Towards a Conceptual Model and Reasoning Structure for Insider Threat Detection

Philip Legg*, Nick Moffat, Jason R.C. Nurse , Jassim Happa,
Ioannis Agrafiotis, Michael Goldsmith, and Sadie Creese
Cyber Security Centre, Department of Computer Science, University of Oxford, UK

Abstract

The insider threat faced by corporations and governments today is a real and significant problem, and one that has become increasingly difficult to combat as the years have progressed. From a technology standpoint, traditional protective measures such as intrusion detection systems are largely inadequate given the nature of the 'insider' and their legitimate access to prized organisational data and assets. As a result, it is necessary to research and develop more sophisticated approaches for the accurate recognition, detection and response to insider threats. One way in which this may be achieved is by understanding the complete picture of why an insider may initiate an attack, and the indicative elements along the attack chain. This includes the use of behavioural and psychological observations about a potential malicious insider in addition to technological monitoring and profiling techniques. In this paper, we propose a framework for modelling the insider-threat problem that goes beyond traditional technological observations and incorporates a more complete view of insider threats, common precursors, and human actions and behaviours. We present a conceptual model for insider threat and a reasoning structure that allows an analyst to make or draw hypotheses regarding a potential insider threat based on measurable states from real-world observations.

Keywords: Insider Threat, Conceptual Model, Reasoning Structure

1 Introduction

The full extent of the insider-threat problem is becoming more apparent and more concerning every day. According to the 2011 CyberSecurity Watch Survey [1], whilst 58% of cyber-attacks on organisations are attributed to outside threats, 21% of attacks are initiated by their own employees or trusted third parties. In the Kroll 2012 Global Fraud Survey [2], they report that 60% of frauds are committed by insiders, up from 55% in the previous year. Likewise, the 2012 Cybercrime report by PwC [3] states that the most serious fraud cases were committed by insiders. Cases such as Robert Hanssen [4], Bradley Manning [5] and most recently, Edward Snowden [6], highlight the severity of insider threat towards government organisations. Of course, this only represents a fraction of the problem since in each of these cases the insider was eventually detected. In reality, however, many organisations fail to detect the presence of an insider threat which can cost billions of pounds per year and cause serious damage to the organisation, much of which therefore goes unreported and so the true extend of the problem is still unknown. For the purpose of this paper, we consider an 'insider' to be anyone with privileged access (e.g., an employee, contractor, client or business partner) to an organisation's data, systems or infrastructure, and an 'insider threat' to be an insider that intentionally abuses this access for some gain.

Whilst there already exists much work on the topic of insider-threat detection, the problem is still persistent within modern society. Previously, much work has focused on monitoring and profiling the

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, volume: 4, number: 4, pp. 20-37

*Corresponding author: Cyber Security Centre, Department of Computer Science, University of Oxford, Parks Road, Oxford, OX1 3QD, UK. Phone: +44-(0)1865-273838, Email: phil.legg@cs.ox.ac.uk

technological actions of insiders. However, it is evident that the problem extends beyond the technological domain and that addressing the behavioural and psychological characteristics of the insider may help to predict or detect a potential insider threat. In this paper we propose such an approach, by developing a conceptual model for insider threat that aims towards an all-encompassing organisational view of problem. The model incorporates a reasoning structure that can provide alerts to an analyst based on measurement data collected from real-world observations, and also allows an analyst to formulate hypotheses and explore the potential of insider threats. We go beyond the current state of the art, with an emphasis on a tiered approach, the reasoning structure within and between tiers, and the synoptic organisational view of problem. The conceptual model has been developed with the practicalities of implementation in mind, whilst maintaining flexibility for elements to be incorporate at a later stage, ensuring that the model is as far as possible future-proof. Most importantly, the conceptual model is deliberately implementation-agnostic, in order to make it widely adoptable.

The remainder of this paper is organised as follows. Section 2 discusses related work, including modelling, detecting and psychosocial analysis of the insider-threat problem. Section 3 details our conceptual model by examining thee three-tiered approach of Hypotheses, Measurements and Real World tiers. Section 4 describes the various elements that feed upwards through the model from the Real World tier, taking a structured four-lane approach to classify Enterprise, People, Technology and Information, and Physical elements within the organisation. Section 5 focuses on the reasoning capability of our model, which connects the three tiers of the model through the use of hypothesis-trees. Section 6 provides a discussion that reflects on our model and compares it against other insider-threat modelling approaches. Finally, Section 7 concludes the paper and indicates directions for future work.

2 Related Work

Insider threat research has attracted a significant amount of attention in the literature due to the severity of the problem within many organisations. Back in 2000, early workshops on insider threat highlighted the many different research challenges surrounding the topic [7]. Since then, there have been a number of proposals to address these challenges. For example, Greitzer *et al.* [8] discuss strategies for combating the insider-threat problem, including raising staff awareness and more effective methods for identifying potential risks. They define an insider to be an individual who currently, or at one time, was authorised to access an organisation's information system, data, or network. Likewise, they refer to an insider threat as a harmful act that trusted insiders might carry out, such as causing harm to an organisation, or an unauthorised act that benefits the individual. The CERT program at Carnegie Mellon University (CMU) has conducted much foundational work surrounding the insider-threat problem, resulting in over 700 case-studies that detail technical, behavioural, and organisational details of insider crimes [9]. They define a malicious insider to be a current or former employee, contractor, or other business partner who has or had authorized access to an organisation's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organisation's information or information systems. Spitzner [10] discusses early research on insider-threat detection using honeypots (decoy machines that may lure an attack). However, as security awareness increases, those choosing to commit insider attacks are finding more subtle methods to cause harm or defraud their organisations, and so there is a need for more sophisticated prevention and detection.

Due to the complexity of the insider-threat problem, much of the previous work focuses on the challenge of modelling the problem, and providing framework solutions that attempt to tackle the issues identified through modelling. Schultz [11] presents a framework for prediction and detection of insider attacks. He acknowledges that no single behavioural clue is sufficient to detect insider threat, and so suggest using a mathematical representation of multiple indicators, each with a weighted contribution.

Wood [12] presents an analytical model of insider threat that evaluates the attributes of the insider. The attributes presented are: access, knowledge, privileges, skills, risk, tactics, motivation, and process. Magklaras and Furnell [13] propose a threat evaluation system that estimates the level of threat that is likely to originate from a particular insider based on certain profiles of user behaviour. Butts *et al.* [14] propose a hierarchical tree model that illustrates the distinct actions that an insider threat could potentially carry out as part of an attack. They consider four types of attack actions: alteration, distribution, snooping and elevation. Maybury *et al.* [15] developed an observables taxonomy for the analysis and detection of insider threat that goes beyond only cyber actions to also incorporate such measures as physical access, violations, finances and social activity. Althebyan and Panda [16] present a model for insider-threat prediction based on the insider's knowledge and the dependency of objects within the organisation. Eom *et al.* [17] propose a framework to create a layered defence against insider threat based upon attack trees and misuse monitors. Likewise, Nithiyanandam *et al.* [18] also suggest a layered defence based on user authentication, activity monitoring, data monitoring, resource monitoring and an over-arching defence manager. Similar to the approach taken by the CERT program, Martinez-Moyano *et al.* [19] also use system dynamics to model the emergence of insider-threat vulnerabilities. Bishop *et al.* [20] discuss the insider-threat problem, and note that the term insider threat is ill-defined, and rightly recognise that there should be a degree of "insiderness" rather than a simple binary classification of insider threat or not. They propose the Attribute-Based Group Access Control (ABGAC) model, as a generalisation of role-based access control, and show its application to three case studies [21]: embezzlement, social engineering, and password alteration. Doss and Tejay [22] propose a model for insider-threat detection that consists of four stages: monitoring, threat assessment, insider evaluation and remediation. Pfleeger *et al.* [23] present a framework for describing insiders and their actions, and use example cases to show how their taxonomy helps understand the situation and how it might have possibly been prevented.

There are a number of proposed models and frameworks that already exist within the literature for characterising, or detecting, insider threat. However, many of these previous approaches neglect the human behavioural and social aspects of insider threat, or may only address a small subset of the human aspects. In particular, most prior work focuses on either the online *or* the offline behaviour — there is limited work that combines the two. Gonzalez and Sawicka [24] develop a system-dynamics framework for assessing human factors in information-security systems, however they do not consider the technological elements of the insider-threat problem. Shaw [25] suggests that there are at least two different types of insiders (thieves and disgruntled) who show different behavioural patterns, following his study of insider threat case studies and theories. Colwill [26] examines the human factors surrounding insider threat in the context of a large telecommunications organisation, remarking that greater education and awareness of the problem is required. Greitzer *et al.* [27] focus on incorporating inferred psychological factors into a modelling framework. Kandias *et al.* [28] present a prediction model that incorporates psychological profiling, real-time usage profiling, and a decision manager that assesses motive, opportunity and capability based on these profiles. Brdiczka *et al.* [29] combine psychological profiling with structural anomaly detection to develop an architecture for insider-threat detection. Sasaki [30] proposes a detection framework based on psychological triggers that impel an insider to act in a particular way (for instance, if the organisation announce an inspection, an insider threat may begin deleting their tracks and other data records). In our work, we build upon the existing literature and aim to develop an all-encompassing model of the insider-threat problem, which accurately captures sufficient technological, social and behavioural threat indicators for implementation within a deployed detection architecture.

3 Conceptual Model

To address the complex and dynamic problem that an insider threat may pose, we have developed a conceptual model for insider-threat detection. Figure 1 illustrates our three-tier model, which incorporates a Hypotheses tier, a Measurements tier and a Real World tier. The analyst tasked with detecting potential insider threats is positioned above the Hypotheses tier with a 'view' that passes through each of the tiers below. Furthest from the analyst is the Real World tier. This tier contains a large set of elements that exist in the real world that correlate with insider threat. Example elements would include an insider's system-activity logs, building-access logs, their physical behaviour and their psychological mindset. Importantly, detection engines will not have direct access to elements within this tier. The higher tiers must rely on measurements of the Real World elements, discussed next.

The middle tier of the model is the Measurements tier, where measurements of Real World elements are recorded. Some of these (particularly the technical ones) are directly observable and therefore we would expect to have a high confidence in the associated values – for instance, system access logs can be used to determine ('measure') definitively whether or not an insider's account downloaded sensitive Intellectual Property (IP). Other Real World elements (particularly the psycho-social ones) are only observable indirectly based on usually small sets of indicators, so will have lower confidences – an example is 'measuring' an employee's stress level as high because they have a large workload and are exhibiting withdrawal from co-workers; confidence here may be quite low, due to the inevitable paucity of information about the insider's psychological state.

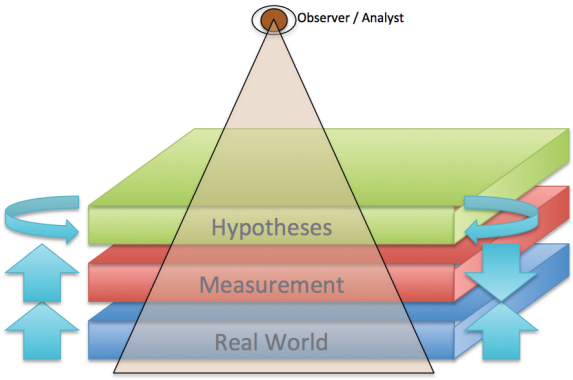


Figure 1: Tiered approach illustrates how the model supports both bottom-up and top-down reasoning

The Hypotheses tier is the top tier and is where the analyst interacts with the model. It represents the top level hypotheses the analyst could explore regarding an insider threat. We allow for hypotheses to be formed either by the underlying reasoning component that operates from the Real World upwards (as highlighted by the arrows on the left of Figure 1), or by an analyst investigating an insider of interest based on suspicions, thus from the Hypotheses tier down the model (right side of Figure 1). Detection and assessment of insider threat involves reasoning within and between the top two tiers of our model, seeded by measurements of Real World elements to determine a degree of confidence in top-level hypotheses (typically that particular individual insiders are currently or potentially threats to the organisation).

Bottom-up reasoning begins with measurements being deduced from Real World elements. Such measurements could include frequency- or time-based counts, however, psychological elements will be more qualitative. These measurements form the basis for generating statistical profiles for each insider, and profiles capturing their traits and behaviours. From these, indirect measurements can be made that

are intended to give an indication of the insider's mindset, potentially including their intentions. As necessary, alerts would then be generated by the system that feed into sub-hypotheses, i.e., low-level hypotheses that can later be used to construct more useful parent hypotheses. The analyst could then formulate the more complete and high-level hypothesis based on the detailed, reliable system alerts and sub-hypotheses; as will be discussed further in Section 5. For example, the system may alert the analyst if the collected measurements indicate that an employee is sending more emails than usual, entering the office at unusual times, has been in email contact with rival corporations, and is disgruntled because of being denied a promotion. The analyst may therefore hypothesise that the individual is acting as an insider threat based on these alerts.

Alternatively, with top-down reasoning, the process begins with the analyst having some cause for concern regarding a particular person. This may be as a result of an observed action (e.g., a persistent attempt to access sensitive company files), or by a suspicion raised by a company employee or customer. It could also be as a result of insight gained through previous bottom-up reasoning, or simply an analyst wanting to conduct a "what-if" analysis under the assumption of a particular scenario (e.g., dismissal or demotion). In these cases, the analyst would first formulate a hypothesis (e.g., contractor X might be an insider threat) – this will typically have a number of related sub-hypotheses and alerts – and using this, the model would attempt to reason on it using the available data from the Measurements tier. The model output would be the probability of the hypothesis being true.

Top-down reasoning can be particularly useful, in that it allows the analyst to identify what additional measurements may need to be taken (or 'switched on') from the real world that would allow thorough reasoning about a potential insider threat, or reasoning with increased confidence. For example, there may be suspicions about an employee being a threat, however, the organisation may not be able to reason about this possibility with any reasonable degree of confidence unless they start monitoring system-calls or begin logging access to secure areas within the company. Although bottom-up and top-down are separately presented, there is close interaction between the two and at some point in each reasoning approach, the other is required. Indeed, one could also envisage the model generating a set of hypotheses, then (based on these) the analyst guides detailed analysis and monitoring.

4 Elements of the Model

At the core of our conceptual model are the elements that provides a comprehensive view of the insider-threat problem. Figure 2 show our elements diagram, that is representative of the "Real World" tier in our conceptual model. Our development has been guided by the related works (Section 2), case studies reported in the literature (e.g., CERT [9]) and case studies learnt from organisations first-hand by ourselves and our collaborators as part of our research. Our motivation is to provide an all-encompassing view of the organisation that considers all factors related to insider threat. From this, there would exist a subset of elements that are actually feasible to assess should an insider-threat system be implemented from our model. Lastly, there would exist a further subset of elements that take account of the ethical and legal considerations that the organisation operate under. For instance, the use of social-media data regarding employees is not a technological limitation, but due to current ethical and legal restrictions this may be deemed as a breach of privacy within an organisation's policy, and so such data would be prohibited for use to detect potential insider threats. By taking an all-inclusive approach to what elements should be incorporated, and not restricting ourselves to what is currently feasible, this should help to future-proof our model as better approaches to measuring or modelling more complex elements are developed. It is important to acknowledge that whilst we aim for an all-encompassing view, we do not claim completeness with respect to the elements, attributes or relationships shown. However, the model would easily support the addition or modification of elements in the future.

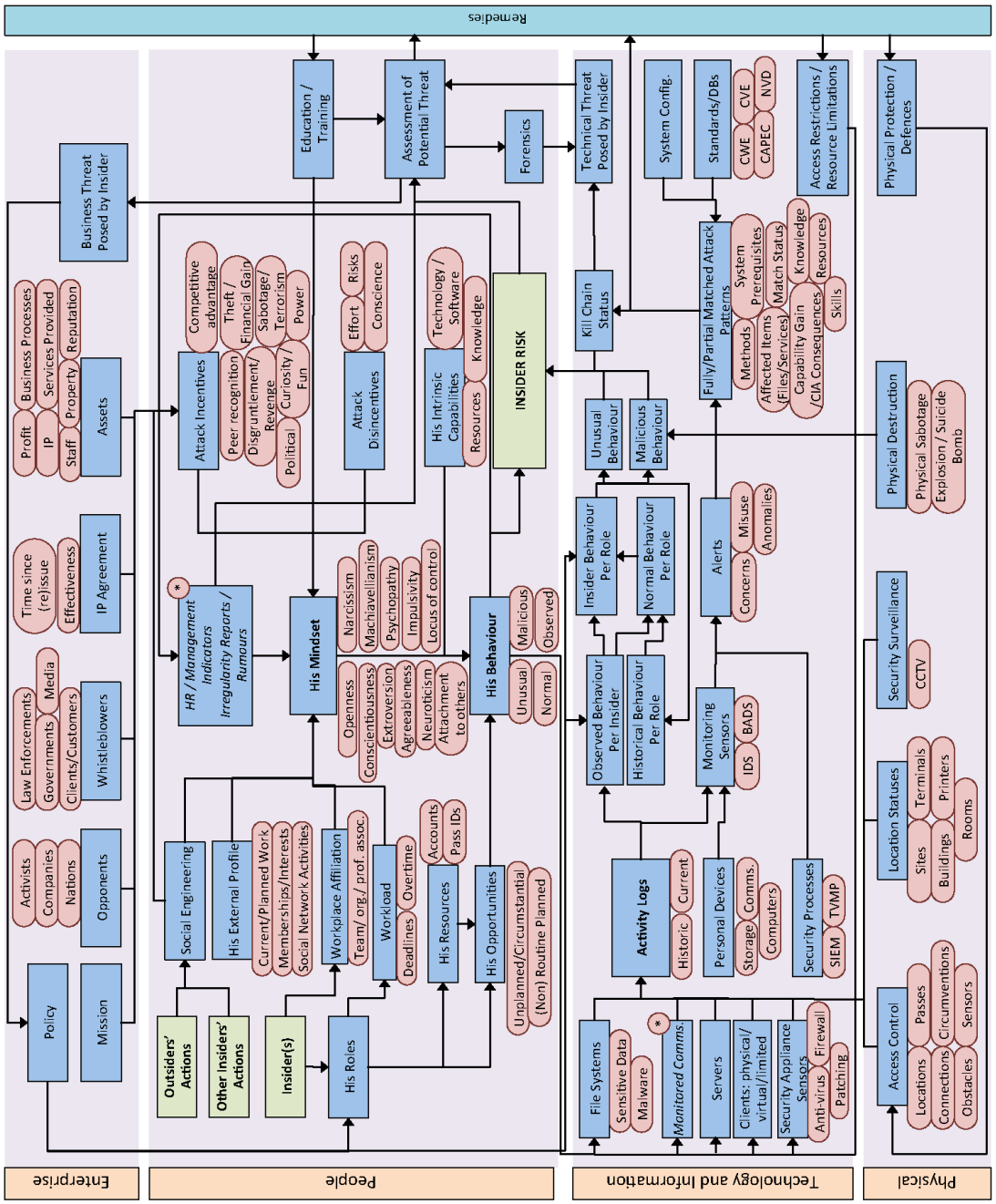


Figure 2: Elements diagram that represents the Real World level of the conceptual model. Elements are shown within their respective lanes: Enterprise, People, Technology and Information, and Physical. Green elements illustrate those that are specific to insider threat. All major elements that contribute towards the respective lane are shown in blue. Element attributes are shown in pink. The arrows represent the influence or impact between elements.

We categorise elements into four distinct categories, known as *lanes*. This lanes approach draws on the four-lane classification proposed by the SATURN project [31] that consists of *Enterprise*, *Information*, *Technology* and *Physical* lanes. We adapt this by combining the Technology and Information lanes, and incorporating a People lane. The combination of the Technology and Information lanes reflects the close connection shared between these within modern enterprises. The addition of the People lane allows for mapping of people behaviour, mindset and motivations, including how a person may act alone or in a group. This is crucial to the insider-threat problem, since there is a need for understanding the human element that motivates the insider to become a threat. It is here that the insider threat resides within the conceptual model. The four lanes of our model are described as follows:

- **Enterprise** – elements that constitute the enterprise on an operational level.
- **People** – elements describing an insider, his motivations and his behaviour within the enterprise.
- **Technology and Information** – elements relating to hardware and software in the enterprise and the digital activities that can be recorded.
- **Physical** – elements that capture physical components (e.g., locations) that exist within the enterprise.

Our strategy has been to identify individual elements from the problem space that may provide useful data for analysis. Many elements will have a number of attributes that characterise that element. For instance, not every insider threat will have the same attack incentives; some may attack for financial gain, others may attack for competitive advantage. We also represent the relationship by arrows between elements that shows the influence or impact that one element may have on another, which in some cases could be causal and in others could be dataflow-related. Finally, the model supports reasoning within and between tiers to allow an analyst to hypothesise about the presence of an insider threat. Elements would serve to inform measurements that are used to perform reasoning throughout the tiers of the model. In the following sections we describe each lane in detail.

4.1 Enterprise Lane

The Enterprise lane shows the elements that directly or indirectly impact the operation of the enterprise. In the context of insider-threat detection, this primarily concerns the business risk posed by an insider threat. Whilst an insider who steals petty cash from a cash register is clearly having a negative impact on the business, the risk is relatively low compare to the insider who sells on IP to a rival organisation which could threaten the existence of the business.

Other aspects of the enterprise include internal and external influences of other corporations and governments, the enterprise's relationship with its competitors, its reputation, its working culture, profits, global span and company mission. Mapping out the operational influences in the enterprise allow a variety of context-dependent attack opportunities to be identified. For instance, a hospital environment is substantially different from that of a bank and so the types of insider threat are likely to differ also. The bank may expect more financially-motivated threats, whereas the hospital may be focused towards data theft or modification of patient records for personal gain.

4.2 People Lane

The People lane shows the elements that relate a potential insider threat to a range of psychological, behavioural and personal factors. For example, *His Mindset* and *His Behaviour* are two key factors that directly relate to what triggers an insider to act as a threat towards the enterprise. Factors such as *Attack Incentives*, *His Intrinsic Capabilities* and *His Opportunities* can also impact on the possibility of the

insider becoming a threat. Similarly, *Irregularity Reports and Rumours*, *HR and Management Indicators*, and *Workload* may result in the mindset and behaviour of an insider shifting towards malicious actions.

Other Insiders' Actions and *Outsiders' Actions* play a significant role in how an insider may become a threat to the enterprise. Whilst many previous models of insider threat concentrate on a single person as a threat, it should be recognised that quite often there may be more than one person operating as a threat inside the organisation. For instance, the organisation may have a policy that requires multiple users to enter their credentials to access sensitive information. Collusion between a group of insiders may also make it more difficult to detect malicious behaviour of a single person. It is therefore vital that the model can account for this. With regards to outsider actions, it is apparent from case studies that insiders may become a threat as a result of being coerced by an outsider [9]. This could range from such activities as another organisation wanting to gain competitive advantage and recruiting the insider to obtain information for them, through to a partner or family member taking advantage of an opportunity and forcing them to steal goods for financial or personal gain. Whilst the outsider's actions cannot be monitored directly, it is important that communication (be it technological-based or otherwise) between the insider and outside parties should be considered within the scope of the model.

An Insider may have one or more *Roles* within the organisation, for which the Insider has a number of responsibilities (*Workload*), a finite set of *Resources*, and a number of *Opportunities*. The workload of an insider will directly impact their mindset, since a low workload may result in boredom or demotivation whilst a high workload may result in stress and frustration. The resources that an insider can access within their role will influence the opportunities that may present themselves. For instance, if the insider has access to financial transactions, then the opportunity for stealing money becomes greater. Other opportunities may be accidental, for instance, if an insider notices that a stock room has been left unlocked. Based on these circumstances, whether an insider chooses to act maliciously is highly dependent on their mindset. In addition, their mindset will be based also on how the insider perceives their affiliation with the enterprise. A weak affiliation suggests that they do not care for their enterprise, whilst a strong affiliation may mean that a denial of request is taken much more personally. Both outcomes could affect the mindset that the insider has towards the enterprise. Within the psychology literature there is existing research that aims to characterise a person's mindset, such as the well-established OCEAN [32], and the Dark Triad [33]. From our research we also found that impulsivity, locus of control, and attachment to others all attribute to the mindset of those who have acted as an insider threat.

Irregularity Reports and *Management Indicators* sit within the People lane and focus on such factors as aggressive or suspicious behaviour, personal circumstances, and employment records. Short employments in similar job roles may be indicative of a high risk person, likewise knowing that a person has financial difficulties or a criminal record should all be factored in to the risk they could pose. Reports would be made as a result of observed behaviour. Likewise, their mindset may be affected by the knowledge that a report has been made, or a review meeting has taken place. It is also within the People lane that an *Assessment of Potential Threat* is made by the analyst, based on the suspicion of Insider Risk. This may be as a result of technological monitoring (bottom-up reasoning), or based on the behavioural observations and intuition (inferred by top-down reasoning). This gives rise to the cycle between *Assessment*, *Forensics* and *Technical Threat Posed by Insider*, which links social, behavioural and technological observations. *Education and Training* of staff would also influence *Assessment*, which in turn indicates the extent of the *Business Threat Posed by Insider*.

4.3 Technology and Information Lane

This lane describes the Technology and Information elements within the enterprise. Information may come from a number of different sources such as activity logs, communication channels, physical sensors, and performance reports. From this, historic and currently observed profiles can be formed to

characterise an employee (likewise, the same could also be done on a per-role or per-organisation basis). These profiles can be used to help determine unusual (potentially malicious) behaviour.

In this lane are the technical facilities that exist within the enterprise, which would also be used to inform the monitoring process. *File Systems* includes the data stored on the enterprise systems. *Monitored communications* could consist of a wide variety of attributes based on the culture of the organisation, including telephone, email, web browsing, web traffic, social media (e.g., Facebook, Twitter), instant messaging, FTP, and remote access. All machines on the network supplied by the enterprise to each insider would reside within either the *Monitored Communications* (e.g., work mobile phone), *Servers* or *Clients* elements. *System Defences* could include anti-virus, firewalls and patching. *Personal Devices* includes BYODs (e.g., mobile phones and computers), and personal storage (e.g., Dropbox, USB sticks).

From these elements, *Activity Logs* are produced. Such logs would feed into the *Observed Behaviour Per Insider*. This can be used to inform the *Insider Behaviour Per Role*, and the *Normal Behaviour Per Role*. The *Insider Behaviour Per Role* may be classified as *Unusual* or *Malicious*, based on the knowledge of *Normal Behaviour Per Role*, or else it is regarded as normal and would feed into the *Historical Behaviour Per Role*. *Activity Logs* could also inform *Monitoring Sensors* such as Intrusion Detection Systems (IDS). This would generate *Alerts* that highlight concern, misuse or anomaly regarding an insider's actions. These alerts could be compared against *Attack Patterns* to determine full or partial matches based on previously known attack strategies. Any *System Configuration* and existing externally developed *Standards* or *Databases* such as CAPEC [34], CVE [35], and CWE [36] could also be used to further assess these matches. Should a match be found, then the *Kill Chain Status* would be initiated. *Unusual Behaviour*, *Malicious Behaviour*, and *Kill Chain Status* all indicate towards an *Insider Risk*.

4.4 Physical Lane

This lane represents the physical elements of an organisation including the access, location and destruction of physical buildings belonging to the organisation. If an insider accesses work premises outside of their usual working hours then this could be deemed as suspicious. Likewise, records of failed access attempts may also prove important. Mapping out the movements of an insider may reveal insight into their actions, for instance, access to unauthorised areas may signify a precursor to sabotage.

Access Control specifies locations, passes, connections, circumventions, obstacles and sensors in place that are able to detect how people move around (in real-world physical space). This information is communicated back to the central file system for storage and access control management. The *Location Statuses* element stores information about sites, terminals, buildings, rooms, printers, faxes and phones available to the enterprise. *Security Surveillance* includes measures such as CCTV for physical security monitoring. The *Physical Destruction* element outlines what physical destruction an insider could potentially cause, which would clearly indicate *Malicious Behaviour*.

4.5 Remedies

Once an *Insider Risk* has been computed per insider, a variety of *Remedies* can be proposed by the model that span across all four lanes. Remedies may include *Education and Training* in the People lane, *Access Restrictions or Resource Limitations* in the Technology lane, and *Physical Protection Or Defences* in the Physical lane. Each of these remedies provides a feedback cycle in the overall monitoring of the Insider. In the case of *Education and Training*, this will impact *His Mindset*, with the intention that they will be moved towards a more positive mindset towards the organisation. Whilst many cases may involve employee dismissal, *Education and Training* may also act as a deterrent so that others do not choose to act out an attack. *Access Restrictions or Resource Limitations* may impact on the *File Systems*, *Monitored Communications*, *Servers*, *Clients*, and *System Defences*, depending on the form of access

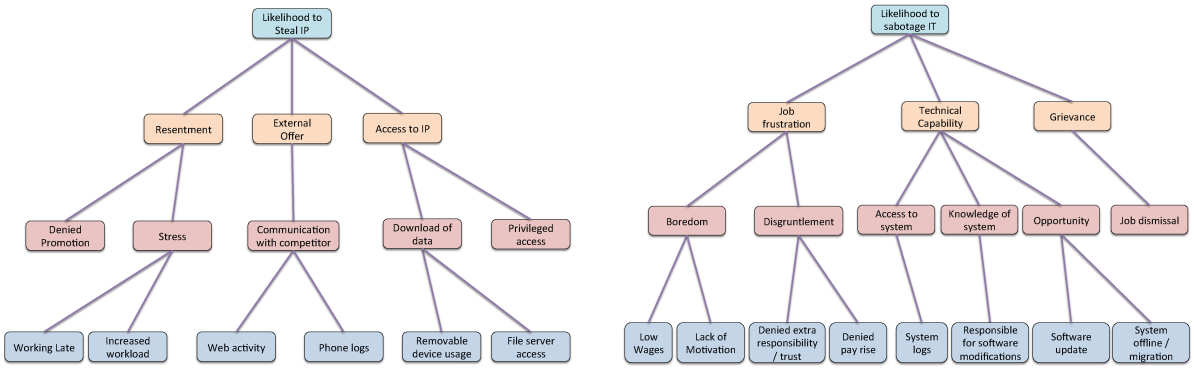


Figure 3: Hypothesis-tree generation. Directly-measured observables are shown at the lowest level of the tree, which then propagate upwards to the parent nodes to inform indirectly-measured elements.

restriction required. Likewise, in the physical lane, *Physical Protection or Defences* may impact on *Access Control* to block an Insider from entering a particular area of the building.

5 Hypothesis-Based Reasoning

As previously discussed, the model is designed to support reasoning within and between the various tiers. In order to achieve this, we construct hypothesis-trees that allow an analyst to derive the probability that a particular insider may pose a threat towards the organisation. The use of established theories from social-science and psychology literature, on empirical evidence and prior knowledge, or on mathematical models that represent relationships between elements, could all help to inform the construction of a hypothesis. The hypothesis-tree forms a connection between directly-measured observables at the Measurements tier, and the hypothesis that the analyst formulates at the Hypotheses tier.

Figure 3 presents two common scenarios of insider-threat activity: IP theft and IT sabotage. For each scenario we construct a hypothesis-tree that could be used to reason whether a particular event may occur based on measured evidence. For the benefit of this discussion we present a simplified subset of nodes; in reality the tree would consist of many more connected nodes. At the lowest level of the tree are the leaf nodes that are directly measurable. In (a), there are six directly-measured observables: *working late*, *increased workload*, *web activity*, *phone logs*, *removable device usage* and *file server access*. *Working late* could be observed through physical building access or system log-in activity. Likewise, web activity, phone logs, removable device usage and file server access can all be monitored in the Technology and Information lane. Increased workload would be observed within the People lane, through social interaction, reporting from other employees or knowledge of current deadlines. Parent nodes are indirectly measured through calculation from their child nodes. For example, the combination of working late and increased workload could define a measure of stress for the insider. Similarly, web activity and phone logs may reveal that the insider has been in frequent contact with a competitor.

As the tree is traversed towards the root node, there is a progression from directly to indirectly-measured elements such as resentment and grievance. Such psychological states alone may or may not be sufficient to indicate whether an insider is a threat, however, combined observations from all four lanes in our model would help to improve the confidence of the indicator. For instance, whilst high resentment towards the organisation is not desirable, it may not necessarily be sufficient to make a person become

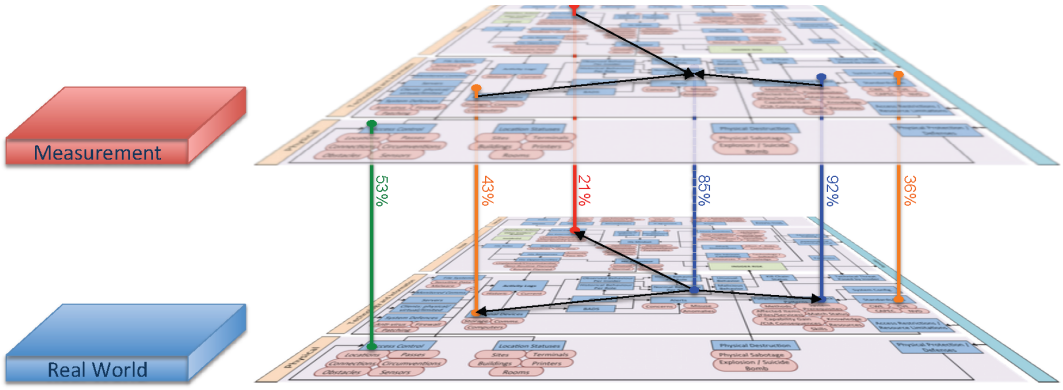


Figure 4: Deriving confidence values between the tiers of the model. Each element will have a confidence value as determined by the sensor that obtains that measurement. Confidences can be assigned for both directly-mesured observables (solid lines) and indirectly-measured elements (dashed lines).

a threat. However, if they also have access to IP then they may have a greater potential to act as a threat, and so should become a higher monitoring priority. If they have also been in contact with a rival organisation, then they may be looking to impress a potential new employer, further increasing the probability of them stealing IP. Likewise, an insider with technical capability, frustration in their current role, and a grievance towards their employer may then decide to cause damage to IT systems, for instance by installing malicious software. There have been numerous documented instances of this in the past [9].

We appreciate that attributes such as stress, resentment and grievance are difficult to directly measure. Therefore we calculate indirect measurements that give an estimate, based on a combination of direct measurements. Ideally the combination of elements used should yield a strong confidence for the resulting measure. It may be possible to assess confidence based on prior case studies, or from observed examples. With our approach, as more cases are observed the confidence of good indicators within the tree should increase, thus revealing which indicators prove reliable for detection. Given a result from the reasoning process, it is crucial that the analyst can be confident that the reasoner has produced a dependable result. Therefore, we also need to consider how we can assess the reasoning process.

Figure 4 shows how one could derive the confidence of elements between the tiers of the model. Each element at the Measurement tier would have a confidence that links this back to one or many Real World sensors (illustrated by the vertical lines). Likewise, each sensor may relate to one or many elements at the Measurement tier. The model would be agnostic towards the source of the confidence values, so as to avoid introducing bias. As the model begins to respond to newly-observed activity, we would expect the accuracy of the reasoning to change over time. The reasoner would assign a confidence score as to how well each element in the model can be measured, for both directly and indirectly-measured elements. As shown in Figure 4, the confidence of indirectly-measured elements would be based upon the confidences of directly-measured observables that contribute towards that measurement. The analyst can also choose to adjust the confidences based on their expert knowledge, so as to reflect the true performance of the reasoner. This in turn would update the confidence values associated with each element and inference within the conceptual model. We envisage that through continual use with the introduction of newly-observed data the confidences would then stabilise over time.

6 Discussion

In this paper we have presented a conceptual model for insider-threat detection. The model is based on a tiered approach that connects Real World, Measurements, and Hypotheses through the use of hypothesis-trees to evaluate the probability of insider threat. We present the elements at the Real World that feed upwards through the tiers, which are split into four lanes: *Enterprise*, *People*, *Technology and Information*, and *Physical*. This allows for a clear understanding of the relationship between elements, spanning across psychological, social and technological domains. It also provides a structured view of the elements that should be explored within the insider-threat problem, and moves towards how these elements should be measured in order to facilitate monitoring and hypothesis generation at higher tiers in the model.

As part of the model development, it is important that we acknowledge and recognise the extensive work on insider threat that has been conducted to date so far. In Section 2 we presented an overview of the related works that surround the topic. Since there are already previously proposed conceptual models for the prevention and detection of insider threat, it is therefore necessary to study how our model extends upon these. The different approaches presented in Section 2 could be categorised in four different types; those that emphasise on threshold, anomaly detection, rule-based and model based methods. There are limitations to each of these approaches. The threshold approach is in essence a collection of alerts which when they appear, cause a flag to be raised signalling possible suspicious behaviour. However this approach is not dynamic and it is highly dependant upon observations. In a similar vain, anomaly detection focuses on highlighting events and behaviours, but here the emphasis is on statistically outstanding data (outliers). This shares the same fate as the threshold model, since behaviours which should be detected as insider threats may not be statistically significant and thus non-detectable by this approach. Rule-based techniques seek for patterns that have already been identified as insider-threat behaviour and as such, are unable to detect any new attack patterns. Even if the database is updated with new patterns of attacks, this approach lacks flexibility as novel attacks or attacks that exhibit slight variation from those stored in the database will remain undetected. Model-based methods, similar to the rule-based approach, seek to identify attack vectors at an abstract level. Although this technique may not be dynamic, it can be used to establish a clear understanding of possible attack patterns.

For the remainder of our discussion, we shall focus on two groups that have contributed extensively to insider-threat research and have helped to inform and motivate the development of our model: the CERT program at CMU [9], and the work of Greitzer *et al.* [27] from the Pacific Northwest National Laboratory (PNNL). We shall address the proposed approach of each research group and identify how we believe our model manages to extend beyond the current state-of-the-art.

The CERT Insider Threat Center has been focused on the challenges that insider threats pose since 2001. They have gathered together a significant number of real-world case studies that show the types of insider threat that exist in the workplace today, including IT sabotage, IP theft and data fraud cases. This provides an invaluable resource for understanding the processes that an insider may engage in prior to becoming a threat, the mindset that the insider is likely to have, the attacks that insiders could conduct, and the detection mechanism that resulted in the insider being caught. The CERT team have focused on their collection of case studies, on which they have built models that represent different insider attacks. Thus far, they have used system dynamics to construct a series of MERIT (Management and Education of the Risk of Insider Threat) models, focusing on IT sabotage [37] and IP theft [38]. The approach CERT take is in essence a model-based method. It is a decisive step to better understand the behaviour of the insiders, however, a detection system which is required to be dynamic to detect potentially novel patterns of attacks, cannot be designed solely based on this approach.

There are some important distinctions between our approach and that taken by CERT. Our model aims to capture an all-encompassing view of an organisation, consisting of three tiers (Hypothesis, Measurement and Real World) and four lanes (Enterprise, People, Technology and Information and Physical),

of which we then drill down to observe individual elements regarding the insider threat, and use hypothesis trees to obtain hypothesis from the real world tier. On the other hand, CERT take the approach of starting with the specific insider-threat problem (e.g., IT sabotage), and building upon this to incorporate wider influences from the organisation. Since their models are specifically tailored towards particular attack patterns observed in the case studies, it is quite possible that this may restrict the models for the detection of novel attack vectors. The need for different models to represent the different attacks observed in their case studies may also prove challenging from an implementation viewpoint.

With our approach, we make no prior assumptions on the enterprise and focus purely on a wide capture of elements that relate to insider threat. In reality, the detection of a particular attack such as IT sabotage may only require a subset of the elements. The proposed model in this paper has the ability to represent the critical paths identified by CERT. The advantage of our conceptual model is the abstract and dynamic notion which supports any possible combination of thresholds, alarms or paths possible; provided that these paths will be well-defined in terms of elements and attributes within the organisation.

The work of Greitzer *et al.* [27, 8] and Bishop *et al.* [39] has contributed to the development a of conceptual model for an insider-threat reasoning framework. In [27], they presented a model that represents a dynamic belief propagation network, fed with raw data. The reasoning logic is modelled by a system named CHAMPION (Columnar Hierarchical Auto-associative Memory Processing In Ontological Networks). This is based upon the use of ontologies, reifiers, memory and AMCs (Auto-associative Memory Components), whereby data is interpreted and used to infer new assertions at each AMC, starting with raw data and feeding up through the hierarchy until a final decision is made. At a high level abstraction the model consists of: data, observations, indicators and behaviours. They adopt the metaphor of constructing a jigsaw, where individual pieces represent data entering the system. Data is grouped together to form observations, which are then grouped together to provide indicators, which are finally grouped to identify behaviours. The last component within their model is Attribute-Based Group Access Control, which is used to prioritise critical resources within the organisation that should be monitored (e.g., file servers with sensitive data), and also to prioritise insiders who currently exhibit concerning behaviours. The objective is to alleviate the workload of the analyst by focusing on priority insiders and resources.

Within their proposed model [27], they also focus on psychosocial profiling of actors, incorporating observables such as disgruntlement, anger management, disregard for authority and stress. They quite rightly acknowledge the importance of psychological and social behavioural monitoring, rather than assuming only a technical detection mechanism, which we also believe to be the case and so address at the People lane in our model. Although the CHAMPION system is closely related to our work, we believe that the elements presented in our conceptual model provide a more extensive viewpoint due to the inclusion of Physical and Enterprise lanes. In addition, we aim to go beyond the commonly-used psychological attributes surrounding insider threat such as disgruntlement and stress, by also incorporating well-established psychological theories such as OCEAN [32], and Dark Triad [33].

In [39], much of the psychological analysis that is discussed focuses on sentiment analysis. Whilst this should certainly be considered, it is important to address that it should not be relied on as the sole approach for psychological analysis due to ethical and legal ramifications. Instead, we explore how measurable actions can relate to psychological assessments such as OCEAN. For instance, the social graph of an insider based on their email usage could relate to extroversion, or their ability to submit work before a given deadline could be used to assess conscientiousness. Likewise, sporadic logging-in times may indicate that a person is neurotic, or their web browsing habits may suggest their openness. By taking an all-encompassing view, we can consider what is conceptually-possible, what is feasibly-possible, and what is ethically or legally possible, should such a system be implemented from our model.

Another significant distinction between the CHAMPION model and our approach is our ability to perform not only bottom-up reasoning but also top-down reasoning. As previously mentioned in Section 3, the top-down approach provides an opportunity for the analyst to explore their instinct and intu-

ition, by conducting "what-if" analyses to represent particular scenarios of interest which could reveal new patterns of attack. Whilst we appreciate the advances in pattern recognition techniques, it is important that the model supports both computer-based and human-based recognition since human intuition remains vital in understanding complex human behaviours. The strength of our model stems from the interaction that an analyst can have with the reasoning level. Furthermore, it is worth mentioning that the abstract and dynamic notion of our model, combined with the two-fold reasoning (bottom-up and top-down) could also provide an efficient mechanism for validating theories from social and psychology sciences. For instance, another advantage that our model offers is that in a simulation environment where the "real world" data is known, the top-down reasoning could allow for testing of theories. The closer the reasoning process is to the real-world tier, the more representative of reality the theory would be.

7 Conclusion

In this work we present our tiered model for insider-threat detection that encompasses elements from enterprise, people, technology and information, and physical domains. Our aim is to provide an all-encompassing view of an organisation, to identify when an insider may conduct threatening behaviour towards an organisation. Reasoning can be performed on elements within the real world, by deriving measurements that then propagate up through the model to provide a confidence value for a particular person being or becoming a threat. We have compared our approach with existing models to highlight where we feel that our model compliments and contributes with previous insider-threat research. Future work will address the development of a prototype detection architecture based on our proposed model.

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] CSO Magazine, CERT Program (Carnegie Mellon University) and Deloitte, "CyberSecurity Watch Survey: Organizations Need More Skilled Cyber Professionals To Stay Secure," 2011, http://www.sei.cmu.edu/newsitems/cybersecurity_watch_survey_2011.cfm.

[2] Kroll and Economist Intelligence Unit, "Annual Global Fraud Survey. 2011/2012," 2012.

[3] PricewaterhouseCoopers LLP, "Cybercrime: Protecting against the growing threat - Events and Trends," 2012.

[4] FBI, "Robert Philip Hanssen Espionage Case," 2001, http://www.fbi.gov/about-us/history/famous-cases/robert-hanssen.

[5] Guardian, "Bradley manning prosecutors say soldier 'leaked sensitive information'," 2013, http://www.guardian.co.uk/world/2013/jun/11/bradley-manning-wikileaks-trial-prosecution.

[6] BBC News, "Profile: Edward Snowden," 2013, http://www.bbc.co.uk/news/world-us-canada-22837100.

[7] R. Anderson, T. Bozek, T. Longstaff, W. Meitzler, M. Skroch, and K. Van Wyk, "Research on Mitigating the Insider Threat to Information Systems," in *Proceedings of the Insider Workshop*, August 2000.

[8] F. L. Greitzer, A. P. Moore, D. M. Cappelli, D. H. Andrews, L. A. Carroll, and T. D. Hull, "Combating the Insider Cyber Threat," *IEEE Security & Privacy*, vol. 6, no. 1, pp. 61–64, 2007.

[9] 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.

[10] L. Spitzner, "Honeypots: catching the insider threat," in *Proc. of the 19th IEEE Annual Computer Security Applications Conference (ACSAC'03), Las Vegas, Nevada, USA*, December 2003, pp. 170–179.

[11] E. E. Schultz, "A framework for understanding and predicting insider attacks," *Computers and Security*, vol. 21, no. 6, pp. 526–531, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016740480201009X

[12] B. Wood, "An insider threat model for adversary simulation," *SRI International, Research on Mitigating the Insider Threat to Information Systems*, vol. 2, 2000.

[13] G. B. Magklaras and S. M. Furnell, "Insider Threat Prediction Tool: Evaluating the probability of IT misuse," *Computers and Security*, vol. 21, no. 1, pp. 62–73, 2002.

[14] J. Butts, R. Mills, and R. Baldwin, "Developing an Insider Threat Model Using Functional Decomposition," in *Proc. of the 3rd International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security (MMM-ACNS'05), St. Petersburg, Russia, LNCS*, vol. 3685. Springer-Verlag, September 2005, pp. 412–417.

[15] M. Maybury, P. Chase, B. Cheikes, D. Brackney, S. Matzner, T. Hetherington, B. Wood, C. Sibley, J. Marin, T. Longstaff, L. Spitzner, J. Haile, J. Copeland, and S. Lewandowski, "Analysis and Detection of Malicious Insiders," in *Proc. of the 2005 International Conference on Intelligence Analysis, McLean, Viginia, USA*. MITRE, May 2005.

[16] Q. Althebyan and B. Panda, "A Knowledge-Base Model for Insider Threat Prediction," in *Proc. of the 2007 IEEE Information Assurance and Security Workshop (IAW'07), West Point, New York, USA*. IEEE, June 2007, pp. 239–246.

[17] J. Eom, M. Park, S. Park, and T. Chung, "A framework of defense system for prevention of insider's malicious behaviors," in *Proc. of the 13th International Conference on Advanced Communication Technology (ICAT'11), Phoenix Park, Gangwon-Do, Korea*. IEEE, February 2011, pp. 982–987.

[18] C. Nithiyanandam, D. Tamilselvan, S. Balaji, and V. Sivaguru, "Advanced framework of defense system for prevetion of insider's malicious behaviors," in *Proc. of the 2012 International Conference on Recent Trends In Information Technology (ICRTIT'12), Chennai, Tamil Nadu, India*, April 2012, pp. 434–438.

[19] I. J. Martinez-Moyano, S. H. Conrad, E. H. Rich, and D. F. Andersen, "Modeling the Emergence of Insider Threat Vulnerabilities," in *Proc. of the 38th Conference on Winter Simulation (WSC'06), Monterey, California, USA*. IEEE, December 2006, pp. 562–568.

[20] M. Bishop, S. Engle, S. Peisert, S. Whalen, and C. Gates, "We have met the enemy and he is us," in *Proc. of the 2008 Workshop on New Security Paradigms (NSPW'08), Lake Tahoe, California, USA*. ACM, September 2008, pp. 1–12.

[21] ——, "Case studies of an insider framework," in *Proc. of the 42nd Hawaii International Conference on System Sciences (HICSS'09), Waikoloa, Big Island, Hawaii, USA*. IEEE, January 2009, pp. 1–10.

[22] G. Doss and G. Tejay, "Developing insider attack detection model: a grounded approach," in *Proc. of the 2009 IEEE International Conference on Intelligence and Security Informatics (ISI'09), Dallas, Texas, USA*. IEEE, June 2009, pp. 107–112.

[23] S. L. Pfleeger, J. B. Predd, J. Hunker, and C. Bulford, "Insiders behaving badly: Addressing bad actors and their actions," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 169–179, 2010.

[24] J. Gonzalez and A. Sawicka, "A framework for human factors in information security," in *Proc. of the 2002 WSEAS International Conference on Information Security, Hardware/Software Codesign, E-Commerce and Computer Networks, Rio De Janeiro, Brazil*, October 2002, pp. 1871–1877.

[25] E. D. Shaw, "The role of behavioral research and profiling in malicious cyber insider investigations," *Digital Investigation*, vol. 3, no. 1, pp. 20–31, 2006.

[26] 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.

[27] 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.

[28] M. Kandias, A. Mylonas, N. Virvilis, M. Theoharidou, and D. Gritzalis, "An insider threat prediction model," in *Proc. of the 7th international conference on Trust, Privacy and Security in Digital Business (TrustBus'10), Bilbao, Spain, LNCS*, vol. 6264. Springer-Verlag, August 2010, pp. 26–37.

[29] 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 *Proc. of the 2012 IEEE Symposium on Security and Privacy Workshops (IEEE SPW'12), San Francisco, California, USA*, May 2012, pp. 142–149.

[30] T. Sasaki, "A framework for detecting insider threats using psychological triggers," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 3, no. 1/2, pp. 99–119, 2012.

[31] S. Creese, M. H. Goldsmith, and A. O. Adetoye, "A logical high-level framework for critical infrastructure resilience and risk assessment," in *Proc. of the 3rd International Workshop on Cyberspace Safety and Security (CSS'11), Milan, Italy*, September 2011, pp. 7–14.

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

[33] 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.

[34] MITRE, "Common Attack Pattern Enumeration and Classification," http://capec.mitre.org/.

[35] ——, "Common Vulnerabilities and Exposures," http://cve.mitre.org/.

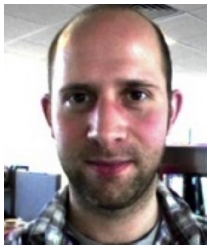
[36] ——, "Common Weakness Enumeration," http://cwe.mitre.org/.

[37] A. P. Moore, D. M. Cappelli, and R. F. Trzeciak, "The "Big Picture" of Insider IT Sabotage Across U.S. Critical Infrastructures," Software Engineering Institute/Carnegie Mellon University, Tech. Rep. CMU/SEI-2008-TR-009, May 2008.

[38] A. P. Moore, D. M. Cappelli, T. C. Caron, E. Shaw, D. Spooner, and R. Trzeciak, "A Preliminary Model of Insider Theft of Intellectual Property," Software Engineering Institute, Carnegie Mellon University, Tech. Rep. CMU/SEI-2011-TN-013, June 2011.

[39] M. Bishop, C. Gates, D. Frincke, and F. Greitzer, "AZALIA: an A to Z assessment of the likelihood of insider attack," in *Proc. of the IEEE 2009 Conference on Technologies for Homeland Security (HST'09), Boston, Massachusetts, USA*. IEEE, May 2009, pp. 385–392.

Author Biography



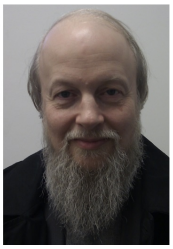
Philip Legg is a post-doctoral research associate in the Cyber Security Centre at the University of Oxford. He works on the Corporate Insider Threat Detection project to develop novel techniques for the detection and prevention of insider threat. His interests span across several research areas, including machine learning, data visualization, human-computer interaction, and computer vision. Previously, he worked at Swansea University exploring data visualization for sports video analysis. Prior to this, he obtained both his BSc and his PhD in Computer Science from Cardiff University, where his research focused on multi-modal medical image registration.

Nick Moffat has a BA in Mathematics from Cambridge University and a doctorate in Software Engineering from Oxford University. He has approximately 20 years of employment by UK government and industry behind him, during which time he developed and applied formal methods techniques and tools for analysis of system safety and security properties. On joining Warwick University 3 years ago, and now at Oxford, his focus has been on developing practical reasoning techniques for assessing privacy and security aspects of systems.

Jason R.C. Nurse received his B.Sc. in Computer Science and Accounting (UWI, Barbados – 2001), M.Sc. in Internet Computing (Hull, UK – 2006), and Ph.D. degree in Computer Science specialising in Web Services Security and e-Business (Warwick, UK – 2010). He has worked within industry and academia throughout his career. This has included various IT roles within industry, and academic posts such as Research Fellow at Warwick University, and more recently, Cyber Security Researcher at the University of Oxford. Jason has published several articles at both journal and conference levels and also sits on the programme committee board of related venues. His research interests include insider threats, online security risks, information security and trust, human aspects of security, services security.

Jassim Happa is a post-doc researcher in the Cyber Security Group at the University of Oxford, where he is currently developing novel visualisation approaches to improve situational awareness during cyber attacks. He obtained his BSc (hons) in Computing Science at the University of East-Anglia in 2006, after which he worked for NorCERT (Norwegian Computer Emergency Response Team) as an Intrusion Detection System analyst. In October 2007 he started a PhD at the University of Warwick where he developed a number of novel approaches to aid cultural heritage research using computer graphics and visualisation. He has since December 2011 worked at Oxford.

Ioannis Agrafiotis is a post-doctoral researcher at the Cyber Security Centre, in the Department of Computer Science at the University of Oxford. Currently, Ioannis is working on CyberVis, a Dstl funded project, aiming to create a tool to increase situational awareness in CERT environments. Before joining Cybervis, Ioannis worked on a research project aiming at developing a system able to reason for and anticipate insider-threat attacks (Corporate Insider Threat Detection: Cyber Security Inside and Out). He has also worked on the TSB and EPSRC funded TEASE project, researching how people understand information quality and provenance, and looking to develop mechanisms that effectively communicate trustworthiness. Prior to his post-doctoral experience, Ioannis completed a PhD in Engineering at the Warwick Manufacturing Centre (WMG), University of Warwick, focusing on formal methods for information privacy. During his doctoral studies Ioannis worked on the EnCoRe project through which he received an EPSRC studentship. He also holds an MSc in Analysis, Design and Management of Information Systems from the London School of Economics and Political Science in the UK and a BSc in Applied Informatics from the University of Macedonia in Greece.



Michael Goldsmith is a Senior Research Fellow at the Department of Computer Science and Worcester College, Oxford. With a background in Formal Methods and Concurrency Theory, Goldsmith was one of the pioneers of automated cryptoprotocol analysis. He has led research on a range of Technology Strategy Board and industrial or government-funded projects ranging from highly mathematical semantic models to multidisciplinary research at the social-technical interface. He is an Associate Director of the Cyber Security Centre, Co-Director of Oxford's Centre for Doctoral Training in Cybersecurity and one of the leaders of the Global Centre for Cyber Security Capacity-Building hosted at the Oxford Martin School, where he is an Oxford Martin Fellow.

Sadie Creese is Professor of Cybersecurity in the Department of Computer Science at the University of Oxford. She is Director of Oxford's Cyber Security Centre, Director of the Global Centre for Cyber Security Capacity Building at the Oxford Martin School, and a co-Director of the Institute for the Future of Computing at the Oxford Martin School. Her research experience spans time in academia, industry and government. She is engaged in a broad portfolio of cyber security research spanning situational awareness, visual analytics, risk propagation and communication, threat modelling and detection, network defence, dependability and resilience, and formal analysis. She has numerous research collaborations with other disciplines and has been leading inter-disciplinary research projects since 2003. Prior to joining Oxford in October 2011 Creese was Professor and Director of e-Security at the University of Warwick's International Digital Laboratory. Creese joined Warwick in 2007 from QinetiQ where she most recently served as Director of Strategic Programmes for QinetiQ's Trusted Information Management Division.