Eliciting Requirements for a Company-wide Data Leakage Prevention System

Stefan Gärtner[1], Svenja Schulz[1], Kurt Schneider[1], and Steffen Förster[2]

[1]Software Engineering Group, Leibniz University Hannover, Germany
{stefan.gaertner,kurt.schneider}@inf.uni-hannover.de
[2]Continental Information Technology Hannover, Germany
steffen.foerster@conti.de

1 Introduction

Sensitive information assets are critical to business as their leakage can harm a company's competitiveness and reputation [1]. To prevent information disclosure, companies increasingly make use of *Data Leakage Prevention (DLP)* systems [6]. As company-wide DLP systems have to operate in highly dynamic and heterogeneous working environments, the challenge is to identify requirements for configuring these systems properly. To approach this challenge, we investigate the DLP system in test at Continental Tires. As a result, we propose an approach to support elicitation of desired requirements for DLP systems and discuss preliminary results.

2 Data Leakage Prevention System

To protect assets properly, companies have not only to focus on external attacks but also on threats from inside [3]. These threats include deliberate information disclosure caused by social engineering or intentional misuse of privileges as well as accidental disclosure by employees due to insufficient security policies or careless behavior [2]. To prevent information from leaking outside the company, DLP systems are used which provide mechanisms for identifying sensitive information by content [6].

The DLP system in test at Continental Tires consists of four major components as presented in Figure 1. The monitoring component contains various sensor elements that capture access information of the regarded environment (e.g. email, HTTP/SSL, file system, printer, fax, removable media). To reduce the amount of information incurred, the content is classified, based on pre-defined keywords, whether it is confidential with respect to the company's business field.

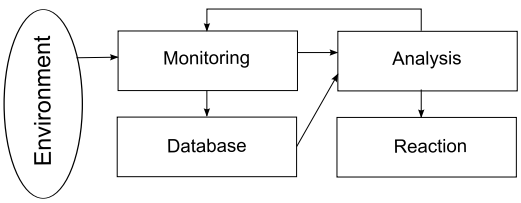

Figure 1: Architecture of the considered DLP system.

Classified information are stored in a database and searched for suspicious access patterns by the analysis component. If the DLP system detects a potential incident, the corresponding access is blocked by the reaction component. Additionally, the incident is reported to the line manager who must approve the access. To consider changing environments, the analysis component is able to adapt sensors of the monitoring component.

Requirements for these components are manifold depending on the working environment, employees, and business processes.

3 Requirements Elicitation Approach

To consider any malicious behavior of employees in advance for configuring DLP systems is a difficult task. Therefore, we need to analyze their normal behavior according to their usual working activities. Sudden deviation from this behavior is considered malicious and corresponding actions must be blocked. The problem is, however, that employees behave differently based on their work (heterogeneous working environment). For example, a product manager has to send sensitive specification documents to the customer or to a subcontractor. Although this is a valid activity for a product manager, it might be forbidden for developers. Thus, a more complete picture of the employees' behavior must be obtained.

As analyzing each employee of a large company is time-consuming, we therefore combine data mining of access information with information flow analysis. Based on the behavior with respect to the working activities and environment, requirements for the DLP system can be derived. Our elicitation process consists of the following activities:

1. **Identification:** Identify different personas within the company (e.g. managers, worker, etc.) and select an employee representing a certain persona. They help to understand the real users in terms of working activities and environment.

2. **Pilot Monitoring:** Record fine-grained access and communication information of the chosen representatives. Determine association rules describing their normal working behavior.

3. **Interviewing:** Conduct interviews with the chosen representatives to elicit coarse-grained information about their current working activities from an information flow perspective.

Within the initial study at Continental Tires, we identified three different personas: (1) executive staff, (2) mid-level staff, and (3) skilled workers. Due to privacy issues within the company, we mainly focused on executive and mid-level staff. In the pilot monitoring phase, we used RapidMiner[1] to determine relevant association rules from recorded data. To acquire the working activities from an information flow perspective, we applied our *FLOW* method as described in [4, 5]. For this purpose, *FLOW* provides an elicitation sheet to guide interviews with employees. To adapt *FLOW* to our needs, we extend the elicitation sheet to gather information about used document types and their confidentiality.

By analyzing recorded data, we identified following primary document types that were used during the study: (1) Portable Document Format (47%), Microsoft Excel Format (34%), Microsoft Word Format (2%), and Microsoft Powerpoint Format (13%). However, it has been stated within the interviews that some activities require extensive cooperation and data exchange with external people. For this purpose, Microsoft SharePoint and ContiView are used to manage and to share confidential documents. Unfortunately, these tools are not monitored by the current DLP system at all.

Based on the interviews, measurement data has been identified as a new document type. Although confidentiality of measurement data has been rated differently, interviewees agreed that this data must be protected from leakage. As measurement data consists of numbers and is rarely described by text, the current DLP system is not feasible to protect this data effectively. In both cases, new requirements of the DLP system have been identified based on the interviews.

Besides using the mentioned elicitation sheet for interviews, we also found it very helpful to use mined rules from recorded data to focus interviews on conspicuous behavior. For example, it has been found in our study that one user frequently uses a backup tool which copies confidential data. In the interview, this behavior has been justified as part of the established business process that should not be blocked for the specific user as long as the tool is used in the same manner as monitored. This example also shows that analyzing recorded data helps to identify work practices facilitating leakage by accident.

Although interviews are time-consuming, they were indispensable to understand the working environment and business processes. However, interviewees could not tell reliably how a specific tool or communication service was used. To obtain these information, recorded data and corresponding association rules has been used successfully. They could be manually assigned to the activities and tools reported during the interviews without significant problems. Thus, both techniques complement each other in order to elicit requirements of company-wide DLP systems efficiently.

In summary, the following preliminary findings from our initial study are highlighted:

- To elicit requirements on what tools and communication services are used within the company, information flow analysis focusing on working activities are indispensable.
- To extract requirements on how tools and services are used, fine-grained access and communication information that have been recorded within the monitoring phase are needed. Moreover, this information can also be used to evaluate whether tools and services support data leakage by accident (e.g. external access privileges, email encryption).
- Executive and mid-level staff has very different tasks which constitute various requirements for DLP systems.

4 Conclusion

In this paper, we proposed a first attempt for requirements elicitation to configure company-wide DLP systems properly. Therefore, we combined fine-grained access and communication information with information flow analysis. Since our preliminary results are promising, we plan to extend the evaluation of our approach and to validate gathered requirements. As information disclosure is difficult to measure in an industrial setting, we need to engage security experts trying to disclose information in a realistic manner.

Acknowledgement

This work was partially supported by the DFG (German Research Foundation) under the Priority Programme SPP1593: Design For Future — Managed Software Evolution.

References

[1] S. Liu and R. Kuhn. Data loss prevention. *IT professional*, 12(2):10–13, 2010.

[2] A. Shabtai, Y. Elovici, and L. Rokach. *A survey of data leakage detection and prevention solutions*. Springer, 2012.

[3] G. J. Silowash and C. King. Insider Threat Control : Understanding Data Loss Prevention (DLP) and Detection by Correlating Events from Multiple Sources. Technical report, Software Engineering Institute (SEI), 2013.

[4] K. Stapel, E. Knauss, and K. Schneider. Using FLOW to Improve Communication of Requirements in Globally Distributed Software Projects. In *Workshop on Collaboration and Intercultural Issues on Requirements: Communication, Understanding and Softskills (CIRCUS '09)*, 2009.

[5] K. Stapel and K. Schneider. Managing Knowledge on Communication and Information Flow in Global Software Projects. *Expert Systems - Special Issue on Knowledge Engineering in Global Software Development*, 2012.

[6] T. Takebayashi, H. Tsuda, T. Hasebe, and R. Masuoka. Data loss prevention technologies. *Fujitsu Scientific and Technical Journal*, 46(1):47–55, 2010.

[1] http://rapidminer.com/