Method and Means of Critical Information Processing in Corporate Networks*

Oleksandr Huralnyk[1,*,†], Andrzej Kwiecien[2,†], Vadym Paiuk[1,†], Olexandr Klein[1,†], Oleksii Lyhun[1,†]

[1] *Khmelnytsky National University, Khmelnytsky, Instytutska Str., 11, 29016, Ukraine*
[2] *Silesian University of Technology, Akademicka str., 2A, Gliwice, Poland*

Abstract

The article studies modern methods and means of processing critical information in corporate networks. Since the growth of data volumes increases the risk of data breaches, the main causes of such incidents are analyzed, in particular, unintentional personnel errors (misdelivery, misconfiguration) and abuse of privileges by insiders.

Data Loss Prevention (DLP) technologies are considered as a key mechanism for data protection, including detection, monitoring, prevention and audit of information flows. A comparative analysis of different models of DLP systems (centralized, distributed, hybrid) and their application in the corporate environment has been carried out. Particular attention is paid to commercial solutions from leading manufacturers, their capabilities, advantages and disadvantages.

The article focuses on the current challenges of DLP systems, such as the high frequency of false positives, limited capabilities for analyzing encrypted traffic, difficulties in detecting insider threats, and scalability issues. A review of recent research demonstrates promising approaches combining anomaly analysis, machine learning, attribute-based encryption (CP-ABE), and Zero Trust principles.

The article discusses the peculiarities of using botnets to steal critical information in corporate networks. A solution for protecting information in case of botnet activity is proposed, based on the integration of data leakage prevention systems (DLP) – a module specially configured to neutralize botnet threats.

The results of the study indicate the need for an integrated approach to information security, which includes a combination of technological, organizational and legal measures. This will significantly reduce the risks of data leakage and ensure reliable protection of critical information in the face of modern cyber threats.

Keywords

critical information, data leakage, Data Loss Prevention (DLP), data protection, botnets

1. Introduction

Modern business operations depend on information technology, which leads to a significant increase in the amount of data processed and stored within corporate networks. On the one hand, this opens up significant opportunities for analytics, improving business processes, and making decisions based on processed data. On the other hand, the growth in the volume of information significantly increases the likelihood of data leaks that can have a serious negative impact on companies.

Losses of critical information can occur for various reasons: from unintentional staff errors to intentional malicious actions. This can mean the loss of confidential information, disclosure of intellectual property or financial information, which ultimately causes financial losses, undermines the trust of customers and partners, and may also entail legal liability.

In the era of globalization and active digitalization of business, ensuring information security is becoming one of the priorities. Corporations must implement comprehensive security strategies that

Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security, Aprile 04, 2025, Khmelnytskyi, Ukraine
* Corresponding author.
† These authors contributed equally.
gurualexua@gmail.com (O. Huralnyk); andrzej.kwiecien@polsl.pl (A. Kwiecien); vadympaiuk@gmail.com (V. Paiuk); olexandrkleyn@gmail.com (O. Klein); oleksii.lyhun@gmail.com (O. Lyhun)
0009-0009-1175-8726 (O. Huralnyk); 0000-0003-1447-3303 (A. Kwiecien); 0000-0002-7253-893X (V. Paiuk); 0000-0002-1896-943X (O. Klein); 0009-0004-5727-5096 (O. Lyhun);

encompass not only technological aspects but also organizational and legal aspects. Knowing how to effectively prevent them and implementing combined security systems helps reduce risks and protect the most valuable data.

2. Errors and problems in information processing

According to the analysis of the report [3], the main problems during information processing are Privilege Misuse and Miscellaneous Errors in the context of the two most common types of such errors - Misdelivery and Misconfiguration.

According to [3], about 25% of successful data leakage incidents in the corporate sector are the result of botnet activity. The risk especially increases if the botnet is used for the targeted theft of critical data - financial, intellectual property, personal.

In modern corporate networks, botnets have acquired the status of one of the most dangerous tools of cyberattacks. A botnet is a collection of compromised devices (computers, servers, IoT devices) united under the control of an attacker (the so-called botmaster). Due to the distributed nature of the attack, botnets effectively hide the origin of malicious actions by using the legitimate resources of the victims.

Privilege abuse is almost always associated with insiders using legitimate access to a system for purposes other than its intended purpose—for example, for financial gain, to damage the company's reputation, or for the "convenience" of employees bypassing security measures to get the job done faster. Most often, personal information (customer, employee, partner data) is stolen, which can be easily monetized. In the healthcare industry, access to medical records can be abused because the volume and sensitivity of patient data creates additional opportunities for attackers.

In recent years, the most prominent Miscellaneous Errors have been Misdelivery (sending or transferring data to the wrong person) and Misconfiguration (incorrect access or system configuration settings that cause a leak). Misdelivery is often associated with email (wrong recipient) or sending physical documents to the wrong address. Misconfiguration most often occurs due to incorrect configuration of web servers, cloud storage, etc., when access is open to a wider range of people or is generally available on the Internet without proper authentication. In most cases, personal information is disclosed. The average amount of data that becomes publicly available due to such errors is often significant, and the fact of the leak is often reported by the affected customers themselves or third-party security researchers [3].

The most common channels of data leakage include:

- Network (browser, cloud services).
- Email.
- Printed documents.
- Stolen or lost equipment.
- Removable media (USB, phones, hard drives).
- Instant messaging programs (Telegram, Viber).

Key recommendations for preventing data leaks include logging and monitoring the actions of privileged users, regular auditing, and revoking access rights, especially if an employee resigns or changes position. Training staff to explain why specific security rules exist to reduce the risk of deliberate violations. Automated verification mechanisms to detect potential errors. Secure default settings, especially in cloud environments, to reduce the likelihood of accidental public availability of confidential data. Checking the correctness of email addresses, introducing warnings or confirmations when sending emails with attachments or a large number of recipients. Regular communication with staff about data transfer policies (especially critical information), including working with physical documents. These measures should be part of a more holistic information security strategy, which should include the implementation of technical solutions using modern Data Loss Prevention (DLP) systems [1].

3. Data Loss prevention (DLP) technologies

DLP systems are becoming increasingly important in the field of protecting critical information from illegal extraction, misuse, unauthorized access or accidental deletion, which can occur at endpoints, in networks and in data stores. Their implementation contributes to increasing the level of information security, compliance with the principle of least privilege, monitoring data flows, compliance with legal requirements on confidentiality and protection of intellectual property.

A DLP system is a set of tools and processes aimed at detecting, monitoring and protecting confidential data from unauthorized access, leakage or theft. Such solutions control the movement of data in the corporate environment (at rest, during use and in transit) and prevent it from reaching unauthorized persons [2].

Advanced DLP solutions use encryption, a notification system and preventive measures to reduce the risks associated with data leaks. In today's market, such systems are actively developing in response to the rapid growth of information volume, which requires protection against cyber espionage and cyberattacks in both the government and corporate sectors [1].

The principal components of Data Loss Prevention (DLP) systems include discovery, monitoring, prevention, reporting, and auditing [10]. Discovery entails periodic or continuous scanning of corporate devices and storage systems to identify the presence of confidential data, irrespective of its location. Monitoring ensures persistent oversight of data transmission and usage within information flows, such as network channels, workstations, and application services. Prevention mechanisms enable the blocking, encryption, or redirection of data transfers in alignment with established security policies. Reporting and auditing functions are responsible for generating comprehensive reports on security incidents, user activities, and policy compliance, while also maintaining an event log to support subsequent analysis.

Such multi-level coverage Figure 1 makes it possible to detect potential leaks in time and respond to threats, in particular through automated scenarios or notifications to responsible persons.

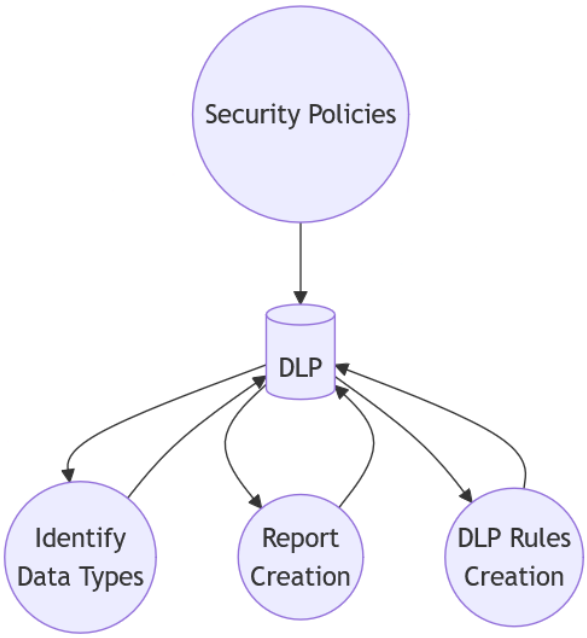

Figure 1: DLP-system interaction levels.

A typical algorithm for a DLP system's operation [4] encompasses several interrelated phases designed to ensure comprehensive data protection. The initial stage, data collection and interception, involves deploying an endpoint agent or network sensor to capture information flows. Subsequently, the system conducts content analysis by searching for specific keywords, phrases, or regular expressions that may indicate the presence of sensitive information. The analyzed data is then

compared against established security policies, prompting preventive measures—such as issuing warnings or blocking transmissions—when a match is detected. Throughout this process, all events, user identifiers, timestamps, and transmission channel details are recorded in a logging module to facilitate both immediate oversight and post-incident investigation. Lastly, the system's reaction phase may halt the sending of information, enforce document encryption, or dispatch notifications to an administrator.

In modern DLP solutions, data is generally categorized into three broad types [5]. The first, Data in Use, pertains to the regulation of data usage on workstations, which includes monitoring actions such as copying, printing, or creating screenshots. Data in Motion focuses on overseeing and managing data transfers within network traffic, including web, email, and FTP channels. Finally, Data at Rest entails scanning storage repositories—such as file servers and databases—to detect and classify confidential files. This tripartite classification enables organizations to implement targeted safeguards for each stage of the data lifecycle.

4. Classification of DLP systems

During the development of DLP systems, researchers commonly differentiate between centralized and distributed models. In the centralized approach, event collection and analysis occur under the control of a single console, with all endpoint agents and sensors (installed on client PCs, servers, or network devices) transmitting data to a central module. Such a configuration facilitates straightforward management and rapid coordination of security policies, but requires sufficient computational capacity to accommodate large traffic volumes. In contrast, the distributed model conducts event analysis locally, within the specific part of the system where a security threat materializes. Although this design alleviates the burden on central nodes, it can introduce complexities in maintaining consistent policies and a unified response methodology [5].

Hybrid solutions are frequently employed in practice, merging the advantages of centralized administration and distributed analysis according to the infrastructure's scale and security requirements. A comparative overview of the key features of centralized, distributed, and hybrid data loss prevention (DLP) systems is presented in Table 1.

Beyond these architectural models, DLP technologies can be grouped by their deployment strategies. Endpoint DLP relies on agents installed on local machines (such as desktops or laptops) to provide fine-grained oversight of user actions, including copying data to external media, printing documents, or capturing screenshots [6]. Network DLP operates through sensors or proxies that intercept and analyze traffic in real time, thereby offering robust control over data in motion while still exhibiting limitations in monitoring localized activity [7]. Discovery DLP targets data at rest by scanning file repositories, mail servers, and databases to identify and label sensitive information, enabling a more precise application of security policies [26]. Cloud DLP, relevant to SaaS environments such as Office 365 and Google Workspace, focuses on controlling uploads and downloads to and from the cloud, and is often integrated with a Cloud Access Security Broker (CASB) to extend protection capabilities [25].

Table 1
Comparison of the main features of data loss protection systems

Criterion	Centralized DLP	Distributed DLP	Hybrid DLP
Architecture	All management and analytics logic is concentrated in a single node (server or service platform).	Each node or region has its own analyzers and decision logic that can communicate with each other.	Combination of a centralized management server and multiple local/regional nodes.

Deployment	Easier to deploy as a "single point of control"; requires agents or sensors to be installed on endpoints or network nodes.	Typically deployed in large geographically distributed networks (branch offices, different data centers).	Central server + local nodes/agents in remote networks or branches.
Management and Monitoring	Easier for the administrator: all configuration and reporting are managed from a single console.	Multiple points of management; more complex synchronization of policies and reporting.	Centralized policies with the ability to fine-tune at the local level.
Scalability	May be limited by the performance of the central server; high load requires scaling out or up.	Scalable: adding new nodes increases computing resources and bandwidth.	Flexible scaling: the main server scales "up" and distributed nodes "down".
Performance	May be worse during peak load as all analytics go through a single node.	High performance due to parallel processing (each node processes its own data).	Balanced performance: some processing is performed centrally, the rest on the nodes.
Reliability	Vulnerability due to "single point of failure": a failure of the central server can bring the entire system down.	More fault tolerance: failure of one node does not disrupt the entire system.	Higher than centralized, but depends on the availability of the central node and remote ones.
Configuration Complexity	Generally lower: just configure a central server and deploy clients.	Higher: requires synchronization of configurations on different nodes, complex network topology.	Medium: centralized management, but local configuration may be required for remote nodes.
Maintenance Cost	Lower in small and medium-sized organizations; increases with scale.	Can be high due to distributed management, especially in large infrastructures.	Balanced costs: more flexibility, but still requires coordination between the central node and local ones.
Flexibility of Policy Updates	Everything is done from a single place, but there may be delays due to high load on the central server.	Policies are updated independently on each node, which allows for faster local response, but more difficult to centralize.	Centralized distribution with the ability to make individual changes on local nodes (balance of speed and consistency).
Application Examples	Typically – small and medium-sized companies, or large companies with a clearly centralized infrastructure.	Large corporations with numerous remote offices, data centers, and global presence.	Organizations with mixed infrastructure: some systems are concentrated in one place, and some are distributed.

Main Advantages	Ease of deployment, centralized control, single source of truth.	High scalability, localized processing, no single point of failure.	Combining the advantages of both approaches: centralized management + distributed detection and processing.
Main Disadvantages	"Single point of failure", there may be problems with bandwidth and performance in large-scale networks.	Complex policy management and synchronization, higher administration costs	Requires careful architecture: it is necessary to configure the interaction between the central server and distributed nodes.

5. Overview of modern DLP solutions

Systems for comprehensive detection and prevention of data leaks for corporate networks can be divided into commercial and research (non-commercial). In this article, we will consider commercial systems, which, unlike research ones, offer more complete and multifunctional options for protecting critical information, and can also provide comprehensive solutions with scalability and integration capabilities.

Below is an overview of the architecture of DLP systems, the main components, the interaction between them and the place of the system in the overall data protection infrastructure are considered.

Broadcom – Symantec Data Loss Prevention [10] is a comprehensive solution composed of several interdependent components. The Enforce Server, also known as the Management Console, serves as the central node where administrators configure policies, dictionaries, and overarching system settings. Detection Servers, each specialized in a particular function, directly analyze network traffic or files; for example, Network Prevent (for Web or Email) integrates with network and mail gateways to monitor and block unauthorized transmissions, Endpoint Prevent operates as an agent on workstations to regulate USB and clipboard usage, and Data Insight or Network Discover periodically scans file repositories and databases to locate confidential data. A dedicated Database stores incident logs, policy definitions, and scanning results, supporting both immediate threat detection and historical analysis.

In operational terms, the Enforce Server generates security policies and synchronizes them with the Detection Servers, which conduct either real-time or scheduled data analysis. Policy triggers are returned to the Enforce Server, recorded as incidents, and included in relevant reports. One of the system's notable features is its clear delineation between the Management (Enforce) tier and the Detection layer, enabling seamless clustering and scalability by adding new Detection Servers as needed. Its broad integration capabilities extend to mail gateways, proxy servers, and cloud services, thereby ensuring multi-layered protection that spans endpoint controls (Data in Use), network traffic (Data in Motion), and storage environments (Data at Rest). The solution offers modules such as Network Prevent (for web or mail), Endpoint Prevent, and Discovery (for scanning local and network resources), supplemented by a considerable library of predefined templates aimed at detecting personal data and financial information, including PCI DSS, SOX, and HIPAA compliance. Deep infrastructure integration and high-accuracy detection of sensitive data position this platform as a powerful tool; however, its deployment can be complex and requires meticulous configuration to minimize false positives.

Forcepoint's Data Loss Prevention Portfolio Suite [11] represents a comprehensive solution encompassing multiple components that collectively provide robust protection against data exfiltration. The Forcepoint Security Manager (FSM) acts as a centralized administrative console, enabling the configuration of DLP policies, analytics, and reports. Endpoint Agents operate on both stationary and mobile workstations, overseeing activities such as file manipulation, clipboard usage,

and interactions with USB devices. Network Agents or Protectors analyze network protocols—among them HTTP, HTTPS, and FTP—and can integrate with existing proxies and gateways, while the Forcepoint CASB module extends DLP functionality to cloud applications. The system's analytics and reporting capabilities include user and entity behavior analysis (UEBA), facilitating more sophisticated monitoring of user actions.

From an operational perspective, FSM transmits defined policies to both Endpoint Agents and Network Agents, ensuring that data in use (encompassing local copying, printing, and other endpoint activities) and data at rest (through Discovery modules) remain under consistent scrutiny. Network Agents further monitor data in motion across network channels, and Forcepoint CASB broadens the protective scope to various cloud services. Among its notable attributes is the system's capacity for elastic scaling according to the number of endpoints and control points present in the network. The solution supports Network DLP, Endpoint DLP, Email DLP, and CASB, overseeing data transfer via email, web, FTP, and removable media. A unified protection policy spanning both DLP and web security, combined with the capability for granular policy tuning and network segmentation, underscores its versatility. At the same time, large-scale deployments may necessitate significant investments and demand meticulous coordination across endpoint and network agents to ensure optimal performance.

Fortra Digital Guardian [12] is a solution offered by Fortra that integrates several interdependent components to provide comprehensive protection against data exfiltration. The primary Management Console acts as a central hub for policy administration, auditing, and reporting, while the Endpoint Agent—installed on Windows, Mac, or Linux workstations—serves as the main mechanism for monitoring files, processes, and data operations in real time. Additionally, Discovery & Classification modules automatically identify and categorize sensitive data, and an optional Network Sensor can be employed to analyze HTTP and SMTP traffic, though the system's emphasis remains on endpoint-level controls.

During operation, the Endpoint Agent continuously observes activities such as copying, forwarding, and printing. If an event conflicts with configured security policies, it can either block the action, encrypt the file, or issue a notification to the Management Console. Upon receiving event logs, the console generates detailed reports and enables the administrator to refine policies in accordance with emerging requirements. The platform distinguishes itself through its robust insight into user behavior at workstations, offering thorough safeguards against insider threats by tracking file, clipboard, printer, and removable-media actions in conjunction with real-time data classification. Rules may also be dynamically adapted on the basis of user risk profiles or active processes, ensuring a high degree of responsiveness. Although its endpoint focus delivers substantial efficacy in securing "Data in Use," organizations seeking a more holistic scope for network or storage protection may need to integrate supplementary tools. As with many endpoint-centric DLP architectures, careful configuration of agents is crucial to minimizing false positives and maintaining optimal performance.

GTB Data Protection Suite [13] from GTB Technologies comprises several interconnected components that collectively provide a robust data protection framework. A hardware or virtual GTB Inspector monitors network traffic (HTTP, SMTP, and FTP) in either inline or tap mode, identifying potential leaks as they occur. The Endpoint Protector, deployed on individual workstations, controls local data operations and can enforce encryption, while the Management Console aligns network and endpoint policies, oversees event logs, and produces comprehensive reports. An Exact Data Matching (EDM) Engine enhances detection accuracy by comparing transmitted information against known patterns—such as a database of payment card numbers—and recognizing even partial matches.

This integrated approach ensures coverage of data in motion and data in use, with the Management Console unifying policies across the Inspector and Endpoint Protector and thereby facilitating immediate detection and response. A patented content-aware analysis method allows the system to recognize confidential data even when fragmented, minimizing false positives. The suite includes Endpoint Protector, Network Protector, Email Protector, and Cloud DLP modules, and supports various detection methods—ranging from regular expressions and lexical analysis to

machine learning algorithms and EDM—ensuring flexibility and precision. While it holds a smaller market share and offers fewer predefined integrations and dictionaries compared to industry leaders, GTB Data Protection Suite distinguishes itself through its advanced detection technologies and capacity for refined data classification.

ManageEngine's Log360 (DataSecurity Plus) [14] offers a unified platform for log collection and analysis, user activity tracking, and DLP functionality. The core Log360 component aggregates and evaluates logs from servers and network devices, while the DataSecurity Plus module monitors file servers, configures DLP policies, identifies confidential data, and tracks user actions. Both elements connect through a central Management Console, where logs and DLP-related incidents converge for reporting and oversight.

This integrated solution provides SIEM capabilities, enables comprehensive audits of the Windows environment, and consolidates incident analysis and data leak prevention within a single user interface. Its Active Directory Audit, File Server Audit, and DLP features form part of a cohesive ecosystem particularly well suited to small and medium-sized organizations. Although the DLP module may lack the nuanced detection accuracy of leading competitors, the system's ease of deployment and centralized management interface render it a convenient option for entities seeking combined log management, security incident analysis, and data loss prevention.

Proofpoint's Sigma Information Protection Platform [15] comprises several interconnected modules designed to safeguard sensitive information across multiple vectors. Proofpoint Email Security, which can be deployed either as a gateway or cloud-based service, inspects both incoming and outgoing email traffic, applying DLP checks to prevent data leaks and detect phishing attempts. Its Cloud App Security Broker (CASB) component extends protection into cloud environments such as Office 365 and Google Workspace, classifying files and regulating public access. An optional Endpoint DLP Agent offers an additional layer of defense for local data operations, monitoring activities related to USB usage and the clipboard. All of these elements converge in the Information Protection Console, which presents a unified interface for managing policies, overseeing content lists, and reviewing log data and incidents.

A core emphasis of the Sigma platform lies in its cloud-centric capabilities, covering email channels, cloud services, and phishing prevention. By employing user and entity behavior analytics (UEBA), the system identifies anomalous actions that deviate from a typical user's activity profile. Its modules include Email DLP, the Cloud App Security Broker, and Endpoint DLP, featuring robust algorithms for content analysis and phishing detection. Although the solution's breadth in network-based scenarios may be comparatively narrower, it delivers comprehensive coverage of email and cloud workflows. Organizations choosing to deploy endpoint modules may require additional licenses, reflecting the product's flexible yet modular approach to enterprise data protection.

A comparison of the main advantages and disadvantages of each of the considered DLP solutions is given in Table 2.

Table 2
Advantages and disadvantages of commercial DLP systems

Solution	Advantages	Disadvantages
Broadcom – Symantec DLP	Clearly separated architecture: Enforce Server + Detection Servers (Endpoint, Network, Email) for scalability. Rich Discovery variety (Network Discover, Data Insight), allows you to cover "Data at Rest". Powerful policy configuration via Enforce Console.	High deployment complexity (need to coordinate between multiple Detection Servers). Tight integration may require a large number of resources (servers, databases). Possible performance issues with a large number of Detection Servers.

Forcepoint – DLP Portfolio Suite	Modular structure (Email, Web, Endpoint, CASB), allowing you to adapt to different environments. Single console Forcepoint Security Manager, where policies are synchronized with agents and network modules. Well integrated with Forcepoint UEBA.	Increased complexity when using multiple modules simultaneously (Network DLP, Endpoint DLP, CASB). Often requires multiple servers or virtual machines for full functionality. Resource intensive with high network traffic
Fortra – Digital Guardian	Strong emphasis on endpoint architecture: all critical events are processed locally by the agent. Deep control of user actions (Data in Use) in real time. Flexible policies applied at the endpoint, fast response.	Smaller network component, if full-fledged Network DLP is required, an additional solution is required. Dependence on endpoint agents, which can complicate support for different platforms. With a large number of endpoints, the load on the management server increases.
GTB Technologies – GTB Data Protection	EDM technology shows high accuracy in content analysis. Single management of Endpoint + Network with the possibility of "tap/mirror" mode. Less "heavy" architecture than some competitors, with an emphasis on accurate content analysis	Fewer ready-made integrations and templates than market leaders. Requires fine-tuning of EDM, without which it may not work correctly. Lack of support for certain environments.
ManageEngine – Log360 / DataSecurity Plus	Combining SIEM functionality with a DLP module. Single console for incident analytics and file management, Active Directory Audit. Easy implementation, especially for medium-sized Windows networks.	DLP capabilities may be inferior in depth and flexibility to specialized DLP solutions. May lack individual implemented cloud or network components for complex environments. Dependence on Windows ecosystem.
Proofpoint – Sigma Information Protection	Powerful Email DLP, anti-phishing, powerful integration with email gateways. CASB module for data control in cloud services. UEBA component for user behavioral analysis	Network scenarios (HTTP) are less covered, other solutions must be used (Network Firewall, Proxy). Endpoint DLP has less functionality than competitors. High cost of the complete set (Email + CASB + Endpoint).

All commercial solutions considered have a common goal - to detect, monitor and protect confidential data from leakage (insider or external). However, they differ in specialization and scope of application. The difference also lies in the depth of integration (endpoint or Network), specialization (Email, insider threats) and deployment method (on-premises, hybrid, SaaS). Each system has a central policy manager and one or more types of agents. In some solutions (Symantec, Forcepoint) the architecture is very large-scale, distributed, while in others (Safetica) it is more compact, with an emphasis on rapid deployment and simplicity.

Each solution has its own "pluses" (area of main expertise) and "minuses" (gaps in integration, complexity of architecture or lack of broad functionality). The choice depends on the scale of the company, the level of risk, budget and existing IT landscape (on-premises or cloud). A significant criterion when choosing a DLP system is the ability to scale and customize policies for specific industries (e.g., finance, healthcare, government). It is also important to pay attention to compatibility with existing corporate network protection and management systems (Firewall, IAM, SIEM, etc.).

6. Disadvantages of existing DLP systems

Modern DLP systems either detect deviations from typical behavior (based on anomalies) or apply signatures of known attacks. Although signature-based methods allow you to accurately identify familiar threats and effectively block them, they are unable to respond to unforeseen attacks or attackers who use unusual intrusion vectors (for example, those known only to insiders). In contrast, anomaly-based methods can detect as yet unknown attacks, but due to frequent false positives, their contribution to detecting suspicious activity is limited. [17]. The system considers plain text to be critical data (for example, a format match with a credit card number). As a result, users receive excessive blocking, and work efficiency decreases.

One of the shortcomings of DLP systems is the problem of working with encrypted traffic. DLP does not see the content of encrypted data. Proxy decryption is required, but this increases the computational load. Integration with endpoint agents that analyze data before encryption is also possible [18].

Working with insiders with legitimate access is problematic. DLP can miss a leak if a user has official access to a file, but sends it through a prohibited channel [22]. It is necessary to monitor anomalous actions, even if they are formally permitted, and monitor upload volumes, time of day, and unusual addresses.

With large volumes of traffic, performance and scalability problems may arise, and analysis may create delays.

7. Review of new research

Research [17] presents a hybrid DLP system that combines anomaly-based and signature-based approaches to facilitate prevention and detection. Using an anomaly-based mechanism, the framework automatically creates a model of typical user behavior that allows it to detect unusual transactions made by insiders. Operator responses to alarms are then used to automatically create and update attack signatures.

In [16], a methodology combining behavior-based DLP, machine learning-based DLP, and network-based DLP is proposed to improve data integrity in cyberspace. This approach provides a comprehensive and integrated solution that detects and prevents data loss across the entire network. The proposed methodology includes identifying sensitive data, deploying endpoint DLP, network DLP, and machine learning DLP solutions, integrating DLP solutions, implementing monitoring and reporting, and continuous system improvement.

[18] considers the use of a ciphertext policy attribute-based encryption algorithm to protect data both during transmission and at rest. The system triggers preventive actions, such as encryption updates or access restrictions, in case of suspected data breaches to mitigate the consequences. By combining anomaly detection techniques with Ciphertext-Policy Attribute-Based Encryption (CP-ABE), a robust framework is created to improve data security and privacy across a range of domains, providing a proactive line of defense against potential data breaches.

In their paper [19], they propose to strengthen organizational security by using a zero-trust security approach that should completely replace traditional solutions. In addition to restricting access to resources through the principle of least privilege, the zero-trust paradigm emphasizes the importance of authenticating and verifying each person and device. Devices are only granted access after they have received their credentials and access rights. User authentication, application security, and device security are some of the factors considered. A zero-trust framework is a practical way to address the challenges of modern network systems and secure legacy systems. The paper analyzes the zero-trust security problem in a number of modern networks, including the financial industry, IoT (Internet of Things), enterprise, and 5G networks.

The paper [20] examines an approach to mitigate cybersecurity risks associated with text data leakage. Unlike traditional DLP products, this solution focuses not only on creating and updating policies, but also on minimizing false positives. A web extension is presented that captures and

securely transmits all data entered in the browser to the organization's server. Using a server-based Bidirectional Encoder Representations from Transformers (BERT) trained on specific organization datasets, the system classifies text to improve the detection of critical information.

In this [22] research paper, a new level of data leak detection based on insider behavior and trust in the insider is proposed. The proposed system is based on a multi-agent paradigm that is used for threat detection because it can solve the problem of behavior logs in terms of self-sufficiency and productivity. A new dimension of analysis is added based on the behavior of an internal employee and his various interactions in the information system.

In [23], an approach to botnet detection for distributed systems is presented. It is based on a developed three-level model that includes botnet components: a control center, control centers, and botnet core elements (bots). The new framework provides the ability to detect known and unknown botnets and consists of host and network levels.

In [24], a new method for detecting DDoS botnets is proposed based on the analysis of the network characteristics of botnets. It uses semi-supervised fuzzy c-means clustering. The analysis is based on characteristics extracted from network traffic that may indicate the presence of DDoS botnets on the network.

In [21], a new information technology for detecting botnets is proposed based on the analysis of botnet behavior in a corporate network. Botnet detection is performed in two ways: using network-level analysis and host-level analysis. One approach allows analyzing the behavior of software on the host, which can indicate the possible presence of a bot directly in the host and identify malicious software, and the other involves monitoring and analyzing DNS traffic, which allows concluding that network hosts are infected with a botnet bot.

In the article [8], a method for detecting botnets in corporate networks is presented. It is based on the use of artificial immune system algorithms. The proposed approach allows distinguishing harmless network traffic from malicious traffic using a clonal selection algorithm taking into account the features of the presence of a botnet in the network.

In [9], information technologies were developed that are used to collect, process, and store large amounts of data from the web. The technology studies the statistical characteristics of different segments of the web space and their cluster structure. To find the optimal number of clusters and cluster centers, two methods are used: the well-known k-core decomposition algorithm and a new method developed by the authors. The new algorithm is based on the distribution of eigenvalues of the stochastic matrix, which describes the process of Markov transitions in the system. The clustering process is carried out using the Power iteration clustering algorithm.

8. Mechanism of stealing critical information through botnets

Botnets typically revolve around a command and control (C&C) center, one or more intermediate nodes, and a set of compromised devices known as bots [23]. The C&C center can function through a centralized server or a peer-to-peer network, enabling attackers to distribute commands to every infected machine under their control. Intermediate nodes—often acting as proxies or relays—serve to obfuscate the true origin of malicious activities, making it more challenging to identify the attacker's location. Each bot, once infected, executes the directives it receives, which may involve scanning the network, collecting files, or exfiltrating sensitive data.

In a common attack sequence, the botnet operator gains unauthorized access by exploiting software or operating system vulnerabilities, frequently aided by social engineering methods such as phishing. Once established on the target system, the malicious code runs in a concealed mode and waits for further commands from its control server. It often secures elevated privileges (root or administrator) and circumvents antivirus solutions, allowing it to search local files, databases, or other repositories for confidential information, such as financial records or intellectual property. When sufficiently valuable data is identified, the malware prepares it for clandestine extraction—sometimes by embedding it into sessions that appear legitimate, for example via HTTP or HTTPS, or by employing hidden tunnels (DNS or P2P) [8]. Attackers may additionally encrypt the outbound

traffic to make network-level detection more difficult. Since these transfers often blend with typical user activity, conventional security tools may fail to recognize them as malicious, thereby allowing the botnet operator to systematically remove sensitive documents without drawing attention.

Traditional security measures demonstrate limited effectiveness against targeted botnet activity. Firewalls, although vital for perimeter defense, rarely provide the depth of content inspection needed to identify data leaks hidden within benign traffic. Antivirus applications rely on known signatures and may prove incapable of detecting new or obfuscated threats. Intrusion detection and prevention systems (IDS/IPS) can flag anomalies but often generate excessive false positives and do not always uncover exfiltration disguised as ordinary communication. Consequently, these solutions can be circumvented by attacks that specifically aim to mimic legitimate network behavior.

To counteract this sophisticated threat, a more comprehensive strategy integrates a specialized botnet detection mechanism with a data loss prevention (DLP) system. This combined approach targets both network-level indicators of compromise—such as irregular traffic patterns or unusual encryption usage—and the handling of sensitive data at endpoints. By correlating information from botnet monitoring with DLP logs, security teams can identify subtle warning signs of exfiltration that might otherwise remain undetected. Ultimately, unifying these technologies enhances the capacity to intercept unauthorized data transfers, narrows the window in which attackers can operate, and strengthens an organization's overall resilience against critical information theft.

9. System architecture diagram with integrated information protection component

In the proposed architecture of the critical information processing system Figure 2, there are four key nodes: a computer or server on which users work, which can be infected by a botnet (Endpoint Host), a small software module that monitors data manipulation (DLP Agent), a central policy and incident management server (DLP Console), and a network traffic monitoring module designed to detect botnet activity (Botnet Detection).

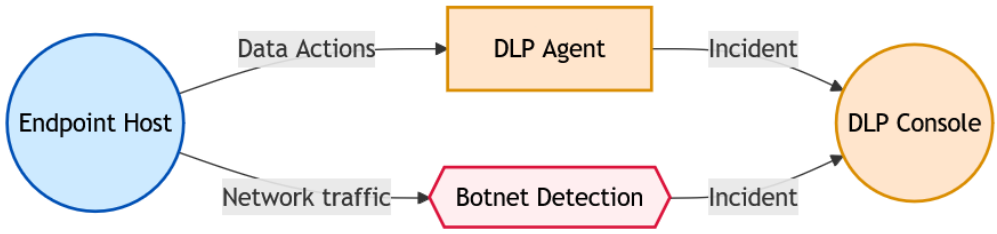

Figure 2: System architecture.

In the process of operation, the system operates as follows: when a user performs certain actions on a workstation or server (in particular, working with documents, copying files, using external media), the DLP Agent tries to recognize whether the files are confidential and whether the actions do not contradict the established policies. If violations are detected, the agent sends a signal (incident) to the DLP Console, where logs can be stored and warnings for administrators can be generated. At the same time, this same Endpoint Host generates network traffic that passes through Botnet Detection. If the internal algorithm or signature analysis detects suspicious behavior typical of botnets (various commands or connections to dubious addresses), the detection module sends a signal (incident) to the DLP Console, signaling possible damage. Then, by comparing data from both sources (agent and botnet detection module), the DLP Console can establish a correlation: if a bot-like program tries to send files silently, this becomes a reason to block the transfer, tighten policies, or inform security. This way, confidential files cannot be transferred without knowledge or approval, and suspicious activities are immediately reported to the notification system, allowing for a rapid response to incidents.

10. Experimental studies

To evaluate the effectiveness of the proposed architecture, where DLP and the botnet detection module work together, we will conduct experiments in which a botnet attack with a data leak attack is simulated and the system is tested to what extent it is able to detect and block unauthorized extraction of critical data. The scenario and parameters by which the system's performance will be evaluated are described below.

In this experiment, test bots (malware emulators) are installed on workstations. They try to send sensitive files via a previously known channel (HTTP, HTTPS, DNS tunneling). The goal is to check how the system captures this activity.

On the host where the DLP Agent operates, a set of files marked as confidential (financial reports) is placed. The bot emulator tries to send these files to a remote server. It is checked whether the DLP Agent blocks copying or sending of prohibited files, and whether the botnet detection module detects a suspicious connection.

Evaluation parameters for this experiment:

Detection Rate: the proportion of successfully detected attempts to transmit critical information among all attacks.

Response Time: how much time passes from the moment the leak begins to blocking or notification.

False Positives: the number of normal operations that the system mistakenly recognized as malicious.

The results of the experiments are presented in Table 3.

Table 3
Experimental results

Test number	Attack scenario	Detection Rate (%)	Response Time (sec.)	False Positives (%)
1	Fast sending of large files	97%	10	2.5%
2	Gradual extraction of small data fragments	94%	20	3.0%
3	Using a non-standard port (TCP/8081)	96%	15	2.8%
4	Hidden encryption of files before sending	92%	25	3.5%
5	Masquerading traffic as legitimate HTTPS	95%	18	3.2%

According to the test results, the average Detection Rate exceeds 90–95%, which indicates a fairly high ability of the system to detect attempts at unauthorized data extraction even under conditions of masking or encryption. The average response time ranges from 10 to 25 seconds depending on the complexity of the attack, which allows for relatively quick blocking of the transfer of sensitive files. The false positive rate, which ranges from 2.5–3.5%, is considered acceptable for real corporate environments: this means that the system does not block or notify about legitimate actions too often. In general, such indicators indicate the effectiveness of the proposed integrated architecture in countering botnet attacks aimed at stealing critical information, while leaving the possibility of further configuring policies to reduce false positive incidents.

11. Conclusions

This article analyzes modern methods and tools for processing critical information in corporate networks, in particular data leakage prevention systems (DLP). The main causes of information leaks are investigated, including technical errors of personnel (misdelivery, misconfiguration) and abuse

of privileges by insiders. It is determined that information security threats require a comprehensive approach, including both technological and organizational measures.

The main types of DLP systems (centralized, distributed, hybrid) and their features in the corporate environment are considered. An analysis of popular commercial solutions is conducted, their advantages and disadvantages are identified. Particular attention is paid to such challenges as a high number of false positives, difficulties in processing encrypted traffic, the complexity of detecting threats from insiders, and the issue of scalability of DLP systems.

A review of current research has demonstrated promising methods for improving DLP solutions, including the use of machine learning algorithms for more accurate threat detection, attribute-based data encryption (CP-ABE), and the implementation of the Zero Trust concept to strengthen access control.

The results of the experiment showed that the interaction of the DLP module and network anomaly detection mechanisms allows you to effectively block unauthorized transmission of critical data and provides a basis for further strengthening the security system.

Thus, to effectively protect critical information, organizations should apply a multi-layered approach that combines advanced technologies, clearly configured security policies, constant monitoring and staff training. The use of modern DLP solutions in combination with other cybersecurity tools can significantly reduce the risks of data leakage, ensure compliance with regulatory requirements, and guarantee the company's information resilience in the face of growing cyber threats.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: grammar and spelling check; DeepL Translate in order to: some phrases translation into English. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

[1] B. S. Shishodia and M. J. Nene, Data leakage prevention system for internal security, Proceedings of the International Conference on Futuristic Technologies (INCOFT), Belgaum India, 2022, pp. 1-6. doi:10.1109/INCOFT55651.2022.10094509.

[2] What is Data Loss Prevention (DLP), 2025. URL: https://www.gartner.com/en/information-technology/glossary/data-loss-protection-dlp.

[3] Verizon data breach investigations report, 2022. URL: https://www.verizon.com/business/en-gb/resources/reports/dbir/2022/summary-of-findings/.

[4] Symantec DLP White Paper, 2022. URL: https://docs.broadcom.com/doc/white-paper-data-protection-where-it-matters.

[5] What are the 3 types of Data Loss Prevention, 2025. URL: https://www.xcitium.com/what-are-the-3-types-of-data-loss-prevention/.

[6] Protect sensitive data on all your endpoints, 2025. URL: https://www.digitalguardian.com/resources/datasheet/endpoint-dlp.

[7] Network DLP, 2025. URL: https://www.forcepoint.com/cyber-edu/network-dlp.

[8] S. Lysenko, K. Bobrovnikova and O. Savenko, A botnet detection approach based on the clonal selection algorithm, Proceedings of the IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv Ukraine, 2018, pp. 424-428. doi:10.1109/DESSERT.2018.8409171.

[9] O. Kyrychenko, Information technology for statistical cluster analysis of information in complex networks, Computer Systems and Information Technologies 4 (2022) 47–51. URL: https://doi.org/10.31891/csit-2022-4-7.

[10] Broadcom – Symantec Data Loss Prevention, 2025. URL: https://www.broadcom.com/products/cyber-security/information-protection/data-loss-prevention.
[11] Forcepoint Data Loss Prevention, 2025. URL: https://www.forcepoint.com/product/dlp-data-loss-prevention.
[12] Fortra – Fortra's Digital Guardian, 2025. URL: https://www.fortra.com/product-lines/digital-guardian.
[13] GTB Technologies – GTB Data Protection Suite, 2025. URL: https://gttb.com/network-enterprise-dlp/.
[14] ManageEngine – Log360 (Data Security Plus), 2025. URL: https://www.manageengine.com/log-management/.
[15] Proofpoint – Sigma Information Protection Platform, 2025. URL: https://www.proofpoint.com/us/products/defend-data.
[16] I. Yadav and H. Gupta, Designing Data Loss Prevention system for the enhancement of data integrity in cyberspace, Proceedings of the 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida India, 2023, pp. 1361-1365. doi:10.1109/ICAC3N60023.2023.10541823.
[17] A. Srivastava, P. Srivastava, A. Pillai and V. S. Sharma, A hybrid framework for Data Loss Prevention and detection, Proceedings of the International Conference on Signal Processing and Advance Research in Computing (SPARC), Lucknow India, 2024, pp. 1-6. doi:10.1109/SPARC61891.2024.10828891.
[18] V. Sasikala, P. Lakshmi, P. Nagaraja, V. Lakshmi, CH. Bhanu, Data Leakage Detection and prevention using ciphertext-policy attribute based encryption algorithm, Proceedings of the 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida India, 2024, pp. 1-5. doi:10.1109/ICRITO61523.2024.10522194.
[19] S. Nagaraj, Shankaramma, Framework analysis and zero trust security issues in contemporary network systems, Proceedings of the 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), Bengaluru India, 2024, pp. 1-6. doi:10.1109/CSITSS64042.2024.10816783.
[20] V. K. Kaliappan, P. U. Dharunkumar, S. Uppili, A. Adhi Vigneshwarar, S. Bharani, Sentinel guard: an integration of intelligent text Data Loss Prevention mechanism for organizational security (I-ITDLP), Proceedings of the International Conference on Science Technology Engineering and Management (ICSTEM), Coimbatore India, 2024, pp. 1-6. doi:10.1109/ICSTEM61137.2024.10560825.
[21] S. Lysenko, O. Savenko, K. Bobrovnikova, A. Kryshchuk, B. Savenko, Information technology for botnets detection based on their behaviour in the corporate area network, in: Gaj, P., Kwiecień, A., Sawicki, M. (eds), Computer Networks. CN 2017. Communications in Computer and Information Science, Springer, Cham, vol 718, 166–181. doi:10.1007/978-3-319-59767-6_14.
[22] M. E. Moudni, E. Ziyati, Data leakage prevention approach based on insider trust calculation, Proceedings of the 10th International Conference on Wireless Networks and Mobile Communications (WINCOM), Istanbul Turkey, 2023, pp. 1-6. doi:10.1109/WINCOM59760.2023.10322935.
[23] O. Savenko, A. Sachenko, S. Lysenko, G. Markowsky, N. Vasylkiv, Botnet detection approach based on the distributed systems, International Journal of Computing, 19(2), 190-198. URL: https://doi.org/10.47839/ijc.19.2.1761.
[24] S. Lysenko, O. Savenko, K. Bobrovnikova, DDoS botnet detection technique based on the use of the semi-supervised fuzzy c-means clustering, CEUR-WS (2018), vol.2104, 688-695. URL: https://ceur-ws.org/Vol-2104/paper_251.pdf.
[25] Microsoft Purview, 2025. URL: https://learn.microsoft.com/en-us/purview/.
[26] Varonis Data Loss Prevention, 2025. URL: https://www.varonis.com/platform/dlp.