Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2026-24.1.0020

RESEARCH ARTICLE

Data Exfiltration Protection Method Usıng a One-Off Dıstrıbutıon Prıvacy Tags

Larisa Cherckesova[1*], Olga Safaryan[2], Evgeniya Roschina[3]

[1, 2, 3] Don State Technical University, Rostov-on-Don, Russia

ARTICLE INFO

Received: Jan 24, 2026

Accepted: Apr 4, 2026

Keywords

Data Protection Method
Privacy
Information Security
Zero Trust

***Corresponding Author**

safari_2006@mail.ru

ABSTRACT

In this paper one of the most developing areas in the field of information technology, namely information security, has been considered. The study identified the main problems of modern security systems, which became the basis for the development of a new approach to information assurance. Due to the huge number of unaccounted vulnerabilities associated with insiders, the Zero Trust security model, supplemented by the preservation of file confidentiality state, became the basis for the development. That feature critically reduces the chance of extracting sensitive data from a protected information system

INTRODUCTION

Internet-connected devices, including Internet of Things (IoT) objects, are used for data theft. Data exfiltration is the unauthorised transfer of data from a system by someone with malicious intent or malware installed on the system (Anderson, 2020).

This can be done either physically or remotely.

Remote exfiltration typically uses a form of backdoor or malware with the ability to covertly monitor any outgoing or incoming traffic and current data stored on the target system (Chung et al., 2023). When someone can gain physical access to the target system, data exfiltration can be done physically, for example, by transferring data to an external storage device that can be easily hidden. The threat of data exfiltration is very high because by the end of the first quarter of 2020, 1196 publicly disclosed data breaches were reported worldwide, resulting in 8.4 billion records or data fragments being compromised (Chung et al., 2023). While many data theft threats are at the network layer and can be addressed with appropriate network security deployments, there are also physical threats that need to be addressed to ensure protection across all areas of a company's infrastructure.

Typically, physical threats can be addressed through physical measures such as fencing systems to prevent unauthorized access, inserting external storage devices and using physical locks to prevent theft of the system from the premises. In IoT, because of the placement of devices in an accessible location, it can be much more difficult to eliminate physical threats. In addition, a network solution must be designed to monitor and observe any physical actions performed on any systems on the network. In this research, we describe such a solution that is able to monitor and prevent unauthorized actions being performed on systems in a network, such as data transfer or modification. With digital fingerprinting and network-wide endpoint rules, any potential data leakage or

exfiltration can be mitigated as any data that is considered sensitive or any action that is not on the list of authorized rules is prohibited and therefore cannot be performed.

Since the focus of this mechanism is to prevent data theft attacks on the internal network, the solution is to use zero-trust security principles because these principles assume that no one can be trusted and everyone, regardless of their clearance. or privileges, must have their actions and file usage verified. The zero trust principle explicitly verifies that authentication and authorisation of all actions are performed regardless of the credentials or permissions of the requesting user. While some users will have more privileges than others, all must be verified using key points such as user account, location, service, and data classification.

This leads to a zero-trust policy where both users and administrators must undergo a certain amount of vetting before they can perform actions, especially actions involving the use of sensitive data.

Preserving the principle of zero trust least privilege access will be achieved by restricting user access rights at each endpoint on the internal network. By using a list of rules deployed on each endpoint, it prevents unauthorized actions including the use of any system tools, programs and data. Groups of endpoints can be tied to a specific set of rules, such as HR and financial systems segmentation, so that only finance staff can access final software and data on a specific set of endpoints. This can work well in conjunction with an existing deployment such as Active Directory as it can reinforce the privileges and access that a user must have to perform certain actions and access data, as well as blocking any prohibited actions (Zhukabayeva et al., 2025).

Finally, setting policies for the value of the data, the zero trust principle will be addressed by using data-driven defence, with data privacy being determined by searching keywords across all the data on the network. Using keyword search, the privacy level can be determined by the number of keyword matches, so that data can be appropriately placed into two classifications: non-confidential or confidential. This can then assist in detecting and preventing any exfiltration attempts using sensitive information, with the classification of files being checked during the authorization of any action involving the use of a particular file or group of files.

The Russian information leakage protection system provided by InfoWatch is very similar in essence. InfoWatch Enterprise Solution includes two main modules - Traffic Monitor and Net Monitor. The first one prevents leaks through e-mail and Internet channels, and the second one prevents leaks through printers and workstation ports. Despite many years of experience and high level of protection, there are still ways to cheat even such an advanced system. The main breach in Enterprise Solution protection is the repeated checking of files for confidentiality, which, under certain conditions, can lead to the distribution of a once confidential file to the status of non-confidential, after which it can be sent to a user who previously had no rights to view the document. The goal of this project is to develop a system to protect against data exfiltration, using the preservation of a confidentiality label throughout the life cycle of a file. (Zhukabayeva et al., 2025)

Our main contribution is as follows: comprehensive overview of the various attacks and the associated methods and technologies they use to steal data; evaluate typical defence mechanisms used to counter or defend against data theft attacks; new method that can effectively resist internal data theft attacks.

Method

The main purpose of this chapter is to explore and provide a complete technical understanding of data exfiltration, including attack methods and defence mechanisms. In addition to examining data theft, more specific areas are explored, including prior knowledge of insider attacks and approaches to countering data exfiltration.

Types of data exfiltration:

Physical

Physical data exfiltration occurs when an attacker can gain physical access to a system or data to conduct an attack, either by initiating a transmission over the Internet, by physically stealing a device or data, or by manually transferring data to an external storage device. Having access to the system opens up many more possibilities, such as installing malware to remotely steal data or even steal the system.

However, even with physical access to the system, this only becomes an effective means of data theft if the attacker can gain access to system data. Consequently, if the system uses any kind of password protection or disc encryption, valid credentials must also be obtained so that the system can be accessed and data can be made available for transmission

Other forms of physical media can also be used by attackers to steal data, such as USB memory sticks and printing out any important documentation. In addition, with the advent of the bring your own device (BYOD) concept, any of the devices brought by employees can pose an immediate security threat as the devices can potentially be stolen, lost or even used as an entry point by attackers (Varsalone, 2024).

Remote

Remote data exfiltration uses a remote connection to allow data to be transferred from a system without physically accessing it. By establishing a remote connection to a system, an attacker can not only extract data from that system, but also perform traffic analysis and monitor/intercept incoming and outgoing communications such as email. HTTP/HTTPS, FTP/SFTP and email are all key examples of data theft channels that can be used for remote data theft (Balisan et al., 2024). For remote data theft techniques to be viable, the attacker uses vulnerabilities or exploits to access the system from outside the network.

This can come from a variety of areas, but generally it can be due to outdated software, outdated security updates, undetected malware, or the use of any currently running services and protocols on the system. While remote exfiltration can provide more freedom for attack methods and data transfer, if the vulnerability an attacker uses to access the system is discovered and remediated, they will no longer be able to access the system and must specify an alternative access method or possible target.

Data exfiltration attack methods

Physical attacks

The first attack method for stealing physical data is to steal the target system. This method requires the attacker to have physical access to both the room where the system is located and the target itself, which may require some form of identification or authorisation before they can gain access. This method can be detected very easily and quickly if not done in advance, especially if the theft occurs during business hours when the device is in constant use.

Devices such as mobile phones and/or tablets are prime targets for this form of data theft as they are much more likely to be identified as a missing device rather than a stolen device, as both are easy to lose, especially if employees participate in BYOD (Balisan et al., 2024). Once stolen, attackers can connect devices to their systems and then use appropriate tools or techniques to compromise the device and then exfiltrate any stored data such as sensitive emails and documents.

USB drives and external hard drives can be easily used for data theft as they just need to be plugged into the target system and then can be used to steal data from the system. While many companies use defence techniques such as disabling USB ports on systems to protect against this form of data leakage, many companies do not and leave themselves open to attack. A McAfee report showed that

removable devices accounted for 31% of all data breaches, showing how effective and viable this method is for exfiltrating data (Chung et al., 2023). Copying data to an external medium will not leave any obvious traces unless the administrator has reviewed the system logs, which further proves that this is an effective attack method. Transferring data from the system to an external medium usually does not require user intervention after startup, so this method can also be left unattended.

Recovery of discarded devices

When companies or consumers modernise their infrastructure or systems, they should ensure that any unneeded devices are properly disposed of. If this is not done, attackers will have the opportunity to access the devices and attempt to extract any data that is still on the systems.

Companies that do not ensure that storage devices such as internal hard drives are cleaned and/or removed from systems before they are disposed of may allow attackers to exfiltrate data with minimal effort. Unless the internal hard drive has been formatted multiple times, data recovery tools can be used on these drives to attempt to recover any residual data still on the drive. Even recovering a small amount of data has the potential for more serious consequences, especially if the recovered data relates to any credentials or clearance information that could be used by an attacker to access sensitive information. Personal devices used by everyday technology consumers also face this threat, as many people simply dispose of their unwanted devices or technology at recycling centres or waste collection points without first checking that all data has been wiped.

Remote attacks

One of the first methods that can be used to steal data remotely is to use websites to exfiltrate data from a system within an organisation to an external network. Websites such as GitHub, Dropbox and Google Drive provide convenient exfiltration channels that attackers can use to upload data, especially as many organisations allow access to these websites. Thus, steganography can be exploited by hiding large amounts of data in images and then deleting those images rather than the data itself. This hides the true nature of the exfiltrated data and makes it easier for the attacker to justify themselves in the event of a claim.

While ransomware used to focus on encrypting the target system and demanding a ransom to decrypt all data on the infected system, data exfiltration is now being used to intimidate victims into paying the ransom. Attackers not only encrypt the target system's data, but also copy large amounts of data from any systems in advance and publish it publicly on the Internet if they do not receive payment. Whether the attacker wants to demand payment or not, this provides an effective method of remotely stealing data, an example of this is Maze, a recent ransomware that has been used to steal and encrypt data on any systems on which it has been installed (Camacho, 2024).

Phishing is another attack method that can be used to steal data remotely, carried out by an attacker who sends multiple emails to different addresses in an organisation to gain access to sensitive information (Varsalone, 2024). The attacker does not need to install anything on the target system, but can use social engineering techniques such as masquerading as a trusted person to coerce the target into divulging sensitive information such as login credentials or personal information. By spoofing the email address of a trusted person, the attacker can easily be deemed trustworthy, meaning that victims will believe they are receiving a legitimate message from a trusted source when in reality they are communicating with the attacker. Targeted phishing is an example of phishing that targets specific individuals or companies, where attackers gather details and personal information about their targets to increase the likelihood of success.

Data encapsulation is another technique that can be used to remotely extract data without detection. Attackers who first encapsulate the target data in a file of a particular type and then proceed to exfiltrate the data from the organization’s internal network to an external destination do this.

File types such as ZIP, RAR and CAB can be used to encapsulate target data in an archive, protecting it from any traffic monitoring services or data loss prevention strategies an organisation may have in

place. Nested zip files can also be used, as many data loss prevention (DLP) systems stop scanning nested zip files after 10-100 archives to avoid zip bomb attacks.

Many companies use a myriad of services and protocols to create a working network infrastructure. Therefore, by examining the protocols used by a company's network and identifying any vulnerabilities, an attacker can remotely exfiltrate data. An example is Domain Name System (DNS) tunnelling, which can be used to steal data. By establishing a command and control channel through DNS, an attacker can encode data from other programmes and protocols into requests and responses.(Brown et al., 2023). However, this requires the attacker to have access to the internal DNS server, and to have an authoritative domain server setup to establish tunnelling from the compromised internal system through the internal DNS server and to the malicious domain server. Without this, any traffic from the compromised system going directly to the malicious authoritative server could potentially be blocked by the network firewall. Insider threats refer to any malicious threats that originate from personnel working within a company, such as employees, contractors, or former employees (Balisan et al., 2024).These threats are difficult to protect against because insiders within a company are the same people who are trusted to handle confidential or otherwise sensitive data and knowledge of the company's network infrastructure.

This means that all forms of insider threats must be identified and appropriate strategies to prevent insider attacks must be clearly defined, as anyone who has access to a company's network can be considered an insider threat (Camacho, 2024).

Types of insider threats:

Compromised insiders

If an insider within a company has compromised their credentials, this can lead to a security breach for anyone who uses those credentials to gain unauthorised access to any sensitive information to which the insider may have access. However, if they do not realise that they have been compromised, this results in attackers gaining continued access to restricted data and potentially compromising more insiders within the company. This type of threat occurs when insiders click on malicious links sent by phishing emails posing as legitimate and trusted sources, causing the insider to inadvertently compromise themselves and their system for the attacker.

Careless insiders

Negligent insiders include any insider who follows security best practices and/or any methods employed by the company to prevent data breaches and compromise, putting themselves and the company at risk. Examples of this type of negligence include leaving a system unlocked when unattended, storing passwords written in the public domain, and using default credentials for services and software, which opens the door for an attacker to gain access to the system and conduct an attack (Huma, 2025).

These types of insiders can easily become prime targets for malicious insiders, especially if their negligence is noticed and monitored by other insiders with malicious intent: 63% of reported incidents in 2024 were caused by insider negligence. (Huma, 2025)

Attackers can use attacks such as social engineering or phishing to coerce this type of insider into divulging key information, which can be particularly effective if the target neglects security concerns and practices.

Malicious insiders

Malicious insiders are people within a company who intentionally want to damage or disrupt the company's network or infrastructure. These may be insiders who will use their authorization or clearance to access data or systems that are otherwise restricted and then launch an attack such as exfiltration and public distribution of sensitive data. Additionally, because these insiders are legitimate users and will understand how the company's systems are structured, it will be easier for

them to identify vulnerabilities that can be exploited. This advantage leads to them having an easier time covering their tracks and hiding any clear evidence that could lead to the discovery of an attack. Malicious insiders can also more severely impact systems because they can identify key systems and data that are targeted. Even an attack as simple as deleting data, rendering a system unusable, or having to suspend normal company operations results in high costs *due to damage and resulting downtime(Camacho, 2024)*

Typically targeted data

The type of organisation and resources available is very important as it can be a determining factor in what data can be available to an attacker to target, and the opportunities associated with what can happen after an attack. This is why many attackers will assess and determine which types of organizations are more valuable targets, taking into account factors such as scale, purpose and infrastructure to select a primary target for attack. An example of this type of targeting would be companies offering financial services, such as banks and other financial institutions. This is because they are often targeted because they are responsible for protecting and managing thousands of records containing sensitive information such as card and account data. Other forms of data, such as personal health information and intellectual property, are prime candidates for attackers because they can be used to blackmail, publish or sell to willing buyers, especially when intellectual property is involved. The main source of concern associated with intellectual property theft is an organization's direct competitors. This can lead to attacks involving organized crime becoming more prevalent as competitors may use various tactics to obtain sensitive information that may allow them to become a leader in their respective field through the use of stolen property.(Huma, 2025).

Approaches to countering data leakage:

Preventive

Preventive measures against data theft include implementing measures or systems that attempt to block any methods that could be used to initiate unauthorized outbound transmission of data from the network.

This can be associated with organizations that use DLP strategies, which include implementing several different systems to monitor and block various methods that could lead to the loss or leakage of any sensitive data. Systems such as IDSs and firewalls are very often used in this way to monitor and inspect any outgoing or incoming traffic. Thus, if an external attacker attempts to access the internal network because they are connecting from an unauthorized device/IP address, they will be blocked from accessing any network resources or systems, regardless of whether they were able to gain physical access. connection to the network (Varsalone, 2024). However. while both IDSs and firewalls can monitor and protect the entire network. antivirus software is used to protect individual systems. which can help identify any infected systems that may have been compromised or targeted by external threats attempting to gain access through attacks. such as phishing or masquerade attacks.

Detective

Detection countermeasures focus on detecting exfiltration attempts but do not prevent them in advance. While preventive countermeasures are more active in defending against exfiltration, detection countermeasures react when unauthorized exfiltration is attempted. This can be centred at two levels, the first of which is the network layer. When implemented at this level, network traffic is monitored and some form of packet inspection is included in the process so that any attempt to exfiltrate sensitive information, whether masked or not, can be inspected and detected before it leaves the network. The second level at which this can be implemented is the host level, which can focus on monitoring the access patterns of users when they are using the system, or any initiated data transfer by monitoring access to the system and any abnormalities in system usage, such as a

user attempting to access multiple file shares to which they do not have access, or a user attempting to copy a large number of files to an external storage device, can then be triggered oops

Investigative

Investigation refers to countermeasures that are used to investigate data leakage incidents after they have occurred in order to investigate and identify key information about the attack to prevent it from happening again. By determining how, when, and who exfiltrated the data, it is possible to not only identify the person responsible for the attack, but also determine what techniques or vulnerabilities were used to carry out the attack. Incident investigation can also help mitigate any effects of an attack that could potentially damage the company or its operations. Any details discovered can be key for lead investigators to determine the potential impact that may occur and from this take the necessary steps to mitigate or minimize the impact it will have (Sobol et al., 2023). Investigative countermeasures can also be useful to other companies and organizations in a large number of different sectors, as victims of a recent attack may disclose key information. Areas such as how the attack was carried out, what vulnerabilities were exploited, and most importantly, recommended security fixes that should be made to prevent this type of attack from happening again. If a victim of a recent exfiltration attack decides to disclose information about the attack, all sensitive security or personnel information should be excluded to ensure confidentiality; otherwise, this may result in additional data leakage (Tari et al., 2023).

Mechanisms to protect against physical data theft:

Networking

Firewall

A firewall is one of the most common methods used to prevent data leakage because it can be used to monitor inbound and outbound connections and allow or deny certain connections based on port, IP address and file type. This means that in a scenario where a malicious insider has physical access to the system and tries to exfiltrate data from the internal network to cloud storage, the firewall will inspect the traffic before it leaves the internal network and determine that it is originating to an IP address or service that is not allowed and will be rejected. (Wei et al., 2024)

User authentication and privileges

Authentication is a network layer mechanism that can be used to counter data theft. It creates a barrier between the attacker and the target system because they must obtain valid credentials to authenticate to the system, whether the attacker has physical access or not. This works well even if the attacker has their credentials, as there is no guarantee that they can access all the data on the system unless they have sufficient permissions or privileges to access anything other than their data. Systems such as Active Directory allow this to be done, whereby privileges can be set for entire groups of users or per user, meaning that authentication and privileges can be fine tuned so that only authorized personnel can access certain programs, files and perform certain actions (Prabhaker et al., 2024)

Physical

Locks

Locks are very important in preventing physical theft as they prevent intruders from stealing devices and taking data outside the premises. Examples include Kensington locks for laptops and computers, which can provide a simple prevention mechanism that prevents any unauthorized personnel from moving systems or stealing them. Whilst some locks can be tampered with, it is unlikely that an intruder would pick the lock as this would lead to a quick detection of the incident, which could lead to the identification of the intruder. As well as being locks for the devices themselves, door locks can prevent unauthorized personnel from entering restricted areas such as server rooms, access to which

could allow an attacker to not only disrupt the network, but also install malware or alter certain network properties to assist them in carrying out future attacks (Zhukabayeva et al., 2025).

Physical authentication and biometrics

Physical authentication is another method that can be used to defend against physical identity theft attacks using authentication methods such as fingerprints, facial recognition and smart card reading that can help provide physical authentication to ensure that only authorized users have access to their systems. This can also be used to support already established authentication on systems in the form of two-factor authentication. When entering a username and password, the user could be prompted to insert an ID smart card or scan a fingerprint to provide another layer of authentication. This type of authentication can make it virtually impossible for attackers to access the system without the correct credentials, biometrics or identification (Camacho, 2024). As mentioned above, the second level of authentication must be tied to the presence of a particular item (the "I have" option), or the user's biometric data (the "I am" option). Each of these methods has its own advantages and disadvantages.

Benefits of authentication using unique items:

- Easy installation and maintenance of the hardware;
- No need for end-users to participate in the creation of the key database;
- Low cost of readout hardware.

Disadvantages of authentication using unique items:

- An item may be stolen or taken away from the user;
- In most cases, specialised equipment is needed to handle the items, which in turn require certain costs for purchase, installation and maintenance;
- It is theoretically possible to make a copy or emulator of an item.

Advantages of a biometric system:

- Reliable and fast authentication: using a fingerprint or iris pattern, electronic analytical devices recognize a person within one or two seconds;
- High level of security: a person's biometric traits are unrepeatable, minimizing recognition errors;
- Biometric data cannot be lost or forgotten;
- Biometric authentication devices are user-friendly and cost-effective (budget-friendly) to operate.

Disadvantages of the biometric system:

- Biometrics cannot be changed in the current database - unlike passwords, they are linked to a specific human individual throughout their lifetime;
- Due to age-related changes, injuries, amputations, and more, the reference comparison models that are stored in the memory of electronic computing devices need to be constantly updated;
- To create biometrics samples, special readers are needed;
- Biometric characteristics cannot be kept secret, so skilled attackers can forge fingerprint or palm print samples;
- High cost of equipment for quality inspection.

Companies or organizations that deploy enterprise-grade devices for their employees to work while travelling and at home greatly benefit from this form of authentication, as it ensures that even if the device is stolen or lost, you can be sure that no one will have access to the system or any data on it.

Protection mechanism against physical data theft

Figure 1 shows the proposed physical data theft protection mechanism. The main design feature of this solution is that the control server is used not only for storage, but also for creating both digital fingerprints and rules used for any endpoints in the network. Only network administrators will be allowed access to the server, and only one set of valid credentials will be used to access the server. This helps to minimise the number of credentials that can be targeted by an attacker who may attempt to steal or obtain credentials to gain unauthorised access to the server (Camacho, 2024).

Algorithm 1 summarises the performance of the proposed physical data exfiltration protection mechanism.

Every file placed on the system or network will be checked to determine if it is considered confidential or non-confidential, looking across all directories and repositories to identify as many items as possible. RegEx and keyword search, are some of the most common techniques used to identify sensitive data, with RegEx focusing on the expressions and context of the data, while keyword search identifies specific words that have been identified as representing sensitive information (Wei et al., 2024).

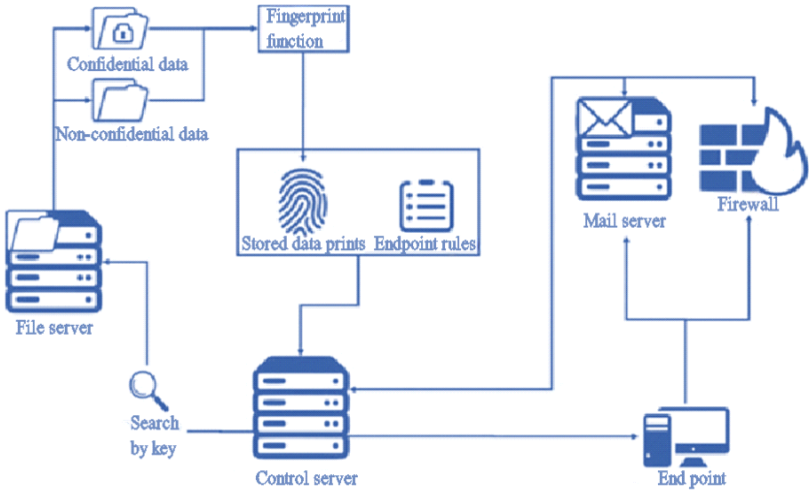

Figure 1. Defence mechanism against identity theft

Algorithm 1. Protection mechanism against exfiltration of physical data

Require: Action(A), File (F), Fingerprint(Fp)

Ensure: Permission (Deny or Allow Request)

While A∈ Rulelist then

If A status = 'Allow'then

Fp = SHA256(F)

If Fp∈ FingerprintList then If Fp is confidential then

OUTPUT "Authorisation is required to use"

Request = F + A

Verify Request

If Verify returns true

OUTPUT "Action is allowed using " + F

```
        Log Action, File, User, System ID, Date
        Permit Request
    Else
        OUTPUT "Action is not allowed using " + F
        Log Action, File, User, System ID, Date
        Deny Request
       Else if Fp is not confidential then
       Permit Request
       Log Action, File, User, System ID, Date
End                                                                    if
Else if Fp∈/ FingerprintList then
      KeyWordSearch(F)
      OUTPUT "File does not match any on system, try again
later."
      Deny Request
     Else if Fp not present then
       Log Action, File, User, System ID, Date
       Permit Request
     End if
   Else if A status = "Deny" then
     OUTPUT "Action is not allowed"
     Log Action, File, User, System ID, Date
     Deny Request
End if
End While
OUTPUT "Action is not allowed"
Log Action, Unknown Tag*, File, User, System ID, Date
Send Log
Deny Request
```

This described mechanism utilises keyword search to identify any sensitive information, with a management server used to scan all data hosted on the network to identify and place each file in the "confidential" or "non-confidential" category (Balisan et al., 2024)

The server can be configured with a keyword search list where administrators can add or modify any existing lists already contained within the system. This makes keyword search much more versatile as it can be modified to meet the needs of different organisations, regardless of their function or the type of data they are working with.

Figure 2 shows that once the server has determined whether a file is confidential or non-confidential, the result is stored until the corresponding digital fingerprint file is created. Once the fingerprint is created, the corresponding result is placed in the digital fingerprint record. This means that when performing a confidentiality search, the server can check the corresponding digital fingerprint record and quickly determine whether or not it has been marked confidential.

Since many of the threats from malicious insiders involve physical threats associated with using internal systems to perform malicious activities, there must be a level of protection for each endpoint to ensure that physical threats can be dealt with while the mechanism resides on the server at the network layer. Endpoint access rules enable this level of protection by deploying a list of rules on the network and ensuring that all systems on the network must follow a set number of rules to prevent unauthorised activity (Figure 3).

Using the principle of zero trust with least privilege ensures that users are granted a minimum number of permissions to ensure that they can use their systems to perform any necessary tasks, ensuring that any other actions are prohibited.

This may include prohibiting the use of inbuilt system tools reserved for administrators, such as system settings, terminal or command line, which can be accessed by an attacker, potentially leading to data theft (Varsalone, 2024).

It can also be used to prohibit actions such as transferring and copying data from certain directories and preventing unauthorised users from accessing software outside their respective department, for example, allowing only users who work in the finance department to access payroll software (Huma, 2025).

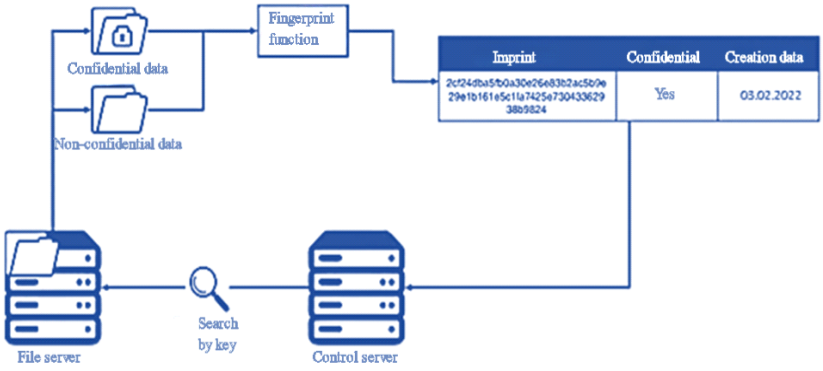

Figure 2. Checking the digital footprint

If an action request is sent to the management server, but it cannot find the appropriate rule for the action received, it will immediately reject the action. In addition to rejecting the request, a log of the request will be created, but it will contain a special unknown tag which will then be sent to the administrators for review. This is to alert the administrators of any action attempts being made that have not yet been accounted for, such as attempts to run newly installed programmes. Digital fingerprinting is the process of taking a file, such as a text file or document, and converting it into a short string of bits that can be used to uniquely identify the original file. This results in the fingerprint being used to identify different files rather than standard file names or file contents, which helps to reduce the amount of sensitive data sent over the network.

One of the problems inherent in this method is the collision that can occur when two unique files are matched and result in the same data footprint.

To maintain unique fingerprints for different files, the probability of collision must be minimised to avoid conflicting digital fingerprints. In order to create a robust and secure fingerprinting function as well as password matching verification for user authentication that could be used as part of this

security mechanism, it was decided to choose a function that generates pseudo-random character sequences as the digital fingerprint function (Mohamed, 2024).

The decision was based on the high probability of the threats being realised:

- This system is thought of to protect against data leakage through insiders. Since they are part of the employed staff, the possibility of an insider accessing the endpoints of other users in conjunction with the data channels cannot be avoided. If an attacker catches the password in unencrypted form through channel sniffing and steals the key card of a certain user, he can gain unauthorised access to the data. Accordingly, the password should be changed already at the endpoint for further transmission for checking its correctness (login-password correspondence). Encryption is not suitable in this case because of the possible reversibility of the process by decryption when the key is found, or by decryption using software.
- If a collision occurs during the creation of a data fingerprint, three scenarios are possible when requesting its processing, depending on the software implementation of the management server:
- There will be a new file in the system, and the permission check of the requested file will be compared exactly with its parameters.
- The old file will be in the system, and, as in the first point, the check of rights to use the requested file will be compared exactly with its parameters.
- The system will compare the access rights to both documents and, if there is a discrepancy in any of them, will refuse the request.

Either of these options may violate the availability state of the file, and the first two paths may also violate the privacy state if one of the files is not allowed to be viewed by a particular user.

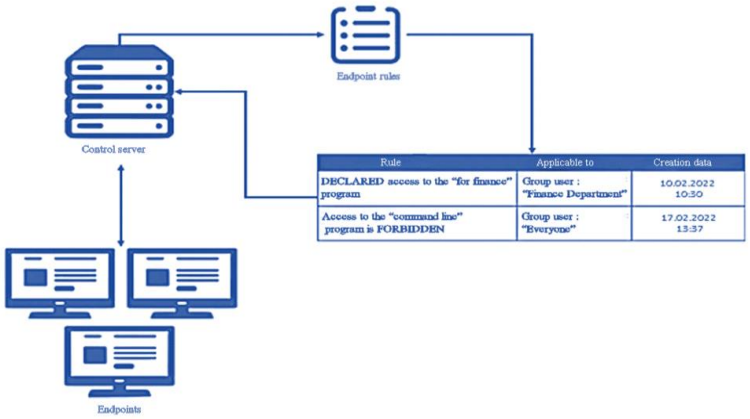

Figure 3. Access rules

Because all requirements were met, it was decided to use the SHA-256 hashing function for both purposes.

With SHA-256, the resulting fixed-size data fingerprints will be 256 bits long, and due to the avalanche effect, even a small change to the original file will create a new hash, helping to avoid any potential fingerprint collisions. Additionally, because it is a one-way function, it becomes much more difficult for any attacker, whether an external attacker or a malicious insider, to decrypt the referenced information from the hash alone. In addition to the above advantages, SHA-256 has all the required properties to be considered a pseudorandom sequence generator, namely:

- Efficiency – speed of the algorithm and low memory costs (high speed of operation will not create large additional delays in the execution of various functions of programmes, which in turn meets the requirement to the security system "ease of use";

- Reproducibility - the ability to replay a previously generated sequence of numbers any number of times (necessary to verify that the entered password matches the data stored in the servers);
- Portability – the same functioning on different equipment and operating systems (allows not to rewrite the program code for each customer's operating conditions);

This is a slower method compared to other digital fingerprinting methods such as Rabin's digital fingerprinting algorithm, which is designed specifically for the task of matching digital fingerprints and analysing collision probability. However, Rabin's algorithm is not secure against attackers because the use of an internal key during fingerprint creation means that if an attacker is able to find and obtain the key, they will be able to modify files without causing the digital fingerprints to change. (Wei et al., 2024)

In this mechanism, digital fingerprints are created and then stored on the management server. This is so that when a user on the endpoint requests to copy, modify or send a file, a request will be sent to the management server to check whether the file involved in the request is classified as confidential or not. By examining the request and matching the hash function to the file requested by the user, it can then determine if the file is confidential and either allow or deny the user to perform the action.

Although this mechanism is network-based, it is still relevant to physical layer protection as it can protect against physical layer threats such as unauthorised access or use of the system and sensitive data. Using a list of endpoint rules, it is possible to create a universal defence against malicious or unauthorised use of any systems on the internal network, with each system following a set of rules with only permitted actions being performed. An example would be an insider attacker attempting to transfer data from the network to an external storage device, and the action would be rejected by the management server because it is not on the list of allowed actions (Brown et al., 2023). Protection at the physical level is also enhanced by network identification of sensitive data, where all network files and data are subjected to a keyword search to determine the level of confidentiality in each file. This means that even if an action is authorized by the management server, such as sending an email, if the relevant file(s) used are marked as confidential, the action will be rejected. This helps prevent malicious insiders from using allowable actions to exfiltrated sensitive data because both the action and the file are checked.

To work harmoniously across different network infrastructures, the mechanism must be able to work seamlessly with any pre-existing deployments that are used to prevent data leakage. Generally, both firewalls and application-based techniques are used to prevent data theft attacks, with firewalls focusing on monitoring and inspecting all communications, while application-based measures are used to prevent personnel from performing any prohibited actions in any applications. (Prabhaker et al., 2024)

Because application-based security measures are tied to a single application, they can help protect applications from malicious or prohibited use and ensure that only personnel with appropriate privileges can perform certain tasks. However, this can still leave an endpoint at risk of malicious use outside of any application. Endpoint rules complement application-based measures by ensuring that restrictions and protections are in place to stop not only prohibited actions within applications, but also in the system as a whole. By deploying a list of rules to all network devices, it ensures that users are granted the least privilege necessary to mitigate any attempted malicious use of the system without hindering the user's ability to perform tasks effectively. Figures 4.A and 4.B show the rules for endpoints and the measures for applications.

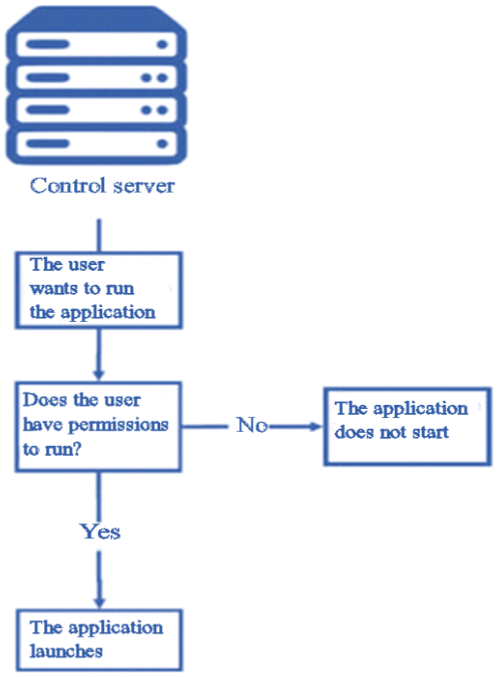

Figure 4A. Rules for endpoints

Since a typical firewall focuses on inspecting traffic as it leaves and enters the network, it typically has its own established set of rules allowing and denying different types of traffic, ports, and addresses. This can work well with the sensitive data identification method, as a confidential or non-confidential tag can be used to initially verify whether the data used in the action contains any sensitive information.

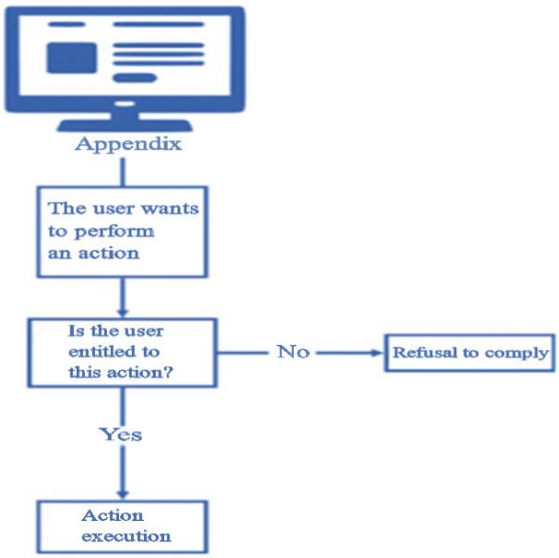

Figure 4B. Rules for endpoints

Then, after verifying that there is no sensitive information or that sensitive information is authorized, the firewall should verify that traffic is allowed to leave the network by checking properties such as destination and source address, port and application (Balisan et al., 2024). This results in an additional layer of security, because even though both the endpoint rule list allows the action and the confidential identification allows the relevant data to be used, if the firewall identifies that traffic is

not allowed to leave the network after it has been inspected, it will be denied. Figures 5.A and 5.B show the confidential data list and the firewall deployment.

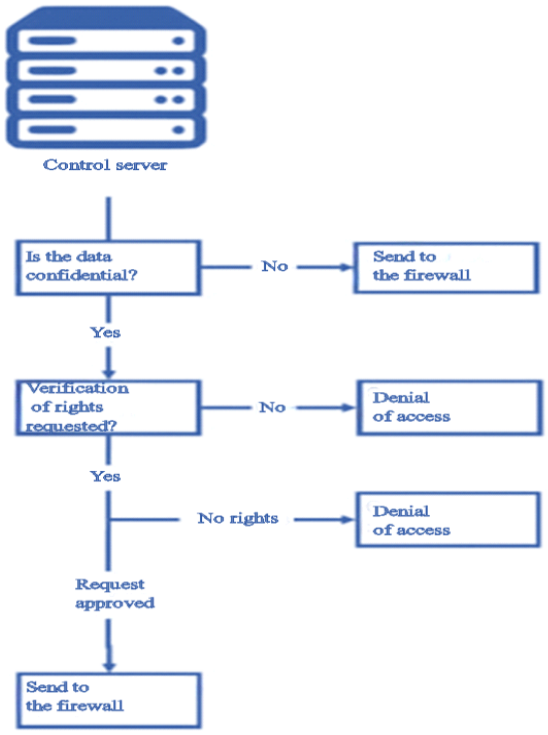

Figure 5.A. List of sensitive data and firewall deployment

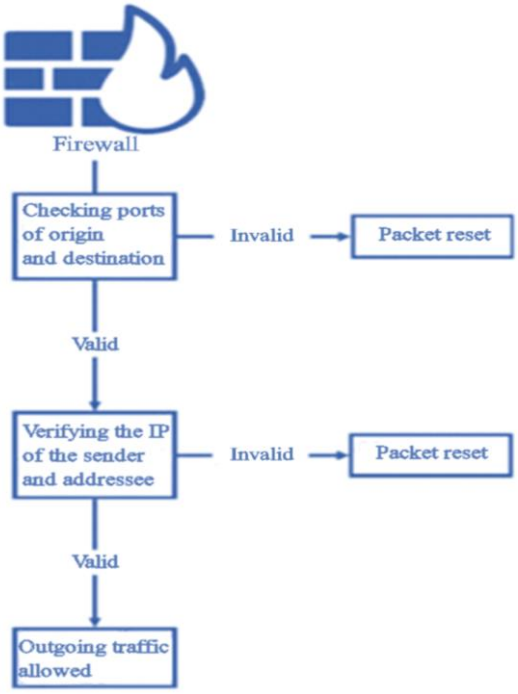

Figure 5.B. List of sensitive data and firewall deployment

Results and Discussion

To test the proposed mechanism and its implementation in an already established network, a simulation tool will be used to create a simulated network containing various systems and

components that can be found in the business network infrastructure. A simulated network will be used as this helps to more accurately depict the implementation process as well as demonstrate the security capabilities of the mechanism once implemented. Graphical Network Simulator (GNS) was chosen to create the simulated network infrastructure as it has a large number of tools and options that can be used to create an accurate representation of the actual network in which the mechanism will be implemented.

Figure 6 shows a modelled network layout with several different areas that have been physically and logically isolated according to business operations sector and system usage. This includes the HR, financial and administrative parts of the network, which are kept separate to improve network scalability and reduce potential risks from placing administrative systems in parts of the network that could potentially be compromised or hacked as a result of accidental or deliberate misuse. The server part of the network is where all the key systems of the network are located, including the file, application, email, and management server that the engine needs to function (Prabhaker et al., 2024). This layout was chosen because it provides a simple yet effective design that not only provides security but also ensures that different and unconnected areas of the network are isolated from each other. However, it can also provide consistent performance for all users, ensuring that the number of points between the user endpoint and key systems is minimised to reduce potential bottlenecks and increase throughput.

For the mechanism to be integrated into the network, the management server must be located in the server portion of the network, connecting it to all other servers on the network that provide key functions for the network and its operation. This ensures that all key systems are located together in one "zone" and physically isolated from other areas of the network, helping to mitigate any attempts to maliciously exploit or disrupt the network.

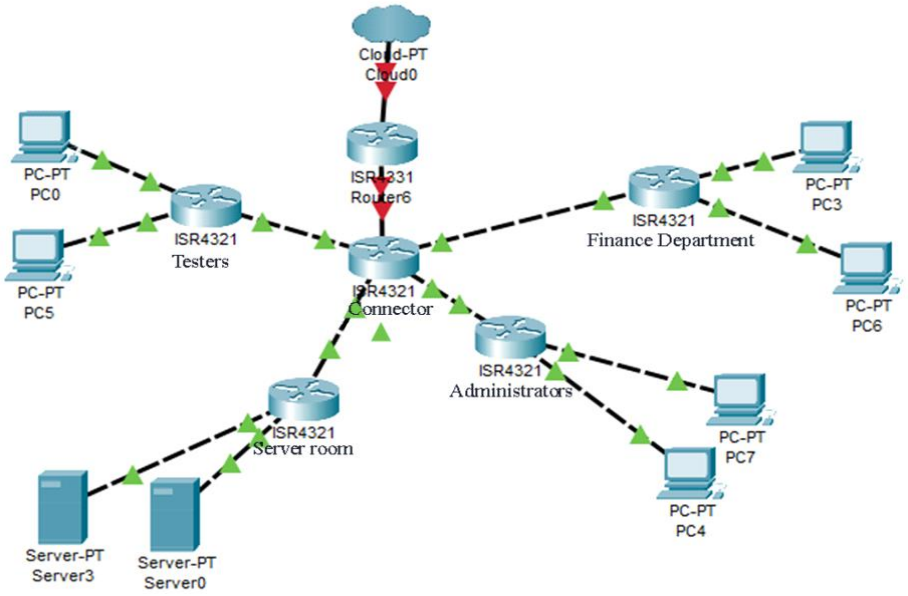

Figure 6. Network diagram with different isolated areas

Threat Scenario

The scenario to be used in conjunction with the simulated network focuses on a malicious insider in the finance department of a company who has decided to try to exfiltrate some sensitive data in order to sell it to the company's competitors. Since the malicious insider works in the finance department, it means that he or she immediately has access to a large amount of sensitive information regarding all the expenses and projected profits of the company, which is the main target for data theft attacks (Varsalone, 2024).To minimise the risk and the chance of being discovered by other employees and administrators, the insider decided to extract only one file by transferring it to a USB drive, which the

insider was able to conceal and bring onto the premises. The company where the malicious insider works has a strict policy regarding the use of removable media and prohibits its use by all employees. Using this scenario will allow the mechanism's approach to be demonstrated in detail on a step-by-step basis, showing how it is able to mitigate physical data theft attacks by being a network-based mechanism. Figure 7 demonstrates the threat scenario.

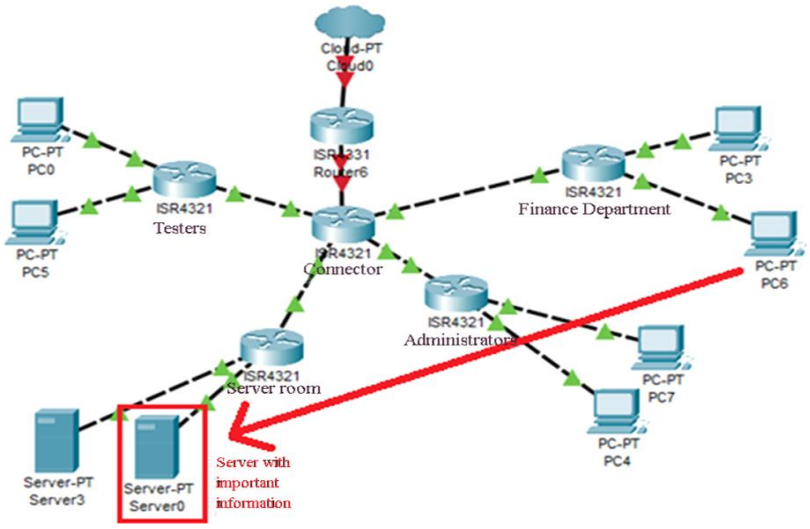

Figure 7. Threat scenario

Scenario implementation and analysis

To begin, the malicious insider will access a file server on the network and identify a file that he/she wants to transfer from the network to removable storage. Because he/she works in the finance department, was able to find a spreadsheet containing a large amount of information about the company's current expenses and projected profits for the fiscal year. After selecting a suitable target, the attacker will attempt to start the exfiltration process by inserting his removable storage device into the USB port of his system. Once this is done, the mechanism will take immediate possession and begin first by determining the action just taken, which in this case is that the device was inserted into the USB port. Once the mechanism has identified the action taken, it will immediately stop the relevant process (e.g. loading the necessary drivers to use the device) and generate an action request that will be sent to the management server.

When creating an action request, the mechanism will use several key pieces of information including the IP address, user account, and workstation from which the action request originated in order to maintain a consistent level of non-repudiation no matter where or when the action is requested. As you can see in Figure 8, the action request was created using these key elements as well as identifying the specific action requested, which is the download of the driver for the external media. It is important to note that if an attacker were to perform an action that involved the use of a specific file, an additional element would be attached to the request that would reference the file or files being used. This is done so that the management server can also check whether the file or files used contain any sensitive data, even if the action is allowed. (Topilin et al., 2025)

When an action request is successfully created, it is immediately sent to the management server, which checks the action request and identifies what action is contained in the request. Once it identifies the action, it will attempt to find the corresponding action in its list of endpoint rules, which is used to determine what actions can and cannot be performed by users on their systems . (Malik et al., 2024)

When the appropriate rule is found, the Management Server will determine whether the action was allowed or denied and, depending on the result, send a response to the appropriate endpoint informing it whether to resume or cancel the action. In this scenario, because the company has a strict policy regarding the use of external storage devices, the Management Server will find the appropriate rule and determine that the action is prohibited.

Once an action has been identified as denied, an action response is created that includes the requested action, the result (allowed or denied), and a date and time stamp, as shown in Figure 9. This is logged by the management server and then sent back to the malicious insider endpoint, where it checks the response and determines whether the requested action is allowed or denied to proceed.

Since the response contains a failure token, the endpoint will immediately stop initialising the device, leaving the external storage device unusable. This prevents a physical data theft attack because the attacker can no longer use the external storage device to transfer sensitive data from the internal network.

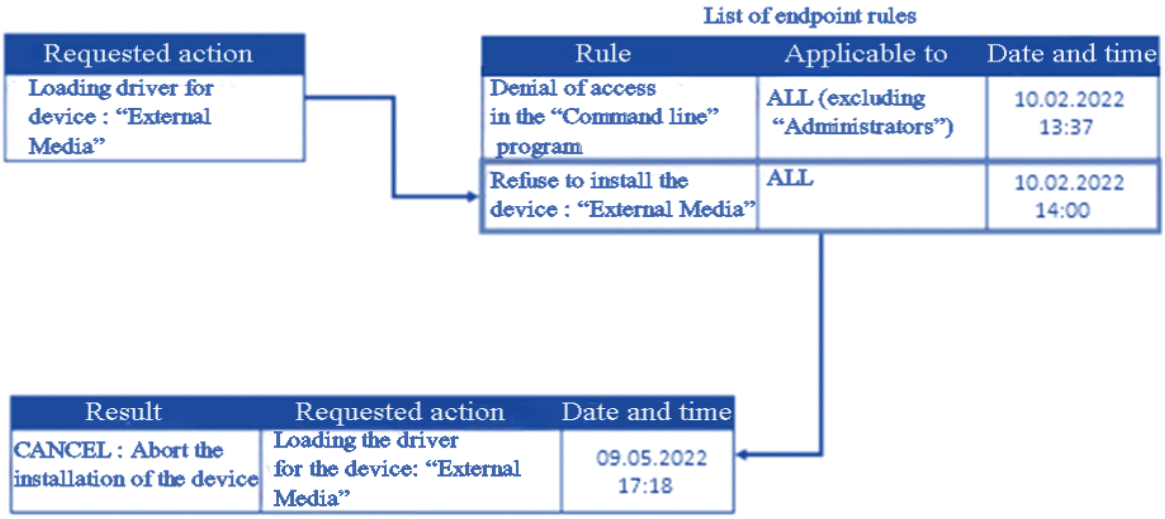

Figure 9: Rejected action response

Scenarios

There are three scenarios used to evaluate the proposed mechanism and its ability to mitigate physical data theft attacks. These scenarios were chosen because they all use an attack method that tests a specific component of the mechanism, including the identification of sensitive data and a list of rules for endpoints. Each scenario uses a different attack method; however, they are all based on the idea that a malicious insider would conduct attacks within the intended network intending to steal sensitive information.

To ensure that the scenarios used for the evaluation are relevant and can properly test the effectiveness of the mechanism, the objectives of each scenario are directly related to a specific design goal that was established earlier in the development process. By creating each scenario and relating it to the design goal, it ensures that each scenario correctly evaluates a key area of the mechanism; therefore, it is possible to determine exactly how the mechanism can work and how it can meet a set of objectives.

Scenario 1

The first scenario focuses on an attacker attempting to use an external storage device to physically steal data from an internal network. Since many external storage devices, such as USB, are very small, they are easy to not only transport but also to hide from others. This scenario would begin with an attacker plugging an external storage device into an available USB port on their system. Once this is done, the mechanism will immediately take effect due to the list of endpoint rules deployed on the

network. First, the system will stop initializing the storage device and send an action request to the Management Server to check whether initializing the new USB device is allowed. The Management Server will then check its list of endpoint rules to determine if there is an appropriate entry for such an action, and if there is, it will check to see if it should be allowed or denied. As the use of external storage devices is prohibited on any systems on the internal network, the Management Server will identify the action as rejected and send a response back to the endpoint warning it that it cannot proceed with the action and will cancel the initialization of the device, leaving it unusable. A log will also be created and stored on the management server. Which will contain key information about the recent request, including the action that was denied, the name of the user who sent the request and the endpoint from which it was sent, as well as the time and date.

Since this is a mechanism that follows the principles of zero trust, any action that does not have a corresponding entry in the guidelines will be immediately rejected. This is to ensure that malicious insiders cannot maliciously use any programmes, functions or other elements before they are included in the list of rules.

Scenario 2

The second scenario focuses on an attacker attempting to use a cloud storage service to exfiltrate data by directly downloading targeted data from the internal network. As cloud storage services such as Dropbox, Google Drive and OneDrive become increasingly popular for storing information, this makes them ideal for data exfiltration. However, while they are widely used by regular consumers, depending on the security policies in place at a business, they can be either restricted or completely banned. To begin, an insider attacker will first identify a cloud storage provider to exfiltrate the data, and determine what data will be targeted for exfiltration. Once an insider has accessed the cloud storage website and selected a piece of sensitive information to exfiltrate, he or she can start the exfiltration process by attempting to upload file. Before the file can be uploaded, the mechanism is initiated by sending an action request to the management server. Since the action to be performed involves the use of a file, before sending it to the management server, it hashes the file using SHA-256 to create a file fingerprint. Once the fingerprint is created, a request is sent. It will first request a list of management server endpoint rules to determine if copying the file is allowed, and it is. However, since the request also includes a file fingerprint, the Management Server will take the file fingerprint and attempt to locate the corresponding fingerprint in its list of stored fingerprints to determine whether it is confidential or non-confidential. The management server will locate the corresponding fingerprint record and, by checking the confidential tag assigned to the record, will be able to determine that it is a confidential file and therefore cannot be used without verification. It will then prompt the insider that verification is required to perform this action using this file, and give the insider the option to either request verification or cancel the action.

Validation is used in this step because in cases where sensitive data is to be used legitimately, such as for emailing to a trusted contact, it can be validated and then sent. A verification request is similar to an action request; however, it is sent to the administrator for verification so that they can determine if it is a legitimate use of sensitive data. In this case, even if an insider decides to send a validation request in the hope that they will be allowed to proceed because the file being used is confidential and the use of cloud storage is prohibited, they will be denied.

If the file is not confidential, it will be allowed to go to the firewall, which will be responsible for checking other information such as source and destination IP addresses and ports, as well as packet inspection.

Scenario 3

This scenario focuses on a malicious insider who was able to obtain a malicious tool that can be used to install a backdoor into the internal network so that important sensitive data can be accessed remotely.

The insider has saved the malicious tool in their downloads folder; however, the attacker must run the tool to begin the backdoor installation process. Once the insider attempts to run the tool, the mechanism will initialise and begin by stopping the file from running and creating an action request. The action request will be sent to the management server where it will be used to determine if the python file is allowed to run. Since this is a python-based tool that uses the ".py" extension, the management server can check the list of endpoint rules for a rule that pertains to files with that specific file extension. If the management server can find a rule with the appropriate file extension, it then determines whether to allow or deny the execution of an action to run the file. However, according to the established assumptions mentioned earlier, the management server does not currently have a rule mandating the use of python files. This causes the management server to be unable to detect the associated rule and therefore immediately rejects the request.

The Management Server will also create a log for this request; however, it will contain an unknown tag in the request because it was unable to find the associated rule for the request and therefore had to immediately deny the request. Once the administrator receives the log of unknown requests, he or she can review it and determine if the new rule should be added to the list of endpoint rules and if the associated action should be denied or allowed. This process of reviewing unknown requests is very useful because it helps administrators not only detect malicious activity and create new rules to prevent it, but also identify problems where users are unable to perform legitimate tasks, such as running a programme because a rule is missing or misconfigured. (Chung et al., 2023)

Analysing the results

Examination of the evaluation results shows that this mechanism is capable of mitigating several different variations of physical data theft attacks by being located entirely at the network layer. Physical security techniques such as locks, covers, and fences can be used to mitigate physical data theft attacks by restricting access to systems; however, they cannot account for any malicious use that occurs within the system itself and can be stopped due to unauthorised access.

This is why the network defence method offers the best of both worlds and can be used not only to prevent malicious network activity, but can also be used to prevent successful malicious physical activity. Having a dedicated management server to host all the major components of the mechanism, as well as endpoint rule lists and file fingerprints, limits the number of ways an attacker can potentially access and disrupt the mechanism. This helps to reinforce the zero trust principle as it assumes that anyone can pose a threat, so by placing all components and data at a single point, it offers a secure method of operation without compromising performance. Chung et al., 2023). This results in a single point of failure, but this is a small price to pay compared to the large number of potential breaches that can occur if the mechanism is distributed across the network.

The versatility of the mechanism is another advantage, as it can be configured in a variety of ways to fit the environment in which the deployment is being performed. Endpoint rules can be configured in any way the administrator sees fit and can be used to separate programme access, configuration and other activities according to the different user groups on the network, so users can access only what they need, be able to perform the required tasks and nothing else (Chung et al., 2023).

Keyword scanning can also be configured so that the determination of whether files contain sensitive data or not can be changed depending on which keywords are labelled as sensitive (Mohamed, 2024). This is important because depending on the type of business in which the engine is deployed, sensitive information may vary, and with it the keywords associated with what is confidential and what is non-confidential.

Logging is a very important factor in helping to not only troubleshoot issues, but also to identify any signs of attempted malicious activity. Logging all allowed and denied requests is extremely useful as it helps administrators track user activity and verify that both endpoints and their users are

complying with established rules and policies. Additionally, the use of unknown requests helps administrators identify any potential loopholes or areas that are being exploited, with all actions not related to the rule explicitly prohibited until the rule is created. This is particularly useful because there are times when vulnerabilities are not properly accounted for and could potentially be exploited by attackers.

One of the main limitations is the actual feasibility of its deployment in the network. As this is a proposal without a developed prototype for evaluation, no formal testing of any real equipment has been carried out, meaning that the feasibility of the mechanism can only be assessed and not properly measured.

While there may be an evaluation of the mechanism in a general, hypothetical sense, it does not take into account any network related issues, implementation issues, or the overall performance that would be seen when implemented on real hardware.

Another limitation is that it has a single point of failure. With all components and data hosted at a single point, if it is compromised, an attacker will have access to the entire mechanism and all information about its operation, including stored endpoint rules and file fingerprints. This can lead to the attacker manipulating the mechanism in their favour, which can lead to the theft of large amounts of data and damage to the business. Additionally, if the management server goes offline, the entire network should be taken offline to mitigate any attempts to exfiltrate sensitive data when the server is offline. This could potentially result in significant downtime until the management server is up and running again, resulting in significant damage due to the network outage. An improvement that could be made to this mechanism is to combine firewall packet inspection with file fingerprints stored on the management server. Using the file fingerprints, the firewall could not only inspect the packet and the source and destination addresses, but also check to see if the data attached to the packet contains sensitive data. This would eliminate the need for the management server to perform file fingerprinting, which would help reduce the number of checks performed because the packet and fingerprint checks would be combined into one. This could potentially help prevent attackers from hiding sensitive data within allowed traffic types, as the data would still be verified even if the protocol associated with the packet is allowed. (Tari et al., 2023)

Conclusion

Because of the study, all the set objectives were achieved: comprehensive review of the various attacks and the associated methods and technologies they use to steal data is produced; typical defence mechanisms used to counter and defend against data theft attacks were assessed; new method that can effectively resist internal data theft attacks has been developed.

This system was developed to increase the level of security of confidential data stored and created in the protected system. Due to the emphasis on security, the system has some disadvantages associated with the inconvenience of using it outside the organisation, namely the inability to work without a connection to the company's local network or the Internet (offline mode).

This inconvenience is not present in the system cited earlier, namely InfoWatch Enterprise Solution, but there is a security gap related to this, namely the inability to track changes in confidential documents by replacing certain words and phrases, after which the document will no longer have the status of confidential and can be sent to persons who are not authorised to view it. Due to the impossibility of full control over the user's offline actions, it was decided to prohibit any actions without the permission of the server, the request to which cannot be sent without connection to the company's local network or the Internet. The goal of this work was achieved by placing a one-time privacy label on files. Subsequent checks will only update the list of private data contained in the file and even if there is no more such information left in the file, the label can only be removed manually by the server operators. Non-confidential data is also constantly checked for the occurrence of words

and phrases from the list, which guarantees security in any attempt to write sensitive data to a file not intended for this purpose.

References

Balisan H., Egho-Promise E., Lyada E., Aina F. (2024). Towards improved threat mitigation in digital environments: a comprehensive framework for cybersecurity enhancement. International Journal of Research -Granthaalayah

Brown, R. & Roberts S. J. (2023)Intelligence-Driven Incident Response. O'Reilly Media, Inc.

Camacho N. (2024). The role of ai in cybersecurity: addressing threats in the digital age. JAIGS, 3(1)

Chung M., Yang Y., Wang L. Cento,G., Jerath K, Raman A., Lie D., Chignell M. (2023). Implementing Data Exfiltration Defense in Situ: A Survey of Countermeasures and Human Involvement. ACM Computing Surveys. 55.

Huma, Z. (2025)Hacking the hackers: Offensive security strategies in modern cyber defense Journal of Data and Digital Innovation (JDDI)

Malik, J., Muthalagu, R., Pawar, P. M. (2024) A systematic review of adversarial machine learning attacks, defensive controls, and technologies IEEE Access

Mohammed, B. (2024). The impact of artificial intelligence on cyberspace security and market dynamics. Brazilian Journal of Technology

Prabhaker N., Bopche, G. S., Arock, M.(2024) Data-level cyber deception in cloud of things: Prospects, issues, and challenges," in Cloud of Things. Chapman and Hall/CRC,

Sobol B.V., Soloviev A.N., Vasiliev P.V., Lyapin A.A. (2023) Modeling of Ultrasonic Flaw Detection Processes in the Task of Searching and Visualizing Internal Defects in Assemblies and Structures. Advanced Engineering Research

Tari, Z., Sohrabi, N., Samadi, Y., Suaboot, J.(2023).Data Exfiltration threats and prevention techniques: Machine Learning and memory-based data security. John Wiley & Sons,

Topilin I.V., Han M., Feofilova A.A., Beskopylny N.A. (2025)Comparative Analysis of Neural Network and Machine Learning Models for Short-Term Traffic Flow Prediction on Shenzhen Expressway. Advanced Engineering Research

Varsalone, J. & Haller, C.(2024) The Hack is Back: Techniques to Beat Hackers at Their Own Games. CRC Press

Wei, D., Sun, W. , Zou, X., Ma, D. , Xu, H., Chen, P., Yang, C., Chen, M., Li, H. (2024)An anomaly detection model for multivariate time series with anomaly perception PeerJ Computer Science

Zhukabayeva, R., Zholshiyeva L., Karabayev, N. ,Khan, S., Alnazzawi, N. (2025)Cybersecurity solutions for industrial internet of things–edge computing integration: Challenges, threats, and future directions Sensors, vol. 25, no. 1