YÖNETİM BİLİŞİM SİSTEMLERİ DERGİSİ
http://dergipark.gov.tr/ybs

Yayın Geliş Tarihi: 07.02.2021
Yayına Kabul Tarihi: 26.05.2021
Online Yayın Tarihi: 07.07.2021

Cilt:7, **Sayı**:1, **Yıl**:2021, **Sayfa**:79-95
e-ISSN: 2630-550X
Araştırma Makalesi/Research Article

DEPLOYING DATA LOSS PREVENTION (DLP) SYSTEMS IN BIG ENVIRONMENTS

Yenal ARSLAN[1]

[1] Social Security Institution, Ankara Turkey

ABSTRACT

Data Leakage or Loss Prevention (DLP) or information leak prevention (ILP) or information protection and control (IPC) technology has been developed to prevent data from intentionally or accidentally leaking out. Data loss prevention systems differ from conventional security controls such as firewalls or intrusion detection systems (IDS) in terms of dedication and proactivity. Conventional security controls have less dedication to the actual content of the data.

Although there are some academic DLP studies in the literature, very few studies on industrial solutions. This study established a DLP system for the Social Security Institution (Sosyal Güvenlik Kurumu, SGK) of Turkey. SGK is Turkey's one of the biggest institutions with 28,000 employees. Installation methods and experienced problems were noted objectively. And important things about the implementation of industrial DLP systems in large institutions have been marked.

Anahtar Kelimeler: Data loss prevention, DLP, Cyber security management

BÜYÜK ORTAMLARDA VERİ KAYBINI ÖNLEME SİSTEMLERİNİN UYGULANMASI

ÖZET

Veri sızıntısı, veri kaybı önleme (DLP), bilgi sızıntısı önleme (ILP), bilgi koruma ve kontrol (IPC) teknolojisi, verilerin kasıtlı veya kazara dışarı sızmasını önlemek için geliştirilmiştir. Veri kaybı önleme sistemleri, adanmışlık ve proaktiflik açısından güvenlik duvarları veya saldırı tespit sistemleri (IDS) gibi geleneksel güvenlik kontrollerinden farklıdır. Geleneksel güvenlik kontrolleri, verilerin mevcut içeriğine daha az odaklanırlar.

Literatürde bazı akademik DLP çalışmaları olmasına rağmen endüstriyel çözümler konusunda çok az çalışma bulunmaktadır. Bu çalışma, Türkiye'deki Sosyal Güvenlik Kurumu (SGK) için kurulan DLP sistemini ele almaktadır. SGK, 28.000 çalışanı ile Türkiye'nin en büyük kurumlarından biridir. Kurulum yöntemleri ve yaşanan sorunlar objektif olarak dikkate alınmış ve endüstriyel DLP sistemlerinin büyük kurumlarda uygulanmasıyla ilgili önemli konulara dikkat çekilmiştir.

Keywords: Veri Sızıntısı Önleme, DLP, Siber Güvenlik Yönetimi

INTRODUCTION

We cannot provide cyber security only with conventional firewalls. We must take a holistic view of cyber security management. Of course, to protect data, we should first start with the conventional protection tools available in institutions and then we must add some specific solutions. Cyber space is like a black hole. This black hole should be closed as much as possible with special solutions. Data leakage prevention (DLP) systems are one of these special solutions.

Data leak the unauthorized disclosure of sensitive information from a corporate network is one of the most significant fears that organizations face today (Gordon, 2007). Preventing data leakages is not always possible because of the need to access, share, and use information, which leads to the inevitable release of confidential data (Alneyadi et al., 2016). The output of sensitive data to other e-mail addresses via corporate e-mail, the output of files to external memory, disks by USB or by writing to CD / DVD ROMs, extracting data from the application by taking print screen, taking

pictures or with copy/paste keys, uploading files to other addresses via FTP, uploading files to any website are some of the known methods of data leakage. Organizational communication happens in a huge number of ways like platforms, applications, and devices. As corporations look for new methods to raise efficiency in productivity and collaboration, they also create new risks to employees that can leak or steal corporate data. Moreover, data breaches can affect customers' perception of a company's image by decreasing its reputation, especially if sensitive personal information is leaked (Huth et al., 2013). Data breaches have increased by nearly two-fold in the five years since 2012, increasing from 68 in 2012 to 130 in 2017 (Faiz et al., 2020). Security professionals believe that insider threat is the greatest risk for enterprises. According to a survey conducted by CA Technologies on the state of insider threat in 2018, 90% of surveyed organizations felt vulnerable to insider attacks and 53% of organizations indicated that they have been the target of inside attack during the year (Alhindi et al., 2019). The other tech giant cisco reveals that organizations experienced about one threat per month (Guevara et al., 2017).

According to recent Ponemon Institute research in 2019, insider threats have three main categories. Employee or contractor negligence, criminal and malicious insiders, and credential theft. They emphasize that %63 of the incidents were the results of negligence, %23 incidents were the results of criminal and malicious insiders, and %15 incidents were results of credential theft. The average cost for credential theft $871,686 criminal and malicious insiders $756,760 and negligence $307,111 (IBM/ObserveIT, 2020).

DLP (Data Loss/Leak Prevention) is a relatively new and increasingly used data protection type in the field of network security. With DLP software, the exit of unwanted data from your system can be prevented. DLP software is aimed at unauthorized use, monitoring, and protection of sensitive data. DLPs differ from conventional security controls such as firewalls, VPNs, and IDSs in terms of dedication and proactivity. Conventional security systems have less dedication to the actual content of the data (Alneyadi et al., 2016). DLP systems have three main properties. Firstly, they can analyze the content of confidential data and the surrounding context. Second, they can be deployed to protect confidential data in different states that are transit, use, and rest. The third attribute is the ability to protect data through various remedial actions such as notifying, auditing, blocking, encrypting, and quarantining.

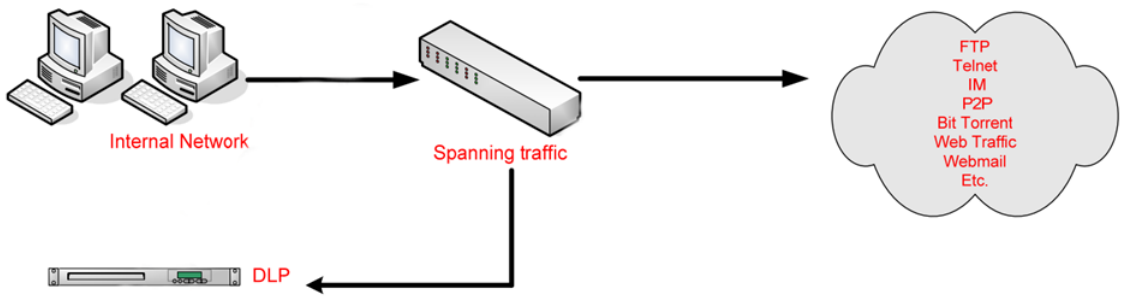
Security violations cannot be tracked in real-time in traditional inspection methods. In addition to this, limited audit resources make DLP systems compulsory in organizations.

Data Loss Prevention technologies can request both structured and unstructured data. For example, structured data;11 digits pattern such as 1234567899876 for a Turkish Social Security Number or regular expression like ^[1-9]{1}[0-9]{10}$ for Turkish Identity number. Unstructured data includes word or pdf files, spreadsheets, and any type of record for which its format is not an essential component of its meaning. First, DLP discovers information assets in areas of most importance or risk across an organization, Then DLP starts to enforce to prevent data with related policies in three important contexts: Motion, Rest, and End Point (Liu, & Kuhn, 2010).

Data in Motion

Information moves immediately by file servers, web applications, electronic mails, instant messages, and many other communication tools. Protection here includes implementing solutions at network gateways to monitor, encrypt, filter, and block sensitive data in outbound direction with no restricting the flow of non-sensitive communications. In figure 1 shows data in motion.

Figure 1. Data in Motion



Usage control generalizes access control to the future handling of data once access to it has been granted. It can be seen as a natural framework for the idea of Data-In-Use DLP, as it aims at regulating the usage of (confidential) data according to specified security policies and thus also allows to prohibit unwanted events of data leakage (Wuchner, & Pretschner, 2012). There are several approaches to protecting data in use. For example, Shu and Yao developed a technique called fuzzy fingerprint in their study and tried to prevent network-based data leakage (Shui & Yao, 2013).

Data at Rest

Data-at-rest is defined as all data in computer storage. To keep data-at-rest from being accessed, stolen, or altered by unauthorized people, security measures such as data encryption and access control are commonly used (Tahboub, & Saleh, 2014).

As its name, this feature is not responsible to hold anything. The Discovery feature has two options. The first responsibility is discovering sensitive data on data repositories.

Information can live forever in servers, databases, desktops, laptops USB drives, and other data repositories—and all their backup copies and archives. Prevention must start here to create an inventory of information by discovery. Then it should be followed by remediation and new controls to stop reoccurrence.

Data at End Point

Electronic information is "in use" when an end-user is working on it at a network endpoint: a laptop, desktop, or other computing platforms. Protection here means restricting use at the endpoint, for example by blocking and reporting attempts to copy information to a USB drive or print it while connected outside the corporate network.

METHODOLOGY

Considering that there is very little information about industrial DLP systems in the literature, the selection criteria of DLP systems, topology drawings related to the installation, installation stages, integration, and performance problems have been evaluated in detail through a real application. This study established a DLP system for the Social Security Institution (SGK) of Turkey. In the next flow of the article, the Social Security Institution will be introduced, and then DLP solution approaches and selection criteria will be mentioned, followed by topology and installation stages and performance improvement metrics. After explaining the problems and constraints experienced, the results and lessons learned will be discussed.

The Social Security Institution (SGK) is the governing authority of the Turkish social security system. Social Security Institution is established with the objective of the realization of a social security system at the contemporary standards that will provide individuals with social insurance and universal health insurance, based on the principles of social insurance, effective, equitable, easily accessible, actuarial, and sustainable in financial terms.

Purpose of Having Data Loss Prevention for Social Security Institution

Achievement goals with DLP under this project;

- Instant follow-up of information infiltration events
- Raising Institutional awareness
- Determination and protection of criticality ratings of information in use
- Identification and protection of criticality ratings of information stored in data fields like databases and file servers
- Detection, classification, and protection of information flowing on the network
- Periodic renewal of data classifications and security policies
- Competence of control and management tools

Choosing a Vendor

There are some features for data loss prevention systems to be installed in large environments such as central management, strong reporting, backup needs, and completing projects with insufficient personnel.

Monitoring and Prevention

Monitoring feature solutions ought to follow all information access instantly to identify unauthorized activities in light of point by point — For instance, an answer of questions; who, what, where, when, and how of every datum access. These arrangements must have the capacity to respond promptly to avert unapproved get to or suspicious action by favored insiders and a potential bad-meaning insider or malicious outsiders, also, to computerize information security administration controls in heterogeneous institutions. With the correct network design, Monitoring can enhance security to minimize the total cost of having a DLP solution.

For the prevention feature, if you want to monitor and block an e-mail that includes sensitive corporate information, many of the vendors require integration between their DLP solution and MTA (Mail Transfer Agent) like Ironport, Proofpoint, and Sendmail. If you want to prevent multiple channels, this increases the cost of data prevention. Apart from cost, this may cause problems like integration and troubleshooting. If you want to block almost all channels, you must consider these.

Centralized Management

Almost every corporation is suffering from missing manpower. If you have a new product, this means you need new stuff. Centralized management can reduce the responsibility of your staff. In order to reduce staff cost, you must create, update or disable policies, reports, or data filters from the same screen.

Backup and Storage Requirements

Almost every association needs data storage. Some of the DLP vendors are software-based – some of them have hardware appliances.

If your data retention policy expresses that data must be kept for a half year, hardware appliances need to deal with terabytes of data. This situation not only might cause some performance problems also will cost more.

Ease of Integration

Each association has its special needs. Therefore, generally, box solutions of vendors do not meet the specific requirements of customers.

If you think about taking prevention mode, ease of integration is a key component to consider apart from your hardware and software needs for DLP. At times, associations are aware of the troubles in deploying DLP in preventive mode simply after the critical measure of work has been finished. If this gets ignored, the general solutions can take extreme time.

Staffing Needs

As mentioned above, almost every corporation suffers from a lack of staff. If you choose a DLP vendor that does not have a centralized management screen, does not have a user-friendly admin console or mature workflows, this means you need more staff and your staff must have experience with DLP deeply.

Result of Vendor Choosing

Based on the above issues and the features in table 1, features were scored from 0 to 4 (4 for best 0 for worst). After the assessment, SGK decided to purchase the Symantec Data Loss Prevention solution.

Table 1. Vendor Comparison

	Symantec	Websense/Forcepoint	MacAfee	Microsoft
Product Capabilities				
Network Monitoring	4	3	3	0
Email and web prevention	4	4	4	2
Data discovery and protection	4	4	4	2
File access and usage monitoring	4	2	0	2
Endpoint monitoring and prevention	4	2	4	0
Cloud email monitoring and prevention	2	2	3	0
Mobile device monitoring and prevention	4	0	2	0
Scan target coverage	4	3	2	2
Self-service remediation	3	1	0	0
Performance and scalability	4	2	4	4
Extensibility (integrations & APIs)	2	0	3	0
Policy Enforcement				
Unified policy management	3	3	3	2
Content-aware detection	4	3	3	2
Incident response workflow	4	2	3	2

Role-based access control	4	1	4	0
Reporting and analytics	4	3	2	0
Management and Security				
User authentication and identity resolution	3	0	3	4
System management and security	4	4	3	2
Market Leadership				
Customer support and success	4	1	2	2
Deployment methodology	4	0	0	1
Research and development	4	1	2	1

Since our goal is not to determine the best product but to exhibit different DLP mechanisms and find the most appropriate solution by discussing their effectiveness, we did not enter into a comparison like the SC Magazine that Alneyadi et al. mentioned in their study (Alneyadi et al., 2016). Because these assessments reflect case studies that lack a comprehensive attack vector classification and a clearly defined threat model.

Also, in figure 2 Gartner's released 2016 Magic Quadrant for Data Loss Prevention presents shows similar results with our comparison table (Gartner, 2016).

Figure 2. Gartner's released 2016 Magic Quadrant for Data Loss Prevention

The infrastructure of the Social Security Institution (SGK)

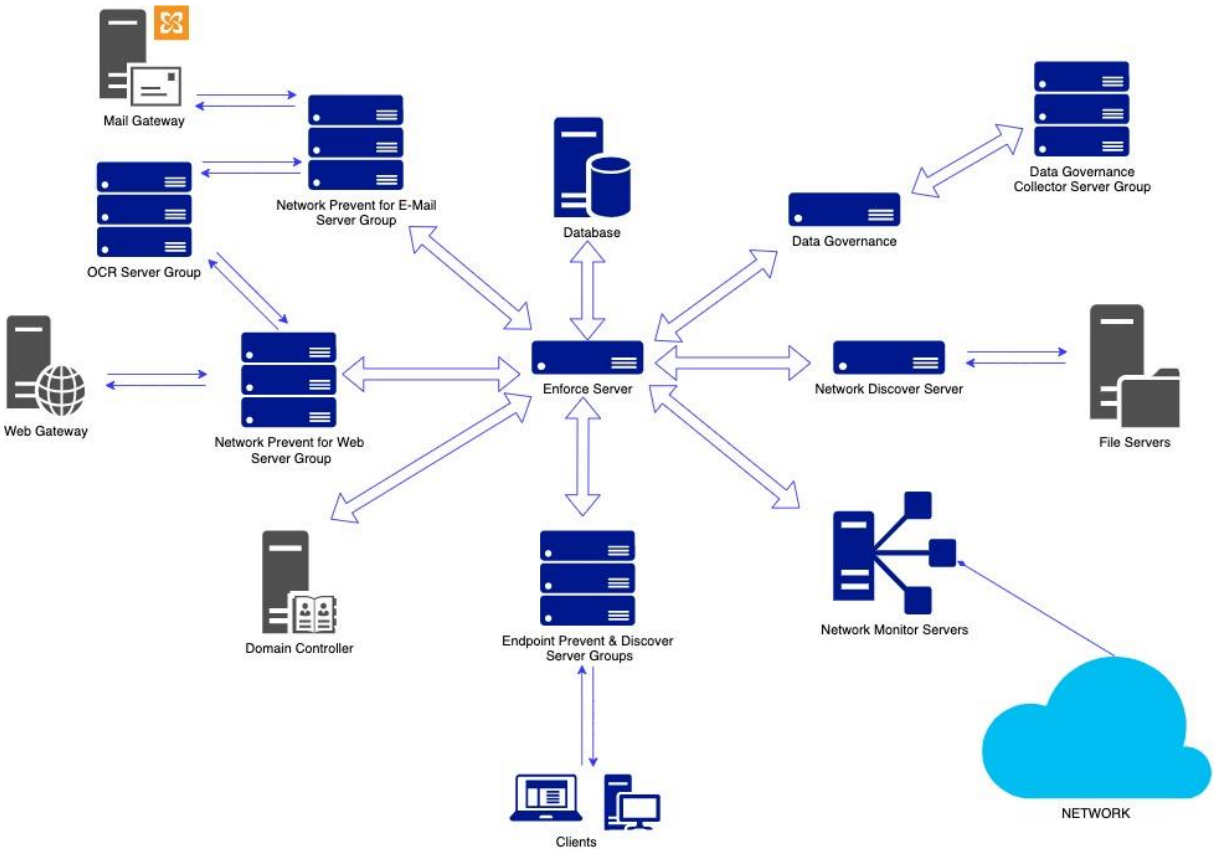
There are 1 headquarter, 2 data centers, and 584 provincials in 81 cities. SGK has 28000 clients in the Data Loss Prevention structure in total.

According to these numbers, SGK has the below servers;

- 1 Database
- 1 Enforce Management
- 4 Endpoint Management
- 1 Network Discovery
- 4 Mail Prevent
- 6 Web Prevent
- 1 Network Monitor

SGK DLP logical topology shows in figure 3.

Figure 3. Logical Network Topology – Data Loss Prevention Deployment



Database Platform: It is the central data point with a centralized database, case, policy, and report & workflow.

Enforce Platform: It is responsible for policies such as writing, reporting, user, and authority/role management. Enforce is a component that generates "fingerprints/indexes" that we can consider as signatures to be used in mathematical expressions of confidential information from documents/databases we have shown as confidential information.

Ocr Server: Ocr servers are central servers created to read the content of image files seen on clients.

Detection Servers:

Endpoint Discover: The component that will scan for steady/steady hidden information on end-user computers. It requires an agent in endpoint systems. This will scan the "fingerprint" and data structures we assume to be confidential information such as credit cards, top-secret and service-specific keywords and send incident information to the relevant Enforce platform. The main purpose is to "discover the hidden information that stood still" on the end user's machine.

Endpoint Prevent: On the end-user side, it also works as an agent structure and follows the use of confidential information, and performs actions defined by the specified policies (user popups, blocking, etc.). Not only the network operations of the user's machine but also all the activities that could result in the infiltration of information, not at the network level; For example; Copy information to USB, write to CD / DVD, print / receive faxes; It can prevent.

Network Monitor: A component that will detect the use of pre-determined confidential information by analyzing incoming / outgoing traffic at network/information output points (eg firewall, LAN interface). A component that will work on port mirroring/sniffing at the L2 level and

will detect all infiltration events in the traffic that was captured by packet capture. It performs passive monitoring on all TCP protocols.

Network Prevent for Web (Web Prevent): A component that will allow you to take actions such as blocking traffic and removing harmful content in case of infiltration of information in web traffic. This component integrates with the applications on the user web traffic and is a component that can prevent users from uploading confidential information to their websites by writing them to blog sites. As a business logic, it works with an ICAP-compliant proxy. Supports HTTP, HTTPS, FTP protocols.

Network Prevent for Mail (Mail Prevent): A component that can prevent the infiltration of information in mail traffic. It can block hacked mails, forward to the encryption gateway, marking, and tagging. Supports TLS protocol

Network Discover: A component that will track whether or not previously identified confidential information is found by scanning fixed stationary information on databases, file servers, websites, SharePoint portals. Thanks to this module, it is possible to determine that confidential information is out of the determined areas with policies and to prevent the infiltration of information.

Ex: Discovery of \\fileserver\common\Salaries.

Network Protect: A component that allows confidential information to be moved to the quarantine, deleted, or replaced with a warning message if the file exists outside of the end-user is detected in the server, database, website, etc. Works with network discover. (Network discover licenses are mandatory.) It provides preventive measures against infiltration of information by preventing confidential information from being located outside the places where it should be found.

Data Insight: A component that allows a complete analysis of who has access to the files containing this information (such as read-write-change) if confidential information is detected outside of a required location. It is complementary to the network to discover and protect modules. This generates invaluable information about file management.

Ex: Who read, reach \\fileserver\common¬\salaries.xls, Top 10 users who read this file, how often they read?

In terms of performance; Physical memory usage of nearly 6 GB, disk usage of nearly 20 GB for Enforce Server is in a healthy status. Physical memory usage nearly 5 GB, disk usage nearly 8 GB for Detection servers are in a healthy status.

Agents

There are around 28000 Agents Registered in the Current version.

There are four (4) Endpoint Prevent Servers in Load Balance to manage the agents.

There is one (1) Endpoint Prevent Server working as Endpoint Discover.

Deployment Process

Deployment steps

To get healthy results and do not disturb users with false positives, SGK decided to implement the DLP project in monitor mode first without an agent. In the beginning, the system got almost 4000 alerts in a week. After that policy fine-tuning which is mentioned below was implemented.

After implementing network, web, and mail, SGK had decided to distribute agents to clients. The main building was chosen as the pilot to respond quickly to the problems.

Agent Deployment for Headquarter

Endpoint visibility provides incredibly advanced information on employee activity, identifying anything that's out of the ordinary. It notices when patterns of activity change, and is even capable of accurately predicting when a member of staff is preparing to leave the business (Hooson, 2015).

To distribute agents, Microsoft System Center Configuration Manager (SCCM) was used, in a while, agents registered themselves to enforce the server. For almost two weeks all system was monitored and did not face any problem. However, it is useful to know that all headquarter client has the same Microsoft Windows image, same configuration, and almost uses the same applications.

Therefore, SGK decided to choose a different pilot provincial to deploy agents.

Conducting Pilot Provincials

After headquarter experience, tried to find provincials that have most various Microsoft Windows operating system images. 3 provincials were found and OSTIM provincial was chosen because of its close location and familiar supervisors of information system team addition to various Microsoft Windows images. There were 87 end-users in there and agent distribution takes almost half of a day. We had faced with some problems for Microsoft Windows 10 operating system, agents affected font collections and sizes of Microsoft Office applications. To solve this, a bug fix was implemented. Secondly, after OSTIM provincial, SGK had started to distribute agents to Uşak provincials because of familiar supervisor and image variety. Then distribute to big cities of Turkey, Eskişehir, Manisa, Bursa respectively. In this period, policy tunings to minimize false positives continued. After Bursa, SGK decided to distribute whole provincials in the country and use the plate code of cities to start with. First, we choose cities whose plates start between 50 – 81, then 35 – 49, and lastly 1 – 33. We did not distribute for Istanbul (34), because Istanbul has almost half of the network traffic and the employee number has equal to the sum of all cities.

After activating in all cities except Istanbul, some performance issues on the web were experienced, and two more webs prevent modules were added to the existing four webs prevent. This time, we noticed that in busy times, ICAP integration problems start to occur. To solve this issue, the maximum number of request limits for the web gateway was reconfigured.

Informing Employees about Data Loss Prevention Project

Alert on windows startup

The main purpose of DLP systems is deterrence. Deterring those who want to leak data is much less than the cost of data leakage. Just like in the defense industry. Because these investments are always much less costly than war.

To notify users about the data loss prevention project, we started to show pop-up alerts on desktops when Windows started up. We expected to see a decrease for alerts in enforcing servers but there were no remarkable changes.

Banners and posters

It is widely agreed that technology alone cannot prevent cyber incidents. We know that the most common behavioral approaches to help prevent data leakage are information security policies and awareness campaigns (van der Kleij et al., 2020). Then SGK decided to use banners and posters that explain the importance of data, how to use/prevent data, and our data prevention project. This action took some attraction from users and people started to ask about the project. This attraction makes us think about alert counts but their principal concern was monitoring. Then, SGK sent an e-mail that explains our main goal is protecting data not monitoring users, we reduce the anxiety of end-users but the expected alert count decrease did not happen.

Written information letter

After Banners and Posters, SGK publishes a circular letter that includes information about data loss prevention projects, sanctions, punishments what if data leakage occurs, and distribute to all of the employees, get their signatures.

DISCUSSIONS

ICAP Integration Problem

After ICAP integration, we saw that there is a problem with integration between DLP and web gateway. In busy times, users reported below the screen in figure 4.

Figure 4. ICAP Integration- Error

To solve this situation, we increased request limits in Network Prevent Module for Web Servers.

Maximum Number of Requests: from 64 to 86016.

Maximum Number of Responses: from 64 to 86016.

System Requirements for Agents

Agents of data loss prevention require 30 MB of memory as a minimum. Also, it generally requires 80MB of storage area in the first installation.

The DLP Agent software temporarily consumes additional memory while it detects content or communicates with the Endpoint Prevent server. After these tasks are complete, the memory usage returns to the previous minimum. Additional disk space is then required to temporarily store incident data on the endpoint computer until the DLP Agent sends that data to the Endpoint Prevent server. If the endpoint computer cannot connect to the Endpoint Prevent server for an extended period, the DLP Agent will continue to consume additional disk System requirements and recommendations 22 Endpoint computer requirements for the DLP Agent space as new incidents are created. The disk space is freed only after the agent software reconnects to the Endpoint Prevent server and transfers the stored incidents." (Symantec, 2017).

Policy Tuning

To distinguish allowed from malicious transactions, DLP systems maintain a model of either allowed (whitelisting) or malicious (blacklisting) behavior. This model can either be specified based on an expert's knowledge or learned from past transactions (Costante et al., 2016).

SSI board decided to take in table 4 below actions.

Table 2. Policy List

Policy Name	Definition	Fine Tuning	Actions
Common Spyware Upload Sites	It detects accesses to common spyware sites and creates an incident. Most used spyware sites are currently added	Can be added if other spyware sites are not listed but known	To be added when the IP address or spyware sites requested by Network Security team
Confidential Documents	If documents includes keywords like "internal use only", "proprietary", "confidential"or "Do not distribute", it creates an incident	Can be added if there is a standard used by SGK as a standard that a document is confidential and should not go out	Keywords like ÇOK GİZLİ, HİZMETE ÖZEL, KİŞİYE ÖZEL (According to the Ministry of Transport and Communication)
Encrypted Data	It detects the encrypted documents.	There is no need to make any changes to this policy.	NES messages were added to the whitelist.
HIPAA and HITECH (including PHI)	If information such as drug names, disease names, treatment names, or drug codes is included with a TC number, it will identify and generate an incident. Drug names, disease names, treatment names, and medication codes are included here.	Turkish drug names, disease names, treatment names, and drug codes can be added	Keywords like "Bursa" were removed because of creating False positives. If there are special Turkish drug names, they will be added
Illegal Drugs	It identifies illegal drug/drug names and creates an incident. Illegal drugs or controlled substances (which can be used in the production of narcotics) are added at the ready.	Turkish illegal drug names can be added	Special Turkish illegal drug names will be added if available.
Merger and Acquisition Agreements	It identifies the keywords related to purchasing and agreements and creates an incident.	There is no need to make any changes to this policy.	
Network Diagrams	It will detect and generate incident information if there is IP address information in Microsoft Visio	There is no need to make any changes to this policy.	

	drawings.		
Network Security	It identifies the keywords which are related to Network Security/Hacking or data which is similar to logs generated by Keyloggers	Turkish terms or program names used by hackers can be added.	To be added when there are security warnings that are specifically requested by the Network Security team
Password Files	If the file is a file named "passwd" or "shadow", or if the contents of the file are in the form of "passwd", "shadow", or "SAM", it creates an incident	There is no need to change this policy .	
Payment Card Industry Data Security	It detects credit card information and generates an incident.	There is no need to change this policy	
Resumes	It detects the keywords that are likely to pass through the resume and creates an incident.	Turkish words related to the resume can be added.	Turkish Sites & keywords were added
Sarbanes	Identify and generate financial data	There is no need to change this policy	
Source Code	It detects source codes of Java, Perl, C, etc. programming languages and creates an incident.	There is no need to change this policy	
Turkish Competition Authority	It detects "Price Association", "Price Stability" etc. Keyword and creates an incident.	There is no need to change this policy	
Turkish Identities	It detects a Turkish Identity number and creates an incident according to the number of IDs. For example; if it detects more than 100 IDs, the incident level is "High"	There is no need to change this policy	

DLP management should not be the responsibility of the IT department. Once they have implemented the technical DLP controls used to enforce DLP policy, the various data owners should be responsible for managing and keeping their data safe (Rogowski, 2013). After meeting with Risk Managers of departments, SSI board decided that Risk managers will manage policies and incidents own by own for their departments to create proper policies and investigate existing incidents according to their business processes.

Also, in this meeting, the SSI board decided to prohibit all data related to GDPR (General Data Protection Regulation) which is sent to all domains except gov.tr domains.

RESULTS

As expected, observed a remarkable decrease in the incident counts. In figure 5 and figure 6, DLP Network and Endpoint incident counts graph according to policy changes and informative actions based on months is below.

Figure 5. DLP Endpoint Incident-Time Graphic

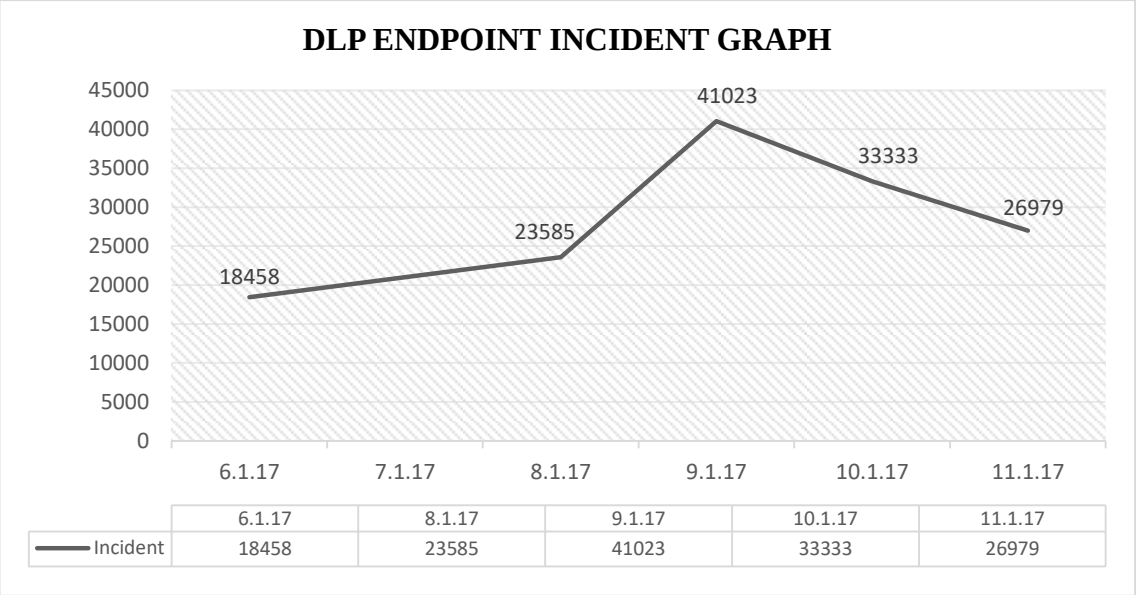
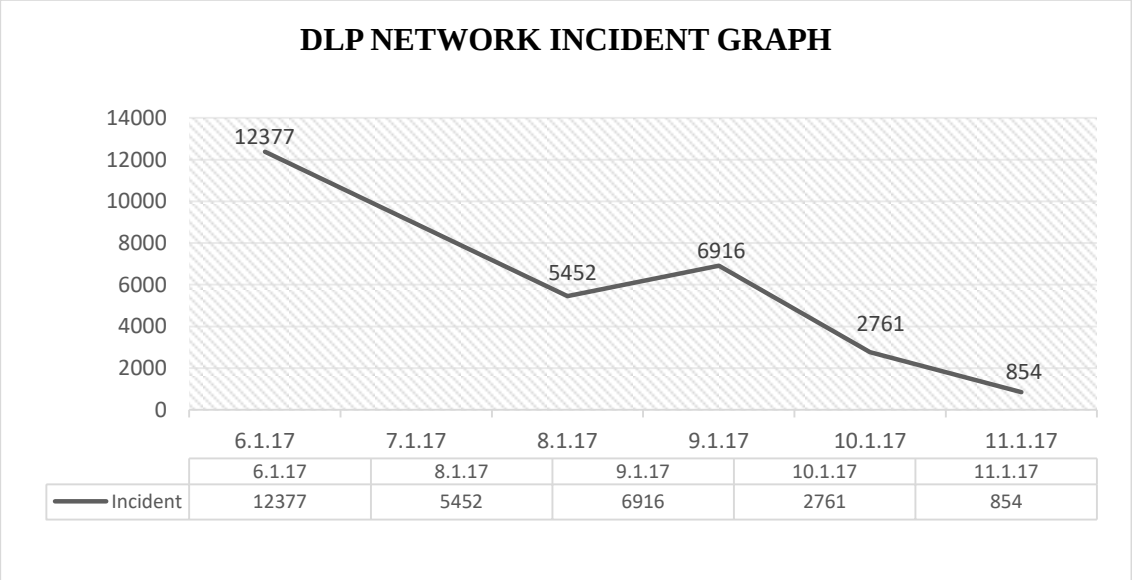


Figure 6. DLP network incident-time graphic



Attention should be paid to the number of new clients when examining incident counts in the graphics because implementation distributed step by step. The circular letter was published on September 20th,2017. A remarkable decrease can be observed. Policy Fine-Tuning for "Hipaa and Turkish Identity Numbers" was done on November 2nd,2017.

The studies listed below, which were incomplete in the study, were proposed to the institution

- Implement a DLP Network monitor to look for data in motion across all protocols.
- Implement DLP Network Discover to look for data at rest.
- Implement DLP Endpoint Discover to look for data at rest in the endpoints.
- Updating the risk scores for every risky action made by users and ensuring that the DLP rules are tightened flexibly for that user according to this score.

To reach a healthy and enduring solution, in the first phase product sends scheduled reports about software services, incidents, and system warnings/errors. In addition to automated scheduled reports, the IT team, especially employees who are responsible for DLP product analyzing incidents and writes reports after eliminating false positives and known issues. After the elimination process, the IT team and Head of IT have meetings to investigate the incident deeper. Every week, IT team members, the Head of IT, and the head of the executive board have a meeting about incidents that are reported by the IT team to investigate them deeper. After the installation, every risk manager and their team should decide to escalate the incident to the law department or notifying and informing the end-user about his neglect. Data loss prevention is one of the most important solutions to protect the sensitive data of organizations. However, Data loss prevention projects not only take time but also have some difficulties. Reactions of end-users, breaking managers, and inspectors' resistance against DLP agents were some of the most important problems faced in this project.

Establishing a team responsible for solving problems and investigating the incident is important. Otherwise, if you do not review incidents and do not control the management console of your product, the DLP project is meaningless. In the solution, OCR operation is not done in the client, for this reason there were slowdowns especially in printer output, this was noted as a matter for the manufacturer to work on. After the covid-19 pandemic, remote working has spread the use of mobile business equipment. DLP systems should focus on creating solutions, especially on mobile phones. Special dictionaries for instance data dictionaries can be used to identify specific data such as medical terms and geographical information. The dictionary-based approaches can help by accelerating and improving detection significantly.

It is important to have strong central management panels in DLP systems. Almost every organization today suffers from a lack of manpower in the IT field. If you have a new product, it

means you need new things. Central management can reduce the responsibility of your staff. You need to create, update or disable policies, reports, or data filters from the same screen to reduce staffing costs.

Backup and storage requirements are important in DLP projects as in many projects. If your data retention policy states that data must be retained for several years, DLP hardware or software devices must deal with terabytes of data. Not only does this cause some performance problems, but it also becomes costly. You must have an archiving and backup policy for your DLP system.

To get healthy results and not to disturb the user with false positives, you should first decide to implement the data loss prevention project in monitor mode without installing an agent for end-users. Once you have control over the network, web, and email (data in motion), you can decide to distribute agents to end users. It would be a good approach to start the distribution of agents from the pilot units or regions and proceed step by step. Endpoint visibility through agents defines everything extraordinary by providing incredibly advanced information on employee activity. You notice when activity patterns are changing, and you can even accurately predict when a staff member is preparing to leave. (If a staff member is copying all the information).

It is important to inform the employees about the Data Loss Prevention Project. A warning can be set when the employees' computers are turned on. When operating systems start, you can show pop-up alerts on the user desktops to inform about the data loss prevention project. You can use posters to explain your data prevention project.

You should send an email to users, explaining that your main goal is not to track them, but to protect data. At the beginning of projects such as DLP, the end user may be concerned. To overcome this situation, you must ease the anxiety of the end-user.

To more effectively reduce data leaks and comply with legislation, you should issue an internal order signed by the highest authority that includes the penalties to be imposed if an employee leaks data. You should then distribute this internal order to all employees and obtain their signatures.

The entire organization must commit to the DLP policy, it cannot be left to the IT department. Too often we see that the implementation and management of DLP (and information security in general) are being left to the IT department (Rogowski, 2013). After the installations and announcements are made, it will be useful to establish data leakage prevention commissions in each unit. Because employees or managers who work in a busy schedule may not be able to fulfill the requirements of the DLP project sufficiently. These commissions, which meet at certain periods, can decide whether to report the incidents to the legal department or to notify the end-user of their negligence.

As the last word, in this study, it is tried to give an idea about how to install industrial DLP solutions that are not handled much in the literature in large enterprises and their possible results. Problems encountered and solutions are expressed. Finally, the missing points in the project and the issues that the DLP vendor fell short of were presented to inspire future studies.

REFERENCES

Sultan Alneyadi, Elankayer Sithirasenan, Vallipuram Muthukkumarasamy, A survey on data leakage prevention systems, Journal of Network and Computer Applications, Volume 62, 2016, Pages 137-152, ISSN 1084-8045, https://doi.org/10.1016/j.jnca.2016.01.008.

Hanan Alhindi, Issa Traore, Isaac Woungang, Preventing Data Leak through Semantic Analysis, Internet of Things, 2019, 100073, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2019.100073.

E. Costante, D. Fauri, S. Etalle, J. den Hartog and N. Zannone, "A Hybrid Framework for Data Loss Prevention and Detection," 2016 IEEE Security and Privacy Workshops (SPW), San Jose, CA, USA, 2016, pp. 324-333, https://doi.org/10.1109/SPW.2016.24.

Faiz, Mohamed & Arshad, Junaid & Alazab, Mamoun & Shalaginov, Andrii. (2019). Predicting likelihood of legitimate data loss in email DLP. Future Generation Computer Systems 110. https://doi.org/10.1016/j.future.2019.11.004.

Mohamed Falah Faiz, Junaid Arshad, Mamoun Alazab, Andrii Shalaginov, Predicting likelihood of legitimate data loss in email DLP, Future Generation Computer Systems, Volume 110, 2020, Pages 744-757, ISSN 0167-739X, https://doi.org/10.1016/j.future.2019.11.004.

Gordon, P. (2007). Data Leakage - Threats and Mitigation. SANS Institute, Tech. Rep.

César Guevara, Matilde Santos, Victoria López, Data leakage detection algorithm based on task sequences and probabilities, Knowledge-Based Systems, Volume 120, 2017, Pages 236-246, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2017.01.009.

Stuart Hooson, Smarten your data security before new EU legislation or risk corporate loss, Network Security, Volume 2015, Issue 6, 2015, Pages 8-10, ISSN 1353-4858, https://doi.org/10.1016/S1353-4858(15)30048-9.

Huth, C.L., Chadwick, D.W., Claycomb, W.R. et al. Guest editorial: A brief overview of data leakage and insider threats. Inf Syst Front 15, 1–4 (2013). https://doi.org/10.1007/s10796-013-9419-8

IBM/ObserveIT (2020). Cost of insider threats 2020 report.

S. Liu and R. Kuhn, "Data Loss Prevention," in IT Professional, vol. 12, no. 2, pp. 10-13, March-April 2010, https://doi.org/10.1109/MITP.2010.52.

Magic Quadrant for Endpoint Protection Platforms, Gartner, Jan. 2016.

Rogowski, W. (2013). The right approach to data loss prevention. Computer Fraud and Security, 8 (2013), 5–7.

Walter Rogowski, The right approach to data loss prevention, Computer Fraud & Security, Volume 2013, Issue 8, 2013, Pages 5-7, ISSN 1361-3723, https://doi.org/10.1016/S1361-3723(13)70070-8.

Shu X., Yao D.. (2013) Data Leak Detection as a Service. In: Keromytis A.D., Di Pietro R. (eds) Security and Privacy in Communication Networks. SecureComm 2012. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 106. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36883-7_14

Symantec™ Data Loss Prevention System Requirements and Compatibility Guide – v14.6(2017).

R. Tahboub and Y. Saleh, "Data Leakage/Loss Prevention Systems (DLP)," 2014 World Congress on Computer Applications and Information Systems (WCCAIS), Hammamet, Tunisia, 2014, pp. 1-6, https://doi.org/10.1109/WCCAIS.2014.6916624.

Rick van der Kleij, Remco Wijn, Tineke Hof, An application and empirical test of the Capability Opportunity Motivation-Behaviour model to data leakage prevention in financial organizations, Computers & Security, Volume 97, 2020, 101970, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2020.101970.

T. Wüchner and A. Pretschner, "Data Loss Prevention Based on Data-Driven Usage Control," 2012 IEEE 23rd International Symposium on Software Reliability Engineering, Dallas, TX, USA, 2012, pp. 151-160, https://doi.org/10.1109/ISSRE.2012.10.