A Survey on Balancing Data Loss Prevention (DLP) with User Privacy in a Data-Driven World

Nurhafizah Binti Ahmad Zaini[1*], Mohamad Fadli bin Zolkipli[2**]

[1,2]School of Computing, Universiti Utara Malaysia, Malaysia

Email: nurhafizah_ahmad_@soc.edu.my[1*], m.fadli.zolkipli@uum.edu.my[2**]

Abstract

Nowadays, data breaches are a major concern for industries and governments, Data Loss Prevention (DLP) solutions have become essential tools to protect sensitive information and uphold data integrity. This study examines the ever-changing field of DLP methods, highlighting the importance of maintaining a balance between protecting data and respecting user privacy in the midst of widespread data circulation. The study discusses the difficult obstacles organizations encounter when merging DLP with privacy protection through analyzing real-life examples, research findings, laws, and technology developments. The results offer useful suggestions for matching DLP projects with privacy principles to improve organizational ability to withstand data breaches while also protecting individual privacy in a connected digital environment.

Keywords

Data Loss Prevention, User Privacy, Data Security, Data Protection

Introduction

During a time of fast technological progress, discussions about security attract the interest of different groups. Mark Zuckerberg thinks that the future of the cyber world will have a significant influence on the next generation, as the internet makes the virtual environment available to all, as stated by Huang, Li, and Cai (2023). This widespread interconnectedness requires a careful attitude towards technological advancements, as it increases the possibility of unauthorized disclosure and sharing confidential data. Understanding user privacy thoroughly and finding the right balance between Data Loss Prevention (DLP) and user privacy is essential in addressing these challenges in a data-driven world.

Previous discussions have established that DLP is more than the reinforcement of technology; lastly it maintains confidentiality and integrity in the cyber space. With the help of careful monitoring and applying strict controls in terms of access, DLP systems serve as reliable

Submission: 6 June 2024; **Acceptance:** 13 August 2024

safeguards against leakage and breaches when it comes to information and guarantee that such data will not be shared with people who do not have the authorization to get this kind of data (Gupta & Kush, 2023; Paracha et.al, 2022; Ahmed et. al, 2022). As explained by Stallings (2020), DLP works in the capacity of a privacy protector analyzing for privacy irritants like social security numbers, and financial data. They encode such data, thus it becomes unreadable to third parties, or refuse to send it to insecure areas, which ensures the privacy of individuals. However, the effectiveness of DLP extends beyond these traditional measures, embracing innovative approaches to uphold user privacy amidst ongoing technological developments.

DLP systems utilize various customizable strategies to cater to the specific needs of users in different situations, all in the aim of improving privacy protection. Tang et al. (2022) illuminated the use of deceptive dummy methods, giving users the ability to discreetly manage the sharing of their location information in location-based services. Additionally, combining statistical analysis and forecasting techniques enhances the predictive abilities of DLP models, allowing for proactive actions to prevent possible breaches (Gupta & Kush, 2023). These adaptive approaches enhance data security but also uphold the fundamental principles of user autonomy and privacy.

However, there are difficulties within the domain of DLP. Conversations regarding vulnerability assessments, intricate technologies, and cybersecurity risks shed light on the continuous challenges in the industry. Companies need to adhere to both regulations and ethical standards when creating successful DLP strategies (Kapiton et al., 2023; Bielova & Byelov, 2023; Bedadhala et al., 2023). This paper focuses on combining DLP with user privacy by summarizing the current situation, obstacles, and potential benefits. By carefully analyzing these factors, we suggest a detailed plan for managing DLP and protecting user privacy, aiming for a safer and fairer data-driven setting.

Literature review

According to research conducted by Guha et al. (2021), DLP systems are implemented to identify, oversee, and stop possible data breaches using established policy guidelines. It is known by various names such as data leakage prevention (DLP) systems, extrusion prevention systems, content monitoring systems, and filtering systems. Data losses can be classified as either external or internal, with internal incidents resulting from deliberate or accidental actions. Identifying and stopping intentionally leaked information from within remains a major obstacle for certain systems. Typically, data can be lost in several ways such as removable storage devices, web apps, emails, instant messaging platforms, and physical or digital documents. The DLP systems are specialized systems utilized to safeguard data while in use, in transit, and at rest.

According to a study by Shahzad, A., & de Sousa, E. M. (2021), the main worry for 20.97% of respondents regarding DLP systems was the potential effect on employee productivity and collaboration. On the other hand, almost 32% of participants identified the intricacy of establishing and upholding policies as a major concern. Additionally, DLP systems need to integrate privacy-enhancing technologies like data anonymization and encryption in order to reduce the chances of sensitive user information being revealed. DLP systems can guarantee that unauthorized parties

cannot read or use exposed data by encrypting it while both at rest and in transit, even if a breach happens.

Discussion 1: The Role of DLP in Protecting Sensitive Data

First of all, DLP systems are intentionally designed to help organizations protect data from leakage and theft. These solutions efficiently manage and regulate the transfer of data while permitting only allowed people to retrieve sensitive data (Gupta & Kush, 2023; Paracha et al., 2022; Ahmed et al., 2022). According to Akhramovych (2023), data security methods are implemented based on the data state, which includes still data, data-in-transit and data-in-use where encryption, management of access rights and digital rights protection technologies should be effective for each stage. Through the recognition and prevention of data leakage through the various communication channels provided by emails, USB connections, and other external means of data transfer, the system makes certain that data does not get outside the company's perimeter (Daubner & Považanec, 2023). This constant monitoring as well as enforcement of the access control mechanisms put in place ensure that the environment is secure dramatically minimizing the possibility of achieving a hostile act against the information.

DLP systems provide user-controlled privacy protection mechanisms, which allow users to manage how and with whom the location data is shared, especially in location-based services (Tang et al., 2022). Such strategies make sure that only identified Parties have the permission to access sensitive location information and that they cannot be tracked and misuse data about an individual person. As for the types of DLP models, advanced models build on this protection by using statistical analysis and forecasting to anticipate the access to the data. Often being the models of their actions, such systems can predict specific risks and apply precautions; thus, the accuracy and efficiency of data protection are increased manifold. It means the security measures are constantly a step forward from the threats, which minimizes risks of data loss (Gupta & Kush, 2023).

In addition, DLP systems are commonly integrated with other security solutions including encryption, digital rights management, and cloud access security brokers (CASB) to effectively safeguard data in all states and uses (at rest, in transit, and in use) (Akhramovych, 2023; Ahmed et al., 2022; Kapiton et al., 2023). This integration ensures that the content is well protected by means of data encryption, access through digital right management, and safety of data in cloud environments. Furthermore, the approach of incorporating DLP with Security Information and Event Management (SIEM) platform enhances the ability of most probably identifying and handling data breaches appropriately. These SIEM systems provide real-time analysis of the security alerts generated by applications and network devices to give an immediate and precise response to threats (Paracha et al., 2022).

Finally, DLP solutions can be deployed in endpoints and cloud platforms hence providing a stricter security on various structures (Daubner & Považanec, 2023; Ahmed et al., 2022). The use of DLP systems in big organizations has been found to have the greatest effect in lessening the likelihood of data breaches and enhancing internal security (Shishodia & Nene, 2022).

Table 1 demonstrates how DLP Systems protect sensitive data.

Function	Description	Example Strategies
Monitoring and Control	Prevents unauthorized access to sensitive data	Data encryption, access rights
Privacy Preservation	Ensures control over data sharing and usage	Location data control
Integration	Works with other security tools for comprehensive protection	Encryption, digital rights management

Discussion 2: Addressing Challenges in Data Loss Prevention

Data loss prevention (DLP) faces complex challenges in vulnerability assessment, data integrity assurance, operational intricacies, and defense against cyber-attacks (Kapiton et al., 2023; Majdoubi et al., 2023; He, 2023; Bielova & Byelov, 2023). Analyzing and identifying vulnerabilities is a crucial part of DLP, ensuring proactive measures against potential breaches (Kapiton et al., 2023). To maintain data integrity and confidentiality, robust encryption protocols and access controls are necessary (Majdoubi et al., 2023). Additionally, managing the ever-changing landscape of technological advancements and operational complexities remains a fundamental challenge in developing effective DLP strategies (He, 2023). Protecting against cyber-attacks requires continuous vigilance and adaptation to emerging threats (Bielova & Byelov, 2023).

Thus, the advance in technology particularly in the field of IT and big data present many challenges to DLP, (Kapiton et al. , 2023). The rapid development of these technologies poses a problem on generating enough protective measures (He, 2023). The addition of AI causes an increase in concerns because it introduces new aspects into data protection methods (Bielova & Byelov, 2023). Therefore, it means that one has to monitor the advancement in technology and come up with ways to ensure that the implemented DLP frameworks are strong (Kapiton et al. , 2023).

Security issues are different for big data and cloud computing compared to other security processes (Majdoubi et al., 2023; Fataftah & Isong, 2022). There is a call for adaptive security solutions since they are dynamic in nature and psychologically scalable (Fataftah & Isong, 2022). A big threat to big data systems include vulnerability issues for instance data alteration and unauthorized data access (Majdoubi et al. , 2023). Thus, the challenges above require that new measures be put in place such as employing machine learning algorithms for risk identification (Majdoubi et al. , 2023).

Challenges such as unauthorized access and identity theft pose a huge threat to DLP, as revealed by Bedadhala et al. (2023) and Niklekaj (2023). From these threats, data integrity and confidentiality can be threatened which requires strong measures in the network (Bedadhala et al, 2023). Effective monitoring, modern firewalls, encryption, and user training are the keywords, which maintain network security strategies (Bedadhala et al, 2023; Niklekaj, 2023). Implementing

these actions decreases the risks of cyber threats and ensures the longevity of DLP frameworks (Niklekaj, 2023).

Table 2 present the primary challenges and solutions in DLP.

Challenge	Study	Description	Proposed Solution
Technological Challenges	Kapiton et al. (2023)	Rapid evolution of technologies complicates protection mechanisms	Keeping pace with advancements
Operational Challenges	He (2023)	Managing complex operational landscapes	Adaptive security solutions
Cybersecurity Threats	Bedadhala et al. (2023)	Unauthorized access, identity theft	Advanced firewall, continuous monitoring

Discussion 3: Impementation of Real-World Data Loss Prevention (DLP) Techniques

It's important that Data Loss Prevention (DLP) techniques are applied to help protect organization's important data. They improve organization internal security systems and make sure that the organization complies with the set rules and regulation. The most recent literature review discusses techniques and solutions, which are emerging to enhance DLP systems, given the increasing threats and risks present in today's computer environment.

The fundamental technologies that underpin DLP solutions are those that target the identity and safeguard of data against internal and external threats. Paracha et al. (2022) stated that there are Two Layered DLP Strategies to be complied. To achieve this, this method seeks to prevent the transfer of data through USB devices and configure the email system to prevent the sending of emails with forbidden contents. It is possible to avoid breaches and meet the requirements of different statutes and rules and improve the internal protection in companies and organizations through prevention strategies.

Therefore, to avoid this, Gupta et al (2022) proposed a method for DLP based on semantics using statistical measures to identify or prevent the loss of such information. Unlike other models, the TF-IDF considers the context and the neighboring goods and services, and thus supplies a sound basis for document categorization even if some of the papers are substituted or changed. This semantics centered approach helps in achieving better classification results and brings variety of other data states, which are in use, in transit, or at rest under protection of DLP solutions.

Thus, the evolution of DLP goes on with the incorporation of new methods and technical instruments improving the protection of data. Kapiton et al. (2023) also review development including Data Encryption Management (DSM) and Data Security Isolation (DSA) alongside the currently prominent DLP-based solutions. Also, the proposal of electrostatic and electromagnetic shielding shows the comprehensive method of elimination of data loss by protection against interceptions and high voltage interferences.

As stated before, real-life implementation of DLP strategies necessitates the combination of the fundamental technologies, statistical methods, and semantic models and working techniques. The application of various facets in models improves the effectiveness of DLP solutions affording comprehensive protection against data leaks and unauthorized disclosures. The current threats can be dealt with and the security of an organization can be well maintained and there can be full compliance with the set regulations.

According to new studies, it is revealed that more attention should be paid to enhance techniques of DLP due to developing security risks. Core technologies, statistical and semantic-driven approaches as well as innovative methodologies can improve organizations' defense for data breaches, safeguard data, and meet regulatory requirements in a more integrated digital environment.

Table 3 illustrates DLP techniques and implementations.

DLP Technique	Description	Implementation	Effectiveness
Two Layered DLP Strategies	Prevent unauthorized data removal and accidental sharing	USB blocking, email filtering	High
Semantic-driven DLP	Uses statistical methods like TF-IDF for sensitive data detection	Contextual analysis, document classification	Moderate to High
Data Encryption Management (DSM)	Encryption-based approach for data protection	Data encryption, access control	High

Conclusion

This research highlights the important role of Data Loss Prevention (DLP) systems in preserving data and an important aspect of user privacy. It goes through various methods of DLP and the actual, real-world usage to depict how they not only prevent such access but also give customized solutions to maintaining the privacy of the users. The results highlight the importance of incorporating advanced technologies like encryption and predictive analytics into DLP solutions to improve data security in various states: of stored being transferred and being utilized for a given period. It is essential to integrate such approaches to create a solid base to meet current and future threats in cyberspace, as well as to ensure the company's compliance with regulation standards and laws.

As for future, the continuous advancement of technology presents both challenges and opportunities for DLP systems. Future research should focus on the formation of the adaptive security models that correspond to the current tendencies and threats. Moreover, adoption of other new approaches like semantic driven DLP along with enhanced shielding methods gives more holistic approach to the problem. Through anticipating for the possible risks that could pose threat to an organization's data security and adopting the use of new technologies in the management of data, organizations can enhance their data security standards while embracing the user privacy standards in the emerging complex world of technological advancement.

Acknowledgements

The authors would like to thank all members of the School of Computing who participated in this study. This study was carried out as part of the System and Network Security Project. This work was supported by Universiti Utara Malaysia.

References

Ahmed, S., Haq, A., Sheeraz, M., & Durad, M. (2022). Design and development of cloud-based QR coded watermarking DLP system. 2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 696-701. https://doi.org/10.1109/IBCAST54850.2022.9990327

Akhramovych, V. (2023). Data protection at the stages of its functioning. *Cybersecurity: Education, Science, Technique*, (21), 149-161. https://doi.org/10.28925/2663-4023.2023.21.149161

Bedadhala, S., Kotteti, C., & Sadiku, M. (2023). Cyber Security: Challenges and Preventive Measures. International Journal of Advances in Scientific Research and Engineering. https://doi.org/10.31695/ijasre.2023.9.2.7.

Bielova, M., & Byelov, D. (2023). Challenges and threats of personal data protection in working with artificial intelligence. Uzhhorod National University Herald. Series: Law. https://doi.org/10.24144/2307-3322.2023.79.2.2.

Daubner, L., & Považanec, A. (2023). Data Loss Prevention solution for Linux endpoint devices. Proceedings of the 18th International Conference on Availability, Reliability and Security. https://doi.org/10.1145/3600160.3605036

Fataftah, F., & Isong, B. (2022). Security Issues and Possible Solutions in Cloud Computing and Big Data: A Review. 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), 1-6. https://doi.org/10.1109/ICECET55527.2022.9872548.

Guha, A., Samanta, D., Banerjee, A., & Agarwal, D. (2021). A Deep Learning Model for Information Loss Prevention From Multi-Page Digital Documents. IEEE Access, 9, 80451–80465. https://doi.org/10.1109/ACCESS.2021.3084841

Gupta, K., & Kush, A. (2023). A forecasting-based DLP approach for data security. Data Analytics and Management. https://doi.org/10.1007/978-981-15-8335-3_1

He, J. (2023). Research on Computer Network Security Prevention in the Context of Big Data. Frontiers in Computing and Intelligent Systems. https://doi.org/10.54097/fcis.v4i3.11147.

Huang, Y., Li, Y. J., & Cai, Z. (2023, June). Security and Privacy in Metaverse: A Comprehensive Survey. Big Data Mining and Analytics, 6(2), 234–247. https://doi.org/10.26599/bdma.2022.9020047

Kapiton, A., Dziuban, O., Franchuk, T., & Yatsenko, I. (2023). Analysis of innovative methods of computer data loss prevention. Èlektronnoe modelirovanie, 45(6), 77-84. https://doi.org/10.15407/emodel.45.06.077

Majdoubi, C., Mendili, S., & Gahi, Y. (2023). Data Security Patterns for Critical Big Data Systems. 2023 IEEE 6th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), 01-08. https://doi.org/10.1109/CloudTech58737.2023.10366149.

Niklekaj, M. (2023). Security in Computer Networks: Threats, Challenges, and Protection. Ingenious. https://doi.org/10.58944/mcne2043.

Paracha, M., Sheeraz, M., Chai, Y., Ahmad, S., Khan, Z., Hussain, S., & Durad, M. (2022). Implementation of two layered DLP strategies. 2022 International Conference on Cyber Warfare and Security (ICCWS), 8-13. https://doi.org/10.1109/ICCWS56285.2022.9998436

Shahzad, A., & de Sousa, E. M. (2021). Data Loss Prevention from a Malicious Insider: Cloud Service Providers' Perspective. PACIS, 153.

Shishodia, B., & Nene, M. (2022). Data leakage prevention system for internal security. 2022 International Conference on Futuristic Technologies (INCOFT), 1-6. https://doi.org/10.1109/INCOFT55651.2022.10094509

Stallings, W. (2020). Data Loss Prevention as a Privacy-Enhancing Technology. Journal of Data Protection & Privacy, 3(3), 323-333. https://www.henrystewartpublications.com/sites/default/files/JDPP3.3Datalossprevention asaprivacyenhancingtechnology.pdf

Tang, J., Zhu, H., Lu, R., Lin, X., Li, H., & Wang, F. (2022). DLP: Achieve customizable location privacy with deceptive dummy techniques in LBS applications. IEEE Internet of Things Journal, 9, 6969-6984. https://doi.org/10.1109/jiot.2021.3115849

Thakur, A., Zhu, T., Abrol, V. et al. (2024). Data encoding for healthcare data democratization and information leakage prevention. Nat Commun 15, 1582. https://doi.org/10.1038/s41467-024-45777-z

Wu, J., Wang, J., Nicholas, S., Maitland, E., & Fan, Q. (2020). Application of big data technology for COVID-19 prevention and control in China: lessons and recommendations. Journal of medical Internet research, 22(10), e21980. https://www.jmir.org/2020/10/e21980/

Zuo, C., Lin, Z., & Zhang, Y. (2019, May). Why does your data leak? uncovering the data leakage in cloud from mobile apps. In 2019 IEEE Symposium on Security and Privacy (SP) (pp. 1296-1310). IEEE.