Data Leakage Prevention
A Position to State-of-the-Art Capabilities and Remaining Risk

Barbara Hauer
Institute of Systems Software, Johannes Kepler University Linz, 4040 Linz, Austria

Keywords: Security, DLP, Data Leakage Prevention, ILP, Information Leakage Prevention, Information Exposure.

Abstract: Organizations from all around the world are facing a continuous increase of information exposure over the past decades. In order to overcome this thread, out of the box data leakage prevention (DLP) solutions are applied which are used to monitor and to control data access and usage on storage systems, on client endpoints, and in networks.

In recent years products from market leaders, such as McAfee, Symantec, Verdasys, and Websense, evolved to enterprise content-aware DLP solutions. However, this paper argues that current out of the box solutions are not able to reliably protect information assets. It is only possible to reduce the probability of various incidents if organizational and technical requirements are accomplished before implementing a DLP solution. To be efficient, DLP should be a concept of information security within the information leakage prevention (ILP) pyramid which is presented in this paper. Furthermore, data must not be equalized with information which requires different strategies for protection. Especially in case of misusing privileges by exploiting an unlocked system or by shoulder surfing, the remaining risk must not to be underestimated after all.

1 INTRODUCTION

Not least Edward Snowden (Greenwald et al., 2013) brought to mind that data leakage is ever-present. And insiders are able to take data, even documents with a lot of information, out of the company office and the organization is unaware of the leakage. Regularly, the leakage is only notices if the data is published. It seems almost incredible that data leakage even affects organizations like the National Security Agency (NSA). The NSA assumed to understand their data lifecycle, to have the most sophisticated information technology (IT) for security, and to operate on the highest thinkable information security level. Hence, it is fascinating that Edward Snowden could easily steal such an amount of data. This and other incidents caused a huge upturn in sales figures because organizations buy data leakage prevention (DLP) solutions to avoid data leakage and information exposure issues.

In general, this paper refers to DLP as data leakage prevention but the definition of the term "DLP" varies and a consistent terminology is absent. Therefore, terms such as data loss prevention (DLP), extrusion prevention (EP), content monitoring and filtering (CMF), content monitoring and protection (CMP), information leakage prevention (ILP), information leak prevention and detection (IDLP), outbound content compliance (OCC), or information protection and control (IPC) are commonly used as synonyms.

Information exposure, which is addressed by ILP, is "the intentional or unintentional disclosure of information to an actor that is not explicitly authorized to have access to that information" (CWE, 2013). This work refers to DLP as a part of ILP since DLP focuses on data and the ISO/IEC 2382 standard specifies data as a "representation of information in a formalized manner, which should be suitable for communication, interpretation, or processing" (JTC 1, 1993).

In terms of DLP, the detection and classification of sensitive data has to consider the state of the data which can be data in motion (DIM), data in use (DIU), or data at rest (DAR). Additionally, the classification according to the information security level influences data handling. For example, data classified as "top secret" is subject to other restrictions than data classified as "restricted" or "public". Therefore, DLP solutions rely on reasonable strategies for detection and classification of sensitive data. Commonly, the strategies to detect and classify data are based on

- key word and regular expression search,
- digital fingerprints,

Hauer B..
Data Leakage Prevention - A Position to State-of-the-Art Capabilities and Remaining Risk.
DOI: 10.5220/0004951703610367
In *Proceedings of the 16th International Conference on Enterprise Information Systems* (ICEIS-2014), pages 361-367
ISBN: 978-989-758-028-4

- data tagging, and
- machine learning techniques.

Several of these strategies are extended by analysis methods based on awareness of context and content. Nevertheless, the detection and correct classification of data are still a major challenge due to the existence of encryption, hidden channels, unsupported content types, as well as large amounts of multimedia and unstructured data.

The following section introduces state-of-the-art enterprise content-aware DLP solutions of well-known market leaders. This work states that DLP should be characterized as a concept of information security and therefore, the ILP pyramid, which is described in section 3, should be considered. The remaining risks of this concept are particularized in section 4.

2 STATE-OF-THE-ART DLP SOLUTIONS

DLP solutions offer different approaches to monitor and to protect confidential data at client endpoints. Typically, these solutions validate and authorize applications before allowing confidential data to be transferred and to be migrated. Furthermore, data usage on client endpoints and network traffic are monitored, and copy and paste operations and taking screenshots can be prevented. In general, current DLP solutions are able to control the access to confidential data and the utilization of this data by the user. Moreover, the systems can prevent unauthorized users or applications to obtain confidential data. But there are several issues and limitations.

According to a report on "Enterprise content-aware DLP" (Ouellet, 2013) which was published in 2013, DLP solutions offering a holistic approach and functionalities for DIM, DIU, and DAR are referred to as market leaders. In this context, the report identified the market leaders EMC - RSA, McAfee, Symantec, Verdasys, and Websense. Compared to the Forrester Wave report published in 2008 (Raschke, 2008), the leadership in the market has been slightly shifted due to various buyouts. These days, the market leaders offer additional functionalities for DIM, DIU, and DAR since they agreed on further incorporations and they implemented findings of academic research.

Some functionalities of DLP solutions, such as restrictions for copying data to unauthorized removable devices, taking screenshots, and printing files as well as the support of data encryption, are standard. However, the DLP solutions of the market leaders differ in the approaches and the implementation details.

2.1 EMC - RSA

EMC offers the "RSA Data Loss Prevention Suite" (EMC Corporation, 2013) which includes the "RSA DLP Enterprise Manager", the "RSA DLP Datacenter", the "RSA DLP Network", and the "RSA DLP Endpoint" in version 9.6. The "RSA DLP Datacenter" performs scans on client endpoints, data repositories, file shares, databases, storage systems (SAN/NAS), and Microsoft SharePoint products. These scans make use of key word and regular expression search as well as digital fingerprints to find sensitive DAR in present described-content and fingerprinted-content. The "RSA DLP Network" addresses DIM by monitoring network traffic. However, its capabilities are limited to the internet protocol (IP) version 4 and 6, and common higher-level protocols such as the hypertext transport protocol (HTTP) and the simple mail transfer protocol (SMTP). DIU and DAR are handled by the "RSA DLP Endpoint" which intercepts application calls and scans the involved content to prevent disallowed user actions. The EMC DLP suite does not support Linux based operating systems, mobile device management (MDM), and cloud infrastructures. Furthermore, the solution is specialized in Microsoft products and, according to the EMC - RSA website, Microsoft is a customer using the "RSA DLP Datacenter" for compliance with payment card industry (PCI) and Sarbanes-Oxley regulations.

2.2 McAfee

Like other McAfee products, "McAfee Data Loss Prevention 9.3.0" (McAfee, Inc. , 2013) is based on the McAfee ePolicy Orchestrator (McAfee ePO) server . The "McAfee DLP Monitor" and "McAfee DLP Prevent" address DIM while "McAfee DLP Discover" deals with DAR and the "McAfee DLP Endpoint" covers DIU. Basically, the "McAfee DLP Monitor" makes use of a switched port analyzer (SPAN) port or a network tap to passively monitor network traffic and to determine the sender, the data type, and the destination. "McAfee DLP Prevent" is able to block or redirect network traffic but the tool is limited to web or e-mail traffic per appliance, 30 concurrent SMTP connections, and 4000 concurrent internet content adaptation protocol (ICAP) connections from a web proxy server. In General, key word and regular expression search as well as digital fingerprints are used for data tagging. In order to improve key word search which includes a dictionary search, McAfee recommends using whole phrase matching or statistically improbable phrases (SIPs). Digital fingerprints are used to create data and file signatures. Furthermore, "McAfee DLP Discover" classifies content by document property definitions which are based on predefined metadata values and filename extensions. User actions on client endpoints are addressed by the "McAfee DLP Endpoint" which monitors data usage and prevents, for example, copying data to removable media, printing files, and taking screenshots. In further consequence, rights management and role based access control are supported. Similar to the EMC solution, the McAfee solution is specialized in Microsoft products and does not offer sustainable support for Linux based operating systems and cloud infrastructures. MDM is offered separately for mobile devices such as Apple iPhones, Apple iPads, Android devices, and Windows Phones. However, McAfee is aware of some limitations and known issues. For example, Windows does not load the host DLP plugin in safe mode, and as a result the web host protection rules do not work and e-mail protection rules are bypassed in some cases (McAfee, Inc., 2013).

2.3 Symantec

Consisting of multiple parts, "Symantec Data Loss Prevention 12" (Symantec Corporation, 2013) can be installed on Red Hat Enterprise Linux operating systems as well as on Microsoft Windows Server operating systems. However, the endpoint agents are limited to Microsoft Windows operating systems. Symantecs DLP solution for DIM is software-based and consists of three products: the "Symantec Data Loss Prevention Network Monitor", the "Symantec Data Loss Prevention Network Prevent for Email", and the "Symantec Data Loss Prevention Network Prevent for Web". Monitoring and prevention only effect protocols which are enabled in the system, such as the HTTP and transport layer security (TLS) protocols. DLP for e-mail involves smartphones as well as tablets running Google Android, Apple iOS, BlackBerry, and Windows Mobile. The support for web services, social media, and cloud infrastructures is limited to specific providers. Symantecs DLP solution for DAR is composed of "Symantec Data Loss Prevention Network Discover", "Symantec Data Loss Prevention Network Protect", and "Symantec Data Loss Prevention Data Insight Enterprise". "Symantec Data Loss Prevention Endpoint Discover", and the "Symantec Data Loss Prevention Endpoint" are responsible for DIU and available for Windows clients endpoints. In contrast to EMC and McAfee, Symantec not only makes use of key word and regular expression search as well as digital fingerprints, but also deploys vector machine learning techniques for building statistical models based on positive and negative example documents. In addition to detect DAR by scanning data repositories including file servers, databases, and web sites, Symantec also tracks the file usage which can be used to enforce access rules and to understand leakage incidents. However, the DLP solution is limited to certain file types, data formats, network protocols, storage systems, service providers, and software vendors.

2.4 Verdasys

Verdasys DLP solution "Digital Guardian (DG) version 6" with "DLP 3.0" (Verdasys, 2013) specializes in unstructured data and extended operating system support to have an advantage over its competitors. Verdasys defines the DG as an enterprise information protection (EIP) solution which implements a data-centric approach. The "Digital Guardian Management Server" is the command center for operating several agents and various add-on modules. The agents are used for context-based data monitoring, classification, and control, and to enforce data policies on Windows, Linux, Mac OS, VMware, Citrix, Hyper-V, BlackBerry Enterprise Server, Exchange ActiveSync, and iOS platforms. Verdasys offers different network agents, such as "DG XPS DIRECT", "DG XPS MAIL", "DG XPS WEB", and "DG NETCOM", which include an agreement with Fidelis Security Systems for using the "Fidelis Extrusion Prevention System (XPS)". These network agents are deployed as out-of-band sniffers or inline layer 2 bridges. Basically, they try to detect unauthorized DIM based on content, application, and/or protocol across all 65,535 ports. Hence, data detection and classification are shifted to the endpoints and the storage systems. In general, the data classification is based on content, context, and user classification (UC) all of which are complementary and can be combined. The content inspection makes use of key word and regular expression search as well as document similarity based on key words and Bayesian analyses. The context, for example, involves the application, data type, user identity, e-mail properties, and network properties. The classification is stored along with policy rules in meta-tags which allows inheritance and reclassification. Unstructured data is identified and classified according to context parameters by considering user, application, and activity, such as the creation, access, revision, or transmission. Continuous logging and auditing can be extended by key logging, and content and screen capturing, which is used to analyze behavior and to reconstruct leakage incidents.

2.5 Websense

Websense combines the “Websense Data Security Suite”, “Websense Email Security Gateway Anywhere”, and “Websense Web Security Gateway Anywhere” to the “Websense Triton Enterprise Suite” (Websense, Inc, 2013). DLP is particularly addressed by the “Websense Data Security Suite” which, in version 7.8, is comprised of the “Websense Data Security Gateway”, the “Websense Data Endpoint”, and “Websense Data Discover”. The “Data Identification and Classification Engine (DICE)” is embedded within the stated solutions. In order to discover sensitive data on specified network and endpoint systems, a scan is performed on network file systems, SharePoint directories, database servers, Exchange servers, Outlook PST files, and IBM Lotus Domino documents. Different agents are used to scan and monitor the endpoint systems which are limited to Microsoft Windows, Apple Mac OS, Apple iOS, Red Hat Enterprise Linux, and Cent OS Linux platforms. Several cloud storage infrastructures are supported, too. The protection differs according to the operating system and the supported applications. Corresponding the vendor, the solution is able to detect custom encryption and pays attention to the geographical location, and the source and destination resource categories when policies and classification rules are applied. The classification itself makes use of key word and regular expression search, fingerprints, file properties, support vector machine (SVM) models, and optical character recognition (OCR). Furthermore, the vendor claims that “Drip DLP” is able to protect data from timing attacks which leak data slowly by cumulative events. However, this capability is restricted to supported network traffic such as e-mail, web, and IM protocols.

2.6 Academic Research

In addition to industrial vendors, academic research also provides various approaches to prevent data leakage. Some of these approaches directly apply to DLP while others have their origin in related research fields such as data mining algorithms, network anomalies, user behavior, and mobile device security.

In general, DLP classification techniques benefit from data mining and knowledge discovery in databases (KDD) techniques when it comes to the analysis and/or classification of large amounts of data. The major problem of such machine learning methods is to provide adequate training data. This can be addresses by certain techniques, e.g. considering text classification from positive and unlabeled documents (Yu et al., 2003) or effective multi-label active learning for text classification (Yang et al., 2009).

Moreover, there are approaches to handle large bulks of data in networks. Such techniques are able to quickly find anomalies in large data streams by visualizing (Hao et al., 2009), to enhance network intrusion detection systems (NIDSs) for almost real-time automatic network control, and to analyze behavior in communication channels to detect malicious network traffic (Beaver et al., 2013).

A further challenge with respect to DLP is created by covert channels. DLP solutions frequently outsource the detection of these channels from network monitoring to endpoint surveillance or even disregard their existence. Nevertheless, there are approaches in academic research which, for example, block network covert timing channels (Wang et al., 2009) or perceive covert channels (Jaskolka and Khedri, 2011).

On the other hand, there are research approaches focusing on DLP itself. Examples are the published design of a framework for detecting an insider’s leak of confidential information (Baek et al., 2008), the anomaly detection based on usage patterns or user profiles according to events in system logs (Corney et al., 2011), or the mechanisms for detecting and preventing data exfiltration by insiders which are based on DBMS-layer anomaly detection and prevention using provenance tracking and virtualization (Bertino and Ghinita, 2011).

Some of these academic research approaches might be able to become part of an industrial product. Of course, this takes time and the gap between the theoretical possibility and the practical implementation has to be closed. Furthermore, the majority of threats can only be prevented by coordinated functional interaction of security mechanisms.

3 DLP AS A CONCEPT OF INFORMATION SECURITY

Based on computability theories like the Kurt Gödel’s incompleteness theorem, the Alan Turing’s halting theorem, and the Rice-Myhill-Shapiro theorem, a perfect IT security solution is not possible. Accordingly, IT security is a matter of probability. On the one hand, DLP solutions should minimize the probability of data leakage. On the other hand, this means to maximize the probabilities of detection, identification, correct classification, and prevention. In order to be efficient, a large amount of topics in the domain

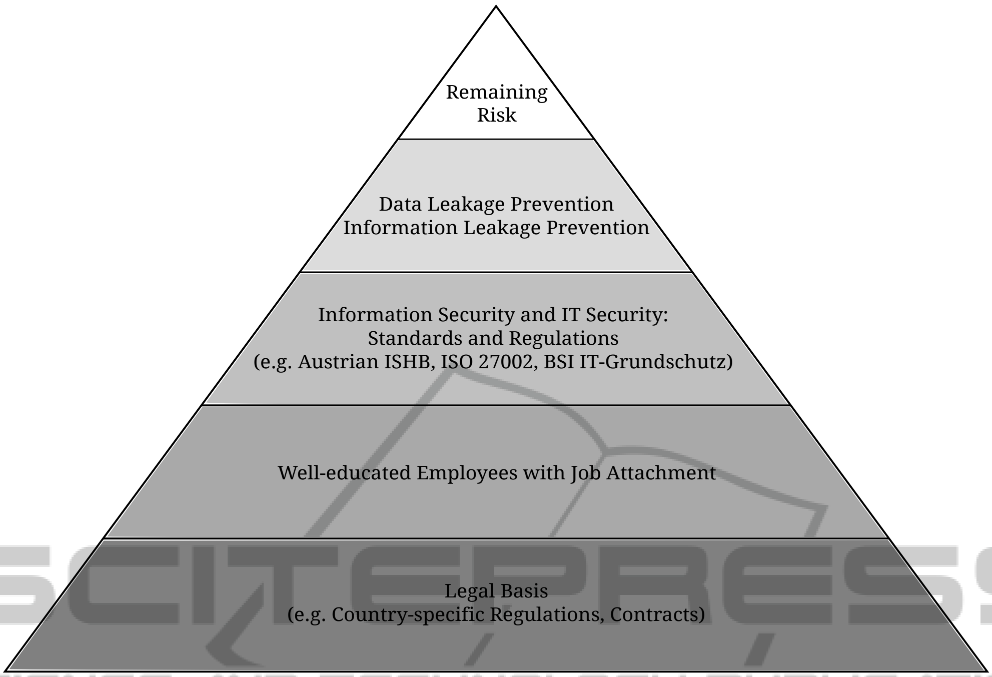
Remaining
Risk
Data Leakage Prevention
Information Leakage Prevention
Information Security and IT Security:
Standards and Regulations
(e.g. Austrian ISHB, ISO 27002, BSI IT-Grundschutz)
Well-educated Employees with Job Attachment
Legal Basis
(e.g. Country-specific Regulations, Contracts)

Figure 1: DLP as a concept of information security in the information leakage prevention pyramid.

of information security have to be considered. Subsequently, a solution based on a single product which covers isolated topics is doomed to failure. By increasing the number topics in the domain of information security covered by a DLP solution, the probability of data leakage can be reduced.

This work proposes a strategy which characterizes DLP as a concept of information security. The ILP pyramid, which is demonstrated in Figure 1, represents the DLP concept as a part of ILP within the domain of information security and information security precautions. The levels of the ILP pyramid are correlated and the objective is to minimize the probability of data leakage which is illustrated as the remaining risk at the top of the pyramid.

In order to establish a basis for DLP, country-specific regulations, such as contracts of employment and business partner contracts, have to be observed. The legal basis for permanent surveillance, for example, may demands to inform involved parties about their rights and obligations, and to obtain their consent. In Austria breaches of security and supervisory measures can be prosecuted by criminal law as well as by civil law. Clear agreements within contracts as well as non-disclosure agreements are strongly recommended.

The next level of the ILP pyramid focuses on employees. Organizations have to pay attention to the continuous education of their employees to kept their competence, training, and awareness up to date. In general, the employers have to be mindful of a positive working environment which benefits employee satisfaction and job attachment. These characteristics increase trust on both sides, the organization and the employees. According to Blank's results, a central role in the organizational context is assigned to trust due to the fact that trust saves expenses and enhances the quality of work (Blank, 2011). Hence, trust is a cost factor when implementing DLP. On the other hand, control mechanisms, which are a major part of DLP solutions, may have a negative influence on the employee's trust since it gives the impression of the organization not trusting its employees. A well defined basis for conversation positively effects the acceptance of regulations and control mechanisms, and therefore the employee's inhibition of offending against and of bypassing those regulations and control mechanisms. Basically, employee satisfaction and job attachment also influence the staff fluctuation and as a consequence the return on investment (ROI). This is because new employees necessitate a period of vocational adjustment, accompanied by coordinated education and training, to be highly productive. Moreover, well-educated and motivated employees are required to implement, configure, manage, supervise, and operate the technical installations within the infrastructure of an organization. All IT security installations, like DLP solutions, are among those technical installations and therefore, organizations have to place confidence in their proper functioning.

Compliance to information security standards and regulations constitutes the third level of the

ILP pyramid. The objective is to gain a high information security level by considering a large amount of threads. For example, the ”BSI IT-Grundschutz-Kataloge“ (BSI IT Baseline Security Catalogs) (BSI, 2013) published by the German Federal Office for Information Security (BSI) specifies threads which can lead to unintentional information exposure, loss of confidentiality of information, loss of data integrity, or data and information leakage. In Austria the ”Österreichische Informationssicherheitshandbuch“ (Austrian Information Security Handbook) (BKA, ISB und A-SIT, 2012) has to be considered. This document is compatible with the ISO/IEC 27001 standard (JTC 1/SC 27, 2005a) and takes ISO/IEC 27002 controls (JTC 1/SC 27, 2005b) into account. Basically, it specifies the confidentiality of information being an asset which has to be protected and preserved, and for that reason organizations have to ensure that information is not made available or disclosed to unauthorized entities. Furthermore, measures to guarantee the confidentiality and integrity of sensitive data are recommended. Of course, these objectives correspond with the objectives of DLP and ILP and therefore, compliance to information security standards and regulations are a prerequisite for DLP and ILP.

The legal basis, well-educated employees with job attachment, and a high information security level provide a reasonable starting point for an ILP concept. At this point an organization should have acquired a decent access control management, security classifications, and the knowledge about the location of each information asset. Furthermore, its employees should be ”aware of information security threats and concerns, their responsibilities and liabilities“ (JTC 1/SC 27, 2005b) and willing to comply with instructions. An ILP concept has to address the remaining information security issues, such as DLP and the advantageous usage of DLP within an existing technical infrastructure. A separate DLP concept can be elaborated to handle the challenging task of meeting the organizational and technical requirements.

However, several intended or unintentional types of information disclosure as well as data leakage cannot be prevented. Hence, there is a remaining risk which can be reduced but not eliminated.

4 REMAINING RISK

According to the book (Shabtai et al., 2012) which was published in 2012, industrial DLP solutions are mainly utilized to prevent accidental leakage incidents. These solutions evolved and recently offer protection against various malicious insider threats by applying more sophisticated detection, classification, and prevention techniques. Nevertheless, information and data exposure incidents are ever-present in the media. There are approaches to detect or prevent data leakage by behavioral analytics. For example, Verdasys advertised to implement sophisticated usage behavior analyses which monitor sensitive data usage up to six months or even longer. Based on the assumption that there is conspicuous behavior, a large bulk of data is collected and analyzed. In fact, preventing data leakage becomes more difficult if the leakage arises from data usage of an authorized person and no abnormal behavior can be detected. For example, there is no forbidden or illegal behavior involved if an authenticated and authorized user displays confidential data on a mobile device outside the company office. In case of misusing privileges by exploiting an unlocked system or by shoulder surfing, state of the art DLP solutions are not able to provide the required protection.

Moreover, it seems that vendors still assume the term ”dumbest assumable user (DAU)“. Considering that administrators of IT security systems and persons with knowledge of all control and IT security measures can cause data leakage, it would be more advantageous to assume the term ”cleverest assumable user“ when it comes to security.

5 CONCLUSIONS

This work argues that organizations cannot buy out of the box DLP solutions and trust in solving data leakage and information exposure issues. In fact, even the most sophisticated state-of-the-art enterprise content-aware DLP solution is not able to do so. In the last years DLP solutions evolved but they are far off being a silver bullet for data leakage and even less for information exposure. They are not able to stand a chance if an organization does not acquire the required organizational and technical qualifications. Therefore, this work proposes the ILP pyramid which characterizes DLP as a concept of information security. Creating a legal basis, taking care of well-educated employees with job attachment, and accomplishing a high information security level is the starting point for a successful ILP concept. This concept can reduce the probability of incidents but there is a remaining risk which cannot be eliminated.

REFERENCES

Baek, E., Kim, Y., Sung, J., and Lee, S. (2008). The Design of Framework for Detecting an Insiders Leak of Confidential Information. In *Proceedings of the 1st International Conference on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop (e-Forensics)*.

Beaver, J. M., Symons, C. T., and Gillen, R. E. (2013). A Learning System for Discriminating Variants of Malicious Network Traffic. In *Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop (CSIIRW)*, Oak Ridge, USA.

Bertino, E. and Ghinita, G. (2011). Towards Mechanisms for Detection and Prevention of Data Exfiltration by Insiders. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (ASIACCS)*, Hong Kong, China.

BKA, ISB und A-SIT (2012). *Österreichisches Informationssicherheitshandbuch*. Bundeskanzleramt Österreich (BKA), Schweizer Informatikstrategieorgan des Bundes (ISB) und Zentrum für sichere Informationstechnologie - Austria (A-SIT), Wien, Austria.

Blank, N. (2011). *Vertrauenskultur: Voraussetzung für Zukunftsfähigkeit von Unternehmen*. Gabler Verlag - Springer Fachmedien, Wiesbaden, Germany.

BSI (2013). *BSI IT-Grundschutz-Kataloge*. Bundesamt für Sicherheit in der Informationstechnik (BSI), Bonn, Germany.

Corney, M., Mohay, G., and Clark, A. (2011). Detection of anomalies from user profiles generated from system logs. In *Proceedings of the Ninth Australasian Information Security Conference (AISC)*, volume 116, pages 23–32, Darlinghurst, Australia.

CWE (2013). CWE-200: Information Exposure. *Common Weakness Enumeration (CWE) for The MITRE Corporation (MITRE)*.

EMC Corporation (2013). RSA Data Loss Prevention (DLP).

Greenwald, G., MacAskill, E., and Poitras, L. (2013). Edward Snowden: the whistleblower behind the NSA surveillance revelations. *The Guardian*.

Hao, M. C., Dayal, U., and Keim, D. (2009). Visual Analytics of Anomaly Detection in Large Data Streams. In *Visualization and Data Analysis (VDA)*, San Jose, USA.

Jaskolka, J. and Khedri, R. (2011). Exploring Covert Channels. In *International Conference on System Sciences (HICSS)*, Hawaii, USA.

JTC 1 (1993). *ISO/IEC 2382-1:1993 Information technology - Vocabulary - Part 1: Fundamental terms*. ISO/IEC Information Technology Task Force (ITTF), Washington D.C., USA.

JTC 1/SC 27 (2005a). *ISO/IEC 27001:2005 Information technology - Security techniques - Information security management systems requirements specification*. ISO/IEC Information Technology Task Force (ITTF), Washington D.C., USA.

JTC 1/SC 27 (2005b). *ISO/IEC 27002:2005 Information technology – Security techniques – Code of practice for information security management*. ISO/IEC Information Technology Task Force (ITTF), Washington D.C., USA.

McAfee, Inc. (2013). McAfee Total Protection for Data Loss Prevention (DLP).

McAfee, Inc. (2013). Data Loss Prevention Endpoint 9.3 Known Issues. Technical Articles ID: KB77168.

Ouellet, E. (2013). Gartner: Magic Quadrant for Content-Aware Data Loss Prevention. *Gartner*.

Raschke, T. (2008). The Forrester Wave: Data Leak Prevention, Q2 2008. *Forrester Research*.

Shabtai, A., Elovici, Y., and Rokach, L. (2012). *A Survey of Data Leakage Detection and Prevention Solutions*. SpringerBriefs in Computer Science. Springer US, New York, USA.

Symantec Corporation (2013). Symantec Data Loss Prevention (DLP).

Verdasys (2013). DLP 3.0 - Data Loss Prevention. Digital Guardian - The Complete Enterprise Information Protection Platform.

Wang, Y., Chen, P., Ge, Y., Mao, B., and Xie, L. (2009). Traffic Controller: A Practical Approach to Block Network Covert Timing Channel. In *International Conference on Availability, Reliability and Security (ARES)*, Fukuoka, Japan.

Websense, Inc (2013). Websense Data Security Suite.

Yang, B., Sun, J.-T., Wang, T., and Chen, Z. (2009). Effective Multi-Label Active Learning for Text Classification. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA.

Yu, H., Zhai, C., and Han, J. (2003). Text Classification from Positive and Unlabeled Documents. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, New York, USA.