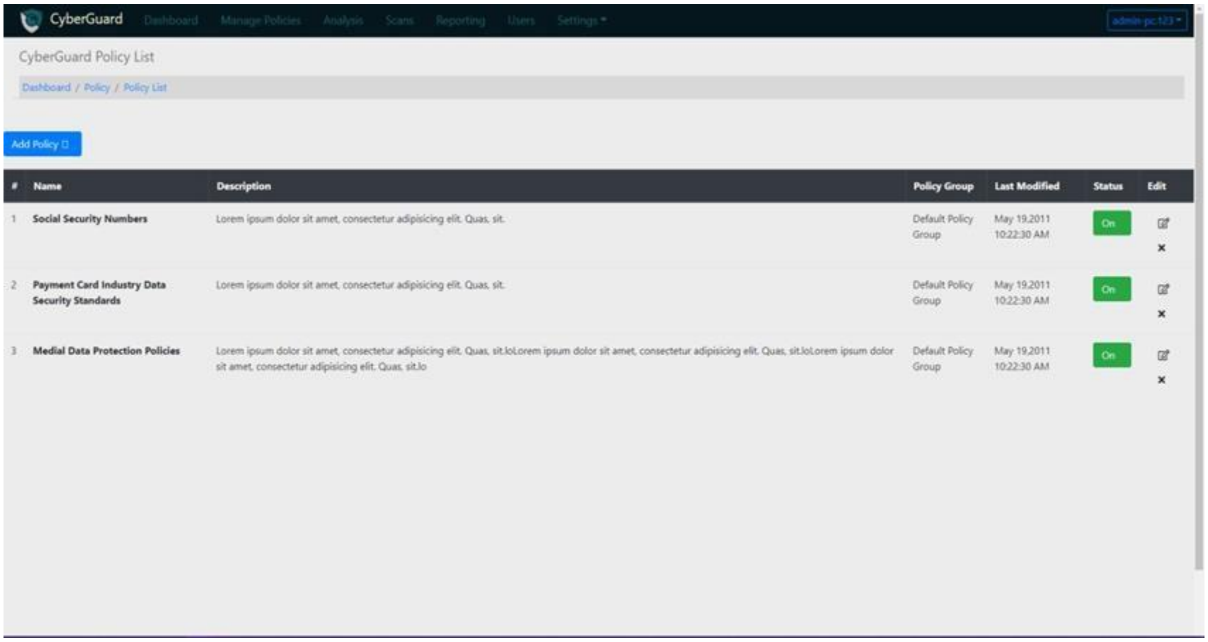
select > Manage Policies

Here all the pre-implemented policies can be seen and edited and new policies can be implemented.

Managing Policies

1. **Navigate to Dashboard > Manage Policies.**

Here we can see the already implemented policies and we can also edit them here.



CyberGuard Policy List

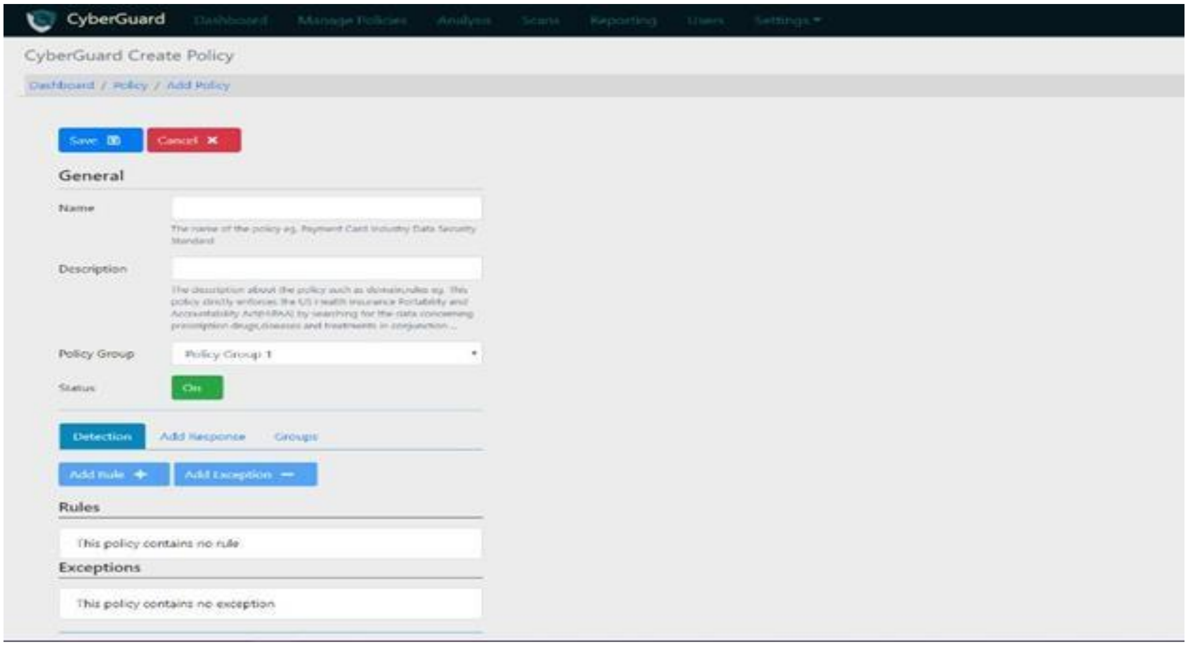
Dashboard / Policy / Policy List

#	Name	Description	Policy Group	Last Modified	Status	Edit
1	**Social Security Numbers**	Lorem ipsum dolor sit amet, consectetur adipisicing elit. Quas, sit.	Default Policy Group	May 19,2011 10:22:30 AM	On	✎ ×
2	**Payment Card Industry Data Security Standards**	Lorem ipsum dolor sit amet, consectetur adipisicing elit. Quas, sit.	Default Policy Group	May 19,2011 10:22:30 AM	On	✎ ×
3	**Medial Data Protection Policies**	Lorem ipsum dolor sit amet, consectetur adipisicing elit. Quas, sit.loLorem ipsum dolor sit amet, consectetur adipisicing elit. Quas, sit.loLorem ipsum dolor sit amet, consectetur adipisicing elit. Quas, sit.lo	Default Policy Group	May 19,2011 10:22:30 AM	On	✎ ×

Figure 6.5. Pre-defined Policy List Screen

But to add new policies navigate

2. Click Add > Policy



CyberGuard Create Policy

General

Name

The name of the policy eg. Payment Card Industry Data Security Standard

Description

The description about the policy such as domain,rules eg. This policy directly enforces the US Health Insurance Portability and Accountability Act(HIPAA) by searching for the data concerning prescription drugs,diseases and treatments in conjunction ...

Policy Group: Policy Group 1

Status: On

Rules

This policy contains no rule

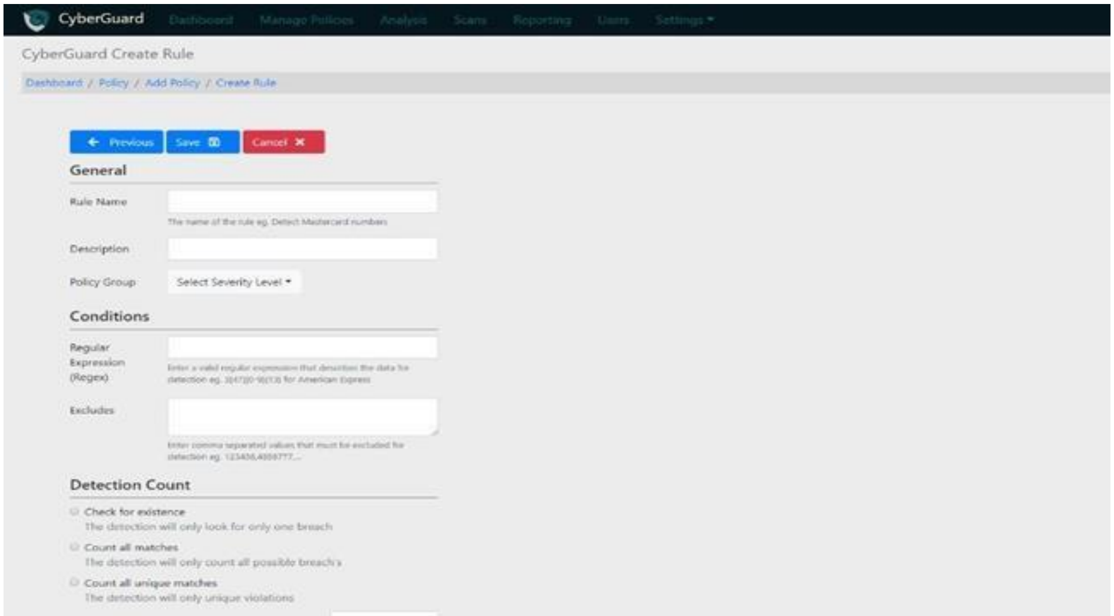
Exceptions

This policy contains no exception

Figure 6.6. Create Policy Screen

3. Click add >Rule

To create the rules of that policy which we implemented. The screens are below



CyberGuard Create Rule

General

Rule Name

The name of the rule eg. Detect Mastercard numbers

Description

Policy Group: Select Severity Level ▾

Conditions

Regular Expression (Regex)

Enter a valid regular expression that describes the data for detection eg. 3[47][0-9]{13} for American Express

Excludes

Enter comma separated values that must be excluded for detection eg. 123456,4559777...

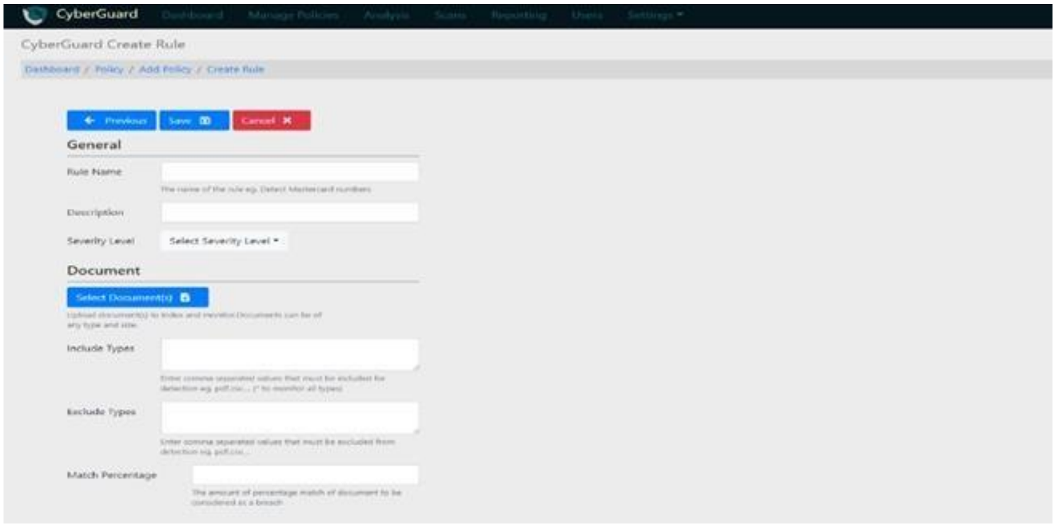
Detection Count

○ Check for existence
The detection will only look for only one breach

○ Count all matches
The detection will only count all possible breach's

○ Count all unique matches
The detection will only unique violations

Figure 6.7. Create Rule Screen 1



CyberGuard Create Rule

Dashboard / Policy / Add Policy / Create Rule

General

Rule Name

The name of the rule eg. Detect Mastercard numbers

Description

Severity Level

Document

Upload document(s) to index and monitor. Documents can be of any type and size.

Include Types

Enter comma separated values that must be included for detection eg. pdf,csv... (* to monitor all types)

Exclude Types

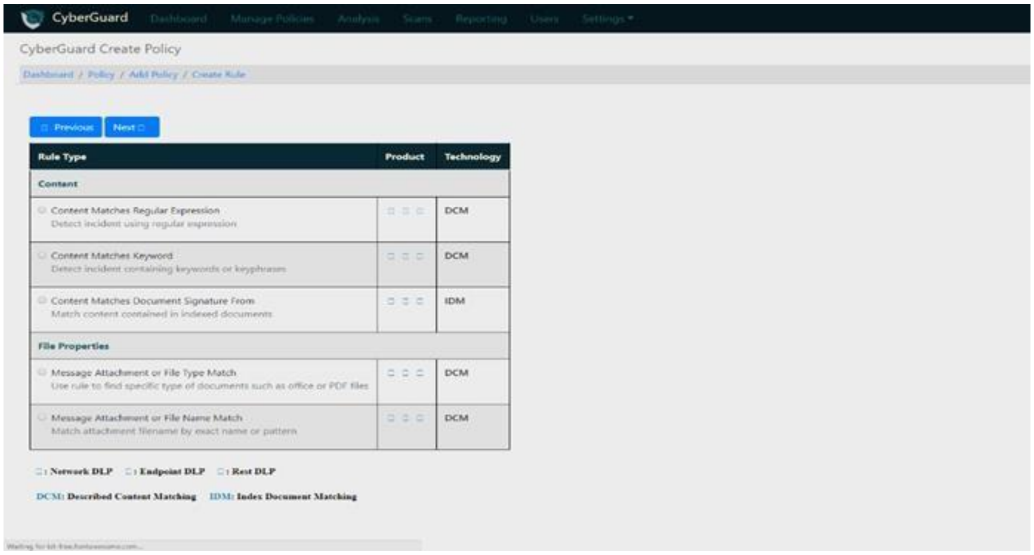
Enter comma separated values that must be excluded from detection eg. pdf,csv...

Match Percentage

The amount of percentage match of document to be considered as a breach

Figure 6.8. Create Rule Screen 2

The screen of all the implemented rules of particular policy DLP



CyberGuard Create Policy

Dashboard / Policy / Add Policy / Create Rule

Rule Type	Product	Technology
Content		
○ Content Matches Regular Expression Detect incident using regular expression	□ □ □	DCM
○ Content Matches Keyword Detect incident containing keywords or keyphrases	□ □ □	DCM
○ Content Matches Document Signature From Match content contained in indexed documents	□ □ □	IDM
File Properties		
○ Message Attachment or File Type Match Use rule to find specific type of documents such as office or PDF files	□ □ □	DCM
○ Message Attachment or File Name Match Match attachment filename by exact name or pattern	□ □ □	DCM

□ : Network DLP □ : Endpoint DLP □ : Rest DLP

DCM: Described Content Matching IDM: Index Document Matching

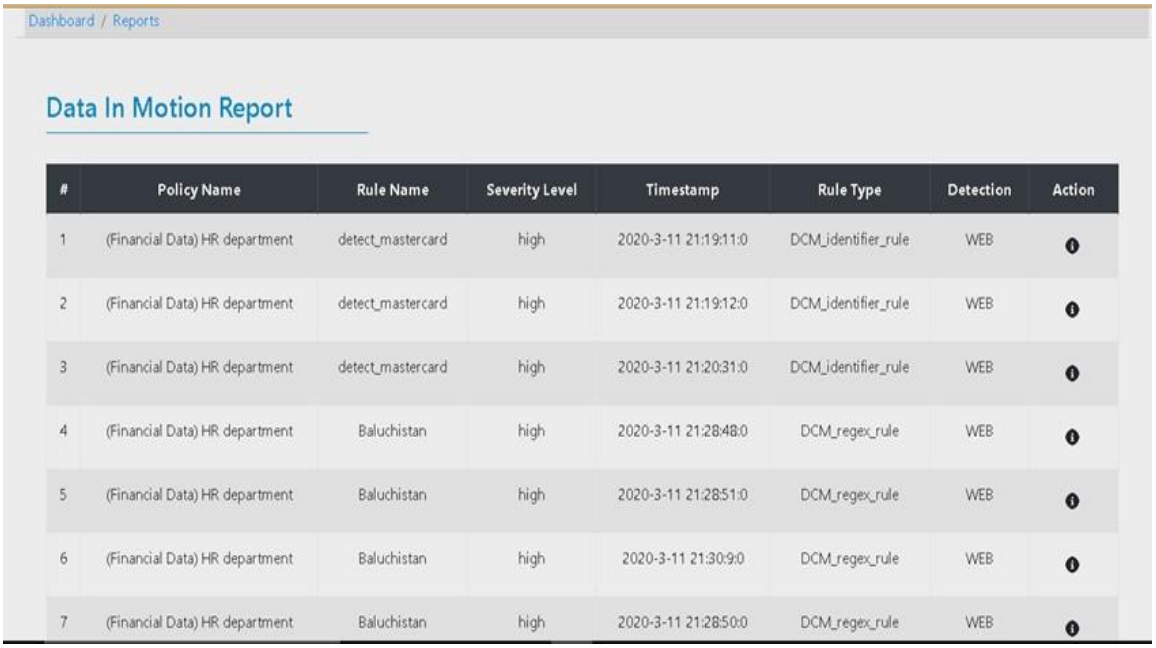
Waiting for kit-free.fontawesome.com...

Figure 6.9. Implemented Policies List Screen

6.2.7. Reports

An active administrator can see and take action on the sensitive data leakage incidents. It gives an administrator a sound understanding of what is going on in the organization. Super administrator usually gets regular reports like weekly or monthly so that one can look for events which are potentially more dangerous to organization.

An administrator can view all the reports which include both new and old ones in his account by using report screen. Cyber Guard local administrator schedule or one-time report. The process of generating reports is from logs which are collected through the daily activities of agents and then the violations done by them are uploaded on management server in form of logs. So, it gives all the information about the information about events of data loss.



Data In Motion Report

#	Policy Name	Rule Name	Severity Level	Timestamp	Rule Type	Detection	Action
1	(Financial Data) HR department	detect_mastercard	high	2020-3-11 21:19:11:0	DCM_identifier_rule	WEB	ⓘ
2	(Financial Data) HR department	detect_mastercard	high	2020-3-11 21:19:12:0	DCM_identifier_rule	WEB	ⓘ
3	(Financial Data) HR department	detect_mastercard	high	2020-3-11 21:20:31:0	DCM_identifier_rule	WEB	ⓘ
4	(Financial Data) HR department	Baluchistan	high	2020-3-11 21:28:48:0	DCM_regex_rule	WEB	ⓘ
5	(Financial Data) HR department	Baluchistan	high	2020-3-11 21:28:51:0	DCM_regex_rule	WEB	ⓘ
6	(Financial Data) HR department	Baluchistan	high	2020-3-11 21:30:9:0	DCM_regex_rule	WEB	ⓘ
7	(Financial Data) HR department	Baluchistan	high	2020-3-11 21:28:50:0	DCM_regex_rule	WEB	ⓘ

Figure 6.10. Reports screen of Cyber Guard

Using this screen, one can view that which policy was violated and which specific rule was related to it. What was its severity level? The information about timestamp and what action was taken all can be seen easily and if you click some event it will give more details like about hostname and machine. To view all this, navigate to reports by clicking on reports on dashboard.

6.2.8. Logs

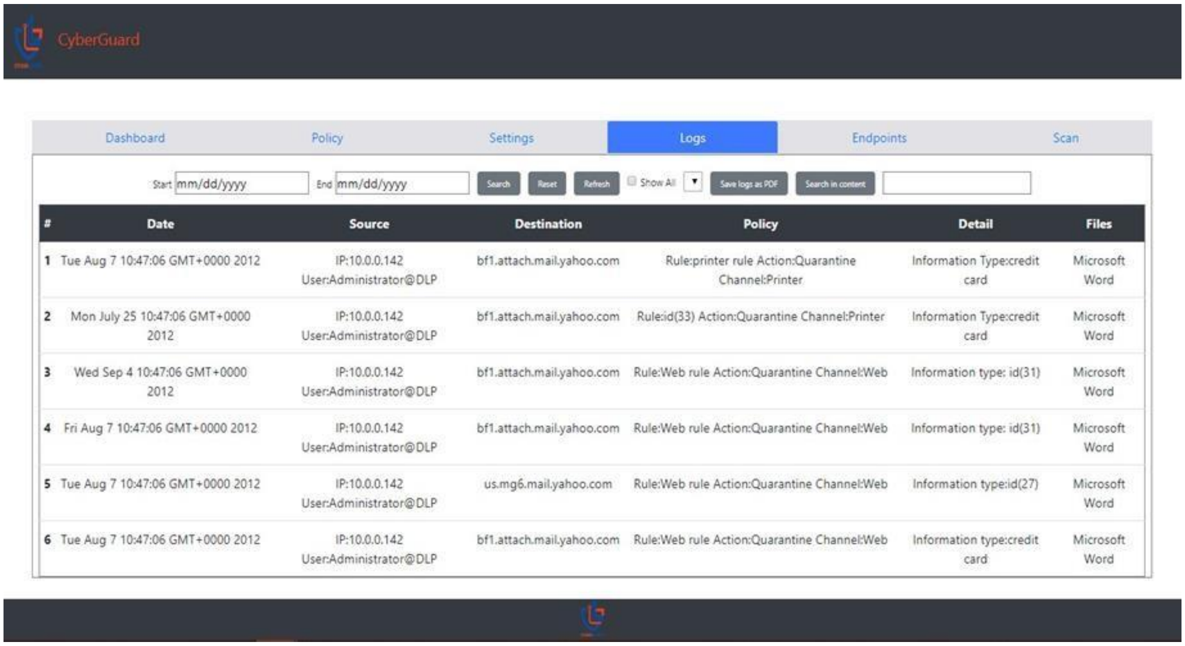
Cyber Guard DLP monitors different types of traffic and the different incidents are being reported to it and at the same it is scanning and storing its results. All information is stored in form of logs in Cyber Guard DLP. These logs help in generating reports later.

Different types of logs generated by DLP over time on Cyber Guard dash board are:

- Incidents
    - Unread Incidents
    - Incident by Product
    - Incident by date
    - Incident by Policy
- Active Agents
- Recent Alerts
- Servers status
    - Analytical Server
    - Database Server
- Policy Violations during entire period
- Proxies

- ○ HTTPS proxy

There is an also a separate tab by navigating to it we can view detailed logs.



#	Date	Source	Destination	Policy	Detail	Files
1	Tue Aug 7 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	bf1.attach.mail.yahoo.com	Rule:printer rule Action:Quarantine Channel:Printer	Information Type:credit card	Microsoft Word
2	Mon July 25 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	bf1.attach.mail.yahoo.com	Rule:id(33) Action:Quarantine Channel:Printer	Information Type:credit card	Microsoft Word
3	Wed Sep 4 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	bf1.attach.mail.yahoo.com	Rule:Web rule Action:Quarantine Channel:Web	Information type: id(31)	Microsoft Word
4	Fri Aug 7 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	bf1.attach.mail.yahoo.com	Rule:Web rule Action:Quarantine Channel:Web	Information type: id(31)	Microsoft Word
5	Tue Aug 7 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	us.mg6.mail.yahoo.com	Rule:Web rule Action:Quarantine Channel:Web	Information type:id(27)	Microsoft Word
6	Tue Aug 7 10:47:06 GMT+0000 2012	IP:10.0.0.142 User:Administrator@DLP	bf1.attach.mail.yahoo.com	Rule:Web rule Action:Quarantine Channel:Web	Information type:credit card	Microsoft Word

Figure 6.10. Logs screen

CHAPTER 7: DEPLOYMENT

7.1. Overview

Cyber Guard DLP is a solution for enterprises to prevent data loss. It is not an application that any consumer with digital device can use to prevent its data theft. It is a complex solution for organizations and should be deployed with extreme care. It keeps data safe if it is deployed properly in organization and if not then can also have negative impact on business daily routines. As the solution designed above is only prototype which contains only basic features of a DLP so its deployment is easy and small in comparison with complex and large solutions in market. Their deployment is complex and should be done in different steps. First small and easy parts should be deployed in it and then when they start working properly then should proceed. So Cyber Guard can be deployed in organizations by dividing the organization into different groups of security. The deployment is started from lower risk levels and cannot not think of deploying in highest security area first. This is recommended because if unfortunately, some breach occurs during the time of deployment the damage should be minimum.

After it is deployed on organization network then start the system with one feature only most recommended is monitoring. The blocking of data is not enabled at once until certain that system is working properly and there are no risks. At this time a support team and help desk is prepared to act quickly in case anything goes wrong.

Deployment of DLP is documented which helps organizations to deploy in case there is no support team from group of developers of DLP or re-deployment of system is needed due to some faut. If there is no documentation of architecture and deployment of DLP then it will be difficult for the organization even the developers of this DLP to deploy again.

The description of how to install and configure cyber guard DLP is below:

7.2. Installation and Configuration

7.2.1. Installation and Configuration of DLP Network Monitor and Management Console

<u>Step1:</u> Install Nodejs

- sudo apt-get install curl
- curl -sL https://deb.nodesource.com/setup_13.x | sudo -E bash -
- sudo apt-get install nodejs
- node -v
- npm -v

<u>Step2:</u> Install Anyproxy:

- npm instsall anyproxy -g (install anyproxy library from npmpackages)

<u>Step3:</u> Installing Mongodb

- sudo apt install -y mongodb (install mongodb database server fromapt package manager)
- sudo systemctl status mongodb (checking service and database)
- sudo systemctl stop mongodb (stop server)
- sudo systemctl start mongodb (startserver)

Step4: install GIT

- sudo apt install g
- git --version

Step5: Download Code for Cyberguard console (Github)

- git clone https://github.com/dawoodaijaz97/Cyberguards-Consoles.git

Step6: Download Code for Web monitoring (Github)

- https://github.com/dawoodaijaz97/web-monitoring-app.git

Step7: Install NPM packages for cyberguard console

- cd Cyberguards-consoles
- npm install

Step8: Install NPM packages for Web monitoring

- cd Web-minotring-app
- npm install

Step9: Run Cyberguard console

- cd Cyberguards-Consoles
- cd bin
- node www

Step10: Run Web monitoring server

- cd Web-monitrin-app
- cd analysis-server
- node server.js

7.2.2. Installation and Configuration of DLP Endpoint

Step1: Download the Endpoint (.exe file) from the Cyber Guard's main server.

Step2: Once downloaded, Double-click on the .exe file. All the required packages are installed and the Endpoint Agent will start running on your system.

Step3: Agent will first download the policies from the main server and store them in the json file.

Step4: When these policies are fetched, Agent will start monitoring activities according to the downloaded policies.

CHAPTER 8: ANALYSIS AND EVALUATION

8.1. Overview

In this chapter, analysis and evaluation performed on Cyber Guard Data Loss Prevention has been described. In analysis part detailed investigation of elements/components of the DLP is done and how they perform their operations. After this analysis, these components are evaluated by defining test cases and checking if those test cases are valid or not. In short, the components are analyzed and checked if they are working according to our requirements. i.e. evaluation.

8.2. Analysis

DLP is comprised of three main features that are:

- **Data in use**
- **Data in motion**
- **Data at Rest**

The workings of each of these modules is checked and analyzed separately.

8.2.1. Data in Use

DLP Agent is installed on every client computer and is running continuously, monitoring the sensitive files and operations on them.

- **Copy/Paste Operation:**

  When a USB is plugged in and someone tries to copy sensitive text from the sensitive files then what the DLP agent does is that it fetches the text from the clipboard, checks whether it contains a sensitive keyword against the pattern defined in the policies and blocks the operation. If any sensitive information is found, the copy/paste operation is blocked and an alert is sent to the DLP main server.

- **Screenshot Operation:**

  When you press “print screen” from the keyboard, the screenshot operation occurs. So that screenshot is held temporary in the clipboard of the computer. DLP agent fetches that screenshot and blocks the screenshot from sending to any removable media. It then sends the alert to the DLP server about the violation of Screenshot policy.

8.2.2. Data in Motion

DLP helps in protecting sensitive data of organization which is going out on internet by deploying HTTPS proxy between the organization’s network and the internet. The network traffic is logged into a database and scanned when required

- **Monitoring of HTTPS Traffic**

  Https proxy monitors all activities on egress traffic. If someone tries to send sensitive data on network, it goes through HTTPS proxy and is stored for later analysis. It will check whether it contains sensitive information or not. The text is scanned for patterns that match with the sensitive keywords. If match found, an alert is generated and sent to the DLP main server by the HTTPS Proxy containing the details of Policy Violation.

- **Analyzing Web Attachments**

  HTTPS Proxy also helps in protecting the sensitive files that are sent on web as attachments e.g. an email attachment. When a file is attached as an attachment in the email, HTTPS proxy scans that file for sensitive keywords and also performs a content matching on that file. If Sensitive keywords are found then an Alert is generated and sent to the DLP main server. On the other hand, Content matching is done on that file against the policies and the content of the original file. A threshold is set on the HTTPS proxy that if the percentage of matching exceeds from 50% then mark that file as sensitive and generate an alert to the DLP core server.

8.2.3. Data at Rest

Monitoring Data at rest means that managing and discovering sensitive data stored across the organization storages. In organization data stored in many digital formats such as Databases, File servers.

- **Scanning Endpoint File System:**

  The endpoint agent is used to scan the file system of the endpoints and discover sensitive data stored on users' systems. Using the CyberGuard console the administrator can select the endpoints to run scans on. The selected endpoints agents start scanning file systems. If any sensitive content is found on any endpoint an alert is generated notifying the administrator.

- **Scanning Organizations Databases:**

  The second feature of "Data at Rest" module is to scan the organizations databases to detect sensitive data so that the sensitive data can be efficiently managed, and the administrator has a better overview of distribution of data across the organization. Unlike data discovery at endpoint which uses an agent to detect sensitive data. The scanning of databases is much simpler. The management consoles connect to the respective database using IP/Port and then fetches all the records from the database and scans it against to rules and policies. If any violations are found alerts are generated and the admin is notified

The administrator can select the database to scans. After the scans are completed the detailed report is generated

8.3. Evaluation

In this part, three main modules of DLP are evaluated and is done by developing test cases and this is done separately for each component.

8.3.1. Evaluation of Data in Use

- **Test Case 1**: Verify if copy/paste of sensitive data is stopped by the agent when sensitive data is copied
- **Test Case 2:** Verify if agent start running on system startup
- **Test Case 2**: Verify if alert in generated when a screen shot of sensitive data is taken

8.3.2. Evaluation of Data in Motion

- **Test Case 1**: Verify how many concurrent connection proxies can handle
- **Test Case 2**: Verify if the traffic is properly saved by the proxy in database

- **Test Case 3**: Verify if scans generate alerts when sensitive data in found in stored traffic
- **Test Case 4:** Verify if sensitive email attachments are detected and alert is generated.

8.3.3. Evaluation of Data at Rest:

- **Test case 1**: Verify if agent runs local scans when scan command is issued by the administrator
- **Test case 2**: Verify if scans take a limited amount of time
- **Test case 3**: Verify if scans do not take too much computational resources
- **Test Case 4:** Verify if sensitive data is found on users' storage respective alert is send to the management console.

8.4. Limitations

- **Data in motion:**
  1. The detection is not in real time the traffic is saved and later scanned for violations.
  2. The scanning process is slow because of single threaded nature of node.js.
  3. SMTP and FTP protocols are currently incomplete.
  4. Saving files in database with size greater than 15mb causes error.
  5. Limited number of files extension are covered (.txt, .pdf, .docx, .ppt, .csv).
  6. The use of https proxy increases the latency in the networks.

7. The https proxy can handle only limited number of concurrent connections with the clients.
8. To intercept and decrypt the https traffic from the client the proxy server certificates need to be installed on the client machine.

- **Data at rest:**

1. Only handles relational database (MySQL) non-relational databases are not supported
2. Scans at endpoints can slow down the endpoints and scan are not paralyzed for optimization
3. only a limited number of file types are scanned (.txt, .pdf, .docx, .ppt, .csv) other file types are not covered yet
4. Limited number of concurrent rest scans (10 sockets connected at a time)
5. The endpoint agent only works for windows systems the Linux systems are not supported

- **Data in Use:**

    1. Agent is only compatible with Windows operating system. (DLP agent does not operate on other Operating systems).
    2. Scanning files with agent is slow and take a lot of time in case of large user storage
    3. Due to the lack of use of multi-threaded programing the scanning process becomes slow
    4. Only Sensitive keywords that are in English are protected by the DLP agent.

CHAPTER 9: CONCLUSION AND FUTURE WORK

9.1. Conclusion

It is apparent that a data loss event can have a huge effect in an organization's normal activities and finance. Surely, clients can lose trust in the company, eventually failure in business. There are various channels through data can be leaked from an organization. This risk of data theft can be reduced, by using a security solution that monitors and blocks the sensitive content from leakage without disrupting normal routine activities. The most effective solution in this scenario is data loss prevention (DLP).

Defining the sensitive data of organization and then finding that sensitive content are the basic requirements to design an effective data loss prevention solution. Recognizing all the data channels through which data flows in and out of organization helps in protecting data through DLP. Defining the sensitive data and then classifying it into different groups of security also enables DLP to perform its functionalities.

So, in short DLP provides organizations a smart, working solution which:

- Identifies sensitive content.
- Protects this data.
- Improves and evaluates DLP rules and policies, after getting more information on that identified sensitive data and communication channels and also classifies this data with help of rules and policies

After designing and implementation, when DLP is deployed, the three basic modules (described below) of data loss prevention start working in the areas of their functionality and cover the security gaps and the output is visible.

Data in motion or data on the network basically checks the traffic going out of organization network against the pre-defined policies. This module ability to check data on HTTPS protocol. In addition, it also checks the data which is uploaded on various web sites over internet and completely analyzes them for sensitive data.

Data at Rest module consists of all the sensitive data which is spread across entire organization network present on different storage areas like file servers. DLP scans these storage areas to locate this sensitive data and take actions according to policies defined.

Data in use module of DLP work on endpoints and monitor and control user actions. If user tries to copy the sensitive content or use snipping tool on it, it takes actions according to pre-defined policies and thus reducing data loss.

The DLP system proposed in this thesis is working prototype consists of all the basic functionalities of DLP. Furthermore, the solution proposed, designed and explained has almost automation that make the system more efficient and also reduces the human error which causes failure in these systems at many points.

Web-based management console is one of the prominent features of this system. It has interactive graphical user interface so that even a lay man can use it. The dashboard is very descriptive and provide real time analytics of the events occurring within organization network. Only administrators have access to this management console and they can define policies and rules for protecting data and they are very easy to create on this console.

Logging of incidents and generating reports is another feature of this DLP which is associated with this console. Real time scanning and monitoring, and in case of any discrepancy alerting all these major features are available in this DLP system.

However, there are still limitations and the proposed DLP system has still room of improvement.

9.2. Future Research on DLP

The DLP solution built is a working protype which consists of three basic components and they are present in most DLPs in market. Th implemented components have some basic functionalities for proof of concept. There is also another part of DLP which is getting popularity now a days known as Cloud DLP. So, in order to expand our knowledge, the next area of research already developed working protype is working of DLP on cloud. In this part, some papers have also been presented to support future research on how cloud computing take part in DLP to prevent data loss.

9.2.1. DLP in Cloud Computing

In this modern world almost, everyone is storing their data on clouds especially the big organizations, but this is also arising some security concerns related to data. Although storing data to a cloud is easy and cost effective. So, if there is a DLP on cloud environment it will give some surety to organizations that their data is safe. But still the main problem is how will DLP work on cloud and how it will be fulfilling cloud requirements?

Distributed storages are used to store the data instead of central model in case of cloud environments so to secure this data methods are needed instead of traditional ones.

As the data of organizations is on various personal digital devices like laptops so the chance of its theft also increases A solution is needed which can monitor data across different distributed storages. So, we have to find how data is stored in such environments. Another important change in this system is that virtual appliances are also involved which can cause many problems.

An important point to be considered is that can cloud service providers identify sensitive content. Does they follow standards or rules and techniques to keep the privacy of data intact. Will the performance of the DLP system on cloud will be efficient?

The main objective of Data loss prevention (DLP) is to locate and block the sensitive data so it cannot be lost. After researching some papers, one effective method to use DLP on cloud for monitoring and blocking identified. Most of the services on cloud use http/https protocol. Thus, if this protocol is checked properly many data loss acts can be detected [13].

SMTP traffic can also be analyzed by using some different techniques like:

- Using an endpoint service on particular instance of cloud.
- Re-routing the traffic to DLP server.

References

[1] E. S. a. V. M. Sultan Alneyadi, "A survey on data leakage prevention systems," *Journal of Network and Computer Applications,* 2016.

[2] C. M. P. a. D. Satyavathy, "A technical review on data leakage detection and prevention approaches," *Journal of Network Communications and Emerging Technologies (JNCET),* 2017.

[3] L. S. L. a. R. Kuhn, "Data loss prevention," *IT Professional,* 2010.

[4] ISACA, "Data Leak Prevention," 2010.

[5] T. Terms, Data.

[6] B. HAUER, "Data and Information Leakage Prevention Within the Scope of Information Security," *IEEE Acsess,* 2015.

[7] L. Securosis, "Understanding and Selecting a Data Loss Prevention Solution," *Securosis,* 2010.

[8] "DATA LOSS PREVENTION GUIDE: LEARN DATA LOSS TIPS".*VERACODE.*

[9] Wikipedia, Data loss prevention software.

[10] C. H. A. G. Jose Ferreira, "10 Reasons Why Your Organization Needs Data Loss Prevention," *SIRUS Edge,* 2018.

[11] E. S. V. M. Sultan Alneyadi, "A survey on data leakage prevention systems," *Journal of Network and Computer Applications,* 2014.

[12] I. I. J. K. A. a. B. O. A. Victor O. Waziri, "Data Loss Prevention and Challenges Faced in their Deployments," in *International Conference on Information and Communication Technology and Its Applications*, 2016.

[13] H. Sethuraman, "Data Loss/Leakage Prevention (DLP)," Luea Unuiversity of Technology, 2012.

Cyber Guard (Data Loss Prevention)

ORIGINALITY REPORT

10%	4%	2%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Higher Education Commission Pakistan Student Paper	2%
2	Barbara Hauer. "Data and Information Leakage Prevention Within the Scope of Information Security", IEEE Access, 2015 Publication	1%
3	www.uop.edu.jo Internet Source	<1%
4	Submitted to Universiti Teknologi Malaysia Student Paper	<1%
5	www.veracode.com Internet Source	<1%
6	Submitted to Federal University of Technology Student Paper	<1%
7	etheses.whiterose.ac.uk Internet Source	<1%
8	medium.com Internet Source	<1%

9	Submitted to Vaasan yliopisto Student Paper	<1%
10	wwm.kualitatem.com Internet Source	<1%
11	www.uday.net Internet Source	<1%
12	Submitted to University of London External System Student Paper	<1%
13	Submitted to Royal Holloway and Bedford New College Student Paper	<1%
14	Submitted to INTI University College Student Paper	<1%
15	Submitted to University of Wales central institutions Student Paper	<1%
16	Submitted to Southern New Hampshire University - Distance Education Student Paper	<1%
17	jssaten.ac.in Internet Source	<1%
18	Submitted to Cankaya University Student Paper	<1%

19	Submitted to University System of Georgia Student Paper	<1%
20	Submitted to University of Ulster Student Paper	<1%
21	Submitted to University of South Australia Student Paper	<1%
22	docplayer.net Internet Source	<1%
23	Submitted to University of Limerick Student Paper	<1%
24	manualzz.com Internet Source	<1%
25	Submitted to University of Leeds Student Paper	<1%
26	Submitted to Eiffel Corporation Student Paper	<1%
27	nextinnovator.com Internet Source	<1%
28	Submitted to National University of Ireland, Galway Student Paper	<1%
29	Submitted to University of Wales Institute, Cardiff Student Paper	<1%

30	Submitted to University of Westminster Student Paper	<1%
31	Submitted to University of Mauritius Student Paper	<1%
32	Submitted to CTI Education Group Student Paper	<1%
33	opus.bath.ac.uk Internet Source	<1%
34	Submitted to Turun yliopisto Student Paper	<1%
35	Submitted to Laureate Higher Education Group Student Paper	<1%
36	eprints.lancs.ac.uk Internet Source	<1%
37	daneshyari.com Internet Source	<1%
38	spectrum.library.concordia.ca Internet Source	<1%
39	www.datis-arad.com Internet Source	<1%
40	www.diva-portal.org Internet Source	<1%
41	senior.ceng.metu.edu.tr Internet Source	<1%

42	hdl.handle.net Internet Source	<1%
43	Submitted to Walden University Student Paper	<1%
44	Submitted to Wright State University Student Paper	<1%
45	alum.cs.sunysb.edu Internet Source	<1%
46	Submitted to Sri Lanka Institute of Information Technology Student Paper	<1%
47	Submitted to Pathfinder Enterprises Student Paper	<1%
48	Submitted to University of Greenwich Student Paper	<1%
49	Submitted to Limerick Institute of Technology Student Paper	<1%
50	umpir.ump.edu.my Internet Source	<1%
51	research.library.mun.ca Internet Source	<1%
52	www.digibib.tu-bs.de Internet Source	<1%

53	"Proceedings of the Third International Scientific Conference “Intelligent Information Technologies for Industry” (IITI’18)", Springer Science and Business Media LLC, 2019 Publication	<1%
54	Submitted to TechKnowledge Student Paper	<1%
55	Submitted to Flinders University Student Paper	<1%
56	Submitted to Middlesex University Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		