3 Data Loss Prevention Systems: Security vs Privacy

Konstantinos Limniotis and Georgia Panagopoulou, Hellenic Data Protection Authority

Abstract. Data loss prevention systems, from a data protection point of view, are studied in this paper. More precisely, data protection issues that arise from the use of data loss prevention systems as a security control are being considered, stating appropriate measures to be taken in order to apply the privacy-by-design principle in the operation of such systems.

Introduction

It is widely known that organisations should put much effort on ensuring the compliance with the personal data protection legislation; apart from the aforementioned obligation, avoiding data breaches is of high importance as the business impact of a data breach could be disastrous. To this goal, several organizational and technical measures are being implemented, whose effectiveness should be constantly evaluated.

Among other security controls, data loss prevention (DLP) systems have a prominent role. By contrast to intrusion detection systems, which aim at scanning incoming traffic, DLPs focus on internal traffic and outgoing data - i.e data that leave the company. Hence, by this way, the company prevents any potential unauthorized use or transmission of proprietary data, independently from whether such an action is unintentional or not. It should be pointed out that the Hellenic Data Protection Authority has received some inquiries, as well as notifications, from data controllers regarding the use of DLPs, mainly as a response to recommendations regarding the need for applying high-level security measures.

However, the special nature of DLP systems poses several risks with respect to the privacy of employees (and, probably, of third parties); this stems from the fact that such systems automatically gather large amounts of personal data and, thus, investigating such data may infringe the privacy. Hence, the adoption of a DLP solution is a decision, which should be taken after careful assessment with all involved stakeholders.

This paper focuses on data protection issues that arise from the use of DLP systems that monitor the outgoing e-mail traffic. More precisely, the aim of the paper is to exhibit the privacy risks that occur when such a DLP system is in place, as well as to propose general rules to mitigate those risks, according to the Privacy-by-Design principle which is a crucial factor for data protection (and is further strengthened in the proposal for the new General Data Protection Regulation [4]}).

The paper is organized as follows; In Section 2 a short overview of the DLP technology is given, describing the main content-analysis techniques that are being met, while Section 3 emphasizes on the effects to the employees' privacy, providing guidelines for a proper use of such systems. Finally, concluding remarks are given in Section 4.

Data Loss Prevention Systems: A short overview

In general, a DLP solution may monitor data in motion (e.g. outgoing e-mails), data at rest (e.g. when users save data on network folders) and data in use (e.g. as users access files) [2]. In any case, a content-analysis is being performed, towards deciding whether a specific action is admissible or not (according to well-defined policies). The widely used content-analysis techniques include (see, e.g. [5]):

1. Keyword matching, which is based on scanning the content for specific kewords from a list.
2. Rule based and regular expression matching, which is the most common technique. It is based on analyzing the content for specific rules — such as 16-digit credit card numbers.
3. Database fingerprinting, which takes either a database dump or live data from a database and seeks for exact matches.
4. Machine learning (or other statistical-based) algorithms, to analyze the content via an appropriate classifier which has been appropriately trained (similarly to the spam filters).

The core of any DLP solution is the network monitoring, whereas email integration is also a main component which strives to protect from data breaches via electronic mails. However, an automated scanning of outgoing e-mails, with further investigation of those that have been characterized (or intercepted) as suspicious, increase the danger of employees' privacy, since e-mails should benefit from the same protection of fundamental rights as traditional paper mail ([1]). Hence, there is a trade-off between security (i.e. customer's personal data protection) and employees privacy (employees personal data protection), which should be appropriately treated.

Protecting Privacy

In this Section, we describe a set of rules that have to be applied when a DLP solution that is based on outgoing e-mail scanning is under consideration. Our approach is based on forcing the well-known data protection principles (such as proportionality) to such systems - which is certainly a nontrivial task. The goal is to apply the Privacy-by-Design approach in the adoption of a DLP solution, characterized by proactive rather than reactive measures ensuring privacy [3].

More precisely, such a processing will be legitimate if it is necessary for the purpose of the legitimate interests pursued by the companies (data controllers), provided that the processing does not violate the rights and freedoms of the data subjects. To this end, the following steps seem to be prerequisite:

- Before implementing a DLP solution, a Privacy Impact Assessment (PIA) should be conducted, to assess the privacy and data protection impacts of such a choice (with the view of examining all possible alternatives to prevent those impacts). The outcome of the PIA should be a fully justified decision regarding the necessity or not of the DLP solution.
- Ensuring transparency of the system is essential. The employees should be provided with full information regarding the operation of the DLP system, whereas accurate internal rules and regulations should be in place. Clearly, the corresponding policy should explicitly refer to the fact that the work e-mail account is intended for work purposes and not for personal communications, so as to avoid monitoring personal communication; hence, the company may scan only e-mails stemming from e-mail addresses which are given by itself (and not e-mails that are exchanged via personal accounts - e.g. via webmail).
- It is crucial to choose appropriate techniques for identifying suspicious e-mails, so as to minimize false positives alarms.
- It is important to force appropriate access rights so as to ensure that only authorized trustworthy persons may have access to data stored by DLPs. Clearly, the need-to-know principle should be applied (for instance, there is no need for a system administrator to have access to these data). To this goal, it should be also pointed out that personal data in the warnings (alarms) should be minimized.

- The DLP log data may, in most cases, contain personally identifiable information, and is therefore subject to the personal data protection legislation, therefore employees must be able to exercise their subject's rights to information, access, deletion to their personal data contained in DLPs log files.
- Third parties' personal data should not be processed by the DLP system. Note that an e-mail may contain such data (for instance, in case of e-mail conversations).
- Data that have been investigated and considered to be innocent, should be immediately deleted.

Note that the above list is not exhaustive; a proper decision of whether a DLP solution of such type is good choice or not should be made, in general, on an ad-hoc basis, according to the aforementioned PIA outcome. The PIA should also determine the appropriate measures that accompany the deployment of the solution.

Conclusions

DLP systems can be an important security control for securing personal data processing but in the same time poses several risks with respect to the processing of employees and third parties personal data. Applying appropriate measures is prerequisite for mitigating those risks and striking the proper balance between legitimate interests of organizations to protect their data and the fundamental right to the protection of individuals’ personal data.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments and suggestions, which helped to improve the presentation of the manuscript.

References

1. Article 29 - Data Protection Working Party: Working document on the surveillance of electronic communications in the workplace. May 2002 (https://www.enisa.europa.eu/activities/cert/support/proactive-detection/proactive-detection-report).
2. Daley, M. J., Fey, L. C., and Fashing, D.: Exploring Data Loss Prevention Systems for Legal Holds and e-Discovery. Information Management, vol. 44. n. 5, pp. 26--30, Sept. 2010, ARMA International.
3. Enisa Report: Proactive Detection of Network Security Incidents. Deliverable - 2011-12-07 (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2002/wp55_en.pdf).
4. European Commission: Proposal for the EU General Data Protection Regulation. (http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf)
5. Rick, M.: Understanding and selecting a data loss prevention solution. Securosis Report, 2010.(https://securosis.com/assets/library/reports/Understanding_and_Selecting_DLP.V2_.Final_.pdf)