Data Loss Prevention

Technical Report · August 2019

CITATIONS
0

READS
4,332

1 author:

Kingston Mwila
University of Zambia
6 PUBLICATIONS **1** CITATION

SEE PROFILE

Data Loss Prevention

Kingstone Ali Mwila
P.G Student. Electrical/Electronics Department
School of Engineering, The University of Zambia
Lusaka, Zambia
KingstonMwila@gmail.com

Dr. Jackson Phiri
Head, Department Of Computer Science
School of Natural Sciences, The University of Zambia
Lusaka, Zambia
Jackson.Phiri@cs.unza.zm

Abstract
Sensitive and confidential data are a requisite for most companies, so protection for this data takes great attention by top management of a company, administrators and IT managers. Data leakage causes negative impact on companies. The traditional security approaches, such as firewalls, can't protect data from leakage. Data leakage/loss prevention (DLP) systems are solutions that protect sensitive data from being in non-trusted hands. This paper is an attempt to survey and study DLP systems that will be conducted as well as a comparison with other security and data protection approaches.

Keywords — data, information technology, professional

I. INTRODUCTION

IT professionals are tasked with the some of the most complex and daunting tasks in any organization. Some of the roles and responsibilities are paramount to the company's livelihood and profitability and maybe even the ultimate survival of the organization. Some of the most challenging issues facing IT professionals today are securing communications and complying with the vast number of data privacy regulations. Secure communications must protect the organization against spam, viruses, and worms; securing outbound traffic; guaranteeing the availability and continuity of the core business systems (such as corporate email, Internet connectivity, and phone systems), all while facing an increasing workload with the same workforce. In addition, many organizations face challenges in meeting compliance goals, contingency plans for disasters, detecting and/or preventing data misappropriation, and dealing with hacking, both internally and externally.

Almost every week, IT professionals can open the newspaper or browse online news sites and read stories that would keep most people up at night. The dollar amounts lost are staggering and growing each year (see sidebar, "Stored Secure Information Intrusions"). Pressures of compliance regulations, brand protection, and corporate intellectual property are all driving organizations to evaluate and/or adopt data loss protection (DLP) solutions [2].

II. DEFINITIONS

Data loss prevention (DLP) is a set of tools and processes used to ensure that sensitive data is not lost, misused, or accessed by unauthorized users. DLP software classifies regulated, confidential and business critical data and identifies violations of policies defined by organizations or within a predefined policy pack, typically driven by regulatory compliance such as HIPAA, PCI-DSS, or GDPR. Once those violations are identified, DLP enforces remediation with alerts, encryption, and other protective actions to prevent end users from accidentally or maliciously sharing data that could put the organization at risk. Data loss prevention software and tools monitor and control endpoint activities, filter data streams on corporate networks, and monitor data in the cloud to protect data at rest, in motion, and in use. DLP also provides reporting to meet compliance and auditing requirements and identify areas of weakness and anomalies for forensics and incident response [1].

[2] Data loss protection is a term that has percolated up from the alphabet soup of computer security concepts in the past few years. Known in the past as information leak detection and prevention (ILDP), used by IDC; information protection and control (IPC); information leak prevention (ILP), coined by Forrester; content monitoring and filtering (CMF), suggested by Gartner; or extrusion prevention system (EPS), the opposite of intrusion prevention system (IPS), the acronym DLP seems to have won out. No matter what acronym of the day is used, DLP is an automated system to identify anything that leaves the organization that could harm the organization.

DLP applications try to move away from the point or niche application and give a more holistic approach to coverage, remediation and reporting of data issues.

One way of evaluating an organization's level of risk is to look around in an unbiased fashion. The most benign communication technologies could be used against the organization and cause harm.

Before embarking on a DLP project, understanding some example types of harm and/or the corresponding regulations can help with the evaluation.

III. EXISTING TECHNOLOGIES USED IN DATA PREVENTION

Even before the Internet and all the wonderful benefits it brings to the world, organizations' data were exposed to the outside world. Modems, telex, and fax machines were some of the first enablers of electronic communications.

Electronic methods of communications, by default, increase the speed and ease of communication, but they also create inherent security risks. Once IT organizations noticed they were at risk, they immediately started focusing on creating impenetrable moats to surround the "IT castle." As communication protocols standardized and with the mainstream adoption of the Internet, Transmission Control Protocol/Internet Protocol (TCP/ IP) became the generally accepted default language of the Internet. This phenomenon brought to light external facing security technologies and consequently their quick adoption. Some common technologies that protect TCP/ IP networks from external threats are:

- Firewalls. Inspect network traffic passing through it, and denies or permits passage based on a set of rules.
- Intrusion detection systems (IDSs). Sensors log potential suspicious activity and allow for the remediation of the issue.
- Intrusion prevention systems (IPSs). React to suspicious activity by automatically performing a reset to the connection or by adjusting the firewall to block network traffic from the suspected malicious source.
- Antivirus protection. Attempts to identify, neutralize, or eliminate malicious software.
- Antispam technology. Attempts to let in "good" emails and keep out "bad" emails.

The common thread in these technologies: Keep the "bad guys" out while letting normal, efficient business processes occur. These technologies initially offered some very high-level, no granular features such as blocking a TCP/IP port, allowing communications to and from a certain range of IP addresses, identifying keywords (without context or much flexibility), signatures of viruses, and blocking spam that used common techniques used by spammers.

IV. THE WAVE OF DLP TECHNOLOGIES

The next wave of technologies that IT organizations started to address dealt with the "inside man" issue. Some examples of these types of technologies include:

- Web filtering. Can allow/deny content to a user, especially when it is used to restrict material delivered over the Web.
- Proxy servers. Services the requests of its clients by forwarding requests to other servers and may block entire functionality such as Internet messaging/chat, Web email, and peer-to-peer file sharing programs.
- Audit systems (both manual and automated). Technology that records every packet of data that enters/leave the organization's network. Can be thought of as a network "VCR." Automated appliances feature post-event investigative reports. Manual systems might just use open-source packet-capture technologies writing to a disk for a record of network events.
- Computer forensic systems. Is a branch of forensic science pertaining to legal evidence found in computers and digital storage media? Computer forensics adheres to standards of evidence admissible in a court of law. Computer forensics experts investigate data storage devices (such as hard drives,
- USB drives, CD-ROMs, floppy disks, tape drives, etc.), identifying, preserving, and then analyzing sources of documentary or other digital evidence.
- Data stores for email governance.
- IM- and chat-monitoring services. The adoption of IM across corporate networks outside the control of IT organizations creates risks and liabilities for companies who do not effectively manage and support IM use.
- Companies implement specialized IM archiving and security products and services to mitigate these risks and provide safe, secure, productive instant-messaging capabilities to their employees.
- Document management systems. A computer system (or set of computer programs) used to track and store electronic documents and/or images of paper documents.

Each of these technologies are necessary security measures implemented in [or "by"] IT organizations to address point or niche areas of vulnerabilities in corporate networks and computer assets.

Even before DLP became a concept, IT organizations have been practicing the tenets of DLP for years. Firewalls at the edge of corporate networks can block access to IP addresses, subnets, and Internet sites. One could say this is the first attempt to keep data where it should reside, within the organization. DLP should be looked at nothing more than the natural progression of the IT security life cycle.

V. DATA LOSS PREVENTION SYSTEMS

DLP is like the layers of an onion. Once the first layer of protection is implemented, the next layer should or could be addressed. There are many different forms of DLP applications, depending on the velocity and location of the sensitive data. Refer to Figure 1.

A. Data in Motion

Data in motion is an easy place to start implementing a DLP application because most can function in "passive" mode, meaning it looks at only a copy of the actual data egressing/ingressing the network. One way to look at data-in-motion monitoring is like a very intelligent VCR. Instead of recording every packet of information that passes in and out of an organization, DLP applications only capture, flag, and record the transmissions that fall within the categories/policies that are turned on. There are two main types of data-in-motion analysis:

Passive monitoring. Using a Switched Port Analyzer (SPAN) on a router, port mirror on a switch or a network tap(s) that feeds the outbound network traffic to the DLP application for analysis.

Active (inline) enforcement. Using an active egress port or through a proxy server, some DLP applications can stop the transmission from happening. The port itself can be reset or the proxy server can show a failure of transmission. The event that keyed off the reset or failure is still recorded.

B. Data at Rest

Static computer files on drives, removable media or even tape can grow to the millions in large multinational organizations. Unless tight controls are implemented, data can spawn out of control. Even though email transmissions account for more than 80% of DLP violations, data-at-rest files that are resting where they are not supposed to be can be a major concern.

Data-at-rest risk can occur in other places besides the personal computer's file system. One of the benefits of networked computer systems is the ability to share files. File shares can also pose a risk because the original owner of the file now has no idea what happened to the file after they share it.

The same can be said of many Web-based collaboration and document management platforms that are available in the market today. Collaboration tools can be used to host Web sites that can be used to access shared workspaces and documents, as well as specialized applications such as wikis, blogs, and many other forms of applications, from within a browser. Once again, the wonderful world of shared computing can also put an organization's data at risk.

DLP application can help with databases as well and half the battle is knowing where the organizations most sensitive data resides. The data-at-rest function of DLP applications can definitely help.

C. Data in Use

DLP applications can also help keep data where it is supposed to stay. Agent-based technologies that run resident on the guest operating system can track, monitor, block, report, quarantine or notify the usage of particular kinds of data files and/or the contents of the file itself. Policies can be centrally administered and "pushed" out of the organization's computer assets. Since the agent is resident on the computer, it can also create an inventory of every file on the hard drives, removable media and even music players. Since the agent knows of the file systems down to the operating system level, it can allow or disallow certain types of removable media [3].

For example, an organization might allow a USB storage device if and only if the device supports encryption. The agent will disallow any other types of USB devices such as music players, cameras, removable hard drives, and so on. Much like the different flavors of DLP that are available (data in motion, data at rest and data in use), conditions of the severity of action that DLP applications take on the event can vary. A good place to diagnose the problems organizations are currently facing would be to start with Monitoring. Monitoring is only capturing the actual event that took place to review at a later time. Most DLP applications offer real-time or near real-time monitoring of events that violated a policy.

Monitoring coupled with escalation can help most organizations immediately. Escalation works well with monitoring as when an event happens, rules can be put into place on who should be notified and/or how the notification should take place. Email is the most common form of escalation.

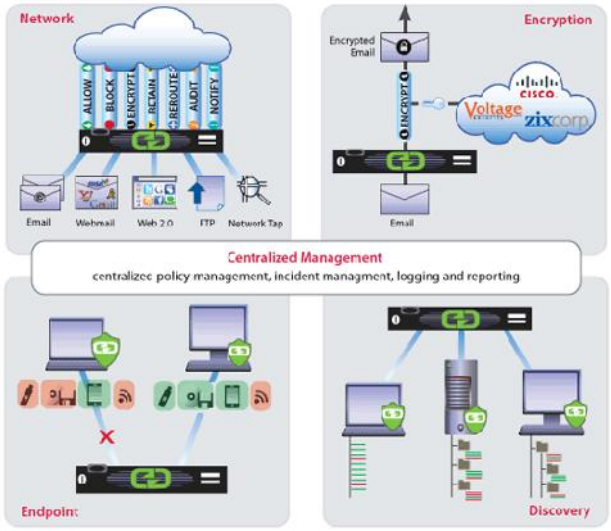

Figure 1.

VI. THE NEED FOR DATA LOSS PREVENTION

Data loss prevention solves three main objectives that are common pain points for many organizations: personal information protection or compliance, intellectual property (IP) protection, and data visibility [1].

Personal Information Protection / Compliance: Does your organization collect and store Personally Identifiable Information (PII), Protected Health Information (PHI), or payment card information (PCI)? If so, you are more than likely subject to compliance regulations, such as HIPAA (for PHI) and GDPR (for personal data of EU residents), that require you to protect your customers' sensitive data. DLP can identify, classify, and tag sensitive data and monitor activities and events surrounding that data. In addition, reporting capabilities provide the details needed for compliance audits.

IP Protection: Does your organization have important intellectual property and trade or state secrets that could put your organization's financial health and brand image at risk if lost or stolen? DLP solutions like Digital Guardian that use context-based classification can classify intellectual property in both structured and unstructured forms. With policies and controls in place, you can protect against unwanted exfiltration of this data.

Data Visibility: Is your organization seeking to gain additional visibility into data movement? A comprehensive enterprise DLP solution can help you see and track your data on endpoints, networks, and the cloud. This will

provide you with visibility into how individual users within your organization interact with data.
While these are the three main use cases, DLP can remediate a variety of other pain points including insider threats, Office 365 data security, user and entity behavior analysis, and advanced threats.

VII. DRENDS THAT ARE DRIVING DLP

The DLP market is not new, but it has evolved to include managed services, cloud functionality, and advanced threat protection amongst other things. All of this, coupled with the upward trend in giant data breaches, has seen a massive uptick in DLP adoption as a means to protect sensitive data. Here are nine trends that are driving the wider adoption of DLP:

The Growth of the CISO Role: More companies have hired and are hiring Chief Information Security Officers (CISOs), who often report to the CEO. CEOs want to know the game plan for preventing data leaks. DLP provides clear business value in this regard and gives CISOs the necessary reporting capabilities to provide regular updates to the CEO.

Evolving Compliance Mandates: Global data protection regulations constantly change and your organization needs to be adaptable and prepared. Within the past couple years, lawmakers in the EU and New York State, respectively, have passed the GDPR of which have tightened data protection requirements. DLP solutions allow organizations the flexibility to evolve with changing global regulations.

There are More Places to Protect Your Data: Increased use of the cloud, complicated supply chain networks, and other services you no longer have full control over has made protecting your data more complex. Visibility into the events and context of events that surround your data before it leaves your organization is important in preventing your sensitive data from getting into the wrong hands.

Data Breaches are Frequent and Large: Adversaries from nation states, cyber criminals and malicious insiders are targeting your sensitive data for a variety motives, such as corporate espionage, personal financial gain, and political advantage. DLP can protect against all kinds of adversaries, malicious or not. Within just the past couple of years, there have been thousands of data breaches and many more security incidents.

Your Organization's Stolen Data is Worth More: Stolen data is often sold on the Dark Web, where individuals and groups can purchase and use it for their own benefit. With certain data types selling for up to a few thousand dollars, there is a clear financial incentive for data theft.

There's More Data to Steal: The definition of what is sensitive data has expanded over the years. Sensitive data now includes intangible assets, such as pricing models and business methodologies. This means your organization has more data to protect.

There's a Security Talent Shortage: The security talent shortage is not going away anytime soon and you've probably already felt its impact on your own organization. In fact, in an ESG and ISSA survey from 2017, 43% of respondents said their organizations had been impacted by the shortage [1]. Managed DLP services act as remote extensions of your team to fill that personnel gap.

VIII. DATA LOSS PREVENTION BEST PRACTICES

It is important to determine primary data protection objective. The organization may be trying to protect its intellectual property, gain more visibility into the data, or meet regulatory compliance? With a main objective in place, it's easier to determine the most appropriate DLP deployment architecture or combination of architectures. The four main DLP deployment architectures are: Endpoint DLP, Network DLP, Discovery, and Cloud.

DLP is not a security-only decision. It can be implemented according the business needs and size of the budget. Leverage the pain points of different business units to show how DLP can address them. For example, the CFO's pain points include efficient use of assets and profitable growth. Managed DLP services address these pain points by eliminating the need for additional staff and CapEx to deploy and maintain a DLP program.

- When researching DLP vendors, establish your evaluation criteria:
- What types of deployment architectures are offered?
- Do they support Windows, Linux, and OS X with feature parity?
- What deployment options do they offer? Do they provide managed services?
- Do you need to defend against mainly internal or external threats? Or both?
- Do you need to perform content- or context-based inspection and classification? Will your users be able to self-classify documents? Do you need a blend of multiple methods?
- Are you most concerned with protecting structured or unstructured data?
- Do you plan to see and enforce data movement based on policies, events, or users?
- What compliance regulations are you bound by?
- What new regulations are on the horizon?
- Who are their technology alliance partners and what technologies would you like to integrate with your DLP?
- How quickly do you need to deploy your DLP program?
- Will you need additional staff to manage your DLP program?

Pragmatic Data Security Cycle

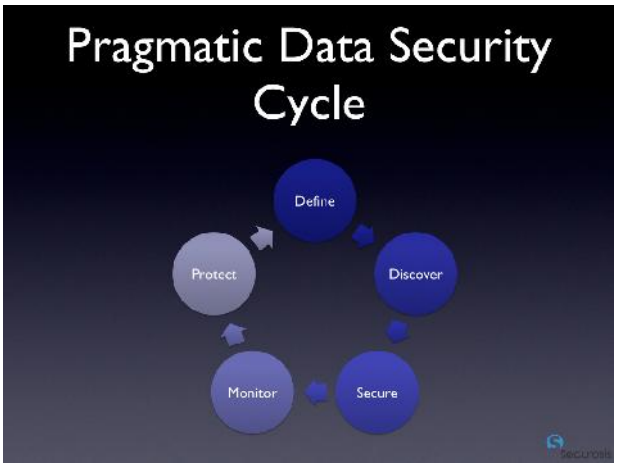

Figure 2.

Clearly define the roles and responsibilities of the individuals involved in your organization's DLP program. Building out role-based rights and duties will provide checks and balances.

Start with a clearly defined quick win. Organizations often try complicated initial rollout plans or try to solve too many use cases at once. Define your initial approach and set objectives that are fast and measurable. You should either take the project approach, where you narrow in and focus on a specific data type, or the data visibility approach, where your primary focus is discovery and automated classification of sensitive data to control egress.

Work together with business unit heads to define the DLP policies that will govern your organization's data. This will help ensure that the different business units are aware of the policies in place and how they might be impacted. Keep in mind that there's no one right way to develop DLP policies. Often, DLP strategy will align with your corporate culture.

Document your processes carefully. This will help you with consistent application of policies, give you a document of record for when reviews are needed, and will also be helpful when onboarding new team members or employees.

Define success metrics and share reporting with business leaders. Determine the key performance indicators (KPIs) you should measure and monitor them closely to determine the success of your DLP program and areas of improvement. Share these metrics with leaders of your organization to show the positive impact of DLP and its business value.

DLP is a program, not a product. Installing a DLP tool is just the first step in Data Loss Prevention. While you can get quick wins, understanding that DLP is a program to be continuously worked on will help you achieve lasting success. DLP is a constant process of understanding your data and how users, systems, and events interact with that data to better protect it.

It requires that agencies implement information security programs that, among other things, include:

- Periodic assessments of the risk
- Risk-based policies and procedures
- Subordinate plans for providing adequate information security for networks, facilities, and systems or groups of information systems, as appropriate
- Security awareness training for agency personnel, including contractors and other users of information systems that support the operations and assets of the agency
- Periodic testing and evaluation of the effectiveness
- of information security policies, procedures, and
- practices, performed with a frequency depending on risk, but no less than annually
- A process for planning, implementing, evaluating, and documenting remedial action to address any deficiencies
- Procedures for detecting, reporting, and responding
- to security incidents
- Plans and procedures to ensure continuity of Operations

An internal risk assessment of what types of "communication," both manual and electronic, that are allowed within the organization can give the DLP evaluator a baseline of the type of transmission that are probably taking place.

Some types of communications that should be evaluated are not always obvious but could be just as damaging as electronic methods. The following list encompasses some of those obvious and not so obvious methods:

- Pencil and paper
- Photocopier
- Fax
- Voicemail
- Digital camera
- Jump drive
- MP3/iPod
- DVD/CD-ROM/3½ in. floppy
- Magnetic tape
- SATA drives
- IM/chat
- FTP/FTPS
- SMTP/POP3/IMAP
- HTTP post/response
- HTTPS
- Telnet
- SCP
- P2P
- Rogue ports
- GoToMyPC
- Web conferencing systems

A. Corporate Culture

Once the corporate culture has established that DLP is worth investigating or worth implementing, the next logical step would be performing a risk/exposure assessment.

Several DLP vendors offer free pilots or proof of concepts and should be leveraged to jumpstart the data risk assessment for a very low monetary cost.

A risk/exposure assessment usually involves placing a server on the edge of the corporate network and sampling/ recording the network traffic that is egressing the organization. In addition, the assessment might involve look for high-risk files at rest and the activity of what is happening on the workstation environment. Most if not all DLP applications have predefined risk categories that cover a wide range of risk profiles. Some examples are [5]:

- Regulations: HIPAA, PCI-DSS
- Acceptable use: Violence, gangs, profanity, adult
- themes, weapons, harassment, racism, pornography
- Productivity: Streaming media, resignation, shopping, Webmail
- Insider hacker activity: Root activity, nmap, stack, smashing code, key loggers

Deciding what risk categories are most important to your organization can streamline the DLP evaluation. If data categories are turned on but not likely to impact what is truly important to the organization, the test/pilot result will contain a lot of "noise." Focus on the "low-hanging fruit." For example, if the organization's life blood is customer data, focus on the categories that address those types of leaks.

B. Precision versus Recall

False positive. A false positive occurs when the DLP application-monitoring or DLP application-blocking techniques wrongly classify a legitimate transmission or event as "uninteresting" and, as a result, the event must be remediated anyway. Remediating an event is a time-consuming process which could involve one to many administrators dispositioning the event. A high number of false positives is normal during an initial implementation, but the number should fall after the DLP application is tuned.

False negative. A false negative occurs when a transmission is not detected as interesting. The problem with false negatives is usually the DLP administrator does not know these transmissions are happening in the first place. An analogy would be a bank employee who embezzles thousands of dollars and the bank does not notice the theft until it is too late.

True positive. Condition present and the DLP application records the event for remediation.

True negative. Condition not present and the DLP application does not record it.

DLP application testing and tuning can involve a trade-off:

- The acceptable level of false positives (in which a no match is declared to be a match).
- The acceptable level of false negatives (in which an actual match is not detected).
- An evaluator can think of this process as a slider bar concept, with false negatives on the left side and false positives on the right. A properly tuned DLP application minimizes false positives and diminishes the chances of false negatives.
- This iterative process of tuning is called thresholding. Creating the proper threshold eventually leads to the minimization of acceptable amounts of false positives with no or minimal false negatives.
- An easy way to achieve thresholding is to make the test more restrictive or more sensitive. The more restrictive the test is, the higher the risk of rejecting true positives; and, the less sensitive the test is, the higher the risk of accepting false positives.

IX. DLP FUTURE DECISION

At the end of the day the DLP market and applications are maturing at an incredible pace. Vendors are releasing new features and functions almost every calendar quarter.

In the past, when monitoring seems sufficient to diagnose the central issue of data security, the marketplace was demanding more control, more granularity, easier user interfaces, and more actionable reports, as well as moving the DLP application off the main network egress point and parlaying the same functionality to the desktop/laptops, servers, and their respective end points to document storage repositories and databases.

In evaluating DLP applications, it is important to focus on the type of underlying engine that analyzes the data and then work up from that base. Next rate the ease of configuring the data categories and the ability to preload certain documents and document types. Look for a mature product with plenty of industry-specific references. Company stability and financial health should also come into play.

Roadmaps of future offerings can give an idea of the features and functions coming in the next release. The relationship with the vendor is an important requirement to make sure that the purchase and subsequent implementation goes smoothly. The vendor should offer training that empowers the IT organization to be somewhat self-sustaining instead of having to go back to the vendor every time a configuration needs to be implemented. The vendor should offer best practices that other customers have used to help with quick adoption of policies. This allows for an effective system that will improve and lower the overall risk profile of the organization.

Analyst briefings about the DLP space can be found on the Internet for free and can provide an unbiased view from a third party of things that should be evaluated during the selection process [2].

X. CONCLUSION

DLP is an important tool that should at least be evaluated by organizations that are looking to protect their employees, customers, and stakeholders. An effectively implemented DLP application can augment current security safeguards. A well thought out strategy for a DLP application and implementation should be designed first before a purchase. All parts of the organization are likely to be impacted by DLP, and IT should not be the only organization to evaluate and create policies. A holistic approach will help foster a successful implementation that is supported by the DLP

vendor and other departments, and ultimately the employees should improve the data risk profile of an organization. The main goal is to keep the brand name and reputation of the organization safe and to continue to operate with minimal data security interruptions.

Many types of DLP approaches are available in the market today; picking the right vendor and product with the right features and functions can foster best practices, augment already implemented employee training and policies, and ultimately safeguard the most critical data assets of the organization.

REFERENCES

[1] What is Data Loss Prevention (DLP)? | Digital Guardian (Online) digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention [Accessed: 2018-10-13]

[2] Tahboub R Saleh Y. Data Leakage/Loss Prevention Systems (DLP). 2014 World Congress on Computer Applications and Information Systems (WCCAIS) Publisher: IEEE 2014 pp: 1-6

[3] Data Loss Prevention Technologies. Tomoyoshi Hiroshi T Takayuki T Ryusuke Masuoka H. FUJITSU Sci. Tech. J. 2010 vol: 46 (1) pp: 47-55

[4] Northcut T. Data loss prevention. (Online) www.ey.com/Publication/vwLUAssets/EY_Data_Loss_Prevention/$FILE/EY_Data_Loss_Prevention.pdf [Accessed: 2018-10-14]

[5] https://iapp.org/tag/data-loss/ [Accessed: 2018-10-14]