Analyzing Human Behavior using Case-Based Reasoning with the help of Forensic Questions

Wolfgang Boehmer*
*Technische Universität Darmstadt, Germany, Hochschulstr. 10, 64289 Darmstadt
Email: wboehmer@cdc.informatik.tu-darmstadt.de

Abstract—**Whether programs are called data loss prevention, content monitoring and filtering, employee activity monitoring, counter corruption, insider trading, or fraud detection, organizations have increasingly implemented projects and initiatives to examine and address insider threats. Insider-perpetrated computer crime is committed by individuals who have permission to use a system, and it is, therefore, based on the actions of trusted users. Most information walks out the front door, not through the firewall.**

This paper presents a theoretical model for analyzing human behavior according to an organization's compliance with legal, institutional, and organizational laws. The theory uses case-based reasoning (CBR) technologies in conjunction with directed acyclic graphs (DAG) and a Hamming similarity function. Defined paths and path deviations in the graphs can be classified to answer automated questions (W7) regarding compliance. The procedure for this model is borrowed from criminology and is referred to as compliance profiling.

Index Terms—**privacy; CBR; DAG; compliance profiling; insider threat**

I. Introduction

The code of practice ISO 27002:2007, Clause 15, explicitly mandates an organization's compliance with institutional and organizational laws. Nevertheless, the number of serious violations recorded in recent years has increased, with the most common violation being accounting scandals. Stock exchanges are sensitive to accounting scandals, because the fraud is often implemented through data manipulation within accounting applications. One response to this problem has been the passage of the Sarbanes-Oxley Act (SOX). Observance of this law can, for example, be achieved by presenting relevant reports to a Certified Public Accountant (CPA). However, monitoring a company's observance of the law or checking whether an application has been manipulated is costly. This article will first investigate what data, questions, and methods are necessary for tracking down manipulations within applications. In particular, a model for automated monitoring of the fulfillment of legal, institutional, and organizational requirements (hereafter referred to in the text as compliance analysis) is presented in this paper.

Von Solms has described the series of developments in information security as having a wave-like structure [1]. The first wave encompasses purely technical aspects, the second wave involves a request for corresponding management, the third wave consists of standardizations. The fourth and last wave is characterized by information security governance. This contribution discusses the fourth and, until now, final wave.

In the next section, the latest challenges to enterprises are explained with respect to the legal and regulatory requirements, and changes in the threat situation are discussed. The new threat situation is characterized by internal attackers that are capable of manipulating certain data. One scenario is that the accounts of a listed enterprise may be altered, which improves the enterprise's current stock market value, for example. Another example may be that an employee of a sales department copies all client contacts prior to quitting his/her job, and the former employee is subsequently employed by a competitor. The internal attacker typically has access to the network and to all applications needed for his/her work. Against this background, it is difficult to distinguish between valid actions and misuse. This scenario clearly delineates two essential sources of uncertainty: the potential for uncontrollable events and the unpredictability of human behavior. In this contribution, we focus on behavioral uncertainty.

In the Section III, we discuss related works in the context of the profiling problem. Profiling is, in essence, the problem of converting data into a model for individuals with the goal of predicting the future behavior of these individuals. If no data is available, no predictions are possible, because human behavior is highly unpredictable. Several standard methods for predicting human behavior are discussed in this section. The model requires definition of two components. First, an identity management system that relates each human activity to an individual must be present. Second, a set of policies must be clearly defined.

Section IV describes compliance analysis from a theoretical standpoint. An internal attacker is difficult to identify if he/she commits no major offenses. Inside offenders can usually be sought only with the help of indicators or behavior patterns. Similar problems arise in criminology. To analyze certain behavior, a profile of the perpetrator may be created. Compliance profiling was designed to support this approach. An independent model is introduced for this purpose, based on case-based reasoning (CBR), in combination with directed acyclic graphs (DAG). This combination allows negative undesirable behavior patterns (data manipulation) to be defined. In a second step, the real behavior of users (e-mail, database access, copy processes, and deletion processes) may be compared with predefined undesirable behavior patterns. When a certain number of predefined patterns (profiles) overlap, a particular user's behavior may be analyzed further (the user is placed on a watch list).

In Section V, we provide conclusions and described some aspects of future investigations.

II. The changing face of threats

This section details the latest challenges faced by enterprises regarding regulatory, contractual obligations, and compliance requirements. These compliance requirements are reactions to new threats that originate specifically within enterprises [2]. Addressing the insider threat requires that information security professionals consider the most complex aspects of the people component of security: human nature and personality. Different facets of human nature come into play depending on whether user actions are taken with or without intent.

A high percentage of any user population will be honest, trustworthy, upstanding individuals who will act with the interests and success of the organization at heart. Human nature is inquisitive, and there will always be users who fall into the category of *explorers*— those who examine networks and servers to see what they can find. Above explorers are the internal intruders, who are the *ones to watch*. The challenge is to identify the *ones to watch*. A limited quantity of research has been conducted to examine the psychological profiles of problematic individuals and their personalities, motives, and the circumstances that contribute to an act of offense.

For example, consider the scenario in which an internal employee has conveniently manipulated (financial) data relevant to the enterprise. An enterprise's perimeter defense systems are unable to respond to these types of threats, because current perimeter defense techniques involve firewall systems, proxy servers, virus defense systems, web servers, etc. A characteristic of these safeguards is the real-time capability of the systems. This topic is discussed by many authors in the literature, e.g. [3]. Most perimeter defenses can be traced back to a formal theoretical foundation, the first-order model theory. This model theory can be an adequate and precise instrument for managing outside threats.

Insider threats, on the other hand, arise from inside intruders who have criminal intentions that do not originate outside the perimeter of the organization. Internal threats are discussed in [4].

The most common information security management system (ISMS) is the ISO/IEC 27001:2005 [5], with more than 5,693 certified enterprises across 79 countries, the top 10 of which countries are Japan, India, UK, Taiwan, China, Germany, Korea, USA, Czech Republic, and Hungary.[1] The ISO standard is particularly relevant to insider threat management, because it is concerned more with process, policy, and security education than technical controls. In Annex A 15.1–15.3 of ISO 27001, compliance is required, although no method for monitoring compliance is prescribed in the ISO 27001 standard itself [5]. In general, it can be said that new threats from inside intruders have, until now, received little consideration [6]. An article by Theoharidou et al. [6] argued that the code of practice, ISO 17799:2005 [7], as well as the information security requirements of the ISO 27001:2005 (formerly BS 7799-2), did not attend to internal threats. Hence, it was suggested in [6] that methods from criminology be used to address internal threats.

Inside intruders have been discussed in the literature, primarily in the fields of computers and networks, see, for example, [8], [9], [10], [11]. This type of intruder is not the focus of this contribution. As the corporate perimeter has become blurred by the outsourcing of services and the introduction of third-party network connections, the definition of the insider has changed. Historically, insiders were simply full- or part-time employees. Today, the term *insider* refers to anyone with in-depth knowledge of internal systems, organizational structure, processes and procedures, or with trusted access to systems, networks, and the information they contain.

In the meantime, a number of legal regulations have been remitted. The primary regulation pertaining to this issue is the Sarbanes-Oxley[2] Act in the USA [12], which requires suitable internal controlling systems for enterprises listed on the New York Stock Exchange. In addition to an ISMS, internal controlling systems (ICS) can be derived [13] from the CobiT or COSO models [14],[15]. The effectiveness of internal controlling systems must be described in reports written periodically (each fiscal year) [16]. This requirement also applies to cases of outsourcing, as Shue has argued [17].

Internal controlling system reports rely on data logging, which proves the integrity of administrators and users. The reports also include logged data that was not collected for perimeter defense [18], and the evaluation of the two types of data log is quite different. A real-time capability, which is required, e.g., for perimeter defense, is not suitable for monitoring insider threats. Historical data must be accessible to carry out compliance analysis. Compliance analysis may be carried out using a SAS 70 Type II assessment [19] by means of a W7 questionnaire. The so-called W7 questionnaire, which forms the basis of a compliance analysis, shows a direct relationship between possible internal intruder behavior (insider threats) and unauthorized actions initiated by users who fall into the category of *explorers*. A given set of questions (W7) are arranged according to the demands of the inquiry, e.g., computer forensics are described by Farmer and Venema in their book [20]. In the case of the SOX criteria, concrete questions can be formulated and designated as the seven Ws (W7 questionnaire). In particular, subjects are analyzed by posing the following questions (in order of importance)

W1 → **w**ho has
W2 → **w**hat
W3 → on **w**hich object
W4 → **w**here
W5 → from **w**here
W6 → **w**hen
W7 → to **w**hich object

to those subjects that have been forwarded specific documents or information, or who have had reading and/or writing access

[1] http://www.iso27001certificates.com (last accessed June 2009)

[2] Sections 302 and 404 are the most applicable.

to information and documents, in accordance with his/her role. The questions can also be posed in reverse order.

In this context, an effective centralized Identity Management System, together with specific policies, becomes important. If not every employee receives centrally-administered and checkable access privileges, an alleged abuse of data cannot be linked unambiguously and retrospectively to one person. Additionally, sufficient detailed policies must be available. Only if these requirements are fulfilled can internal audits be carried out using the W7 questionnaire. The W7 questionnaire transforms the role of internal auditor from a checking role to a monitoring role; in other words, from the role of a watchdog to the role of a bloodhound. With this type of active approach, the obligations of auditors have shifted, as was argued by Sarup in his paper [21], and discussed by von Solms [22].

III. Related works

In general, profiling is widely understood and used in the context of data mining applications. However, the crux of the profiling problem is the need to identify the elements of an undesirable profile relative to a valid profile. Normally, profiling consists of using historical transaction data from individuals to construct a model for each individual's behavior, so that their future behavior may be predicted. The field of user profiling, in the context of information security, is not uniform. Different approaches are applied in different fields, but only those conditions concerning information security may be used. The following discussion provides an overview of the various approaches used in different areas.

Simple profiling techniques, such as histograms, do not generalize well from sparse transaction data. In work by Cadez et al., a special flexible probabilistic mixture model for transactions was proposed [23]. This model fit easily into the mixture model and inferred a probabilistic profile for each individual. A Bayesian Network was used by Sebastiani et al. to build profiles of individuals [24].

In the present paper, we discuss the use of a directed acyclic graph and the CBR method to classify behavior. Typically, a node in a directed acyclic graph represents a stochastic variable, and the directed arcs represent conditional dependencies between these variables. Schuurmans et al. discussed questions relating to the content, scope, and application life cycle of a user profile [25]. They also outlined the issues necessary for forensics (W7) to create a user profile.

Another interesting area in which profiling methods could be applied is the field of linguistics. Author verification using linguistic profiling has been developed by Halteren [26]. In this approach, potential authors, or the selection of one author from a set of known authors (authorship recognition) is determined using a score calculation. This calculation is also used in forensic linguistics, where there is a need to determine whether a suspect did or did not write a specific, potentially incriminating text. Of interest in Halteren's article is the use of a false rejection rate and a false acceptance rate for assessing the quality of recognition. This method uses the fraction of similar choices made by two authors to provide a comparison to previous works. In our approach, we use a similar technique for measuring uncertainty, the Hamming distance. The trade-offs, between information gathering and overhead, that are implemented in a profiling process have been discussed by Moseley et al. [27]. During the aggressive optimization of profiling techniques, the necessity for detailed information and the costs of gathering profiles can outweigh the benefits reaped from profiling. A cost/benefit analysis of profiling has been examined with respect to the various compliance requirements and laws [28]. In the analysis, a specific model was designed by Kerrigan and Law [28] and Lau et al. [29]. Squicciarini et al. discussed policy compliance checking among federated service providers [30]. To minimize these costs, it is recommended that a shadow profiling system be used in a probe-based application monitoring scheme.

A quite different method was developed by Spitzner, in which honeypots were constructed to catch insider threats [8]. However, one of the disadvantages of honeypots is that they have a limited view; they see only what interacts with them. Simply deploying honeypots on one's internal network, then, is unlikely to uncover advanced insiders.

A precise model for compliance analysis techniques is presented in the following section. Using this precise model, an automatic monitoring system observes the behavior of employees with respect to laws and compliance requirements. In particular, the actual behavior of employees is compared to certain undesirable behavior patterns.

IV. Compliance profiling– an active compliance monitoring system

In this section, we take a closer look at compliance analysis and a specific model based on the methodologies of data mining [31]. A method for knowledge discovery within a data set is also presented. The hypothesis is that compliance analysis can find success through application of methods from machine-based learning. The learning system is defined by a learning strategy, as well as a representation of knowledge, a definition of the environment, and a range of use. Mathematical methods, such as case-based reasoning (CBR) and graph theory (DAG), can answer compliance analysis queries.

CBR is a highly integrative process. The manner in which similar cases are represented within CBR depends on the reasoning behind its employment. CBR uses a cycle of four steps (selecting, re-using, checking, and assigning). The cases that compose a CBR system are not simply a list of indicators; The cases serve specific purposes. For instance, in compliance analysis, a CBR system is able to discover deviations (misuse) from regulatory and compliance requirements by assisting with the design of the W7 questionnaire. The process by which CBR designs the W7 questionnaire consists of the following elements:

1) problem and/or situation description; *(Retrieve)* from a similar case;
2) *(Reuse)* results from similar cases to solve the new case;
3) Check the proposed solution: *(Revise)*;

4) Include the new solution in the existing knowledge base: *(Retain)*.

Figure 1 illustrates the concept of a CBR system. The underlying assumption is that similar problems have similar solutions. Case-based reasoning methods are based on the retrieval of past experiences to solve a new problem. Figure 1 shows the steps of a CBR system. Given a new case (P3) and a case base, the first step is the retrieval of a subset of cases (e.g., P, P1, P2 from the problem space) that are similar in some respects to the new case, P3. This subset of cases is used to formulate a new solution for P3, by adapting the solutions P', P1', P2' in the solution space, to the new case, P3. The suggested solution is revised to provide a final solution P3', and the new case and solution are retained for further use.

The key components of the CBR method are the assessment of the similarity between cases in order to retrieve appropriate precedents, and the adaptation of old solutions to new cases. Our analysis adapts the CBR system idea that similar human misuses arise from similar human behaviors. The four elements listed above show the framework for each case representation.

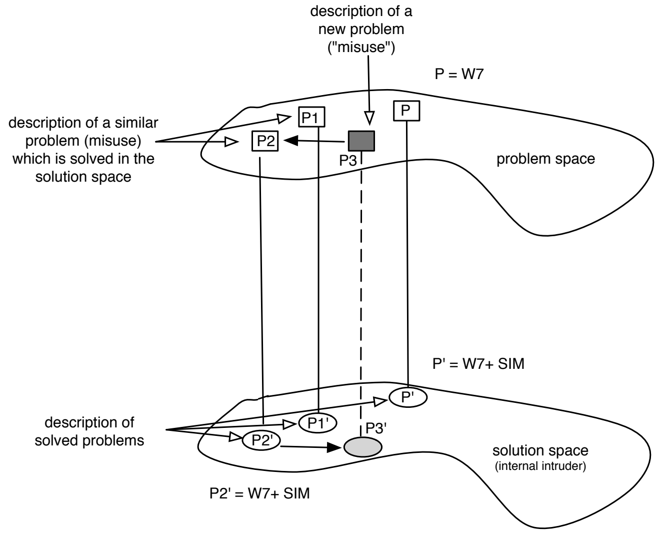
description of a new problem ("misuse")
P = W7
description of a similar problem (misuse) which is solved in the solution space
P1
P
P2
P3
problem space
P' = W7+ SIM
description of solved problems
P'
P1'
P3'
P2'
solution space (internal intruder)
P2' = W7+ SIM

Fig. 1: Problem and solution space of a CBR system.

First, a series of undesirable cases is defined in connection with identity management and policies. For this purpose, a formal representation of the cases P = {W1, W2, W3, W4, W5, W6, W7} is described. We interpret P as a path through a set of data (logging data). If one element of the descriptors, e.g., W1 (a user) is fixed, then every case P = W1(x) corresponds to an n-tuple within a specific time period, with $x = (x_1, \ldots x_n)$ (x denote an asset like a phone, eMail, etc., e.g. Fig. 2). These are the synthetic undesirable cases, and they define a profile in the problem space (P) of an undesirable human behavior (see Fig.1).

To carry out compliance analysis on a real data set,[3] a similarity function is introduced. The undesirable cases, defined in advance, can appear in modified or similar form in the data set (logging database). The data set is sampled from the user activities (see Fig. 2). Each subject and IT-object is represented by $y_1, \ldots y_n$. Taking $y = (y_1, \ldots y_n)$, e.g., real W1(y) cases from this data set, the similarity function (sim) can be applied to check W1(y) for similarities with the undesirable cases W1(x),

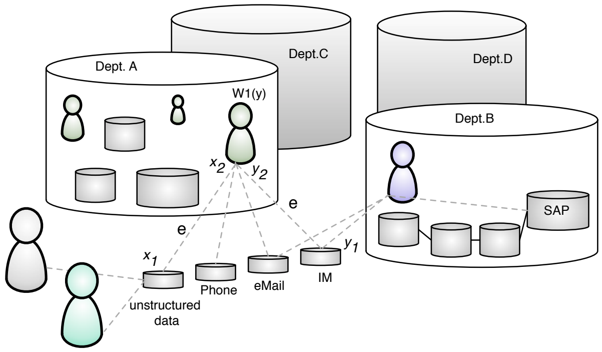

Fig. 2: Overview of an enterprise landscape with users activities.

$$sim(x, y) = function(sim_1(x_1, y_1), \ldots sim_n(x_n, y_n)). \quad (1)$$

The Hamming similarity function, which is suitable for indicating bivalency, defines the Hamming distance measures ($dist_H$) between the two cases, e.g., W1(x, y):

$$dist_H(x, y) = \sum_{i=1}^{m} |x_i - y_i|. \quad (2)$$

A correspondence between x and y constitutes a violation, if the Hamming distance, $dist_H(x, y) \approx 0$ (see Fig. 4). Cases of compliance (in which the security policy has been followed) yield a Hamming distance, between x and y in Eq. (3), of $dist_H(x, y) \approx 1$, and

$$sim_H = 1 - \frac{dist_H(x, y)}{m}. \quad (3)$$

Intermediate values may be evaluated by defining stimulus thresholds.

Analysis of an entire log database requires introduction of a directed graph G, in which the n-tuples (V, E) are defined to indicate a set of vertices (V) and a collection of edges (E). All $x \in V$ and $y \in V$ are represented as dots, and $e = (x_1 x_2) \in E$ and $e = (y_1 y_2) \in E$ are represented as junctions (see Fig. 3, Fig. 4). We interpret $e \in E$ as a communication between the vertices (x_1, x_2) and (y_1, y_2) (see Fig. 2).

A W7 questionnaire is formulated based on the elements $x_1, \ldots x_n$, which provide nodes within the set V. The edges show the relations between the nodes for a case W1(y_i). A case $x_1, \ldots x_n \in V$ in a data set under investigation is checked by modeling or compliance analysis by comparing the $sim_H(x, y)$ function and W7 path for $P = (V, E)$ against the W7 path for $P' = (V', E')$.

[3] e.g., a logging database that has saved suitable information about the behavior of subjects.

A path is defined in this model using $V = x_0, x_1, \ldots x_k$ and $E = x_0x_1, x_1x_2, \ldots x_{k-1}, x_k$, where x_i are disjoint pairs. The edges x_0 and x_k are, respectively, the starting edge and ending edge of the path P, and are joined by P. The edges $x_1, \ldots x_{k-1}$ are the internal edges of the path P. The number of edges along a path is its length. The length k of a path is designated P^k. A path P from x_0 can pass through to x_k several times. For $0 \leq i \leq j \leq k$, we define,

$$\begin{aligned} def: \quad Px_i &= x_0 \ldots x_i; \quad P'y_i = y_0 \ldots y_i; \\ def: \quad x_iP &= x_i \ldots x_k; \quad y_iP' = y_i \ldots y_k; \\ def: \quad x_iPx_j &= x_i \ldots x_j; \quad y_iPy_j = y_i \ldots y_j. \end{aligned} \tag{4}$$

We can also define suitable partial paths. The paths that contain compliances are determined by the policy and, finally, by the procedure. The procedure enforces the policy. With this model, an automatism can be designed that identifies application manipulations or offenses against policies within a data set (log data).

Figure 3 shows the graph of an undesirable case within a data set, the profile of an abuse (violation of rights and policies). The profile permits similar undesirable cases to be learned by means of a CBR cycle. This classification system permits characterization of certain undesirable behaviors among employees in an enterprise. If a real path (P') in a data set is found to be similar to an undesirable synthetic path (P), with a Hamming similarity that exceeds a predefined threshold, this path may be flagged for investigation into potential application manipulations. The Hamming similarity function may take into account fuzzy behavior in the analysis of the edges. Each edge depicts one part of the W7 questionnaire.

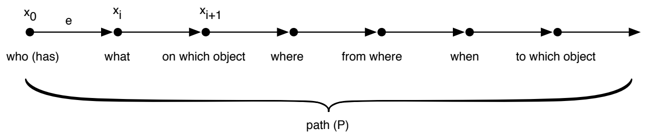

Fig. 3: Synthetic undesirable case of the path (P).

Figure 4 shows the undesirable profile from Fig. 3, supplemented with the automated evaluation of the subject's behavior. The behavior of a subject is, therefore, linked to a possible violation of a law. This comparison is defined as compliance profiling, and enables an early response to a violation on the basis of indication and appearance. With compliance profiling, an active controlling system, in terms of SOX, is produced. CBR works under the assumption that similar situations have similar explanations. We refer to the example in Section 1 (Introduction, page 2), in which a change in the accounts of a listed enterprise causes its current stock market value to increase. This type of compliance violation can be defined by comparison to a synthetic path (P) (see Fig. 3). Real cases will be very similar to synthetic paths, but not necessarily identical. The Hamming distance between P and P' is used to decide if the path P' is of type P, such that a subset of the solution space may be applicable (see Fig. 1).

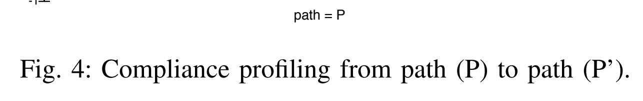

Fig. 4: Compliance profiling from path (P) to path (P').

V. Conclusions and further investigations

Solms pointed out in his paper that the fourth wave of information security focuses on Information Security Governance [1]. Following the structure articulated by Solms, internal monitoring strategies must clearly differ from those associated with perimeter defense monitoring. Perimeter defense strategies search for vulnerabilities within operating systems, firewall systems, browsers, and web servers. Possible malicious codes are identified by pattern and signature recognition. If such activities are present on a formal reduced basis, then first order logic is sufficient for identifying the threats. For internal monitoring problems (compliance analysis), a different method is required.

This article has shown, theoretically, that compliance analysis, with the help of the W7 questionnaire and case-based reasoning using graph theory (compliance profiling), is a suitable approach for assessing compliance issues. A proof of concept is expected to follow the development of this method, pending the availability of appropriate data sets.

In summary, we identified specific theoretical paths, using a sufficient number of real cases, to build the problem space. Four steps were required to build the knowledge base: select, re-use, clarify, and assign [32]. The knowledge base can be used later to identify compliance violations in the solution space (see Fig. 1).

Auditing the trail of events (paths) across all user activities provides unique insight into how users interact with sensitive information. Not only is this a powerful *pre*-forensics resource, but regular reviews of event types permit the ongoing tuning of rules for addressing new threats or strengthening and improving controls.

The next step focuses on the implementation of this method in a Security Event Management (SEM) System that collects the set of user data for a proof of concept analysis. An SEM framework using collection, analysis, integration, event-correlation, and scenario-analysis techniques processes the raw data gathered from the hybrid network [33].

References

[1] S. H. von Solms, "Information Security - The Fourth Wave," *Computers & Security*, vol. 25, no. 3, pp. 165–168, 2006.

[2] T. Robinson, "Data security in the age of compliance," *netWorker*, vol. 9, no. 3, pp. 24–30, 2005.

[3] W. Yan, E. S. H. Hou, and N. Ansari, "Frame-based attack representation and real-time first order logic automatic reasoning," in *ITRE*, pp. 225–229, 2005.

[4] L. Boral, M. Disla, S. Patil, J. Williams, and J. S. Park, “Countering insider threats in personal devices,” in *ISI*, p. 365, 2007.

[5] SC27, *ISO/IEC 27001:2005 – Information security management systems - Requirements*, vol. Information technology of *Security techniques*. Beuth-Verlag, Berlin, 10 2005.

[6] M. Theoharidou, S. Kokolakis, M. Karyda, and E. A. Kiountouzis, “The insider threat to information systems and the effectiveness of iso17799,” *Computers & Security*, vol. 24, no. 6, pp. 472–484, 2005.

[7] SC27, *ISO/IEC 17799:2005– Code of practice for information security management*, vol. Information technology of *Security Techniques*. Beuth-Verlag, Berlin, 2nd. ed., 06 2005.

[8] L. Spitzner, “Honeypots: Catching the insider threat,” in *ACSAC '03: Proceedings of the 19th Annual Computer Security Applications Conference*, (Washington, DC, USA), p. 170, IEEE Computer Society, 2003.

[9] Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, *Intruders pattern identification*, IEEE Computer Society, 2009.

[10] S. T. Teoh, K.-L. Ma, S. F. Wu, and T. Jankun-Kelly, “Detecting flaws and intruders with visual data analysis,” *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 27–35, 2004.

[11] S. Jang, H. Kim, ed., *An intruder tracing system based on a shadowing mechanism*, Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on, IEEE Computer Society, 2002.

[12] Sarbanes and Oxley, “Sarbanes-Oxley Act, U.S. Congress,” 2002.

[13] J. S. Broderick, “SMS, security standards and security regulations,” *Information Security Technical Report*, vol. Volume 11,, pp. Pages 26–31, March 2006.

[14] I. G. Institute, *CobiT, Control Objective in Information and related Technology*. ISBN 1-933284-37-4.: ITGI, 4th. ed., 2006.

[15] N. Li, J. C. Mitchell, and W. H. Winsborough, “Beyond proof-of-compliance: security analysis in trust management,” *J. ACM*, vol. 52, no. 3, pp. 474–514, 2005.

[16] G. Hardy, “Using IT governance and COBIT to deliver value with IT and respond to legal, regulatory and compliance challenges,” *Information Security Technical Report*, vol. Volume 11,, pp. Pages 55–61, March 2006.

[17] L. Shue, “Sarbarnes Oxley and IT Outsourcing,” *Information Systems Control Journal, ISACA*, vol. 5, 2004.

[18] B. Karabacak and I. Sogukpinar, “A quantitative method for ISO 17799 gap analysis,” *Computers & Security*, vol. 25, no. 6, pp. 413–419, 2006.

[19] AICPA, “SAS 70: Typ I und typ II; Statement on Auditing Standards (SAS) No. 70, AU 324,,” 2002.

[20] D. Farmer and W. Venema, *Forensic Discovery*. Upper Saddle River, NJ: Addison-Wesley., 2005.

[21] D. Sarup, “Watchdog or bloodhound? the push and pull toward a new audit model.” Information Systems Audit and Control Association, 2004, vol. 1, pages 23-26.

[22] R. von Solms and S. H. von Solms, “Information Security Governance: A model based on the Direct-Control Cycle,” *Computers & Security*, vol. 25, no. 6, pp. 408–412, 2006.

[23] I. V. Cadez, P. Smyth, and H. Mannila, “Probabilistic modeling of transaction data with applications to profiling, visualization, and prediction,” in *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, (New York, NY, USA), pp. 37–46, ACM, 2001.

[24] P. Sebastiani, M. Ramoni, and A. Crea, “Profiling your customers using Bayesian networks,” *SIGKDD Explor. Newsl.*, vol. 1, no. 2, pp. 91–96, 2000.

[25] J. Schuurmans, B. de Ruyter, and H. van Vliet, “User profiling,” in *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, (New York, NY, USA), pp. 1739–1740, ACM, 2004.

[26] H. V. Halteren, “Author verification by linguistic profiling: An exploration of the parameter space,” *ACM Trans. Speech Lang. Process.*, vol. 4, no. 1, p. 1, 2007.

[27] T. Moseley, A. Shye, V. J. Reddi, D. Grunwald, and R. Peri, “Shadow profiling: Hiding instrumentation costs with parallelism,” in *CGO '07: Proceedings of the International Symposium on Code Generation and Optimization*, (Washington, DC, USA), pp. 198–208, IEEE Computer Society, 2007.

[28] S. Kerrigan and K. H. Law, “Logic-based regulation compliance-assistance,” in *ICAIL '03: Proceedings of the 9th international conference on Artificial intelligence and law*, (New York, NY, USA), pp. 126–135, ACM, 2003.

[29] G. T. Lau, S. Kerrigan, K. H. Law, and G. Wiederhold, “An e-government information architecture for regulation analysis and compliance assistance,” in *ICEC '04: Proceedings of the 6th international conference on Electronic commerce*, (New York, NY, USA), pp. 461–470, ACM, 2004.

[30] A. C. Squicciarini, M. C. Mont, A. Bhargav-Spantzel, and E. Bertino, “Automatic Compliance of Privacy Policies in Federated Digital Identity Management,” in *POLICY*, pp. 89–92, 2008.

[31] C. Beierle and G. Kern-Isberne, *Methoden wissensbasierter Systeme*. ISBN-10 3-8348-0010-4.: Vieweg Verlag, 3. ed., 2006.

[32] E. Armengol, S. O. nón, and E. Plaza, “The Explanatory Power of Symbolic Similarity in Case-Based Reasoning,” in *Artificial Intelligence Review* (P. Gervás and K. Gupta, eds.), vol. 24 of *ISSN 0269-2821*, pp. pp. 145–161(17), Springer-Verlag, Springer-Verlag, 2005.

[33] L. Liu and et. al., “A Security Event Management framework using Wavelet and Data-Mining Technique,” *Communications, Circuits and Systems Proceedings, 2006 International Conference on*, vol. 3, pp. 1566 – 1569, June 2006.