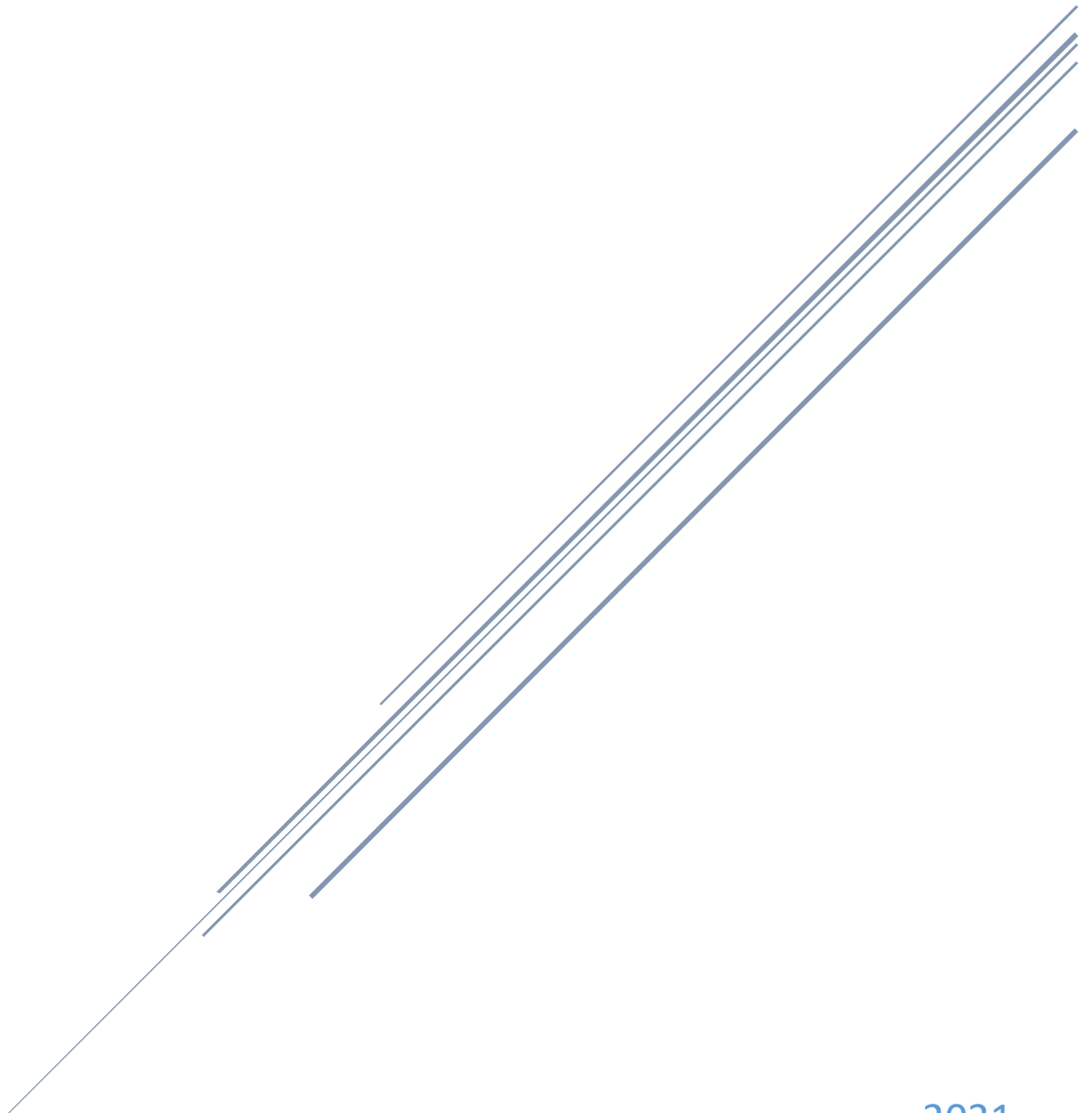


# DATA LEAK PREVENTION (DLP)

## Report



2021

National Defense Complex – Cybersecurity Internship

## Table of Contents

1	Introduction .....	1
1.1	Background .....	1
1.2	Definition .....	2
1.3	DLP Solutions .....	3
1.4	DLP Features vs. DLP Solutions: .....	4
2	Objectives .....	4
3	Data Leak Prevention Problem .....	5
3.1	Data Loss/Breaches:.....	6
3.1.1	Insider Threat: .....	6
3.1.2	UBER Data Breach: .....	7
3.1.3	SONY Data Breach:.....	8
3.1.4	International oil and gas company, domestic financial institute, international software company – By ERNST & YOUNG .....	8
3.1.5	Case Study - Dash Board View .....	9
3.1.6	Case study - HR sending sensitive info to webmail: .....	9
3.1.7	Case study - Employee Salary Info Sent to Webmail:.....	10
3.1.8	Case study - Credit Cards Numbers Sent Unencrypted: .....	10
4	DLP Core activities .....	11
4.1	Policies .....	12
4.1.1	Policy Violations:.....	12
4.2	Discover / Identification of Data.....	13
	Data Scanning methods: .....	13
4.2.1	Pre-study of sensitive data of the organization: .....	13
4.2.2	Content Discovery Techniques .....	14
4.3	Monitoring of Data .....	16

4.3.1	Regular expressions .....	16
4.3.2	Fingerprint.....	17
4.3.3	Dictionaries: .....	17
4.3.4	File system protection: .....	17
4.3.5	Network protection: .....	17
4.3.6	Kernel protection: .....	18
4.3.7	Content Analysis Techniques .....	18
4.4	Protect/Prevention/Management of Data Leakage .....	21
4.4.1	Blocking Copy Functionality .....	21
4.4.2	Blocking Inputs and Devices .....	22
4.4.3	Blocking Application Input.....	22
4.4.4	Protecting E-Mail.....	22
4.4.5	Automatic Data Removal.....	22
4.4.6	Dealing with Encryption .....	23
5	DLP Implementation/Deployment Methods .....	24
5.1	Protecting Data in Motion, at Rest, and in Use.....	25
5.1.1	Keep data safe while in use on endpoints .....	26
5.1.2	Protect data in motion over the network .....	27
5.1.3	Protect data at rest across storage repositories.....	28
5.1.4	Protect data in the cloud.....	29
5.2	DLP depending on the state of data: .....	30
5.3	Implementation of DLP on Endpoint and Network .....	30
5.4	Endpoint Features and Integration .....	31
5.4.1	USB/Portable Device Control:.....	31
5.4.2	Endpoint Protection Platforms: .....	31

5.4.3	Non-Antivirus EPP: .....	32
5.5	Networks .....	32
5.5.1	Network Monitor .....	32
5.6	Storage Features and Integration .....	33
5.6.1	Database Activity Monitoring and Vulnerability Assessment: .....	34
5.6.2	Vulnerability Assessment: .....	34
5.6.3	Document Management Systems:.....	34
5.6.4	Content Classification, Forensics, and Electronic Discovery: .....	34
5.7	Other Features and Integrations .....	34
5.7.1	SIEM and Log Management: .....	34
5.7.2	Enterprise Digital Rights Management:.....	35
5.7.3	Email Encryption: .....	35
5.8	Distributed and Hierarchical Deployments:.....	35
6	DLP Products/Solutions available in Market.....	36
6.1	Some Views of the DLP Products: .....	37
6.1.1	SYMANTEC.....	37
6.1.2	McAfee .....	38
6.1.3	Check Point.....	38
6.1.4	EndPoint DLP.....	38
7	DLP Benefits.....	38
8	DLP Limitations.....	40
9	DLP Challenges .....	40
9.1	Encryption.....	40
9.2	Collaboration.....	40
9.2.1	Access Control .....	40

9.2.2	Semantic Gap in DLP .....	41
10-	References .....	41

# 1 Introduction

## 1.1 Background

The last decade of the 20th century offered warnings of what was to come in the next century. The 1990s brought the first bombing of the World Trade Center, the bombing of the Murrah Federal Building in Oklahoma City, the first war with Iraq, crimes resulting from the Internet, the increased value of proprietary information, and attention to violence in the workplace. As we know, not long into the 21st century, on September 11, 2001, terrorists attacked the World Trade Center and the Pentagon. Following the attacks, a crisis in confidence in government occurred. Citizens asked: How could the most powerful nation on earth be subject to such a devastating attack? What went wrong? Who is to blame? In response to the crisis, President George Bush declared war on terrorism. He appointed a new Cabinet position, the Office of Homeland Defense, to coordinate counterterrorism. The attacks also led to greater police powers for search and seizure and electronic surveillance, and the age-old question of how to balance police powers and constitutional rights. These bold, surprise attacks, subsequent bioterrorism (i.e., anthrax attacks through the U.S. Postal System), the war in Afghanistan, and the second war in Iraq show the difficult challenges facing our world in this new century. The United States and its allies are not only faced with conflict in Iraq, Afghanistan, and other regions, but also old and emerging state competitors and the proliferation of weapons of mass destruction.[2]

Although existing prior to 2006, DLP started blooming around the fall of this year. At this time larger vendors started to acquire smaller companies specializing in data security, a trend that continued far into 2007. During 2008 and the beginning of 2009 DLP was on everyone's lips, as illustrated by Gartner's 2008 hype cycle seen in Figure 3. After this, the technology slowly faded into the background, as is common for many security trends. [6]

The largest vendors today are Websense, Symantec, RSA, Palisade Systems, NextLabs, McAfee, Fidelis Security Systems, Code Green Networks and CA Technologies. All these vendors offer products with the core features associated with a DLP. This includes network monitoring, e-mail monitoring, file system monitoring, and endpoint protection. To discover content the product use

wordlists, regular expressions, and partial file hashing. As for machine learning algorithms, only Symantec offers this feature in their DLP product [6]

## 1.2 Definition

A few years ago, the DLP market was dominated by startups with only a couple major acquisitions by established security companies. The entire market was probably smaller than the leading one or two providers today. Even the term ‘DLP’ was still under debate, with a menagerie of **terms** like

- ⇒ Extrusion Prevention,
- ⇒ Anti-Data Leakage,
- ⇒ and Information Loss Protection still in use

While we have experienced maturation of the products, significant acquisitions by established security firms, and standardization on the term DLP, in many ways today’s market is even more confusing than a few years ago.

DLP definition provided by **Securosis** – A leading research and advisory firm that has been following the development of DLP closely.

**Definition:** “Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use through deep content analysis” [8]

Thus, the defining characteristics are:

- Deep content analysis
- Central policy management
- Broad content coverage across multiple platforms and locations

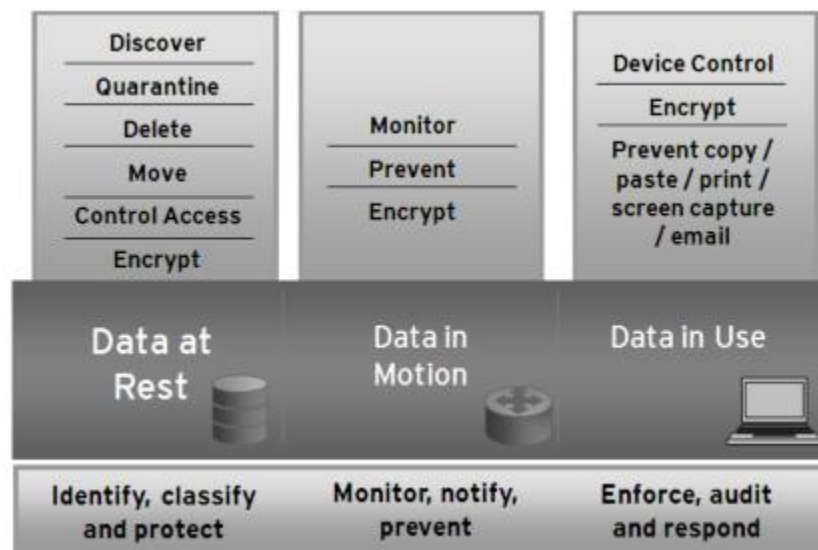
### **Another definition of DLP states that:**

“DLP [Data Loss Prevention] is a system that performs real-time scanning of data at rest and in motion, evaluates that data against existing policy definitions, identifies policy violations and automatically enforces some type of pre-defined remediation actions such as alerting users and administrators, quarantining suspicious files, encrypting data or blocking traffic outright.” -

451 Research, “The Data Loss Prevention Market by the Numbers,” July 2015

According to **MOGULL Rich**, “DLP is the current technological solution to protect an organization from data leaks. DLP technology enforces the criteria of how the information will flow in and out of the company's electronic network including audit trails, notifications, and response actions [4]”.

DLP which was first **presented to the market in 2006**, is a solution that can inspect the content of the electronic data of the organization specializing in looking for sensitive or valuable information, which is moving without permission through the company.



## 1.3 DLP Solutions

DLP solutions both protect sensitive data and provide insight into the use of content within the enterprise. Few enterprises classify data beyond public vs. everything else. DLP helps organizations better understand their data and improves their ability to classify and manage content. Full-suite solutions provide complete coverage across your network, storage repositories, and endpoints, even if you aren't using the full capabilities. There are three other possible approaches:

Partial-suite DLP solutions are dedicated DLP tools that cover two potential channels (e.g., network and storage) and contain full workflow (such as incident management) and content analysis capabilities. There are very few partial suites available these days.



Single-channel DLP solutions cover only one channel, but still include full DLP workflow and content analysis capabilities. While we tend to see more single channel offerings than partial suites, there are still only a few products on the market — almost all either network or endpoint.

DLP features are now included in a variety of products, offer a subset of coverage and content analysis capabilities, and typically lack dedicated DLP workflow. For example, we have seen network firewalls with basic pattern-matching capabilities, vulnerability assessment scanners that look for data types (such as credit card numbers), and limited content analysis in an email security gateway.

## 1.4 DLP Features vs. DLP Solutions:

When evaluating options, it's sometimes difficult to characterize the real differences between DLP features and dedicated DLP solutions, and the value of each. The key differences are:

- A DLP product or solution includes centralized management, policy creation, and enforcement workflow, dedicated to the monitoring and protection of content and data. The user interface and functionality are dedicated to solving the business and technical problems of protecting content through content awareness.
- DLP features include some of the detection and enforcement capabilities of DLP products but are not dedicated to protecting content and data. This latter approach is sometimes called “DLP Light” to reflect its less-robust nature, and it's becoming extremely common in a variety of other security tools. (Gartner calls this “channel DLP”).

## 2 Objectives

**Objective** of Data Leakage/Loss Prevention is to minimize the data loss and business impact at all times due to a data breach event that could potentially become an incident.

**In this report, we** have reviewed the DLP field. Specifically, data leak prevention problem, describe core activities, Implementation approaches, and outline some potential benefits and challenges while using DLP.

### 3 Data Leak Prevention Problem

There are numerous ways through which sensitive data can be revealed to untrusted third parties, as depicted in Figure 1. [3]

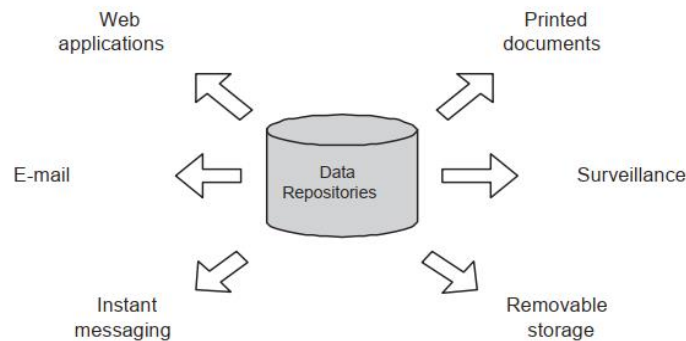


Fig. 1. Data leak channels.

As shown in Figure 1, data leaks can occur in many ways. Hardware theft, social engineering, surveillance, the mismanagement of printed documents is a few of the more traditional data leak channels. Additionally, electronic communications such as instant messaging, web applications and email provide additional challenges. These electronic channels are highly utilized in organizations and provide means to send data quickly and easily to a third party. While traditional data leaks can be more suitably defended with traditional approaches [10] lightweight and context aware techniques, which can infer who is communicating and what is being communicated, are needed to prevent data leaks in electronic communications [3]

***Data breach incidents cost US companies \$204 per compromised record, with an average total per-incident cost of \$6.75m – Ponemon Institute***

According to the latest Cost of Data Breach Study<sup>1</sup>, human errors and system glitches caused nearly two-thirds of data breaches. a recent Symantec study found that more than half of employees admit to emailing business information from the office to their personal accounts. Employees are the leading cause of information leaving your organization. Some are looking for a head start when moving to a new company, while others simply feel a sense of ownership toward something they create. In fact, half of departing employees keep confidential information when they leave. They

believe it won't harm the company they are stealing from, and the majority does not believe it is a crime to use confidential data taken from a previous employer.

Despite the number of technological steps that have been taken to reduce data losses, cases of data breaches are not decreasing. A recent study conducted by the Open Security Foundation shows that over 502 million records, including credit card numbers, access credentials and other personal information, were leaked in the first half of 2014[20]

Data loss prevention (DLP) is critical to stop accidental and malicious data leaks—whether it's customer information, financial data, intellectual property or trade secrets. Today's enterprise must be able to identify, track, and secure all confidential data at rest, in use, and in motion. This is increasingly difficult due to growing risk factors, including mobile workers and the widespread use of USB drives, webmail, IM, and CDs/DVDs. According to some stats.

- Insider & Partners Cause Most Breaches – 880% of breaches.
  - Insiders make mistakes handling data.
  - Broken business processes increase risk.
- Compliance mandates data protection – 81% of companies breached were not PCI compliant.
  - Increased focus on data privacy
  - Need to demonstrate data controls.
- More Complex Threats to Your Data - \$6.7. million average cost of breach.
  - External threats target high value data.
  - Limited visibility of where data is.

## **3.1 Data Loss/Breaches:**

### **3.1.1 Insider Threat:**

Employees are the backbone of your organization, but they are also one of the most pressing dangers to your information – more so than hackers – due to their legitimate access to internal systems and applications.

Some examples of data loss are **WikiLeaks** founded by Julian Assange which exposed top secret information from governments and military organizations causing a big impact for political

regimes and private companies. Also, the case of Edward Snowden, who has been responsible for the most important leak in the **NSA's history** [4].

Another Big Example of Data breach is **YAHOO**. In September 2016, the once dominant Internet giant, while in negotiations to sell itself to Verizon, announced it had been the victim of the biggest data breach in history, likely by “a state-sponsored actor,” in 2014. The attack compromised the real names, email addresses, dates of birth and telephone numbers of 500 million users. The company said the "vast majority" of the passwords involved had been hashed using the robust b-crypt algorithm.

A couple of months later, in December, it buried that earlier record with the disclosure that a breach in 2013, by a different group of hackers had compromised 1 billion accounts. Besides names, dates of birth, email addresses and passwords that were not as well protected as those involved in 2014, security questions and answers were also compromised. In October of 2017, Yahoo revised that estimate, saying that, in fact, all [3 billion user accounts](#) had been compromised. [21]

Citing the 2<sup>nd</sup> example is **E-BAY** The online auction giant reported a cyberattack in May 2014 that it said exposed names, addresses, dates of birth and encrypted passwords of all of its 145 million users. The company said hackers got into the company network using the credentials of three corporate employees, and had complete inside access for 229 days, during which time they were able to make their way to the user database. CEO John Donahue said the breach resulted in a decline in user activity but had little impact on the bottom line. [21]

### **3.1.2 UBER Data Breach:**

The scope of the **Uber breach** alone warrants its inclusion on this list, and it's not the worst part of the hack. The way Uber handled the breach once discovered is one big hot mess, and it's a lesson for other companies on what not to do. The company learned in late 2016 that two hackers were able to get names, email addresses, and mobile phone numbers of 57 users of the Uber app. They also got the driver license numbers of 600,000 Uber drivers. As far as we know, no other data such as credit card or Social Security numbers were stolen. The hackers were able to access Uber's GitHub account, where they found username and password credentials to Uber's AWS account. Those credentials should never have been on GitHub.

Here's the really bad part: It wasn't until about a year later that Uber made the breach public. What's worse, they paid the hackers \$100,000 to destroy the data with no way to verify that they did, claiming it was a "bug bounty" fee. Uber fired its CSO because of the breach, effectively placing the blame on him. [21]

### **3.1.3 SONY Data Breach:**

On April 20, 2011 77 million PlayStation Network- SONY accounts hacked; estimated losses of \$171 million while the site was down for a month.

This is viewed as the worst gaming community data breach of all-time. Of more than 77 million accounts affected, 12 million had unencrypted credit card numbers. Hackers gained access to full names, passwords, e-mails, home addresses, purchase history, credit card numbers and PSN/Qriocity logins and passwords. "It's enough to make every good security person wonder, 'If this is what it's like at Sony, what's it like at every other multi-national company that's sitting on millions of user data records?'" said eIQnetworks' John Linkous. He says it should remind those in IT security to identify and apply security controls consistently across their organizations. For customers, "Be careful whom you give your data to. It may not be worth the price to get access to online games or other virtual assets."

In 2014, Sony agreed to a preliminary \$15 million settlement in a class action lawsuit over the breach.[21]

### **3.1.4 International oil and gas company, domestic financial institute, international software company – By ERNST & YOUNG**

#### **Ernst & Young solutions and results**

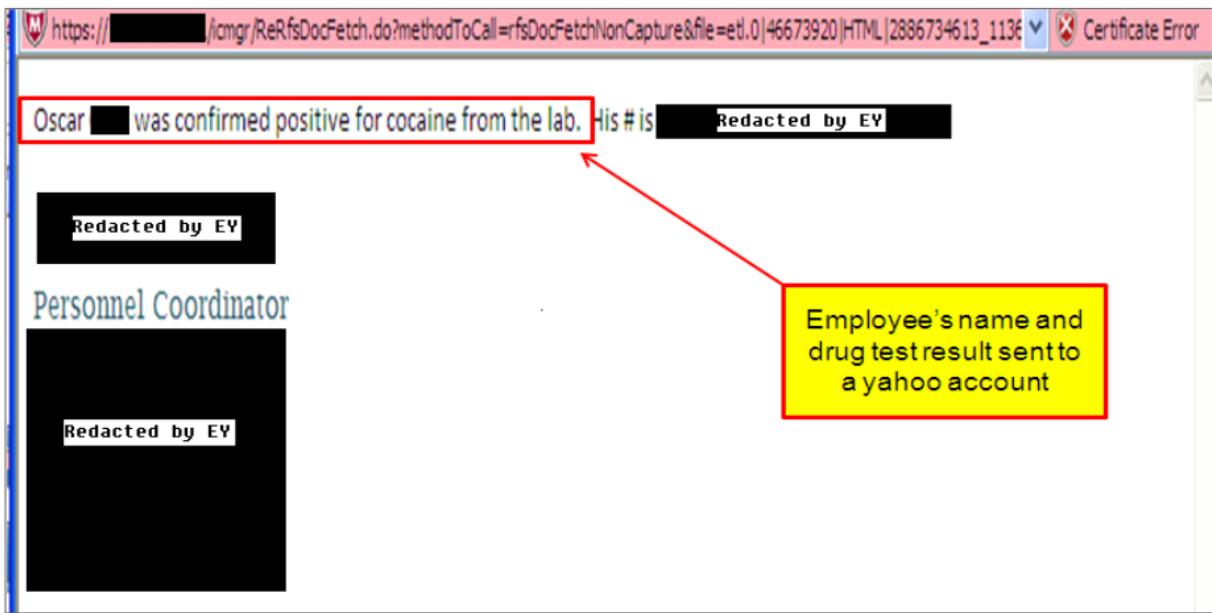
- ▶ Ernst & Young deployed the DLP appliances to gain a better understanding of the company's management of sensitive information as it is stored and transferred
- ▶ The assessments highlighted weaknesses in the overall DLP program and evidenced the loss of sensitive data such as IP, PII and PHI
- ▶ Technology was accompanied with business process assessment to provide process-level root cause for data loss
- ▶ Assessment provided CIO, IA, Legal, HR, CFO, Audit Committee, with insight to the company's data loss exposure and detailed recommendations on how to address DLP

### 3.1.5 Case Study - Dashboard View



### 3.1.6 Case study - HR sending sensitive info to webmail:

Provides evidence of data loss of sensitive employee information



3.1.7 Case study - Employee Salary Info Sent to Webmail:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
18	Base grable													
19														
20	JUAN GONZALEZ													
21														
22	JUNIO MAYO DE													
23	Con Benef													
24	Ingreso		6.402											
25	Fondo Solididad		6.692											
26	Aporte a Pension		5.779											
27	Voluntarios		0											
28	Aporte AFC		0											
29	Subtotal		3.931											
30														
31	Excento 25%		0.983											
32	Base Neta		2.948							9.667				
33														
34	Base grable		2.948											
35	SUMAS BASES		2.948											
36	BASE MES		8.250			191	6.367							
37	Alivo tributario		8.596											
38	ACUMULADO ANUAL		6.402											
39	BASE PENSION		2.471											
40	ACUMULADO AÑO		3.931											
41	BASE MES CON ALIVO		5.664											
42														
43														
44	I. Definición de datos Básicos													
45	a) Valor de la UVT durante el año 20													
46														
47	b) Valor del salario promedio gravable del trabajador													
48	durante los 12 meses anteriores													
49														
50														
51														
52														
53														
54														
55														
56														
57														
58														
59														
60														
61														
62														
63														
64														
65														
66														
67														
68														
69														
70														
71														
72														
73														
74														
75														
76														
77														
78														
79														
80														
81														
82														
83														
84														
85														
86														
87														
88														
89														
90														
91														
92														
93														
94														
95														
96														
97														
98														
99														
100														

3.1.8 Case study - Credit Cards Numbers Sent Unencrypted:

Credit card number used for payment sent in the body of an email

ID	36550
Protocol	SMTP_Request
Capture Device	
Time Sent	
User Id	unknown
Source IP/Port	
Source Location	United States
Destination IP/Port	
Destination Location	United States
Policy	PCI Compliance (admin)
Rule	Unencrypted Mastercard Numbers
Sender	
Recipients	
Subject	
Attachments	
Message	Show Message in Separate Window !

Good Morning

I would like to order the following:

- 10 sheets of Daily stamps in the total amount of \$600.00

Please charge my credit card: 5400 1234 5678 9101 2345 6789 Please leave the validation stickers and invoice w

Thanks,

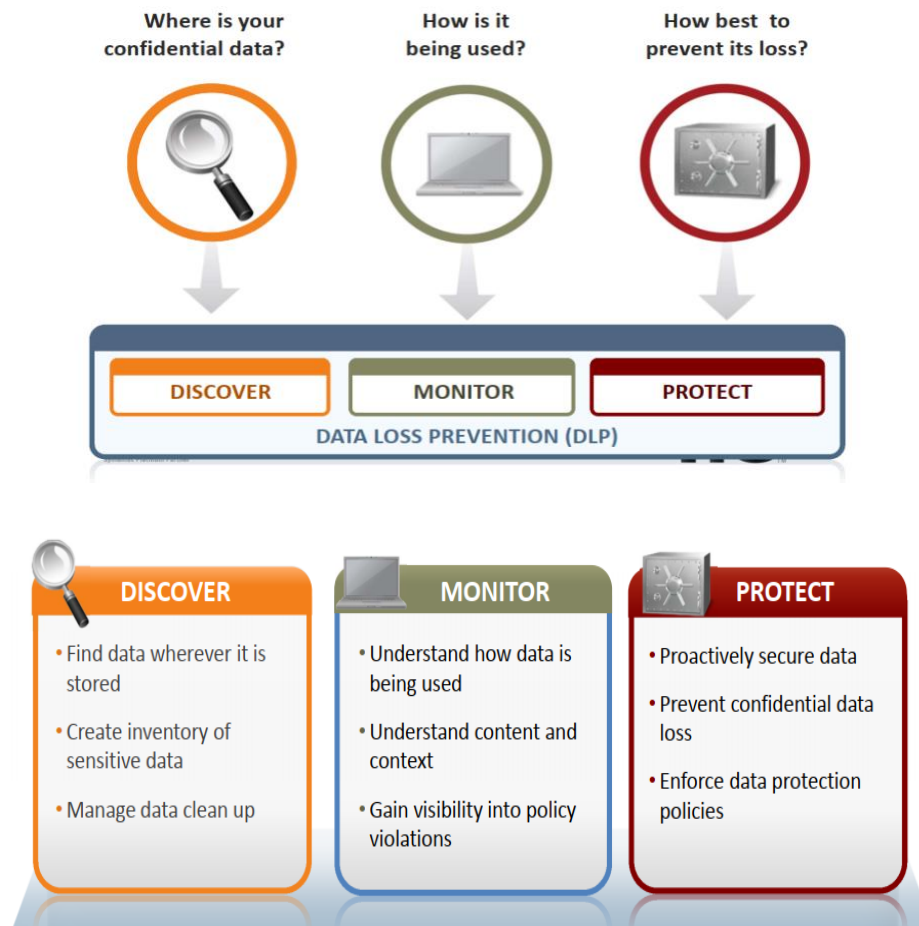
Credit Card number clearly listed in this email

## 4 DLP Core activities

DLP works by performing four key functions:

- **Discover** data wherever it is stored and identify data owners to make data clean-up easy
- **Monitor** how data is being used and where it is going to provide visibility into broken business processes and high-risk users
- **Protect** data by automatically enforcing data loss policies; educating users about security; securing exposed data; and stopping data leaks
- **Manage** data loss policies, incident remediation, and risk reporting from an easy-to-use central management console

**Data loss prevention identifies** potential **data** breaches / **data** ex-filtration transmissions and prevents them by monitoring, detecting and blocking sensitive **data** while in-motion (network traffic), at-rest (**data** storage), and in-use (endpoint actions).





## 4.1 Policies

In the heart of dlp are the policies. Without them there would be no differentiation between public and sensitive data. Policies can be based on the organizations own specifications, but also external requirements, such as pci dss and similar. These policies are then converted into rules that the dlp can enforce during operation.

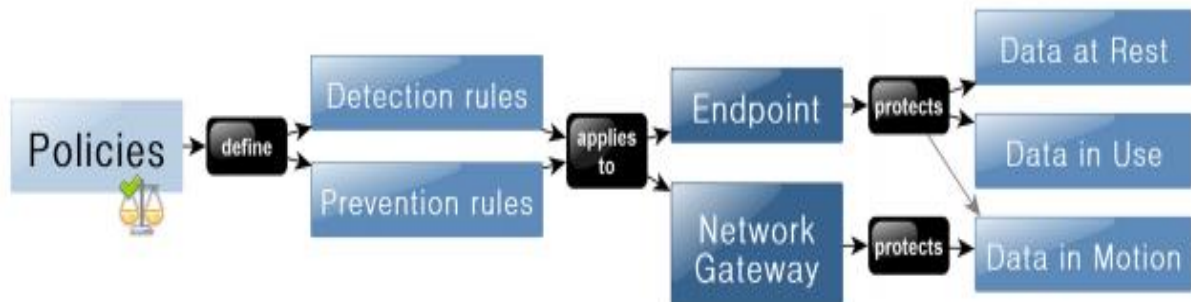


Figure 6: Policy overview

Figure 6 shows how regular policies are converted into rules the DLP software can use to enforce said policies. The detection rules specify how to detect sensitive content, while the prevention rules specify how the detected content should be treated. These rules are then deployed to endpoint agents and the network DLP so they can be used when monitoring the different data states.

### 4.1.1 Policy Violations:

Once a policy violation is discovered, a DLP tool can take a variety of actions:

- **Alert/Report:** Create an incident in the central management server just like a network violation.
- **Warn:** Notify the user via email that they may be in violation of policy.
- **Quarantine/Notify:** Move the file to the central management server and leave a text file with instructions on how to request recovery of the file.
- **Quarantine/Encrypt:** Encrypt the file in place, usually leaving a plain text file describing how to request decryption.
- **Quarantine/Access Control:** Change access controls to restrict access to the file.
- **Remove/Delete:** Either transfer the file to the central server without notification, or simply delete it.

## 4.2 Discover / Identification of Data

The DLP software after capturing the data uses file cracking technology for content analysis, which is used to read and understand the information that is inside the file. Before a DLP system is implemented, organizations often face a situation where their digital data is spread over multiple locations with no control of where sensitive data is located. In DLP terms, content identification is applying policy to identify where sensitive files are located. This can be in databases, on files shares, the local storage on laptops and workstation etc. The identification works similarly to an anti-virus scan, but instead of looking for virus and malware, it looks for sensitive documents and logs their location. From the results administrators and management can decide how to consolidate the files. Content identification deals with data at rest.



**Data Scanning methods:** The scan is done with the help of multiple methods, including simple file crawlers that can be installed on servers and workstation, remote file scans where the management server scans network shares, and endpoint scans where the DLP agents check the local storage of the machines they are installed on. Although content identification is commonly used when initially deploying a DLP, running the scan regularly is not uncommon. Beware that scans can take up system resources and takes time to complete, so it should not be run during work hours.

### 4.2.1 Pre-study of sensitive data of the organization:

In this section, sensitive data are defined and identified. The information in an organization, that is considered sensitive, will differ depending on the business of the organization. Sensitive data may include business records, product designs, personal information, company information, and so on (John, 1996; Eric, 2003; Matt and Sathyanarayana, 2005).

There is some information that will be sensitive in all organizations. Examples are personal information for employees, payroll information, phone numbers and home addresses for employees. [18]

XXXXX Department		
1- Data Manipulation Requirements for Entire Department		
Type of Data	Storage Location ( shared folder on network, employee computer...etc)	Data owner
Data Classification (Strictly Confidential, Confidential, Internal, Public)	Other Departments that need to access this data	Current Security Controls in place

**Figure 3.1: Data Manipulation Requirements**

## 4.2.2 Content Discovery Techniques

Content discovery consists of **three components**, based on where information is stored:

- 1. Storage:** scanning mass storage including file servers, SAN, and NAS.
- 2. Applications/Servers:** application-specific scanning of stored data on email servers, document management systems, and databases.
- 3. Endpoints:** scanning workstations and laptops for content.

There are four basic techniques for content discovery:

- 1. Remote Scanning:** Either the central policy server or a dedicated scanning server accesses storage repository via network shares or other administrative access. Files are scanned for content violations. Connections are often made using administrative credentials, and any content transferred should be encrypted, but this may require reconfiguration of the storage repository and isn't always feasible. Most tools allow bandwidth throttling to limit network impact, and scanning servers are often placed close to the storage to increase speed and limit network impact. This technology supports scanning nearly any storage repository, but even with optimization performance is limited by the network.

2. **Agent-Based Scanning:** An agent is installed on the system (server) to be scanned and scanning is performed locally. Agents are platform specific and use local CPU cycles, but can potentially perform significantly faster than remote scanning, especially for large repositories.

3. **Memory-Resident Agent Scanning:** Rather than deploying a full-time agent, a memory-resident agent is installed which performs a scan and then exits without leaving anything running or stored on the local system. This offers the performance of agent-based scanning in situations where you don't want an agent running all the time.

4. **Application Integration:** Direct integration (often using an agent) with document management, content management, or other storage repositories. This integration not only supports visibility into management content, but allows the discovery tool to understand local context/metadata and possibly enforce actions within the system.

Any of these technologies can work for any of the modes, and enterprises typically deploy a mix depending on policy and infrastructure requirements. We currently see deployments guided by technology limitations of each approach:

- Remote scanning can significantly increase network traffic and has performance limitations based on network bandwidth and target and scanner network performance. Some solutions can scan gigabytes per day (sometimes hundreds, but not terabytes per day) per server due to these practical limitations — which may be inadequate for very large storage. The advantage is that you only need access to a file share for scanning.
- Agents, ephemeral or permanent, are limited by processing power and memory on the target system, which may translate into restrictions on the number of policies that can be enforced, and the types of content analysis that can be used.
- Agents don't support all platforms.

## 4.3 Monitoring of Data

---

NINETY PERCENT OF  
DATA LOSS PREVENTION  
IS ABOUT WHAT YOU  
DO AFTER YOU FIND  
CONFIDENTIAL DATA

---

When sensitive information is identified, DLP triggers alerts and actions to prevent this issue. The critical information should have been previously analyzed and identified by the company in a risk analysis, where it was applied controls to the risks with unacceptable level [2]. The unacceptable risks related to data loss will be reduced to an acceptable level by the DLP deployment. In addition, DLP technology supports the protection of sensitive data not only from intruders, it also protects from internal personnel, who may be braking a process. For instance, if an employee stores sensitive information in his computer and it is not being encrypted, a DLP will be able to recognize this nonconformity After the DLP has access to the content of the information, it generally uses regular expressions, file fingerprinting and dictionaries, or a mix employing a context understanding to detect policy violations related to sensitive information [9]; [8].

### 4.3.1 Regular expressions

define patterns of possible inputs strings [10], which will be related to sensitive information. For example if it is needed to look for the word 'secret' or 'Secret', the regular expression will be '[S/s]ecret'. This will identify files or traffic with the word secret in its content/metadata or payload respectively. Regular expressions are powerful, but they could get really complex as well. IDS technology, such as SNORT [11] uses regular expressions for detection too [12]. Even though regular expressions are a helpful mechanism to detect critical information, they are only useful when the sensitive information target is structured data, i.e. it has a defined format, like a national insurance number, passport number, or credit card numbers, but when the sensitive information does not have a regular pattern or it is unstructured, the only option is to fingerprint the data since it is an undefined format [8].

### 4.3.2 Fingerprint

The **Fingerprint** is a secure hash that is saved in the database of the DLP system. The hash will be compared with the files hash to verify the possible existence of classified data. The DLP combines its mechanism of detection with the context of data location in order to check if the data is in a place that should not be. For instance, if personal data of students is in the secure data base system of the university, it will be not an incident, but if the same data is found in the secretary's computer of the faculty, it should raise a flag and alert about the incident. The fingerprinting technology has the option to perform an exact document matching or a partial document matching [8]. The DLP policies can specified the area where it is going to be applied, i.e. staff network or shared folders.

### 4.3.3 Dictionaries:

Also, the DLP solutions used **Dictionaries** to detect sensitive data. Dictionaries are lists of key terms predefined or user defined, which are related to a sensitive category, such as personal data, strategic documentation, confidentiality terms, etc. In some cases, the DLP technology is able to read these terms in real time from a database giving a more accurate source of information [4]

DLP endpoint protection is installed on workstations and other devices in the form of an agent. The agent enforces policies by monitoring all data activity and scanning all locally stored files. Usually the agent also allows physical input to be controlled. This means an administrator can centrally disable USB, FireWire and other interfaces with ease. Also, the act of burning CDs or DVDs can be prevented.

DLP endpoints feature different types of protection that can be categorized as follows: [11]

### 4.3.4 File system protection:

Monitors all file operations similarly to real-time protection found in anti-virus software. This is to make sure files are not copied to unauthorized locations or that encryption is automatically applied by the DLP when saving data to said locations. Scans can also be performed on stored data to discover policy violations.

### 4.3.5 Network protection:

Monitors data being transmitted over the network when the endpoint is away from the corporate network. Otherwise, the network DLP is responsible for this functionality.

### 4.3.6 Kernel protection:

DLP integrated in the OS and applications to prevent actions such as copying to the clipboard, taking screenshot or typing sensitive data into chat programs.

Various methods to analyze and discover sensitive data exists. The most common methods are the use of **keyword matching, regular expressions and data fingerprinting/hashing**. Although not widely implemented, statistical methods - like those used for spam mail - can also be used to identify sensitive data and is an area of high interest to the DLP industry.

### 4.3.7 Content Analysis Techniques

Once the content is accessed, following major analysis techniques will be used to find policy violations, each with its own strengths and weaknesses.

#### 4.3.7.1 Rules-Based/Regular Expressions:

This is the most common analysis technique available in both DLP products and other tools with DLP features. It analyzes the content for specific rules — such as 16 digit numbers that meet credit card checksum requirements, medical billing codes, and other textual analyses. Most DLP solutions enhance basic regular expressions with their own additional analyses (*e.g.*, a name in proximity to an address near a credit card number).

***What it's best for:*** As a first-pass filter, or for detecting easily identified pieces of structured data like credit card numbers, Social Security numbers, and healthcare codes/records.

***Strengths:*** Rules process quickly and can be easily configured. Most products ship with initial rule sets. The technology is well understood and easy to incorporate into a variety of products.

***Weaknesses:*** Prone to higher false positive rates. Offers very little protection for unstructured content such as sensitive intellectual property.

#### 4.3.7.2 Database Fingerprinting:

Sometimes called Exact Data Matching, this technique takes either a database dump or live data (via ODBC connection) from a database and only looks for exact matches. For example, you could generate a policy to look only for credit card numbers in your customer base, thus ignoring your own employees buying online. More advanced tools look for combinations of information, such as the magic combination of first name or initial, with last name, with credit card or Social Security

number, that triggers most US state-level breach disclosure laws. Make sure you understand the performance and security implications of nightly extracts vs. live database connections.

***What it's best for:*** Structured data from databases.

***Strengths:*** Very few false positives (close to 0). Allows you to protect customer & sensitive data while ignoring other, similar, data used by employees, such as their personal credit cards for online orders.

***Weaknesses:*** Nightly dumps don't include transaction data since the last extract. Live connections can affect database performance. Large databases affect product performance.

#### **4.3.7.3 Partial Document Matching:**

This technique looks for a complete or partial match to protected content. You could build a policy to protect a sensitive document, and the DLP solution will look for either the complete text of the document, or even excerpts as small as a few sentences. For example, you could load up a business plan for a new product and the DLP solution would alert if an employee pasted a single paragraph into an instant message. Most solutions are based on a technique known as cyclical (or overlapping) hashing, where you take a hash of a portion of the content, offset a predetermined number of characters, then take another hash, and keep adding hashes of document segments until done. Outbound content is run through the same hash technique, and the hash values compared for matches. Some products use cyclical hashing as a base, then add more advanced linguistic analysis.

***What it's best for:*** Protecting sensitive documents, or similar content with text, and source code. Unstructured content that's known to be sensitive.

***Strengths:*** Ability to protect unstructured data. Generally low false positives (some vendors will say zero false positives, but any common sentence/text in a protected document can trigger alerts). Doesn't rely on complete matching of large documents; can find policy violations on even a partial match.

***Weaknesses:*** Performance limitations on the total volume of content that can be protected. Common phrases/verbiage in a protected document may trigger false positives. Must know exactly which documents you want to protect. Trivial to avoid (ROT 13 encryption is sufficient for evasion).



#### 4.3.7.4 Statistical Analysis:

Use of machine learning, Bayesian analysis, and other statistical techniques to analyze a corpus of content and find policy violations in content that resembles the protected content. This category includes a wide range of statistical techniques which vary greatly in implementation and effectiveness. Some techniques are very similar to those used to block spam. Of all the techniques listed, this is the least commonly supported by different products.

***What it's best for:*** Unstructured content where a deterministic technique, such as partial document matching would be ineffective. For example, a repository of engineering plans that's impractical to load for partial document matching due to high volatility or extreme volume.

***Strengths:*** Can work with more nebulous content where you may not be able to isolate exact documents for matching. Can enforce policies such as “Alert on anything outbound that resembles the documents in this directory.”

***Weaknesses:*** Prone to false positives and false negatives. Requires a large corpus of source content — the bigger the better.

#### 4.3.7.5 Conceptual/Lexicon:

This technique uses a combination of dictionaries, rules, and other analyses to protect nebulous content that *resembles* an ‘idea’. It’s easier to give an example — a policy that alerts on traffic that resembles

insider trading, which uses key phrases, word counts, and positions to find violations. Other examples are sexual harassment, running a private business from a work account, and job hunting.

***What it's best for:*** Completely unstructured ideas that defy simple categorization but are like known documents, databases, or other registered sources.

***Strengths:*** Not all corporate policies or content can be described using specific examples — conceptual analysis can find loosely defined policy violations other techniques can’t even try to be monitoring for.

***Weaknesses:*** In most cases these are not user-definable, and the rule sets must be built by the DLP vendor with significant effort (costing more). Because of the loose nature of the rules, this technique is very prone to both false positives and false negatives.

#### 4.3.7.6 Categories:

Pre-built categories with rules and dictionaries for common types of sensitive data, such as credit card numbers/PCI protection, HIPAA, etc.

**What it's best for:** Anything that neatly fits a provided category. Typically easy to describe content related to privacy, regulations, or industry-specific guidelines.

**Strengths:** Extremely simple to configure. Saves significant policy generation time. Category policies can form the basis for more advanced enterprise-specific policies. For many organizations, categories can meet a large percentage of their data protection needs.

**Weaknesses:** One size fit all might not work. Only good for easily categorized rules and content.

These techniques serve as the basis for most of the DLP products on the market. Not all products include all techniques, and there can be significant differences between implementations. Most products also support chaining techniques — building complex policies from combinations of different content and contextual analysis techniques.

## 4.4 Protect/Prevention/Management of Data Leakage

### 4.4.1 Blocking Copy Functionality

Although blocking the ability to copy data to the clipboard is present in traditional endpoints, the DLP version is more intelligent. With DLP, when you copy data, the action is only blocked when sensitive data is copied. If non-sensitive data is being handled, the DLP will not interfere and as a result be less of an annoyance. For users commonly working with sensitive data, this functionality can be disabled and at the same time enabled for other groups of users.

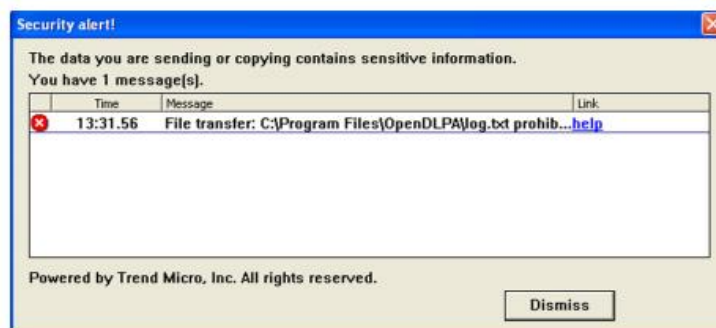


Figure 8: Trend Micro DLP blocking copying of the sensitive word "masteroppave".

#### **4.4.2 Blocking Inputs and Devices**

To disable USB inputs various methods are used. One way of doing this is to disable the devices in Windows' device manager. But if more functionality is needed, such as allowing employees to use company issued USB-drives, disabling the USB controller is not a viable solution. A more practical solution is to hook into the USB mass storage library and as soon as a new device is discovered, the ID can be compared against a list of allowed devices (see Figure 9, next page). The device is allowed to mount if it matches the ID of a company issued USB drive. If not, the device is either ejected or mounted as read-only.

#### **4.4.3 Blocking Application Input**

Many DLP products advertise protection of data channels provided by popular communication applications such as Skype or Yahoo Messenger. In this case the DLP monitors all files the given process (e.g., skype.exe) accesses, and when sensitive files are involved, all read access is blocked (see Figure 10). Depending on the DLP product, this does not necessarily affect the text input of the program, meaning that sensitive data can still be sent by manually typing it in. On the other hand, network traffic containing sensitive data can still be intercepted by the network DLP and get blocked at this point. This is of course only relevant when the traffic is unencrypted, which does not apply to Skype traffic.

#### **4.4.4 Protecting E-Mail**

DLP offers two ways of protecting e-mail communication; integration with the mail server and endpoint integration with the local e-mail client. Figure 11 below illustrates a typical scenario where a local user sends the e-mail (marked as 1) to the mail transfer agent (MTA), using for example SMTP. Before sending the message out of the network, it is first forwarded to the DLP server for inspection. The DLP can then block or allow the message depending on the sensitivity of the content. How this integration is done varies between vendors. In many cases the DLP can even encrypt or mark messages for encryption as this is a requirement by many businesses. [15]

#### **4.4.5 Automatic Data Removal**

When defining a policy in the DLP, you first define a detection rule, and then an action that is to be taken if the detection rule is triggered. This action depends on the DLP product, but some common ones are listed here:

- **Block:** The operation related to the sensitive data is blocked.
- **Allow:** The operation related to the sensitive data is allowed.
- **Alert:** An alert dialog is displayed when the user triggers the DLP.
- **Encrypt:** The data is automatically encrypted when it is for example saved to a USB drive or transferred to a specific location.
- **Justify:** The user must give a reason when the DLP is triggered. If it was accidental the user can cancel the original operation.
- **Log:** Logs the incident.

All the above actions can be combined so that when the DLP triggers, multiple actions are executed. Having these actions makes it easier for users to work where a DLP is in place. In a situation where you trust the user, having the “justify” option enabled, means accidental leaks can be avoided, while still allowing sensitive documents to be legitimately sent and false positives to be bypassed. In large complex environments, the actions can be set up on a per group basis, to be as little intrusive to the employees as possible while still providing data protection [6]

#### 4.4.6 Dealing with Encryption

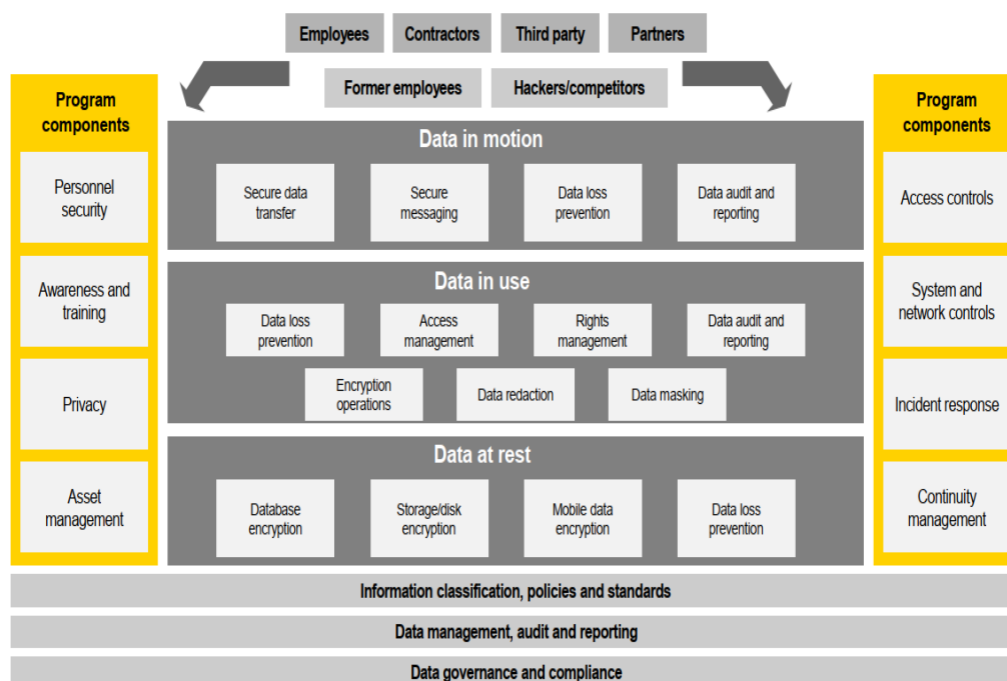
To ensure secure communication over the internet, it is essential that companies take advantage of some form of encryption. A recent trend is that web services redirect users to their TLS protected page (HTTPS), which in return offers better protection of the data being transmitted. For companies, using HTTPS and other encrypted communication channels offers a secure way of transmitting sensitive data. From the viewpoint of DLP, encryption can be difficult to work with. Dealing with encryption is especially problematic for the network DLP, which cannot analyze encrypted data in a comprehensible manner. If sensitive documents are leaked through an encrypted connection, the DLP will be unable to block it. For endpoint DLP the same applies to encrypted files. If a file is encrypted, the DLP will not be able to read the contents and enforce policy. The good news then is that loss of sensitive data via encryption is relatively rare in practice.

For secure communication it is possible to use proxy servers to intercept the communication. Basically, the proxy, or network DLP in our case, acts as a reverse proxy and launches a man-in-the-middle attack as soon as a client wants to establish a connection with an HTTPS server [16].

Normally, the client's browser would issue a certificate warning, but since it is configured with the company's certificate authority (CA), the certificate is trusted. A drawback to this is that any browsers missing the CA in its configuration will still issue a warning. Additionally, the DLP will still not be able to interpret any files that were encrypted or obfuscated before transmission. With endpoint DLP the action can be blocked before the document is encrypted and transmitted. Basically, any file access to an untrusted program (e.g., encryption tool) can be denied. Of course, not giving local users permission to install and use third party tools will go a long way in combatting this problem. As the next sub-chapter demonstrates, users can still obfuscate documents if they have write privileges to them, so being able to encrypt files or the communication is not necessarily required.

## 5 DLP Implementation/Deployment Methods

In addition to protecting sensitive data, a modern DLP should be adaptive, mobile and as minimally intrusive as possible [7]. Adaptive means that it can work in different environments and be configured to meet the needs of a wide range of different businesses. Mobile means that it can still protect the data, even when the device is used outside the company network. The products today only fulfill this to a certain degree. DLP is still maturing, but unlike a few years ago, most vendors have standardized on the core functionality that defines a modern DLP solution. [6]

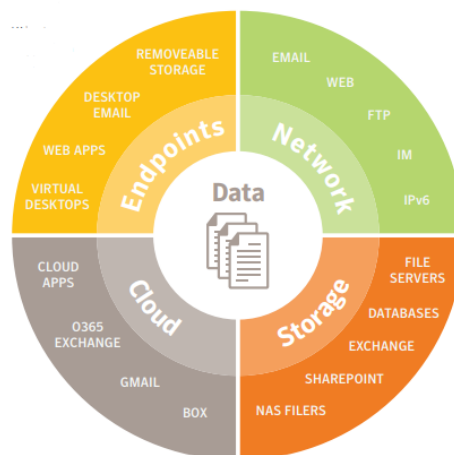


## 5.1 Protecting Data in Motion, at Rest, and in Use

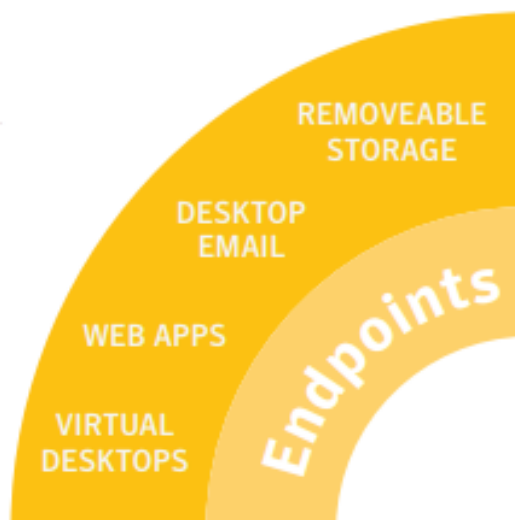
The goal of DLP is to protect content throughout its lifecycle — on the network, in storage, and on endpoints. In terms of DLP, this includes three major aspects:

- Data in Motion protection is monitoring (and potentially filtering) of traffic on the network (passively or inline via proxy) to identify content being sent across specific communications channels. This would include monitoring email, instant messages, and web traffic for snippets of sensitive source code. In motion tools can often block based on central policies, depending on the type of traffic.
- Data at Rest is protected by scanning of storage and other content repositories to identify where sensitive content is located. We often call this content discovery. For example, you can use a DLP product to scan your servers and identify documents with credit card numbers. If the server isn't authorized for that kind of data, the file can be encrypted or removed, or a warning sent to the file owner.
- Data in Use is addressed by endpoint solutions that monitor data as the user interacts with it. For example, they can identify when you attempt to transfer a sensitive document to a USB drive and block it (as opposed to blocking use of the USB drive entirely). Data in use tools can also detect things like copy and paste, or use of sensitive data in an unapproved application (such as someone attempting to encrypt data to sneak it past the sensors).

## KEEPING DATA SAFE:



### 5.1.1 Keep data safe while in use on endpoints



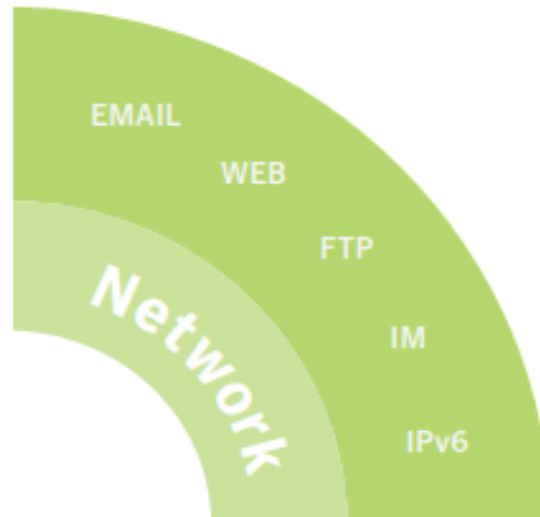
DLP endpoint protection is installed on workstations and other devices in the form of an agent. The agent enforces policies by monitoring all data activity and scanning all locally stored files. Usually the agent also allows physical input to be controlled. This means an administrator can centrally disable USB, FireWire and other interfaces with ease. Also, the act of burning CDs or DVDs can be prevented. DLP endpoint integrates with the OS kernel and program functionality.

**DLP Endpoint Discover** scans local hard drives and gives you deep visibility into sensitive files that users are storing on their laptops and desktops. It provides a wide range of responses including local and remote file quarantining, and policy-based encryption and digital rights management enabled by the DLP Endpoint Flex Response API

**DLP Endpoint Prevent** monitors users' activities and gives you fine-grained control over a wide range of applications, devices and platforms. It provides a wide range of responses including identity-based encryption and digital rights for files transferred to USB. With Endpoint Prevent, you can alert users to incidents using on-screen popups or email notifications. Users can also override policies by

providing a business justification or canceling the action (in the case of a false positive).

### 5.1.2 Protect data in motion over the network



DLP for Network solution monitors and prevents sensitive data from being leaked over a wide range of communication protocols across your network.

**DLP Network Monitor** captures and analyzes outbound traffic on your corporate network, and detects sensitive content and metadata over standard, non-standard and proprietary protocols. It is deployed at network egress points and integrates with your network tap or Switched Port Analyzer (SPAN). Network Monitor performs deep content inspection of all network communications with zero packet loss, unlike other solutions that sample packets during peak loads and put you at high risk for false negatives.

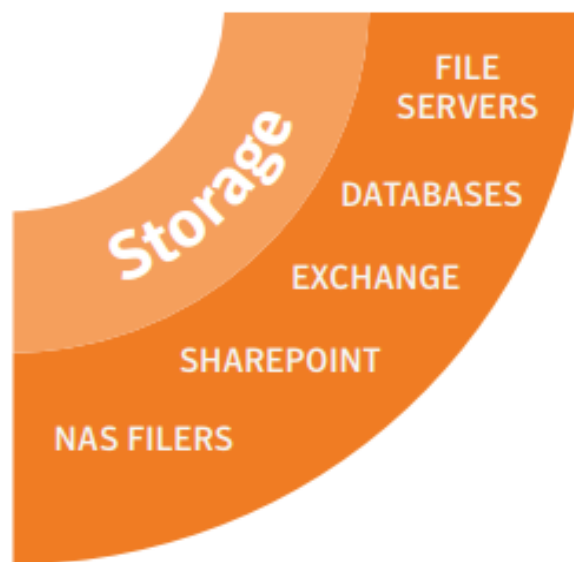
**DLP Network Prevent for Email** protects sensitive messages from being leaked or stolen by employees, contractors, and partners. It monitors and analyzes all corporate email traffic, and optionally modifies, redirects, or blocks messages based on sensitive content or other message attributes. Network Prevent for Email is deployed at network egress points and integrates with mail transfer agents (MTAs) and cloud-based email including Microsoft® Office 365 Exchange. Network Prevent for Email is available as software or virtual appliance.



**DLP Network Prevent for Web** protects sensitive data from being leaked to the Web.

It monitors and analyzes all corporate web traffic, and optionally removes sensitive HTML content or blocks requests. Network Prevent for Web is deployed at network egress points and integrates with your HTTP, HTTPS or FTP proxy server using ICAP. Network Prevent for Web is available as software, hardware appliance, or virtual appliance.

### 5.1.3 Protect data at rest across storage repositories

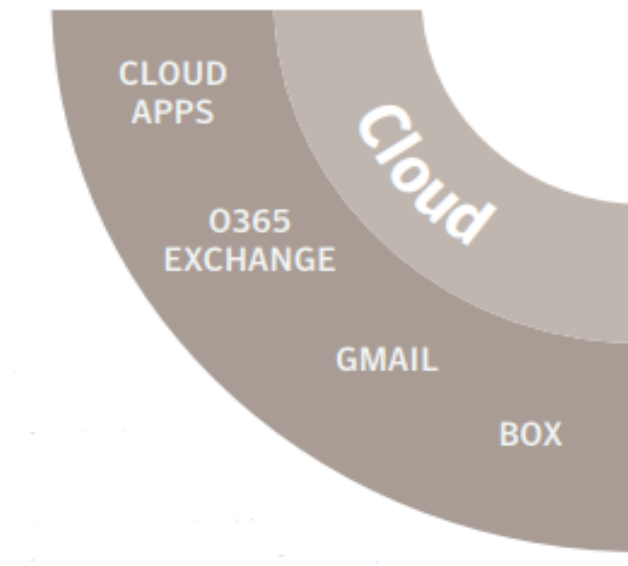


With **Symantec DLP for Storage**, you can discover and secure sensitive data at rest – the data stored on file servers, endpoints, cloud storage, network file shares, databases, SharePoint, and other data repositories.

First, **DLP Network Discover** finds confidential data by scanning network file shares, databases, and other enterprise data repositories. This includes local file systems on Windows, Linux, AIX, and Solaris servers; Lotus Notes and SQL databases; and Microsoft Exchange and SharePoint servers. DLP Network Discover recognizes custom file types—based on the binary signature of the file. It also provides high-speed scanning for large, distributed environments, and it optimizes performance by scanning only new or modified files.

Next, **DLP Network Protect** adds robust file protection capabilities on top of Network Discover. Network Protect automatically cleans up and secures all of the exposed files Network Discover detects, and it offers a broad range of remediation options, including quarantining or moving files, copying files to a quarantine area, or applying policy identity-based encryption and digital rights to specific files. Network Protect even educates business users about policy violations by leaving a marker text file in the file's original location to explain why it was quarantined.

#### 5.1.4 Protect data in the cloud



With DLP Cloud Services, you can extend powerful data protection controls to the cloud with the convenience of cloud-delivered DLP. They provide rich discovery, monitoring and protection capabilities for a wide range of cloud applications as well as on-premises applications.

The **DLP Cloud Detection Service** inspects content extracted from cloud app and web traffic, and automatically enforces sensitive data policies. It offers enhanced cloud-to-cloud integration with Symantec CloudSOC, our industry leading cloud access security broker solution, to protect data in motion and data at rest across more than 100 in sanctioned and unsanctioned cloud apps such as Office 365, G-Suite, Box, Dropbox, and Salesforce. The integration allows to extend existing policies and robust detection to cloud applications and manage all incidents from the DLP console. Controls include un-share sensitive files, quarantine, block them from leaving the application, and also apply identity-based encryption and digital rights automatically to specific files shared with third parties.

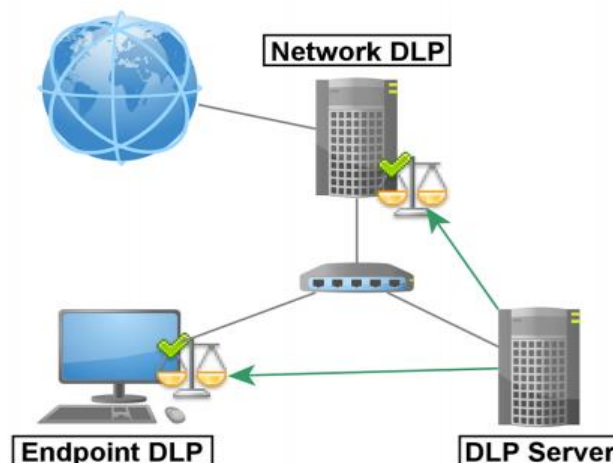
The **DLP Cloud Service for Email** provides accurate, real-time monitoring of corporate email traffic by leveraging built-in intelligence and advanced detection capabilities that minimize false positives. It also provides real-time protection against data leaks with automated messaging blocking, or message modification to enforce downstream encryption or quarantining. When data is shared with third parties it can automatically enable identity-based encryption and digital rights for email bodies and attachments. The DLP Cloud Service for Email supports Gmail for Work, Microsoft Office 365 Exchange Online as well as Microsoft Exchange Server.

## 5.2 DLP depending on the state of data:

DLP is applied differently depending on what state the data is in. For data at rest, the content of stored data is scanned; for data in use, the DLP interacts with input given to programs and OS; and for data in motion each data packet is analyzed. To effectively protect an organization, all these channels have to be monitored and managed, which makes DLP a bit more complex than the average firewall or anti-virus solution.

## 5.3 Implementation of DLP on Endpoint and Network

Figure 5 below gives a basic overview of the physical parts common to a DLP system. The Endpoint DLP is installed directly on a workstation and keeps track of how data is stored (data at rest) and used (data in use). The network DLP is often placed between the LAN and WAN as a proxy which monitors network traffic (data in motion). The DLP server manages both these components and is mainly responsible for policy deployment (illustrated with green arrows) and logging policy violations.



## **5.4 Endpoint Features and Integration**

DLP features have appeared in various endpoint tools aside from dedicated DLP products since practically before there was a DLP market. This continues to expand, especially as interest grows in controlling USB usage without onerous business impact.

### **5.4.1 USB/Portable Device Control:**

A frequent inhibitor to deployment of portable storage management tools is their impact on standard business processes. There is always a subset of users who legitimately needs some access to portable storage for file exchange (e.g., sales presentations), but the organization still wants to audit or even block inappropriate transfers. Even basic content awareness can clearly help provide protection while reducing business impact. Some tools include basic DLP capabilities, and we are seeing others evolve to offer somewhat extensive endpoint DLP coverage — with multiple detection techniques, multivariate policies, and even dedicated workflow. This is also a common integration/partner point for full DLP solutions, although due to various acquisitions we don't see those partnerships quite as often as we used to. When evaluating this option, keep in mind that some tools position themselves as offering DLP capabilities but lack any content analysis; instead relying on metadata or other context. Finally, despite its incredible usefulness, we see creation of shadow copies of files in many portable device control products, but almost never in DLP solutions.

### **5.4.2 Endpoint Protection Platforms:**

For those of you who don't know, EPP is the term for comprehensive endpoint suites that include antivirus, host intrusion prevention, and everything from remote access and Network Admission Control to application whitelisting. Many EPP vendors have acquired full or endpoint-only DLP products and are in various stages of integration. Other EPP vendors have added basic DLP features — most often for monitoring local files or storage transfers of sensitive information. So there are options for either basic endpoint DLP (usually some preset categories), all the way up to a DLP client integrated with a dedicated DLP suite.

### **5.4.3 Non-Antivirus EPP:**

There are also endpoint security platforms that are dedicated to more than just portable device control, but not focused on antivirus like other EPP tools. This category covers a range of tools, but the features offered are generally comparable to the other offerings. Overall, most people deploying DLP features on an endpoint (without a dedicated DLP solution) are focused on scanning the local hard drive and/or monitoring/filtering file transfers to portable storage. But as we described earlier you might also see anything from network filtering to application control integrated into endpoint tools.

## **5.5 Networks**

Many organizations first enter the world of DLP with network-based products that provide broad protection for managed and unmanaged systems. It's typically easier to start a deployment with network products to gain broad coverage quickly. Early products limited themselves to basic monitoring and alerting but all current products include advanced capabilities to integrate with existing network infrastructure to provide protective, not just detective, controls.

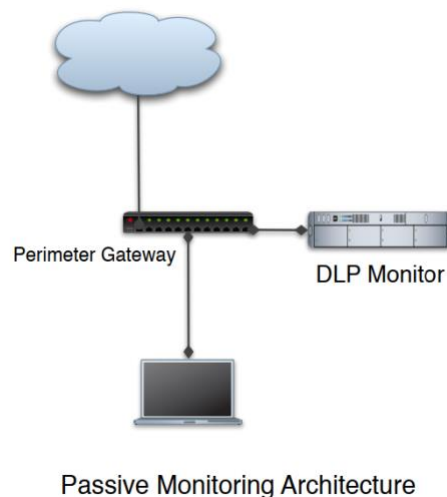
### **5.5.1 Network Monitor**

At the heart of most DLP solutions lies a passive network monitor. The network monitoring component is typically deployed at or near a gateway on a SPAN or mirror port (or a similar tap). It performs full packet capture, session reconstruction, and content analysis in real time. Performance is more complex and subtle than vendors normally discuss. First, on the client expectation side, most clients claim they need performance to match their full peak bandwidth, but that level of performance is unnecessary except in very unusual circumstances because few organizations are really running that high a level of communications traffic. DLP is a tool to monitor employee communications, not all network traffic. Realistically we find that small enterprises normally run under 50mbyte/sec of relevant traffic, medium enterprises run in the 50-200mbyte/s range, and large enterprises around 300mbyte/s (as high as 500 in a few cases). Because of the content analysis overhead, not every product runs full packet capture. You might have to choose between pre-filtering (and thus missing non-standard traffic) or buying more boxes and load balancing. Also, some products lock monitoring into pre-defined port and protocol combinations, rather than using service/channel identification based on packet content. Even if full application channel identification is included, you will need to make sure it's enabled. Otherwise,

you might miss nonstandard communications such as connecting over an unusual port. Most of the network monitors are dedicated general-purpose server hardware running DLP software. A few vendors deploy true specialized appliances.

Also keep in mind, especially when testing, that performance is often tied to the number and scope of DLP policies you deploy. If you perform large amounts of partial document matching or database fingerprinting (both of which rely on repositories of hash values) you might find performance slowing and need to move toward load balancing or separating traffic streams. **For example**, you can offload email to a dedicated monitor because, as we'll discuss in a moment, email monitoring using a different architectural model than web and most other traffic.

Finally, there are some organizations which exceed the average monitoring requirements in terms of both bandwidth and port/protocol coverage due to either their size or the nature of threat they face (advanced attackers more commonly exfiltrate data over unusual ports & protocols). If you fall into this category, make sure you include it in your DLP requirements and test it in the selection process. You should also coordinate your DLP program with any other egress filtering projects, which may be able to reduce your DLP load.



## 5.6 Storage Features and Integration

We don't see nearly as much DLP Light in storage as in networking and endpoints — in large part because there aren't as many clear security integrations points. Fewer organizations have any sort of storage security monitoring, whereas nearly every organization performs network and

endpoint monitoring of some sort. But while we see less DLP Light, as we have already discussed, we see extensive integration on the DLP side for different types of storage repositories.

#### **5.6.1 Database Activity Monitoring and Vulnerability Assessment:**

DAM products, many of which now include or integrate with Database Vulnerability Assessment tools, now sometimes include content analysis capabilities. These are designed to either find sensitive data in large databases, detect sensitive data in unexpected database responses, or help automate database monitoring and alerting policies. Due to the high potential speeds and transaction volumes involved in real time database monitoring, these policies are usually limited to rules/patterns/categories. Vulnerability assessment policies may include more options because the performance demands are different.

#### **5.6.2 Vulnerability Assessment:**

Some vulnerability assessment tools can scan for basic DLP policy violations if they include the ability to passively monitor network traffic or scan storage.

#### **5.6.3 Document Management Systems:**

This is a common integration point for DLP solutions, but we don't see DLP included as a DMS feature.

#### **5.6.4 Content Classification, Forensics, and Electronic Discovery:**

These tools aren't dedicated to DLP, but we sometimes see them positioned as offering DLP features. They do offer content analysis, but usually not advanced techniques like partial document matching and database fingerprinting/matching.

## **5.7 Other Features and Integrations**

The lists above include most of the DLP Light, feature, and integration options we've seen; but there are few categories that don't fit quite as neatly into our network/endpoint/storage divisions:

#### **5.7.1 SIEM and Log Management:**

All major SIEM tools can accept alerts from DLP solutions and possibly correlate them with other collected activity. Some SIEM tools also offer DLP features, depending on what kinds of activity they can collect to perform content analysis on. Log management tools tend to be more

passive, but increasingly include some similar basic DLP-like features when analyzing data. Most DLP users tend to stick with their DLP solutions for incident workflow, but we do know cases where alerts are sent to the SIEM for correlation or incident response, as well as when the organization prefers to manage all security incidents in the SIEM.

#### **5.7.2 Enterprise Digital Rights Management:**

Multiple DLP solutions now integrate with Enterprise DRM tools to automatically apply DRM rights to files that match policies. This makes EDRM far more usable for most organizations, since one major inhibitor is the complexity of asking users to apply DRM rights. This integration may be offered both in storage and on endpoints, and we expect to see these partnerships continue to expand.

#### **5.7.3 Email Encryption:**

Automatic encryption of emails based on content was one of the very first third-party integrations to appear on the market, and a variety of options are available. This is most frequently seen in financial and healthcare organizations (including insurance) with strict customer communications security requirements.

### **5.8 Distributed and Hierarchical Deployments:**

All medium to large enterprises, and many smaller organizations, have multiple locations and web gateways. A DLP solution should support multiple monitoring points, including a mix of passive network monitoring, proxy points, email servers, and remote locations. While processing/analysis can be offloaded to remote enforcement points, they should send all events back to a central management server for workflow, reporting, investigations, and archiving. Remote offices are usually easy to support because you can just push policies down and reporting back, but not every product offers this capability.

The more advanced products support hierarchical deployments for organizations that want to manage DLP differently between multiple geographic locations, or by business unit. International companies often need this to meet legal monitoring requirements which vary by country. Hierarchical management supports coordinated local policies and enforcement in different regions, running on their own management servers, communicating back to a central management server. Early products only supported one management server but now we have options to deal with these



distributed situations, with a mix of corporate/regional/business unit policies, reporting, and workflow. With so much sensitive information moving around, it's important each tier in the hierarchy is well secured and all communications between DLP nodes is encrypted.

## 6 DLP Products/Solutions available in Market

The table below shows features of different DLP products available in market: [11]

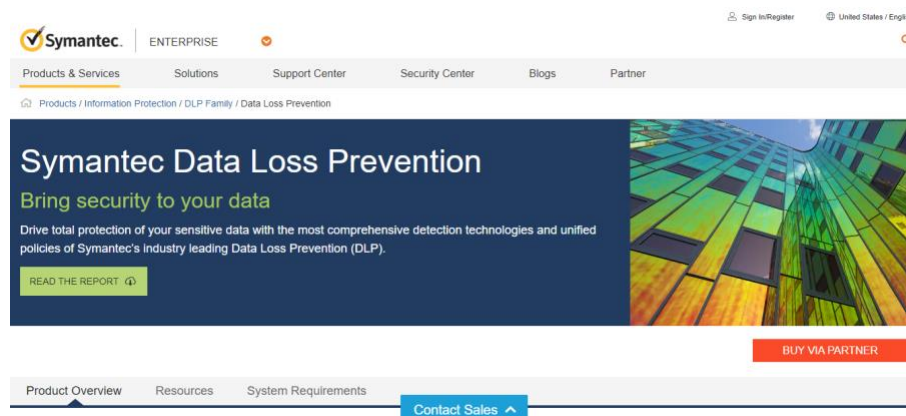
Software	Platform	Features	Cloud Compatibility	Pros	Cons
<b>OpenDLP (free)</b>	Web	Data Management, Tracking	No	Open-source versatility, free	a bit hard to configure, scarce support
<b>MYDLP by Comodo (free)</b>	Desktop	Monitor, Discover, Prevent data leaks	Unknown	Centralized Management, Google-like search engine, easy to use	No pricing information available, limited documentation available
<b>McAfee</b>	Desktop	Excellent forensic analysis, Data management	Yes	Intelligent data prioritization, forensic data analysis	Complicated to setup and manage, it does not offer free trial
<b>Digital Guardian Endpoint DLP</b>	Desktop	Great encryption, Data management, Tracking	Yes	Versatility, can be on-premise, cloud or hybrid, works with windows, mac & linux endpoints	Expensive licensing
<b>Symantec Data Loss Prevention</b>	Desktop	Data management, Tracking	Yes	Excellent cloud compatibility, extreme scalable	Could be too-enterprise oriented

<b>Check Point Data Loss Prevention</b>	Desktop	Data remediation, Education	No	Simple interface, Easy to use, a single management console	Some might find it too simple
<b>Safetica Data Loss Prevention</b>	Desktop	Data leak prevention	No	Easy to use, proper monitoring tool	A bit expensive, support is limited to online

In conclusion, as time passes, new technologies have found their way into DLP, classification of data by content and by the user is of them. Also, new safety features had to be added regarding the cloud and online file storage services, Therefore finding the best data loss prevention solution for your organization could be a top priority as no any DLP software gives complete protection of the data of your organization, each has its own specialized features, but just in case Hybrid DLP Software can come in handy.

## 6.1 Some Views of the DLP Products:

### 6.1.1 SYMANTEC



## 6.1.2 McAfee



**McAfee**  
Together is power.

Get Total Protection for All Your Devices with McAfee® Total Protection™

Stay protected from the latest threats - ransomware, viruses, malware, spyware, unwanted programs, and more.

**\$49.99** ~~\$\$\$9.99~~ **\$70 OFF\***

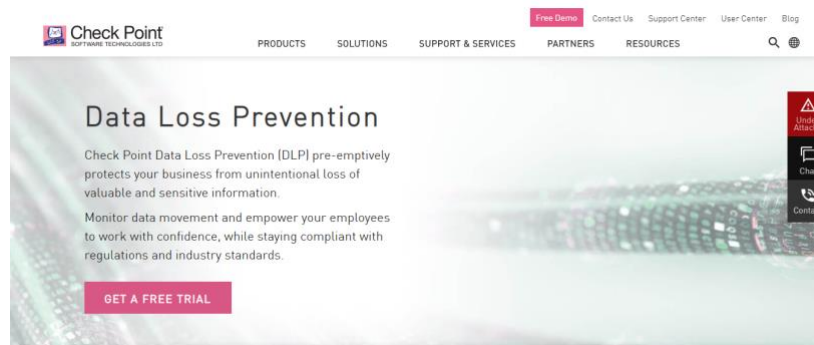
Protection for 10 Devices

[BUY NOW](#)

\*Price shown is for 1 year. See offer details below.

The advertisement features a family of three (a man, a woman, and two children) looking at a laptop together, suggesting a home environment. The background is a soft-focus image of a bookshelf.

## 6.1.3 Check Point



**Check Point**  
SOFTWARE TECHNOLOGIES LTD.

PRODUCTS SOLUTIONS SUPPORT & SERVICES PARTNERS RESOURCES

[Free Demo](#) [Contact Us](#) [Support Center](#) [User Center](#) [Blog](#)

### Data Loss Prevention

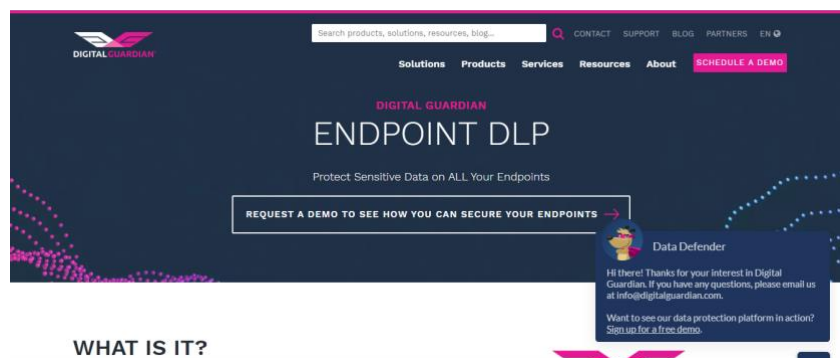
Check Point Data Loss Prevention (DLP) pre-emptively protects your business from unintentional loss of valuable and sensitive information.

Monitor data movement and empower your employees to work with confidence, while staying compliant with regulations and industry standards.

[GET A FREE TRIAL](#)

The webpage has a clean, professional design with a light blue and white color scheme. A sidebar on the right contains icons for 'Join Attack!', 'Chat', and 'Contact'.

## 6.1.4 EndPoint DLP



**DIGITAL GUARDIAN**

Search products, solutions, resources, blog... [CONTACT](#) [SUPPORT](#) [BLOG](#) [PARTNERS](#) [EN](#)

[Solutions](#) [Products](#) [Services](#) [Resources](#) [About](#) [SCHEDULE A DEMO](#)

### DIGITAL GUARDIAN ENDPOINT DLP

Protect Sensitive Data on ALL Your Endpoints

[REQUEST A DEMO TO SEE HOW YOU CAN SECURE YOUR ENDPOINTS](#)

**Data Defender**

Hi there! Thanks for your interest in Digital Guardian. If you have any questions, please email us at [info@digitalguardian.com](mailto:info@digitalguardian.com).

Want to see our data protection platform in action? [Sign up for a free demo.](#)

[WHAT IS IT?](#)

The webpage features a dark blue background with pink and white text. A search bar and navigation links are at the top. A prominent call-to-action button is in the center. A chatbot window is open on the right side.

# 7 DLP Benefits

The biggest benefit of an enterprise data loss prevention software is that it can display all the monitoring information on a single console (centrally managed). Business organizations go

through major financial losses and reputational damage when they experience loss of sensitive data and other forms of enterprise information. Companies are now very much aware of these dangers and hence data protection has become the most trending topic, however many organizations fail to completely understand the business case for Data Loss Prevention (DLP) initiatives. Given below are some of the key reasons why an organization needs DLP: [17]

- DLP technology provides IT and security staff with a 360-degree view of the flow, location, and usage of data across the enterprise. It is capable of checking network actions against an organization's security policies, and enables you to protect and control sensitive data, including personally identifiable information (PII), financial data, customer information, and intellectual property.
- When used along with complementary controls, DLP enables preventing the accidental exposure of personal information across all devices. Wherever data lives, DLP has the potential to monitor it and majorly reduce the risk of data loss.
- Technology controls are becoming essential to attain compliance in specific areas. DLP provides these controls, including policy templates and maps that automate compliance, address particular requirements, and enable the collection and reporting of metrics.
- DLP provides updated policy templates and maps that address specific requirements, help in the collection and reporting of metrics, and automate compliance. After a policy need is detected, DLP can make the modification as simple as helping a suitable policy template on your system.
- When organizations fail to adopt the necessary steps to detect sensitive data and protect it from misuse or loss, they are actually risking their potential to compete. Companies that obtain data protection and privacy right can boost their brand reputation and resilience going forward. However, those that get it wrong are likely to end up in financial loss and reputational damage. DLP thus enables protecting critical data and preventing negative publicity and loss of revenue that certainly follow data breaches.

Data loss is an umbrella term that describes the program, governance, policy intantiation, management controls and solution implementationof people, process and technology measures to prevent the loss of, or unauthorized access to sensitive data.

## **8 DLP Limitations**

Many of the features of DLP overlap with the security features of firewalls, intrusion detection systems and certain endpoint software. [6]

For an attacker, disabling the endpoint DLP will help in stealing data, but at the same time considerations have to be taken not to trigger any suspicion. Depending on the DLP product, uninstalling an endpoint agent might get reported to management server, but if the attacker is able to block such communication the server will think the endpoint is offline. [6]

## **9 DLP Challenges**

### **9.1 Encryption**

Different prevention mechanisms are needed to cover different states of data. In particular, detecting and preventing data leaks in transit have major challenges due to encryption and the high volume of electronic communications. While encryption provides means to ensure the confidentiality, authenticity, and integrity of the data, it also makes it difficult to identify the data leaks occurring over encrypted channels. [3]

### **9.2 Collaboration**

To be able to identify the ‘outsider’ in a communication, the collaborating parties should be identified. However, identifying the communities of collaboration is not a straightforward task. the temporal nature of the collaborations should be addressed. As time passes, new collaborations are formed, and existing ones disappear. Thus, identifying the communities of collaboration should not be regarded as a one-time task but as a continuous task to be carried out on regular intervals. [3]

#### **9.2.1 Access Control**

Access control provides the first line of defense in DLP. However, it does not have the proper level of granularity and may be outdated. While access control is suitable for data at rest, it is difficult to implement for data in transit and in use. [3]

### 9.2.2 Semantic Gap in DLP

When the communicating group defines a data leak as well as the data exchanged during the communication, a simple pattern matching or access control scheme cannot infer the nature of the communication. Therefore, data leak prevention mechanisms need to keep track of *who*, *what* and *where* to be able to defend against complex data leak scenarios. [3]

## 10- References

- [1] Data Loss Prevention: Considerations from an IT Audit Perspective, ISACA November Luncheon, *EARNST& YOUNG*
- [2] Security and Loss Prevention: An Introduction, 5th Edition, Philip P. Purpura
- [3] Understanding Data Leak Prevention, Preeti Raman, Hilmi Gunes Kayacik, Anil Somayaji
- [4] Methodology for Data Loss Prevention Technology Evaluation for Protecting Sensitive Information, López G.\*; Richardson N.\*\*; Carvajal J.\*
- [5] MOGULL Rich (2010). Understanding and Selecting a Data Loss Prevention Solution. [online]. Websense. Last accessed 01 October 2014  
at: [https://securosis.com/assets/library/reports/Understanding\\_and\\_Selecting\\_DLP.V2\\_Final\\_.pdf](https://securosis.com/assets/library/reports/Understanding_and_Selecting_DLP.V2_Final_.pdf)
- [6] Data Loss Prevention Systems and Their Weaknesses, Tore Torsteinbø
- [7] BUNKER, Guy and FRASER-KING, Gareth (2009). Data leaks for dummies. Chichester; Hoboken, N.J, Wiley Publishing, Inc.
- [8] Mogull, Rich. Best Practices for Endpoint Data Loss Prevention. s.l. : Securosis, L.L.C., 2009.
- [9] BS ISO/IEC 27001:2013: Information technology. security techniques. information security management systems. requirements. (2013)
- [10] J. Livingston, "Tips and Strategies to Protect Laptops and the Sensitive Data They Contain," *Information Systems Control Journal*, vol. 5, 2007.
- [11] The Best Data Loss Prevention (DLP) Software ( Free & Paid ), ITECHTICS Making Technology Accessible
- [12] Mogull, Rich. Best Practices for Endpoint Data Loss Prevention. s.l. : Securosis, L.L.C., 2009.
- [13] Autonomy Corp. KeyView IDOL & Connectors. autonomy.com. [Online] [Cited: March 28, 2012.]  
<http://www.autonomy.com/content/Products/idol-modulesconnectors/index.en.html>.
- [14] Symantec. What does kvoop.exe do for Symantec DLP? Symantec Connect. [Online] August 18, 2010. [Cited: March 28, 2012.] <http://www.symantec.com/connect/forums/whatdoes-kvoopexe-do-symantec-dlp>

- [15] M, Margaret. MTA Integration for SMTP Prevent. Symantec Connect. [Online] October 13, 2010. [Cited: May 22, 2012.] <http://www.symantec.com/connect/forums/mtaintegration-smtp-prevent>.
- [16] Mogull, Rich. Implementing DLP: Deploying Network DLP. Securosis. [Online] February 13, 2012. [Cited: May 9, 2012.] <https://securosis.com/blog/implementing-dlpdeploying-network-dlp>.
- [17] What is Data Loss Prevention? February 12, 2019 | By Comodo
- [18] A New Approach for Sensitive Data Leakage Prevention Based on Viewer-Side Monitoring By Muneer Yousef Fareed Hasan (Thesis)
- [19] <https://www.imperva.com/learn/data-security/data-loss-prevention-dlp/>
- [20] Open Security Foundation, "Data breach trends during the first half of 2014," Report, 2014.
- [21] CSO Online - <https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html>
- [22] Understanding and Selecting a Data Loss Prevention Solution – SECUROSION, Version 2.0 Released: October 21, 2010