# Human-Machine Decision Support Systems for Insider Threat Detection

**Philip A. Legg**

**Abstract** Insider threats are recognised to be quite possibly the most damaging attacks that an organisation could experience. Those on the inside, who have privileged access and knowledge, are already in a position of great responsibility for contributing towards the security and operations of the organisation. Should an individual choose to exploit this privilege, perhaps due to disgruntlement or external coercion from a competitor, then the potential impact to the organisation can be extremely damaging. There are many proposals of using machine learning and anomaly detection techniques as a means of automated decision-making about which insiders are acting in a suspicious or malicious manner, as a form of large scale data analytics. However, it is well recognised that this poses many challenges, for example, how do we capture an accurate representation of normality to assess insiders against, within a dynamic and ever-changing organisation? More recently, there has been interest in how visual analytics can be incorporated with machine-based approaches, to alleviate the data analytics challenges of anomaly detection and to support human reasoning through visual interactive interfaces. Furthermore, by combining visual analytics and active machine learning, there is potential capability for the analysts to impart their domain expert knowledge back to the system, so as to iteratively improve the machine-based decisions based on the human analyst preferences. With this combined human-machine approach to decision-making about potential threats, the system can begin to more accurately capture human rationale for the decision process, and reduce the false positives that are flagged by the system. In this work, I reflect on the challenges of insider threat detection, and look to how human-machine decision support systems can offer solutions towards this.

P.A. Legg (✉)
Department of Computer Science and Creative Technologies,
University of the West of England, Bristol, UK
e-mail: phil.legg@uwe.ac.uk

I. Palomares et al. (eds.), *Data Analytics and Decision Support for Cybersecurity*,
Data Analytics, DOI 10.1007/978-3-319-59439-2_2

# 1 Introduction

It is often said that for any organisation, “employees are the greatest asset, and yet also the greatest threat”. The challenge of how to address this *insider threat* is one that is of increasing concern for many organisations. In particular, as our modern world is rapidly evolving, so to are the ways in that we conduct business and manage organisations, and so to are the ways in that those who choose to attack can do so, and succeed. In recent times there have been many high profile cases, including Edward Snowden [1], Bradley Manning [2], and Robert Hanssen [3]. According to the 2011 CyberSecurity Watch Survey [4], 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 [5], they report that 60% of frauds are committed by insiders, up from 55% in the previous year. Likewise, the 2012 Cybercrime report by PwC [6] states that the most serious fraud cases were committed by insiders. Of course, in all of these cases, these figures may not truly reflect the severity of the problem given that there are most likely many more that are either not detected, or not reported publicly. To define what is an ‘insider’, it is often agreed that this is somebody who compared to an outsider, has some level of knowledge and some level of access in relation to an organisation. Whilst employees are often considered to be the main focal point as insiders, by this definition there may be many others, such as contractors, stakeholders, former employees, and management, who could also be considered as insiders.

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. In their work, 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. Carnegie Mellon University has conducted much foundational work surrounding the insider-threat problem as part of their CERT (Computer Emergency Response Team), 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.

In this chapter, I discuss and reflect on my recent research that addresses the issues that surround insider threat detection. Some of this work has been previously published in various journals and conference venues. The contribution that this chapter serves is to bring together previous work on developing automated machine-based detection tools, and to reconsider the problem of insider threat detection with regards to how the human and the machine can work in tandem to identify malicious activity. Neither the human alone, nor the machine alone, is sufficient to address the problem in a satisfactory manner.

## 2 Related Works

There are a variety of published works on the topic of insider threat detection that range from theoretical frameworks for representing the problem domain, through to practical implementations of detection systems. As a research area, it is multi-disciplinary in nature, including computational design of detection algorithms, human behavioural modelling, business operations management, and ethical and legal implications of insider surveillance.

### *2.1 Models for Understanding the Problem of Insider Threat*

Legg et al. propose a conceptual model that can help organisations to begin thinking about how to detect and prevent insider attacks [11]. The model is based on a tiered approach that relates real-world activity, measurement of the activity, and hypotheses about the current threat. The model is designed to capture a broad range of attributes related to insider activity that could be characterised by some means. The tiered approach aims to address how multiple attributes from the real-world tier can contribute towards the collection of measurements that may prove useful for forming hypotheses (e.g., heavy workload, working late, and a developing disagreement with higher management, could result in a possible threat of sabotage). Nurse et al. [12] also propose a framework, this time for characterising insider threat activity. The framework is designed to help an analyst identify the various traits that surround insider threats, including the precipitating events that then motivate an attacker, and the identification of resources and assets that may be exploited as part of an attack. By considering these attributes, analysts may be able to ensure a full and comprehensive security coverage in their organisation.

Maybury et al. [13] developed a 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. Similarly, Colwill [14] 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, whilst Greitzer et al. [15] focus on incorporating inferred psychological factors into a modelling framework. The work by Brdiczka et al. [16] combine such psychological profiling with structural anomaly detection, to develop an architecture for insider-threat detection that demonstrates much potential for solving the problem.

In terms of measuring behaviours that may indicate a threat, Roy et al. [17] propose a series of metrics that could be used based on technical and behavioural observations. Schultz [18] 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. Althebyan and Panda [19] present a model for insider-threat prediction based on the insider's knowledge and the dependency of objects within the organisation. In the work of Sasaki [20], a trigger event is used to identify a change of behaviour, 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).

Bishop et al. [21] 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 [22]: embezzlement, social engineering, and password alteration. Other work such as Doss and Tejay [23] propose a model for insider-threat detection that consists of four stages: monitoring, threat assessment, insider evaluation and remediation. Liu et al. [24] propose a multilevel framework called SIDD (Sensitive Information Dissemination Detection) that incorporates network-level application identification, content signature generation and detection, and covert communication detection. More recently, Bishop et al. [25] extend their work to examine process modelling as a means for detecting insider attacks.

## 2.2 Approaches for Detecting Insider Threat

Agrafiotis et al. [26] explore the sequential nature of behavioural analysis for insider threat detection. The sequence of events is a critical aspect of analysis, since a single event in isolation may not be deemed as a threat, and yet in conjunction with other events, this may have much greater significance. As an example, an employee who is accessing sensitive company records would be of more concern if they had recently been in contact with a rival organisation, compared to an employee who may be acting as part of their job role requirement. They extend the work on sequential analysis in [27], where this scheme is then applied to characterise a variety of insider threat case studies that have been collated by the Carnegie Mellon University CERT.

Elmrabit et al. [28] study the categories and approaches of insider threat. They categorise different types of insider attack (e.g., sabotage, fraud, IP theft) against the CIA security principles (confidentiality, integrity, availability), and also against human factors (motive, opportunity, capability). They discuss a variety of tools in the context of insider threat detection, such as intrusion detection systems, honeytokens, access control systems, and security information and event management systems. They also highlight the importance of psychological prediction models, and security education and awareness, both of which are required by organisations in order to tackle the insider threat problem effectively. It is clear that technical measures alone are not sufficient, and that 'security as a culture' should be practiced by organisations wishing to address this issue successfully.

Parveen et al. [29] use stream mining and graph mining to detect insider activity in large volumes of streaming data, based on ensemble-based methods, unsupervised learning and graph-based anomaly detection. Building on this, Parveen and Thuraisingham [30] propose an incremental learning algorithm for insider threat detection that is based on maintaining repetitive sequences of events. They use trace files collected from real users of the Unix C shell, however this public dataset is relatively dated now. Buford et al. [31] use situation-aware multi-agent systems as part of a distributed architecture for insider threat detection. Garfinkel et al. [32] propose tools for media forensics, as means to detecting insider threat behaviour.

Eldardiry et al. [33] also propose a system for insider threat detection based on feature extraction from user activities, although they do not consider role-based assessments as part of their system. Senator et al. [34] propose to combine structural and semantic information on user behaviour to develop a real-world detection system. They use a real corporate database, gather as part of the Anomaly Detection at Multiple Scales (ADAMS) program, however due to confidentiality they can not disclose the full details and so it is difficult to compare against the work.

McGough et al. [35] propose a beneficial software system for insider threat detection based on anomaly detection of a user profile and their job role profile. Their approach also aims to incorporate human resources information, for which they describe a five states of happiness approach to assess the likelihood that a user may pose a threat. Nguyen and Reiher [36] propose a detection tool for insider threat that monitors system call activity for unusual or suspicious behaviour. Maloof and Stephens [37] propose a detection tool for when insiders violate need-to-know restrictions that are in place within the organisation. Okolica et al. [38] use Probabilistic Latent Semantic Indexing with Users to determine employee interests, which are used to form social graphs that can highlight insiders.

### *2.3 Insider Threat Visualization*

With regards to insider threat visualization, the technical report by Harris [39] discusses some of the issues related to visualizing insider threat activity. Nance and Marty [40] propose using bipartite graphs to identify and visualize insider threat activity where the nodes in the graph represent two distinct groups, such as user nodes and activity nodes, and the edges represent that a particular user has performed a particular activity. This approach is best suited for comparative analysis once a small group of users and activities have been identified, as scalability issues would soon arise in most real-world analysis tasks. Stoffel et al. [41] propose a visual analytics application for identifying correlations between different networked devices, based on time-series anomaly detection and similarity models. They focus primarily at the network traffic level, and so they do not currently consider other attributes related to insider threat such as file storage systems and USB connected devices. Kintzel et al. [42] use scalable glyph-based visualization using a clock metaphor to present an overview of the activity over time of thousands of hosts on a network. Zhao et al. [43] looked at anomaly detection for social media data and presented their visualization tool FluxFlow. Again, they make use of the clock metaphor as part of their visualization, which they combine with scaled circular glyphs to represent anomalous data points. Walton et al. [44] proposed QCATs (Multiple Queries with Conditional Attributes) as a technique for understanding and visualizing conditional probabilities in the context of anomaly detection.

### *2.4 Summary of Related Works*

From the literature it becomes clear to see that the topic of insider threat has been extensively studied from a variety of viewpoints. A number of models have been put forward for how one could observe and detect signs that relate to whether an insider is posing a threat, or has indeed already attacked. Likewise, a number of detection techniques have been proposed. However, it is difficult to assess their true value when some only consider a sub-set of activities, or do not provide validation in a real-world context. In the following sections, I discuss work that has been conducted in recent years on insider threat detection by colleagues and myself. In particular, I address both machine-based and human-based approaches for decision-making on the current threat posed by an individual. As part of this, I also describe the real-world validation study of the machine-driven decision process that was performed, and an active learning approach for combining human-machine decision-making using visual analytic tools. These contributions set the work apart from the wider body of research that exists on insider threat detection, by supporting both human and machine in the process of identifying malicious insiders.

## 3 Automated Detection of Insider Threats

The process of detecting insiders that pose suspicious or malicious activity is a complex challenge. Given the large volume of data that may exist about all users activity within an organisation, human methods alone will not prove scalable.

Instead, there is a need for the machine to make a well-informed decision about the threat posed by an individual, based on their observed activity, and this differs from what is deemed as normal behaviour.

### 3.1 Automated Detection Using User and Role-Based Profile Assessment

In the paper by Legg et al. [45], "Automated Insider Threat Detection System using User and Role-based Profile Assessment", an insider threat detection system is proposed that is capable of identifying anomalous activity of users, in comparison to their previous activity and in comparison to their peers . The detection tool is based upon the underlying principles of the conceptual model proposed in [11]. The paper demonstrates the detection tool using publicly-available insider threat datasets provided by Carnegie Mellon University CERT, along with ten synthetic scenarios that were generated by an independent team within the Oxford Cyber Security group. In the work, the requirements of the detection system are given that:

- The system should be able to determine a score for each user that relates to the threat that they currently pose.
- The system should be able to deal with various forms insider threat, including sabotage, intellectual property theft, and data fraud.
- The system should also be able to deal with unknown cases of insider threat, whereby the threat is deemed to be an anomaly for that user and for that role.
- The system should assess the threat that an individual poses based on how this behaviour deviates from both their own previous behaviour, and the behaviour exhibited by those in a similar job role.

The system comprises of five key components: data input streams, user and role-based profiling, feature extraction, threat assessment, and classification of threat. From the data streams that were available for the CMU-CERT scenarios, and for those developed by the Oxford team, the data typically represented the actions of 1000 employees over the period of 12 months, with data that captured login and logout information for PC workstations, USB device insertion and removal, file access, http access, and e-mail communications. Each user also has an assigned job role (e.g., technician, receptionist, or director), where those in a similar role are expected to share some commonality in their behaviour. The first stage of the system is to connect to the available data streams, and to receive data from each stream in the correct time sequence as given by the timestamp of each activity.

As data is received, this is utilised to populate a profile that represents each individual user, as well as a combined profile that represents a single role. The profiles are constructed in a consistent hierarchical fashion, that denotes the devices that have been accessed by the user, the actions performed on each of these devices,

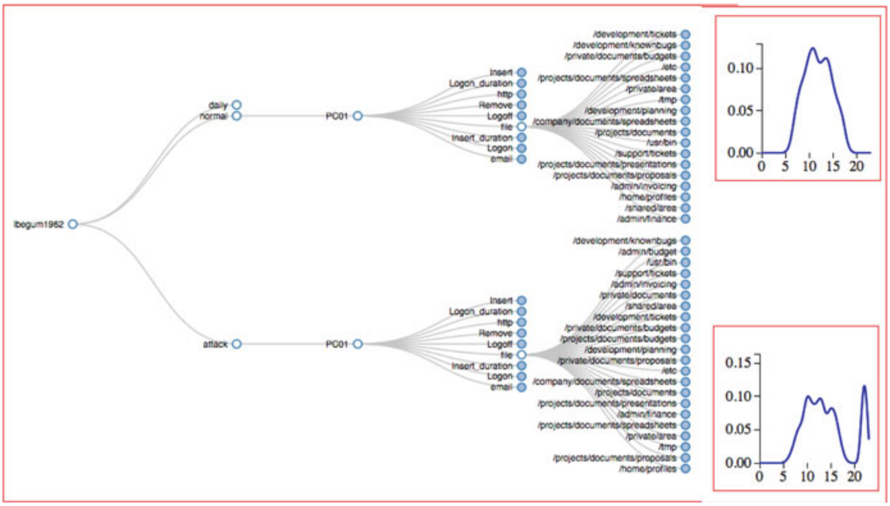


**Fig. 1** Tree-structured profiles of user and role behaviours. The root node is the user ID, followed by sub branches for ‘daily’, ‘normal’, and ‘attack’ observations. The next level down shows devices used, then activities performed, and finally, attributes for those activities. The probability distribution for normal hourly usage is given in the *top-right*, and the distribution for the detected attack is given in the *bottom-right*. Here it can be seen that the user has accessed a new set of file resources late at night

and the attributes associated with these actions. At each of these nodes in the profile, a time-series is constructed that denotes the occurrence of observations on a 24-h period. Figure 1 shows an interactive tree view of an individual user’s profile.

Once the system has computed the current daily profile for each user and for each role, the system can then extract features from the profile. Since the profile structure is consistent and well-defined, it means that comparisons between users, roles, or time steps can be easily made. In particular, the feature sets consists of three main categories: the user’s daily observations, comparisons between the user’s daily activity and their previous activity, and comparisons between the user’s daily activity and the previous activity of their role. The full set of features that are computed for each user is provided in [45]. These features include a variety of measurements that can be derived from the profiles, such as *New device for user*, *New attribute for activity for device for role*, *Hourly usage count for activity*, *USB duration for user*, and *Earliest logon time for user*. This set of features intends to be widely applicable for most organisations, although of course, there may be more bespoke features that are relevant for specific organisations that could also be incorporated. To perform the threat assessment, the system aims to identify variance between related features that may be indicative of a particularly anomaly. This is performed using Principal Component Analysis (PCA) [46]. PCA performs a projection of the features into lower dimensional space based on the amount of variance exhibited by each feature. From the user profiles, an $n \times m$ matrix is constructed for each

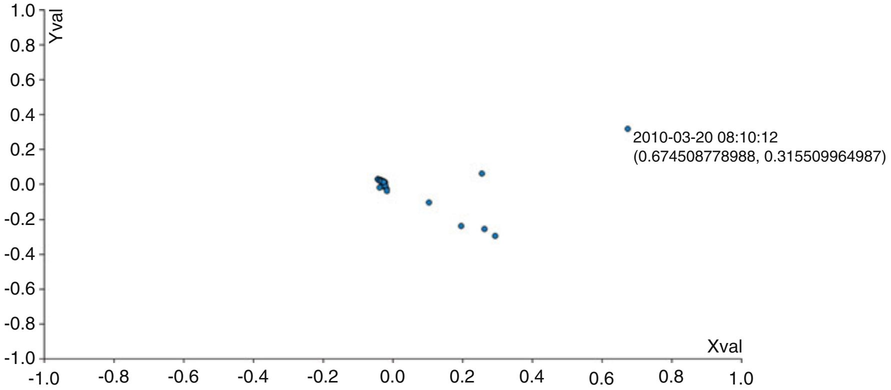
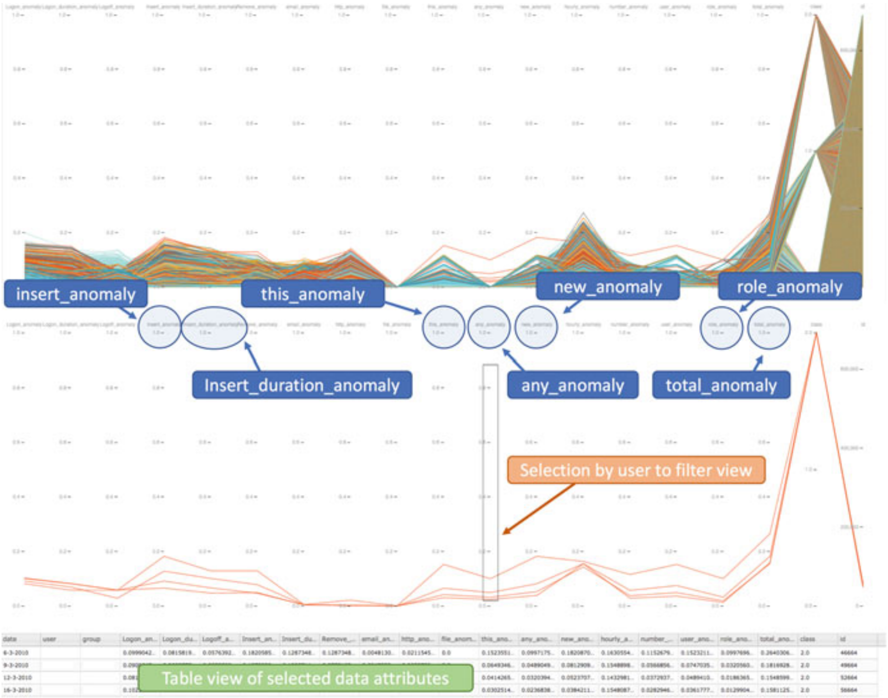


**Fig. 2** Example of using PCA for assessing deviation in user activity. Each point represents a single user for a single day (observation instance). Here, only a single user over time is shown to preserve clarity. The majority of points form a cluster at the centre of the plot. There are five observations that begin to move away from the general cluster. At the far right is a point observed on the 20th March 2010, that exhibits the most deviation in the user's behaviour

user, where $n$ is the total number of sessions (or days) being considered, and $m$ is the number of features that have been obtained from the profile. The bottom row of the matrix represents the current daily observation, with the remainder of the matrix being all previous observation features. Essentially, this process reduces an $n$-dimensional dataset to $n - 1$ dimensionality, based on the vector of greatest variance through the data. By performing this successively, we can reduce to 2 or 3 dimensions. Similar instances would be expected to group together, whilst instances that exhibit significant variation would appear far from other points in the space, where each point represents a single user on a single day. The system performs PCA using a variety of different feature combination that relate to a particular area of concern (e.g., web activity). Figure 2 shows the PCA decomposition for a detected insider threat. It can be seen that the most-part of activity clusters towards the centre, however over time, there are activities that diverge from this cluster, that represent daily observations where the user has performed significantly different. By considering the Euclidean distance of points from the centroid of the cluster, or from the point given by the role average, a measure of anomaly or deviation can be obtained for a given observation.

The score for each anomaly metric can then be analysed, for each user, for each day (e.g., *file_anomaly*, *total_anomaly*, *role_anomaly*). A parallel co-ordinate plot is used (Fig. 3), where each polyline shows a single user for a single day, against the various anomaly metrics (where each axis is a separate anomaly metric). In the example shown in Fig. 3, there is an observation that appears separate on the *any_anomaly* metric (this relates to activity that has been observed on **any** device—rather than just **this** device that it may have been observed on). By brushing the axis, the analyst can filter the view to show only this result. This reveals activity performed by a particular user of interest, who was found to be the malicious insider



| date | user | group | Logon_an… | Logon_du… | Logoff_a… | Insert_an… | Insert_du… | Remove_… | email_an… | http_ano… | file_anom… | this_ano… | any_ano… | new_ano… | hourly_a… | number_… | user_ano… | role_ano… | total_ano… | class | id |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6-3-2010 | | | 0.0999042… | 0.0815819… | 0.0576392… | 0.1820585… | 0.1287348… | 0.1287348… | 0.0048130… | 0.0211545… | 0.0 | 0.1523551… | 0.0997175… | 0.1820870… | 0.1630554… | 0.1152679… | 0.1523211… | 0.0997696… | 0.2640306… | 2.0 | 46664 |
| 9-3-2010 | | | 0.090[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0649346… | 0.0489049… | 0.0812909… | 0.1548898… | 0.0566856… | 0.0747035… | 0.0320563… | 0.1816828… | 2.0 | 49664 |
| 12-3-2010 | | | 0.081[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0414265… | 0.0320394… | 0.0523707… | 0.1432981… | 0.0372937… | 0.0489410… | 0.0186365… | 0.1548599… | 2.0 | 52664 |
| 16-3-2010 | | | 0.102[illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.0302514… | 0.0236838… | 0.0384211… | 0.1548087… | 0.0282946… | 0.0361777… | 0.0129904… | 0.1581125… | 2.0 | 56664 |

**Fig. 3** Parallel Coordinates view to show the corresponding profile features. An interactive table below the parallel co-ordinates view shows a numerical data view of the profile features that have been selected. Here, a particular user scores significantly higher than other users on one metric. Interactive brushing allows this to be examined in further detail

in the set of 1000 employees, who was accessing systems and using a USB storage device early in the morning. This was found to be different from the other users in the role of ‘Director’, who did not use USB storage devices, and very rarely used systems at this time of day.

This approach was found to be successful for the test scenarios from CMU-CERT and from Oxford. Unlike other supervised machine learning techniques, this approach requires no labelling of instances, making it easier to be deployed quickly and effectively within an organisation. Given the variety of ways that a user may exhibit activity that could be deemed a threat, classifying instances may be quite difficult in any case. Classification also assumes that future instances of a particular threat will closely match the currently-observed case, which may not be the case (e.g., exfiltration of data could be performed in a variety of ways). The challenge with the proposed approach is ensuring that the available data streams can capture the occurrence of the activity that is required to identify the threat. It also requires that the feature extraction supports all possible attack vectors that could be imagined that relates to the available data. Whilst a comprehensive set of features are provided, organisations may well find that they wish to incorporate additional features, or refine the existing features. By adopting an extensive approach for obtaining features, modification and creation of new features can be achieved with minimal reconfiguration.

## *3.2 Validation of Decision Support in Insider Threat Systems*

Whilst the previous section describes the detection tool in detail, perhaps the biggest challenge with developing insider threat tools is actually validating their performance in the real-world. Previously, synthetic data scenarios were used for developing and testing the tool. The paper by Agrafiotis et al. [47], "Validating an insider threat detection system: A real scenario perspective" extends this to report on the deployment of the detection tool in a real organisation.

The head of security for the particular organisation in question (not disclosed for confidentiality purposes) indicated that there had recently been an incident, which meant that there was a known insider that the system could be trialled against. From discussions with the head of security, the detection system was modified to account for three particular areas of interest: File-access logs, Patent DB interactions, and Directory DB interactions. Compared to the previous work of [45], the real-world organisation also presented scalability challenges. Here, file access logs provided more than 750,000 data entries per day, compared to approximately 20,000 in the synthetic examples. However, by only considering authenticated data entries resulted in a significant reduction in the amount of data from 750,000 to 44,000 entries per day. This was deemed as appropriate by the head of security, since this then provided user details, whereas the unauthenticated attempts were simply denied access to the system. They use five anomaly metrics (which are anonymised in their paper due to non-disclosure agreements), based on combinations of the features derived from the user activity profile.

For testing the system, they deployed the detection system over two different time periods (1 September to 31 October, and 1 December to 31 December), accounting for 16,000 employees. The December period contained no known cases, and served as a training period for establishing a baseline of normal activity to compare against. The period in September and October contained one known case of insider threat activity. When testing on this dataset, a number of false positives were generated as either medium or high alert, for 4129 individuals. However, on closer inspection, what the authors actually found was that the system produced approximately 0.5 alerts per employee per day. Yet, for one particular user, they generated 12 alerts in a single day. Sure enough, this particular user was the insider. Given the nature of a multi-national organisation, working times are likely to change significantly, and it is recognised by the head of security that users do not conform to strict working patterns on a regular basis. However, the fact that the system is capable of identifying the repeat occurrence of alerts for a user shows the strong potential of this system. Further work aims to consider how combinations of alerts across multiple days can be accumulated to better separate this particular individual from the other alerts that were generated. Nevertheless, the importance of this study is crucial for the continued development of insider threat detection tools, and demonstrates a real-world validation of how a system can be deployed in a large complex organisation.

# 4 Visual Analytics of Insider Threat Detection

The previous section looked at machine-based approaches for detecting insider threat activity from a large group of users. The capable of the machine to make such well-informed decisions is, by large, limited by the data that is available to the system, and how the system can understand and make sense of features that are derived from the user profiles. This section looks to explore how the human can utilise this knowledge that the machine generates, to further improve the decision-making process. Realistically, the disciplinary action of an insider would not be enforced until a security analyst and management have gathered the facts and can confidently identify that the user is a threat. Therefore, the machine-based approach serves to reduce the search space that the human analyst needs to consider, and then the human can explore this further, to understand *why* the machine may have arrived at such a decision, and whether the human agrees or disagrees with this decision.

## *4.1 Supporting Human Reasoning using Interactive Visual Analytics*

In the paper by Legg [48], "Visualizing the Insider Threat: Challenges and tools for identifying malicious user activity", it is shown how visualization can be utilised to better support the decision-making process of the detection tool. The system makes use of a visual analytics dashboard, supported by a variety of linked views including a interactive PCA (iPCA) view (as originally proposed by Jeong et al. [49]). The proposed dashboard, shown in Fig. 4, allows for overview summary statistics to be viewed, based on selection of time, users, and job roles. The iPCA view shows the measurement features on a parallel coordinates plot, and a scatter plot that represents the 2-dimensional PCA. In particular, what this offers is the ability to observe how the PCA space relates back to the original feature space. By dragging points in the scatter plot, a temporary black polyline is displayed on the parallel co-ordinates that shows the inverse PCA for the new dragged position, giving an interactive indication of how the 2-dimensional space maps to the original feature space. For the analyst, this can be particularly helpful to strengthen their reasoning for a particular hypothesis, such as for understanding what a particular cluster of points may be indicative of. The tool also features an activity view, where activities are plotted by time in a radial view (Fig. 5). This can be particularly useful for examining the raw

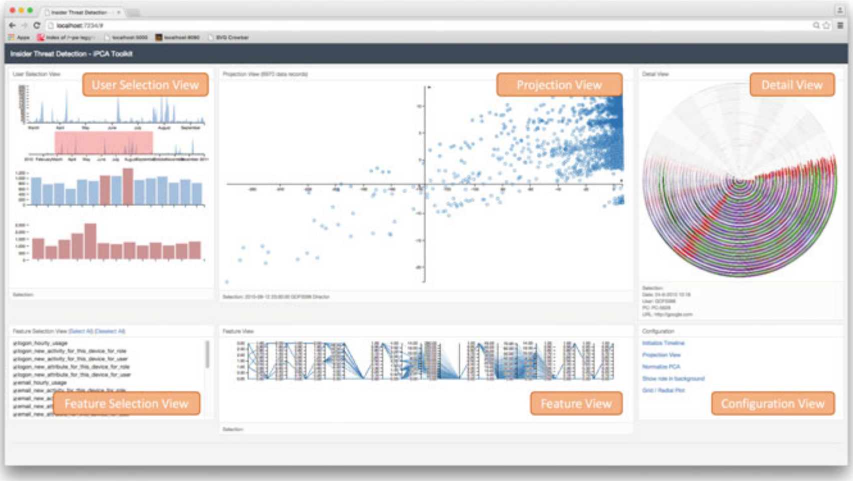
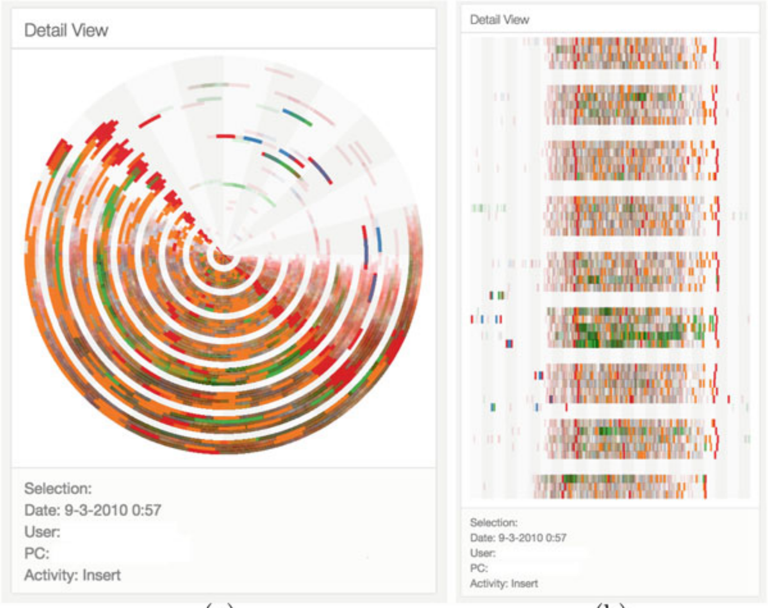


**Fig. 4** Layout of the visual analytics dashboard. The dashboard consists of four visualization views: User Selection, Projection, Detail, and Feature. The dashboard also has two supporting views for feature selection and configuration


Detail View

Selection:
Date: 9-3-2010 0:57
User:
PC:
Activity: Insert

(a)

Detail View

Selection:
Date: 9-3-2010 0:57
User:
PC:
Activity: Insert

(b)

**Fig. 5** Two variants of the detail view for exploring user activity, using (**a**) a circular plot (where time maps to angle and day maps to the radius), or (**b**) a rectangular grid plot (where time maps to the x-axis and day maps to the y-axis). *Colour* denotes the observed activity, and the selection pane provides detail of attributes. The role profile can be shown by the translucent *coloured* segments

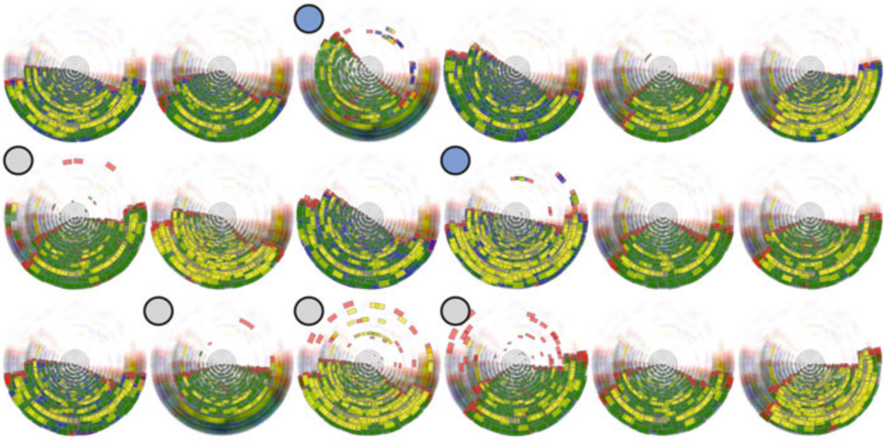


**Fig. 6** Assessment of 18 different user profiles within the same job role. Of the profiles, six profiles exhibit activity that occurs outside of the typical time period (marked by a *circle top-left* of profile). Two of the users also use USB devices (marked by a *blue circle*) during this non-typical time period, which may be of potential interest to the analyst. This view provides a compact and comparable overview of similar users

activity for days where there is significant deviation. Again, this links with the PCA view, so that when a user hovers on a point, the corresponding ring in the radial view is highlighted, and visa versa.

The detail view also forms the basis for a role overview mode, where the analyst can inspect the detail view of all users that exist within the same role. Figure 6 shows 18 users, where red indicates login/logout activity, blue indicates USB insertion/removal, green indicates e-mail activity, and yellow indicates web activity. As previously, a translucent background is used to represent the role profile, so that comparisons can be made between how the user compares against this. From this view, it can be seen that six of the users access resources outside of typical working hours for that particular role (marked by a circle top-left of profile), and two of these are making use of USB devices during these non-typical hours (marked by a blue circle). By visualizing the activity by means of this overview, it allows analysts to gain a clearer understanding of how the other users in the role perform, which can help further support their decision-making about the threat that is posed.

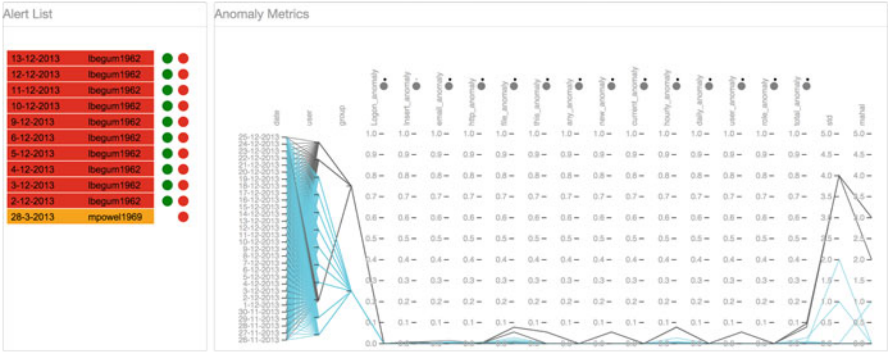
Visual Analytics provide an interface for analysts to visual explore the analytical results produced by the machine decision process. The ability to utilise the machine-based detection allows for an initial filtering process that alleviates the workload for the analyst, whilst the visual analytics approach then enables analysts to obtain a richer understanding of why the machine has given a particular result, without obscuring the data should the analyst decide that further analytics of additional data is required to fulfil their own decision process.

## 4.2 Active Learning for Insider Threat Detection

The visual analytics dashboard is a powerful interface that links the user to intuitive visual representations of the underlying data. Through interaction, the analyst can explore and delve deeper into this data, to support the development of their hypotheses on the intentions of an insider who may pose a threat to the organisation.

By exploiting the concept of a visual analytics loop [50], the user interaction can be utilised to inform the system, based on the new knowledge that they have obtained from viewing the current result. From a machine learning viewpoint, this is akin to the human providing online training labels, based on instances of particular interest. This concept of training on a small sample of key instances, as determined by the current result of the system (rather than providing a complete training set like in supervised learning) is referred to as *active learning* [51].

In the paper by Legg et al. [52], "Caught in the act of an insider attack: detection and assessment of insider threat", an active learning approach is proposed for refining the configuration of the detection system. Figure 7 shows the approach for introducing active learning into a visual analytics tool. As seen previously, a parallel co-ordinates plot is used to depict each user for each day. The plot can be configured to show a historical time windows (e.g., the last 30 days). The left-side shows a user alert list, where minor and severe alerts are show as orange and red respectively, and the date and user are given as the text label. If the analyst clicks on an alert, a tree-structured profile is displayed (Fig. 1) that allows them to explore deeper into *why* a particular observation has been flagged. In the tree profile, all previously-acceptable activity is shown under the *normal* node, whilst the current attack is shown under the *attack* node. In this example, it appears that the user has accessed a set of files late at night that they would not typically work with, hence why they have been flagged up in this case.



Alert List

| | | | |
|---|---|---|---|
| 13-12-2013 | lbegum1962 | ● | ● |
| 12-12-2013 | lbegum1962 | ● | ● |
| 11-12-2013 | lbegum1962 | ● | ● |
| 10-12-2013 | lbegum1962 | ● | ● |
| 9-12-2013 | lbegum1962 | ● | ● |
| 6-12-2013 | lbegum1962 | ● | ● |
| 5-12-2013 | lbegum1962 | ● | ● |
| 4-12-2013 | lbegum1962 | ● | ● |
| 3-12-2013 | lbegum1962 | ● | ● |
| 2-12-2013 | lbegum1962 | ● | ● |
| 28-3-2013 | mpowel1969 | | ● |

**Fig. 7** Detection system as a result of active learning. The analyst has rejected the alert on *mpowel1969* (shown by the removal of the accept option). This reconfigures the detection system to downgrade the anomaly associated with this result—in this case *insert_anomaly*—which can be observed by the *circular dials* by each anomaly metric. In addition to the alert list, the parallel co-ordinates can be set to present only the 'last 30 days', which provides a clear view of the detected insider *lbegum1962*

For the active learning component, the key here is that each label also has an accept or reject option (shown by the green and red circles to the right of the label). The user does not necessarily have to provide this information, however if they do, then the system is able to incorporate this knowledge based into the decision-making process. This is done by taking a weighted contribution from each feature, so that if a rejected result scores highly on a particular feature, then this feature can be down-weighted for this particular user, or role, or entire group, for a particular period of time. In this fashion, the burden of false positives can be alleviated for the analysts.

## 5 Future Directions for Insider Threat Detection

It should be apparent by now that there is substantial interest in the area of insider threat detection as a research discipline. Yet despite many proposed solutions, the problem continues to persist in many organisations. So why is this? Part of the challenge is security awareness. Many organisations are simply ill-equipped to gather and analyse such activity data. Others may choose not to invest in security until it is too late. Part of the challenge here is also how to transition the work from academic research into industrial practice. A number of spin-out companies are beginning to come about from insider-threat research, for which may begin to address this problem. So, what is it that organisations can be doing to protect themselves?

Perhaps the key element is for organisations to identify their most precious assets, and identifying features to represent these. Machine Learning routines can only form useful insight if working with appropriate data that is representative of the problem domain. A system is unlikely to detect that a user is about to steal many sensitive records if it knows nothing about a user's access to such records. Therefore, identifying which activity features an organisation is most concerned about is a vital step in any application of insider threat detection. It is also vital that appropriate visual analytic tools are in place for assessing the results of automated detection routines, so that the analyst can fully understand the reasoning behind why a particular individual has been flagged up as suspicious. Without this, humans are merely taking a machine's word for whether a individual should be disciplined. Given the severe consequence of false accusations, it is vital that the analyst has full confidence in a given decision.

Another emerging area of interest in combating the insider threat is analysing text communications. This raises many ethical and privacy concerns, although in a corporate environment it could be argued that this is a requirement of the role (e.g., employees working in national security would be expected to abide by such regulations). One proposal to provide analytics on textual data without exposing privacy concerns is to perform online linguistics analysis that can then be used to characterise the communication, rather than the raw text alone. In [53], the Linguistic Enquiry Word Count (LIWC) tool was used as a means of characterising psychological traits through use of language. The LIWC tool essentially provides dictionaries that relate particular words (or parts of words), to independent features (e.g., love, friend, hate, self). There has been much work in the psychology domain of relating LIWC features to OCEAN characteristics (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism) [54], and also the Dark Triad (Narcissism, Machiavellianism, Psychopathy) [55]. A visual analytics dashboard was developed for analysing the communications of multiple users against these features, that can be used to identify when there is significant change in a user's communication, and how this could imply a change in their psychological characteristics. Whilst this initial study demonstrated potential in this area, there is much work that remains to be done in how feasible a solution this can provide.

Another consider to make is who should be responsible for decision making in insider threat—human or machine? Given the severity of disciplinary action, it could be argued that a human should always need to intervene to inspect the result to ensure that this is valid before any disciplinary or legal action is taken. Then there is the issue of at what stage does the system intervene—should analysts operate as proactive or reactive? In a proactive environment, systems may attempt to predict the likelihood that a user will become an attacker, rather than a reactive environment that is detecting an already-conducted attack. Again, ethical concerns are raised such as whether a user would have conducted the attack if the system had not intervened at that time? Lab-based experimentation on such scenarios can only take us so far in understanding these problems, and the concept that an employee can be disciplined for an action that they are yet to perform is very much seen as the work of science fiction (much like that of *Minority Report*). However, what is required is the collaboration and cooperation between organisations, and also with academia, to continue to experiment and continue to develop tools that can alleviate and support the demands of the analyst in their decision-making process.

## 6 Summary

In this chapter, I have considered the scope of insider threat detection, and how systems can be developed to enable human-machine decision support such that well-informed decisions can be made about insider threats. Whilst a wide range of work exists on the topic, there is still much work to be done to combat the insider threat. How can detection systems be provided with accurate and complete data? How can detection systems extend beyond 'cyber' data sources, to build a more complete representation of the organisation? How should the psychology of insiders be accounted for, to understand their motives and intentions in their normal practice, and understand how these may change and why? Then there are the ethical concerns that need to be addressed—if employees are being monitored, how will this affect staff morale? Will they simply find alternative ways to circumvent protective measures?

The research shows much potential in being able to combat this problem, however, it also reveals the importance of the human aspects of security. As stated earlier, “employees are the greatest asset, and yet also the greatest threat”, and it is fair that this has never been so true as it is in our modern society today. Technology is enhancing how society operate, and yet it is also providing new means for disgruntled insiders to attack. At the same time, insiders acting in physical space are also becoming more creative in their planning. This begins to illustrate how the boundaries between online and offline worlds are beginning to blur, and cyber is just another factor in the much larger challenge of organisation security. Yet, with continued efforts in the development of new security technologies, we can better support the decision-making process between man and machine, to combat the challenge of insider threat detection.

**Acknowledgements** Many thanks to my colleagues from Oxford Cyber Security, Dr. Ioannis Agrafiotis, Dr. Jassim Happa, Dr. Jason Nurse, Dr. Oliver Buckley (now with Cranfield University), Professor Michael Goldsmith, and Professor Sadie Creese, with whom my early work on insider threat detection was carried out with.

## References

1. BBC News. Profile: Edward Snowden, 2013. http://www.bbc.co.uk/news/world-us-canada-22837100.
2. Guardian. Bradley manning prosecutors say soldier ‘leaked sensitive information’, 2013. http://www.guardian.co.uk/world/2013/jun/11/bradley-manning-wikileaks-trial-prosecution.
3. FBI. Robert Philip Hanssen Espionage Case, 2001. http://www.fbi.gov/about-us/history/famous-cases/robert-hanssen.
4. 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.
5. Kroll and Economist Intelligence Unit. Annual Global Fraud Survey. 2011/2012, 2012.
6. PricewaterhouseCoopers LLP. Cybercrime: Protecting against the growing threat - Events and Trends, 2012.
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, Arlington, Virginia, USA*. RAND, 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. *Security & Privacy, IEEE*, 6(1):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*. Addison-Wesley Professional, 1st edition, 2012.
10. L. Spitzner. Honeypots: catching the insider threat. In *Proc. of the 19th IEEE Computer Security Applications Conference (ACSAC’03), Las Vegas, Nevada, USA*, pages 170–179. IEEE, December 2003.
11. 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*, 4(4):20–37, 2013.

12. J. R. C. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. T. Wright, and M. Whitty. Understanding insider threat: A framework for characterising attacks. In *Security and Privacy Workshops (SPW), 2014 IEEE*, pages 214–228, May 2014.
13. 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 International Conference on Intelligence Analysis, McLean, Viginia, USA*. MITRE, May 2005.
14. C. Colwill. Human factors in information security: The insider threat who can you trust these days? *Information Security Technical Report*, 14(4):186–196, 2009.
15. F. L. Greitzer and R. E. Hohimer. Modeling human behavior to anticipate insider attacks. *Journal of Strategic Security*, 4(2):25–48, 2011.
16. 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 IEEE Symposium on Security and Privacy Workshops (SPW'12), San Francisco, California, USA*, pages 142–149. IEEE, May 2012.
17. K. R. Sarkar. Assessing insider threats to information security using technical, behavioural and organisational measures. *Information Security Technical Report*, 15(3):112–133, 2010.
18. E. E. Schultz. A framework for understanding and predicting insider attacks. *Computers and Security*, 21(6):526–531, 2002.
19. Q. Althebyan and B. Panda. A knowledge-base model for insider threat prediction. In *Proc. of the IEEE Information Assurance and Security Workshop (IAW'07), West Point, New York, USA*, pages 239–246. IEEE, June 2007.
20. T. Sasaki. A framework for detecting insider threats using psychological triggers. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 3(1/2): 99–119, 2012.
21. 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*, pages 1–12. ACM, September 2008.
22. M. Bishop, S. Engle, S. Peisert, S. Whalen, and C. Gates. Case studies of an insider framework. In *Proc. of the 42nd Hawaii International Conference on System Sciences (HICSS'09), Waikoloa, Big Island, Hawaii, USA*, pages 1–10. IEEE, January 2009.
23. G. Doss and G. Tejay. Developing insider attack detection model: a grounded approach. In *Proc. of the IEEE International conference on Intelligence and security informatics (ISI'09), Richardson, Texas, USA*, pages 107–112. IEEE, June 2009.
24. Y. Liu, C. Corbett, K. Chiang, R. Archibald, B. Mukherjee, and D. Ghosal. SIDD: A framework for detecting sensitive data exfiltration by an insider attack. In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pages 1–10, Jan 2009.
25. M. Bishop, B. Simidchieva, H. Conboy, H. Phan, L. Osterwell, L. Clarke, G. Avrunin, and S. Peisert. Insider threat detection by process analysis. In *IEEE Security and Privacy Workshops (SPW)*. IEEE, 2014.
26. I. Agrafiotis, P. A. Legg, M. Goldsmith, and S. Creese. Towards a user and role-based sequential behavioural analysis tool for insider threat detection. *J. Internet Serv. Inf. Secur.(JISIS)*, 4(4):127–137, 2014.
27. I. Agrafiotis, J. R. C. Nurse, O. Buckley, P. A. Legg, S. Creese, and M. Goldsmith. Identifying attack patterns for insider threat detection. *Computer Fraud & Security*, 2015(7):9 – 17, 2015.
28. N. Elmrabit, S. H. Yang, and L. Yang. Insider threats in information security categories and approaches. In *Automation and Computing (ICAC), 2015 21st International Conference on*, pages 1–6, Sept 2015.
29. P. Parveen, J. Evans, Bhavani Thuraisingham, K.W. Hamlen, and L. Khan. Insider threat detection using stream mining and graph mining. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 1102–1110, Oct 2011.
30. P. Parveen and B. Thuraisingham. Unsupervised incremental sequence learning for insider threat detection. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pages 141–143, June 2012.

31. J. F. Buford, L. Lewis, and G. Jakobson. Insider threat detection using situation-aware mas. In *Proc. of the 11th International Conference on Information Fusion*, pages 1–8, 2008.
32. S. L. Garfinkel, N. Beebe, L. Liu, and M. Maasberg. Detecting threatening insiders with lightweight media forensics. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 86–92, Nov 2013.
33. H. Eldardiry, E. Bart, Juan Liu, J. Hanley, B. Price, and O. Brdiczka. Multi-domain information fusion for insider threat detection. In *Security and Privacy Workshops (SPW), 2013 IEEE*, pages 45–51, May 2013.
34. T. E. Senator, H. G. Goldberg, A. Memory, W. T. Young, B. Rees, R. Pierce, D. Huang, M. Reardon, D. A. Bader, E. Chow, et al. Detecting insider threats in a real corporate database of computer usage activity. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1393–1401. ACM, 2013.
35. A. S. McGough, D. Wall, J. Brennan, G. Theodoropoulos, E. Ruck-Keene, B. Arief, C. Gamble, J. Fitzgerald, A. van Moorsel, and S. Alwis. Insider threats: Identifying anomalous human behaviour in heterogeneous systems using beneficial intelligent software (ben-ware). In *Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats*, MIST '15, pages 1–12, New York, NY, USA, 2015. ACM.
36. N. Nguyen and P. Reiher. Detecting insider threats by monitoring system call activity. In *Proceedings of the 2003 IEEE Workshop on Information Assurance*, 2003.
37. M. A. Maloof and G. D. Stephens. elicit: A system for detecting insiders who violate need-to-know. In Christopher Kruegel, Richard Lippmann, and Andrew Clark, editors, *Recent Advances in Intrusion Detection*, volume 4637 of *Lecture Notes in Computer Science*, pages 146–166. Springer Berlin Heidelberg, 2007.
38. J. S. Okolica, G. L. Peterson, and R. F. Mills. Using plsi-u to detect insider threats by datamining e-mail. *International Journal of Security and Networks*, 3(2):114–121, 2008.
39. M. Harris. Visualizing insider activity and uncovering insider threats. Technical report, 2015.
40. K. Nance and R. Marty. Identifying and visualizing the malicious insider threat using bipartite graphs. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, pages 1–9, Jan 2011.
41. F. Stoffel, F. Fischer, and D. Keim. Finding anomalies in time-series using visual correlation for interactive root cause analysis. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, VizSec '13, pages 65–72, New York, NY, USA, 2013. ACM.
42. C. Kintzel, J. Fuchs, and F. Mansmann. Monitoring large ip spaces with clockview. In *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, VizSec '11, pages 2:1–2:10, New York, NY, USA, 2011. ACM.
43. J. Zhao, N. Cao, Z. Wen, Y. Song, Y. Lin, and C. Collins. Fluxflow: Visual analysis of anomalous information spreading on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1773–1782, Dec 2014.
44. S. Walton, E. Maguire, and M. Chen. Multiple queries with conditional attributes (QCATs) for anomaly detection and visualization. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, VizSec '14, pages 17–24, New York, NY, USA, 2014. ACM.
45. P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese. Automated insider threat detection system using user and role-based profile assessment. *IEEE Systems Journal*, PP(99):1–10, 2015.
46. I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.
47. I. Agrafiotis, A. Erola, J. Happa, M. Goldsmith, and S. Creese. Validating an insider threat detection system: A real scenario perspective. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 286–295, May 2016.
48. P. A. Legg. Visualizing the insider threat: challenges and tools for identifying malicious user activity. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on*, pages 1–7, Oct 2015.
49. D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. ipca: An interactive system for pca-based visual analytics. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'09, pages 767–774, Chichester, UK, 2009. The Eurographs Association; John Wiley & Sons, Ltd.

50. P. A. Legg, D. H. S. Chung, M. L. Parry, R. Bown, M. W. Jones, I. W. Griffiths, and M. Chen. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2109–2118, Dec 2013.
51. B. Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.
52. P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese. Caught in the act of an insider attack: detection and assessment of insider threat. In *Technologies for Homeland Security (HST), 2015 IEEE International Symposium on*, pages 1–6, April 2015.
53. P. A. Legg, O. Buckley, M. Goldsmith, and S. Creese. Visual analytics of e-mail sociolinguistics for user behavioural analysis. *Journal of Internet Services and Information Security (JISIS)*, 4(4):1–13, 2014.
54. J. S. Wiggings. *The five factor model of personality: Theoretical perspectives*. Guilford Press, 1996.
55. D. L. Paulhus and K. M. Williams. The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of research in personality*, 36(6):556–563, 2002.