DLP Systems as a Modern Information Security Control

Victor Morozov[1,2] and Natalia Miloslavskaya[1]

[1]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)

[2]OOO SearchInform

Abstract The importance of using modern protection tools against internal information security (IS) threats is proved. The advantages of DLP systems over alternative solutions are disclosed. The principles and technologies underlying the operation of DLP systems are discussed. The architecture, application features and analytical capabilities of the SearchInform Information Security Perimeter (SearchInform) DLP system are described in detail.

Introduction

Today, information is one of the most critical and valuable assets for success and prosperity of any company. The complexity of modern organizations and the trend to move to the cloud and outsource are increasing. At the same time the wide range of new ever-growing information security (IS) threats, especially those related to new information, communication and network technologies, services and devices, are all around us. The well-publicized attack on the Sony Playstation Network, resulted in the loss of user names, passwords, addresses, birth dates and financial details of 77 million users and Sony's financial loss around $171 million (including estimates for customer support costs, legal costs and the impact on future profits), left the online gaming network suspended for weeks in 2011. 2016 Ponemon Cost of Data Breach Study [https://www-03.ibm.com/ security/data-breach/] found that the average consolidated total cost of a data breach grew from $3.8 million in 2015 to $4 million in 2016. And this monetary value is secondary to the impression of a leaky and untrustworthy business.

The main sources of internal IS threats are malicious insiders, privileged users (employees using their high access levels to steal sensitive data), negligent insiders (such employees may store their password on a piece of paper stuck to the computer screen, plug in a USB flash drive found on a parking lot, send sensitive data to a wrong email, etc.) and exploited insiders (they can be lured into providing classified information or even making payments to attackers' accounts as a result of blackmailing, social engineering and other pressure). And among the major data leak channels are e-mail, social networks (Facebook, Twitter, etc.), Internet message boards, web blogs, instant messengers (ICQ, MSN, Jabber, etc.), remov-

able media, mobile devices, printers, FTP servers and Skype. If you do not control the above-mentioned channels or control only one or two data transmission links, your company's sensitive information may be easily transmitted to rival companies. State-of-the-art IS systems should allow all data communication channels, while intercepting and analyzing data flows transmitted over these channels. Comprehensive approach to IS ensurance is impossible even if only one potential data leak channel is not controlled.

About the terminology

DLP, which stands for Data Loss Prevention, is a strategy for making sure that end users do not send sensitive or critical information outside an organization's intranet [1]. From another point of view, DLP is primarily a set of technologies that prevent a leakage of confidential information. The term first hit the market in 2006 and gained some popularity in early part of 2007 [2].

A DLP system refers to a software product that helps a network administrator to control what data end users can transfer and in particular to mitigate the risks pertaining to data leaks, to manage employee loyalty and positive attitude, and to prevent fraud, kickbacks, sabotage and other harmful activities. From a technical point of view, a DLP system is a complex of software and hardware that guarantees information protection against threats of illegitimate data transmission from a protected segment of an automated system by analyzing and blocking outgoing traffic. There were certain difficulties with DLP systems classification and name in the early years of introduction of such systems in the market. Thus, for the designation of such systems, in addition to the previously mentioned Data Loss Prevention, Data Leak Prevention or Data Leakage Protection, the following names were used: Information Leakage Protection (ILP) – the term was suggested by Forrester Research in 2006, Information Leak Detection & Prevention (ILDP) – by the company International Data Corporation in 2007, Content Monitoring and Filtering (CMF) by Gartner, Anti-Leakage Software (ALS) by Ernst & Young. There were other options, for example, Extrusion Prevention System (EPS) – the term apparently created under the influence of Intrusion Prevention System.

DLP System's Functionality

The main task of DLP systems is to provide protection against accidental or intentional dissemination of confidential information by employees who have access to information due to their job responsibilities [2]. The application of DLP systems ranges from ordinary leakage through common information transmission (e-mail, Instant Messenger, etc.) to IS incident analysis and staff activity's control. These tasks can include the following: control of the use of working time and company's resources by its employees; preventive check of employees' loyalty; carrying out official investigations of forgery and theft; monitoring the compliance

of employees' actions with the requirements of the applicable IS policies, etc. For that purpose the DLP systems are shipped with many predefined IS policies: universal (which are relevant for every organization for detecting kickbacks and bribery, negative attitudes among the staff, risk groups engaged in alcohol and drug abuse, large debt and others) and industry-specific (like agriculture, gas supply, etc.).

The DLP systems identify, monitor and protect sensitive data in use (e.g. end-point actions), data in motion (e.g. network actions), and data at rest (e.g. data storage) through deep content inspection, contextual analysis of transaction (attributes of originator, data object, medium, timing, recipient/destination, etc.) and a centralized management framework. These systems use various rules to classify and protect information in a way that unauthorized end users cannot accidentally or maliciously share data, whose disclosure could put the organization at risk. To tackle these problems DLP systems fulfill the following actions:

- Defines and enforces consistent organization's IS policies, as well as gets alerts about their breaches;
- Places all information flows under control, filters data streams and protects data in motion, controls virtually all information channels, captures and analyzes the contents of all communications and transmitted files and end-point activities and detects sensitive content in documents of virtually all types;
- Allows to monitor user screens and employee conversations retrospectively or in real time;
- Performs monitoring in stealth mode both inside and outside the office;
- Includes repository of captured data and investigates and prevents the breaches with the help of retrospective analysis capabilities.

The DLP system's operation is based on interception and subsequent analysis of data flows that cross the perimeter to outside or circulate within the protected corporate network. The captured information is analyzed using various search algorithms. And if the data matching the selected criteria are found, the active component of the system is triggered, alerting the incident to the IS officer. In some cases, the transmission of a message (packet, flow, session) can be blocked [3].

The data analysis methods used in the DLP systems are attribute method (for example, using the properties of system objects) and semantic method (based on the semantic analysis of information, for example, through identification of key data combinations). Thus, DLP systems should provide the following searches:

- Keyword search: to find queried words, their forms and synonyms scattered throughout documents and other data;
- Phrase search: to search data for a phrase, for example first and last name or other set expressions;
- Dictionary search: to find documents that contain particular lexicon and slang pertaining to a specific topic, such as drug and alcohol abuse, gambling, etc.;
- Similar-context search: to track modified documents. Entire text or a text extract can be used as a query. The search hits include documents that are similar to the original document not only formally but also meaning-wise;

- Search by attributes: to search for documents by type, recipient, sender and other attributes. You can track activity of domain users, IP addresses, specific email addresses, documents and more;
- Regular expression search: to find data that conform to specific alphanumeric patterns (e.g., first name and last name, passport series and number, etc.);
- Search by fingerprints: to quickly detect files containing sensitive data;
- Complex queries: to combine simple queries into a complex query by using the logical operators AND, OR and NOT.

Most of the corporate DLP systems used are complex, i.e. they unite a *network subsystem* (Network DLP), designed to intercept data sent by users on different network protocols; *a subsystem to protect endpoints* (Endpoint DLP), designed to intercept data at the level of workstations (including data copied to external media and sent to print); *a subsystem for the protection of static data* (Content Discovery), designed to detect and protect sensitive data stored in various databases or file storages. Also, the mentioned systems generally include sophisticated means for analyzing captured data, implemented through a variety of search algorithms. On the basis of these algorithms, it is possible to create criteria for detecting IS incidents and IS policies. Figure 1 shows an example of a deployment scheme for DLP solution modules in an organization's infrastructure.

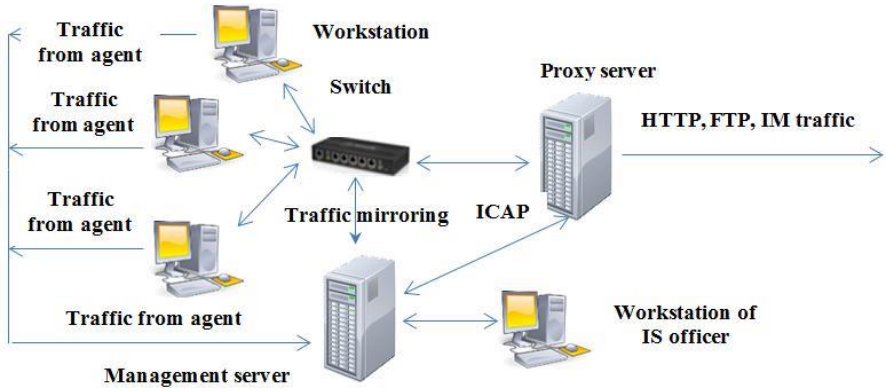

Fig. 1. Example of a scheme for deploying DLP modules in a company's network.

One Example of DLP System

The Russian SearchInform Company is now among the world's best DLP developers as the globally acknowledged company Gartner included this solution in January 2017 Magic Quadrant for Enterprise DLP [4], considering its level of technology implementation and strategic vision of a company. Experts of Gartner noted a modular approach to DLP and strong analytical capabilities, including speech-to-text transcription capabilities and strong image analysis capabilities. Other technological advantages of the SearchInform DLP [http://searchinform.com/products/SearchInform] are the following:

- Full integration with the Windows domain structure: it unambiguously identifies the user and computer sending data over email, ICQ or Skype, regardless of the email account, ICQ or account id used;
- Unlimited investigation capabilities: it archives all captured data, which allows to investigate an incident according to brand-new IS policies;
- Visibility into internal and external connections: it analyzes internal and external contacts of employees. A detailed graph of interactions enables to perform internal investigations swiftly and efficiently;
- Control of employee conversations: it performs voice recording by means of any detected microphone (headset, laptop, webcam, etc.) on or off the intranet. The captured recordings can be searched by attributes. The *LiveSound* mode allows listening to conversations in real time;
- Computer screen activity under control: it reveals who views which information on the workstation during work time. It captures screenshots or video of screen activity and also allows to monitor screen activity in real;
- Protection from malicious sysadmin activity: it analyzes Active Directory log events to detect suspicious actions performed by the company system administrator: account creation/deletion, privilege escalation/de-escalation, etc.;
- Control of data-at-rest on computers and shared network resources: IS officers can track occurrences of sensitive data in locations where such data is not supposed to be;
- Hardware and software reports: The system keeps track of software and hardware on corporate workstations and reports about any changes made. The reports facilitate hardware and software inventory tracking, and thus improve IT department performance and protect companies from losses.

The SearchInform DLP performs real-time inspection of all information flows and alerts about suspicious events. It stores all captured data and lets to reconstruct details should an investigation be required. It operates at two levels: controls network users' activity (up to 1000-1500 employees) and the data outgoing to the Internet as it mirrors traffic at the intranet level and controls endpoint users' activity and monitors the events on workstations, including the laptops used outside of intranet, when employees go on business trips or work from home (using logging of actions by means of software agents installed on their computers).

As soon as the system detects a suspicious activity or an IS policy violation, it sends a notification to the designated IS officer who then initiates an investigation. Traffic is captured on the level of network protocols (Mail, HTTP, IM, FTP and Cloud). Information can be filtered by domain names, computer names, IP and MAC addresses. All intercepted messages are stored in the SQL database, which is indexed by Search Server. Indexes allow quick search in the database.

The SearchInform DLP monitors email communications, voice and text messages as well as files (sent via Skype, Viber, ICQ, etc.), data transferred to/from cloud storage services, posts and comments on web forums and blogs, external devices (flash drives, hard drives, CDs and others), documents sent to printers (Fig. 2), etc. As a rule, for the normal SearchInform DLP functioning only two or three physical or virtual servers are needed. In cases where high performance is required, each server component can be installed on a separate server.

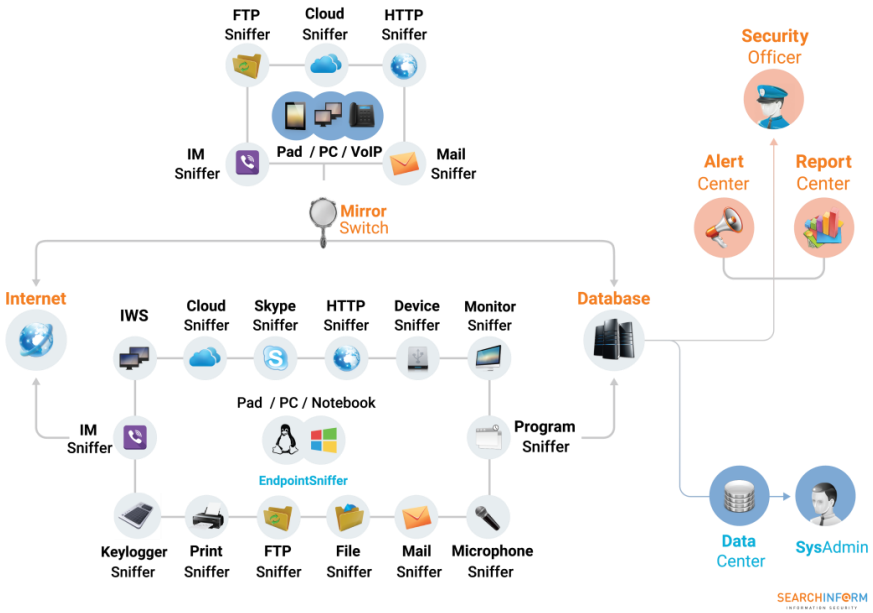

Fig. 2. SearchInform main modules

Conclusion

The presented information allows us to conclude that the analytical capabilities of modern DLP systems in whole are sufficient to prevent leakages of confidential information, control document circulation, detect violations in the distribution of data access rights, monitor staff activity, and address other IS challenges. At the same time, we should not forget that technology solutions are only one part of preventing data loss. Once the robust business focused prevention policies have been established, the organization can then think about applying the right technology to back these up. Their aligning keeps that data secure exactly in that state.

References

[1] Roebuck K. Data loss prevention (DLP). 2011. Tebbo. 418 p.

[2] Kanagasingham P. Data Loss Prevention. SANS Institute. 2008. URL: https://www.sans.org/reading-room/whitepapers/dlp/data-loss-prevention-32883 (Accessed 07.03.2017).

[3] Vasiliev V. DLP systems at the stage of transforming IT to cloud platform. PC Week Review: IT Security. 2011. Vol. 3. Pp. 14–16. (In Russian)

[4] Gartner 2017 Magic Quadrant for Enterprise Data Loss Prevention. URL: https://www.gartner.com/doc/3606038/magic-quadrant-enterprise-data-loss (Accessed 07.03.2017).