Decision Support Systems 46 (2009) 853–864

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem

Carolyn Holton

Management Information Systems, College of Business and Legal Studies, Southeastern University, 1000 Longfellow Blvd, Lakeland, FL 33801, United States

ARTICLE INFO

Article history:
Received 25 January 2007
Received in revised form 7 November 2008
Accepted 13 November 2008
Available online 24 November 2008

Keywords:
IS security
Occupational fraud
Text mining
Design science
Disgruntled employee
Organizational communication

ABSTRACT

Occupational fraud is a $652 billion problem to which disgruntled employees are a major contributor. Much security research addresses reducing fraud opportunity and increasing fraud detection, but detecting motivational factors like employee disgruntlement is less studied. The Sarbanes–Oxley Act requires that companies archive email, creating an untapped resource for deterring fraud. Herein, protocols to identify disgruntled communications are developed. Messages cluster well according to disgruntled content, giving confidence in the value of email for this task. A highly accurate naïve Bayes model predicts whether messages contain disgruntled communications, providing extremely relevant information not otherwise likely to be revealed in a fraud audit. The model can be incorporated into fraud risk analysis systems to improve their ability to detect and deter fraud.

1. Introduction

The Sarbanes–Oxley Act [36] was created in the wake of a series of prominent financial scandals to protect investors from their recurrence. The Act's provisions endeavor to reveal and prevent corporate fraud. Rules issued by the Securities and Exchange Commission (SEC) to enforce the Act are being construed to require all public companies to store every document that influences financial reporting, including all email messages sent and received, for a number of years [5,28,38,47]. Managing the huge volumes of text employees create every day has been called the biggest challenge for companies seeking Sarbanes–Oxley compliance [34]. Industry-specific mandates such as Securities and Exchange Commission rules for brokers and traders, Medicare requirements for healthcare companies, and many other regulations pose their own email retention requirements [45,53].

The Sarbanes Oxley Act's focus is financial reporting and certification as a fraud deterrent, or failing that, to enable fraud discovery [9]. Legislative and regulatory requirements to store email, along with techniques for analyzing unstructured text data, create a less obvious path for fraud deterrence and detection: finding non-financial predictors and indicators of fraud risk or actual fraud in employees' email communications.

A national survey reports that 75% of organizations experienced fraud in the three months prior to the study, with employee fraud being the most prevalent [15]. Occupational fraud losses to companies in the United States are estimated to be around $652 billion per year, equivalent to an average of about 5% of total corporate revenues and a far greater share of profit [33]. Globally, the average fraud loss per company in the 2004–2007 period is estimated to be $8.2 million [25]. Despite attempts to curtail fraud, its incidence continues to grow both at home and abroad [10,17].

1.1. Detecting and deterring fraud

Auditors are charged with uncovering and deterring fraud. In the United States, the American Institute of Certified Public Accountants (AICPA) issues Statements on Auditing Standards (SAS) to guide the work of its members. SAS No. 99, Consideration of Fraud in a Financial Statement Audit, was issued in October 2002, partly in response to the same scandals that led to the Sarbanes–Oxley Act. SAS No. 99 makes identifying and investigating fraud risks an integral part of continuous audit processes [32].

The AICPA endorses a fraud risk detection model in accordance with criminology theory that is much like any good crime novel's means, motive and opportunity test. Three conditions commonly accepted as pre-requisites for fraud, opportunity, rationalization, and incentive [2,3], are sometimes referred to as the fraud triangle. By decomposing fraud risk assessment into these factors, the fraud triangle reduces the cognitive effort required for the activity, which may promote accuracy [51].

Much existing academic work on IS security risks addresses how to secure systems through deterrent (e.g. security awareness to promote appropriate safeguarding of passwords or safer use of

Tel.: +1 863 667 5665; fax: +1 863 667 5200.
E-mail address: cfholton@seuniversity.edu.
URL: http://www.seuniversity.edu.

0167-9236/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2008.11.013

Bluetooth devices) and preventive (e.g. physical locks, password access controls) activities to reduce the opportunity for malfeasance using a computer system [14,18], an important component of systems risk [42]. Another major stream of work relates to detection controls [10]. Information security journals have been described as saturated with articles promoting adherence to standards in these areas [41]. The discussion of cognitive factors is largely limited to knowledge and skills of the would-be perpetrator [52]. Components of the rationalization and incentive sides of the fraud triangle are little studied.

Opportunity is also a common focus of fraud auditing software tools [c.f. [1]]. For example, a typical procurement audit tool will analyze purchase orders, order receipts, invoices, payment amounts, quantities, dates, and the like to confirm that all money going out is accounted for in legitimate transactions corresponding to goods and services received in the amounts received. Duplicate payments, payments exceeding authority levels, or payments generated on weekends might be particular targets of this analysis activity. Audit tools flag unusual values for investigation, sometimes attaching a risk score to each potential fraud indicator identified.

Auditing packages may also analyze and assign risk scores for non-process data, such as the existence of potentially fraudulent entities. For instance, fictitious vendors may use post office boxes to receive payments and withhold physical addresses to make it harder to track down people associated with fraudulent transactions. They may not provide telephone numbers or may use answering services exclusively as they have no legitimate operating hours and want to limit links between the fictitious company and the person or people behind it. They may not provide tax identification numbers, which are difficult to fabricate without detection. Each of these signals a potential fraud opportunity, but does not necessarily indicate fraud. A risk assessment score can be assigned to each identified fraud opportunity factor. Typically only entities or items with total risk assessment scores over some threshold trigger investigation.

1.2. The case for incorporating disgruntled employee fraud risk indicators

Although a predominant focus of both IS security research literature and fraud audit tools is the opportunity to commit fraud, employee dissatisfaction has been found to be a far more powerful predictor of fraud risk than opportunity [48]. A large study of nearly 5000 employees concluded that employees' deviant behavior, including property deviance like workplace fraud, is a function of conditions inside the organization [22]. This finding was further narrowed in a study of more than 9000 employees that concluded the more dissatisfied an employee, the more likely he or she was to commit property deviance [23]. This study thus considers a key factor in the rationalization and incentive components of the fraud triangle: whether an employee is disgruntled with his or her employer (Fig. 1).

At the base of the fraud triangle, incentives motivate fraud. As being disgruntled has a positive relationship with property deviance towards one's employer, causes for that disgruntlement, for instance layoffs, the perception of inadequate compensation, or other dissatisfactions with an employer, serve as fraud incentives. Since the perception of unfair rewards or other dissatisfiers may lead to self-justification of fraud as taking what's owed, it may also have a role in the rationalization point [17].

Rationalization is the process of aligning an act of fraud with one's personal code of ethics [32]. Prior work has found that age, gender, and the "domain of morality" in operation are relevant to attitudes and behaviors in the context of ethical computer use [16]. The domain of morality is determined by what standards are in operation. These may be personal standards irrelevant to others; social norms, values and attitudes for the domain; or domain-independent standards like justice and fair allocation of resources [39]. Which of these drives behavior is one determinant of the rationalization component of the fraud triangle. Auditors may be more sensitive to opportunity and incentive than to rationalization [51]. They are not trained in discerning morality, and rationalization has been a source of consternation for them since its indicators are often not observable [43]. Automated methods for discovering rationalization thus hold high promise.

Organizations' abundant email archives provide a path for detecting fraud incentives and potential for rationalization. Disgruntled employee emails appear to be common: examples have been made public in lawsuits, managerial advice websites, industry journals, and trade press. Fraud prevention literature cautions that certain employee comments are predictive of criminal action, for instance blaming executives for things that go wrong, displaying excessive anger, and making threats. [31] Consistent with this

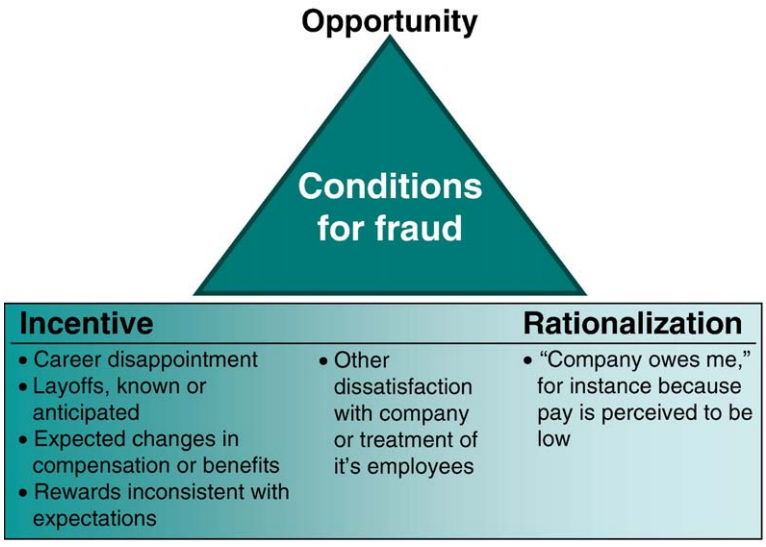

Fig. 1. Fraud triangle employee disgruntlement drivers.