Jaakko Mansikka

DATA LOSS PREVENTION

For Securing Enterprise Data Integrity

Thesis
CENTRIA UNIVERSITY OF APPLIED SCIENCES
Degree Programme
May 2023

centria
University of Applied Sciences

ABSTRACT

Centria University of Applied Sciences	**Date** October 2023	**Author** Jaakko Mansikka
Degree programme Bachelor of Engineering, Information Technology		
Name of thesis DATA LOSS PREVENTION. For Securing Enterprise Data Integrity		
Centria supervisor Jari Isohanni		**Pages** 44
Instructor representing commissioning institution or company Recion Oy		

This final thesis was commissioned by Recion Oy, a regional high-pressure piping company, in order to address a mutually beneficial subject that enhances both the organization's cybersecurity and my personal expertise.

The primary focus of this study was to introduce the concept of Data Loss Prevention systems by initially exploring the theoretical foundations and operational mechanisms of such systems. This was followed by a concise market analysis, offering insights to prospective buyers. Delving further into the realm of cybersecurity, the most significant threat, namely human factors, was discussed, along with strategies to mitigate the risks associated with personnel. Moreover, the importance of optimizing operational security and contemplating a comprehensive approach that encompasses multiple systems in achieving an adequate cybersecurity standard was emphasized.

In the concluding section, the implemented Data Loss Prevention system is presented, accompanied by a brief overview of the enabling process and recommendations for those considering a similar implementation, including potential pitfalls to avoid.

Key words
authentication, azure, cybersecurity, data loss prevention, encryption, market analysis

CONCEPT DEFINITIONS

DLP

(Data Loss Prevention) is a security solution that identifies and helps prevent unsafe or inappropriate sharing, transfer, or use of sensitive data.

AIP

(Azure Information Protection) is part of Microsoft Purview Information Protection (formerly Microsoft Information Protection or MIP).

API

(Application Programming Interface) is a way for two or more computer programs to communicate with each other.

LDAP

(Lightweight Directory Access Protocol) is a directory service protocol that runs directly over the TCP/IP stack.

IDP

(Identity proofing) can authenticate any entity connected to a network or a system, including computers and other devices.

SAML

(Security Assertion Markup Language) is a standardized way to tell external applications and services that a user is who they say they are.

XML

(Extensible Markup Language) is a markup language and file format for storing, transmitting, and reconstructing arbitrary data.

CAGR

(Compound Annual Growth Rate) is the mean annual growth rate of an investment over a specified period longer than one year.

IDS

(Intrusion Detection System) is an application that monitors network traffic and searches for known threats and suspicious or malicious activity.

IPS

(Intrusion Prevention System) is a form of network security that works to detect and prevent identified threats.

NBAD

(Network Behaviour Anomaly Detection) is a security technique that provides network security threat detection.

SIEM

(Security Information and Event Management) is a solution that helps organizations detect, analyse, and respond to security threats before they harm business operations.

IAM

(Identity Access Management) is a security and business discipline that includes multiple technologies and business processes to help the right people or machines to access the right assets at the right time.

PAM

(Privileged Access Management) is the combination of tools and technology used to secure, control and monitor access to an organization's critical information and resources.

RMS

(Azure Rights management) is the cloud-based protection technology used by Azure Information Protection. Azure RMS helps to protect files and emails across multiple devices, including phones, tablets, and PCs by using encryption, identity, and authorization policies.

AD

(Active directory) It runs on Windows Server and enables administrators to manage permissions and access to network resources. Active Directory stores data as objects.

FIGURES

TABLES

1 INTRODUCTION

Data Loss Prevention (DLP) software is an important tool for organizations to protect sensitive information from leaving the organization. This software focuses on monitoring both incoming and outgoing data and preventing unauthorized copying. DLP software is classified into two categories: Enterprise DLP solutions and Integrated DLP solutions. The former monitors network traffic and emails while the latter includes data classification tools, content management platforms, cloud access security brokers, and data discovery tools.

There are various techniques used by DLP software to detect policy violations, including the Database Fingerprinting technique, Regular Expressions, exact fingerprint matching, partial Document Matching, concept/lexicon approach, and machine learning or statistical strategies. Data classification is an important aspect of data loss prevention as it helps organizations identify the types and location of sensitive data. Authentication is also an important part of data security, where the operating system stores credentials securely for future use.

This thesis report aims to discuss the various techniques and strategies used by DLP software to prevent data loss and the importance of data classification and authentication in data security. Recion Oy produces a lot of information every day, for example in the form of electronic documents. Documents are processed in different places such as network disks, Microsoft Sharepoint and in file storage services. The organization estimated that classifying the data would help employees to separate public and secret information from each other. It also clarifies how information may be processed in different services regulated by law or by the requirements of external partners and the organization’s own data.

Data classification is not completely new topic for the organization, as there has been a Document Management System where data classification was utilized. Azure Information Protection (AIP) is a solution that provides data classification and rights management services aimed at protecting documents and preventing unauthorized access or modification. It is part of Microsoft Purview Information Protection and can integrate with other Microsoft cloud applications and services as well as third-party information protection solutions.

2 DATA LOSS PREVENTION

Using data classification, we can identify the exact types of data we have, where they are located, and the degree to which they are sensitive. Although it is crucial to have this knowledge, it is often overlooked as part of data security. A data loss prevention program (DLP) is one of the methods used by companies to prevent unauthorized disclosure of internal data – especially sensitive data. In order to ensure that our confidential data is kept out of the wrong hands, data classification and data loss prevention work together.

2.1 Data Loss Prevention Systems

Software that monitors, detects, and blocks sensitive information from leaving an organization is called Data Loss Prevention (DLP). A DLP software product typically focuses on blocking actions, so it monitors both data entering and exiting corporate networks. The software also prevents unauthorized copying by blocking employee computers from reading and writing to USB (Universal Serial Bus) thumb drives. The primary method of detection for phishing attacks is to monitor incoming emails and search for any suspicious attachments or hyperlinks (Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Tatbul, N., Zdonik, S. and Stonebraker, M., 2002).

Inconsistent content is typically flagged for a manual review or blocked by DLP software. It is easy to recognize critical data about an organization and take remediation action to prevent incidents when a DLP software engine utilizes a number of major advancements. Different technologies are used today to protect against data loss. Integrated DLP solutions and Enterprise DLP solutions are the two main classifications among DLP software. (Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Tatbul, N., Zdonik, S. and Stonebraker, M., 2002.)

Enterprise DLP agents can observe network traffic and emails through various means, including desktop and server software, and physical and virtual machines. The Integrated Data Loss Prevention solution includes a range of tools, such as a data classification tool, a content management platform, a cloud access security broker, and a data discovery tool. (Letia et al., 2001.)

Policy violations can be identified using a variety of content analysis techniques. For instance, the Database Fingerprinting technique can be used to identify precise matches from a database dump or a live database. This approach, also known as Exact Data Matching, is useful for organizing data from databases, despite the potential impact on execution resulting from database dumps or live database connections. Another widely used technique in the field of data loss prevention is the rule-based/Regular Expressions approach. This method includes an engine that filters content based on explicit rules, such as identifying IBANs with 24 digits or debit card numbers with 16 digits. While it may be prone to generating false positives, this approach is an effective way of identifying principles quickly. (Microsoft 2022.)

Within the realms of digital forensics and cybersecurity, hashing holds a crucial importance. Rather than contrasting file contents directly, a more effective method entails comparing their unique digital imprints. These imprints, referred to as hashes, are distinctive markers of a file's content, typically derived using cryptographic techniques. By assessing these hashes instead of the actual data, determinations about the sameness of files or records can be made promptly. This technique offers notable benefits. In instances where files or records are somewhat different but not completely, using hash comparisons greatly lessens the chances of wrongly assuming two files are the same when they aren't. Minimizing these mistakes is essential, particularly in high-stakes areas like digital forensic examinations or vital data handling. Utilizing hash comparisons makes it possible to attain remarkable precision in data validation endeavors. Owing to the distinct nature of every hash, even slight alterations in content will produce considerably divergent hash outputs. Hence, even when two file versions display minor variances, their hashes will be unmistakably different, providing clear evidence of their likeness or difference (Tahboub & Saleh, 2014).

The partial Document Matching technique is another approach that matches certain records, such as versions of a spreadsheet modified by different clients, to find a complete or partial match. Additionally, a concept/lexicon approach has been developed to handle completely unstructured data that may have difficulty categorizing. This approach should be modified according to the DLP solution. Categories with specific rules and word references that assist with analysis, such as Mastercard numbers and PCI protection, can also be used. (Carney et al., 2002.)

In the dynamic world of digital safety, pinpointing policy breaches concerning unsecured content is a daunting task. Thanks to modern computational tools, such as machine learning and statistical approaches like Bayesian analysis, there have been notable progressions in detection and rectification measures. These techniques utilize algorithms and math-based models to discern anomalies, patterns, and possible risks within extensive data sets. Specifically, machine learning utilizes evolving computational algorithms, which become more proficient as they encounter and process more information. This allows for the progressive improvement of its ability to detect breaches concerning unsecured content. Conversely, Bayesian analysis, grounded in the principles of the Bayes theorem, determines an event's probability by correlating prior knowledge with new insights. When focusing on unsecured content, Bayesian analysis gauges the risk of a security violation by evaluating past records and assimilating recent discoveries. Nevertheless, these advanced methodologies come with their set of obstacles. A foremost issue is the emergence of false positives and negatives. False positives arise when the system incorrectly flags safe content as unsecured, and false negatives occur when genuinely unsecured content is overlooked. Such misidentifications can result in unforeseen repercussions, ranging from baseless warnings to real security compromises. Addressing these issues requires an emphasis on gathering expansive and varied data. A more extensive data pool equips the algorithms with a broader perspective, fostering a thorough and precise comprehension of looming risks. By encompassing a diverse range of data inputs and perpetually refreshing the dataset, it's possible to better educate the algorithms, subsequently honing their ability to identify threats (Kaur et al., 2017).

2.2 Classification

Organization should determine where the sensitive data resides if it is the ultimate objective of its data security strategy to prevent breaches. A user cannot secure what it cannot see. As part of a data discovery strategy (Faiz, Arshad, Alazab and Shalaginov, 2020), sensitive data should be grouped into categories that will assist users in prioritizing risks by combining it with a robust and logical data classification strategy.

There are numerous methods of categorizing data that can assist users in determining where they need to focus on monitoring and security (Polozova & Anashkina, 2017). Therefore, data classification is a starting point for users' data loss prevention strategies. (Hart, Manadhata and Johnson, 2011.) It is then possible for users to determine who has access to and changes are being made to their sensitive and most at-risk data once they have located and classified it. The user will be able to reduce risk in the most critical areas (Tahboub & Saleh, 2014).

2.3 Authentication

Most DLP systems use Windows credentials management, a service or user provides credentials to the operating system and securely stores those credentials so that they can be presented to the authentication target in the future (Ma, 2017). Authentication targets on domain-joined computers are domain controllers.

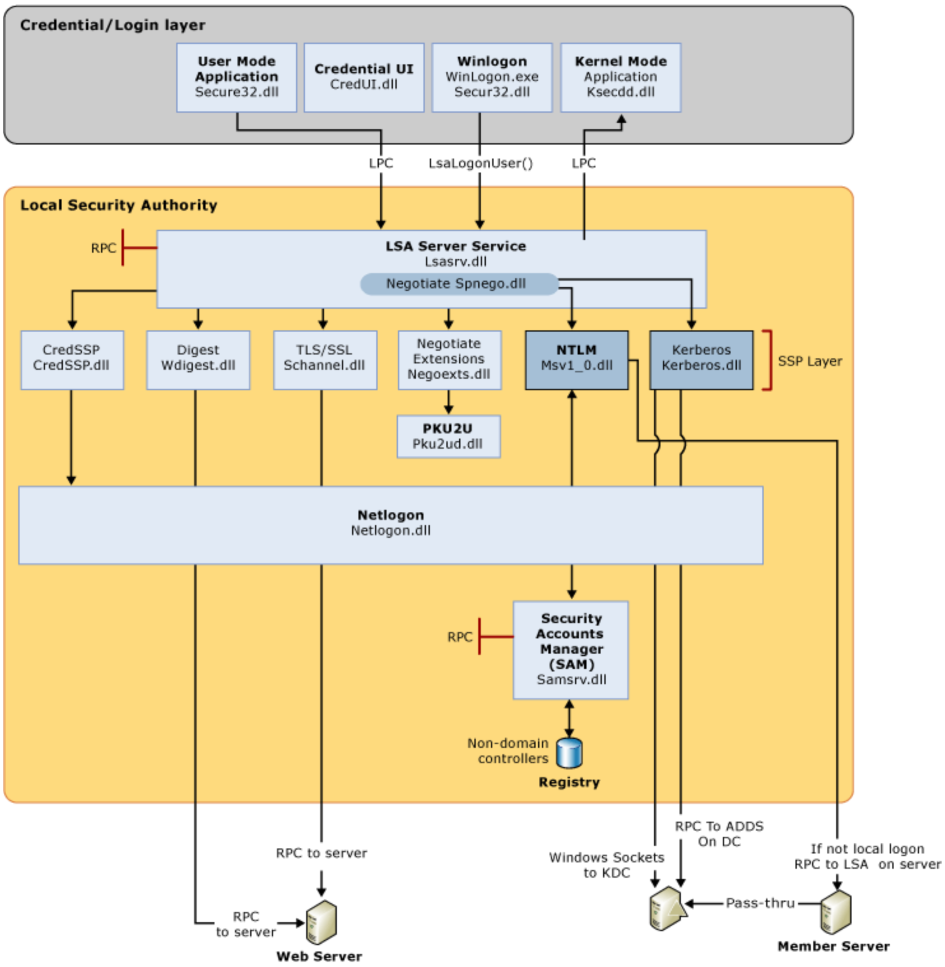

Figure 1 Windows authentication process (Microsoft, 2021)

As part of the registry, HKEY_LOCAL_MACHINE/SECURITY (Ma, 2017) contains local security information. As well as storing the SAM database, it also holds cached log-in credentials, policy settings, and default security settings. The diagram above illustrates the steps taken by credentials during a successful logon as well as the components that are required for authentication.

A credential is a digital document that proves a person's identity by demonstrating authenticities, such as a certificate, a password, or a personal identification number. In order to implement the Winlogon service, users have to provide their Windows credentials in order to automatically verify them with the Security Account Manager database, or, on domain-joined computers, with Active Directory.

Authentication targets receive credentials either by using the APIs provided by the applications or by incorporating input generated by the users on the logon user interface. (Microsoft, 2021.)

Alternatively, DLP systems commonly use LDAP (Lightweight Directory Access Protocol). It is a protocol for accessing and maintaining directory services. LDAP can be used for user authentication by querying a directory server, such as Microsoft Active Directory, to verify a user's credentials. When a user logs into a DLP system, the system can use LDAP to check the user's credentials against the AD server and grant access to classified files based on the user's permissions. (Microsoft, 2023.)

Other authentication methods used by DLP systems are: SAML (Security Assertion Markup Language) which is an XML-based protocol for exchanging authentication and authorization data between parties, such as a DLP system and an identity provider (IDP). When a user logs into a DLP system, the system can redirect the user to an IDP for authentication. The IDP then generates a SAML assertion, which includes information about the user's identity and permissions, and sends it back to the DLP system. The DLP system can then use the SAML assertion to authenticate the user and grant access to classified files. (Johansson & Cantor, 2018.)

3 MARKET OVERVIEW

Over the past few years, there has been an exponential rise of cybersecurity threats, and the market is reacting. The threat concerns all parties from state actors to small commercial businesses. Implementing a data loss prevention system is an easy and affordable solution to monitor and secure businesses data.

3.1 Market Definition and Scope

The technologies capable of performing both content inspection and data analysis forwarded over messaging applications, data in motion, data from endpoint managers, and server solutions (whether on-premises or cloud storage) are known as Data Loss Prevention (DLP) (Trellix, 2022). DLP solutions respond based on predefined rules to mitigate risks from accidental leaks or unauthorized data exposure to external channels. (Trellix, 2022). The user-defined rules are usually influenced by regulatory compliance, such as General Data Protection Regulation (GDPR), Payment Card Industry Data Security Standard (PCI-DSS), Health Insurance Portability and Accountability Act (HIPAA), and more (Groot, 2020). Therefore, the users first classify confidential data driven by regulatory compliances to follow national, international, and organizational regulations. The DLP software helps to identify policy violations based on the parameters to enforce remediation alerts, encryption, and other actions to mitigate or potentially eliminate data risks.

3.2 Types of DLP

DLP solutions focus on mitigating operation risks by covering how particular data flows: data in use, data in motion, and data in rest. Furthermore, they are classified into three categories: Network, Endpoint, and Cloud DLP. (Crowdstrike, 2022.)

3.2.1 Network DLP

The Network DLP focuses on monitoring and protecting all three data flows, including the cloud. It performs its function by tracking and analysing the network activity within an organization to detect any discrepancy under the company’s data security policies by monitoring employee email, messages, and file transfers. Network DLP provides complete details to the data security team as it acts as a ledger establishing a database filled with information on how and where the data flows. (Crowdstrike, 2022.)

3.2.2 Endpoint DLP

With the help of Endpoint DLP, the information security team can monitor all endpoints ranging from servers, laptops, smartphones, or other devices where data flows to prevent data misuse. It tracks both the data stored on and off the network and assists in streamlining regulatory requirements by classifying business-critical data as defined by the information security team. (Crowdstrike, 2022.)

3.2.3 Cloud DLP

The Cloud DLP is a segment under the Network DLP which is specially designed for companies highly dependent upon cloud repositories. The Cloud DLP solutions provide high visibility for all data activity in the cloud. It performs its functions by scanning and auditing the cloud servers to detect sensitive information. The Cloud DLP then encrypts the sensitive information before storing in the cloud. The information security team is alerted by the Cloud DLP solution when regulatory or policy violations occur by maintaining authorized users and applications lists. (Crowdstrike, 2022.)

3.3 Market Share and Key Players

The DLP solutions industry is fragmented, with many key players sharing the market share. The global DLP market is forecasted to grow at a CAGR of around 24% to USD 6 billion by 2026. Furthermore, 40% of its growth is expected to be fuelled by the North American market. Some key players in the DLP market are McAfee Corp, Trustwave Holdings Inc, Broadcom Inc, GTB Technologies Inc, CoSoSys Srl, and Digital Guardian Inc. (Technavio, 2022.)

In terms of DLP market share by application, cloud storage leads with 35%, followed by 25% of policy standards and procedures applications and 20% of encryption applications. The rest of the market share is divided equally among web & email protection DLP solutions and other applications, as highlighted in the figure below. These DLP solutions are used in healthcare, retail, logistics, defence, government bodies, telecommunication, and more industries. (IMAC Group, 2021.)

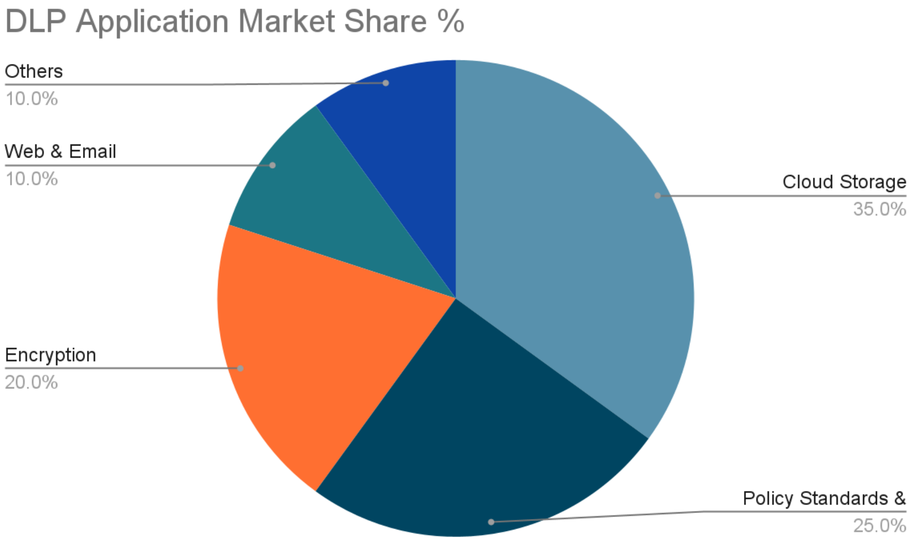

Figure 2 DLP Market share segments (IMAC Group, 2021)

According to data compiled and analysed by SoftwareReviews, a division of an IT research and advisory firm Info-Tech Research Group, the McAfee DLP is the market leader in DLP solutions based on user & vendor experiences and product features as shown in the figure 3 below. The figure below indicates that McAfee DLP has a high ease of implementation, ease of data integration, product strategy, development rate, and vendor support, making it the leading solution. (version2hk, 2021.)

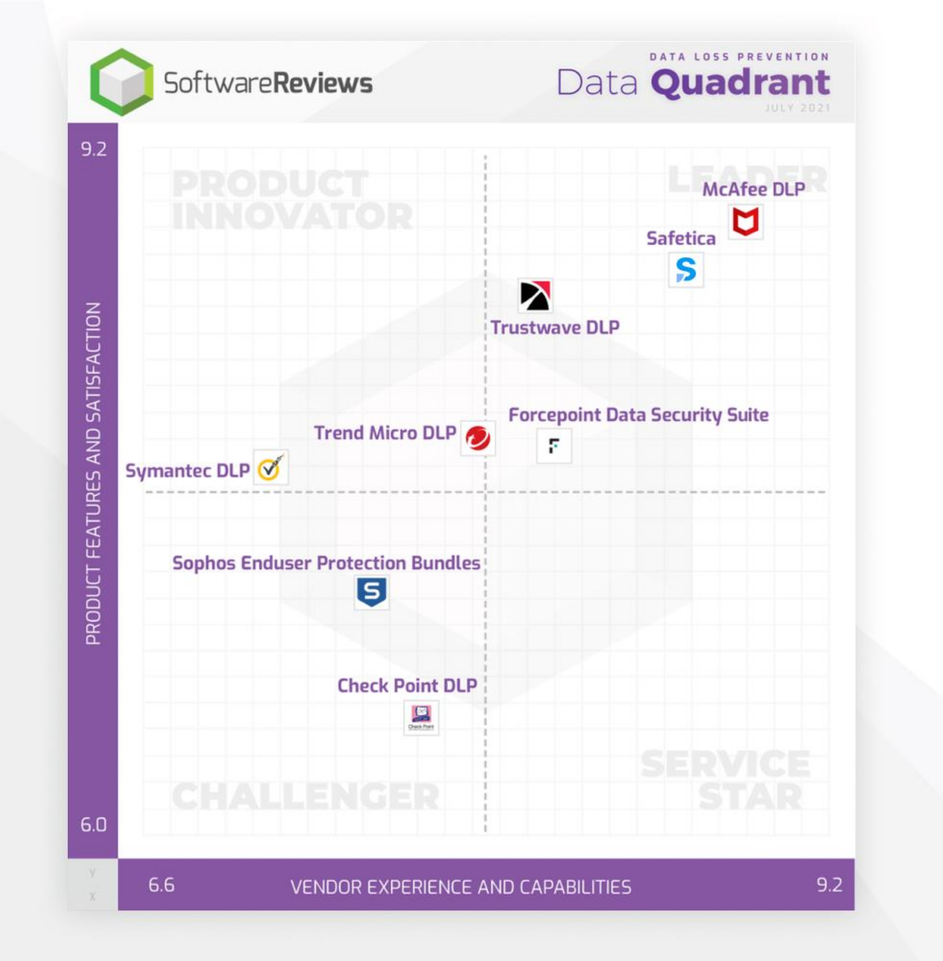

Figure 3 DLP Magic Quadrant (version2hk, 2021)

3.4 Key Players Products

The DLP solutions market comprises many players providing a wide range of services and products, as mentioned earlier. The top 8 DLP solutions based on reviews are as follows:

3.4.1 Digital Guardian

The Digital Guardian data protection solution works across endpoints, corporate networks, and cloud applications by giving deployment flexibility as it is available on a service or deployment model. The information security team can gain data transparency as the Digital Guardian solution provides insights on how the data flow without policies. The DLP solution is available for all endpoints and applications, so users can benefit from its rapid deployment and instant scalability. The AWS-powered web-based console analyses data activities from endpoints and network systems to provide remedies to internal and external threats. Furthermore, the Digital Guardian solution also protects intellectual property. (Robb, 2022.)

3.4.2 Fidelis

The Fidelis network provides a clear context of the data flow in a single platform. It detects, analyses threats, and halts sessions that violate the set policies. The users can use real-time sensors and a customizable policy engine to increase their data protection capability through Fidelis. The DLP solution gives clear context on bi-directional data, such as email handling to block threats that are coming and going from the organization. (Robb, 2022.)

3.4.3 Check Point

The Check Point DLP solution provides active Data Loss Prevention by empowering users to remediate real-time events. It focuses on educating the users on proper data handling policies and frees the information security team from the constant need to monitor data flow. To prevent data loss, the Check Point DLP solution is directly integrated into the firewalls and security gateways, helping to secure work applications, web traffic, email exchanges, and browsers. (Robb, 2022.)

3.4.4 Clumio

The Clumio DLP solution Protect & Discover provides a backup and recovery facility. It helps to meet regulatory requirements backed by global policies and provides ransomware protection through immutable and end-to-end encryption. The DLP solution provides maximum ease to the information security team as data can be restored within minutes and receive real-time recommendations to decrease data threats. As the data are stored out of the AWS accounts, Clumio solution does not require additional actions to protect backup data from malicious threats. Furthermore, users have the flexibility to automate and streamline backup plans as it provides backup lifecycle management (Robb, 2022.)

3.4.5 Trellix

The Trellix Data and User Security is a DLP solution established after the merger of McAfee Enterprise and FireEye. Trellix DLP solution leverages the power of Artificial Intelligence (AI) and Machine Learning (ML) to identify non-typical user behaviour across the enterprise with multi-vector DLP. The solution provides data protection through its intelligent threat identification, dynamic access adjustment system, and automated responses to the users. (Robb, 2022.)

3.4.6 Code42

The Code42 DLP focuses on high-risk activities of employees using contextual Incydr Risk Indicators allowing the information security team to protect data in risky events such as employee exit cases. The Code42 DLP solution monitors all the areas where data resides to identify incidents where files move outside of trusted sources. It helps users to identify the highest potential risks that require immediate attention through contextual risk scoring based on file vector and user behaviour. (Robb, 2022.)

3.4.7 Forcepoint

The Forcepoint DLP solution comprises predefined templates, policies, and classifications applicable to regulatory requirements of more than 80 nations enabling better control and visibility to the information security team. It allows users to collaborate securely as the DLP solution uses a policy-based auto-encryption system protecting data when it flows outside the company. (Robb, 2022.)

3.4.8 Proofpoint

The Proofpoint DLP solution is well known for its people-centric data protection approach. It allows granular level visibility into user interactions with confidential data in real-time. The solution has simplified deployment in either Software-as-a-Service (SaaS) or lightweight endpoint agent architecture. (Robb, 2022.)

3.5 Buyers Guide

As organizations' sheer volume of data is growing every day, their protection has become of utmost importance, especially since data loss can have devastating consequences for every business. It was estimated in 2019 that businesses would suffer from a cyberattack every 11 seconds in 2021, and the most likely threat are internal sources (Morgan, 2019). Data loss can have dire effects on an organization's financial and public image health. According to a report by IBM, the cost of a data breach reached more than USD 4.2 million in 2021 alone (IBM, 2022). Therefore, companies should use DLP solutions to prevent financial and reputational mishaps from data loss.

Things to consider before using DLP for organizations, to bring out the best value from DLP, users should consider the following best practices: firstly, DLP users should clearly define their objectives for using the solution, whether they want to protect their intellectual property, increase data flow visibility, prevent malicious attacks, or meet regulatory compliance. A clear objective will help users achieve the best out of DLP solutions (Spanning Cloud, 2021.)

To better use DLP solutions, the nature of data requiring protection should be identified, classified, and prioritized as it helps streamline and program the solutions better. Automated data classification technology can be used to save time and increase efficiency in classifying data. (Brooks, 2020.)

As DLP solutions can only work based on pre-determined policies to track and analyse malicious activities, developing comprehensive information security rules is vital to program them. The user roles and responsibilities should be clearly defined to restrict access to sensitive data. With the help of DLP solutions, information security teams can assign required authorization levels to users depending upon their needs. If the access levels are not pre-determined, then programming authorizations can be time-consuming. Simply deploying DLP solutions is not enough to protect and prevent data loss. Since humans are considered the weakest links in cybersecurity, educating and training them should also be a part of the information security process. (Spanning Cloud, 2021.)

Apart from the above-mentioned best practices, users should set particular evaluation requirements for DLP vendors. These requirements include understanding the deployment architectures, options, and managed service models offered by the vendors. They should also consider the support for a particular or multiple operating systems that the vendors provide. It's crucial to determine whether they offer internal or external threat protection. Users should be aware of the compliance regulations that bind their organization. Furthermore, they should consider the vendor’s technological partners and the availability of DLP technology integration. The deployment speed and ease provided by the vendors is another key consideration. Lastly, understanding the training and workforce requirements to use the solution is a crucial factor in choosing a vendor. (Digital Guardian, 2020.)

4 PSYCHOLOGICAL ASPECT

There is a misconception that cyber-attackers are computer geniuses who can hack their way into their target’s most complicated security systems and firewalls. Admittedly, there are weaknesses that can be exploited in poorly designed cybersecurity software and hardware. However, human beings are equally vulnerable, as they can be tricked into unknowingly opening a gateway to cybercriminals.

4.1 Overview

Attackers rely heavily on social engineering schemes to exploit the weaknesses of individuals who are their most effective backdoor entry pass into well-guarded infrastructure. These are people who are already armed with user IDs and passwords that can access precious data, trade secrets, proprietary blueprints, and customer lists, among others. An important part of the scheme is to get into the employees’ mindset in order to piggyback their way into restricted areas in the system. Therefore, aside from the technical aspects of DLP and cyber security in general, it is important to look into the psychology behind employees and attackers. (Emilianov, M. n.d.)

4.2 Social Engineering

Almost all cyber-attacks have a social engineering component where individuals are deceived in order to gain access to and obtain confidential data. According to Crawley (2021), there are five common types of social engineering. Firstly, phishing, the most popular type, mimics legitimate businesses via emails or text messages as well as social media posts and even websites. An example is when a victim clicks a misleading link contained in an email, which routes to an unsafe website to obtain user ID and password, such as a personal bank account. Secondly, pretexting involves fabricating a scenario, or a pretext, in order to steal information, such as someone pretending to call from the bank’s security team to confirm the target’s personal banking details. Thirdly, baiting promises inducements, usually free goods, or prizes, to trick users into handing in personal credentials. Fourthly, quid pro quo promises a service in exchange for information. Finally, tailgating involves an attacker piggybacking on someone’s security pass to enter a building, for instance.

4.3 Understanding the Mindset of Employees Towards Cybersecurity

Emilianov (n.d.) claims that human beings are the weakest link in cybersecurity because they have backdoor access, can expose confidential information, and do damage to infrastructure. La Huis and Salihoglu (2019) pointed out two critical psychological factors that affect employees’ attitudes towards cybersecurity. The first factor is a lack of understanding, where some employees may not be aware of the important role they play in protecting their company’s critical data from cyber-attacks. These employees value efficiency, i.e., getting the job done in the shortest time possible, rather than taking extra steps to protect corporate data. They also believe that cybersecurity is the sole responsibility of their IT department. The second factor is desensitization or cybersecurity fatigue, which arises from a growing apathy towards cybersecurity due to the proliferation of security breaches often seen in the daily news. In other words, there’s a perception that these attacks are simply a normal part of a computer-reliant world, and they would happen no matter what. This increasing complacency is an inherent weakness in any system.

On the other hand, Crawley (2021) explains the three psychological phases that cybersecurity professionals go through. Initially, they use all their cybersecurity know-how to build a system that they aim to be 100% secure. Eventually, they realize that 100% security is unattainable without sacrificing productivity, leading to blame placed on users for their ignorance and foolishness. In the final phase, they accept that everyone, including experts, has weaknesses and can make lapses in technology. For example, an emergency evacuation order should trigger a network-wide shutdown of all computers to protect sensitive data, instead of relying on panicked employees.

5 RISK MANAGEMENT

Improving cybersecurity is a double-edged sword. While increasing cybersecurity measures should be one of the businesses top priorities, there can be such thing as too much security. It comes from exaggerated understanding of the risk businesses faces and trying to implement policies too harsh. This will end up hindering the business performance, lowering competitiveness and agility. The solution is to take all the workload from IT staff and redistribute the responsibility and knowledge to as many staff as possible. (Kissoon, 2021.)

5.1 Preventing Cyber Attacks

Aside from the technical cyber-attack preventative measures such as more encryption and multi-factor authentication, Crawley (2021) emphasizes a series of steps to help improve cybersecurity. An important long-term goal is to create and maintain a security culture where proper vigilance and scepticism become second nature to every employee. This culture will take a long time to develop through regular training, but it will undoubtedly be worth it.

Awareness of the important role of employees in cybersecurity needs to be promoted, emphasizing the detrimental impact of breaches and providing easy to remember tips on how to prevent them. Tools should be provided to help employees identify threats, detect them, and understand what actions to take. Mandatory cybersecurity training, as well as random phishing tests, need to be conducted to keep employees on alert for any potential threat. Regular reviews and evaluations of security systems and protocols should be carried out to identify areas for improvement. Pairing strong and dynamic technological tools with a psychological approach to cybersecurity can prove to be formidable defences against cyber-attacks both inside and outside any organization. (Prasad & Moon, 2022.)

5.2 DLP Deployment

Security teams set the rules for detection and blocking DLP in the early days, but these rules were often circumvented due to their simplistic nature. With newer software, the detection and blocking approaches can be improved over time by using machine learning-based artificial intelligence. The user should verify whether SharePoint Online supports encrypted (unified) labels in its tenant. The encrypted label option is disabled by default (Microsoft, 2022). Users can enable it via Security Centre or check its status through PowerShell if it is not visible there. The tenant should be given 24 hours after the option has been enabled.

6 OVERALL SECURITY

DLP tool by itself is not sufficient for a business to secure their data. DLP should be used in conjunction with other cyber security systems to further improve the OPSEC. As mentioned in chapter 3.2 it is important to recognize what kind of data the business has, where the data resides, who is allowed to access it and who is really accessing it. These are the foundation to data governance. In this chapter some examples are given which systems businesses should consider using in combination with DLP.

6.1 Network Security Systems

6.1.1 IDS

A security product known as an Intrusion Detection System (IDS) uses security policies to notify administrators when malicious activity is detected. Detecting computer malware and attacks is the main goal of IDS. (SOC/SIEM, 2017.)

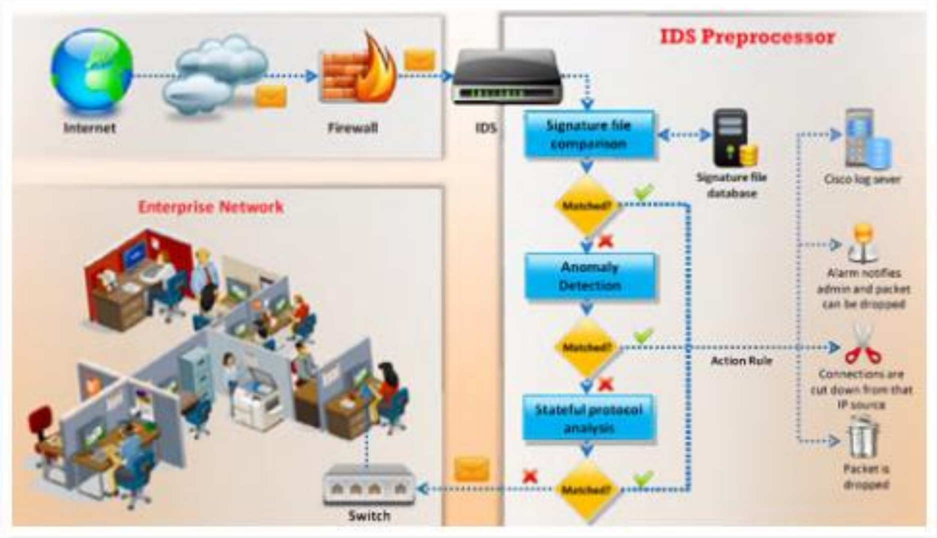

Figure 4 IDS Communication Diagram (SOC/SIEM, 2017)

Advanced IDS have sensors for detecting signatures, as well as behavioural activity to identify malicious behaviour. When the signature matches or the packet is dropped, a notification will be sent to the administrator if the connection from the source IP has been lost. If the signature matches, sensors will forward anomaly detection, regardless of whether or not the packet or request received matches. When a packet passes the anomaly stage, stateful protocol analysis is carried out, and then packets are passed on the network through switches. In the event of a mismatch, an alarm is raised and the administrator is notified after the connection from the source IP address is terminated and packets are dropped. (Ashoor and Gore, 2011.)

6.1.2 IPS

By monitoring a network and identifying unusual activity, intrusion prevention systems (IPS) protect a company's network against intrusion. Its main purpose is to detect and report unusual activity. (Wallarm 2022.)

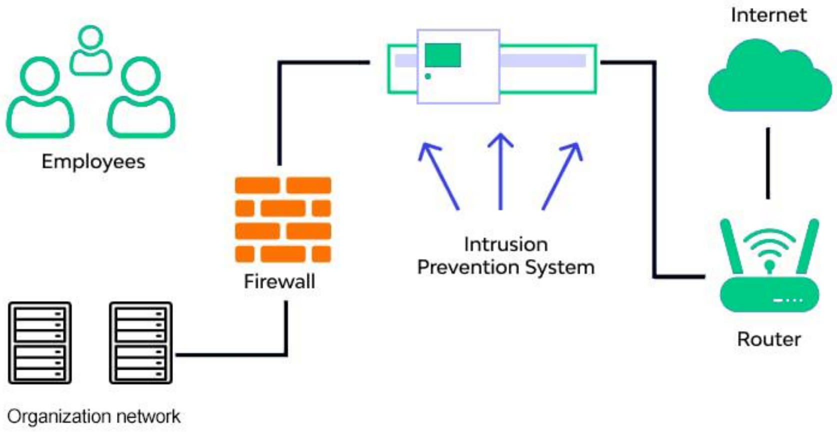

Figure 5 IPS communication diagram (Wallarm 2022.)

The information contained in directed network information allows an intrusion prevention system to distinguish between malignant movement and perceived assault designs. As IPS motors inspect network traffic, they analyse it consistently for realized assault designs based on their inward signature data. The IP address or port of an aggressor might be dropped by an IPS if it is not determined in stone to be vindictive. Real traffic can keep on streaming without any apparent assistance interference. An IPS normally logs and notifies security directors of identified occurrences, in addition to deterrents and security updates, an IPS can constantly monitor and combat Internet threats, helping organizations stay safe on the internet. (Patel, Taghavi, Bakhtiyari and Celestino Júnior, 2013.)

6.1.3 NBAD

Network Behaviour Anomaly Detection uses network behaviour anomalies to detect malware or attacks in real time. It monitors individual networks and detects attacks. When an anomalous behaviour is detected on enterprise networks (Shah et al., 2016), network behaviour anomaly detection notifies security teams or initiates an automated response.

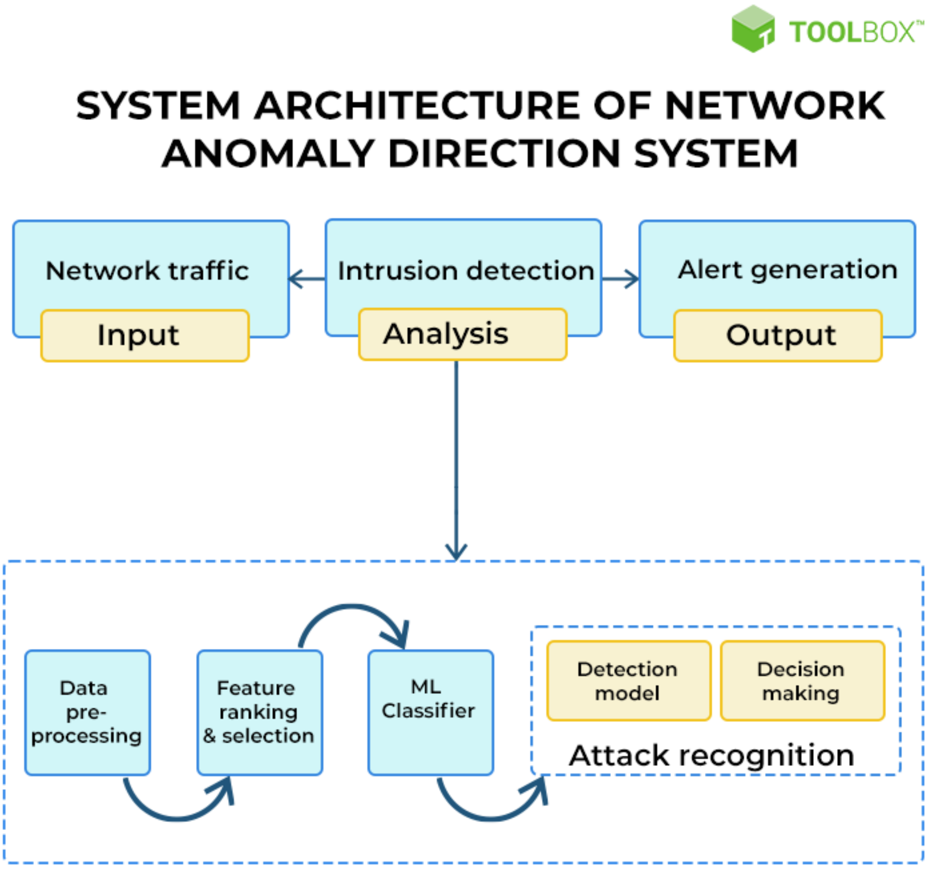

Figure 6 NBAD communication (Ashtari, 2022)

6.2 Security Information and Event Management

6.2.1 SIEM

Security Information and Event Management (SIEM) serves as a fundamental component in today's cybersecurity framework. This robust mechanism integrates several features, such as gathering, observing, and evaluating security incidents, into a unified system (Swanagan, n.d.). Through SIEM, IT managers obtain an all-encompassing insight into operations within the IT domain, allowing them to act preemptively against looming threats. The core operation of SIEM systems involves accumulating extensive log information from a range of elements in the IT structure, including servers, protective barriers, and various network components. After gathering, this information is carefully scrutinized to pinpoint irregular behaviors or actions suggesting security compromises or malevolent activities. Essentially, SIEM operates as a watchful guardian, persistently overseeing the digital terrain for deviations. A standout benefit of SIEM lies in its real-time surveillance capabilities. In a time characterized by rapidly changing cyber dangers, promptly recognizing and addressing these hazards is essential. SIEM facilitates this by notifying IT managers about emerging concerns instantaneously, curtailing the chances for ill-intentioned individuals to take advantage of system weaknesses. Furthermore, many SIEM solutions incorporate a series of standard compliance outlines that adhere to multiple regulatory benchmarks. With a growing focus on global data safeguarding directives, entities can utilize these outlines to confirm their IT procedures conform to recognized guidelines, thereby minimizing potential regulatory infringements. By consolidating security-related data from assorted origins, SIEM grants IT managers a comprehensive perspective of their technological landscape. This consolidated view proves crucial in making data-driven choices, formulating protective measures, and allocating resources efficiently for security-related endeavors (Schultz, 2009).

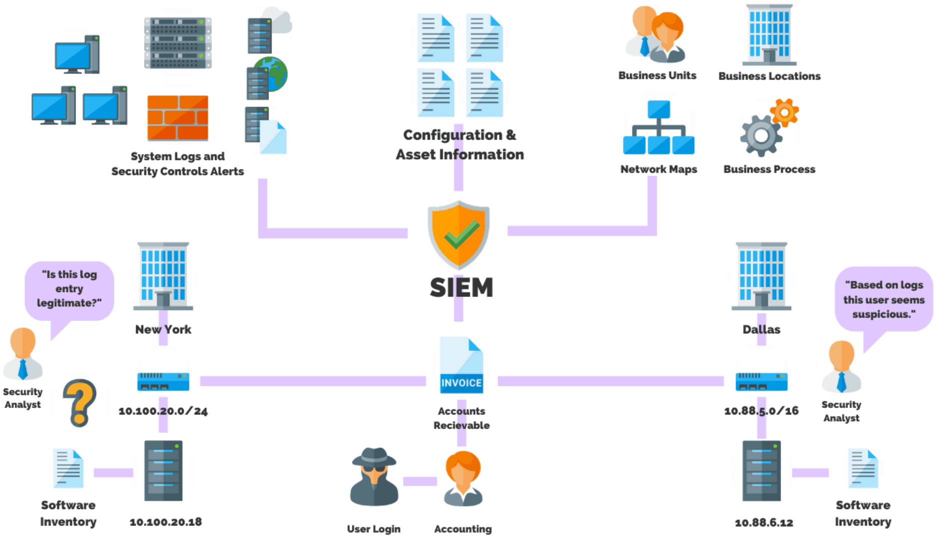

Figure 7 SIEM communication diagram (Swanagan, n.d.)

Security Information and Event Management (SIEM) systems can detect threats, comply with regulations (Anastasov and Davcev, 2014), and manage security incidents by collecting and analysing security events, as well as a variety of other event and contextual data sources. One of the core capabilities is the collection and management of log events, as are the capabilities for analysing log events, as well as operational capabilities (like incident management, dashboards, and reporting) across disparate sources (Anastasov and Davcev, 2014). With SIEMs, IT staff can quickly detect attacks before, during, and after they happen, resulting in a more effective incident response. In order for a cyber security program to be successful and continuously monitored, SIEM technology is essential. (Swanagan, n.d.).

6.3 Identity and Access Management Systems

6.3.1 IAM

A user's identity and authorization are managed through Identity Access Management (IAM). An integrated identity management solution combines features compatible with a zero-trust approach to cybersecurity (Thakur, Gaikwad, 2015). In this approach, users are prompted to verify their identity when requesting access to company data, servers, applications, or services. Every member of an organization has access to IAM solutions. It does not matter what device a user uses to access the company's infrastructure; all users are identified. An array of applications and services are accessible to users through IAM. (Ruchini, 2021).

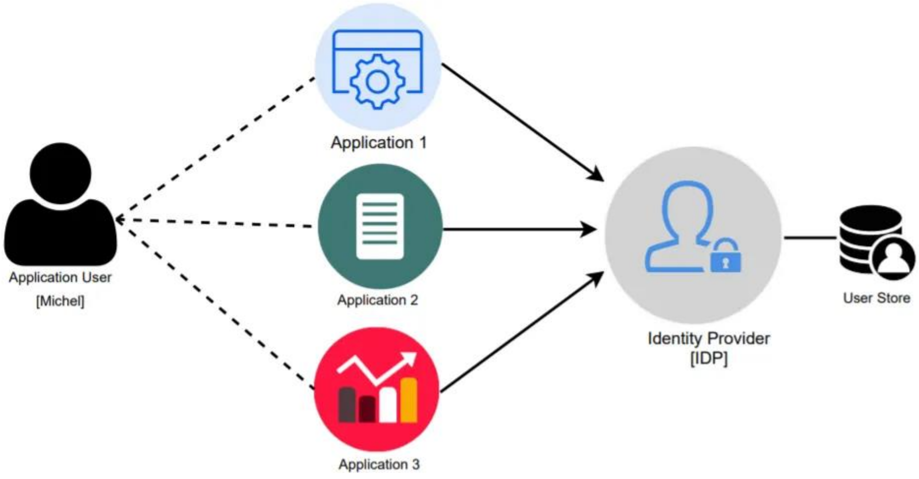

Figure 8 IAM communication diagram (Ruchini, 2021)

The Centralized Access Management system maintained by IAM addresses these issues. Users are authenticated and accounts are controlled here. The Identity Provider (IDP) is responsible for managing users (Thakur and Gaikwad, 2015). Logging through this IDP is trusted by all applications. This Centralized Access management system has the main advantage of relieving the application developer from managing users and reducing the burden on users to manage their own accounts.

6.3.2 PAM

Privileged Access Management provides permission for more sensitive data to be accessed by privileged users as a subset of IAM. Users' actions and access to sensitive information can be controlled and limited through PAM. (Purba and Soetomo, 2019). User access to privileged information is the primary purpose of PAM solutions. Users with administrative privileges can access highly sensitive systems through PAM. (Biswas, 2020).

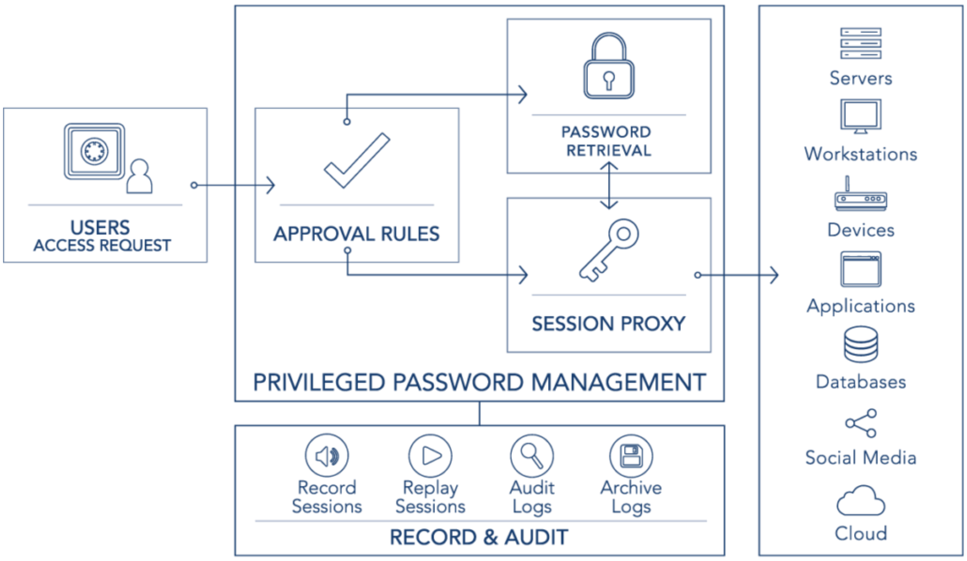

Figure 9 PAM communication diagram (Beyondtrust)

By having a tamper-proof password safe, all privileged credentials that provide elevated access can be discovered, onboarded, and managed, which provides a secure, centralized location for discovering, onboarding, and managing credentials. In order to manage and secure credentials used for application-to-application and application-to-database interactions, application-to-application password management (AAPM). (Purba and Soetomo, 2019). capabilities are vital. In addition to removing embedded credentials from code, vaulting them, and following best practices, embedded credentials can be managed by standalone tools or as part of privileged credential management / PASM software. (Biswas, 2020). There are some tools that provide secret management capabilities for DevOps and Continuous Integration / Continuous Delivery workflows (Purba and Soetomo, 2019). Managing privileged sessions (PSM) involves monitoring and managing all sessions involving elevated access and permissions for users, systems, applications, and services. (Biswas, 2020).

7 AZURE INFORMATION PROTECTION

Azure Information Protection is part of Microsoft Purview Information Protection. The solution consists of data classification and rights management services, its aim is to protect documents, tackling the challenges of unauthorized access or modification. To better understand how AIP works, understanding the product lifecycle is necessary.

7.1 Licensing

The organization chose Azure Information Protection to enable their DLP needs. The business already had an active enterprise licensing for Microsoft, and they upgraded the subscription to E5 Security Addon. There are three tiers of Azure Information Protection: AIP for Office 365, Premium P1 or Premium P2. These options have varying feature sets and the P2 features are available by default in the E5 and its Compliance and Security add-ons.

Management of Core Identity & Access (Bhardwaj, Bannerjee and Roy, 2021)	**Free**	**OFFICE 365 Apps**	**Premium P1**	**Premium P2**
Active Directory Objects	50.000 Object limit	No object limit	No object limit	No object limit
Synchronization of Device objects between on-premises directories and Azure AD (Device write-back)	Not available	Available	Available	Available
Password Protection (custom banned password)	Not available	Not available	Available	Available
Windows Server Active Directory password protection (global and custom banned passwords)	Not available	Not available	Available	Available
Self-service BitLocker recovery, enterprise state roaming with Azure AD Join (Bhardwaj, Bannerjee and Roy, 2021)	Not available	Not available	Available	Available
Advanced Security and Usage Reports	Not	Not	Available	Available

	available	available		
Application Proxy	Not available	Not available	Available	Available
Microsoft Identity Manager CAL5	Not available	Not available	Available	Available
Group naming policy	Not available	Not available	Available	Available
Azure Information Protection integration	Not available	Not available	Available	Available
Sharepoint Limited Access	Not available	Not available	Available	Available
Terms of Use (set up terms of use for specific access)	Not available	Not available	Available	Available
Multi-Factor Authentication with Conditional Access (Bhardwaj, Bannerjee and Roy, 2021)	Not available	Not available	Available	Available
Microsoft Cloud App Security integration	Not available	Not available	Available	Available
3rd party Identity governance partners integration	Not available	Not available	Not available	Available
Entitlement Management	Not available	Not available	Not available	Available
Price	Free	M365 E1, E3, E5, F3	$6/user/month	S9/user/month

Table 1 Licensing options

7.2 Rights Management

Using Azure Rights Management Service (Azure RMS), Azure Information Protection protects user data. A SharePoint-based information management system that integrates with other Microsoft cloud applications and services, such as Office 365 and Azure Active Directory, as well as with users' own applications and third-party information protection solutions, is Azure RMS. Cloud and on-premises solutions can be integrated with Azure RMS. Using Azure RMS, users can encrypt, identify, and authorize users (Azure Information Protection (AIP) labeling, classification, and protection, 2022). When using Azure RMS, the protection maintained with documents and emails stays with them regardless of where they are stored. (Jin and Stivers, 2017). This provides users with control over their content even when they share documents or emails with others.

Azure Information Protection (AIP) protects users' data through several key steps. Firstly, data is classified and labelled to establish its sensitivity level. Secondly, a label-based policy, access control, and encryption are used to protect the data. Thirdly, access to documents can be revoked if necessary, and the documents can be tracked to monitor their usage. Fourthly, documents that contain sensitive information are protected using a content key that is stored in the Azure Information Protection tenant root key in the file header for each document. An authorized user or service can open the document using the content key. Users can manage their own tenant keys, or Microsoft can manage them on their behalf. Finally, Azure Rights Management Services (RMS) ensures that the secret formula is not sent to Azure during the encryption, decryption, authorizing, and restriction processes. These steps work together to provide comprehensive protection for users' data and prevent unauthorized access to sensitive information. (Lakshmi, 2019.)

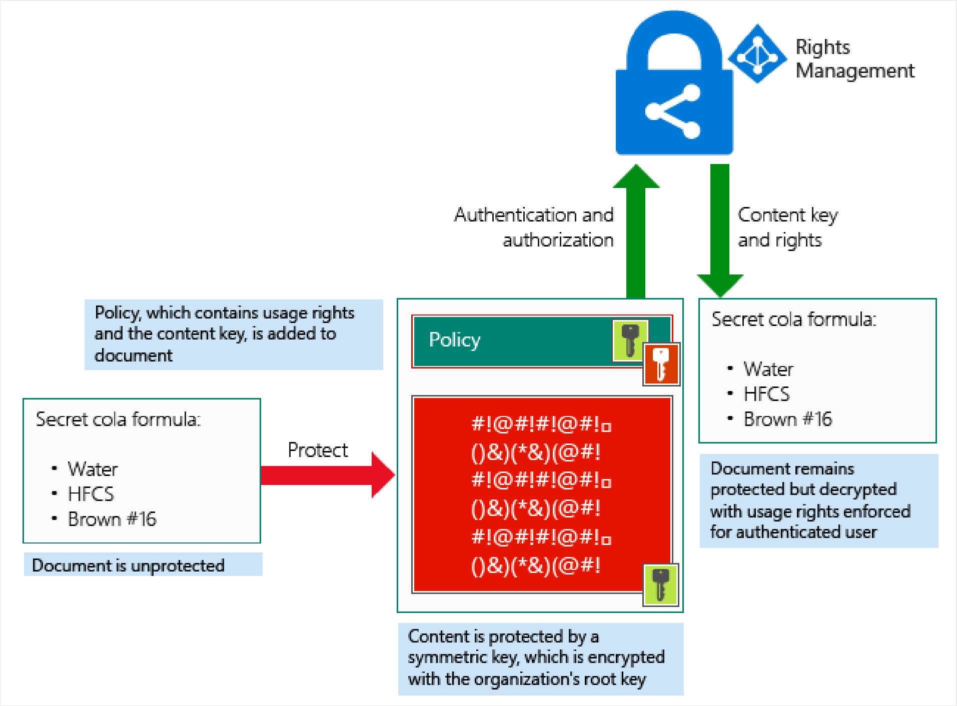
Rights Management
Authentication and authorization
Content key and rights
Policy, which contains usage rights and the content key, is added to document
Policy
Secret cola formula:
• Water
• HFCS
• Brown #16
Protect
#!@#!#!@#!¤
()&)(*&)(@#!
#!@#!#!@#!¤
()&)(*&)(@#!
#!@#!#!@#!¤
()&)(*&)(@#!
Document is unprotected
Content is protected by a symmetric key, which is encrypted with the organization's root key
Secret cola formula:
• Water
• HFCS
• Brown #16
Document remains protected but decrypted with usage rights enforced for authenticated user

Figure 10 AIP workflow diagram (Microsoft, 2022)

7.3 Encryption

According to Mojžiš and Balogh (2020), customers have the ability to classify, label, and encrypt their data with AIP. AIP can be leveraged by IT administrators in several ways. First, preset rules can automatically classify emails and documents. Second, watermarks, headers, and footers can be added to content. Third, confidential files can be protected using Rights Management, which involves signing operations using SHA-256 and RSA 2048-bit keys, encrypting files, and sending them to designated recipients within and outside the organization. To restrict the use of the file (Greenacre, 2018), a specific set of rights can be applied. The rights policy determines how content should be encrypted based on a user's identity and authorization.

In the real world, IT administrators can reduce the amount of protected content. AIP prevents unauthorized access to content, and a controlled environment free of malicious apps can be achieved using products such as Microsoft Intune, Microsoft Endpoint Configuration Manager, and others (Kesan, Hayes, Bashir, 2013). Users can manage and control their environment with AppLocker to prevent the installation of malicious applications. Azure Active Directory (AD) Identity Protection enhances trust in the identity of users, while Enterprise Mobility + Security (EMS) Conditional Access enhances device and identity trust. Data that contains personally identifiable information is the most valuable asset an organization can have in the digital age. These capabilities enable better control of data from end to end, and AIP facilitates information security and classification. Organizers have benefited greatly from AIP since its introduction. (Michaelis, 2012.)

7.4 DLP circumvention

The copy usage right is not granted by Rights Management in several Windows platforms (Windows 7, Windows 8.1, Windows 10, Windows 10 Mobile, and Windows 11) which prevents the use of screen capture tools, and screen captures cannot be prevented by browsers. The web versions of Outlook and Office use browsers and screen captures are not prevented by apps on iOS, Mac, or Android devices. Users can avoid accidental or negligent disclosure of sensitive or confidential information by using screen captures to share data on a screen. It is critical to recognize that users can share data displayed on a screen in many ways, such as taking screenshots. (Device Advice, 2021.)

In spite of supporting the Rights Management API, platforms and software could not stop users from sharing data they should not (Microsoft, 2022). Authorization and usage policies in rights management can help protect users' critical data, but they should be used in conjunction with other controls. When documents are labelled, classified, and protected, they cannot be distributed outside of your organization (Business Tech Planet, 2022). When people use AIP, information can be kept private in several ways.

Users are not allowed to simply go into a document, copy and paste the content, and send it across to another email address simply because a protected document prevents them from doing so. If a user attempts to email a protected document or sensitive information, they will receive an email telling them that their email cannot be sent. Protected documents or sensitive information can be blocked before they are sent. When a user attempts to send a sensitive email, an IT department and administrator will be notified via the Azure portal. (Business Tech Planet, 2022.)

8 IMPLEMENTATION

According to International cybersecurity standards ISO27001 and 27002 (SFS-ISO/IEC 2013a, 2013b) and National recommendations (VAHTI 2010, 51) organisations should have documented principles about classifying data. Organization should implement an approach and guides to handle classified data. Their purpose is to give details of what classified data is allowed to be handled, moved, stored or spread. (SFS-ISO/IEC 2013b.)

8.1 Classification

Building a classification plan or classification model starts with identifying the need. The purpose of classification is essential to consider. Company leaders and board members need to give their full support towards the implementation of data classification. If the implementation gets created without going through the executive committee, it might lead to situation where it does not suit all departments of the business. For the project to progress, responsibilities and duties should be defined at the very beginning. One other part of making successful classifications is to involve employees to the creation process. This way they have a better understanding of what is confidential information, and which requires protection. In the data identification phase, everything should be documented from data location to file content and owner.

Cybersecurity standards (SFS-ISO/IEC 2013a, 2013b) suggests classifying data with legal demands, as well as businesses needs to different categories. Labels should be marked in a way, that the one handling data knows clearly under which label the data belongs to. Marking can be done to the document by a text mark or a technical metadata. (SFS-ISO/IEC 2013b.)

8.2 Labeling

The classifications decided for the implementation of Azure Information Protection are the preconfigured default labels. The initial idea was to set the Highly Classified label to follow Finnish Defence Ministry's (Katakri, 2020) Protection Level 4 (ST IV) guidelines, but it was not carried out due to the strict requirements demanded. The amount of project data consisting of so called “red-label” documents was so small it was a better choice to keep such classified data on offline storage inside a safe within a room with access control and camera surveillance, thus meeting the necessary security requirements.

8.3 Proposed Labels

Proposed labels	**Label description**
Personal	Non-business data, for personal use only.
Public	Business data that is specifically prepared and approved for public consumption.
General	Business data that is not intended for public consumption but can be shared with external partners as required. Examples include a company internal telephone directory, organizational charts, internal standards, and most internal communication.
General \ Anyone (unrestricted)	Organization data that isn’t intended for public consumption but can be shared with external partners if appropriate. Examples include customer conversations that don’t include sensitive info or released marketing materials.
General \ All Employees (unrestricted)	Organization data that isn’t intended for public consumption. If you need to share this content with external partners, confirm with other data owners that it's OK to share and then change the label to General \ Anyone (unrestricted). Examples include a company internal telephone directory, organizational charts, internal standards, and most internal communication.
Confidential	Sensitive business data that could cause damage to the business if shared with unauthorized people. Examples include contracts, security reports, forecast summaries, and sales account data.
Confidential \ Anyone (unrestricted)	Confidential data that doesn’t need to be encrypted. Use this option with care and appropriate business justification.
Confidential \ All Employees	Confidential data that requires protection, which allows all employees full permissions. Data owners can track and revoke content.
Confidential \ Trusted People	Confidential data that can be shared with trusted people inside and outside your organization. These people can also reshare the data as needed.
Highly Confidential	Very sensitive business data that would cause damage to the business if it was shared with unauthorized people. Examples include employee and customer information, passwords, source code, and pre-announced financial reports.
Highly Confidential \ All Employees	Highly confidential data that allows all employees view, edit, and reply permissions to this content. Data owners can track and revoke content.
Highly Confidential \ Specific People	Highly confidential data that requires protection and can be viewed only by people you specify and with the permission level you choose.

Table 2 Proposed Labels (Microsoft, 2023)

8.4 Deployment

To begin the deployment, executive committee had to go through the labels with all departments. Once the green light was received, planning began for the different steps of deployment. In this scenario, within the Microsoft environment, the only steps required were creating the unified labels under AIP and deploying the Azure Information Protection Client to all users with Group Policy. One decision made was not to the mass classification of data, but to give users the freedom to classify after they had acknowledged the internal training material. To test the classification, the Client was installed only to a small test group. The Labels were deployed to all users under the tenant simultaneously because Microsoft does not recommend partial implementations of unified labels.

8.5 Training

Training and guiding are a key part of data loss prevention systems, employees should be told the impact, goals and benefits as well as responsibilities and duties. If the organization does not have a classification model or plan of action, then employees can't verifiably know how information is being handled. Additionally, employees should be notified why data is being classified, what data is beneficial to the business, what classes there are, and which information belongs to which classification.

Defining responsibilities and duties aims to clarify employees' roles. Especially the role of the data's owner is crucial. It must remain disclosable who owns the data and what duties the owner has towards the classification. An untrained employee might misuse classified data, and it might lead to under classifying as well as false sense of security. Over classifying data is also a threat because the data could not be reached in time. (SFS-ISO/IEC 2013b)

9 CONCLUSION

The thesis described the initiation of data loss prevention work in the organization's IT unit, from comparing products to deploying one. During the sub-studies that formed the thesis, how the public classification of electronic documents can be understood, implemented, and utilized in the IT unit was explored. The goals set for the sub-studies were achieved.

The thesis was a long process, but completing it taught me many new things. Classifying information turned out to be a broad and challenging topic. Thanks to the sub-studies, I understand the subject better and can apply the knowledge in practice. The expertise gained through the thesis is important, as the classification of information has been a topic that industry experts should master for years.

I also feel that I am better prepared to develop my professional skills and engage in lifelong learning. Additionally, I learned to utilize research information and methods more efficiently as part of working life development. The full construction and implementation of Azure Information Protection in the organization were not carried out because such an implementation can take a significant amount of time. However, based on the results, the organization can assess whether it is sensible and worthwhile to continue and expand the use of data loss prevention systems.

To progress towards the classification of information resources, other significant tasks must still be carried out in both the IT unit and the entire organization. The organization must define the security requirements associated with each classification category, implement the rules for marking and processing information, train employees in the new operating model, and introduce supportive security mechanisms. Furthermore, the need for classification based on information integrity and availability should be considered.

The organization now possesses the necessary general knowledge, understanding, and skills to implement information classification. The journey towards the classification of information resources has begun.

REFERENCES

Anastasov, I. and Davcev, D., 2014. SIEM implementation for global and distributed environments. 2014 World Congress on Computer Applications and Information Systems (WCCAIS).

Ashtari, H.T. (2022) *What is Network Behavior Anomaly Detection? definition, importance, and best practices for 2022*, *Spiceworks*. Available at: (https://www.spiceworks.com/tech/networking/articles/network-behavior-anomaly-detection/) Accessed: 07 June 2023.

Ashoor, A. and Gore, S., 2011. Difference between Intrusion Detection System (IDS) and Intrusion Prevention System (IPS). Advances in Network Security and Applications, pp.497-501.

Bhardwaj, N., Banerjee, A. and Roy, A., 2021. Case Study of Azure and Azure Security Practices. Machine Learning Techniques and Analytics for Cloud Security, pp.339-355.

Biswas, S. (2020) Integrating BeyondTrust password safe to secure Nutanix prism using privilege and identity..., Medium. Available at: https://saptarshi-biswas-999.medium.com/integrating-beyondtrust-password-safe-to-secure-nutanix-prism-using-privilege-and-identity-9be32c57892c (Accessed: 09 October 2023).

Business Tech Planet. 2022. How Does Azure Information Protection Work? | Business Tech Planet. Available at: (https://businesstechplanet.com/how-does-azure-information-protection-work/#:~:text=You%20can%20prevent%20users%20from,be%20blocked%20prior%20to%20transmission) Accessed 20 September 2022.

Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Tatbul, N., Zdonik, S. and Stonebraker, M., 2002. Monitoring Streams — A New Class of Data Management Applications. VLDB '02: Proceedings of the 28th International Conference on Very Large Databases, pp.215-226.

Computernetworksecuritis.blogspot.com. 2017. Intrusion-Detection-System (IDS) And Its Detailed Working Function -SOC/SIEM. Available at: (https://computernetworksecuritis.blogspot.com/2017/10/intrusion-detection-system-ids-and-its.html) Accessed 20 September 2022.

Crowdstrike, 2022. What is Data Loss Prevention (DLP)? [beginners guide]: Crowdstrike. crowdstrike.com. Available at: (https://www.crowdstrike.com/cybersecurity-101/data-loss-prevention-dlp/) Accessed: 20 June 2023.

Crawley, K. (2021, October 5). 8 steps to better security: A simple cyber resilience guide for business. John Wiley & Sons. Available at: (https://learning.oreilly.com/library/view/8-steps-to/9781119811237/c07.xhtml#head-2-53) Accessed: 20 June 2023.

Device Advice. 2021. Block screenshots using Microsoft Information Protection - Device Advice. [online] Available at: (https://deviceadvice.io/2021/11/01/block-screenshots-using-microsoft-information-protection/) Accessed 20 September 2022.

Digital Guardian, 2020. What is Data Loss Prevention (DLP)? A definition of data loss prevention. Digital Guardian. Available at: (https://digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention) Accessed: 15 September 2022.

Docs.microsoft.com. 2022. Azure Information Protection (AIP) labeling, classification, and protection. Available at: (https://docs.microsoft.com/en-us/azure/information-protection/aip-classification-and-protection) Accessed 11 September 2022.

Emilianov. M. (n.d.), Why the Real Key to Cybersecurity Is Psychology, Cloudindustryforum.Org, Available at: (https://www.cloudindustryforum.org/content/why-real-key-cybersecurity-psychology) Accessed: April 11, 2022

Faiz, M., Arshad, J., Alazab, M. and Shalaginov, A., 2020. Predicting likelihood of legitimate data loss in email DLP. Future Generation Computer Systems, 110, pp.744-757.

Gartner Inc., Data Loss Prevention Reviews 2022: Gartner Peer insights. Gartner Inc. Available at: (https://www.gartner.com/reviews/market/data-loss-prevention) Accessed: 20 June 2023.

Greenacre, M., 2018. Compositional Data Analysis in Practice. Available at: (https://www.taylorfrancis.com/books/mono/10.1201/9780429455537/compositional-data-analysis-practice-michael-greenacre) Accessed 11 September 2022.

Groot, J.D., 2020. What is Data Loss Prevention (DLP)? A definition of data loss prevention. Digital Guardian. Available at: (https://digitalguardian.com/blog/what-data-loss-prevention-dlp-definition-data-loss-prevention) Accessed: 20 June 2023.

Hart, M., Manadhata, P. and Johnson, R., 2011. Text Classification for Data Loss Prevention. Privacy Enhancing Technologies, pp.18-37.

Huis, T., & Salihoglu, M (2021, October 30). The Psychology of Cybersecurity. Crowe. Available at: (https://www.crowe.com/cybersecurity-watch/psychology-of-cybersecurity#:~:text=Two%20important%20psychological%20factors%20are,become%20desensitized%20to%20the%20threat) Accessed: 20 June 2023.

IBM, 2022. Cost of a data breach report 2022. IBM. Available at: (https://www.ibm.com/in-en/security/data-breach) Accessed: 20 June 2023.

IMAC Group, 2021. Data Loss Prevention Market: Global Industry Trends, share, size, growth, opportunity and forecast 2022-2027. Data Loss Prevention (DLP) Market Size, Share, Trends 2022-2027. Available at: (https://www.imarcgroup.com/data-loss-prevention-market) Accessed: 20 June 2023.

Jin, G. and Stivers, A., 2017. Protecting Consumers in Privacy and Data Security: A Perspective of Information Economics. SSRN Electronic Journal.

Johansson, L. and Cantor, S. (2018) The entity category security assertion markup language (SAML) attribute types [Preprint]. doi:10.17487/rfc8409.

Kaur, K., Gupta, I. and Singh, A.K. (2017) ‘A comparative evaluation of data leakage/loss prevention systems (DLPS)’, Computer Science &amp; Information Technology (CS &amp; IT) [Preprint]. doi:10.5121/csit.2017.71008.

Katakri 2020 - Tietoturvallisuuden auditointi työkalu viranomaisille (2020) Katakri 2020. Available at: (https://um.fi/documents/35732/0/Katakri+-+2020_1218.pdf/ab9c2d4a-5031-3670-6743-3f8921dce8c9?t=1608) Accessed: 07 June 2023.

Kissoon, T. (2021) 'Cybersecurity risk-management framework', Optimal Spending on Cybersecurity Measures, pp. 95–112. doi:10.4324/9781003200895-7.

Lakshmi, V., 2019. Beginning security with Microsoft technologies. [Berkeley, CA]: Apress.

learn.microsoft.com. 2022. Enable sensitivity labels for Office files in SharePoint and OneDrive. Available at: (https://learn.microsoft.com/en-gb/microsoft-365/compliance/sensitivity-labels-sharepoint-onedrive-files?view=o365-worldwide) Accessed 20 September 2022.

learn.microsoft.com. 2022. Screen capture protection. [online] Available at: (https://learn.microsoft.com/en-us/azure/virtual-desktop/screen-capture-protection) Accessed 20 September 2022.

learn.microsoft.com. 2023. Credentials processes in Windows authentication. Available at: (https://learn.microsoft.com/en-us/windows-server/security/windows-authentication/credentials-processes-in-windows-authentication) Accessed 15 June 2023.

learn.microsoft.com. 2023. Microsoft 365 compliance center. Available at: (https://learn.microsoft.com/en-us/microsoft-365/compliance/mip-easy-trials?view=o365-worldwide) Accessed 15 June 2023.

learn.microsoft.com. 2023. How does Azure Information Protection work?. Available at: (https://learn.microsoft.com/en-us/azure/information-protection/how-does-it-work) Accessed 15 June 2023.

Letia, I.A., Craciun, F. and Köpe, Z. (2001) 'Norms for DLP agents working in a warehouse scenario', Engineering of Intelligent Systems, pp. 728–733. doi:10.1007/3-540-45517-5_80.

Ma, Z., 2017. CPSec DLP: Kernel-Level Content Protection Security System of Data Leakage Prevention. Chinese Journal of Electronics, 26(4), pp.827-836.

Mojžiš, J. and Balogh, Š., 2020. Breaking Microsoft Azure Information Protection Viewer Using Memory Dump. Software Engineering Perspectives in Intelligent Systems, pp.913-920.

Michaelis, P. (2012) 'Enterprise mobility – a balancing act between security and usability', ISSE 2012 Securing Electronic Business Processes, pp. 75–79. doi:10.1007/978-3-658-00333-3_8.

Morgan, S., 2019. Cybersecurity Ventures. Available at: (https://www.herjavecgroup.com/wp-content/uploads/2018/12/CV-HG-2019-Official-Annual-Cybercrime-Report.pdf) Accessed 15 June 2023.

Patel, A., Taghavi, M., Bakhtiyari, K. and Celestino Júnior, J., 2013. An intrusion detection and prevention system in cloud computing: A systematic review. Journal of Network and Computer Applications, 36(1), pp.25-41.

P. Kesan, J., M. Hayes, C. and N. Bashir, M., 2013. Information Privacy and Data Control in Cloud Computing: Consumers, Priv Consumers, Privacy Preferences, and Mark ences, and Market Efficiency. Available at: (https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=4311&context=wlulr) Accessed 11 September 2022.

Polozova, E. and Anashkina, N. (2017) 'Analysis of Information Security Threats for developing DLP-systems', Production Engineering Archives, 17(17), pp. 24–27. doi:10.30657/pea.2017.17.05.

Prasad, R. and Moon, Y. (2022) 'Architecture for preventing and detecting cyber attacks in cyber-manufacturing system', IFAC-PapersOnLine, 55(10), pp. 2246–2251. doi:10.1016/j.ifacol.2022.10.042.

Purba, A. and Soetomo, M., 2019. Assessing Privileged Access Management (PAM) using ISO 27001:2013 Control. ACMIT Proceedings, 5(1), pp.65-76.

Reeves, A., Delfabbro, P., & Calic, D. (2021, March 10). Encouraging employee engagement with cybersecurity: How to tackle cyber fatigue. SAGE Open, 11(1), 215824402110000. https://doi.org/10.1177/21582440211000049

Robb, D., 2022. Top 8 data loss prevention (DLP) solutions: Esecurity Planet. eSecurityPlanet. Available at: (https://www.esecurityplanet.com/products/data-loss-prevention-dlp-solutions/) Accessed 15 June 2023.

Ruchini, C. (2021) Introduction to identity and Access Management, Medium. Available at: (https://medium.com/identity-beyond-borders/introduction-to-identity-and-access-management-2f3b80862647) Accessed 07 June 2023.

Schultz, E. (2009) 'Security Information and Event Management (SIEM) technology', Information Security Management Handbook, Sixth Edition, Volume 3 [Preprint]. doi:10.1201/9781420090956-c9.

Shah, N., Beutel, A., Hooi, B., Akoglu, L., Gunnemann, S., Makhija, D., Kumar, M. and Faloutsos, C., 2016. EdgeCentric: Anomaly Detection in Edge-Attributed Networks. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW).

Spanning Cloud, 2021. Data loss prevention: What is DLP & why is it important? Spanning. Available at: (https://spanning.com/blog/data-loss-prevention-dlp/) Accessed 07 June 2023.

SFS-ISO/IEC. 2013a. SFS-ISO/IEC 27001:2013 ”Information technology - Security techniques - Information security management systems – Requirements”. 2. painos. Helsinki: Suomen Standardisoimisliitto SFS.

SFS-ISO/IEC. 2013b. SFS-ISO/IEC 27002:2013 ”Information technology - Security techniques - Code of practice for information security controls”. 2. painos. Helsinki: Suomen Standardisoimisliitto SFS.

Swanagan, M., n.d. What Is A SIEM Solution? Benefits, Tools, & Strategies. Purplesec.us. Available at: https://purplesec.us/siem-solutions/ Accessed 11 October 2023.

Tahboub, R. and Saleh, Y. (2014) ‘Data leakage/loss prevention systems (DLP)’, 2014 World Congress on Computer Applications and Information Systems (WCCAIS) [Preprint]. doi:10.1109/wccais.2014.6916624.

Technavio, 2022. Data loss prevention market by deployment and geography - forecast and analysis 2022-2026. Technavio. Available at: (https://www.technavio.com/report/data-loss-prevention-market-industry-analysis) Accessed 20 September 2022.

Thakur, M. and Gaikwad, R., 2015. User identity and Access Management trends in IT infrastructure-an overview. 2015 International Conference on Pervasive Computing (ICPC).

Trellix, 2022. What is DLP and how does it work? Trellix. Available at: (https://www.trellix.com/en-us/security-awareness/data-protection/how-data-loss-prevention-dlp-technology-works.html) Accessed 15 June 2023.

version2hk, 2021. Safetica is a leader in softwarereviews' Data Loss Prevention Data quadrant. Version 2. Available at: (https://version-2.com/en/2021/10/safetica-is-a-leader-in-softwarereviews-data-loss-prevention-data-quadrant/) Accessed 15 June 2023.

What is an intrusion prevention system (IPS) - wallarm (2022) wallarm.com. Available at: https://www.wallarm.com/what/intrusion-prevention-system (Accessed: 09 October 2023).

What is Privileged Access Management (PAM)? (no date) Privileged Access Management (PAM) . Available at: (https://www.beyondtrust.com/resources/glossary/privileged-access-management) Accessed: 07 June 2023.