Digital Investigation 22 (2017) 142–146

Contents lists available at ScienceDirect

Digital Investigation

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

The use of communicated negative sentiment and victimization for locating authors at-risk for, or having committed, insider actions

Eric Shaw*, Maria Payri, Ilene Shaw

Stroz Friedberg/Aon, Washington, DC, 20036, USA

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form
21 June 2017
Accepted 22 June 2017
Available online 24 August 2017

ABSTRACT

This article examines the challenge of locating persons at-risk for insider actions through communications by identifying psychological states and attributions associated with past insider violations. We specifically seek to expand and replicate two earlier studies (Shaw et al., 2013a, 2013b) which examined the relationship between negative sentiment and insider risk and the utility of psycholinguistic software targeting negative sentiment and feelings of victimization for locating communications of at-risk criterion groups within an organization's communication cache. The first study attempted to replicate Shaw et al. 2013a, in which two new rating tools for negative sentiment and insider risk were applied to 1000 randomly selected Enron archive communications to determine the frequency and overlap of messages containing negative sentiment and insider risk. In the present work the data set was expanded from 1000 to 10,000 messages. The second study attempted to replicate Shaw et al. 2013b by inserting 100 communications from authors at-risk for, or having committed, insider acts into the expanded Enron sample. The software used to operationalize the search for indicators of negative sentiment and victimization located 95% of the criterion group emails and all of the targeted authors inserted into the Enron sample. In Study 3, the repetition of this search using only indicators of negative sentiment caused a significant decrement in search results, reducing the percent of located emails from 95% to 58%. This result supported earlier findings (Shaw et al. 2013a, 2013b) emphasizing the dangers of using negative sentiment alone to locate indicators of insider risk. Finally, we described the true and false positive rates obtained when an enterprise version of the psycholinguistic software was deployed to locate similar at-risk communications in an actual mail cache containing over 50 million messages from over 63,000 Senders. The implications of all three studies and this field deployment for the relationship between negative sentiment and insider risk and the ability of investigators and analyst to locate potential insiders within organizational communication caches are discussed.

Background

Insider violations in industry and government pose an increasing concern to security, law enforcement, compliance and counter-intelligence staff, as well as the leaders of these organizations who answer to their respective constituents and the public. It is extremely rare to find an insider case that has not involved the use of information systems to plan, access, steal or exfiltrate the information involved, whether it is corporate or government secrets, or employee or customer information. It is equally rare to not find evidence of the risk of impending violence committed by insiders stored in relevant information systems. Shaw and Sellers (2015) have described the Critical Pathway to Insider Risk framework, which summarizes the "path" travelled by corporate and government insiders who have committed these violations with emphasis on early manifestations of disgruntlement within the organization.

While many types of technical indicators of insider risk exist, these systems tend to focus on anomalous employee use of information systems and related behaviors (hours of attendance, copying or downloading, patterns of information accessed and sent, etc.). Unfortunately, the fact that most insiders steal or damage information to which they have authorized and regular access and the amount of information and false positives these systems produce have limited their effectiveness. These systems also do not address the underlying psychological states, personality and decision-making by these individuals that accompany increased risk.

* Corresponding author. 5225 Connecticut Avenue NW, Suite 514, Washington, DC, 20015, USA.
E-mail address: eshaw@msn.com (E. Shaw).

http://dx.doi.org/10.1016/j.diin.2017.06.014
1742-2876/

Until recently, there have been no automated systems designed to detect the negative emotions and attitudes that are frequently associated with insider actions, despite the fact that employee frustration and anger have long been associated with aggression and violence in the workplace (Glomb and Liao, 2003; Hershcovis et al., 2007; Hershcovis and Barling, 2010), as well as turnover, absenteeism, accidents on the job, alcohol consumption, and other high-risk health behaviors (O'Neil et al., 2009). Holton (2009) also found an association between anger and fraud, and Band et al. (2006) found similar links to sabotage and espionage. Occupational health researchers who study a range of counter-productive work behaviors (CWBs), from taking long lunches to workplace violence, have consistently found a strong link between negative emotions and CWBs (Sakurai and Jex, 2012; Dalal, 2005; Brief and Weiss, 2002; Schat and Kelloway, 2005). More recently, Taylor et al. (2013) examined language associated with simulated insider activities and found that assigned insiders used language indicating greater self-involvement and more negative emotion compared to themselves prior to the simulated insider assignment and compared to a control group not assigned to simulate insider activity.

Ideally, a combination of detection approaches sensitive to both anomalous system behavior as well as the underlying disgruntlement that may motivate such violations might improve our ability to prevent, detect and intervene in insider risk cases. For example, information security personnel with numerous technical risk indicators might better prioritize the limited resources available to investigate these leads by starting with persons who also display signs of disgruntlement.

However, a critical but unanswered question for investigators and analysts concerned with the possible psychological precursors of insider violations is what percent of communications containing negative sentiment also contain indicators of insider risk. Are communications with negative sentiment or other alterations in language associated with insider activities a pathway for locating disgruntled at-risk individuals or a wild goose chase of false positives and unethical invasions of privacy?

Previous research on locating individuals at-risk for insider acts from their communications

Only a few previous studies have addressed the issue of what to look for in employee communications as an indicator of insider risk and whether insider risk can be effectively differentiated from negative sentiment in general. For example, Shaw et al. (2013a,b) introduced two new scales for the identification and measurement of negative sentiment and insider risk in communications by actual insiders, in order to examine the unexplored relationship between these two constructs. The inter-rater reliability and criterion validity of the Scale of Negativity in Texts (SNIT) and the Scale of Insider Risk in Digital Communications (SIRDC) were established with a random sample of emails from the Enron archive and a criterion group of established insiders, disgruntled employees, suicidal, depressed, angry, anxious, and other sampled groups. In addition, the sensitivity of the scales to changes over time as the risk of digital attack increased and transitioned to a physical attack was also examined in an actual case study. Inter-rater reliability for the SNIT was extremely high across groups (.944, $p \leq 0.05$) while the SIRDC produced lower, but acceptable levels of agreement (.823, $p \leq 0.05$). Both measures also significantly distinguished the criterion groups from the overall Enron sample. The scales were then used to measure the frequency of negative sentiment and insider risk indicators in the 940 random Enron email sample and the relationship between the two constructs. While low levels of negative sentiment were found in 20% of the sample, moderate and high levels of negative sentiment were extremely rare, occurring in less than 1% of communications. Less than 4% of the sampled emails displayed indicators of insider risk on the SIRDC.

Emails containing high levels of insider risk comprised less than one percent of the sample. Of the 222 emails containing negative sentiment in the sample, only 36, or 16.3%, also displayed any indicators of insider risk. The odds of a communication containing insider risk increased with the level of negative sentiment and only low levels of insider risk were found at low levels of negative sentiment. All of the emails found to contain insider risk indicators on the SIRDC also displayed some level of negative sentiment. While this research established the relative rarity of negative sentiment and signs of insider risk, as well as the differences between the two constructs, it involved a relatively small sample size.

Shaw et al. (2013a,b) subsequently tested the effectiveness of a psycholinguistic software program[1] previously used for investigations for locating the full range of these communications. After significant testing to determine the correct combination of psychological states and attributions (linguistic indicators of anger, blame, victimization) an additional randomized sample of communications from actual insiders and persons at-risk of insider actions previously coded for their SNIT and SIRDC values were inserted into a sample from the Enron archive to determine the software's ability to locate communications high and low in negative sentiment and insider risk. The software proved less effective in locating emails Low in negative sentiment on the SNIT and Low in insider risk on the SIRDC. However, the software performed extremely well in identifying communications from actual insiders randomly selected from case files and inserted in this email sample. In addition, it appeared that the software's measure of perceived Victimization was a significant supplement to using negative sentiment alone, when it came to searching for actual insiders. Previous findings (Shaw et al., 2013a,b) indicate that this relative weakness in identifying Low levels of negative sentiment may not impair the software's usefulness for identifying communications containing significant indications of insider risk because of the very low base rate and low severity of insider risk at Low levels of negative sentiment.

This preliminary review indicated that the software may not be effective for early identification of persons with Low levels of negative sentiment that may subsequently turn into individuals at-risk for insider activity. The low base rate for insider risk measured on the SIRDC of 16.3% for communications low in negative sentiment and the exclusively low level of insider risk contained in these emails indicates that the vast majority of these subjects present either little or no risk of insider actions. Further time series research will be necessary to determine whether this group Low in negative sentiment and insider risk ever converts to more concerning risk levels. Until such time, there are significant validity, resource allocation and ethical questions surrounding a focus on such individuals. The software's relative lack of sensitivity to lower levels of negative sentiment and insider risk in search mode would not limit its use in monitoring previously identified individuals with any level of risk or other sources of concern. Although many of the "false positives" acquired in the successful search for actual insiders in this experiment were shown to be true positives for other forms of insider risk, the software still produced fairly high rates of false positives that could burden analysts. An informal survey of the true positive rates of conventional insider detection software solutions focusing on technical anomalies and use of limited key words

[1] For more information on WarmTouch software (subsequently renamed Scout) see Shaw and Stroz (2004).