Understanding Insider Threat: A Framework for Characterising Attacks

Jason R.C. Nurse†, Oliver Buckley†, Philip A. Legg†, Michael Goldsmith†, Sadie Creese†,
Gordon R.T. Wright§, Monica Whitty§
† Cyber Security Centre, Department of Computer Science, University of Oxford, UK
§ Department of Media and Communications, University of Leicester, UK
†{*firstname.lastname*}*@cs.ox.ac.uk*, §{*grw9, mw229*}*@leicester.ac.uk*

***Abstract*—The threat that insiders pose to businesses, institutions and governmental organisations continues to be of serious concern. Recent industry surveys and academic literature provide unequivocal evidence to support the significance of this threat and its prevalence. Despite this, however, there is still no unifying framework to fully characterise insider attacks and to facilitate an understanding of the problem, its many components and how they all fit together. In this paper, we focus on this challenge and put forward a grounded framework for understanding and reflecting on the threat that insiders pose. Specifically, we propose a novel conceptualisation that is heavily grounded in insider-threat case studies, existing literature and relevant psychological theory. The framework identifies several key elements within the problem space, concentrating not only on noteworthy events and indicators – technical and behavioural – of potential attacks, but also on attackers (e.g., the motivation behind malicious threats and the human factors related to unintentional ones), and on the range of attacks being witnessed. The real value of our framework is in its emphasis on bringing together and defining clearly the various aspects of insider threat, all based on real-world cases and pertinent literature. This can therefore act as a platform for general understanding of the threat, and also for reflection, modelling past attacks and looking for useful patterns.**

***Index Terms*—insider threat; threat framework; technical, psychological indicators; attack chain; case studies**

I. Introduction

Traditional notions of cybersecurity place an emphasis on protecting against attacks that arise from external threats. However, it is becoming increasingly apparent that the greater threat to an organisation's security may well lie within, as evidenced in many recent surveys (for example the 2012 Cybercrime survey [1] and the Risk of Insider Fraud study [2]). To underline this point, a recent study by Clearswift [3] reports that 58% of reported security incidents were as a result of insider threat. This point is further supported by a number of recent high-profile, highly publicised cases of data exfiltration and whistleblowing; for example Edward Snowden [4], Xiang Dong Yu [5] and Michael Woodford [6]. Of course, these reports probably only present a small percentage of the cases of insider-threat. It is widely accepted that there are a myriad of insider incidents that will go unreported (for fear of damage to the reputation of the company, for instance) or that will go unnoticed as the attacks simply avoid detection.

An insider can be thought of as an individual who is an employee (past or present), contractor or other trusted third party, who has privileged access to the networks, systems or data of an organisation [7]. In this paper we will consider two categories of insider threat. The first is a malicious insider threat, where the insider uses their privileged access to intentionally cause a negative impact to the confidentiality, integrity or availability of the organisations' information, systems or infrastructure [7]. It is typically understood that a malicious insider will seek to exploit their privileged access for some inappropriate gain, whether it be personal, financial or for revenge. The attempted attack by a Fannie Mae employee after being dismissed is a perfect example of an insider threat likely motivated by revenge [8].

The second form of threat is that of an accidental, or non-malicious, insider; this is actually reckoned to be the most common type of threat [9, 10]. Carnegie Mellon University's Computer Emergency Response Team (CMU-CERT) pioneers one of the most comprehensive research programs on insider threat, and they define the accidental threat as an insider who, without malicious intent and through action or inaction, causes harm or increases the probability of future harm to the confidentiality, integrity or availability of the organisation's assets or resources (e.g., information or systems) [11]. This therefore covers human mistakes, errors, and a barrage of other mishaps that may compromise the organisation, many arguably the fault of bad system design as much as negligence of insiders. For the purpose of characterising attacks, such compromises are included in the broad range of attacks that can potentially occur. Real-world examples of the unintentional threat include employees losing their work devices [12], accidentally leaking sensitive company information on social networks [13], and falling for phishing and other disguised malware attacks [14].

As the insider-threat problem has grown, so to has the attention it has received within the research community. There have been in-depth discourses on everything from what exactly an insider threat is [15] and what the range of human and psychological factors involved are [16, 17], to how threats can be predicted, detected and effectively addressed with appreciation of technological and behavioural advances and theories [17–21]. These approaches have resulted in numerous models and frameworks for insider threat, each with its distinct perspective on the problem and specific area which it aims to address. In spite of these advances in research and the various proposals, however, there is arguably still no unifying

DOI 10.1109/SPW.2014.38

framework which seeks to fully characterise the insider-threat problem space. That is, defining which insiders attack, why they attack, the human factors that lead to accidental threats, how one's background may impact likelihood of attack, what behaviour may be exhibited before or during an attack, what the common attack vectors and steps within an attack are, and what assets and vulnerabilities are typically targeted.

The focus of this paper is therefore to address this gap and present a framework for understanding and characterising insider threat that is grounded in real-world threat data and pertinent literature. We draw on insider-threat cases from CMU-CERT and the UK's Centre for the Protection of National Infrastructure (CPNI), broad survey data and existing research, and apply a grounded-theory approach [22] to deduce the framework. To evaluate its ability to capture and allow analysis of a variety of attack scenarios, we use an additional set of cases collected directly [23] within our broader research project. Overall, the most closely related work to what we propose is that by CMU-CERT (i.e., MERIT models for fraud, IP sabotage, etc. [7]). The distinguishing factor of our work is its broad nature and ability to capture all types of insider attack in a single comprehensive framework, while also remaining simple enough to facilitate understanding and discourse on what is, at times, a very complex problem.

We expect the framework to be useful to security practitioners and researchers alike. It provides a basis for elucidating the threat that practitioners' enterprises face and the important elements (e.g., precursors, indicators, attack types and steps) that are worth taking note of within the insider-attack chain; while for researchers, the framework supplies a well grounded conceptualisation of the insider-threat domain that can form the basis for understanding, and future research. A central part of this utility is the framework's capacity to retrospectively analyse documented insider-threat cases, particularly for purposes of identifying patterns of attack.

The remainder of this paper is structured as follows. Section II presents the scope of our work, methodology used to define the framework and the context in which this research should be viewed. Section III discusses the characterisation of insider attacks, with a focus on the threat framework developed, its components and the important relationships identified. In Section IV, we demonstrate how our framework can be used to capture and reflect upon several types of insider attacks, be they malicious or accidental. Section V engages in a critical discussion of the framework, first comparing it to related work, such as models by CMU-CERT, and then reflecting on other related concerns and challenges. Finally, we conclude the paper in Section VI, presenting avenues for future work.

II. Scope, methodology and context

Our framework for characterising insider attacks was born out of the need for a better approach by which the various components of the insider-threat problem could be easily understood and reflected upon. This objective served to guide our research and scope the creation of the framework. To build the framework itself, we adopted a grounded-theory approach. Originating in the field of sociology, grounded theory has become a popular research method through which new frameworks, models and theories can be developed, by a process of data-gathering, categorisation and coding, followed by various comparative and theoretical analyses of findings [22]. The objective of grounded theory is to collect data for analysis until saturation is reached in order to develop a new theory. It is important to note that in grounded theory data-collection and analysis are interrelated and analysis commences as soon as data starts being collected. To apply the approach to our work, we first collected a data-set of 80 insider-threat cases. There was no specific inclusion criteria, thereby allowing us to use cases from CMU-CERT [7, 24], the UK's CPNI [25] and various news articles (e.g., [5, 8, 26–28]). To further enrich the theory-development task, we also gathered as many relevant publications – academic or industry-based – as we could find (for instance, [2, 11, 17, 29–36]), and thus followed the more informed theory-generation approach [37] (considering data, developing theory and then comparing with previous research).

Starting with the cases, we assessed each one and noted emerging categories and themes (this was the categorisation and coding process). An example of a theme that arose in malicious-insider cases was '*Motivation to Attack*', which describes the reason why an individual might have attacked their organisation. As the themes were defined, we also took the opportunity to analyse relevant literature, both for additional themes and to better interpret the ones that we were finding. The next task was comparing and reflecting on the themes identified across the cases, first for consistency and then with the aim of identifying relationships between themes. The latter of these tasks was an iterative process that involved first hypothesising about each potential relationship, then validating that hypothesis with the cases, and also against the literature; we aimed for at least 70% agreement with the data to support validation. One relationship identified, for instance, was the strong influence that work and personal events (e.g., demotion or financial problems) had on an individual's personality state; a link further supported by psychological research [38].

Having identified a set of themes and relationships pertaining to insider threat, we then constructed a diagrammatic representation – this resulted in the first draft of the full framework. To allow evaluation and further analysis and refinement, we collected an additional set of 99 cases through direct means [23], within our broader research project. These were assessed and coded according to the procedure given above, and the themes and relationships arising were compared to the ones already identified in the framework. This proved to be a very successful exercise, as a majority of the concepts found were already present. One insight offered by the new cases, however, was a greater range of the potential values associated with the themes and framework components. For instance, we found several previously unrecorded events which could trigger an insider to attack (e.g., loyalty to friends or family, and cultural pressures), and we were also able to further detail the variety of psychological traits that may contribute to an

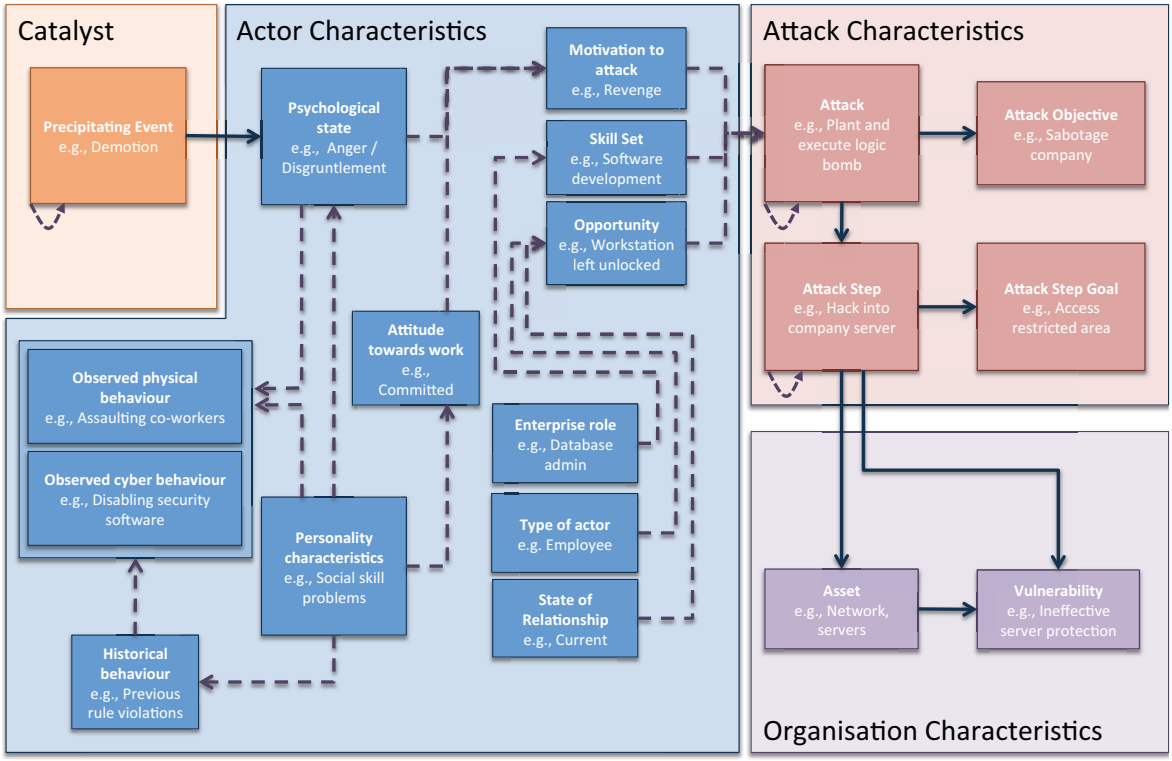

Fig. 1. A framework for characterising insider attacks

insider attack, such as overly impulsive behaviour [23].

A final point about our framework is that, as with all proposals developed using grounded-theory, it is grounded in the data that we assessed. Therefore, even though we sought to be thorough in our investigation and framework development, there may be aspects not yet represented, for instance, influences on threat currently undiscovered. Additionally, it is important that readers appreciate that our conceptualisation aims at defining and connecting the various main components of insider threat, but also the various bodies of knowledge (in Computer Science, Psychology, Organisational Behaviour) that are imperative in fully appreciating and studying the problem. This therefore seeks to define a central component of the foundation for future understanding and work within this space. At this stage, we are not aiming towards aspects such as how the framework could be directly used to detect or predict insider attacks. Moreover, the proposed framework does not address practical details on: (i) how live data could be gathered for framework components (e.g., *Personality Characteristics* or *Motivation to Attack*) as part of protective monitoring against attacks; (ii) how such data could meaningfully be measured in an enterprise context to predict likely attacks; nor (iii) what intensity of an element (e.g., desire for revenge) may be needed to push an insider to the next stage (e.g., to launch an attack). These are all very interesting and topical research problems, but are not within the scope of this current report; we do, however, for completeness, engage in some further brief discussion on them in Section V-B.

III. Characterising Insider Attacks

The framework presented in Figure 1 consists of several classes of component (or *Element*), depicted in four areas, namely, attack *Catalyst*, *Actor Characteristics* (i.e., those of a potential insider threat), *Attack Characteristics* and *Organisation Characteristics*. In the figure, boxes are used to represent specific elements, while solid arrows indicate a definite relationship between the elements and dashed lines potential relationships. To assist in our discussion below, we have further broken our consideration of these areas into the following sections: *Understanding the Propensity to Attack*; *Observing Behaviour of Trusted Personnel*; *The Actor / Insider*; *Dissecting the Attack*; and *Assets Under Attack and their Vulnerabilities*.

A. Understanding the Propensity to Attack

To explain the elements of the framework and their relationships, we begin with the behavioural and psychological aspects relating to the *Actor*; in many ways, these can be seen as the antecedents or key initial factors to understanding an individual's propensity to attack. The topic of an insider's psychology has received substantial research and practitioner emphasis over the last few years (e.g., [17, 18, 39]), after being somewhat overlooked in early enterprise-security work. Based on our research into these articles and the collected case data, we identify eight elements that may be especially

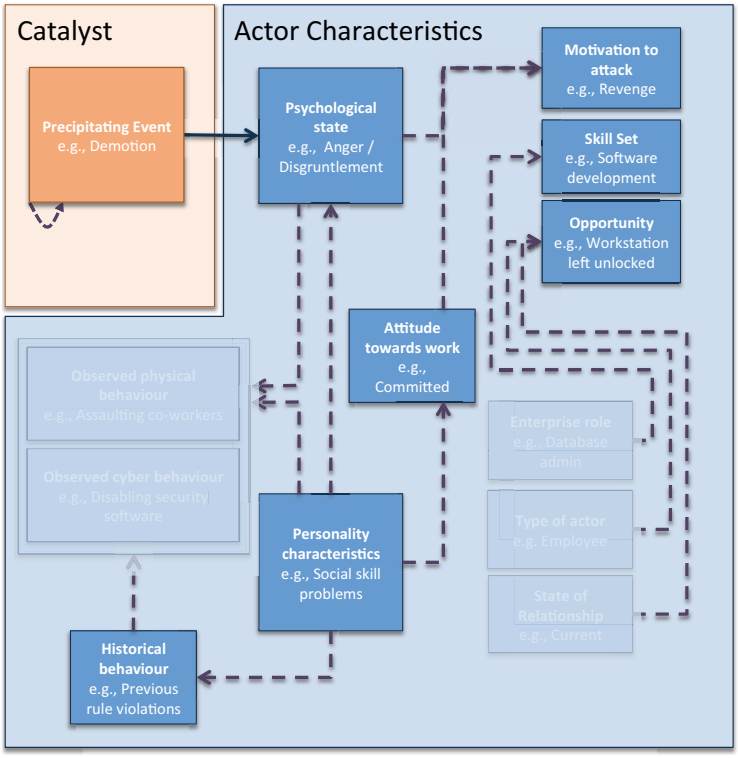

Fig. 2. 'Propensity to attack' elements

useful in modelling and analysing this aspect of insider threats. These are the *Precipitating Event* or catalyst, the individual's *Personality Characteristics*, *Historical Behaviour*, *Psychological State*, *Attitudes Towards Work*, *Skill Set*, *Opportunity* and lastly, *Motivation to Attack*. Below, we discuss these elements in further detail, and then consider the relationships between elements; this discursive approach is also adopted for the remainder of this Section.

The *Precipitating Event* is the key event or catalyst that has the potential to tip the insider over the edge into becoming a threat to their employer. This term was initially seen in the insider-threat literature in Moore *et al.* [40], and has also aptly been called the 'tipping point' [33]. Examples of such events include employee dismissal, disputes with employers (e.g., regarding IP rights), perceived injustices, negative company acts (e.g., lay-offs), family problems (divorce, child custody issues, health problems), coercion, new opportunities (e.g., job offer from a competing company), or even lack of training in the case of accidental attacks. Research in the field of Counterproductive Workplace Behaviour (CWB) [35, 41] was crucial to our definition and understanding of these events, and their general link to human behaviour and aggression at work.

A significant point that arose from our cases and CWB literature (e.g., [35]) is that a negative event need not have happened, as perception or rumours of something bad can be just as damaging. The case of the systems administrator that began constructing a logic bomb [7, p. 257]) based on rumours of lower bonuses is a perfect example. Moreover, being mindful of only employees' work-related activities might result in missing other events that could be the catalyst for attacks. Similarly, as accidental attacks become more detrimental to the enterprise [42], there will be a growing need to understand better how they come about, and the related tipping points. Generally, however, this *Precipitating Event* element reflects the need for better appreciation of the range of events that could set an insider along the path to an attack.

Personality Characteristics, including psychological traits and dispositions, capture the features of an *Actor's* personality based both on their innate self (thus, the static aspects) and their life experiences (therefore, the more responsive and dynamic aspects). General personality traits can include their OCEAN (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism) [43], the Dark Triad (Machiavellianism, Narcissism and Psychopathy) [44], and Sensation-Seeking [41]. Other Characteristics may include: maturity, aggressiveness, social-skill problems, superficiality, (lack of) self-esteem and personal integrity [17, 45].

As they pertain to insider threat, *Personality Characteristics* are central to how we as humans think and act, and therefore have a strong influence on whether or not an individual is likely to get involved in malicious activities or threat-enhancing behaviour at work (be it intentional or unintentional). Our cases highlighted the importance of this, especially the impact that *Personality Characteristics* may have on future

actions. We were able to identify *Personality Characteristics* as being a factor in many of the cases, and validated the associated relationship against behaviours using literature. Personality traits such as Machiavellianism, excitement-seeking, and Narcissism were found to relate to insider threats and antisocial behaviour [17,41,46]. Likewise, from an accidental-threat perspective, some OCEAN (especially, agreeableness and openness) traits can relate closely with an individual's susceptibility to scams [47].

It is important to understand, however, that although some *Personality Characteristics* are worthy of note, they are not sufficient in isolation to identify a potential attacker rather, we need to assess clusters of *Personality Characteristics* together with catalysts (*Precipitating Events*) and even the individual's environment. For example, an employee who is highly Narcissistic, in some cases might be the perfect choice for a particular role in an organisation; however, in combination with a stressful event (e.g., being over-looked for a promotion) or opportunity (e.g., being offered a new job elsewhere with better benefits), this may lead to a specific psychological state which then results in an increased risk being posed to the organisation. Aldrich Ames, for instance, reportedly suffered from a narcissistic personality disorder which lead him to "believe he was bulletproof" [48], a likely factor in his espionage. As mentioned above, clusters of *Personality Characteristics* might also be useful to consider. For example, an individual who scores high on all three personality traits that constitute the Dark Triad might, theoretically, pose a greater threat than those who do not.

Historical Behaviour documents the kinds of activities the Actor has engaged in during the past and, as with most behaviour, is likely to be influenced by their Personality Characteristics. There are, of course, an infinite range of behaviours, but from a malicious, insider-threat perspective, examples of notable behaviours include: addictive practices (e.g., gambling or alcohol abuse), previous rule violations (e.g., harassment or company policy violations), criminal history, or a history of serious mental problems. Of course, when considered in isolation, this again is not alone indicative of an individual becoming an insider threat. For unintentional threats, behaviours discovered to be relevant were typically the result of human error (e.g. carelessness or absent-mindedness) [49].

Overall, many of the above are linked to the person's personality. A CMU-CERT case [7, p. 257] illustrates this element's importance, though, when a system administrator with a history of electronic crimes used similar malicious techniques to attack his employer, first using blackmail and then sabotage. This suggests that past actions have some influence on future actions – a point further supported by CWB literature [50]. There is the argument that if proper checks had been conducted beforehand then appropriate measures could have been taken. Yet, to actually conduct such checks may be challenging given the mobility of today's workforce, and companies may not have the resources available to perform extensive investigations. Either way, behaviour remains a significant factor in understanding (if not necessarily diagnosing) insider attacks.

Psychological State represents the *Actor's* psychological and emotional state (e.g., happy, depressed, stressed or anxious). This might be, in part, due to their psychological make-up (e.g., some individuals have clinical depression), or as a result of their environment (e.g., a stressful event, such as forced job transfer, which leads to depression [38]), which therefore explains the relationships with *Personality Characteristics* and *Precipitating Event* respectively. If the environment causes the state it might arise from an event outside the workplace, or from within it. In the insider-threat literature, it is commonly a disgruntled employee responsible for attacks; but disgruntlement is only one of a set of states found to be a compelling precursor in our case research. Others include stress, fear (e.g., of dismissal or group exclusion), a lack of appreciation, a feeling of entitlement (to customer contact data for instance), feeling opportunistic, or, from an accidental threat perspective, carelessness, boredom or dissatisfaction.

Further insight was also attained from the research on CWB, where several other *Psychological States* of concern can be found which have (validated) relationships with counterproductive behaviours such as perceptions of organisational injustice and workplace inequalities, or revenge cognitions [35]. An interesting point here is that like the *Precipitating Event*, the *Actor's* state can be influenced by opportunities that happen to present themselves. For example, in one case [7, p. 272] the system administrator happened upon several unprotected files on an FTP server, an opportunity he found too great to pass up, as he later cracked the encryption passwords in the files, thereby gaining unauthorised access to an extensive amount of customer data. Although undocumented, it is plausible that certain traits or experiences in his background (e.g., impulsive behaviour) in conjunction with the Event provoked him to seize the opportunity and act maliciously.

Motivation to Attack captures the reason that an *Actor* might desire to attack the enterprise. The notion of attack motivation is well understood in the threat-assessment field and therefore, in addition to case reflection, we drew heavily on existing work [34] for our general categories. Namely, motivations can be: financial, political, for revenge, curiosity or fun, power, competitive advantage, or peer recognition. Based on our case findings, we posit that an *Actor's* current *Psychological State* would be a significant influence on their *Motivation to Attack*. For instance, a contractor's disgruntlement because of mistreatment might give rise to some desires for revenge; or, fear of being excluded from the office 'in crowd' or loyalty to friends/family/country [25] may motivate an employee to engage in an attack (not dissimilar to the situation with theft in Greenberg and Barling [51]). This general State-to-Motivation relationship was also apparent when reflecting on the related CWB literature, insofar as it relates to anger, frustration and various perceptions (e.g., of unfairness or injustice) [35].

Interestingly, *Psychological State* can also be coupled with *Attitude Towards Work* in its influence on *Motivation*. For example, if an employee was constantly overlooked for a

promotion, therefore possibly feeling aggrieved (Psychological State), he might feel that his strong commitment to his employer (Attitude) might have been misplaced, and hence become highly motivated to attack. This idea, albeit anecdotal at this stage, ties in closely with existing work on the individual relationships, from State to Motivation [35] and Attitude to Motivation [36]. For this *Motivation* element, we also consider the possibility of accidental insider attacks. To describe these, we maintain two overarching classifications of motivation, one 'deliberate' (capturing the categories above) and one 'accidental' (to accommodate human factors, mistakes, etc.). This ensures that our framework is inclusive enough to capture all types of threats.

Skill-Set captures the *Actor's* capability or the requisite skills needed to conduct an attack [34]. Arguably it is reasonable to make the link between the role that an *Actor* carries out, within an enterprise, and the *Skill-Set* that they possess; although there is certainly scope for a skill set outside of a job role. This is highlighted by a CMU-CERT case [7, p. 253]: a software developer at an organisation who was angry at the lack of company bonuses inserted malicious code into the enterprise's premier product, an inter-network communication interface. In this instance it was the employee's software-development background and associated development skill that allowed him to initiate and carry out the attack.

Opportunity captures the *Actor's* chance to initiate an attack on the enterprise. The notion of the opportunity to attack is well defined within related literature relating to threat-assessment and risk management. Clearly, in order for an *Actor* to carry out a malicious attack on an organisation, they will need an opportunity to initiate the attack [34]. For example, CMU-CERT present the case [7, p. 266] of an insider, employed by the police, to communicate information about drivers' licenses. The insider was then recruited by a third-party to provide license information, and later progressed to creating fraudulent licenses. The opportunity in this case was the insider's authorised access to the license database; without this access, the attack could not have taken place.

The progression from *Psychological State* to *Motivation to Attack* is one stage in particular where employers (via the Human Resources (HR) department or security practitioners) could step in and address potentially negative states (e.g., observed feelings of anger, dissatisfaction, or irresponsibility) before the individual becomes motivated to attack. This could mean HR meeting with the employee to discuss and correct any misunderstandings in a more relaxed manner, offering additional support if an employee is going through personal problems, or from a more technical perspective, changing policies, or implementing new security measures meant to deter thoughts of attack.

B. Observing behaviour of trusted personnel

Two areas that are key to understanding and modelling insider attacks are observations about an *Actor's* Physical and Cyber behaviour. *Observed Physical Behaviour* captures the actual physical behaviour that may have been exhibited by an Actor, for example in the office, with colleagues or accessing buildings and resources. *Observed Cyber Behaviour* places the focus on the technologically-related behaviour that an Actor may exhibit over the enterprise information infrastructure, such as the usage of Internet, e-mail, and workstations.Both of the observed sets of behaviour may be indicative of an attack either currently being conducted, or soon to be conducted. Moore *et al.* [40] propose similar aspects in their analysis of insider threats, explicitly naming them behavioural and technical precursors. Many of these can also be seen in work by the FBI [27]. In our research, we are compiling (as far as possible) a comprehensive list of behaviours that will be of interest when trying to understand insider threats.

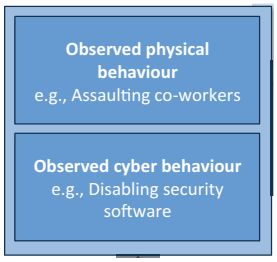

Fig. 3. Behaviour elements

For the physical domain, examples of potentially concerning behaviours include: assaulting or intimidating co-workers, clients or business partners, expressing a negative attitude towards the company, violating company policy, and poor job performance. On the other hand, concerning cyber behaviour includes: violating technology usage policies, attempts to gain access to data of systems beyond their job's responsibilities, and deactivating security tools. In terms of relationships, all these behaviours may be influenced by *Personality Characteristics* and *Historical Behaviour*. In the first case, the type of person someone is can undoubtedly influence their actions, while in the second case, although not inevitable, previous behaviour does often reoccur, potentially because of the link to personality.

It is important to remember that observing precursor behaviours in isolation may not necessarily be indicative of an insider attack. Whilst some behaviours may clearly be directly-related to an attack (e.g., breaking into a safe), other behaviours may require a more holistic view of the inter-play with related factors from our framework to establish whether an attack is present.

C. The actor

In Section I we defined the concept of an insider within an organisation. The *Actor* element within our framework is used to define a number of generic types of individual that could be considered as part of an insider attack (using the definition provided in Section I). The types of individual we have identified are: employee, contractor or consultant, client or customer, joint venture partner, vendor and external attacker; most of these individuals represent positions held by

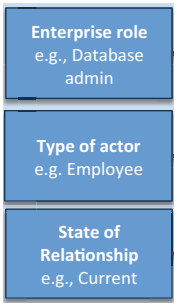

Fig. 4. Actor elements

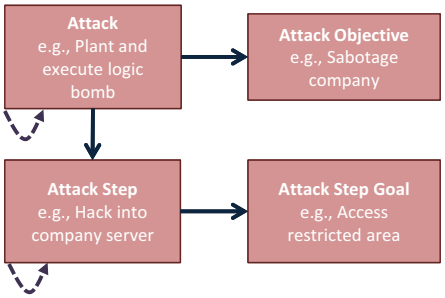

Fig. 5. Attack elements

trusted personnel, with the exception of an external attacker. The external attacker is included in this list because of the fact that individuals that are external to an organisation may recruit and collaborate with any of the trusted personnel to assist in conducting an attack against the enterprise. This collaboration could be voluntary (e.g. mutual desire for financial gain), coerced (e.g. blackmail) or unknowing (e.g. as a result of a phishing or social-engineering attack).

We make the assumption that an *Actor* is innocent until proven otherwise, and so there may or may not be a *Motivation for Attack*. This list of types of individuals who can be considered insiders is by no means novel, but it should serve to stress the fact that employees are not a company's only concern. There have been several documented attacks (e.g., CMU-CERT [7, pp. 248, 269, 271]) originating from a variety of trusted third parties, sometimes with much higher privileged access than the average employee.

The *State of Relationship* and *Enterprise Role* are directly related to the *Actor*. The first identifies the current state of the relationship between the enterprise and the *Actor*, while the second captures the role that the *Actor* may have in the enterprise. There are four states of relationship which we distinguish: current, former, serving notice and temporary. Although people in any of these states may be dangerous to an organisation, cases have shown that individuals in the last three states may be especially risky. In several cases for example, Actors serving notice, especially those that work with sensitive information, are of significant concern because they may attempt to take it with them to boost their value to a competing organisation. The types of *Enterprise Role* vary considerably but knowledge of an *Actor's* role is useful as certain roles have been shown to tend towards specific attacks, with set attack objectives in mind. According to Cappelli *et al.* [7], for instance, scientists, engineers, programmers and salespeople are typically the roles that steal IP, and usually for the purposes of setting up their own business, carrying to a new job, or giving to foreign organisations or governments.

D. Dissecting the attack

An *Attack* represents an activity that is conducted by an *Actor*, either deliberately or accidentally, that will have a negative impact on the enterprise. The *Attack* will typically have a result associated with it, i.e. an *Attack Objective*. An example of an *Attack* is a former IT administrator planting and executing a logic bomb on his previous employer's network; while the objective, in this case, may be sabotage. There are a large array of possible attacks, especially when considering the disparate areas of sabotage, IP theft and fraud, and therefore we will not seek to enumerate them in this article; these will, however, be available at some point on our website [52].

Attack Objectives are much more constrained, and tend to consist of: financial gain, personal gain, competitive gain, political gain and damage to the organisation. *Attack Objectives*, as one would expect, are usually closely tied with the motivations behind the threat. For instance, desire for revenge usually leads to an attack seeking personal gain or to damage the company. In the case where an attack is accidental or unintentional, we consider the Objective to be task or activity that is the reason for the incident. Therefore, having to complete a task at work under strict time constraints could be the Objective that was the reason for an employee taking a USB stick with sensitive data home, which was later lost during their commute. By modelling accidental threats in this way, we are able to gain further insight into the reasons linked to attacks, and thus facilitate better understanding.

While *Attacks* aim to be generic, *Attack Steps* define, in detail, the specific activities undertaken to conduct the attack. As such, an *Attack* can be composed of several chained Steps. To steal sensitive IP, an insider threat may: (i) determine which of their colleagues has the credentials to access the desired IP (reconnaissance); (ii) coerce those individuals - possibly via financial means, charm or physical threats - to assist in the task; (iii) use the ill-gotten credentials to access the IP; (iv) download the IP to portable media; and (v) delete the related log files. These *Attack Steps* can also be thought of in terms of the value gained from each step, namely, the *Attack Step Goal*. Thus, for the steps above, we would have: (i) gathering intelligence; (ii) recruiting accomplices; (iii) accessing restricted data; (iv) exfiltration of a volume of data; and finally (v) covering tracks. We believe that these goals can be particularly helpful when discussing the insider-threat problem with top management, and effectively communicating the attack, inclusive of what would be gained through each step, without going into excessive technical detail.

Attacks Steps share some similarity with the pre-existing

notion of Attack Trees [53], in that both methods can describe how a particular target or asset might be attacked. The value of our *Attack Steps* is that they allow for the clear sequencing and ordering of actions. We envisage that this added value with Steps is particularly useful when it comes to applying our framework to understand and assess an insider attack.

The idea of Intrusion Kill Chains [54] has particular relevance when considering our Attack Steps. These Kill Chains provide a means of describing the different phases of an intrusion and are modelled as a chain to emphasise the idea that if there is a breakdown at any one stage then the entire process is disrupted. Intrusion Kill Chains are designed to model attacks, with the aim of highlighting patterns within individual intrusions and how they may fit into part of a larger threat. It is easy to imagine how a similar aim could be achieved using our Attack Steps; when enough attacks have been collected and modelled then they could be used to establish common Attack Steps. The concept of building a library of Attack Steps is similar to the idea of Common Attack Pattern Enumeration and Classification (CAPEC) [55]. Attacks are recorded there in a similar fashion to our Attack Steps, and then CAPEC is used to identify opportunities for increasing the 'robustness and defendability' of software.

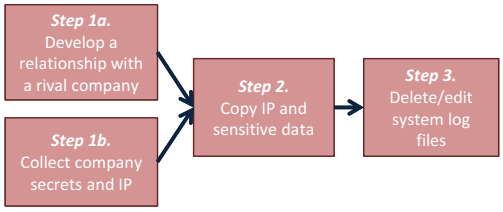

Fig. 6. Attack Steps

Figure 6 shows an example sequence of Attack Steps within an Attack, and highlights the strong sense of ordering within the attack. All steps at the same level in an Attack Tree are essentially in parallel, and so do not preserve any ordering. Figure 6 highlights an idea of concurrency in our Attack Steps, as the initial stage of the attack sees the Actor both developing a relationship with a rival company while at the same time gathering company secrets and IP in tandem. This example of Attack Steps can than be followed through, sequentially, to the IP being copied and finally the Actor editing and deleting log files in order to cover up the evidence of their attack.

E. Assets under attack and their vulnerabilities

The last two classes of element are *Assets*, items of value to the enterprise and of interest to the threat (e.g., company data, hardware, and personnel), and *Vulnerabilities*, weaknesses in assets or controls protecting them (e.g., weak passwords on administrative accounts, unpatched Web servers, and inadequate building security).

These are well-understood and frequently modelled aspects, and therefore are not covered in any detail here; Jones and Ashenden [34] offer further insight. In the framework, we

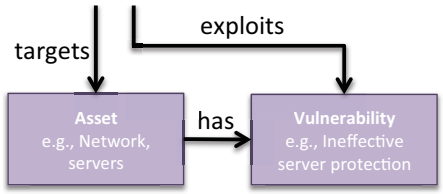

Fig. 7. 'Assets and vulnerability' elements

link these elements to Attack Steps instead of to the Attack, in order to allow a more detailed definition of the assets and vulnerabilities associated with each attack step. The advantage for a security practitioner with this configuration is the ability to see exactly what assets may be or have been targeted by each step of an attack, and the respective vulnerabilities that could be or were exploited. With the theft of IP example above, assets targeted were personnel, data, access credentials and log files, while vulnerabilities included a lack of security-awareness training and employee support, failure to monitor or block portable-media downloads, and inadequate protection of sensitive files. If conducting an investigation into a line of potential insider attacks after the fact, the way we define the attack steps could also allow an analyst to spot patterns, for example, certain assets usually being targeted or vulnerabilities exploited en route to more comprehensive attacks.

The various elements and relationships described in this section allows us to bring together precipitating events, individual's personality traits (and predispositions), behaviours (both historic and current), enterprise states and roles, attacks (and their detailed steps and goals) and targeted assets. With this framework, we can begin to characterise insider attack as we illustrate farther below.

IV. Using the Framework to Capture Attacks

To demonstrate the framework's use in assessing attacks, we apply it to three new cases. The first is based on a CMU-CERT case [56, p. 20] involving fraud, the second is based on a case of accidental leakage [57], and the third case is a scenario that we have devised. The aim in the last case is illustrate the framework's ability to handle concerns and threats resulting from human-factor-related issues associated with both poor systems setup and human error.

A. Case 1 - Tax office manager engaged in fraud

A tax office employed the insider as a manager. The insider had detailed knowledge of the organisation's systems and helped design the organisation's newly implemented computer system. The insider convinced management that her department's activities should be processed outside of this new system. All records for the insider's department were maintained manually, on paper, and were easily manipulated. Over 18 years, she issued more than 200 fraudulent checks, totalling millions of dollars. The insider had at least nine accomplices, insiders and outsiders, with unspecified roles in the scheme. The incident was detected when a bank teller

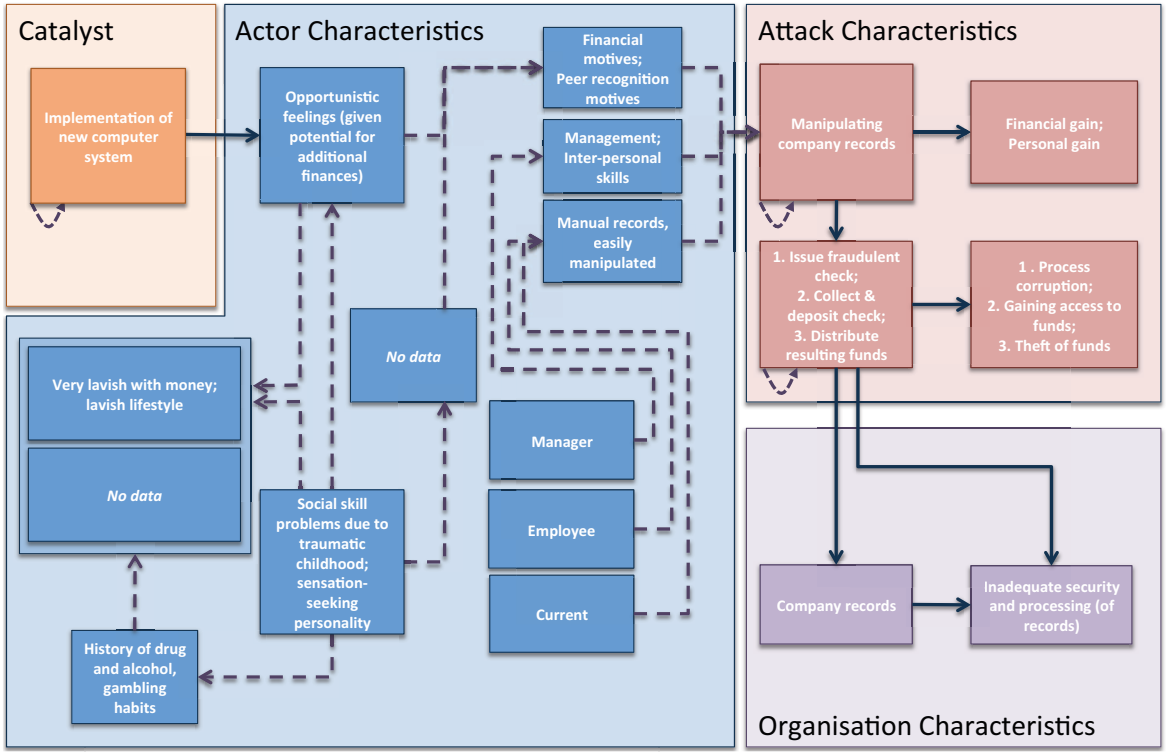

Fig. 8. Applying the framework to Case 1

reported a suspicious check for more than $400,000. The insider was arrested, convicted, and ordered to pay $48 million in restitution, $12 million in federal taxes, and $3.2 million in state taxes. One of the insider's motivations was that she enjoyed acting as a benefactor, giving co-workers money for things like private school tuition, funerals, and clothing; this was in addition to her own lavish lifestyle with luxury cars and designer clothing. The insider avoided suspicion by telling her co-workers that she had received a substantial family inheritance. The insider apparently endured a traumatic childhood, leading her to abuse drugs and alcohol and develop a substantial gambling habit.

Figure 8 shows this case mapped against the framework; we use *'No data'* where we have no further information on an element. The framework allows a security practitioner or researcher to clearly visualise the situation surrounding the insider and the attack conducted. Here, for instance, we see the catalyst for the attack (implementation of a new system), and the insider's *Psychological State* as influenced by the catalyst and their *Personality Characteristics*. We also are able to model the person's historical actions relating to drugs, alcohol and gambling, and the affect that had on their currently observed behaviour. Most importantly, the framework links three key elements relating to the actor that ultimately led her to the attack, i.e., her financial needs, her management and inter-personal skills (these allowed her to convince upper management that her department should be treated differently) and the opportunity present, in use of easily manipulated manual records.

This case shows one example of how our framework can be used to map a case of malicious insider threat, and more specifically, Data Fraud. The framework can also be applied to other malicious cases including IT Sabotage and IP Theft. For instance, if we consider a case of IT Sabotage where an employee (*Actor*) has deleted critical system files shortly after being made redundant, this would map well to our framework. This could be seen in the precipitating event (the employee being made redundant), the relationship status (the employee is serving notice) and possibly their emotional state (angry, dissatisfaction or anxiety). In addition to this, there may or may not be changes to the employee's physical behaviour. One might imagine, for example, that as he is angry or anxious, that he lashes out or snaps at co-workers. From the technical perspective, there is also a reasonable chance that there was a change to their cyber behaviour. In the case of deleting essential system files then it might well be the case that this was an area of the system that was not normally used by the employee. This discussion just covers a few of the ways in which general malicious attacks can be modelled.

B. Case 2 - Booking clerk accidentally leaking sensitive data

A booking clerk at a prison became an 'accidental insider' by unintentionally pasting the sensitive details of over 1000 inmates into an email, in response to a visitation booking request. The insider was new to the role and did not have a

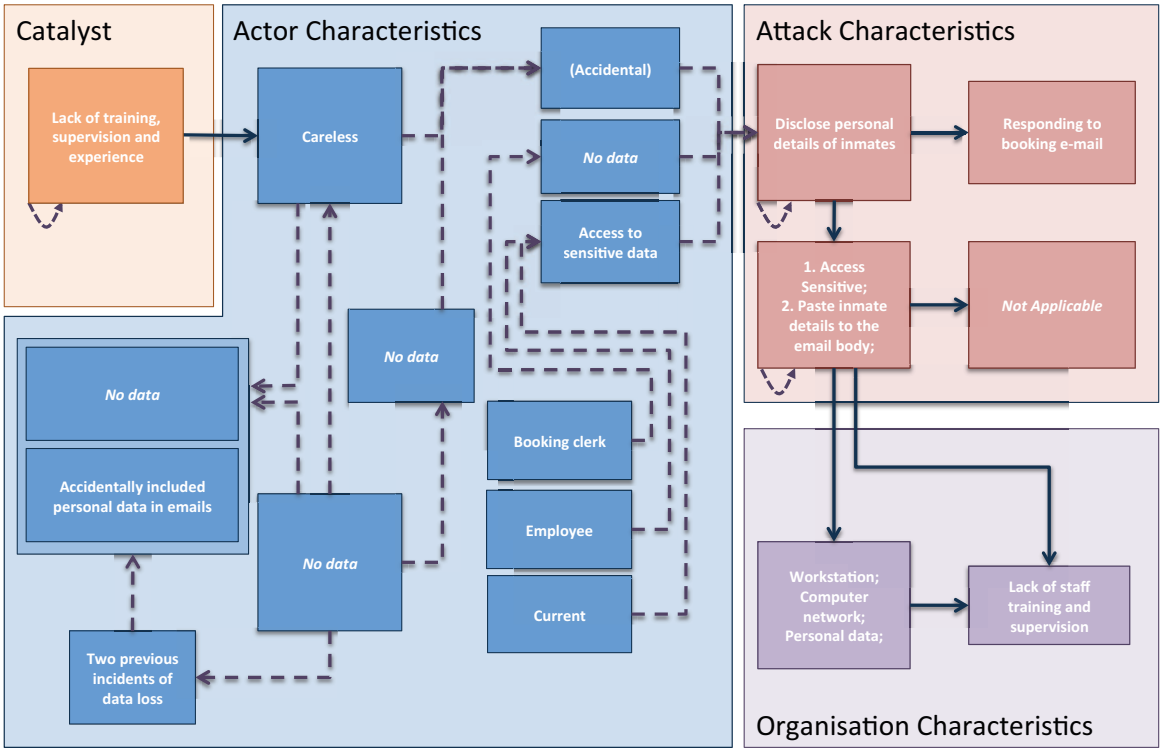

Fig. 9. Applying the framework to Case 2

sufficient level of experience, training or supervision. Prisoner data is stored on a network database, called Quantum; this is a secure network system that meets the UK Government's IT standards, and to which access is highly controlled. There is a separate biometrics system, which is not networked, that is used to process and book visits to the prison. The two systems are entirely separate and the existing process to update the biometrics system was to perform a 'profile dump' of inmate details, essentially copying and pasting a file, via the Windows Explorer, to an unencrypted removable disk. The clerk accidentally pasted this sensitive personal data into an email and sent it to the family of an inmate, who had requested a visit. This was not the first time that this had happened; in fact, it is reported that the same employee had made the same mistake twice before. The Ministry of Justice was fined £140,000 by the Information Commissioner's Office as a result of the breach.

In Figure 9, we show how this case of an unintentional attack can be modelled; we use 'Not Applicable' for elements that typically do not apply in accidental cases. The first point of note is that several of the elements in the framework are unpopulated, and this can to some extent be expected. The reason behind this is that as it is an accidental threat, the 'attack' is likely to be very sudden (in this case, a momentary lapse in concentration or judgement) and as such, there are probably behaviours to be concerned about (physical or cyber) or notable characteristics that are immediately relevant. Moreover, there is arguably no *Attack Objective* or *Attack Step Goal* as these assume malicious intent. On the other hand, accidental cases can often offer some useful information – this case for instance, highlights the carelessnesses of the individual in terms of their current state and historical behaviour and actions. For security practitioners using this framework, the catalyst is also of interest, as this hints to the real problem at hand.

The above case serves to illustrate the usefulness of the framework, in that this is an incident of data loss that had happened on two previous occasions, in very similar circumstances and with the same employee involved. Using the framework to reflect on the two previous incidents might well have highlighted the employee's lack of training or need for supervision, or it might have served to highlight flaws in the day-to-day procedures when transferring sensitive data.

The framework is capable of mapping a number of other accidental-insider scenarios: for example, if we think of the fairly common situation of a USB drive being lost (or forgotten) in a public place. There might be no changes to the employee's physical behaviour, and they may not have any prior history of data loss, but there are still factors that our framework would be able to capture. In this case it could be that there were external stresses in the employee's personal life that have had an impact on their psychological state, this in turn could have caused them to forget to pick up all of their belongings after a long train journey. These are some of the

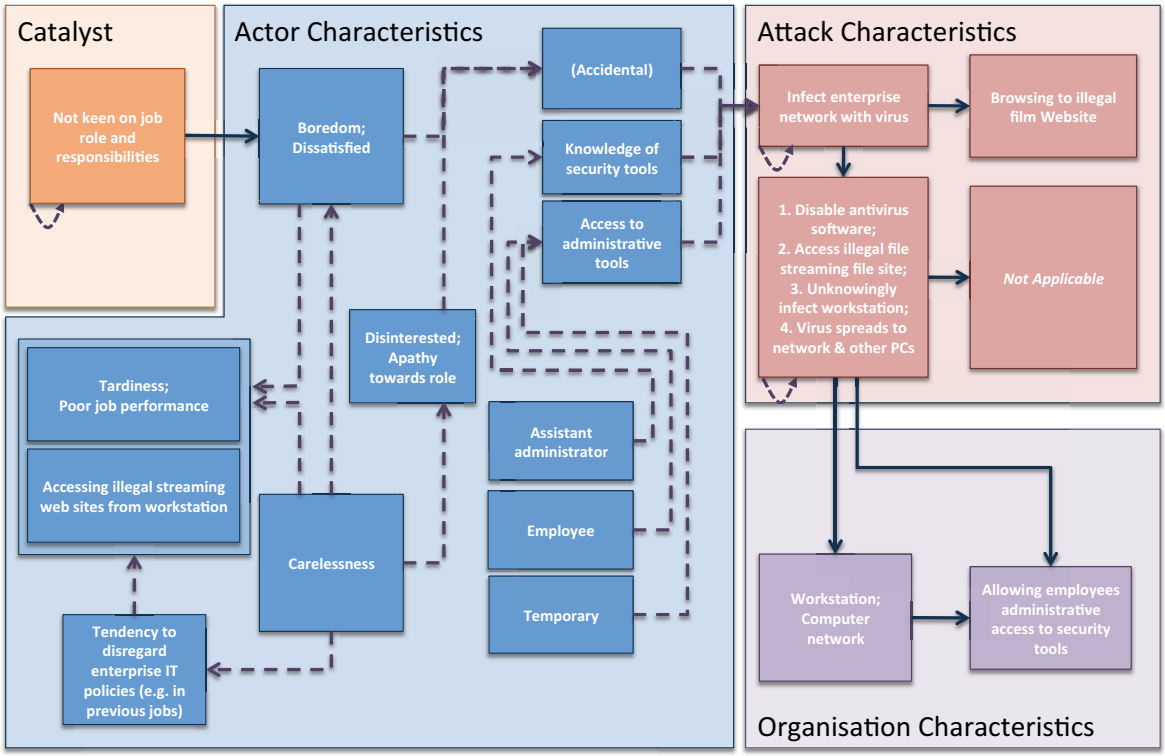

Fig. 10. Applying the framework to Case 3

ways in which we could look to model and further understand accidental insider attacks.

C. Case 3 - Inappropriate browsing resulting in attack

The third 'attack' that we cover is a case that we have created based on a combination of cases assessed. The attack that we are modelling is an accidental-insider case, which also contains elements of suspicious behaviour. This serves to highlight that our framework is capable of mapping attacks that span several categories.

As temporary cover for an administrator assistant role, a medium-sized logistics company hired a highly qualified graduate. During his interview, the HR manager was concerned about his interest in the role, but because his father was a friend of the director, her concerns were overruled. Initially the insider was punctual to work and appeared keen on the job. From week three, however, this changed. Co-workers observed a decline in the quality of his work, an increasing number of late arrivals, and a general apathy to the job and its responsibilities. His immediate supervisor noticed his carelessness and the low-quality work, but as only two weeks remained on his contract, he largely ignored them. After lunch one day, the insider was bored and instead of working, browsed to an illegal film streaming website; he did this a few times a week and avoided the company's block on the sites by disabling his workstation's security software. On this occasion, he was the victim of a 'drive-by download' which, unbeknown to him, infected his workstation with a virus. The virus then quickly spread to the enterprise network and from there to other workstations, eventually causing a full shutdown of the network and a three-day pause in logistical operations. An estimated $600,000 in damage and lost profit was caused by this inadvertent attack on company systems. An investigation into the event found that the insider had broken IT policies in his previous employment, and he was perceived to have a generally careless attitude. The case is modelled in Figure 10.

As mentioned in the section above, modelling accidental 'attack' incidents is difficult for many reasons, most notably the potentially sudden nature of the attack and likely lack of precursors, although there will be repeat (and so to some extent predictable) offenders. The case here is one where the insider has displayed a number of concerning indicators, such as tardiness, apathy towards roles, and accessing illegal sites, that should have been picked up on by his employer. Failure to recognise and deal with these aspects, coupled with his possession of the right skill-set and the opportunity, led to the unfortunate accidental attack on the enterprise and substantial loss. The benefit of the framework to this case would be identifying the key elements along the attack path and the behaviours that should have concerned security practitioners and HR. Assuming we had an attack-pattern database to compare against, these behaviours might have highlighted the increased possibility of an attack. Of course, there is also the fact that some of these aspects by themselves should have caused concern, as they are a direct breach of company policy.

V. Discussion

Thus far, we have presented a framework for characterising insider attacks, and demonstrated how real-world cases can be mapped on to it to facilitate deeper understanding. This is particularly through the linking of various critical elements that play a part in an insider attack. Whilst we believe that this framework provides a rich foundation for the analysis of insider threats, we also acknowledge that there are many other proposed frameworks and models relating to this threat. In this section, we reflect on the most relevant contributions from the literature, and critically compare our framework against these. A point to note here is that we do not cover proposals dedicated exclusively towards the detection of insider threats, as this is not our aim here and is therefore out of the scope of this paper. After reflecting on our work as compared to other's proposals, we consider the challenges to the use of the framework within organisations today and in the future.

A. Reflection against related works

The CERT project conducted by CMU [7] is without question one of the most comprehensive contributions to the field. As part of their research, CERT have proposed a series of MERIT (Management and Education of the Risk of Insider Threat) models using System Dynamics, to describe different types of attack (IT sabotage, IP theft, data fraud). As part of our discussion, we shall focus on the approach they have adopted, and consider how this very relevant work compares against our method and framework.

System Dynamics provide a mechanism for simulating complex environments, through the use of positive or negatively-reinforced causal loops. For instance, as an employee becomes disgruntled, the more likely that they will under-perform, which in turn will result in more disciplinary action, that will mean that the employee becomes more disgruntled. By their very nature as a tool for modelling these intricate environments, these conceptualisations can quickly become quite difficult to understand as the model becomes overly-large [58]. In addition, since System Dynamics models tend to describe a cyclic process, there is no immediately obvious starting point for mapping knowledge to a model, which could even act as an initial challenge for practitioners.

Compared to the CERT models, our framework was deliberately designed to strike a balance between ease of use and extensive coverage of the fundamental components of the problem. We also wanted to preserve the natural chain of events that typically surround insider attacks, in that, the catalyst triggers the individual to, at some point in the future, conduct an attack against the organisation. Again, our main aim is to facilitate an improved understanding between the various components of the insider-threat problem by clearly identifying the relationship that exists between the different attributes. In doing this, we aim to provide a pragmatic tool that could potentially be adopted by any organisation, regardless of their technical capability. As will be discussed later, as part of our future work we will be working alongside security practitioners to obtain feedback on the framework, and on its perceived utility and coverage. Whilst the MERIT models certainly provide a comprehensive view, there exist attributes that may well be difficult for an organisation to know, even during a post-mortem examination of the attack: for instance, detailed access paths unknown to the organisation.

What makes System Dynamics particularly powerful as a modelling tool is its ability to easily translate a model into a mathematical simulation, and there exist software tools such as Vensim (`http://vensim.com`) to achieve this. This also poses another significant challenge, however, namely how to accurately quantify attributes, and how to quantify the impact that one attribute may have on another. Whilst other modelling tasks may have well-quantifiable attributes, this is clearly not the case with insider-threat assessment, since most of the problem is centred around the mind-set of the individual. In our framework, we purposely do not aim to quantify attributes, as this is not the intended use of it at this stage.

Another notable difference in the approach taken in our work compared to that of the MERIT models is that each MERIT model is specifically designed for particular form of attack. In contrast, our framework is designed to characterise any form of insider-based attack. A key benefit therefore, is that our framework allows a variety of attacks to be assessed using the same basis, and moreover, it is broad enough to cope with attacks that merge different types of attack. MERIT models also seem to concentrate on the problem from the viewpoint of the organisation. Here, we have taken an insider-centric approach to developing our framework since the insider is going to be at the very core of any insider-threat incident. Each model contains attributes that do not feature in the other; for example, CERT address the organisation's trust of the insider, whereas we address attributes such as the insider's attitude towards work. Due to the differences in how our framework is positioned, it is indeed quite possible that the models could complement each other.

In other works, Pfleeger *et al.* [39] present a framework for describing insider threats and their actions, which is based upon four key attributes: the organisation, the environment, the system and the individual. The framework is designed to allow an analyst to question how attributes interact, for example, 'Do the actions of the individual violate de jure or de facto policy?', or 'W hat was the intent of the action?'. By formulating answers to such questions, this can be used to classify different threats by how they relate to the defined attributes. This methodology provides a useful platform on which to base an initial investigation. However, as the authors acknowledge, their framework lacks the full scope of detail, such as attributes that focus on the insider's perspective, and events that lead up to the attack. Instead, their model provides a much higher-level overview of the problem space. In much the same fashion, our framework could be considered at an overview level to assess the catalyst (environment), the actor (individual), the attack (system), and the organisation. Our framework then allows one to delve much deeper into formulating understanding surrounding each of these core components, by defining the individual attributes that make up each component.

Sarkar [30] discusses the factors that contribute towards the creation of an insider threat, and identifies capability, motivation and opportunity as the three key attributes that if an insider threat possesses then they have a high potential to attack. These three attributes are also present in our framework, and in addition, we illustrate how an insider may come to possess each of these through the linked relationship with other elements. For instance, if it is known that the individual is highly neurotic (a personality characteristic), are they more likely to react badly to the news of a demotion, and therefore be motivated to attack? Since our framework shows the link between elements, establishing whether such relationships exist may become much clearer.

B. Challenges for the framework

We have designed the framework to facilitate the understanding and consideration of the factors associated with an insider threat and the execution of an attack. By characterising previously executed or publicly documented attacks with our framework, one could begin to assess the prevalence of specific elements (e.g., Personality Characteristics, Psychological States and Attack details) and their values (e.g., Social Skill Problems, Disgruntlement or Stress), towards identifying patterns of attack.

There are, however, a few challenges to the use of the framework now and in the future. One of these difficulties is in characterising aspects within insider threats, especially when trying to understand the mind-set of the individual that attacks. For instance, their intent can be based on both static and dynamic personality traits. Static traits such as the traditional OCEAN profiles [43] can be measured by using established psychological surveys that are quite often used by HR staff. Unfortunately, the reality is that it is difficult to definitively know whether information provided by surveys is truly accurate, particularly if the subject is already non-compliant. Dynamic traits are even more difficult to identify, and mostly can only be inferred by qualified staff (e.g., personnel in HR with an appropriate psychology or psychoanalysis background) from the actions of the individual; not all companies will have such specialist staff at hand. Often there is the possibility of relying on the co-operation of other individuals within the organisation to report suspicious behaviour observed in the workplace, or outside of it. The problem here, however, is that this may be highly subjective or pure conjecture. With all of these aspects in mind, in reality, it is extremely difficult to effectively collect useful data, and in many cases, this even applies after an attack.

As an exercise in understanding how much data – especially on the precursors to an attack – is currently captured in reported cases, we reflected on the 179 cases we gathered throughout our study. From our analysis, we found that there were some parts of the framework that were well-documented in all cases, but others that were mostly unknown. For instance, Attack and Organisation Characteristics, as well as Actor Roles, Types and Relationship States are typically well-known elements; these were present in over 90% of the cases. To some extent, this is unsurprising, given that it is easily observable data: i.e., the attack is typically now known and the perpetrator has been identified. In the cases where this information was unknown, the attacks featured unknown attackers or accidents that were unattributable, including loss, theft or sabotage.

On the other hand, elements where data was sparse included *Personality Characteristics* (present in only 32% of cases), *Historical Behaviour* (11% of cases), *Attitude Towards Work* (31% of cases), and *Observed Cyber* (8%) and *Physical* (37%) Behaviours. The first point that one will note is that a number of these elements are closely associated with the insider's mind-set. Our findings here, therefore, reiterate the difficulty in gathering this psychological information on insider threats; in many of the cases, we found that much more emphasis was placed on the attack itself, rather than on gathering data regarding indicators or precursors in order to learn from it. An encouraging finding with regards to the framework itself was that elements such as *Precipitating Event*, *Psychological State* and *Motivation to Attack* were either very often stated or easily inferrable from cases (present in 87%, 78% and 90% of cases respectively). This gives some hope in terms of precursors to attack that are currently captured, and therefore can be described and modelled, and later used for pattern identification. More generally, this mapping of the full set of case data has also been useful for the framework as it has validated its ability to capture and describe a large set of data.

Although it is useful to have detailed information on employees to better engage and understand the insider threat, the gathering of such information also raises issues about ethical and legal usage of information. As stated in the UK Data Protection Act, personal data can only be used for the specified purposes that it was collected for. Therefore, an organisation would need to declare the introduction of employee monitoring to the individuals within the organisation, if they choose to use information for this purpose. Employees may well show resentment against the idea of monitoring, feeling that it invades their privacy, or induces a lack of trust between the organisation and them. Kiser *et al.* [59] and Greitzer *et al.* [60] are two insightful articles that explore the ethical issues with monitoring and its broader impact in the enterprise. If companies do decide to engage in more comprehensive monitoring, it would be essential that they consider the impacts and possibly even run educational campaigns to assure employees that those who act within the acceptable behaviour for their job role should not be alarmed or concerned.

As a final discussion point regarding the future development of the framework, we address the issue of quantifying actions and their impact. Foundational work in risk management suggests that if an individual has motive, capability and opportunity, then they are likely to conduct an attack. However, a crucial question here is, what constitutes as 'enough' motive, or 'enough' capability? Likewise, somebody may well exhibit all these, and yet still choose not to attack. Much previous literature also discusses the concept that if an individual is disgruntled then they may choose to act out. Again though, it is difficult to know just how disgruntled an individual needs

to become in order for them to pose a threat. Within our framework, we do not aim to associate a particular value with an element, such as high stress or medium-to-high disgruntlement. Instead, we focus on defining and capturing the relationships between elements, for instance, how their psychological state will impact their motivation to attack. The framework could then be used further to assess past attacks for how often individuals have exhibited a particular set of attributes and what the outcome of this was (e.g., how the attack was initiated, or the flow of the attack). Such analysis, once mature and supported by identified attack patterns, may then allow one to infer the risk associated with observing a series of states within the framework. For example, if somebody has previous history of disruption, their psychological state is disgruntled, and they are about to be made redundant, then the organisation may choose to take appropriate measures.

VI. Conclusion and Future Work

Insiders who constitute a threat can have a significant impact on the systems, processes and data of an organisation, and ultimately, cause irreparable damage to its activities and reputation. To tackle this problem properly, enterprises need to have a good understanding of the threat that they face, and the various aspects that are likely to be involved. In this paper, we aim to facilitate a better understanding of the threat at hand, through the presentation of a unifying framework to fully characterise insider attacks. The framework has been developed by a successful application of grounded theory to a comprehensive set of cases and a substantial amount of relevant literature, and has been further evaluated and refined using a second independent set of cases (e.g., [23]). As such, we have some confidence that our framework is able to capture and identify a large proportion of the key elements that make up the insider-threat problem, from significant events and indicators (e.g. behavioural and technical elements), to the human factors that are behind (even unintentional) attacks.

We envisage that a principal use of the framework would be to analyse past attacks and allow the identification of patterns that may exist between them. For instance, a security practitioner who has documented insider-attack cases for their enterprise can map these cases using the framework to look for patterns across the set. In addition to identifying any patterns (e.g., common attack routes), they may be able to highlight areas where further information might be desirable (that is, gathering more data to fill in elements) in order to facilitate more complete mapping, again leading to better understanding. The problem of limited data is clearly a challenge to overcome within research on insider-threat, as has been discussed in Section V. We actually posit that the framework can be used as a guide for collecting and organising case data in the short term, and then once sufficient data has been gathered, or could apply it to spot patterns across the case set.

Looking towards the future, there are several avenues being explored to advance the research presented in this paper. The first is on the use of the framework itself. That is, the framework that we have developed has been created with enterprises, security practitioners and researchers in mind. In the future, therefore, we aim to conduct several formal and informal feedback exercises with security practitioners and researchers, in order to gain insight into the utility they perceive for the framework and, especially, how easy it would be for a typical practitioner to describe and model their own cases. Another area of interest to us is the use of the framework as a basis for an insider-threat detection approach. Although there are several challenges in terms of the direct application of the framework, as highlighted in our Discussion, there is arguably scope for enhancing detection via the attack patterns. That is, if an on-going case matches with a known attack pattern in the framework, this could highlight the need for investigation, or at least additional monitoring.

Acknowledgements

This research was conducted in the context of a collaborative project on Corporate Insider Threat Detection, sponsored by the UK National Cyber Security Programme in conjunction with the Centre for the Protection of National Infrastructure, whose support is gratefully acknowledged. The project brings together three departments of the University of Oxford, the University of Leicester and Cardiff University.

References

[1] PricewaterhouseCoopers. (2012) Cybercrime: Protecting against the growing threat - Events & Trends. [Online]. Available: http://www.pwc.tw/en/publications/events-and-trends/e256.jhtml

[2] Ponemon Institute and Attachmate Corporation. (2013) The risk of insider fraud second annual study: Executive summary. [Online]. Available: http://www.attachmate.com/resources/analyst-papers/bridge-ponemon-insider-fraud-survey.htm

[3] Clearswift. (2013) The enemy within. [Online]. Available: http://www.clearswift.com/sites/default/files/images/blog/enemy-within.pdf

[4] BBC News. (2013) Profile: Edward Snowden. [Online]. Available: http://www.bbc.co.uk/news/world-us-canada-22837100

[5] Reuters. (2011) Ex-Ford engineer sentenced for trade secrets theft. [Online]. Available: http://www.reuters.com/article/2011/04/13/us-djc-ford-tradesecrets-idUSTRE73C3FG20110413

[6] BBC News. (2011) Former Olympus boss Woodford blows whistle on company. [Online]. Available: http://www.bbc.co.uk/news/15742048

[7] 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*, 1st ed. Addison-Wesley Professional, 2012.

[8] FBI. (2010) Fannie Mae corporate intruder sentenced to over three years in prison for attempting to wipe out Fannie Mae financial data. [Online]. Available: http://www.fbi.gov/baltimore/press-releases/2010/ba121710.htm

[9] IDC. (2009) Insider risk management. [Online]. Available: http://www.imerja.com/files/file/White_Papers/RSA/IDC%20Report%20-%20Insider%20Risk%20Management.pdf

[10] SC Magazine. (2012) Danger within: Insider threat. [Online]. Available: http://www.scmagazine.com/danger-within-insider-threat/article/245432/

[11] CERT Insider Threat Team. (2013) Unintentional insider threats: A foundational study. [Online]. Available: http://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=58744

[12] The Telegraph. (2013) 'Boozing Brits' lose work devices when drinking. [Online]. Available: http://www.telegraph.co.uk/technology/mobile-phones/10468026/Boozing-Brits-lose-work-devices-when-drinking.html

[13] The Economic Times. (2011) Corporate business secrets getting leaked on social media websites. [Online]. Available: http://articles.economictimes.indiatimes.com/2011-11-21/news/30424900_1_social-media-social-networking-employees

[14] CSO. (2013) Spear phishing poses threat to industrial control systems. [Online]. Available: http://www.csoonline.com/article/740396/spear-phishing-poses-threat-to-industrial-control-systems

[15] J. Hunker and C. W. Probst, "Insiders and insider threats – an overview of definitions and mitigation techniques," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 2, no. 1, pp. 4–27, 2011.

[16] C. Colwill, "Human factors in information security: The insider threat who can you trust these days?" *Information Security Technical Report*, vol. 14, no. 4, pp. 186–196, 2009.

[17] E. D. Shaw and H. V. Stock, "Behavioral risk indicators of malicious insider theft of intellectual property: Misreading the writing on the wall," Symantec, Tech. Rep., 2011.

[18] F. L. Greitzer and R. E. Hohimer, "Modeling human behavior to anticipate insider attacks," *Journal of Strategic Security*, vol. 4, no. 2, pp. 25–48, 2011.

[19] O. Brdiczka, J. Liu, B. Price, J. Shen, A. Patil, R. Chow, E. Bart, and N. Ducheneaut, "Proactive insider threat detection through graph learning and psychological context," in *IEEE Symposium on Security and Privacy Workshops (SPW)*, 2012.

[20] P. A. Legg, N. Moffat, J. R. C. Nurse, J. Happa, I. Agrafiotis, M. Goldsmith, and S. Creese, "Towards a conceptual model and reasoning structure for insider threat detection," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 4, no. 4, pp. 20–37, 2013.

[21] J. R. C. Nurse, P. A. Legg, O. Buckley, I. Agrafiotis, G. Wright, M. Whitty, D. Upton, M. Goldsmith, and S. Creese, "A critical reflection on the threat from human insiders - its nature, industry perceptions, and detection approaches," in *International Conference on Human Aspects of Information Security, Privacy and Trust*. Springer, 2014, to appear.

[22] C. Willig, *Introducing qualitative research in psychology*, 3rd ed. Open University Press, 2013.

[23] M. Whitty and G. R. Wright, "Corporate Insider Threat Detection Internal Deliverable 3.1 Report of findings from Case Studies," 2013.

[24] Dark Reading. (2013) 8 egregious examples of insider threats. [Online]. Available: http://www.darkreading.com/insider-threat/slide-show-8-egregious-examples-of-insid/240152563

[25] CPNI. (2013) Insider data collection study. [Online]. Available: http://www.cpni.gov.uk/Documents/Publications/2013/2013003-insider_data_collection_study.pdf

[26] BBC. (2009) Previous cases of missing data. [Online]. Available: http://news.bbc.co.uk/1/hi/uk/7449927.stm

[27] FBI. (n.d.) The insider threat: An introduction to detecting and deterring an insider spy. [Online]. Available: http://www.fbi.gov/about-us/investigate/counterintelligence/the-insider-threat

[28] Dark Reading. (2012) Five significant insider attacks of 2012. [Online]. Available: http://www.darkreading.com/insider-threat/five-significant-insider-attacks-of-2012/240144559

[29] M. Bishop, S. Engle, S. Peisert, S. Whalen, and C. Gates, "We have met the enemy and he is us," in *Workshop on New security Paradigms*. ACM, 2009, pp. 1–12.

[30] K. R. Sarkar, "Assessing insider threats to information security using technical, behavioural and organisational measures," *Information Security Technical report*, vol. 15, no. 3, pp. 112–133, 2010.

[31] P. R. Sackett and C. J. DeVore, "Counterproductive behaviors at work," in *Handbook of industrial, work, and organizational psychology (Vol 1)*, N. Anderson, D. S. Ones, H. K. Sinangil, and C. Viswesvaran, Eds. London: Sage, 2001, pp. 145–164.

[32] PricewaterhouseCoopers, "Cybercrime: Protecting against the growing threat," http://www.pwc.tw/en/publications/events-and-trends/e256.jhtml, 2012.

[33] W. R. Claycomb, C. L. Huth, L. Flynn, D. M. McIntire, and T. B. Lewellen, "Chronological examination of insider threat sabotage: preliminary observations," *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications*, vol. 3, no. 4, pp. 4–20, 2012.

[34] A. Jones and D. Ashenden, *Risk management for computer security*. Butterworth-Heinemann, 2005.

[35] M. J. Martinko, M. J. Gundlach, and S. C. Douglas, "Toward an integrative theory of counterproductive workplace behavior: A causal reasoning perspective," *International Journal of Selection and Assessment*, vol. 10, no. 1-2, pp. 36–50, 2002.

[36] J. M. Stanton, K. R. Stam, I. Guzman, and C. Caledra, "Examining the linkage between organizational commitment and information security," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3. IEEE, 2003, pp. 2501–2506.

[37] D. Vaughan, "Theory elaboration: The heuristics of case analysis," in *What is a case? Exploring the foundations of social inquiry*, C. Ragin and H. Becker, Eds. Cambridge: Cambridge University Press, 1992, pp. 173–202.

[38] L. A. Clark and D. Watson, "Mood and the mundane: relations between daily life events and self-reported mood," *Journal of personality and social psychology*, vol. 54, no. 2, pp. 296–308, 1988.

[39] S. L. Pfleeger, J. B. Predd, J. Hunker, and C. Bulford, "Insiders behaving badly: Addressing bad actors and their actions," *Trans. Info. For. Sec.*, vol. 5, no. 1, pp. 169–179, 2010.

[40] A. P. Moore, D. M. Cappelli, and R. F. Trzeciak, "The "Big Picture" of insider IT sabotage across U.S. critical infrastructures," CMU-CERT Program, Tech. Rep., 2008.

[41] B. Marcus and H. Schuler, "Antecedents of counterproductive behavior at work: a general perspective." *Journal of Applied Psychology*, vol. 89, no. 4, p. 647, 2004.

[42] McAfee. (2011) Data loss by numbers (Whitepaper). [Online]. Available: http://www.mcafee.com/us/resources/white-papers/wp-data-loss-by-the-numbers.pdf

[43] J. S. Wiggins, *The five factor model of personality: Theoretical perspectives*. Guilford Press, 1996.

[44] D. L. Paulhus and K. M. Williams, "The dark triad of personality: Narcissism, Machiavellianism, and psychopathy," *Journal of research in personality*, vol. 36, no. 6, pp. 556–563, 2002.

[45] R. R. McCrae and P. T. Costa Jr., "Toward a new generation of personality theories: Theoretical contexts for the five-factor model," in *The five-factor model of personality: Theoretical perspectives*, J. Wiggins, Ed. NY: Guilford Press, 1996, pp. 51–87.

[46] E. T. Axelrad, P. J. Sticha, O. Brdiczka, and J. Shen, "A Bayesian network model for predicting insider threats," in *IEEE Symposium on Security and Privacy Workshops (SPW)*, 2013.

[47] T. Halevi, J. Lewis, and N. Memon, "A pilot study of cyber security and privacy related behavior and personality traits," in *22nd International Conference on WWW*, 2013, pp. 737–744.

[48] 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," CMU-CERT Program, Tech. Rep., 2006.

[49] D. Liginlal, I. Sim, and L. Khansa, "How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management," *Computers & Security*, vol. 28, no. 3, 2009.

[50] L. Greenberg and J. Barling, "Predicting employee aggression against coworkers, subordinates and supervisors: The roles of person behaviors and perceived workplace factors," *Journal of Organizational Behavior*, vol. 20, no. 6, pp. 897–913, 1999.

[51] ——, "Employee theft," *Trends in organizational behavior*, vol. 3, pp. 49–64, 1996.

[52] The Corporate Insider Threat Detection Project, n.d., http://www.cs.ox.ac.uk/projects/CITD.

[53] B. Schneier, "Attack trees," *Dr. Dobbs journal*, vol. 24, no. 12, pp. 21–29, 1999.

[54] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *Leading Issues in Information Warfare & Security Research*, vol. 1, p. 80, 2011.

[55] The MITRE Corporation. Common Attack Pattern Enumeration and Classification (CAPEC). [Online]. Available: http://capec.mitre.org

[56] G. J. Silowash, D. M. Cappelli, A. P. Moore, R. F. Trzeciak, T. Shimeall, and L. Flynn, "Common sense guide to mitigating insider threats," CMU-CERT Program, Tech. Rep., 2012.

[57] The Register. (2013) MoJ fined 140K for emailing privates of 1,000 inmates. [Online]. Available: http://www.theregister.co.uk/2013/10/22/inmate_detail_mailout_data_breach

[58] Y. Barlas, "Leverage points to march upward from the aimless plateau," *System Dynamics Review*, vol. 23, no. 4, pp. 469–473, 2007. [Online]. Available: http://dx.doi.org/10.1002/sdr.389

[59] A. I. Kiser, T. Porter, and D. Vequist, "Employee monitoring and ethics: Can they co-exist?" *International Journal of Digital Literacy and Digital Competence (IJDLDC)*, vol. 1, no. 4, pp. 30–45, 2010.

[60] F. L. Greitzer, D. A. Frincke, and M. Zabriskie, "Social/ethical issues in predictive insider threat monitoring," *Information Assurance and Security Ethics in Complex Systems: Interdisciplinary Perspectives*, pp. 132–161, 2010.