Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 15th International Scientific and Practical Conference. Volume II, 253-258

Data Leakage Prevention and Detection in Digital Configurations: A Survey

Svetlana Syarova
CS Department
ULSIT
Sofia, Bulragia
s.syarova@unibit.bg

Stefka Toleva-Stoimenova
CS Department
ULSIT
Sofia, Bulragia
s.toleva@unibit.bg

Aleksandar Kirkov
CS Department
ULSIT
Sofia, Bulragia
a.kirkov@unibit.bg

Samuel Petkov
ULSIT
Sofia, Bulragia
46732r@unibit.bg

Krasimir Traykov
ULSIT
Sofia, Bulragia
k.traykov@unibit.bg

Abstract. **As a result of the development of information and communication technologies (ICT) and Internet electronic interaction at all levels in the organizations and the use of various electronic services has become part of our everyday life. The past few decades have been characterized by a tremendous growth in the amount of data generated. At the same time, digital data are subject to malicious and accidental threats, due to the presence of vulnerabilities in the protection of information systems. Unauthorized access, malware, zero-day attack, data leakage, denial of service (DoS), and phishing have increased exponentially in recent years. Data leakage occurs when sensitive data and confidential information is revealed to unauthorized parties. Data leakage is one of the main targets of any insider threat. Over the last few years, the challenge of dealing with insider threats has been recognized and various methods have been proposed to address such problems. Therefore, most proposed internal threat detection methods work towards data leakage prevention (DLP).**
This paper addresses the data leakage prevention and detection (DLPD) as some of the most critical cybersecurity issues nowadays. The used DLP techniques and technologies have been explored briefly. As the study aims to reveal the scientific interests in the DLP domain we tried to provide a comprehensive overview of academic publications. Finally, the paper focuses on what drives the DLP domain, the challenges and opportunities the digital configurations are faced in the context of data flow monitoring, prevention and detection.

Keywords: data, leakage, prevention, security.

I. INTRODUCTION

Recent decades are characterized by an increased growth in data generated by humans and machines. It is a result of advances in information and communication technologies (ICT), the digitalization of production processes, the increasing use of electronic devices and networks, including the Internet of Things, cloud computing, etc. Simultaneously, the challenge of dealing with data leakage has been recognized and various methods have been proposed to address related insider threats.

"Insiders" are defined by the Cyber Security and Infrastructure Agency (CISA) [1] as: "any person who has or has had authorized access to or knowledge of an organization's resources, including personnel, facilities, information, equipment, networks and systems." An "insider threat" is defined by CISA as: "the threat that an insider will use their authorized access, knowingly or unknowingly, to harm the mission, resources, personnel, facilities, information, equipment, networks or systems" Insiders have all the necessary knowledge about internal systems and their topology and have legitimate access to sensitive and valuable information assets [2], [3]. As such, they can inflict much more damage than outsiders [4], [5]. A joint study by the U.S. intelligence community was presented, which included characterizing and analyzing the methods used to counter malicious insider threats [6]. It has proposed a general model of malicious internal behavior, distinguishing motivations, actions, and relevant observables. In [7] authors focused their study on the risks of insider threats in the field of information technologies (IT) through an organization's external partners. This study suggests reducing these risks by using non-deceptive techniques such as

Print ISSN 1691-5402
Online ISSN 2256-070X
https://doi.org/10.17770/etr2024vol2.8045

intrusion detection systems, but also fraudulent techniques such as honeypots. In 2023, Rosenthal estimated the average cost of an insider threat incident to be $11.45 million, up from $8.76 million in 2018. Organizations are rightly concerned about this threat because insiders can threaten their survival [8].

This paper addresses the DLP domain as one of the most challenging cybersecurity issues today that help identifying, monitoring, protecting and reducing the risks of sensitive-data leakage. A lot of scientific interests have been shown and many related scientific works have been published in academic literature. Various technical approaches have been used in different causes of data leaks. This study is based on a systematic literature review in a way to provide a comprehensive analysis of the current state, challenges and opportunities of data flow monitoring, prevention and detection in the digital configurations. A series of questions arise:

(1) What techniques and technologies are used?
(2) What are the most explored research fields?
(3) What are the challenges and opportunities?

The paper is organized as follows. The second section looks into DLP nature and the techniques and technologies used. The third section explores academic research in the DLP domain and discusses the challenges and opportunities the digital configurations are faced in the context of data flow monitoring, prevention and detection.

II. Materials and Methods

According to [9], DLP solution is used to detect and prevent unauthorized attempts to copy or send sensitive data, both intentionally or/and unintentionally, without authorization, by people who are authorized to access the sensitive information. Some of the used DLP's synonyms are: data loss prevention (DLP), information leakage prevention (ILP), information leakage detection and prevention (ILDP), extrusion prevention (EP), etc. The DLP domain addresses data leaks in the following three states of data throughout their lifecycle by applying specific set of technologies, as shown in Fig. 1 [10], [11]:

- Data-at-Rest (DAR): Data that resides in files system, databases and other storage methods, e.g. a company's financial data stored on the financial application server.
- Data-in-Use (DIU): Data at the endpoints of the network, e.g. USB devices, external drives, MP3 players, laptops, and other highly-mobile devices).
- Data-in-Motion (DIM): Data transmitted on (wire or wireless) network, e.g. customer purchasing details sent over the Internet.

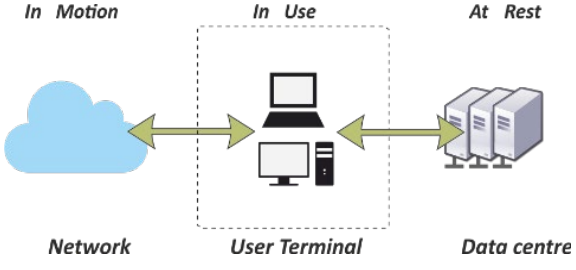

Fig. 1 Data states throughout their lifecycle

DLP solutions can be grouped according to the taxonomy that incorporates the features [12], [13]:

- Data state: DAR, DIU, and DIM.
- Deployment scheme: endpoint and network.
- Leakage handling approach: preventive and detective mechanisms
- Remedial actions: audit, block/remove, notify, encrypt, quarantine.

DLP solutions are used to detect, monitor and protect confidential or sensitive data wherever are stored or used, across endpoint, network, and storage systems. Two main leakage handling approaches - preventive and detective, are shown in Fig. 2:

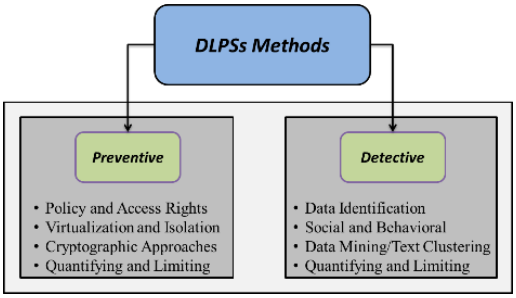

Fig. 2 DLP's methods [13]

Leakage occurrences can be treated by using a detective approach. The system detects any possible leakage incidents and applies the corrective action that is capable of handling the identified leakage incident [12]. Data are categorized as confidential (sensitive) and non-confidential data and subsequently are used for detective purposes. The used techniques are divided into two main groups [13], [14], [15]:

A. Content-based analysis technique

Examines data content to detect sensitive data and protect from accidental exposure and loss in different data states (DAR, DIU, and DIM). In this case, the DLP techniques are mainly based on three content analysis types, which are data fingerprinting, regular expression, and statistical analysis.

B. Context-based analysis technique

Explores only metadata or other properties of the monitored data, for example source, destination, size, recipients, header/metadata information, time stamps, file type, location, format, application, and queries or transactions. Such DLP techniques include social and behavior analysis, data identification, and data mining and text clustering.

The most popular techniques are presented and compared in the following Table 1:

Existing DLP systems, both open source and commercial products, are: Websence DLP, Open DLP, Symantec DLP, Trustwave DLP, MyDLP, RSA DLP, MacAfe DLP, Furtinet DLP, etc. [16].

TABLE 1 SUMMARY OF EXISTING DLP TECHNIQUES

Techniques	Comparison		
	Analysis	***Advantage***	***Disadvantage***
Fingerprinting (exact/partial matching)	Content	Simple; Low false positive rate.	Very sensitive to data modification
Regular Expression	Content	Simple; Allow complex pattern matching	High false positive rate
Statistical analysis (N-gram/ Term weighting)	Content	Detect sensitive content in unstructured data	Large amount of data; High false positive rate and high false negative rate
Social and behavior analysis	Context	Proactive prevention technique; Mitigate insider threats	High false positive rate; Administrator involvement
Data mining and text clustering	Context	Perform a complicated task; ML techniques	High false positive rate; Limited scalability; Complicated
Data identification	Context	Very robust to detect unaltered data	Cannot understand data semantics

III. RESULTS AND DISCUSSION

Determining the exact number of academic research papers in the DLP domain is challenging due to several factors:

Volume of Research: The field of cybersecurity, including DLP, is vast, and numerous academic conferences, journals, and research institutions worldwide contribute to the body of knowledge.

Multidisciplinary Nature: DLP research intersects with various disciplines, including computer science, information security, cryptography, data privacy, and behavioral science. As a result, research papers on DLP may appear in a wide range of academic venues covering these areas.

Diverse Topics and Approaches: DLP research encompasses a broad range of topics, including data classification, content inspection, policy enforcement, behavioral analytics, encryption, cloud security, and regulatory compliance. Researchers employ diverse methodologies and approaches to address different aspects of DLP.

Publication Venues: Academic research on DLP may be published in peer-reviewed journals, conference proceedings, workshop papers, technical reports, and dissertations.

As the study aims to reveal the scientific interests related to the DLP techniques and technologies, various data collections organized by publication types such as Conferences, Journals, Magazines, Early Access Articles, and Books have been examined.

For this research, the search was focused on papers in scientific databases such as Google Scholar, Science Direct, IEEE Xplore, Web of Science, Scopus and ACM Digital Library, as these databases cover relevant scientific information in multiple engineering fields, allowing access to articles published in scientific and academic journals, repositories, archives and other collections. The following keywords were used for the literature search: "Security" AND ("DLP" OR ("Data AND ("Leak" OR "Loss") AND ("Prevention OR Protection"). These terms are searched in Abstract/Title/Keywords of the papers.

The papers to be analyzed are selected by reading the titles of the results obtained. As selection criteria in the analysis of the abstracts of the papers we used: (1) Studies related to DLP techniques and technologies, (2) Studies related to subdomains of DLP domain, (3) Studies related to challenges and benefits of DLP.

This paper reviews the academic research in the DLP landscape grouped into the following categories:

A. Misuse Detection in Database

Various investigations have been conducted the detection of unusual access to databases. Two approaches are distinguished - syntax-oriented and data-oriented. Both involve mapping between users, searches, and search results. The syntax-oriented approach is based on the syntax of the SQL statements of the query to create a user profile. A data-driven approach focuses on what the user is trying to get, usually by extracting features from the search result set, such as the number of searches as well as the minimum, maximum, and average values of the search attributes.

In [17] authors evaluated a syntax-centric approach to data abuse detection in databases management systems (DBMS) that manage SQL query logs to profile the normal access behavior of users in databases. In [18] proposed a method to create a statistical profile of the normal user's database access pattern to see when the user deviates from his routine. The authors used the data-driven approach and considered its composition an irrelevant search expression to recognize the user's intent, giving importance only to the received data. In [19] also used the data-driven approach to model the knowledge an insider can extract from a given set of records. Given that the insider has legitimate access tables, attributes, and files, he can apply his knowledge to create new knowledge. The method uses dependency graphs based on domain-expert knowledge. Fonseca et al. [20] proposed a Malicious Data Access Detector (MDAD). It aims to protect database applications from data attacks and web applications from SQL injection attacks. This is achieved by representing the profile of valid transactions through a graph that describes different sequences of SQL queries (SELECTs, INSERTs, UPDATEs, and DELETEs), from the beginning of the transaction to the commit or rollback command. The DEMIDS system detects intrusions by building user profiles based on their working scopes which consist of feature/value pairs representing their activities. The system uses audit log data to derive profiles describing typical patterns of accesses by database users [21].

In particular, number of methods and systems have been developed for misuse detection in information

retrieval (IR) systems. In [22] compares user behavior in terms of content rather than in terms of commands issued to a developed user profile, learned through clustering, relevance feedback, and fusion methods. Thus, a new dimension was created to profile-based misuse detection for search systems. In [23] has been proposed a relevance feedback approach based on building a user profile containing both query and feedback terms from prior queries. This method compares user's actions with existing profile.

B. Email Leakage Protection

Many authors concern the aim to study the content and headers of email messages for detecting abuse (e.g., spam) and digital forensic analysis. In [24] the authors propose to use stylometry, the statistical analysis of variations in literary style between users. Using machine learning, they were able to verify the authorship of the emails in a majority of cases. This gives a general idea of the ability to identify and use basic email content to gain insight, and in this case, useful attribution intelligence. Nurse's research [25] investigates the extent to which potentially sensitive information could be leaked, in even blank emails, by considering the metadata that is a natural part of email headers. Through findings from auser-based experiment, we demonstrate that there is a noteworthy level of exposure of organizational and personal identity information, much of which can be further used by an attacker for reconnaissance or develop a more targeted and sophisticated attack. According to [26], an electronic message is identified as a leak based on its content and the likelihood that the recipient of the message will receive it. Messages sent to previous recipients are modeled as message-recipient pairs. Such a pair is considered a potential leak if the message is significantly different from previous messages sent to the recipient. To improve performance, Carvalho and Cohen [27] use various features of social networks. They presented an implementation of their solution in Mozilla Thunderbird. They have also expanded their system to not only detect spam recipients but also suggest recipients that the user may have forgotten to include. In [15] an approach is proposed based on analysis of emails exchange among members of the organization and the identification of groups based on common topics. When a new email is composed and about to be sent, each email recipient is analyzed. A recipient is approved if the email's content belongs to at least one of the topics common to the sender and the recipient.

C. Network / Web-based Protection

In [28] authors introduce a method for computing bandwidth in outbound HTTP traffic that involves discarding expected header fields. However, they use a stateless approach and therefore are unable to discount information that is repeated or constrained from previous HTTP messages. Later researchers present leak measurement algorithms for the Hypertext Transfer Protocol (HTTP), the main protocol for web browsing [29]. Instead of trying to detect the presence of sensitive data, they measure and constrain its maximum volume. They take advantage of the insight that most network traffic is repeated or determined by external information, such as protocol specifications or messages sent by a server. By discounting this data, true information leakage has been isolated and quantified.

A system called Elicit (Exploit Latent Information to Counter Insider Threats) is presented in [30]. Its aim is to help analysts identify insider threats. This system takes advantage of network traffic and contextual data both. ELICIT uses a naive Bayes detection approach, using 72 features based on the searching, browsing, downloading and printing behavior of users. Examples of features used include the number of remote print jobs, the number of queries made during a suspicious time, and the number of queries that resulted in high document retrievals. They are combined with contextual information and processed by various rule-based and statistical detectors that issue alerts [31].

D. Encryption and Access Control

Cryptography refers to secure information and communication techniques related to the conversion of data from a readable format to an encrypted format. The main purpose is to ensure that content can only be accessed by authorized devices and users. In [32] a framework for protecting sensitive data share between collaborating organizations has been proposed. Their solution is based on trusted computing, which provides a hardware base trust. The trusted computing ensures that the shared data in encrypted form and the encrypted key is accessible only to authorized devices. [33] presented a web-based framework for preventing leakage of confidential information. It is transparent to the user and ensures the safety of confidential data while they are at -rest, in-motion and in-use. Digital Rights Management (DRM) systems refers to a set of policies, techniques and tools that guide the proper use of digital content and ensure vulnerability management in an organization. In [34], [35], [36] the enterprise DRM system is presented, which provides persistent protection for documents using cryptographic methods.

E. Honey Pots for Detecting Malicious Insiders

A honeypot is a unique security resource. It is an information system resource whose value lies in unauthorized or illicit use of that resource [37]. There are two key types of honeypots that play a role in indicating and capturing an advanced insider threat, honeynets and honeytokens. In [38] authors proposed a prototype honeypot to automatically generate signatures for intrusion detections without hard coding any clue in advance to achieve zero-day detections of unknown malware. [37] presented techniques for detecting insider threats using honeypots and honey tokens. Insider threats have challenges different from outsider attacks, as that the malicious insiders are given access to the system and are much more familiar with it. To help catch such malicious insiders, honeypots should be moved into the network and can take up all unused IP addresses. In [39], [40] has been described a procedure of information assurance forensics using honeypots. It consists of network activity analyses, system and file analyses, and evidence gathering. [41] integrated intrusion tolerance into network security forensics using honeypots, called dynamic forensics‖. The solution makes sure that data gathered for forensic analysis is reliable even if those attacks have tried to modify the data.

In the context of data flow monitoring, prevention, and detection, digital configurations face both challenges and opportunities. Here are some of the key challenges:

Complexity of Digital Environments: Modern digital environments are often complex and dynamic, consisting of diverse systems, applications, and devices interconnected across networks and cloud platforms.

Data Volume and Velocity: Traditional monitoring tools may struggle to keep pace with the continuous flow of data, leading to gaps in coverage and potential security blind spots.

Encryption and Anonymization: Encrypted data traffic obscures the contents of communications, making it difficult to inspect data packets for signs of malicious activity or policy violations. Similarly, anonymized data can obscure the identities of users or devices involved in data transactions, hindering attribution and forensic analysis.

Insider Threats: Employees, contractors, or partners with legitimate access to data may abuse their privileges or inadvertently mishandle sensitive information, leading to data leakage. Detecting and mitigating insider threats requires a combination of technical controls, user monitoring, and behavioral analytics.

Regulatory Compliance Requirements: Organizations must ensure that their monitoring practices comply with relevant regulations such as GDPR, HIPAA, PCI DSS, and others, which often impose strict requirements for data security, privacy, and breach notification.

Some opportunities and trends are:

Advanced Analytics and Machine Learning: Machine learning algorithms can analyze large volumes of data in real-time, identify patterns of normal behavior, and detect anomalies indicative of security threats or policy violations.

Behavioral Analysis: By monitoring and analyzing user behavior patterns, organizations can identify suspicious activities and proactively intervene to prevent data breaches.

Integration with Security Ecosystem: Integration with other security technologies such as SIEM (Security Information and Event Management) systems, endpoint detection and response (EDR) solutions, and threat intelligence platforms, organizations can gain valuable insights into potential security incidents and respond more effectively.

Cloud-Native Solutions: Cloud access security brokers (CASBs), cloud workload protection platforms (CWPPs), and cloud security posture management (CSPM) tools provide centralized monitoring and enforcement of security policies across cloud services and applications.

IV. Conclusions

This paper summarizes our survey on the recent advances and the current trends in DLP research. We recognized major areas and reviewed the academic research in the DLP domain. Significant progress is seen in DLP techniques and technologies to address related insider threats. Overall, while digital configurations face challenges in data flow monitoring, prevention, and detection, advancements in technology, analytics, and security solutions offer opportunities for organizations to improve their ability to safeguard sensitive data and protect against emerging threats. By adopting a holistic approach to data security and leveraging innovative solutions, organizations can mitigate risks and ensure compliance with regulatory requirements in an increasingly complex digital landscape.

Acknowledgments

This work has been supported by the Ministry of Education and Science in implementation of the National Scientific Program "Security and Defense," adopted by RMS No. 731 of 21.10.2021 under Agreement No. D01-74/19.05.2022.

References

[1] Cybersecurity & Infrastructure Security Agency. Defining insider threats, no date. https://www.cisa.gov/defining-insider-threats [Accessed 25 February 2024].

[2] G. Mazzarolo and A.D. Jurcut, "Insider threats in cyber security: The enemy within the gates," Eur. Cybersecur. Journal, vol.6, no.1, pp. 57-63, 2019. 10.48550/arXiv.1911.09575 [Accessed 25 February 2024].

[3] R. Willison and M. Warkentin, "Beyond deterrence: An expanded view of employee computer abuse," Manage. Inform. Syst. Quart., vol. 37, no.1, pp. 1-20, 2013. https://www.jstor.org/stable/43825935 [Accessed 25 February 2024].

[4] PWC, "US cybercrime: rising risks, reduced readiness – KEy findings from the 2014 US State of cybercrime survey," https://www.pwc.com/us/en/increasing-it-effectiveness/publications/assets/2014-us-state-of-cybercrime.pdf [Accessed 25 February 2024].

[5] M.R. Randazzo, M. Keeney, E. Kowalski, D. Cappelli, and A. Moore, "Insider threat study: Illicit cyber activity in the banking and finance sector" Technical Report ADA441249, Carnegie-Mellon Univ Pittsburgh Software Engineering Inst, 2005. https://apps.dtic.mil/sti/citations/ADA441249 [Accessed 25 February 2024].

[6] M. Maybury, "Analysis and detection of malicious insiders," Proceedings of International Conference on Intelligence Analysis 2005.

[7] P. Gaonjur and C. Bokhoree, "Risk of Insider Threats in Information Technology Outsourcing: Can Deceptive Techniques be Applied?," in Security and Management, CSREA Press, p.522, 2006.

[8] Software Engineering Institute, "Common Sense Guide to Mitigating Insider Threats, Seventh Edition, " Carnegie Mellon University, Software Engineering Institute's Digital Library. Software Engineering Institute, 7-Sep-2022 [Online]. Available: https://insights.sei.cmu.edu/library/common-sense-guide-to-mitigating-insider-threats-seventh-edition/ [Accessed: 25-Feb-2024].

[9] A.V.Kale, Sh.P. Dubey and V. Bajpayee, "A revew on Data Leakage Prevention," International Journal of Computer Science and Mobile Computing, vol. 4, no. 4, April 2015, pp. 513-518.

[10] R. Tahboub and Y. Saleh, "Data Leakage/Loss Prevention Systems (DLP)," 2014 World Congress on Computer Applications and Information Systems (WCCAIS), Hammamet, Tunisia, pp. 1-6, 2014. doi: 10.1109/WCCAIS.2014.6916624.

[11] S. Peneti and B. P. Rani, "Data Leakage Detection and Prevention Methods: Survey," in Discovery, vol. 43, no. 198, pp. 95-100, 2015.

[12] A. Shabtai, Y. Elovici and L. Rokach, "A survey of data leakage detection and prevention solutions," in Springer Briefs in Computer Science, Springer, 2012.

[13] S. Alneyadi, E. Sithirasenan and V. Muthukkumarasamy, "A survey on data leakage prevention systems," Journal of Network and Computer Applications, vol. 62, pp. 137-152, 2016. https://doi.org/10.1016/j.jnca.2016.01.008. [Accessed 25 February 2024].

[14] S. Alneyadi, E. Sithirasenan and V. Muthukkumarasamy, "Detecting Data Semantic: A Data Leakage Prevention Approach", pp.910-917, 2015.

[15] P. Zilberman, S. Dolev, G. Katz, Y. Elovici and A. Shabtai, "Analyzing group communication for preventing data leakage via email", pp.37 - 41, 2011.

[16] V.O. Waziri, I. Idris, J.K. Alhassan and B.O. Adedayo, "Data Loss Prevention and Challenges Faced in their Deployments", 2017.

[17] A. Kamra, E. Terzi and E. Bertino, "Detecting anomalous access patterns in relational databases," in International Journal on Very Large Databases, vol.17, pp.5, pp.1063–1077, 2008.

[18] M. Sunu, P. Michalis, N. Hung, and U. Shambhu, "A Data-Centric Approach to Insider Attack Detection in Database Systems", Technical Report, 2009.

[19] Q. Yaseen and B. Panda, "Knowledge acquisition and insider threat prediction in relational database systems," Proceedings, 12th International IEEE Conference on Computational Science and Engineering, pp.450–455, 2009.

[20] J. Fonseca, M. Vieira and H. Madeira, "Online detection of malicious data access using DBMS auditing," ACM Symposium on Applied Computing, pp.1013–1020, 2008.

[21] C. Y. Chung, M. Gertz and K. Levitt, "Demids: A misuse detection system for database systems," in Working Conference on Integrity and Internal Control in Information Systems, pp. 159-178, Boston, MA: Springer US, 1999.

[22] R. Cathey, L. Ma, N. Goharian and D. Grossman, "Misuse detection for information retrieval systems," in Proceedings, 12th ACM Conference on Information and Knowledge Management (CIKM), 2003.

[23] L. Ma, N. Goharian, "Using Relevance Feedback to Detect Misuse for Information Retrieval Systems," ACM CIKM, 2004.

[24] O. De Vel, A. Anderson, M. Corney and G. Mohay, "Mining E-mail Content for Author Identification Forensics," SIGMOD Record, vol.30, no.4, pp. 55-64, 2001.

[25] J. Nurse, A. Erola, M. Goldsmith and S. Creese, "Investigating the leakage of sensitive personal and organisational information in email headers," in Journal of Internet Services and Information Security, vol.5, 2015.

[26] V.R. Carvalho, and R. Balasubramanyan, Information leaks and suggestions: a case study using Mozilla Thunderbird. Proceedings, 6th Conference on Email and Anti-Spam, 2009.

[27] V.R. Carvalho and W. Cohen, "Preventing information leaks in email," in Proceedings, SIAM International Conference on Data Mining, 2007.

[28] K. Borders and A. Prakash, "Web Tap: Detecting Covert Web Traffic," Proceedings of the 11th ACM Conference on Computer and Communications Security (CCS), 2004.

[29] K. Borders and A. Prakash, "Towards quantification of network-based information leaks via HTTP," Proceedings of 3rd conference on Hot Topics in Security, 2008.

[30] D.D Caputo., G.D. Stephens and M.A. Maloof, „Detecting insider theft of trade secrets," IEEE Security and Privacy, vol.7, no.6, pp.14–21, 2009.

[31] G. Mazzarolo and A. Jurcut, "Insider threats in Cyber Security: The enemy within the gates," arXiv preprint arXiv:1911.09575, 2019.

[32] I.M. Abbadi, and M. Alawneh, "Preventing insider information leakage for enterprises," International Conference on Emerging Security Information, Systems and Technologies, pp.99–106, 2008.

[33] K. Yasuhiro and S. Yoshiki, "A Web-based system for prevention of information leakage," (poster), Proceedings of 11th International World Wide Web (WWW) Conference, 2002.

[34] R.S. Reddy, S.R. Gopu, "Enterprise Digital Rights Management for Document Protection," in 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA). pp. 321–326, IEEE, 2017.

[35] M. Munier, V. Lalanne and M. Ricarde, "Self-protecting documents for cloud storage security," Proc. of the 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012, pp. 1231–1238, IEEE, 2012.

[36] M. Munier, "A Secure Autonomous Document Architecture for Enterprise Digital Right Management," in 7th International Conference on Signal Image Technology & Internet-Based Systems. pp. 16–23, IEEE, 2011.

[37] L. Spitzner, "Honeypots: Catching the Insider Threat," Proceedings of the Computer Security Applications Conference, pp. 170-179, 2003.

[38] C. Kreibichi and J. Crowcroft, "Honeycomb – Creating Intrusion Detection Signatures Using Honeypots," ACM SIGCOMM Computer Communication Review, vol. 34, no. 1, pp. 51-56, 2004.

[39] F. Raynal, Y. Berthier, P. Biondi, and D. Kaminsky, "Honeypot Forensics" Part I: Analyzing the Network, IEEE Security and Privacy, vol. 2, no. 4, pp. 72-78, 2004.

[40] F. Raynal, Y. Berthier, P. Biondi, and D. Kaminsky, "Honeypot Forensics" Part II: Analyzing the Compromised Host, IEEE Security and Privacy, vol. 2, no. 5, pp. 77-80, 2004.

[41] T. M. Chen and J. Buford, "Design Considerations for a Honeypot for SQL Injection Attacks," Proceedings of IEEE Local Computer Networks, pp. 915-921, 2009.