Computer Assisted Methods in Engineering and Science
31(2): 187–212, 2024, doi: 10.24423/cames.2024.492

This article belongs to the *Special Issue on Scientific Computing and Learning Analytics for Smart Healthcare Systems* edited by Dr. C. Chakraborty, Dr. S. Barbosa and Dr. L. Garg

Survey on Effective Disposal of E-Waste to Prevent Data Leakage

Akila VICTOR, Gurunathan ARUNKUMAR, Rajendran KANNADASAN, Soundrapandiyan RAJKUMAR, Ramani SELVANAMBI*

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
** Corresponding Author e-mail: ramani.s@vit.ac.in*

E-waste refers to electronic products that are of no use, not working properly, and either close to or at the end of their "useful life". Companies generate large amounts of e-waste when they replace old and outdated IT hardware with new technologies. Disposing of this e-waste is not so simple, as it may contain a significant amount of intellectual property in the form of data. Timely elimination of these records and data is very crucial to secure it. E-waste cannot just be discarded due to associated data security, confidentiality, compliance and environmental risks and policies. Even after deleting data, it can still be prone to social engineering attacks by malicious individuals. Data leakage is the unauthorized transmission of data from within an organization to an external destination or recipient, and it can be transferred electronically or physically. Nowadays, protecting data is of upmost importance for organizations. However, organizations still fail at destroying confidential data from their end-of-life equipment. This article focuses on how to detect data leakage and try to find those responsible for doing so. Different Data Loss Prevention (DLP) techniques that are currently being used by many organizations are discussed and some suggestions are provided for developing more consistent DLP and overcoming the weaknesses prevalent in these techniques. Furthermore, this article discusses various algorithmic, logical, and methodological foundations and procedures followed for large-scale data disposal, determining when the life of data comes to an end.

Keywords: e-waste, data leakage, data leakage detection, data leakage prevention, data disposal, data destruction, data security, end of life of data.

1. Introduction

The National Institute of Standards and Technology (NIST) Computer Handbook defines computer security as "protection afforded to an automated information system in order to attain the applicable objectives of preserving the integrity,

availability, and confidentiality of information system resources (includes hardware, software, firmware, information/data, and telecommunications)" [1]. The security concepts of integrity, confidentiality and availability of data are also known as the CIA triad.

Data is a very important and intellectual asset of any organization. Nowadays, every organization collects and keeps records of a significant amount of sensitive data that may include personal details of a person (such as name, mobile number, address, etc.), financial information (such as credit card and debit card details), employee details and more, depending on the organization and its business. With such vast amounts of data, there is always a threat of it falling into the wrong hands, and there are numerous malicious individuals or groups ready to exploit any opportunity and retrieve this data. Throughout its lifecycle, data is under constant threat of being leaked or exploited.

Data leakage is defined as "the accidental or unintentional distribution of private or sensitive data to an unauthorized entity" [2]. Data leakage may occur at any stage of the data lifecycle, posing a serious threat to companies. The number of such incidents and the cost endured due to those incidents continue to rise. Data transmission, including both outbound and inbound such as instant messaging, emails, website forms, browsing, file transfers, etc., is often unregulated and unmonitored during the transit. This primarily increases the risk of data leakage. Moreover, a lot of sensitive data is sent outside organization's premises, to employees, business partners and other clients. This increases the risk that confidential information will be within the reach of unauthorized individuals. Data leakage can occur not only through digital media but it can also be caused by employees. The leakage of sensitive information can have severe consequences for an organization, regardless of whether it is caused by unintentional mistake, malicious intent, by insider or outsider, or other reasons.

Data leakage can cause a lot of damage to all parties involved, including organizations, individuals and even governments. Losses due to data leakage incidents can be categorized as direct or indirect losses. Direct losses are tangible damages that are not difficult to estimate quantitatively. These include violation of standards and regulations (customer privacy protection laws) leading to fines, customer compensation fees or settlements, litigations involving legal action, declines in sales numbers, remedial fees, and the costs of investigation and restoration. Indirect losses are not easily quantifiable. Thus, they have a greater impact in terms of time, cost, and place. These losses include negative publicity, drop in share price, damage to the organization's name and reputation, exposure of very important assets (such as business plans, code, financial reports, etc.) to competitors, and customer abandonment.

To protect data from such leakages, organizations employ some leakage detection and prevention techniques. The aim of the data leakage detection systems is

to detect instances when any distributor's (owner of the data) sensitive and important data has been leaked by any malicious agents (third parties with whom the data is shared), and try to identify the agents who are responsible for the leaked data. Data leakage prevention systems (DLPS), on the other hand, are systems which have the ability to monitor, identify and protect the confidentiality of data. In addition they can detect any misuse based on standard rules and guidelines. The concept of data leakage prevention is relatively new compared to conventional and prevalent security solutions such as virtual private networks (VPN), intrusion detection systems, firewalls and intrusion prevention systems.

Although organizations try their very best to protect sensitive data when it is being used, stored or in transit, most of them fail to dispose and delete this data properly with the help of suitable end-of-life equipment. In recent times, there have been many high-profile data losses because individuals or large organizations knowingly or unintentionally release massive amounts of information and data when they dispose computers, mobiles and other devices that have reached their end of life. The results of research carried out over four years, examining "more than 1000 computer disks and 160 hand-held devices, have provided an insight into the very poor protection that both organizations and individuals give to data when they dispose of these types of equipment" [3].

In this digital age, we routinely store on the devices extremely sensitive data, ranging from bank account information to medical records, credit card details, and a variety of other highly sensitive data. Even if you have cleared your drives and deleted everything from your devices, savvy hackers know how to recover data, and there are software tools that can recover data from deleted files [4]. While some businesses may attempt to erase data on their own, it is preferable to contact specialists with the necessary knowledge and equipment to eliminate data and destroy hard drives.

The rest of the paper is organized as follows. Section 2 discusses various data leakage detection techniques. Section 3 classifies between traditional and new DLP techniques. Section 4 explains the importance of data disposal, studies all the available techniques and gives suggestions for conducting them. Section 5 concentrates on and presents the end of life of data and its last phase. Section 6 presents the case studies of data leakage in some organizations. Section 7 discusses the conclusion and future scope.

Study's contribution

The threat of data leakage from enterprises and organizations has become increasingly significant as the frequency of leakage occurrences and costs they inflict continue to rise. Various methods and approaches have been created to

handle the problem of data leakage prevention, given the magnitude of the problem. This paper discusses the various methods and models proposed by different researchers to understand the critical problem of data leakage and how to prevent data loss.

Additionally, the paper discusses various methods for detecting and preventing data leakage to protect the valuable assets of individuals and organizations and how to effectively dispose unwanted data.

2. Data leakage detection

2.1. Challenges faced during data leakage detection

One of the most difficult aspects of data loss mitigation is that there are so many causes of data loss in an organization, and there is no single tool or easy solution that can appropriately handle them. However, in order to manage the risks, a solution that addresses the various causes of data loss must be established.

2.1.1. Encryption. It is the initial stage of the data transfer phase. Data leakage prevention (DLP) is hindered by encryption and excessive amounts of digitally transferred data. Due to encryption, the security of the data is maintained. It also makes it difficult to detect data leaks in encrypted systems. Some security mechanisms, such as network-based ones, are not very effective against data leakage caused by stealthy software, since attackers can use strong encryption while transferring data, and the key and encrypted algorithm could be available to data leakage detection administrators. Further, DLP mechanisms must be provided for extra leakage detection of data in encrypted email and file transfer protocols, including SSH file transfer protocol (SFTP).

2.1.2. Access control. In a computing environment, access control is a security approach that regulates who or what can view or utilize resources. It is a basic security concept that reduces the risk to a company or organization. Physical and logical access control are the two types of access control. Access to campuses, buildings, rooms, and physical IT assets is restricted via physical access control. Connections to computer networks, system files, and data are all restricted by logical access control. Access control was one of the initial techniques used for data leakage prevention solutions. However, it is now no longer popular. While this method is appropriate for data at rest, it poses challenges when enforced during data communication. Especially, when the data is fetched from the database it becomes tough to implement this technique. The structures of access control are regularly reviewed with low expectations in mind.

Data sharing between several subjects can be controlled using access control policies. Such policies must be fit for their purpose if high-assurance data security is to be achieved.

2.1.3. Semantic gap in DLP. A common problem in strategies, such as forensics and signatures, is the lack of connotation in the activities under monitoring. When information leakage is defined through the methods of the communicating corporations and the information exchanged within the course of the communication, smooth pattern matching or control access schemes cannot infer the character of the communication. Therefore, information leak prevention mechanisms need to keep track of who, what and where to protect against complex information leak scenarios [5].

2.2. DLD techniques

2.2.1. Watermarking: embedding & extraction. Watermarking is a method of detecting and preventing data leak. Using this technique, each document or record receives a separate and unique identity. The reason being that if a copy of the original data is discovered at unknown suspicious locations, the probability of identifying the responsible agent becomes very high. This technique is highly useful for companies that are involved in creating digital information security products and other market applications [6]. However a few changes can also occur in the original data. The major drawbacks of watermarking method include:

- It introduces a few variations or modifications into records through modifying some of their attributes, therefore making the records less sensitive. This change of records is called perturbation. On the other hand, in a certain situations, actual records cannot be changed under any circumstances. For example, when an agent requires precise revenue figures to carry out payroll, the salary cannot be changed here.
- Another problem is that if the recipient is malicious, watermarks can be easily damaged.

2.2.2. Steganography. Steganography is a method of concealing a secret message within a bigger one in such a way that the presence or contents of the hidden message are undetectable to others. A plain text communication can be hidden in one of two ways: steganography conceals the message's existence altogether, while cryptography renders the message illegible to others through encryption. Steganography is a method of communication that is private, secure, and sometimes even harmful.

2.3. Survey of methods for data leakage detection (DLD)

Some effort is placed in reviewing research papers on DLD and prevention methods considering different states of data. The DLD approach is basically content-based or hybrid one, with variations in data state handling, deployment location, and techniques used.

Shu *et al.* [7] put forward a solution for DLD by implementing a fuzzy fingerprint technique. This technique basically provides more safety and privacy to data during DLD operations. Using this technique, the owner of data can perform detection without revealing the data to the providers. It is mainly beneficial for internet service providers as they can detect traffic and provide DLD solutions to customers. In addition, customers can identify suspicious data and request the provider to detect any leakage of data. Many models have been proposed by the authors to give an overview of security of data and leakage detection. The efficiency and accuracy of the technique were analysed through various experiments, yielding great results.

Costante *et al.* [8] provided a DLP framework by combining both anomaly and signature-based solutions. This framework detects and identifies threats arising from suspicious user activity and behaviour. Usually, in an anomaly-based system, the engine autonomously learns user activities and behaviour so that when an insider carries out suspicious transactions, it gets flagged and the transaction process fails. However, in this framework, the authors extend this to more security and privacy by updating traces of attacks used in transactions before they can cause any harm. This framework automatically builds and updates traces based on alert feedback from the administrator. Experiments showed positive results, hence reducing the amount of data leakage.

Papadimitriou and Garcia-Molina [9] studied modest techniques for identifying data leakage in records and provided a methodology that deals with agents "guilt" and focuses on how much the agent is guilty. The presented algorithm helps distributors as it can identify and detect data leakers. By comparing the data of different agents, the probability of identifying the agent who leaked the data is quite high, even with fake objects in the set. It is a content-based approach, but it has its limitations including the omission of consideration of scenarios involving sufficient data leakage.

Shu *et al.* [10] provided solutions for data leakage patterns that are complex in nature. In order to compare the similarity of two separate data patterns, a sampling algorithm was used. The system can effectively detect customized leaks. Many models have been proposed by the authors to give an overview of security of data and leakage detection. Inferences drawn from the experimental analysis reveal that this method is quite effective in identifying scenarios involving data leakage.

3. Data leakage prevention

3.1. Traditional DLPs

Over the years, various methods have been proposed for data leakage prevention. Firewalls use techniques available for filtering network packets [11]. Still, they can only filter packets based on IP addresses, which will only be helpful if the IP address of the attacker is known. Some of the traditional methods use deep packet inspection (DPI) [12]. DPI architectures find anomalies in the network packets and traffic, and alert the administrator. The process of DPI uses a pattern definition language interpreter (Fig. 1) to detect and prevent attacks from unknown protocols. It also reassembles the network packets that arrive in the wrong order. Traditional data leakage prevention techniques are as follows:

- Watermarking. It is the process in which a unique code is embedded in confidential documents.
- Fingerprinting. In this process, sentences are converted into hash values and stored in a database. After the database creation, each document's hash values are compared with the ones in the database. If a certain number of hash values match, it implies that the document is confidential.
- Intrusion detection system/intrusion prevention system (IDS/IPS). The two approaches for this system are:
  - A pattern matching system: This approach is used to defend a system against known attacks. The signatures of known attacks are loaded in the system. Hence, the major disadvantage of IPS is that unknown attacks will not be detected and it can only protect a system from known threats.

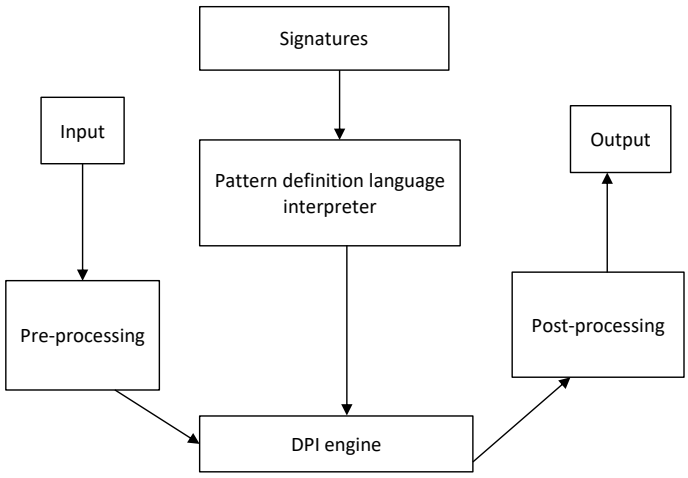

Fig. 1. Deep packet inspection flowchart.

  - Anomaly-detection system: This approach requires an administrator to either authorize or allow the intrusion detection and prevention system to learn about regular activity. It can be used to detect abnormal activities that differ significantly from regular activity [12].
- Anti-malware. Software designed to counteract malware that collects, damages, and accesses confidential data without authorization.
- Firewalls. These are software/devices that can block networks based on given rules and protect networks from attacks of unauthorized sources.
- Security information event management. This is a tool that maintains logs of all networks and storage devices. It can analyze logs and detect suspicious activities.

3.2. New techniques for data leakage prevention

3.2.1. Techniques based on data states. Tahboub and Saleh [4] proposed a new architecture based on deep content inspection (DCI). In this architecture, they provided solutions for different data states such as:

- Data at rest (DIR): DIR is basically all data stored in computer storage. To protect data from being accessed and tampered with, data encryption and access control are usually used. Content discovery is needed for these methods to work effectively, which can be achieved by using the content discovery features of DLP.
- Data in motion (DIM): It is the data that is being transferred through a network. This data is monitored when it is sent across networks using communication channels such as emails, messages, known protocols, and unknown protocols.
- Data in use (DIU): DIU is the data that users interact with. To prevent data leakage, DIU tools monitor the following:
  - screen capture and copy-pasting sensitive data,
  - transfer of confidential data from one storage device to another,
  - printing or sharing sensitive information.

DLP methods classify content based on rules. The two approaches used are:

- Content matching: It is used to detect data loss incidents by matching keywords, file types, regular expressions, etc. It uses match-join algorithms to match the content.
- Learning method: To determine the confidentiality of a given message, DLP systems use machine learning algorithms such as the vector space model (VSM). Messages are represented as vectors, and vector features represent the frequency of occurrence of particular vectors, and a model is built [12].

Ghouse and Nene [13] discussed the progress made in data leakage prevention. The paper briefly describes the vector-based method. In this method, VSM is used. Sentences are converted into vectors, and the vector space is the result of the words within the sentence. The angle cosine value is calculated to check the similarity between two sentences from different documents.

3.2.2. Techniques based on graph neural networks. Graph-based approaches are ideal for text-based categorization. The major advantage of the graph-based approach is that it can categorize based on content, context as well as semantics, which helps in achieving good accuracy.

Ghouse *et al.* [14] proposed a novel D-SeGATe (data leakage prevention using secure gateway analysis technique) architecture. This architecture is only applicable between a secure and an insecure domain. In this architecture, a common key is generated by the key management server (Fig. 2) using the trusted platform module (TPM) and shared with all intranet devices. The TPM ensures that the key is generated automatically, it is random and not preconfigured. The same key is used for encryption and decryption across all intranet devices. New keys are generated after a fixed period of time and circulated to all the intranet devices. The key is stored in a central database.

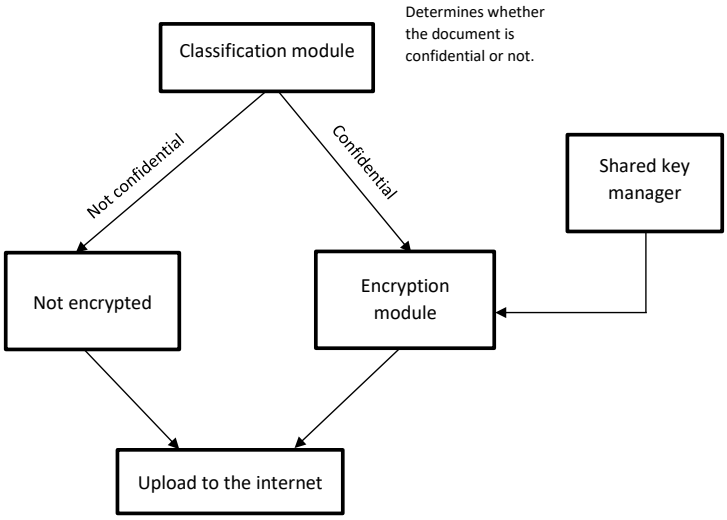

Fig. 2. SeGATe architecture for file upload.

Data leakage approach for two types of data:

- Confidential data: In case of confidential data, categorization takes place at the start of the secure domain. The classification module analyses, preprocesses, and classifies the testing data. After the data is classified as

confidential, the document is encrypted by using the common shared key. The receiver will not be able to read/ decipher the data as it is encrypted and the user does not have the shared key. Once the system receives the encrypted document, it can decrypt it using the common key provided to all intranet devices.

- Non-confidential data: The classification process is the same. After classification, the document is transmitted in its original form and received in the same form without any encryption. This method is helpful as it avoids any additional delays. That is why the classification model plays an important role in determining whether encryption is needed or not.

The proposed architecture in the secure gateway (SeGATe) is the continuation of work in [14] with proper implementation and results. The main focus of this architecture is the data flow from internet end devices to the internet. This architecture consists of two modules: the classification module and the encryption module. The classification/categorization module uses the graph convolutional network (GCN) model because it provides better performance for text classification. The encryption module uses the AES encryption method [15]. The data that is transferred from intranet devices to the internet passes via SeGATe. To protect the intranet devices from the risk of viruses, a data diode is implemented. In this architecture, a document undergoes analysis in the classification module to check for confidentiality, and a file is encrypted based on its content. During decryption, the file is decrypted using a suitable decryption key from the key management module, which is only available to the security administrator. The results of this architecture are phenomenal across various datasets. Training and validation loss decrease exponentially with an increasing number of epochs. The accuracy is over 96% in all datasets and various window sizes.

3.2.3. Technique based on named entity recognition (NER). NER, also known as entity identification and entity extraction, is a system that identifies and classifies words into categories by using linguistic grammar-based techniques and statistical models. Gómez-Hidalgo *et al.* [16] proposed a data leak prevention system using NER. NER is typically used in fields such as journalism and biology. NER can be implemented using supervised learning. The prototype that the authors created employed NER techniques supported by **Freeling**. This prototype prompts the user to enter confidential information, such as personal details and credit card information. The predefined categories store this information, and whenever it finds a website sharing data that matches the pattern, it can alert the user that a data leak is taking place. The software also includes a feature of automatic pattern learning; basically, the system continuously looks for patterns in the transmitted data. It sends the user a report and an alert

if a new pattern is found, and the user can then choose to block or allow this activity. The system continuously learns new patterns and stores them in a separate list. The patterns have an expiry date, and if the user wishes for the new patterns to be matched, they can allow this feature. This software can be used as a personal firewall, i.e., trained over time with the data of a single person. A drawback to this system is that the data to train this data is limited, and the training set needs to be improved to get better results.

3.2.4. Technique to prevent physical transfer of data. Many companies prevent sharing data using external storage devices to mitigate data leakage. The system proposed in [17] consists of a file system mini-filter driver that stops input-output requests sent by the operating system (OS) when certain events occur. This system blocks all external storage devices that can be used to transfer files/data. It creates a list of these storage devices, and if an input/output (I/O) request is generated from these devices, the request automatically gets blocked, and the IRP, suspecting a data leak, sends a STATUS_ACCESS_DENIED message. Another feature of this system is that it blocks any process that attempts to access sensitive and confidential information. The file path for sensitive information is stored in the system during setting up the system. After the process is blocked, a network rule is added so that network data cannot leave the process.

3.2.5. Technique to tackle purposeful evasion attack. Traditional DLPs are only effective for accidental data leaks. Hence, we need other DLP solutions to tackle purposeful evasion attacks. Mustafa [18] stated that malicious data leak (MLD) is one of the most serious threats that cannot be tackled by conventional DLP solutions and proposed a "3-D correlation" (Fig. 3) as an effective security system against data contamination. MDL involves scenarios where a human, malware or bot (agent/ actor) causes deliberate exfiltration of confidential information without authorization. An advanced persistent threat (APT) is an MDL where the agent is non-human, such as malware, bots, etc. This paradigm correlates actors (Moles), information, and operations using identity, roles and

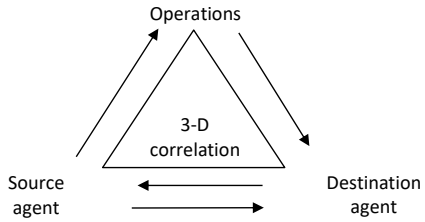

Fig. 3. Definition of 3-D correlation.

the security clearance of actors. This paradigm consists of four major components:

- Actors. These are the moles and the sources of evasion attacks. They consist of identity, access roles, and security profiles.
- Information elements (IE). IE refers to confidential and important data. It has two key attributes: confidentiality level and classification type.
- Operations. Actions that can take place on the data (IE).
- Accessibility map. It is the most important component specifying which IE can be accessed by which actors. Any actor that attempts to access an IE that it does not have access to is identified as a potential violation.

This paradigm can also automatically classify, correlate, and identify potential violations on zero-day documents [18]. Evasion attacks are cyber-attacks against DLP systems. These attacks are very difficult to detect as they exploit the underlying vulnerabilities of DLP systems at algorithmic and systematic levels. There are two types of evasion attacks:

- Identity-based evasion attacks: In such attacks, the mole is either a person within the organization or an APT with proper identity and access rights. Once the mole acquires the identity and access rights, it becomes difficult for the DLP system to defend against the attack, as a major aspect of DLP is to control user access.
- Content-based evasion attacks: These are very challenging attacks as they target the most vulnerable aspect of the algorithm. They involve manipulation of the structural, lexical, or temporal composition of content and attack on the known vulnerabilities in DLP algorithms. There are various types of content-based evasion attacks, and they target fingerprinting algorithms, latent semantic indexing (LSI) algorithms, natural language processing (NLP) algorithms, etc. [18].

3.2.6. Technique to prevent data leakage while collaborating. A major threat to DLP system is data sharing. When collaborating with others, the risk of data leakage increases because of data sharing. Lu *et al.* [19] have come up with a collaborative graph-based mechanism for distributed big data leakage prevention to prevent data leakage. When multiple parties share their data during collaborative efforts, it is important to detect and prevent data leakage. However, training models while preserving the privacy of collaborators is a challenge, and a new architecture is proposed to address it. The proposed architecture consists of two modules: privacy-preserving collaborative training (to train the model using the dataset of all the data providers) and graph matching. Privacy-preserving collaborative training involves employing weighted graphs rather than word vectors, as the first ones capture more contextual structure. These weighted graphs are

merged into one and then masked to encrypt the information. For masking, the authors use the hash function (SHA256) on each node for encryption (Fig. 4).

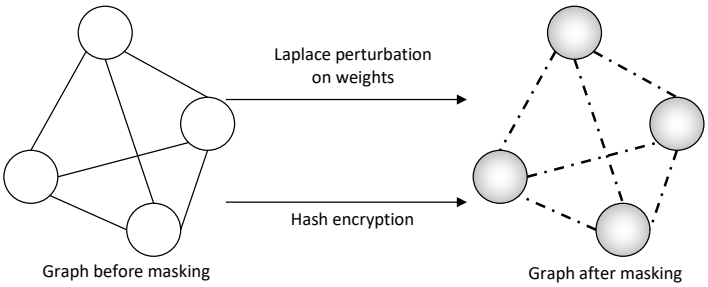

FIG. 4. Graph masking technique.

The encryption process does not affect the training and detection. But, if there are common nodes with the same values, the hash values will match. Therefore, Laplace mechanism is used. The proposed architecture uses graph-matching detection. A score-matching algorithm is used to compare the pre-trained graph G and tested graph Gd. Graph Gd is traversed using breadth-first traversal. Whenever a node or edge matches, a reward is given, and in case of a mismatch, a penalty is imposed. The final sensitivity score comprises the rewards and penalties and can detect transformed data leakage. The overall accuracy of this architecture is better than that of a context-based model (CoBAn) presented in [20]. For a test size of over 2000, the accuracy is over 0.9.

3.3. Applications of DLP systems

Michael [21] briefs about four data leakage prevention methods used in cloud computing:

- My DLP is open-source software designed for data leakage prevention. It is easy to use and provides extraordinary performance in combating data leakage. It can block outgoing information through e-mail and archive suspicious documents. Additionally, it can disable print and transfer functions so that data transfer cannot take place using external storage devices such as USB, hard disks, or printers.
- Watermarking data. A unique signature is robustly embedded into files, images, videos, etc. This makes it easier to track files, check for any data leakage or alterations, and maintain data integrity. Watermarking also helps users to provide proof of ownership by verifying the signature.
- Michael [21] proposes a new technique called Impeding Data Leakage (IDL) method. The method is extended to avert the outflow of the records throughout conversation inside the cloud services. IDL is based on swarm intelli-

gence technologies like artificial bee colony (ABC) and ant colony optimization (ACO) to redirect paths and prevent data leakage. The transmission route is generated by using ABC. After transmitting requests using the path generated by ABC, server responses are sent to the client using ACO. If the leakage takes place, the ACO chooses the next optimized path to transfer the statistics. This technique helps in finding the optimal data transfer path using stigmergy.

- Cloud services provide protection of information and intrusion detection systems on the cloud to make certain that the product used by the user is safe and the data should be recoverable from the cloud.

4. Data disposal

Data needs to be appropriately managed throughout the complete information lifecycle, from capture to destruction, because it is prone to leakage throughout its lifecycle. Data of any kind moves through throughout its useful life and ultimately it is either stored for future use or destroyed when it is no longer needed, as shown in Fig. 5. Hence, information destruction becomes a necessary part of an extensive information control program.

Fig. 5. Data lifecycle.

Data disposal or destruction is the process of eliminating information so that it becomes unreadable (for paper documents) or unrecoverable (for digital documents). In the case of digital documents, this means that neither an operating system nor an application can read it. To destroy data, we can either overwrite the current data with random data until it becomes impossible to retrieve the original data or simply destroy the storage medium.

When files are deleted incorrectly or inadequately from storage media, it is often still possible to rebuild or recover data. A study led by British Telecommunications and other partners, revealed that from 52% of the total disks collected for the study, they could recover information from which the organization that had previously owned the disks could be identified and from 51% of the disks, they could recover information from which an individual could be identified [33]. Data on only 31% of the disks had been disposed of to an extent where it was not easily recoverable. Another study of second-hand hard drives sold over the internet, revealed that about 10% of these hard drives still contained personal information. There have been a lot of such incidents where due to improper

data disposal, a lot of sensitive data has been leaked. Here are few such incidents [32]:

- Idaho Power Co. (Boise, ID): "Four hard drives sold on eBay in 2006 contained hundreds of thousands of confidential documents, employee names and SSNs, and confidential memos to the CEO".
- At many warehouses across the USA, photocopiers used to copy sensitive information such as medical records, bank records, and even documents from police departments were found, ready to be resold, without their hard drives wiped properly.
- Loyola University: "A computer at Loyola University containing names, Social Security numbers, and some financial aid information for 5800 students was disposed of before the hard drive was wiped".

And more such incidents still keep on happening due to lack of knowledge and insufficient policies and procedures for proper data disposal. Setting up policies and procedures that govern the management and use of data enables organizations to manage data more efficiently.

4.1. Data disposal techniques

Data disposal can be of two types: destructive and non-destructive [22]. A destructive technique of data disposal is one which involves the physical destruction of the storage unit. On the other hand, a non-destructive technique preserves the physical media but the data it holds is erased by processes like overwriting or degaussing.

Data disposal methods or techniques are referred to by using various terms such as secure sanitization and data sanitization. According to the National Institute of Standards and Technology (NIST), data sanitization is used to refer to "all data elimination methods, including block-by-block overwrite, drive internal secure erase (SE), and physical, chemical, thermal, or magnetic destruction". There are several techniques for data disposal but none of the methods can be considered perfect or fool proof. Organizations and individuals have to choose a suitable technique according to their requirements.

According to NIST's Guidelines for Media Sanitization publication [23], the following actions can be taken to destroy (or sanitize) data:

1) **Clear.** This involves logical techniques applicable to user-addressable storage. Clearing provides protection against simple non-invasive data recovery techniques. It can be carried out through standard read/write commands to the storage device. Resetting the device to the factory state (performing a factory reset) or rewriting with a new, non-sensitive value are some ways of clearing a device.

2) **Purge.** It includes applying state-of-the-art laboratory techniques that render target recovery infeasible. These techniques can be either physical or logical. Non-destructive purging includes methods such as cryptographic erase, block erase, and overwriting. These can be carried out by using standardized commands for device sanitization, which are media-specific to bypass the inbuilt abstraction found in classic read-and-write commands. Destructive data disposal through methods such as shredding, disintegrating, incineration, degaussing, and pulverizing also renders a device purged.
3) **Destroy.** This also includes using state-of-the-art laboratory techniques to render target recovery infeasible but it also leaves the media unusable for further data storage. Methods to completely destroy media include disintegrating, pulverizing, melting and incinerating. To destroy flexible storage media such as diskettes, shredding is a destructive technique that can be used. Generally, clear and purge actions are used first to sanitize the media. The destroy action is only used if the storage unit fails and other clear or purge methods cannot be applied, or if the clear or purge methods used cannot be verified [23].

We will now look into the various techniques mentioned above in detail:

- **Delete/Reformat.** Deleting a file is the most rudimentary action a person or an organization performs to dispose of their data, but it does not actually destroy the data. It only removes it from a directory. The data remains on the memory chip or the hard drive of a device. In the case of reformatting a disc, the same holds true, as reformatting does not completely dispose of the data. Rather, it places a new file system over the existing one.
- **Data wiping** is the process of overwriting data on any electronic storage media such that it cannot be read anymore. It is carried out by using a bulk wiping device to which the storage medium has to be physically connected. Data wiping allows the reuse of the media that has been wiped without losing any of its original storage capacity. Data wiping is a long, time-consuming process, often taking an entire day for a single device. Thus, it is not a very practical solution for an organization with several devices that need to be wiped. However, it proves to be a very beneficial solution for individuals.
- **Overwriting** data is a form of data wiping. By overwriting data on an electronic device, we simply write over the existing data with a pattern of ones and zeros. This pattern can be random or predefined. For high security data, overwriting is performed multiple times to ensure total data destruction. One of the major concerns with overwriting is the presence of a bit shadow left after the process is performed. A bit shadow is the imprint of the overwritten data that can be observed under an electron microscope.

In addition, using an electron microscope to recover such data is very costly and time-consuming which makes overwriting a strong contender while choosing the right data disposal technique for low-risk businesses. However, it is not recommended for high-security operations [24]. Overwriting is the most common technique used for non-destructive data disposal. However, it can be time-consuming and is only applicable to devices that have not been physically damaged. Another concern with overwriting is that there is no security protection during the process. It also does not work on hard drives that use advanced storage management components. The DoD 5220.22-M is one of the standards used in the industry for data sanitization but its last revision was conducted in 2006, and therefore it may no longer be relevant due to technological advancements since then. NIST has established some newer standards for overwriting data in NIST 800-88 which, if followed, reduce the chance of data recovery from overwritten data (Table 1).

- **Erasure** is a method very similar to overwriting. Erasure completely destroys all data stored on a hard drive. Often, a certificate of destruction is also provided to show that the data on the storage media has been successfully erased. Erasure is useful for organizations when the equipment they use, such as laptops, desktops and data centres, is purchased off-lease. It is also a suitable option for those who want to redeploy or reuse their hard drives for storing other data.
- **Degaussing**: In this method, a high-powered magnet is used to disrupt the magnetic field of an electronic medium. This in turn destroys the device's data. While degaussing can effectively and quickly destroy the data on a device storing a large amount of information or sensitive data, it has some **drawbacks**. Degaussing rearranges the structure of the hard disk drive (HDD) and renders a device inoperable. It is considered a destructive disposal technique. This method is not suitable if devices, such as a laptop, computer or mobile phone, are to be reused. Additionally, it also

Table 1. DoD 5220.22-M vs NIST 800-88 Rev. 1 [34].

Description/Parameters	DoD 5220.22-M	NIST 800-88 Rev. 1
Last revised date	2006	2012
Considerations for new technology (e.g.: SSD)	No	Yes
Recommended overwriting passes	3	1
Applicable sector	Government	All organizations
Recommends specific data erasure methods	No	Yes
Is the method of erasure verifiable?	Yes (only for HDD)	Yes
Maximum ecological conservation	No	Yes

becomes impossible to verify if all of the data has been destroyed as one cannot check it due to the inoperable hard drive. The verification of proper data destruction, in this case, can only be done by using an electron microscope, which is expensive and impractical unless there is some high-security information being destroyed. Furthermore, as technology advances and the size of hard drives increases, it is observed that degaussing capabilities are diminishing over time.

- **Physical destruction** of data can be done in a number of ways, such as melting, incinerating, pulverizing, disintegrating, disk shredding or any other method that renders the physical media unreadable and unusable. This method provides the highest assurance of complete data disposal. This makes the method very advantageous, as it is absolutely impossible to recover or reconstruct the data from a drive or disk that has been physically destroyed. However, the main disadvantage of physical destruction is that it is costly, due to the excessive capital expenses involved.
- **Disintegration, melting, pulverization and incineration.** These sanitization methods are intended to destruct data completely and, in the process, also completely destroy the device. They are typically carried out at an outsourced metal destruction or licensed incineration facilities with the specific capabilities to perform these activities effectively, securely, and safely [25].
- **Shredding** is the most secure and cost-effective way to dispose of all types of end-of-life devices such as media tapes, solid state drives, and hard drives. It is also very effective for tablets, smartphones, motherboards, thumb drives, optical drives and credit card swipe devices. Shedding reduces electronic devices to very small pieces that cannot be put back together for data recovery. Organizations sometimes mix non-sensitive shredded material with sensitive material, which makes the data impossible to be retrieved or reconstructed.
- **Total annihilation.** Some solid-state drives can automatically overwrite data or even physically destroy itself through the press of a button. In such solid state drives (SSDs), a current is applied to the NAND flash memory upon pressing of a designated button, which physically destroys the drive and produces only a puff of smoke when the damage is done.
- **Outsourcing.** While organizations can outsource data destruction, this approach has many risks. Some potential issues with outsourcing are transit breaches, third parties using outdated or uncertified methods of destruction, or even data theft. If an organization wants to outsource data destruction, they should perform some fundamental steps such as deleting or formatting to mitigate risks.

- **Encryption.** Usage of encryption tools is one of the recent disposal trends in the industry for data disposal. These methods use an encryption key for handling drive requests such as erasing or keeping data. The encryption key resides in the hard drive. In the past, this process used to take 30 to 60 minutes, but nowadays, secure discarding of data takes up only a few microseconds, as changing the encryption key is all that is required [26]. This method also ensures that nobody except the individual aware of the original master key (or a password) can perform the operation.

4.2. Choosing a disposal technique

Some data destruction methods are more complicated and they take more time or are more resource intensive than others. Hence, it is common to choose a method based on the underlying sensitivity of data being destroyed or the potential damage it can cause if it is recovered or accidentally leaked, as shown in Fig. 6. Smart healthcare is evolving as a result of improvements in smart networks and the cloud computing paradigm. However, obstacles persists, such as storing sensitive data in untrustworthy and uncontrolled infrastructure and ensuring safe medical data transmission, to name a few. The watermarking's rapid development opens up new possibilities for smart healthcare.

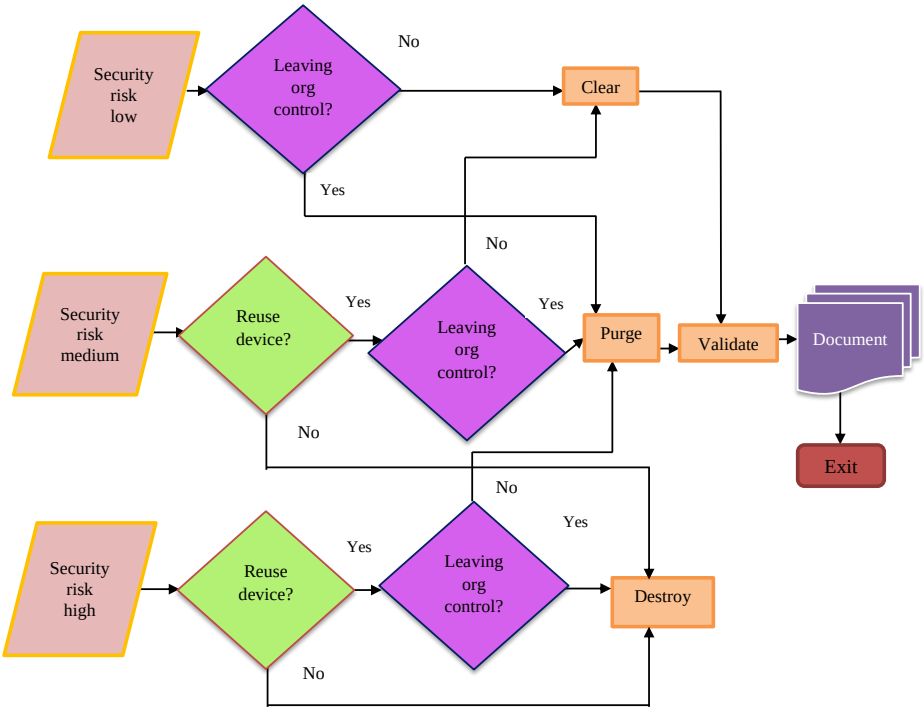

Fig. 6. Sanitization decision flow.

Some factors that should be considered while choosing a data disposal technique are:

1) **Time.** Each method discussed above operates on a different timescale. An organization or an individual must identify the amount of time they are willing to invest in data disposal. The frequency at which data disposal is performed must also be kept in mind because it influences expenses associated with carrying out data disposal.
2) **Cost.** Sanitization actions such as clearing and purging allow to reuse the device after disposing of current data whereas destructive methods physically destroy the equipment and render it unusable. A company or an individual must decide if they can afford destruction of devices and buying new devices or are they prefer to reuse older devices for new purposes. This also influences their choice of data disposal technique.
3) **Validation and certification.** If an organization or an individual must dispose data to comply with regulations or legal requirements, they must choose a disposal technique that is approved for such standards.
4) **Risk level of information.** Information can be classified into various risk categories. For very low risk information, deleting electronic files or using a desk document shredder can be a suitable method. However, such types of destruction can be undone by determined individuals, making these methods unsuitable for more sensitive data [27]. For more sensitive data, more effective destruction methods may be required at a more granular level to be used to ensure that data is completely unrecoverable.

4.3. Data disposal standards and policies

There are many data breach regulations, such as the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, the EU Fair and Accurate Credit Transactions Act of 2003 (FACTA) in the US, the EU General Data Protection Regulation (GDPR), and Gramm-Leach-Bliley Act (GLBA) in the US, but very few standards for destroying or sanitizing data.

The DoD 5220.22-M manual, also known as the National Industrial Security Program Operating Manual (NISPOM), was used earlier. However, as it does not specify any particular sanitization technique, and it is no longer considered acceptable. The NIST guidelines are very well-known in the field of information security and they recommend good standards for data destruction. However, adherence to these recommendations is not compulsory. An organization needs to cultivate a culture of compliance to gain the trust of its clients. In order to have such a culture, the right policies for secure data disposal and destruction must be established. This is very crucial, as there are no compul-

sory standards to be adhered to for secure data disposal. Setting up policies for data disposal holds employees accountable for their actions. Such a policy must include:

- Mentioning of personnel overseeing the process of data disposal and destruction.
- Defining specific best practices to ensure that secure data disposal and destruction methodologies are used. This should be known to the aforementioned personnel appointed.
- Steps for handling media devices that have reached the end of their useful life for the organization, but do not need to be destroyed completely.
- Requirements for updating asset inventory lists.
- Information outlining steps if the data disposal policy is not complied with.

4.4. Data breaches due to poor data disposal

There are 1 in 4 data breaches due to negligence. These breaches can lead to financial losses, reputational damage, identity theft, etc. Some of the real-time incidents that have occurred in the past are listed:

- NHS Trust Hospitals: "In June 2012, the University of Brighton and Sussex fined NHS Trust Hospitals £325 000 after the discovery of highly sensitive personal data belonging to tens of thousands of patients, including information related to HIV and genitourinary drugs (GUM) patients, on hard drives sold on an internet auction site" [35].
- Information Commissioner's Office (UK): "A research through the ICO found that one in ten second-hand hard drives bought on-line may contain residual private information. The ICO requested a PC forensics company – NCC Group – to source around 200 hard drives, 20 memory sticks and 10 cell phones. These devices were examined, initially without any extra software, and then analysed using forensic equipment freely available on the Internet. The studies determined that, while 52% of the hard drives investigated were unreadable or wiped of data, 48% contained data and 11% of this data was personal. In total, 34 000 documents containing private or company statistics were recovered from the devices. At least two hard drives contained enough statistics to allow a person to steal the previous owner's identity." [35].
- From the studies on handheld gadgets, 23% of operational devices contained organizational records that were recoverable, and 19% of those gadgets had recoverable personal records. Other examples of found data included personnel's personal information including salary, home address,

national insurance number, contact number as well as current business plans for large multinational companies, including turnover breakdown by factory [28].

5. End of life of data

End of life of data basically involves removing duplicate data, useless data or the data which is no longer of any use. One such example is in cloud environments, where to maximize resource usage we need to relocate data and remove previous instances of data usage following the rules/regulations for it. Also, sometimes there may be duplicate data presence and this could result in decreased efficiency of the device, so removing such data permanently becomes required. Once data is considered of no use, proper guidelines of data disposal to remove it from all the places must be followed.

There should be no chance of data being retrieved by anyone through any means. The following points have to be considered for the-end-of-life of data [29]:

- Inactive, duplicate, useless data that has reached its full lifespan must be destroyed according to the existing rules and regulations.
- Data centres, servers and all the other places where there is a vast storage of data should provide suitable end-of-life methods, such as disk shredding, demagnetization, or disk replication to their clients to prevent any sensitive data leakage to the public.
- Ensure that all the unnecessary data is permanently removed and there should be no chance of data retrieval from the storage medium to prevent any disclosure of sensitive information.
- Make certain that the data present in the cloud is also removed through proper methods. Ensure that information cannot be disclosed or recovered by any means.
- Assure proper wiping of data from hard drives and other storage devices.

This marks the end of the data lifecycle. It is not an easy process but is the most important part in the data lifecycle, given the present situations where even the tiniest amount of data can uncover an individual's personal life or an organization's business model.

6. Case studies

The case study involving student records at Strathmore College highlights the importance of protecting sensitive records. A staffer at Strathmore Secondary College mistakenly disclosed more than 300 students' records on the school's

intranet in August 2018. Students' medical and mental health issues such as Asperger's, autism, and ADHD were documented in these records. The exposed pupils' medications, as well as any learning or behavioural challenges, were also documented, according to The Guardian. The records were on Strathmore's intranet for about a day in total. Students and parents could access and/or download the information during that time [30].

Customer records were compromised through an unsecured database, according to Veeam. The Shodan search engine indexed an Amazon-hosted IP at the end of August 2018. On September 5th, Bob Diachenko, director of cyber risk research at Hacken.io, came across the IP and instantly deduced that it resolved to a database that had been left unprotected due to a lack of a password. Veeam, a backup and data recovery firm, had 200 terabytes of data exposed in the exposed database. Customer records, including names, email addresses, and some IP addresses, were among the exposed data [30].

The city of Dallas suffered major data losses due to staff incompetence in a series of incidents in March and April 2021. An employee erased 8.7 million critical police files, including video, pictures, audio, case notes, and other items, that the Dallas Police Department had collected as evidence for its cases. The family violence unit owned the majority of the erased files [36].

With rapid technological improvements in the medical industry, clinical data is constantly being created for processing and monitoring. This biomedical data must be translated into usable knowledge in the actual world. The primary objective of high-quality healthcare is biomedical research, which is achieved through constant clinical monitoring and accurate treatment and diagnosis [31]. To protect such sensitive healthcare data, advanced data leakage detection and prevention techniques are required.

7. Conclusion and future scope

The purpose of this paper was to review all effective techniques and methodologies used to detect and prevent data leakage at various phases of the data lifecycle. This paper also explains the importance of proper data disposal at the end of its lifecycle and lists possible methods for accomplishing this. The paper covers all such methods that an individual or an organization can follow to ensure the safety of their data. Based on the analysis presented, it can be concluded that data should be securely managed at all stages of its lifecycle. This paper discusses the importance of data and how data can be effectively destroyed to prevent tracing by hackers. As technology advances, more threats to data security will also emerge and thus newer, more effective methods and techniques should be developed in the future to effectively handle and destroy data.

References

1. B. Guttman, E.A. Roback, *An Introduction to Computer Security: The NIST Handbook*, Diane Publishing, International Institute of Standards and Technology, Gaithersburg, MD, 1995.

2. K. Kaur, I. Gupta, A.K. Singh, A comparative evaluation of data leakage/loss prevention systems (DLPS), [in:] *Proceedings of 4th International Conference on Computer Science & Information Technology (CS & IT-CSCP)*, pp. 87–95, 2017, doi: 10.5121/csit.2017.71008.

3. A. Jones, Why are we not getting better at Data Disposal?, [in:] *Annual ADFSL Conference on Digital Forensics, Security and Law*, Vol. 7, pp. 89–94, 2009, https://commons.erau.edu/adfsl/2009/thursday/7.

4. R. Tahboub, Y. Saleh, Data leakage/loss prevention systems (DLP), [in:] *2014 World Congress on Computer Applications and Information Systems (WCCAIS)*, pp. 1–6, IEEE, 2014, doi: 10.1109/WCCAIS.2014.6916624.

5. R.S. Kadu, V.B. Gadicha, Review on securing data by using data leakage prevention and detection, *International Journal on Recent and Innovation Trends in Computing and Communication*, **5**(5): 731–735, 2017, doi: 10.17762/ijritcc.v5i5.597.

6. C. Bhatt, R. Sharma, Data leakage detection, *International Journal of Computer Science and Information Technologies*, **5**(2): 2556–2558, 2014.

7. X. Shu, D. Yao, E. Bertino, Privacy-preserving detection of sensitive data exposure, *IEEE Transactions on Information Forensics and Security*, **10**(5): 1092–1103, 2015, doi: 10.1109/TIFS.2015.2398363.

8. E. Costante, D. Fauri, S. Etalle, J. den Hartog, N. Zannone, A hybrid framework for data loss prevention and detection, [in:] *Proceedings of 2016 IEEE Security and Privacy Workshops*, San Jose, CA, USA, pp. 324–333, 2016, doi: 10.1109/SPW.2016.24.

9. P. Papadimitriou, H. Garcia-Molina, Data leakage detection, *IEEE Transactions on Knowledge and Data Engineering*, **23**(1): 51–63, 2011, doi: 10.1109/TKDE.2010.100.

10. X. Shu, J. Zhang, D.D. Yao, W.-C. Feng, Fast detection of transformed data leaks, *IEEE Transactions on Information Forensics and Security*, **11**(3): 1–16, 2016, doi: 10.1109/TIFS.2015.2503271.

11. S. Chhabra, A.K. Singh, Dynamic data leakage detection model based approach for MapReduce computational security in cloud, [in:] *Proceedings of 2016 Fifth International Conference on Eco-friendly Computing and Communication Systems (ICECCS-2016)*, Bhopal, India, pp. 13–19, 2016, doi: 10.1109/Eco-friendly.2016.7893234.

12. A. Shabtai, Y. Elovici, L. Rokach, *A Survey of Data Leakage Detection and Prevention Solutions*, Springer, Boston, MA, 2012, doi: 10.1007/978-1-4614-2053-8_4.

13. M. Ghouse, M.J. Nene, Graph neural networks for prevention of leakage of secret data, [in:] *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp. 994–999, 2020, doi: 10.1109/ICCES48766.2020.9137957.

14. M. Ghouse, M.J. Nene, VembuSelvi C., Data leakage prevention for data in transit using artificial intelligence and encryption techniques, [in:] *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, Mumbai, India, pp. 1–6, 2019, doi: 10.1109/ICAC347590.2019.9036839.

15. M.N.A. Wahid, A. Ali, B. Esparham, M. Marwan, A comparison of cryptographic algorithms: DES, 3DES, AES, RSA and blowfish for guessing attacks prevention, *Journal Computer Science Applications and Information Technology*, **3**(2): 1–7, 2018.

16. J.M. Gómez-Hidalgo, J.M. Martín-Abreu, J. Nieves, I. Santos, F. Brezo, P.G. Bringas, Data leak prevention through named entity recognition, [in:] *2010 IEEE Second International Conference on Social Computing*, Minneapolis, MN, USA, pp. 1129–1134, 2010, doi: 10.1109/SocialCom.2010.167.

17. A. Buda, A. Coleşa, File system minifilter based data leakage prevention system, [in:] *2018 17th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, Cluj-Napoca, Romania, pp. 1–6, 2018, doi: 10.1109/ROEDUNET.2018.8514147.

18. T. Mustafa, Malicious data leak prevention and purposeful evasion attacks: An approach to Advanced Persistent Threat (APT) management, [in:] *2013 Saudi International Electronics, Communications and Photonics Conference*, Riyadh, Saudi Arabia, pp. 1–5, 2013, doi: 10.1109/SIECPC.2013.6551028.

19. Y. Lu, X. Huang, D. Li, Y. Zhang, Collaborative graph-based mechanism for distributed big data leakage prevention, [in:] *2018 IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, United Arab Emirates, pp. 1–7, 2018, doi: 10.1109/GLOCOM.2018.8647746.

20. G. Katz, Y. Elovici, B. Shapira, CoBAn: A context based model for data leakage prevention, *Information Sciences*, **262**: 137–158, 2014, doi: 10.1016/j.ins.2013.10.005.

21. G. Michael, Data leakage in cloud computing, *International Journal of Pure and Applied Mathematics*, **116**(9): 273–278, 2017.

22. S.B. Alkhadhr, M.A. Alkandari, Cryptography and randomization to dispose of data and boost system security, *Cogent Engineering*, **4**(1): 1300049, 2017, doi: 10.1080/23311916.2017.1300049.

23. R. Chandramouli, D. Pinhas, Security guidelines for storage infrastructure, *NIST Special Publication*, **800**: 209, 2020, doi: 10.6028/NIST.SP.800-209.

24. T. Liquori, Methods of Data Destruction, Dispose of Data Securely, Accessed on Nov 10, 2021 at https://dataspan.com/blog/what-are-the-different-types-of-data-destruction-and-which-one-should-you-use/.

25. H. Hammouchi, O. Cherqi, G. Mezzour, M. Ghogho, M. El Koutbi, Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time, *Procedia Computer Science*, **151**: 1004–1009, 2019, doi: 10.1016/j.procs.2019.04.141.

26. S. Alneyadi, E. Sithirasenan, V. Muthukkumarasamy, A survey on data leakage prevention systems, *Journal of Network and Computer Applications*, **62**: 137–152, 2016, doi: 10.1016/j.jnca.2016.01.008.

27. K.S. Wagh, A survey: Data leakage detection techniques, *International Journal of Electrical and Computer Engineering*, **8**(4): 2247–2253, 2018, doi: 10.11591/ijece.v8i4.pp2247-2253.

28. A. Jones, Lessons not learned on data disposal, *Digital Investigation*, **6**(1–2): 3–7, 2009.

29. K. Rahul, R.K. Banyal, Data life cycle management in big data analytics, *Procedia Computer Science*, **173**: 364–371, 2020, doi: 10.1016/j.procs.2020.06.042.

30. D. Bisson, 7 Data Breach Case Studies Involving Human Error, Venafi, Accessed on Nov 15, 2021 at https://venafi.com/blog/7-data-breaches-caused-human-error-did-encryption-play-role/.

31. C. Chakraborty, A. Kishor, J.J.P.C. Rodrigues, Novel enhanced-grey wolf optimization hybrid machine learning technique for biomedical data computation, *Computers and Electrical Engineering*, **99**: 107778, 2022, doi: 10.1016/j.compeleceng.2022.107778.

32. S. Acharya, Security Injection: Mobile Risk Management – Introduction, Towson University, Accessed on Dec 12, 2021 at https://cisserv1.towson.edu/∼cyber4all/modules/nanomodules/Mobile_Risk_Management-Introduction.html.

33. A. Jones, C. Valli, I. Sutherland, P. Thomas, The 2006 analysis of information remaining on disks offered for sale on the second hand market, *Journal of Digital Forensics, Security and Law*, **1**(3): 2, 2006, doi: 10.15394/jdfsl.2006.1008.

34. Blancco, Data Sanitization in the Modern Age: DoD or NIST?, Accessed on Dec 26, 2021 at https://www.blancco.com/resources/bp-data-sanitization-in-the-modern-age-dod-or-nist/.

35. T. Caldwell, Seek and destroy, *Network Security*, **2012**(9): 15–19, 2012, doi: 10.1016/S1353-4858(12)70083-1.

36. Report on Data Loss, Dallas City Hall, Accessed on Jan 10, 2022 at https://dallascityhall.com/departments/ciservices/Pages/Report-on-Data-Loss.aspx.

Received February 2, 2022; revised version June 12, 2022; accepted July 31, 2022; published online April 16, 2024.