A framework for Internal Fraud Risk Reduction at IT Integrating Business Processes

Non Peer-reviewed author version

JANS, Mieke; LYBAERT, Nadine & VANHOOF, Koen (2009) A framework for Internal Fraud Risk Reduction at IT Integrating Business Processes. In: International Journal of Digital Accounting Research, 9. p. 1-29.

Handle: http://hdl.handle.net/1942/7890

A framework for Internal Fraud Risk Reduction at IT Integrating Business Processes: The IFR² Framework

Mieke Jans[a] Nadine Lybaert[a] Koen Vanhoof[a]

[a] Hasselt University, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium

{mieke.jans, nadine.lybaert, koen.vanhoof}@uhasselt.be

Abstract

Fraud is a million dollar business and it is increasing every year. Both internal and external fraud present a substantial cost to our economy worldwide. A review of the academic literature learns that the academic community only addresses external fraud and how to detect this type of fraud. Little or no effort to our knowledge has been put in investigating how to prevent ánd to detect ínternal fraud, what we call 'internal fraud risk reduction'. Taking together the urge for research in internal fraud and the lack of it in academic literature, research to reduce internal fraud risk forces itself on. Only after having a framework in which to implement empirical research, this topic can further be investigated. In this paper we present the IFR² framework, deduced from both the academic literature and from current business practices, where the core of this framework suggests to use a data mining approach.

Keywords: IFR² framework, internal fraud, risk reduction, data mining

I. Introduction

Internal fraud is a significant problem to the world economy of today. Organizations allocate lots of resources to internal control, a framework implemented in business practice to prevent internal fraud. These costs, together with the costs of internal fraud itself, represent a large economic cost for the business environment and did not go unnoticed. A US fraud standard (SAS 99) and an international counterpart (ISA 240) were created. Meanwhile, the CEO's of the International Audit Networks released a special report in November 2006. This report, issued by the six largest global audit networks, was released in the wake of corporate scandals. The authors of this report express their believe in mitigating fraud, as they name it "one of the six vital elements, necessary for capital market stability, efficiency and growth".

In academic literature however, there is almost no attention for this huge problem. Based on the absence of a methodological framework to mitigate internal fraud in the academic literature, the cost internal fraud nevertheless presents, and the clear interest the business environment shows, the research objective in this paper is to present a framework for internal fraud risk reduction.

For this purpose, two courses are followed, resulting in our framework for internal fraud risk reduction, the IFR² framework. In Section III we first have a look at what already exists in the business environment to prevent and detect internal fraud. Next, in Section IV, we turn to the methodology followed in the academic field. We start with an extended literature review on corporate fraud detection and prevention in different disciplines. We summarize this review in an overview table with the most important characteristics of each study, being the domain in which it is executed, whether it concerns internal or external fraud, whether it focusses on fraud detection or prevention and which technique is used. By looking at this overview table, we arrive at the conclusion that merely all research is conducted in the field of external fraud. Concerning internal fraud, there is a gap in the academic literature. Another observation is that the bulk of articles apply a data mining approach. In the overview table a last column is added about which kind of data mining task was performed. Because this data mining approach has proven its value in mitigating external fraud and is the methodology of existing fraud detection research, we

provide in a next Section V an introduction in data mining. What we find in business practice and what existing research in external fraud exposes is the foundation of our framework for internal fraud risk reduction, the IFR² framework, presented in Section VI. We start this paper however with a general section about fraud, handling both external and internal fraud.

II. Fraud

What is Fraud?

Fraud is deception. Whatever industry the fraud is situated in or whatever kind of fraud you visualize, deception is always the core of fraud. There are many definitions of fraud, depending on the point of view considering. According to *The American Heritage Dictionary, (Second College Edition)*, fraud is defined as "*a deception deliberately practiced in order to secure unfair or unlawful gain*".

In a nutshell, "Fraud always involves one or more persons who, with intent, act secretly to deprive another of something of value, for their own enrichment" (Davia et al. 2000). Also Wells (2005) stresses deception as the linchpin to fraud. The kind of fraud as subject matter of his book is *occupational fraud and abuse*. This is a delineation of fraud, which is also periodically investigated by the Association of Certified Fraud Examiners[1] (ACFE). In their *2006 Report to the Nation on Occupational Fraud and Abuse*, the ACFE defines occupational fraud and abuse as: "*The use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets*." (ACFE, 2006). This definition encompasses a wide variety of conduct by executives, employees, managers, and principals of organizations. Violations can range from asset misappropriation, fraudulent statements and corruption over pilferage and petty theft, false overtime, using company property for personal benefit to payroll and sick time abuses (Wells, 2005). Although this type of fraud encompasses

[1] The ACFE is the world's premier provider of anti-fraud training and education. Together with nearly 40,000 members, the ACFE is reducing business fraud world-wide and inspiring public confidence in the integrit and objectivity within the profession. (www.acfe.com)

many kinds of irregularities, mind that it does not cover all kind of frauds. Only internal corporate fraud is included. For example fraud against the government (non corporate fraud) or fraud perpetrated by customers (external fraud) are not included.

Classifying Fraud

The delineation of fraud to 'occupational fraud and abuse' is one way to categorize fraud. There are numerous other ways of classifying fraud. A classification that resembles however this first delineation, is the distinction Bologna and Lindquist (1995) make between *internal* versus *external* fraud. This classification, applied in the field of corporate fraud (fraud in an organizational setting), is based on whether the perpetrator is internal or external to the victim company. Frauds committed by vendors, suppliers or contractors are examples of external fraud, while an employee stealing from the company or a manager cooking the books are examples of internal fraud. What is seen as internal fraud, following this definition, is in fact occupational fraud and abuse, since one has to be internal to a company and abuse its occupation to commit internal fraud. We put internal fraud and occupational fraud and abuse as equivalents. A combination of internal and external fraud can also occur, for example when an employee collaborates with a supplier to deprive the company.

Bologna and Lindquist (1995) mention, in addition to other classifications, another way of classifying fraud: *transaction* versus *statement* fraud. The authors define statement fraud as *"the intentional misstatement of certain financial values to enhance the appearance of profitability and deceive shareholders or creditors."* Transaction fraud is intended to embezzle or steal organizational assets. Davia et al. (2000) distinguish two related types of fraud: *financial statement balance* fraud and *asset-theft* fraud. The authors state that the main difference between the former and the latter is that there is no theft of assets involved in financial statement balance fraud. Well known examples of this type of fraud are Enron and Worldcom. We see this classification (financial statement balance fraud vs. asset-theft fraud) as an equivalent of Bologna and Lindquist (1995)'s statement and transaction fraud.

Bologna and Lindquist (1995) give two more classifications of fraud - all classifying

corporate fraud. A first classification is fraud *for* versus *against* the company. The former contains frauds intended to benefit the organizational entity, while the latter encompasses frauds that intend to harm the entity. Examples of fraud for the company are price fixing, corporate tax evasion and violations of environmental laws. While these frauds are in the benefit of the company at first, in the end the personal enrichment stemming from these frauds are the real incentives. Frauds against the company are only intended to benefit the perpetrator, like embezzlement or theft of corporate assets. The authors draw attention to the fact that not all frauds fit conveniently into this schema, for example arson for profit, planned bankruptcy and fraudulent insurance claims.

A last distinction Bologna and Lindquist (1995) refer to is *management* versus *non-management* fraud, also a classification based on the perpetrator's characteristics.

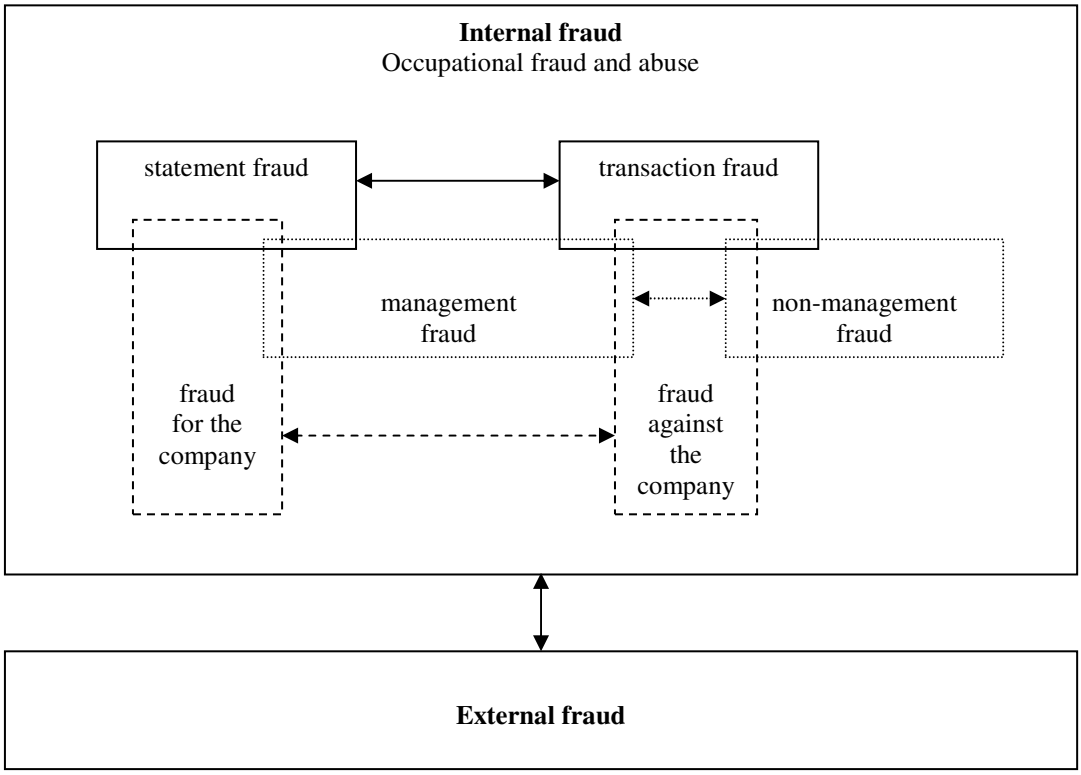

Figure 1. Fraud classification overview

These different classifications all present another dimension and can display some overlap. In Figure 1 we present an overview of how we see the different classifications and their relations to each other, hereby making some assumptions.

The most prominent classification is the internal versus external fraud, since all other classifications are situated within internal fraud. As already pointed out, we see occupational fraud and abuse as an equivalent of internal fraud. Figure 1 also shows that all classifications left, apply only to corporate fraud. This explains why all are embedded in internal fraud.

Within internal fraud, three different classifications occur. We start with a distinction between statement fraud and transaction fraud, respectively financial statement balance fraud and asset-theft fraud in terms of Davia et al. (2000). A second distinction is based upon the occupation level of the fraudulent employee: management versus non-management fraud. We assume that managers can commit both statement and transaction fraud, yet non-management is in our view restricted to transaction fraud only. The last classification we introduce in this overview is fraud for versus fraud against the company. Although fraud for the company does not necessarily need to be statement fraud (for example breaking environmental laws), an overlap is realistic. With the classification for versus against, we again make an assumption. Contrary to fraud against the company, we believe only managers are in an advantageous position to commit fraud for the company, hence the overlap with only management fraud. Whereas fraud against the company is believed to be committed both by managers and non-managers. A last assumption is made concerning the nature of statement fraud. We assume all statement fraud is committed to improve the company's appearance and never to harm the company. Therefor we assume statement fraud is always profiled as fraud for the company, never against the company.

Cost of Fraud: Some Numbers

Fraud is a million dollar business, as several research studies on this phenomenon report shocking numbers. Concerning internal fraud, two elaborate surveys, one conducted in the United States by the ACFE, (ACFE 2006)[2] and one worldwide by PricewaterhouseCoopers (PwC 2007)[3], yield the following information about corporate

[2] 1.134 cases of occupational fraud, reported by a Certified Fraud Examiner between January 2004 and January 2006, are subject of this report.

[3] 5.428 companies (all PwC clients) across 40 countries around the world are subjected to the Global Economic Crime Survey 2007, a biennial survey conducted by PwC.

fraud:

Forthy-three percent of companies surveyed worldwide (PwC-survey) has fallen victim to economic crime in the years 2006 and 2007. The average financial damage to companies subjected to the PwC survey, was US$ 2.42 million per company over the past two years. No industry seems to be safe and bigger companies seem to be more vulnerable to fraud than smaller ones. Participants of the ACFE study estimate a loss of five percent of a company's annual revenues to fraud. Applied to the 2006 United States Gross Domestic Product of US$ 13,246.6 billion, this would translate to approximately US$ 662 billion in fraud losses for the United States only.

The numbers mentioned above all concern forms of internal fraud. There are however also large costs from external fraud. Four important domains afflicted by fraud are regularly discussed: telecommunications, automobile insurance, health care and credit cards. On these domains, we found the following numbers:

Globally, telecommunications fraud is estimated at about US$ 55 billion. (Abidogum 2005) For the second domain, the automobile insurance fraud problem, Brockett et al. (1998) cite an estimation of the National Insurance Crime Bureau (NICB) that the annual cost in the United States is US$ 20 billion. At the website of the NICB we read: "Insurance industry studies indicate 10 percent or more of property/casualty insurance claims are fraudulent." (NICB 2008) Concerning health care insurance claims fraud, the United States National Health Care Anti-Fraud Association (NHCAA) estimates conservatively that of the nation's annual health care outlay, at least 3% is lost to outright fraud. This is $68 billion. Other estimates by government and law enforcement agencies place the loss as high as 10% of their annual expenditure. (NHCAA 2008) Concerning the fourth domain, credit card fraud, Bolton and Hand (2002) cite estimates of US$ 10 billion losses worldwide for Visa/Mastercard only.

Prevention versus Detection

A lot has been written about how to detect fraud. However many authors, like Bologna and Lindquist (1995), state that prevention should take precedence over detection. The

authors mean by fraud prevention creating a work environment that values honesty. This includes hiring honest people, paying them competitively, treating them fairly, and providing a safe and secure workplace.

In the *Accountant's Guide to Fraud Detection and Control*, Davia et al. (2000) state that it is management's responsibility to allocate resources and emphasis to fraud-specific internal controls and to proactive fraud-specific examinations. These approaches are examples of prevention on one hand and detection on the other. The authors point out that it is a mistake to think in terms of one versus the other. Strong internal controls as fraud prevention are very important, but they are best reinforced by following fraud-specific examinations.

In the above mentioned studies of PwC and the ACFE, one speaks only about detection. The studies investigate by means of surveys which are the most occurring means or methods that lead to fraud detection, or are believed to do so by the CFO's. The following are the findings of both studies.

About the way fraud is detected, both studies of PwC and the ACFE stress the importance of tips and chance. According to the ACFE report, an anonymous fraud hotline anticipates a lot of fraud damage. In the cases reviewed, organizations that had such hotlines, suffered a median loss of US$ 100.000, whereas organizations without hotlines had a median loss of US$ 200.000. At the PwC study, no less than 41% of the fraud cases was detected by means of tip-offs or by accident. Internal audit and internal control systems can have a measurable impact on detecting fraud after chance related means. The more control measures a company puts in place, the more incidents of fraud will be uncovered.

Another recent study, performed by Ernst&Young, mentions preventing and detecting fraud. The global survey by Ernst&Young in 2006 revealed similar insights on fraud prevention factors. Respondents identify internal controls as the key factor to prevent and detect fraud. (Ernst&Young, 2006)

Beware that all above mentioned suggestions concerning detection and prevention of fraud, concern internal fraud detection/prevention and further, are the results of non-academic research.

The framework presented in this paper will aim at the combination of fraud detection and prevention, which will be referred to as "fraud risk reduction". This decision is corresponding with the ideas of Davia et al. (2000) and Bologna and Lindquist (1995), that fraud prevention and fraud detection should complement each other. Further, the scope of our research is transaction fraud, a particular form of internal fraud (see Figure 1).

III. Mitigating Internal Fraud in Practice: The Value of Internal Control

The studies of PwC and the ACFE mentioned before, reveal some information concerning the detection of internal fraud. The number one detection tool is chance related, like tip-offs and detection by accident. This kind of tool is not easily influenced by corporate governance, because it is linked with corporate culture, and not with controls. The second best detection tool seems to be internal control and is a better candidate for mitigating internal fraud, since it lends itself better to govern. Internal control is currently the most prevalent mean companies use to mitigate fraud. In this section some history and a brief overview of what internal control encompasses is given.

The Committee of Sponsoring Organizations of the Treadway Commission (COSO)[4] was formed to commission the Treadway Commission to perform its task (study the causes of fraudulent reporting and make recommendations to reduce its incidence). In response to this recommendation, COSO developed an internal control framework, issued in 1992 and entitled *Internal Control - Integrated Framework*. According to the COSO framework, internal control is defined as:

[4] The sponsoring accounting organizations include the American Institute of Certified Public Accountants (AICPA), the American Accounting Association (AAA), the Financial Executives Institute (FEI), the

a process, effected by the entity's board of directors, management, and other personnel, designed to provide reasonable assurance regarding the achievement of objectives in the following categories:

- *Effectiveness and efficiency of operations*
- *Reliability of financial reporting*
- *Compliance with applicable laws and regulations*

If we look at the definition, it is clear why internal control is important as a protection against fraud. The achievement of the first category is to encounter transaction fraud, the second to encounter statement fraud and the third category achievement is to protect the organization against fraud for the company. Following this broad definition, internal control can both prevent and detect fraud. And although this definition is stemming from the foundation of the National Commission on Fraudulent Financial Reporting, also other classes of fraud than fraudulent financial reporting can be encountered. However, the definition is clear about its *reasonable* - not absolute - assurance regarding the objectives. We can conclude that internal control is a means to protect an organization against internal fraud, but given the raising prevalence of fraud it is still not sufficient as a stand-alone tool. Also the numbers provided by the PwC and ACFE surveys reveal that internal control comes off worse than chance means as a detection tool. However, these studies also emphasize the extra value of well functioning internal control systems.

The internal control framework of COSO is the broadest existing framework on this topic. Some industries have taken this framework and customized it to their specific needs, for instance the banking industry. In this environment, Basel II is created, with its own internal control section. It is however based on COSO and hence is a variant of this framework. It is beyond the scope and the goal of this paper to address all existing internal control frameworks. We believe that by addressing the settings of COSO, the general business practice in terms of internal control are covered.

Institute of Internal Auditors (IIA), and the Institute of Management Accountants (IMA).

IV. Fraud Detection/Prevention Literature Review

In this section, an overview of the academic literature concerning fraud prevention and detection is given. Although the subject of fraud prevention is taken into account, almost all articles found address the problem of fraud detection. To gain a clear view of the current situation of research, Table 1 is created. This will provide us with some insights of the implicitly followed methodology in current literature. The table provides us with the author(s) in alphabetical order, the application domain, whether it concerns internal or external fraud, whether the objective is fraud detection or prevention, and which technique is used. The information about the last column (Task) will be discussed later and is of no importance yet.

Concerning the techniques used, an intensively explored method are neural networks. The studies of Davey et al. (1996) (telecommunications fraud), Dorronsoro et al. (1997) (credit card fraud), and Fanning and Cogger (1998), Green and Choi (1997) and Kirkos et al. (2007) (financial statement fraud) all use neural network technology for detecting fraud in different contexts. Lin et al. (2003) apply a fuzzy neural net, also in the domain of fraudulent financial reporting. Both Brause et al. (1999) and Estévez et al. (2006) use a combination of neural nets and rules. The latter use fuzzy rules, where the former use traditional association rules. Also He et al. (1997) apply neural networks: a multi-layer perceptron network in the supervised component of their study and Kohonen's self-organizing maps for the unsupervised part. (the terms *supervised* and *unsupervised* will be explained in a following pragraph). Like He et al. (1997) apply in their unsupervised part, Brockett et al. (1998) apply Kohonen's self-organizing feature maps (a form of neural network technology) to uncover phony claims in the domain of automobile insurance. This is also what Zaslavsky and Strizhak (2006) suggest later, in 2006, in a methodological paper to detect credit card fraud. Quah and Sriganesh (2007) follow this suggestion in an empirical paper on understanding spending patterns to decipher potential fraud cases. A Bayesian learning neural network is implemented for credit card fraud detection by Maes et al. (2002) (aside to an artificial neural network), for uncollectible telecommunications accounts (which is not always fraud) by Ezawa and Norton (1996), for financial statement fraud by Kirkos et al. (2007) and for automobile insurance fraud detection by Viaene et al. (2005) and Viaene et al. (2002).

In Viaene et al. (2005)'s field of automobile insurance fraud, Bermúdez et al. (2007) use an asymmetric or skewed logit link to fit a fraud database from the Spanish insurance market. Afterwards they develop Bayesian analysis of this model. In a related field Major and Riedinger (2002) presented a tool for the detection of medical insurance fraud. They propose a hybrid knowledge/statistical-based system, where expert knowledge is integrated with statistical power. Another example of combining different techniques can be found in Fawcett and Provost (1997). A series of data mining techniques for the purpose of detecting cellular clone fraud is hereby used. Specifically, a rule-learning program to uncover indicators of fraudulent behavior from a large database of customer transactions is implemented. From the generated fraud rules, a selection has been made to apply in the form of monitors. This set of monitors profiles legitimate customer behavior and indicate anomalies. The outputs of the monitors, together with labels on an account's previous daily behavior, are used as training data for a simple Linear Threshold Unit (LTU). The LTU learns to combine evidence to generate high-confidence alarms. The method described above is an example of a supervised hybrid as supervised learning techniques are combined to improve results. In another work of Fawcett and Provost (1999), Activity Monitoring is introduced as a separate problem class within data mining with a unique framework. Fawcett and Provost (1999) demonstrate how to use this framework among other things for cellular phone fraud detection.

Another framework presented, for the detection of healthcare fraud, is a process-mining framework by Yang and Hwang (2006). The framework is based on the concept of *clinical pathways* where structure patterns are discovered and further analyzed.

The fuzzy expert systems are also experienced with in a couple of studies. So there are Derrig and Ostaszewski (1995), Deshmukh and Talluru (1998), Pahtak et al. (2003), and Sanchez et al. (2008). The latter extract a set of fuzzy association rules from a data set containing genuine and fraudulent credit card transactions. These rules are compared with the criteria which risk analysts apply in their fraud analysis process. The research is therefor difficult to categorize as 'detection', 'prevention' or both. We adopt the authors' own statement of contribution in both fraud detection and prevention. Derrig and

Ostaszewski (1995) use fuzzy clustering and therefor apply a data mining technique performing a descriptive task, where the other techniques (but Sanchez et al. (2008)) perform a predictive task.

Stolfo et al. (2000) delivered some interesting work on intrusion detection. They provided a framework, MADAM ID, for Mining Audit Data for Automated Models for Intrusion Detection. Although intrusion detection is associated with fraud detection, this is a research area on its own and we do not extend our scope to this field. Next to MADAM ID, Stolfo et al. (2000) discuss the results of the JAM project. JAM stands for Java Agents for Meta-Learning. JAM provides an integrated meta-learning system for fraud detection that combines the collective knowledge acquired by individual local agents. In this particular case, individual knowledge of banks concerning credit card fraud is combined. Also Phua et al. (2004) apply a meta-learning approach, in order to detect fraud and not only intrusion. The authors base their concept on the science fiction novel *Minority Report* and compare the base classifiers with the novel's 'precogs'. The used classifiers are the naive Bayesian algorithm, C4.5 and backpropagation neural networks. Results from a publicly available automobile insurance fraud detection data set demonstrate that the stacking-bagging performs better in terms of performance as well as in terms of cost savings.

Cahill et al. (2000) design a fraud signature, based on data of fraudulent calls, to detect telecommunications fraud. For scoring a call for fraud its probability under the account signature is compared to its probability under a fraud signature. The fraud signature is updated sequentially, enabling event-driven fraud detection.

Rule-learning and decision tree analysis is also applied by different researchers, e.g. Kirkos et al. (2007), Fan (2004), Viaene et al. (2002), Bonchi et al. (1999), and Rosset et al. (1999). Viaene et al. (2002) actually apply different techniques in their work, from logistic regression, k-nearest neighbor, decision trees and Bayesian neural network to support vector machine, naive Bayes and tree-augmented naive Bayes. Also in Viaene et al. (2007), logistic regression is applied.

Link analysis comprehends a different approach. It relates known fraudsters to other individuals, using record linkage and social network methods (Wasserman and Faust 1998). Cortes et al. (2002) find the solution to fraud detection in this field. The transactional data in the area of telecommunications fraud is represented by a graph where the nodes represent the transactors and the edges represent the interactions between pairs of transactors. Since nodes and edges appear and disappear from the graph through time, the considered graph is dynamic. Cortes et al. (2002) consider the subgraphs centered on all nodes to define communities of interest (COI). This method is inspired by the fact that fraudsters seldom work in isolation from each other.

To continue with link analysis, Kim and Kwon (2006) report on the Korean Insurance Fraud Recognition System that employs an unsupervised three-stage statistical and link analysis to identify presumably fraudulent claims. The government draws on this system to make decisions. The authors evaluate the system and offer recommendations for improvement.

Bolton and Hand (2001) are monitoring behavior over time by means of Peer Group Analysis. Peer Group Analysis detects individual objects that begin to behave in a way different from objects to which they had previously been similar. Another tool Bolton and Hand (2001) develop for behavioral fraud detection is Break Point Analysis. Unlike Peer Group Analysis, Break Point Analysis operates on the account level. A break point is an observation where anomalous behavior for a particular account is detected. Both the tools are applied on spending behavior in credit card accounts.

Also Murad and Pinkas (1999) focus on behavioral changes for the purpose of fraud detection and present three-level-profiling. As the Break Point Analysis from Bolton and Hand (2001), the three-level-profiling method operates at the account level and it points any significant deviation from an account's normal behavior as a potential fraud. In order to do this, 'normal' profiles are created (on three levels), based on data without fraudulent records. To test the method, the three-level-profiling is applied in the area of telecommunication fraud. In the same field, also Burge and Shawe-Taylor (2001) use behavior profiling for the purpose of fraud detection by using a recurrent neural network

for prototyping calling behavior. Two time spans are considered at constructing the profiles, leading to a current behavior profile (CBP) and a behavior profile history (BPH) of each account. In a next step the Hellinger distance is used to compare the two probability distributions and to give a suspicion score on the calls.

A brief paper of Cox et al. (1997) combines human pattern recognition skills with automated data algorithms. In their work, information is presented visually by domain-specific interfaces. The idea is that the human visual system is dynamic and can easily adapt to ever-changing techniques used by fraudsters. On the other hand have machines the advantage of far greater computational capacity, suited for routine repetitive tasks.

Two last studies we would like to mention is that of Tsung et al. (2007) and Brockett et al. (2002), Hoogs et al. (2007) and Juszczak et al. (2008). Tsung et al. (2007) apply manufacturing batch techniques to the field of fraud detection. They use the batch library method. Brockett et al. (2002) use a principal component analysis of RIDIT scores to classify claims for automobile bodily injury. Hoogs et al. (2007) present a genetic algorithm approach to detect financial statement fraud. They find that exceptional anomaly scores are valuable metrics for characterizing corporate financial behavior and that analyzing these scores over time represents an effective way of detecting potentially fraudulent behavior. Juszczak et al. (2008) at last apply many different classification techniques in a supervised two-class setting and a semi-supervised one-class setting in order to compare the performances of these techniques and settings.

Table 1: Fraud detection/prevention literature overview

Author	Application Domain	Internal/ External	Detection/ Prevention	Technique	Task
Bermúdez et al. (2007)	Automobile Insurance Fraud	External	Detection	Skewed Logit Link and Bayesian Analyses	Predicitve
Bolton and Hand (2001)	Credit Card Fraud	External	Detection	Peer Group Analysis and Break Point Analysis	Predictive
Bonchi et al. (1999)	Fiscal Fraud	External	Detection	Decision Tree	Predictive
Brause et al. (1999)	Credit Card Fraud	External	Detection	Rules and Neural Network	Predictive
Brockett et al. (1998)	Automobile Insurance Fraud	External	Detection	Kohonen's Self-Organizing Map	Predictive
Brockett et al. (2002)	Automobile Insurance Fraud	External	Detection	Principal Component Analysis	Predictive
Burge and Shawe-Taylor (2001) and Shawe- Taylor	Telecommunications Fraud	External	Detection	Unsupervised Neural Network	Predictive
Cahill et al. (2000)	Telecommunication Fraud	External	Detection	Profiling by means of signatures	Predictive
Cortes et al. (2002)	Telecommunications Fraud	External	Detection	Dynamic Graphs	Predictive
Cox et al. (1997)	Telecommunications Fraud	External	Detection	Visual Data Mining	Descriptive
Davey et al. (1996)	Telecommunications Fraud	External	Detection	Neural Network	Predictive
Derrig and Ostaszewski (1995) and Ostaszewski	Automobile Insurance Fraud	External	Detection	Fuzzy Set Theory	Descriptive
Deshmukh and Talluru (1998)	Financial Statement Fraud	Internal	Detection	Rule-based Fuzzy Reasoning System	Predictive
Dorronsoro et al. (1997)	Credit Card Fraud	External	Detection	Neural Network	Predictive
Estévez et al. (2006)	Telecommunications Fraud	External	Detection and Prevention	Fuzzy Rules and Neural Network	Predictive
Ezawa and Norton (1996)	Uncollectible Telecommunications Accounts	External	Detection	Bayesian Neural Network	Predictive
Fan (2004)	Credit Card Fraud	External	Detection	Decision Tree	Predictive
Fanning and Cogger (1998)	Financial Statement Fraud	Internal	Detection	Neural Network	Predictive
Fawcett and Provost (1997)	Telecommunications Fraud	External	Detection	Rules, Monitors and Linear Threshold Unit	Predictive
Fawcett and Provost (1999)	Telecommunications Fraud	External	Detection	Activity Monitoring	Predictive
Green and Choi (1997)	Financial Statement Fraud	Internal	Detection	Neural Networks	Predictive
He et al. (1997)	Health Care Insurance Fraud	External	Detection	Neural Network	Predictive

Continued on next page

Author	Application Domain	Internal/ External	Detection/ Prevention	Technique	Task
He et al. (1997)	Health Care Insurance Fraud	External	Detection	Kohonen's Self-Organizing Map	Descriptive
Hilas and Mastorocostas (2008)	Telecommunications Fraud	External	Detection	Neural Network and Clustering	Predictive
Hoogs et al. (2007)	Financial Statement Fraud	Internal	Detection	A Genetic Algorithm Approach	Predictive
Juszczak et al. (2007)	Credit Card Fraud	External	Detection	Many different classification techniques	Predictive
Kim and Kwon (2006)	Insurance Fraud	External	Detection	Insurance Fraud Recognition System (Korea)	Predictive
Kirkos et al. (2007)	Financial Statement Fraud	Internal	Detection	Decision Tree, Neural Network and Bayesian Belief Network	Predictive
Lin et al. (2003)	Financial Statement Fraud	Internal	Detection	Fuzzy Neural Network	Predictive
Maes et al. (2002)	Credit Card Fraud	External	Detection	Neural Network and Bayesian Belief Network	Predictive
Major and Riedinger (2002)	Health Care Insurance Fraud	External	Detection	Electronic Fraud Detection (EFD)	Predictive
Murad and Pinkas (1999)	Telecommunications Fraud	External	Detection	Three Level Profiling	Predictive
Pathak et al. (2003)	Insurance Fraud	External	Detection	Fuzzy logic based expert system	Predictive
Phua et al. (2004)	Automobile Insurance Fraud	External	Detection	Meta-classifiers	Predictive
Quah and Sriganesh (2007)	Credit Card Fraud	External	Detection	Self-Organizing Maps	Descriptive
Rosset et al. (1999)	Telecommunications Fraud	External	Detection	Rules	Predictive
Sánchez et al. (2008)	Credit Card Fraud	External	Detection and Prevention	Fuzzy Rules	Descriptive
Stolfo et al. (2000)	Credit Card Fraud and Intrusion	External	Detection	Meta-classifiers	Predictive
Tsung et al. (2007)	Telecommunications Fraud	External	Detection	Batch Library Method	Predictive
Viaene et al. (2005)	Automobile Insurance Fraud	External	Detection	Bayesian Neural Network	Predictive
Viaene et al. (2002)	Automobile Insurance Fraud	External	Detection	Logistic Regression, k-Nearest Neighbor, Decision Tree, Bayesian Neural Network, SVM, Naive Bayes, and tree-augmented Naive Bayes	Predictive
Viaene et al. (2007)	Automobile Insurance Fraud	External	Detection	Logistic Regression	Predictive
Yang and Hwang (2006)	Health Care Insurance Fraud	External	Detection	Frequent Pattern Mining	Predictive

If we summarize existing academic research by looking at Table 1, we arrive at the conclusion that merely all research is conducted in the field of external fraud. There is clearly a gap in the academic literature concerning internal fraud. Only six articles on internal fraud were found and they address only one kind of internal fraud: statement fraud. This is not even the number one internal fraud. Following the studies mentioned in Section II by PwC and ACFE, asset misappropriation, which is a form of transaction fraud, is the most prevalent kind of internal fraud. Transaction fraud is however no subject of existing research. Further it is confirmed by Table 1 that the bulk of literature aims at providing a detection tool; only two articles incorporate the importance of prevention. As a last observation, one notices that all articles found apply data mining techniques. This is a remarkable divergence of the non-academic research, where internal control was pointed as an effective detection tool, after chance related means (PwC 2007). Internal control does – to date - not include data mining approaches to mitigate fraud.

V. Mitigating External Fraud in Academic Research: The Value of Data Mining

In Table 1 the added value of a data mining approach in the context of fraud detection became clear. It is this approach that we wish to implement in our framework for internal fraud risk reduction. Before turning to the framework itself, this section deals with the most important aspects of the data mining research field. This background information is needed in order to make some non-trivial decisions for our framework, especially because our framework is oriented to internal fraud as opposed to the orientation to external fraud in academic research.

The current information age is overwhelmed by data. More and more information is stored in databases and turning these data into knowledge creates a demand for new, powerful tools. Data analysis techniques used before were primarily oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations and can help to get better insights into the processes behind the data. These interpretations and insights are the sought knowledge. Although the

traditional data analysis techniques can indirectly lead us to knowledge, it is still created by human analysts. (Michalski et al. 1998) The current situation however needed a new way to deal with these never ending databases and new methods to analyze this huge amount of data. A new area came into being: Knowledge Discovery in Databases, also known as KDD. The process of KDD can be mapped out as in Figure 2, a representation based on Tan et al. (2006).

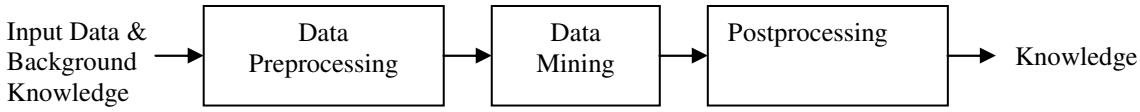

Figure 2. The process of knowledge discovery in databases, based on Tan et al. (2006)

As we can see in this figure, an integral part of the process of KDD is data mining. Together with KDD, data mining was born as a new research field. Data mining is a reaction to overcome the above limitations of data analyzing techniques used before (read: before there was this overwhelming amount of data). A data analysis system now has to be equipped with a substantial amount of background knowledge, and be able to perform reasoning tasks involving that knowledge and the data provided (Michalski et al. 1998). This is what data mining has an answer to. According to Witten and Frank (2000), data mining can be defined as

> *"…the process of discovering patterns in data. The process must be automatic or (more usually) semi-automatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities."*

In effort to meet this goal, researchers have turned to ideas from different disciplines. The machine learning field for example is often mentioned in the same breath as data mining, since it has provided lots of input to data mining. However, data mining also relies on statistics, artificial intelligence, and pattern recognition. Data mining is a

confluence of these disciplines.

With the coming of data mining as a new field of data analysis, data analyzing techniques can be divided into two groups: reporting techniques and data mining techniques. With reporting techniques we refer to the techniques used before, where quantitative and statistical data characteristics are extracted from data and human analysts turn this information into knowledge. (Think for examle at reports with some maximum, minimum and average numbers on sales or purchases.) These are the techniques currently used in internal control settings. With data mining techniques we emphasize the (semi-) automatic process to discover meaningful patterns in large data sets. Especially the data mining characteristic of revealing latent knowledge is very typical and valuable. This characteristic comes forward in the fact that no hypotheses are needed to mine the data, as opposed to pure statistics or data reporting. This is the main reason why these techniques are selected in previous research for detecting external fraud.

An important step in applying data mining is that of data engineering. What data do we have, what kind of information does it capture and what knowledge do we want to extract from it? Depending on the field you (exa)mine, you have information about accounts. An account can involve several things, like a customer's account, an invoice, a calling account and so on. In fact, we start from data about these accounts, we call this account data. For example, for a customer's account, what is the name of the customer, where does he live, what is his telephone number, when did he become a customer and so on. We do not only have account data, we also have operational information about an account. This kind of data describes the behavior of an account, like what was bought on an account, when, if there were any reductions… So actually we have two kinds of information available: account data and operational data on the account. A data mining approach links this information and attempts to alter technical data into behavior since the purpose of a data mining approach is to discover patterns in data.

There are many techniques the field of data mining encompasses, like K-means clustering, decision trees, neural networks etc. These techniques serve different tasks,

like for example classification, clustering, and anomaly detection. Mainly, data mining tasks can be divided in two subgroups: predictive tasks and descriptive tasks. With predictive tasks, the objective is to predict the value of one attribute, based on the values of other attributes. This is what classification techniques pursue. Predictive tasks make a prediction for every observation. Descriptive tasks however, do not pronounce upon every observation, but describe the data set as a whole. It aims to describe the underlying relationships in the data set. Examples of descriptive tasks are pattern recognition, anomaly detection, and correlations. (Tan et al., 2006)

In Table 1 an additional column is provided, stating what kind of task is used in a particular article. In the case of academic fraud detection literature, it appears that mainly predictive tasks are executed. Many different techniques serve this end. The class to be predicted is the label 'fraudulent'/'non-fraudulent'.

Aside from dividing data mining tasks in the groups predictive versus descriptive, there is yet another dimension to classify learning algorithms. Based on the input data, there are two categories of learning: supervised and unsupervised learning. In supervised learning, the class to be learned is present in the data set. In the fraud detection problem, this translates in a data set containing examples of both fraudulent and non-fraudulent records. This means that all the records available are labeled as 'fraudulent' or 'non-fraudulent'. After building a model using these training data, new cases can be classified as fraudulent or non-fraudulent. Of course, one needs to be confident about the true classes of the training data, as this is the foundation of the model. Another practical issue is the availability of such information. Furthermore, this method is only able to detect frauds of a type which has previously occurred. In contrast, unsupervised methods don't make use of labeled records. These methods seek for accounts, customers, suppliers, etc. that behave 'unusual' in order to output suspicion scores, rules or visual anomalies, depending on the method. (Bolton and Hand 2002)

Whether supervised or unsupervised methods are used, note that the output gives only an indication of fraud likelihood. No stand alone statistical analysis can assure that a particular object is a fraudulent one. It can only indicate that this object is more likely to

be fraudulent than other objects.

Mainly supervised data is used in the external fraud detection literature. With Bolton and Hand (2001), Murad and Pinkas (1999), Burge and Shawe-Taylor (2001), Brockett et al. (2002), Kim and Kwon (2006), and Cox et al. (1997), the most important studies concerning unsupervised learning in fraud detection are quoted. Although this list may not be exhaustive, it is clear that research in unsupervised learning with respect to fraud detection is due for catching up. This is also a possible explanation for the 'transaction fraud gap' in the literature. There is no supervised data available on this kind of fraud. The only internal fraud with supervised data available is statement fraud, not coincidentally the only kind of internal fraud investigated in the academic literature. We have to take this difference into consideration when constructing our framework for internal fraud risk reduction.

VI. The IFR² Framework

Internal fraud is currently dealt with by internal control. Internal control is embedded in a well elaborated framework, established by the COSO. Internal control encompasses a wide variety of tasks and settings. Next to a qualitative approach (like for example creating a control environment), quantitative data analyzing is required. It is at this point we believe there lies an opportunity to combine academic research with practical insights. Data mining tools are currently not implemented in the internal control framework. We are however convinced that a framework, based on data mining techniques, can be of additional value to internal control in mitigating fraud. Starting from the academic literature review and current practice, we introduce the IFR² framework as a complement of the existing internal control environment.

Since Table 1 shows the use of data mining for fraud detection/prevention is already explored by academics, we can continue on these insights. However, this research is not conducted in the field of internal fraud, or at least not covering all kinds of internal fraud. Because there are elements of distinction between found academic research and our aim, we cannot just copy existing methods of working. Instead, we present a

framework in which we implement data mining techniques in the area of mitigating internal fraud. Two major differences between our objective and existing work is that we 1) focus on internal fraud which typically involves unsupervised data, and 2) focus on fraud risk reduction instead of fraud detection. This is a contribution to the existing literature, where the use of data mining for (especially external) fraud detection only is investigated. These differences will have their effect on our framework, which will differ from the framework (although never explicitly registered!) used in existing literature. The IFR² framework is presented in Figure 3.

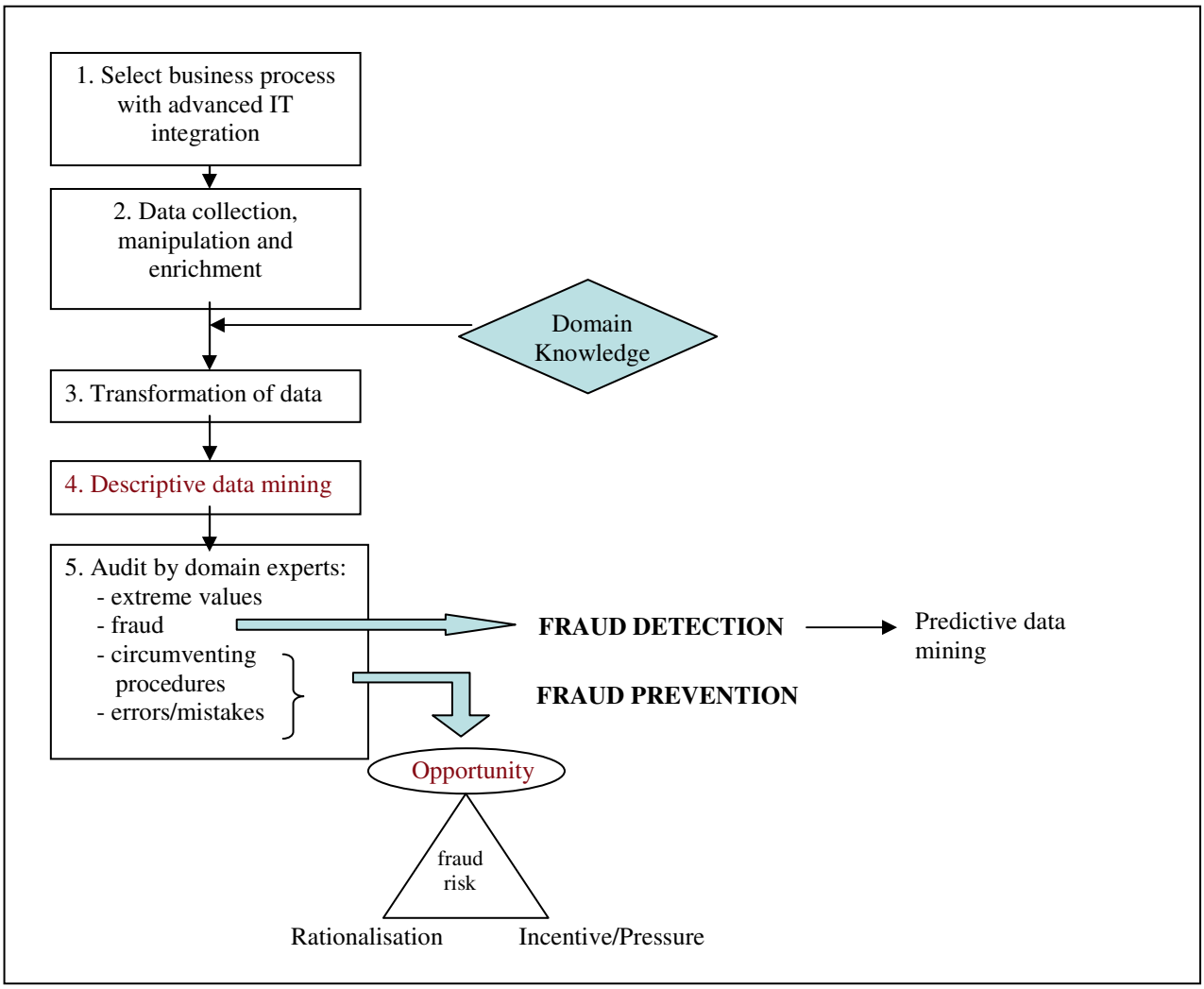

Figure 3. The IFR² framework

The IFR² framework starts with **selecting a business process with an advanced IT integration**. An organization should select a business process which it thinks is

worthwhile investigating. This selection can be motivated by different aspects: a business process that has a great cash flow, one that is quite unstructured, one that is known for misuses, or one that the business has no feeling with and wants to learn more about. Also the implementation of advanced IT, according to Lynch and Gomaa (2003), is a breeding ground for employee fraud. So selecting a business process with an advanced IT integration is a good starting point to encounter this stream of frauds.

After the selection of an appropriate business process, **data has to be collected, manipulated and enriched** for further processing. This is comparable to the step ”Data preparation” in Chien and Chen (2008)’s framework for personnel selection. The manipulation of data refers to the cleaning of data, merging connected data, transforming data into interpretable attributes and dealing with missing values. Although background knowledge may be required for executing this step, these are mainly technical transactions in that they still present operational data.

During the third step, **transformation of the data**, the operational data will be translated into behavioral data. This translation builds - even more than the second step - upon **domain knowledge** and is not just a technical transformation.

The core of the framework is then to apply a **descriptive data mining** approach for getting more insights in this behavioral data. This is where the IFR² framework remarkably differs from the followed methodologies in the existing literature. In the existing academic literature, almost all research applies a data mining technique with a predictive task. The explanation for the IFR² approach is twofold. Existing work predicts whether an observation is fraudulent or not. This can be explained by their focus on fraud detection. We however broaden our intentions, and are interested in all information, captured in the data, that helps us reducing the fraud risk, and not only the class ‘fraudulent/legal’. In order to retrieve more information and patterns in data, a descriptive data mining approach has to be pursued.

Another characteristic of internal fraud risk reduction is the presence of unsupervised data sets, liable to this stream of research. There are almost no supervised data sets

available in the context of internal fraud. This fact also accounts for the use of descriptive data mining instead of predictive data mining. An advantage of the use of descriptive data mining techniques is that it is easier to apply on unsupervised data. Thus for overcoming the exclusion of types of fraud where supervised data is difficult to obtain, the use of descriptive data mining techniques is recommended.

The core of this methodology -to use descriptive data mining- is also motivated by the higher intrinsic value a description of the data set under investigation provides than just a prediction of fraudulent versus legal. A description of the data set as a whole can bring insights to light, that were not clear before. All extra insights an analyst can gain are valuable to better understand what is going on, leading to a better position to mitigate internal fraud. When one only focuses on predicting the fraud class, one is not open minded enough to notice other interesting patterns. Association rules, clustering and anomaly detection are appropriate candidates for describing the data set. These can ultimately lead to observations or outliers, seeming interesting to take a closer look at. This is what happens in the fifth step of our methodology.

The fifth step is the **audit of interesting observations by domain experts**. The descriptives should provide the researchers a recognizable pattern of procedures of the selected business process. In addition some other patterns of minor groups of observation in the data can arise, interesting to have a closer look at. By auditing these observations, one can acquire new insights in the business process. As a general rule, one will always select outliers or extreme values to take a closer look at. Observations defined as outlier can normally be brought back to one of the following four cases: the observation is an extreme value but very logic when looked into, the observation is fraudulent, the observation is the result of circumventing procedures or it is simply a mistake. The regular observations will not draw our attention.

Observations defined as an outlier because they contain extreme values -but can be explained- are not of interest for our purpose. (Think for example at the purchase of a mainframe at the same department as the purchases of CDs.) Nevertheless, they can occur. The other three categories (fraud, circumventing procedures and mistakes) on the

other hand are of interest. If a fraudulent observation comes to our attention as an outlier, this is part of fraud detection. A fraud case can be interesting for adjusting current practice in the business process. If enough similar fraud cases are uncovered, a supervised fraud detection method can be elaborated for this specific fraud, based on a new data set. In this particular case, one can find well elaborated and tested methods in the existing literature. At this stage of investigating, **predictive data mining** tasks are recommended to search specifically for this type of fraud. The other two categories which can be at the origin of an outlier, circumventing procedures and making mistakes, are important in the light of fraud prevention. By making a mistake and realizing nobody notices or by circumventing procedures, a window of opportunity to commit fraud can develop. Opportunity, aside from rationalization and incentive or pressure, is one of the three elements of Cressey's fraud triangle[5]. Also according to Albrecht et al.'s "fraud scale" and even according to Hollinger and Park's theory, opportunity is an element of influence on fraud risk. Being able to select those cases where procedures are circumvented or mistakes are made, is an important contribution to taking away this opportunity and hence to prevent future fraud. The way in which this is dealt with, is up to the company. Internal controls can be adapted, persons can be called to account, procedures can be rewritten or other measures can be taken. This follow-up is not part of our framework anymore.[6]

Conclusion

In this conceptual paper, mitigating internal fraud plays a central role. To put this problem in the right context, we start with an elaborated fraud section about fraud in general. A definition, classifications, costs and other related information is provided. In two following sections, both the business practice in this context and existing academic literature is reviewed. Taking all information together, we deduce and present a

[5] Cressey's hypothesis, better known as the "fraud triangle", sees three elements necessary for someone to commit fraud. There has to be pressure, a perceived opportunity and the perpetrator must be able to rationalize its acts. The fraud triangle is cited many times in fraud literature and has become an important hypothesis. 'Opportunity' is the only element a company has influence on.

[6] Tennyson and Salsas-Forn (2002) show that claims auditing, in the field of automobile insurance fraud, works as fraud detection and as fraud deterrence (a way of preventing) as well. This proves the value of the fifth step of our framework.

framework to reduce internal fraud risk, the IFR² framework. This is prompted by the lack of such a methodology in academic literature, the severe costs internal fraud nevertheless presents and the important role it plays in the business environment.

To build our framework, the methodology followed by academics to fight external fraud inspired us, especially the application of data mining techniques. We also had a look at the current practical framework in the business environment to fight internal fraud: the internal control framework.

The IFR² framework has four major contributions. Firstly, the framework concentrates on mitigating *internal* fraud risk. This was not present yet in the academic literature there almost all research was conducted on external fraud. Secondly, the core of the IFR² framework builds upon a *data mining* approach. When future research investigates this suggestion further, this can be of significant value for organizations, where the current framework of internal control does not apply data mining techniques. We are convinced however that this can deliver additional insights to reduce internal fraud risk. Thirdly, the framework includes *descriptive* data mining techniques, as opposed to the use of predictive techniques in the existing external fraud methodology. This difference presents the benefit of not focussing on fraud detection only, but on detection and prevention. Hence fourhtly, fraud *risk is reduced* instead of only detected when it already took place.

We hope future work will use the IFR² framework to investigate the usefulness of particular analyzing techniques for internal fraud risk reduction. Also a uniform evaluation framework could be the subject of future research. Implementing this framework and its methodology as a complement of an internal control system within a cooperating company, could evaluate the added value for business practices.

References

Abidogum O. A. *Data mining, fraud detection and mobile telecommunications: Call pattern analysis with unsupervised neural networks*. PhD thesis, University of the Western Cape (2005).

ACFE. *2006 ACFE Report to the nation on occupational fraud and abuse.* Technical report, Association of Certified Fraud Examiners.

Albrecht W., Albrecht C., and Albrecht C. Fraud and corporate executives: Agency, stewardship and broken trust. *Journal of Forensic Accounting* 2004; 109-130.

Albrecht W.S., Howe K.R., and Romney M.B. *Deterring Fraud: The Internal Auditor's Perspective*. Institute of Internal Auditors Research Foundation (1984).

Bermúdez L., Pérez J., Ayuso M., Gómez E., and Vázquez F. A Bayesian dichotomous model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics 2007;* doi:10.1016/j.insmatheco.2007.08.002.

Bologna G. and Lindquist R. *Fraud Auditing and Forensic Accounting*. John Wiley & Sons, 1995.

Bolton R. and Hand D. Unsupervised profiling methods for fraud detection 2001.

Bolton, R. and Hand, D. Statistical fraud detection: A review. *Statistical Science* 2002; 17(3):235-255.

Bonchi F., Giannotti F., Mainetto G., and Pedreschi D. A classification-based methodology for planning audit strategies in fraud detection. In *KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA. ACM Press, 1999.

Boyer M.M. Resistance (to fraud) is futile. *The Journal of Risk and Insurance* 2002; 74(2):461-492.

Brause R., Langsdorf T., and Hepp M. Neural data mining for credit card fraud detection 1999.

Brockett P.L., Derrig R.A., Golden L.L., Levine A., and Alpert M. Fraud classification using principal component analysis of RIDITs. *The Journal of Risk and Insurance* 2002; 693:341-371.

Brockett P.L., Xia X., and Derrig R.A. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. *The Journal of Risk and Insurance* 1998; 652:245-274.

Burge P. and Shawe-Taylor J. An unsupervised neural network approach to profiling the behavior of mobile phone users to use in fraud detection. *Journal of Parallel and Distributed Computing* 2001; 61:915-925.

Cahill M., Lambert D., Pinheiro J., and Sun D. Detecting fraud in the real world. (2000)

Chien C.-F. and Chen L.-F.. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with Applications* 2008; 34:280-290.

Choo F. and Tan K. An "American Dream" theory of corporate executive fraud. *Accounting Forum* 2007; 31:203-215.

Cortes C., Pregibon D., and Volinsky C.. Communities of interest. *Intelligent Data Analysis* 2002; 6:211-219.

Cosserat, G.W. *Modern Auditing*. John Wiley & Sons, Ltd, 2 edition, 2004.

Cox K., Eick S., Wills G., and Brachman R.J. Viual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery* 1997; 1:225-231.

Davey N., Field S., Frank R., Barson P., and McAskie G. The detection of fraud in mobile phone networks. *Neural Network World* 1996; 64:477-484.

Davia H.R., Coggins P., Wideman J., and Kastantin J. *Accountant's Guide to Fraud Detection and Control*. John Wiley & Sons, 2 edition, 2000.

Derrig R.A. and Ostaszewski K.M. Fuzzy techniques of pattern recognition. *The Journal of Risk and Insurance*1995; 623:447-482.

Deshmukh A. and Talluru L. A rule-based fuzzy reasoning system for assessing the risk of management fraud. *International Journal of Intelligent Systems in Accounting, Finance & Management* 1998; 74:223-241.

Dorronsoro J., Ginel F., Sanchez C., and Santa Cruz C. Neural fraud detection in credit card operations. *IEEE Transactions on Neural Networks* 1997; 84:827-834.

Ernst&Young. *9th global fraud survey, fraud risk in emerging markets.* Technical report, Ernst&Young, 2006.

Estévez P., Held C., and Perez C. Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. *Expert Systems with Applications* 2006; 31:337-344.

Ezawa K.J. and Norton S.W. Constructing Bayesian networks to predict uncollectible telecommunications accounts. *IEEE Expert: Intelligent Systems and Their Applications* 1996; 115:45-51.

Fan W. Systematic data selection to mine concept-drifting data streams. *Proceedings of SIGKDD04*, 2004; 128-137.

Fanning K. and Cogger K. Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management* 1998; 7:21-41.

Fawcett T. and Provost F. Adaptive fraud detection. *Data Mining and Knowledge Discovery* 1997; 13:291-316.

Fawcett T. and Provost F. Activity monitoring: Noticing interesting changes in behavior. In Chaudhuri and Madigan, editors, *Proceedings on the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 53-62, San Diego, CA, 1999.

Green B. and Choi J. Assessing the risk of management fraud through neural network technology. *Auditing* 1997; 161:14-28.

He H., Wang J., Graco W., and Hawkins S. Application of neural networks to detection of medical fraud. *Expert Systems with Applications* 1997; 134:329-336.

Hilas C.S., and Mastorocostas P.A. An application of supervised and unsupervised learning approaches to telecommunications fraud detection. *Knowledge-Based Systems* forthcoming 2008.

Hoogs B., Kiehl T., Lacomb C., and Senturk D. A genetic algorithm approach to detecting temporal patterns indiciative of financial statement fraud. *Intelligent Systems in Accounting, Finance and Management* 2007;15:41-56.

Jensen M.C. and Mecklink W.H. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 1976; 3:305-360.

Juszczak P., Adams N.M., Hand D.J., Whitrow C., and Weston D.J. Off-the-peg and bespoke classifiers for fraud detection. *Computational Statistics and Data Analysis* forthcoming 2008.

Kim H. and Kwon W.J. A multi-line insurance fraud recognition system: a government-led approach in Korea. *Risk Management and Insurance Review* 2006; 92:131-147.

Kirkos E., Spathis C., and Manolopoulos Y. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* 2007; 32:995-1003.

Laub J.H. Edwin H. Sutherland and the Michael-Adler report: Searching for the soul of cirminology seventy years later. *Criminology* 2006; 442:235-257.

Lin J., Hwang M., and Becker J. A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal* 2003; 188:657-665.

Lynch A. and Gomaa M. Understanding the potential impact of information technology on the susceptibility of organizations to fraudulent employee behaviour. *International Journal of Accounting Information Systems* 2003; 4:295-308.

Maes S., Tuyls K., Vanschoenwinkel B., and Manderick B. Credit card fraud detection using Bayesian and neural networks. *Proc. of the 1st International NAISO Congress on Neuro Fuzzy Technologies, January 6-19, 2002*.

Major J. and Riedinger D. EFD: a hybrid knowledge/statistical-based system for the detection of fraud. *The Journal of Risk and Insurance* 2002; 693:309-324.

Michalski R.S., Bratko I., and Kubat M. *Machine Learning and Data Mining - Methods and Applications*. John Wiley & Sons Ltd., 1998.

Murad U. and Pinkas G. Unsupervised profiling for identifying superimposed fraud. *Lecture Notes in Computer Science* 1999; 1704:251-262.

NHCAA 2008. http://www.nhcaa.org/, consulted September 25, 2008.

NICB 2008. https://www.nicb.org/, consulted September 25, 2008.

Pathak J., Vidyarthi N., and Summers S. A fuzzy-based algorithm for auditors to detect element of fraud in settled insurance claims. *Odette School of Business Administration Working Paper No. 03-9*, 2003.

Phua C., Alahakoon D., and Lee V. Minority report in fraud detection: classification of skewed data. *SIGKDD Explorations* 2004; 61:50-59.

PwC. *Economic crime: people, culture and controls. the 4th bi-ennial global economic crime survey*. Technical report, PriceWaterhouse&Coopers, 2007.

Quah J.T. and Sriganesh M. Real-time credit card fraud detection using computational intelligence. *Expert Systems with Applications* 2007; doi:10.1016/j.eswa.2007.08.093.

Rosset S., Murad U., Neumann E., Idan Y., and Pinkas G. Discovery of fraud rules for telecommunications: Challenges and solutions. In *KDD '99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 409-413, New York, USA. ACM Press, 1999.

Sànchez D., Vila M., Cerda L., and Serrano J. Association rules applied to credit card fraud detection. *Expert Systems with Applications* 2008 forhtcoming.

Stolfo S., Fan W., Lee W., Prodromidis A., and Chan P.K. Cost-based modeling for fraud and intrusion detection: Results from the JAM project. *DARPA Information Survivability Conference & Exposition*, 2:1130-1144. IEEE Computer Press, 2000.

Tan P.-N., Steinbach M., and Kumar V. *Introduction to data mining*. Pearson Education, Inc. 2006.

Tennyson S. and Salsas-Forn P. Claims auditing in automobile insurance: fraud detection and deterrence objective. *The Journal of Risk and Insurance* 2002; 693:289-308.

Tittle C.R., Burke M.J., and Jackson E.F. Modeling Sutherland's theory of differential association: Toward an empirical clarification. *Social Forces* 1986; 652:405-432.

Tsung F., Zhou Z., and Jiang W. Applying manufacturing batch techniques to fraud detection with incomplete customer information. *IIE Transactions* 2007; 396:671-680.

Viaene S., Ayuso M., Guillén M., Gheel D.V., and Dedene G. Strategies for detecting fraudulent claims in the automobile insurance industry. *European Journal of Operational Research* 2007; 176:565-583.

Viaene S., Dedene G., and Derrig R. Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications* 2005; 29:653-666.

Viaene S., Derrig R., Baesens B., and Dedene G. A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk and Insurance* 2002; 693:373-421.

Wasserman S. and Faust K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1998.

Wells J. *Principles of Fraud Examination*. John Wiley & Sons, 2005.

Whittington O.R. and Pany K. *Principles of Auditing*. Irwin McGraw-Hill, 12 edition, 1998.

Witten I. and Frank E. *Data mining: practical machine learning tools and techniques with Java implementations*. Morgen Kaufmann, San Francisco, Calif., 2000.

Yang W.-S. and Hwang S.-Y. A process-mining framework for the detection of healthcare fraud and abuse. *Information and Security* 2006; 18:48-63.

Zaslavsky V. and Strizhak A. Credit card fraud detection using self-organizing maps. E*xpert Systems with Applications* 2006; 31:56-68.