Insider Threat Detection: A Solution in Search of a Problem

Jordan Richard Schoenherr[1,2] and Robert Thomson[1]
[1]Army Cyber Institute / Behavioural Science and Leadership Department, US Military Academy
[2]Department of Psychology / Institute of Data Science, Carleton University
jordan.schoenherr@carleton.ca; robert.thomson@westpoint.edu

***Abstract*— Insider threats (IT) reflect a growing concern in security communities. Despite a rapid increase in the number of papers examining IT, definitions, research methods, models, and critical evaluations are rare. The present paper provides a critical review of these issues. Definitions of insider threat have varied from general: focusing on all forms of organizational deviant behavior, to specific: focusing on individual difference and social context variables. Research methods are based on deductive principles and intuitions of subject matter experts, computational models based on social media activity, and empirical observations based on synthetic or inaccessible data sets, i.e., black data. Following a review of these considerations, we demonstrate that many existing approaches within the behavioral and social sciences can provide more solid foundations to the IT literature. Using insight from research on organizational deviant behaviour and workplace incivility, we conclude by proposing a multidimensional classification system for insider threat SIEVE: severity (S), intentionality (I), type of employee norm violation (EV), and ethicality (E).**

Keywords— insider threat, incivility, workplace deviant behaviour, cyber crime

I. Introduction

Organizations have become increasingly concerned with 'malicious insiders': employees that act against their organizations' best interests. Malicious insiders present a threat due to their knowledge of the affordances of the software, systems, data, and operations within an organization. Insider threats (IT) have been attributed to internal (e.g., personality traits, personal crises) and external factors (e.g., a focus on perimeter-based defenses, inadequate security and screening procedures). [1] Studies suggest that a large proportion of intruders and unauthorized users within a network are organizational insiders (e.g. 34% of breaches [2]) and that many threats are the results of accidents (e.g., 5,830 attacks that were caused by malware/spyware downloaded by employees [3]). Recent surveys suggest that upwards of 27% of cybercrimes are committed by insiders. [4] The perception of threat along with reported statistics and unreported internal observations has resulted in the development of numerous detection approaches. [5]

Despite a growing consensus about the threat posed by malicious insiders and the development of purported detection methods, insider threat remains a poorly understood phenomenon: e.g., "Insider threat is a hard problem; there is no ground truth, innumerable variables, and sparse data," (p. 1.1; [6]). A lack of data reflects a significant problem for research and validation. [7, 8] Equally important, approaches to insider threat often neglect, or make superficial reference to, underlying psychological processes that might give rise to this behaviour. [9, 10] Similarly, simply detecting anomalous behaviour does not imply that the behaviour constitutes an insider threat. [11] Finally, evidence suggests that there are differences in perception in the prevalence of IT behaviours. In recent survey of 500 IT leaders and 4000 employees, significant discrepancies were observed in perceptions of insider threat. [12] Whereas 61% of IT leaders believed that employees have placed information at risk maliciously, 91% of employees believe that they have not intentionally done so.

In order to understand the methodological issues faced by researchers and developers of IT detection methods, we provide a critical review of the concepts, data and models developed in IT. Academic and nonacademic (grey literature) were surveyed. Our results suggest that there is little consensus over definition, that data sets are sparse or synthetic, and purportedly reliable data sets are inaccessible to the wider research community (i.e., 'black data'). Consequently, models of IT are limited in terms of their validity and generalizability. Finally, we consider the workplace incivility literature as a model of future studies of insider threat detection and propose a multidimensional model that differentiates and categorizes threats based on severity, intentionality, norm violation, and ethical motivations.

II. Behavioural and Definitional Issues

Research requires clearly defined constructs. [13, 14] Once defined, validity arguments are confined to the research context in which they were made. [15] For instance, an issue with defining insider threats stem from how individuals and organizations define an 'insider', i.e., group member. Partnerships between public and private sector further blur the boundary between insiders and outsiders [16] making discrete analytic categories problematic. Thus, how IT is operationalized in a given study is crucial for determining both their validity and generality.

Definitions of Insider Threat. Definitions of insider threat vary substantially (for a review, see [17, 18]). Frequently, the concept is left undefined or underspecified, with studies referring to maladaptive organizational behaviour with little qualification. For instance, "The insider threat refers to harmful acts that trusted insiders might carry out," (p. 1.2; [19]), "The malevolent insider manifests when a trusted user of the information system behaves in a way that the security policy defines as unacceptable." [20] Similarly failure to report suspicious behaviour has also been defined as a 'passive' insider threat. [21] Broad definitions present serious issues for accurate identification and prediction such that all employees could be defined as insider threats.

Definitions have also considered intentionality, emphasizing deliberation [22] or malice. Other definitions also include unintentional behaviour that is not malicious [23, 17,

24] or do not address intentionality directly. [7] According to these definitions, insider threats reflect behaviours that run counter to an organization established policies, procedures, and priorities either intentionally or unintentionally. Examples of insider threat include destruction/deletion of critical information assets, making unauthorized changes to data or records, unauthorized extraction or duplication of data or records, intentional or unintentional disruption of networks (e.g., malware), loss of equipment, and eavesdropping or packet sniffing. From this perspective, typologies are less important than the fact that they deviate from accepted norms and conventions (e.g., for a discussion of this in national security [25]).

Researchers have also attempted to provide basic distinction between kinds of insider threats based on data derived from user profiles. [26] As Salem et al. [27] note, three broad types of approaches exist: host-based user usage profiling (Unix, Windows, program, and web), network-based sensors (network observable user actions and honeypots), and integrated approaches that combine features of user profiling and sensors. The adoption of a given approach will depend on the kind of insider threat detection. For instance, user profiling might be effective at detecting insider masquerading attempts whereas 'honeypots' [28] might be more effective for leaker detection. [18] Similarly, this will differ from unintentional insider threats that can be the result of social engineering attempts (e.g., phishing/ spear-phishing, fraudulent websites, or reverse social engineering [19, 29, 30]).

Using a case-based approach, Band et al. [31] claim that the analytic categories of spy and saboteur are not supported by differences in behaviour. However, insider behaviour differs in terms of whether it reflects masquerading or leaking. [27] Masqueraders are those that assuming the identity of an authorized network user. [32, 27, 33] Assuming that users behave in a relatively consistent manner, masqueraders are detected by observing changes in a user's behaviour that are inconsistent with an existing profile. Alternatively, leakers are authorized user that behave with relative consistency within a network. These individuals will access the same file locations making detection more problematic. Crucially, leaking is likely to occur outside a network (e.g., providing information to an interested party after work).

While 'intentions' are sometimes referenced, there is little elaboration as to what these intentions are or how they might change behaviour. Understanding the psychology of attacker is also required. [10] As Andersson and Pearson [34] note in their account of workplace incivility, considering only the extent and nature of a 'deviant' workplace behaviour ignores individual motivations and the social context which are required to accurately predict behaviour within a group. Many models of insider threat have failed to consider this possibility or specify the underlying social dimensions (cf. [17, 19]).

Using interviews, Searle and Rice [39] incorporate aspects of motivation into their insider threat typology. Following 12 interviews concerning three critical incidence within an organization, they identify four different kinds of insiders: those who failed to regulate their own interpersonal behaviour, unintentionally violating rules (*omitters*), unsystematically engaged in deviant behaviour that might reflect a serious violation of workplace norms (e.g., removal of documents; *slippers*), systematic engaged in major and minor violations of workplace norms (*serial transgressors*), and engaged in small reciprocal active (revenge) or passive (withdrawal) aggressive behaviours directed toward individuals or the organization (*retaliators*). Moreover, they also consider individuals who failed to report suspicious behaviour (e.g., moral disengagement; *passive insider threats*). While providing informative categories, their studies does not systematically relate these categories to underlying psychological dimensions such as motivation.

III. Insider Threat: Models and Factors

A. Case-Based Approach to Insider Threat Modelling

Researchers and analysts have attempted to identify a number of indicators of IT. For instance, the US Department of Homeland Security identifies a number of characteristics of "insiders at risk of becoming a threat" (Table 1).[1] Many of the characteristics identified by the DHS were outlined by Shaw et al. [35] presented a "critical path" model of Critical Information Technology Insiders (CITIs) using a case study-based approach to identify any factors that associated with insider threat. [36] This approach selectively presented psychological studies to identify a number of factors that *might* be associated with IT in information science professionals. As we will show, evidence from psychological research suggests that indicators such as introversion and ethical flexibility are not reliable.

Characteristics of Insiders at Risk of Becoming a Threat	
Introversion, Greed/financial need, Vulnerability to blackmail, Compulsive and destructive behaviour, Rebellious, passive aggressive, Ethical "flexibility", Reduced loyalty, Entitlement – narcissism (ego/self-image).	Minimizing their mistakes or faults, Inability to assume responsibility for their actions, Intolerance of criticism, Self-perceived value exceeds performance, Lack of empathy, Predisposition towards law enforcement, Pattern frustration and disappointment, History of managing crises ineffectively.

Table 1. Purported variables related to insider threat. Adapted from DHS. [37]

Introverts. Introversion occupies the first position in the DHS characteristics list and a frequent topic of discussion in the IT literature. Introversion reflects a personality dimension identified in a number of prominent psychological models. [38] However, levels of introversion are not always stable. Observed and self-reported behaviour is not always a good indicator of extroversion or introversion as individual can change their

[1] The document contains a disclaimer, noting that the DHS "does not provide any warranties of any kind regarding any information contained within"

responses depending of situational demands. [39] Moreover, Shaw et al.'s discussion of introversion does not highlight it as a risk factor for IT. Rather, they claimed that computer programmers were generally introverted and that this makes detection of IT behaviour more difficult. Moreover, in one of only three case studies examined by Searle and Rice [21] was the perpetrator identified as an introvert. Consequently, there is no evidence that is publicly accessible that supports a systematic relationship between introversion and insider threat.

Ethical Flexibility. Shaw et al. additionally discussed ethical "flexibility", attributing it to "a subgroup [of computer professionals] whose members do not object to acts of cracking, espionage and sabotage against information." However, ethical flexibility reflects another normal feature of human social cognition. Despite early theories of the development of moral judgment, ethical decision-making is influenced by many factors including time constraints and context. [40] More generally, deviation from ethical norms appears to be associated with anonymity with both adults and children engaging in fewer prosocial and more antisocial behaviours. [41] Thus, ethical flexibility alone is not necessarily a predictor of IT although it is possible that it is mediator.

Subject Matter Experts and Bayesian Models. A highly cited model is provided by Greitzer et al. [42]. In their Bayesian belief network model, that was 'informed' by the five-factor model of personality. [43] In their model, prior probabilities were obtained by asking HR experts to estimate the number of cases per year that were associated with a given psychosocial indicator. Scenarios were then provided to the HR experts and the Bayesian model. Using these values in their Bayesian model, Greitzer et al. [42] observed a strong, positive correlation between expert ratings and model predictions of IT ($R^2 = 0.920$).

Despite a strong correlation, the basis for this result is questionable. It is not clear that HR experts were accurate detectors of insider behaviour, especially in case of behaviours related to specific personality traits (see below). Acknowledging this, they note that "psychosocial risk… involves a large number of complicated judgments in which various numbers of factors are combined," leading them to acquire "expert judgments for only about 3% of the total number of cases" (p. 4.12). Moreover, their results only demonstrate that a Bayesian model can perform like an HR expert. Thus, if the HR expert and Bayesian model are correctly calibrated, there is likely a high probability of detection. If they are not, the Bayesian model merely reflect the folk theories of HR professionals.

Similar problems are evidenced in other approaches. Greitzer et al. [19] assessed twenty-four case studies to develop a model of unintentional insider threat, 25 of which were derived from newspaper reports. Consequently, this study reflects an extension of a case-study based approach based on abstract, publicly accessible reports. Factor inclusion criterion are also unclear. For instance, while attention and stress/anxiety were included in the model, the former was included due to "at least of case" and the latter factor was included because there was a single case wherein the knowledge the victim had of phishing attacks received by their organization "may have increased his desire to accept an offer of mitigation that appeared legitimate, though it was actually another phishing attack," (p. 243).

Linguistic Indicators. Another proposed method for detection is the use of linguistic indicators. For instance, following Shaw et al.'s proposals between IT and revenge, Kandias et al. [20] explored whether negative attitudes toward law enforcement and authorities might support insider threat detection. Using a web crawler, they obtained YouTube users comments on videos and classified them into positive and negative attitudes using machine learning algorithm (Naïve Bayes Multinomial, Support Vector Machines, and Logistic Regression). With caveats concerning ethical and legal considerations, they suggest that this approach can be used predict IT. While 'dark' personality characteristics (e.g., sadism, psychopathy) likely predispose individuals to engage some form of antisocial behaviour, it is not necessarily the case that general attitudes will result in specific behaviour within an organization.

Dark personality traits (e.g., Machiavellianism) could also impede detection. In a study of workplace behaviour that examined the influence of Dark Triad traits, Jonason et al. [44] found that co-workers high in Machiavellianism and narcissism were more likely to engage in 'soft tactics' (offering compliments) while those high in psychopathy and Machiavellianism were more likely to use 'hard tactics' (e.g., aggression) to obtain goals. As a method, this presents challenges for linguistic analysis in a group setting. For instance, in a study by Ho et al. [45] casts doubt on the use of linguistic markers can be used to identify malicious insiders. They provided a collaboration task wherein half the leaders of teams were offered an opportunity to defect after the game started with a corresponding incentive to do so. They then analyzed chat log content for differences between leaders of teams who had been offered the defection incentive and those who had not. They additionally looked for changes before and after the incentive was offered. Crucially, they did not find evidence to support differences in language used by 'insiders.' Whether these results generalize to real world scenarios is an open question. For instance, the interactions observed here were on a brief time-scale, resulting in limited knowledge of other players and limited interpersonal and group cohesion.

IV. Predictors of Insider Threat: Using Black Data and the Prospects of Incivility

Despite many potential blind alleys, studies have identified a number of promising factors related to intentional and unintentional IT behaviours. For instance, males appear to be less susceptible than females in phishing attempts. [46] Kenney et al. [47] reviewed forty-nine cases of insider threat and found 85% of insiders threats held a prior grievance and 84% were associated with the desire for revenge. In an analysis of sixty-two security breaches, Shropshire [8] found that both financial and relationship changes were strongly correlated with espionage. Moreover, relationship and job changes as well as

substance abuse were strongly correlated with information technology espionage.

Using correlational logic, Capelli et al. [23] analyzed 116 cases of insider threat that were broadly classified as fraud, IT sabotage, and theft of confidential or proprietary information. They note that "Over half of the insiders [who committed IT sabotage] were perceived as disgruntled, and most of them acted out of revenge for some negative event," (p. 7). They also note that the majority of the insiders who committed fraud (93.2%) or information theft (75%) were current employees. With 45% of insiders who committed information theft having accepted a position elsewhere at the time of their theft. However, while individuals who have committed insider attacks might no longer feel committed to an organization (i.e., psychological exit) or have intentions to quit (i.e., organizational exit), this does not mean that all individuals who are 'disgruntled' or intend to leave represent an insider threat.

Limitations of Datasets. However, both Shropshire [8] and Capelli et al. [23] used CERT and other US databases (USSS) suggesting that the results come from the same, or highly related, datasets. Moreover, these studies failed to contrast IT cases with normal organizational behavior and other misbehavior that deviates from workplace norms but that fail to reflect an IT. A concern is evidenced with another approach to data in insider threat detection, synthetic data. The CERT website provides synthetic data for model development. [4] Despite the validity of social simulation as an approach and the potential utility of such datasets, [7] they are only valid to the extent that their structured reflects properties of insider threats and their environment (i.e., network traffic). In that they are not widely accessible, black data and models developed on it, cannot be validated.

Uncivil Behaviour. Research and commentaries on IT have suggested a relationship with behaviours associated with incivility. [48, 49, 42, 6] Incivility reflects behaviour that deviates from workplace place norms. [34, 50, 51] Unlike antisocial (e.g., aggression) or prosocial behaviours (e.g., supportiveness), the intentions behind this behaviour are ambiguous, potentially resulting from uncertainty in processing available social cues. [52]

If a causal relationship exists between incivility and IT, it is likely complex. For instance, uncivil behaviour is likely an indicator of an employee who is no longer invested in their work (reduced task cohesion) or their co-workers (reduced social cohesion). However, these reductions in cohesion might be attributable to the experience of incivility. Namely, Andersson and Pearson [34] note that there is a 'spiral of incivility' that results from experiencing disrespect in their workplace. Consequently, incivility might be an indicator of a dysfunctional *workplace* that might lead insiders to engage in malicious behaviour. However, this might not be an individual-level phenomenon. Instead, organization should consider how workplaces might promote insider threats.

Schoenherr and Nguyen [52] present a simple model of incivility that provides one means to understand insider threat. Their Multi-agent Accumulator-based model of Decision-making of Incivility (MADI) assumes that members of working groups accumulate prosocial and antisocial cues from interpersonal interaction. When the amount of either prosocial or antisocial cues reaches a threshold, an employee responds with the corresponding overt behaviour (i.e., respectful or disrespectful). However, agents additionally retain an experience of uncertainty from one interaction to another based on the presence of both prosocial and antisocial cues, i.e., 'mixed feelings.' This provides a model of 'the spiral of incivility' that unfolds over time and could be used to predict reduction in organization commitment (psychological exit) as quitting (organizational exit). [53]

Using MADI, Schoenherr and Nguyen [53] simulated small group interactions that would occur in the workplace. Social agents were generally able to identify social cues accurately as respectful or disrespectful, however, given that some mistakes could occur, respectful behaviour could be misconstrued as disrespectful behaviour. During the first simulation period, a model was presented with respectful behaviour. Social agents were programmed to either have a bias to respond to disrespectful social cues or were somewhat tolerant. At the end of 12 simulated interactions, highly levels of predicted incivility indicative of psychological and organizational exit (Figure 1). The level of both predicted psychological and organization exit were higher for groups with social agents who were less tolerant of disrespectful cues.

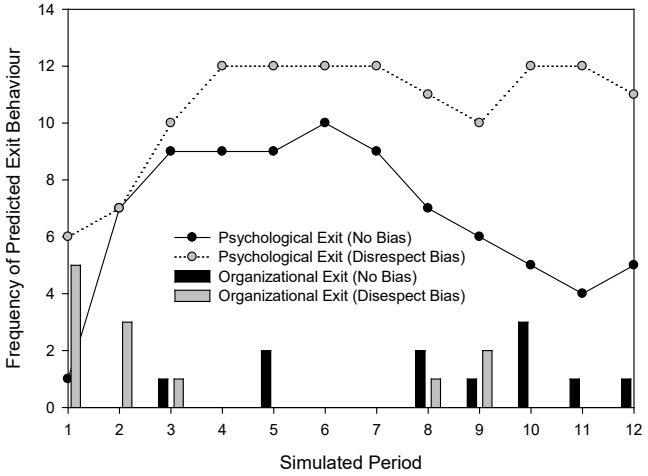

Figure 1. Organizational exit predicted by a model of group incivility adapted from Schoenherr and Nguyen (2019).

While a simulation, this same approach can be used to understand the conditions in which IT is observable as well as a possible relationship with incivility. Namely, the prevalence of incivility and disrespectful behaviour within a group should predict possible psychological distress and psychological exit, which would increase the likelihood that any given individual might leave an organization or leak information. For instance, if incivility is associated with increased stress and avoidance of other employees, [51, 50] an employee might be more inclined to attempt to find employment elsewhere, and develop greater cohesion with individuals and groups outside the organization.

Given that the limited empirical evidence on IT suggest that employees often have the intention to leave an organization when they commit these acts, [47] incivility might reflect an important marker of potential malicious insider behaviour that can be easily detected. Moreover, in that Schoenherr and Nguyen [52, 53] demonstrate how individual differences can affect responses to perceived incivility, organization can attempt to identify 'hotspots' in groups by assessing employees social competencies and personality traits. Such an approach to insider threat detection places a greater emphasis on proactive organizational responses rather than *solely* focusing on *post hoc* responses to malicious insiders.

A. SIEVE: A Multidimensional Approach to IT Behaviours.

The preceding review of the IT literature suggests that approaches adapted from the information and computer science need to be complimented with theories from the behavioural and social sciences. Following from work on organizational deviant behaviour, we assume that insider threat can be understood in terms of a number of social-psychological dimensions. For instance, Robinson and Bennett [54] found that employees perceive deviant behaviour in terms of severity (major or minor) and the type of norm violation (interpersonal or organizational). They found that members of an organization might target a specific individual (gossip or harass), or they might engage engaged in actions against their organization (theft or sabotage). This suggests that 1) the severity of norm violation as well as 2) the type of employee norm violation are likely relevant dimensions in terms of understanding insider threat.

However, Robinson and Bennett were only interested in intentional behaviour (p.556). Following the suggestion of Andersson and Pearson [34] in their work in workplace incivility, 3) *intentionality* would likely be an important third dimension. For instance, some forms of insider threats are clearly intentional such as when an employee shares proprietary information to a competing organization or adversarial country. In contrast, other forms of insider threats clearly lack intentionality, such as the unintentional loss of data due to downloading a virus. The intentionality dimension is also important from a legal perspective in that unintentional loss of data would likely be associated with a lesser penalty relative to intentional theft, destruction, or sabotage or corporate assets. [25]

Finally, if intentional, malicious behaviour can be further subdivided by 4) the *ethical motivation*. Namely, aggression can be used for personal gain (instrumental aggression) or as a means to correct perceive transgression of members within the group (hostile aggression). Indeed, as Fiske and Rai [55] note, aggression is often used to correct perceived injustice. This difference can help further distinguish between malicious insider behaviour that is directed toward 'hurting' an individual within an organization (e.g., supervisor) or organization as a whole to 'get even' (e.g., for terminated an employee, or for inadequate remuneration). For instance, Keeney et al. [47] found that insider behaviour frequently preceded by a prior grievance (85%) or termination (47%) and that the majority (84%) were associated with revenge motivation. In contrast, while whistleblowing might be classified as 'malicious behaviour' within an organization, it might be perceived as virtuous for an external group, i.e., 'society'.

Taken together, we believe that these four dimensions can be used to differentiate IT behaviour from other organizational deviant behaviour and incivility: **severity (S)**, **intentionality (I)**, **type of employee norm violation (EV)**, and **ethicality (E)**, or SIEVE. SIEVE can be used to classify organizational behaviours and identify specific forms of insider threat that arise as a result of distinct motivations. For instance, an employee caught in a phishing scam (e.g., minor, unintentional, organizational), a terminated employee stealing information to gain employment elsewhere (e.g., major, intentional (instrumental), organizational), or deleting information that would be important to their direct supervisor to gain revenge (e.g., moderate, intentional (hostile), interpersonal).

V. CONCLUSIONS

Research on IT has often suffered from inadequate empirical and analytic methods and equally problematic approaches to modelling. IT research needs to identify reliable indicators and consider typologies of insider threats. These typologies reflect important categories due to legal ramifications of insider behaviour. [25] Namely, the kind of behaviour engaged in whistleblowers such as Jeffrey Wigand in the US [56] or Nancy Oliveri in Canada [57] based on public health interests, unintentional leaks due to unsecured networks or email sent to the wrong address, from intentional corporate espionage. Detection mechanisms alone might not be equipped to differentiate these categories. Moreover, even if suspect behaviour is detected using employee and network monitoring, it does not necessarily suggest that an insider threat has been detected [11], i.e. a false positive.

In order to better understand the nature of possible insider threats, we propose model for IT that differentiates the severity, kind, and intentions of behaviour. When intentional, we further differentiate the motivations of the malicious insider in terms of the goals being interpersonal or instrumental aggression. By emphasizing employee motivations, [10] organizations should be capable of developing a more effective understanding of insider threats as well as developing network and employee monitoring systems. For instance, observations of incivility within the workplace might indicate that individuals might be experiencing less organizational commitment and a desire to leave an organization, making them more susceptible to external influences or engage in inadvertent leaking of sensitive information. By monitoring workplace climate, organizations could take more proactive steps to reduce insider threats.

Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under the Cooperative Agreement Number W911NF-19-2-0223. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for the Government purposed notwithstanding any copyright notation herein.

VI. References

[1] I. Homoliak, F. Toffalini, J. Guarnizo, Y. Elovici and M. Ochoa, "Insight into insiders and it: A survey of insider threat taxonomies, analysis, modeling, and countermeasures," *ACM Computing Surveys (CSUR),* vol. 52, pp. 1-40, 2019.

[2] Verizon, "Insider Threat Report: Out of Sight Should Never be Out of Mind," Verizon Business Study, 2019.

[3] I. Grant, "Insiders cause most IT security breaches, study reveals," ComputerWeekly.com, 26 August 2009. [Online]. Available: https://www.computerweekly.com/news/1280090551/Insiders-cause-most-IT-security-breaches-study-reveals. [Accessed 1 December 2019].

[4] CERT, "Insider Threat Dataset," 2020. [Online]. Available: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=508099. [Accessed 1 February 2020].

[5] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis," *Big Data Analytics,* vol. 1, 2016.

[6] C. F. Noonan, "Spy the Lie: Detecting Malicious Insiders (No. PNNL-SA-122655).," Pacific Northwest National Lab.(PNNL), Richland, 2018.

[7] J. Glasser and B. Lindauer, "Bridging the gap: A pragmatic approach to generating insider threat data," in *IEEE Security and Privacy Workshops*, IEEE, 2013, pp. 98-104.

[8] J. Shropshire, "A canonical analysis of intentional information security," *Information Management & Computer Security,* vol. 17, pp. 221-234, 2009.

[9] M. Maasberg, J. Warren and N. L. Beebe, "The dark side of the insider: detecting the insider threat through examination of dark triad personality traits," in *48th Hawaii International Conference on System Sciences*, IEEE, 2015, pp. 3518-3526.

[10] K. R. Sarkar, "Assessing insider threats to information security using technical, behavioural and organisational measures," *Information Security Technical Report,* vol. 15, pp. 112-133, 2010.

[11] I. Agrafiotis, A. Erola, J. Happa, M. Goldsmith and S. Creese, "Validating an insider threat detection system: A real scenario perspective," in *2016 IEEE Security and Privacy Workshops (SPW)*, IEEE, 2016, pp. 286-295.

[12] O. Matters, "Insider Data Breach Survey 2019," Egress, 2020. [Online]. Available: https://pages.egress.com/rs/344-XTD-684/images/egress-opinionmatters-insider-threat-research-report-a4-uk-digital.pdf. [Accessed 2 Feburary 2020].

[13] M. H. Marx, "The general nature of theory construction," in *Theories of Contemporary Psychology*, London, MacMillan, 1963, p. 3–46.

[14] S. Messick, "Validation of inferences from persons' responses and performances as scientific inquiry into score meaning," *American Psychologist,* vol. 50, pp. 741-749, 1995.

[15] M. T. Kane, "An argument-based approach to validity," *Psychological Bulletin,* vol. 112, p. 527–535, 1992.

[16] I. Crinson, "Assessing the 'insider–outsider threat' duality in the context of the develop-ment of public–private partnerships delivering 'choice' in healthcare services: A socio-material critique," *Information Security Technical Report,* vol. 13, pp. 202-206, 2008.

[17] L. Coles-Kemp and M. Theoharidou, "Insider threat and information security management," in *Insider threats in cyber security* , Boston, Springer, 2010, pp. 45-71.

[18] J. Hunker and C. W. Probst, "Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques," *JoWUA,* vol. 2, no. 1, pp. 4-27, 2011.

[19] F. L. Greitzer, J. R. Strozer, S. Cohen, A. P. Moore, D. Mundie and J. Cowley, "Analysis of unintentional insider threats deriving from social engineering exploits," in *IEEE Security and Privacy Workshops*, IEEE, 2014, pp. 236-250.

[20] M. Kandias, V. Stavrou, N. Bozovic and D. Gritzalis, "Proactive insider threat detection through social media: The YouTube case," in *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society*, ACM, 2013, pp. 261-266.

[21] R. Searle and C. Rice, "Assessing and Mitigating the Impact of Organisational Change on Counterproductive Work Behaviour: An Operational (Dis)trust Based Framework," The Centre for Research and Evidence on Security Threats (CREST), 2018.

[22] T. O. Oladimeji, C. K. Ayo and S. E. Adewumi, "Review on Insider Threat Detection Techniques," *Journal of Physics: Conference Series ,* vol. 1299, p. 12046, 2019.

[23] D. M. Cappelli, A. P. Moore and R. F. Trzeciak, "The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes," Addison-Wesley Professional, 2012.

[24] J. R. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. Wright and M. Whitty, "Understanding insider threat: A framework for characterising attacks," in *IEEE Security and Privacy Workshops*, IEEE, 2014, pp. 214-228.

[25] M. R. Papandrea, "Leaker traitor whistleblower spy: National security leaks and the first amendment," *BUL Review,* vol. 94, p. 449, 2014.

[26] S. Mathew, M. Petropoulos, H. Q. Ngo and S. Upadhyaya, "A data-centric approach to insider attack detection in database systems," in *International Workshop on Recent Advances in Intrusion Detection* , Berlin, Springer, 2010, pp. 382-401.

[27] M. B. Salem and S. J. Stolfo, "Modeling user search behavior for masquerade detection," in *International Workshop on Recent Advances in Intrusion Detection*, Berlin, Springer, 2011, pp. 181-200 .

[28] L. Spitzner, "Honeypots: Catching the insider threat. In," in *Proceesings of the 19th Annual Computer Security Applications Conference*, IEEE, 2003, pp. 170-179.

[29] L. Laribee, "Development of Methodical Social Engineering Taxonomy. Master's Thesis," Naval Postgraduate School, Amazon Digital Services, 2006.

[30] T. R. Peltier, Social Engineering: Concepts and Solutions, Information Systems Security, 2006, pp. 13-21.

[31] S. R. Band, D. M. Cappelli, L. F. Fischer, A. P. Moore, E. D. Shaw and R. F. Trzeciak, "Comparing insider IT sabotage and espionage: A model-based analysis (No. CMU/SEI-2006-TR-026)," CARNEGIE-MELLON UNIV PITTSBURGH, 2006.

[32] R. A. Maxion, "Masquerade detection using enriched command lines," in *In the Proceedings of the International Conference on Dependable Systems and Networks*, IEEE., 2003, pp. 5-14.

[33] J. E. Tapiador and J. A. Clark, "Masquerade mimicry attack detection: A randomised approach," *Computers & Security,* vol. 30, pp. 297-310, 2011.

[34] L. Andersson and C. Pearson, "Tit for Tat? The spiraling effect of incivility in the workplace," *The Academy of Management Review,* vol. 24, pp. 452-471, 1999.

[35] E. D. Shaw, K. G. Ruby and J. M. Post, "The insider threat to information systems," *Security Awareness Bulletin,* vol. 2, pp. 27-46, 1998.

[36] E. D. Shaw and L. Sellers, "Application of the Critical-Path Method to Evaluate Insider Risks.," *Studies in Intelligence,* vol. 59, pp. 1-8, 2015.

[37] N. C. a. C. i. Center, "Combating the Insider Threat," NCCIC/US-CERT, 2014.

[38] R. Cuperman and W. Ickes, "Big Five predictors of behavior and perceptions in initial dyadic interactions: Personality similarity helps extraverts and introverts, but hurts "disagreeables"," *Journal of Personality and Social Psychology,* vol. 97, p. 667, 2009.

[39] R. Sanitioso, Z. Kunda and G. T. Fong, "Motivated recruitment of autobiographical memories," *Journal of Personality and Social Psychology,* vol. 59, pp. 229-241, 1990.

[40] N. Mazar, O. Amir and D. Ariely, "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of Marketing Research,* vol. 45, pp. 633-644, 2008.

[41] E. Yoeli, M. Hoffman, D. G. Rand and M. A. Nowak, "Powering up with indirect reciprocity in a large-scale field experiment," *Proceedings of the National Academy of Sciences,* vol. 110, pp. 10424-10429, 2013.

[42] F. L. Greitzer, C. F. Noonan, I. J. Kangas and A. C. Dalton, "Identifying at-Risk Employees: A Behavioral Model for Predicting Potential Insider Threats," US Deparment of Energt, 2010.

[43] P. T. J. Costa and R. R. McCrae, "Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual," Psychological Assessment Resource, Odessa, 1992.

[44] P. K. Jonason, S. Slomski and J. Partyka, "The Dark Triad at work: How toxic employees get their way," *Personality and Individual Differences,* vol. 52, pp. 449-453.

[45] S. Ho and e. al., "Demystifying insider threat: Language-action cues in group dynamics," in *The Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)* , 2016, pp. 2729-2738.

[46] S. Sheng, M. Holbrook, P. Kumaraguru, L. Cranor and J. Downs, "Who Falls for Phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions," in *In the Proceedings of the 28th ACM Conference on Human Factors in Computing Systems,* , ACM, 2010.

[47] M. Keeney, E. Kowalski, D. Cappelli, A. Moore, S. T. and S. Rogers, "Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors," National Threat Assessment Center, Washington DC, 2005.

[48] R. S. Dalal and A. K. Gorab, "Insider threat in cyber security: What the organizational psychology literature on counterproductive work behavior can and cannot (yet) tell us.," in *S J Zaccaro, R S Dalal, L E Tetrick, & J A Steinke (Eds), in Series in Applied Psychology. Psychosocial Dynamics of Cyber Security*, Routledge/Taylor & Francis Group, 2016, p. 92–110.

[49] A. Furnham and E. M. Siegel, "Reactions to organizational injustice: Counter work behaviors and the insider threat," in *Justice and Conflicts*, Berlin, Springer, 2011, pp. 199-217 .

[50] C. M. A. L. M. &. P. C. L. Pearson, "Assessing and attacking workplace incivility," *Organizational Dynamics,* vol. 29, no. 2, pp. 123-137, 2000.

[51] L. Cortina and e. al., "Incivility in the workplace: Incidence and impact," *Journal of Occupational Health Psychology,* vol. 6, pp. 64-80, 2001.

[52] J. R. Schoenherr and K. Nguyen, "Multi-agent accumulator-based decision-making model of incivility (MADI).," in *In International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling an Prediction and Behavior Representation in Modeling and Simulation*, Springer, 2018, pp. 76-81.

[53] J. R. Schoenherr and K. Nguyen, "Modelling the Workplace Incivility with Prosocial and Antisocial Cues to Predict Psychological and Organizational Exit.," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, Springer, 2019.

[54] S. L. Robinson and R. J. Bennett, "A typology of deviant workplace behaviors: A multidimensional scaling study," *Academy of Management Journal,* vol. 38, pp. 555-572, 1995.

[55] A. P. &. R. T. S. Fiske, Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships, Cambridge University Press., 2014.

[56] B. S. Weinstein, "In Defense of Jeffrey Wigand: A First Amendment Challenge to the Enforcement of Employee Confidentiality Agreements Against Whistleblowers," *SCL Review,* vol. 49, pp. 129-, 1997.

[57] A. Schafer, "Biomedical conflicts of interest: a defence of the sequestration thesis-learning from the cases of Nancy Olivieri and David Healy," *Journal of Medical Ethics,* vol. 30, pp. 8-24, 2004.