Insider-Threat Detection using Gaussian Mixture Models and Sensitivity Profiles

Kholood Al Tabash[a,b], Jassim Happa[a]

[a]*Department of Computer Science, University of Oxford, Oxford, UK*
[b]*EXPEC Computer Center, Saudi Aramco, Dhahran, Saudi Arabia*

Abstract

The insider threat is one of the most challenging problems to detect due to its complex nature and significant impact on organisations. Insiders pose a great threat on organisations due to their knowledge on the organisation and its security protocols, their authorized access to the organisation's resources, and the difficulty of discerning the behaviour of an insider threat from a normal employee's behavior [1]. As a result, the insider-threat field faces the challenge of developing detection solutions that are able to detect threats without generating a great number of false positives, and are able to take into consideration the non-technical aspect of the problem. This paper introduces a novel automated anomaly detection method that uses Gaussian Mixture Models for modelling the normal behaviour of employees to detect anomalous behaviour that may be malicious. The paper also introduces a novel approach to insider-threat detection that capitalises on the knowledge of security experts during analysis using visual analytics and sensitivity profiles which is a novel approach to re-contextualise detection output by considering outside, qualitative, non-technical factors that analysts may be privy to, but not the detection method. A feasibility study with experts in threat detection was conducted to evaluate the detection performance of the proposed solution and its usability. The results demonstrate the success of designing a solution that builds on the knowledge of security experts during analysis and reduces the number of false positives generated by automated anomaly detection. The work presented in the paper also demonstrates the potential of introducing more methods for capitialising on the knowledge of security experts to improve the false negative rate, and the potential of designing sensitivity profiles.

Keywords: insider-threat detection, machine learning, gaussian mixture models, visual analytics, sensitivity profiles, feasibility study

1. Introduction

An insider is a *"current or former employee, contractor, or business partner"* with authorized access to an organisation's resources [2]. An insider threat poses a risk to the organisation in question. Insider threat is one of the challenging areas in cyber security due to its complexity and massive impact on organisations. In a research report by the Ponemon Institute on the costs of cyber-security attacks, insider threat ranks as top in terms of cost [3]. However, its impact is not limited to financial losses but may jeopardize the safety of individuals and the reputation of organisations [2]. The problem of detecting insider threats is especially challenging for the difficulty of detecting and confirming insider-attacks. The challenge arises because of the knowledge insiders have on the organisation and its security protocols, their authorized access to the organisation's resources, and the difficulty of discerning the behavior of an insider threat from a normal employee's behavior [1]. As a result, extensive research has been dedicated to developing solutions for insider-threat detection.

There are three main categories of insider-attacks which are: theft of intellectual property or confidential information, sabotage of IT resources, and fraud. Insider-attacks are executed in various ways depending on several factors, such as the type of attack, and the insider's level of access [2, 4]. An insider aiming at sabotaging company resources will have a different attack pattern than one who is after intellectual property. Furthermore, there are various types of indicators of insider threat because attacks vary in execution and intent. For example, indicators of insider fraud are unusual or frequent access to financial systems and personnel data. On the other hand, indicators of a potential IT sabotage are file deletions or alterations (data integrity), and download of executable files. A common insider-threat indicator is a change in the normal behaviour of an insider.

Among the main challenges faced in insider-threat detection is the high rate of false positives, and the incorporation of the human and non-technical aspect of the problem [1, 5]. Due to the nature of the problem, challenges arise in distinguishing anomalies triggered by malicious insiders and those genuinely reflecting a change in behavior posing a challenge in handling false positives and confirming attacks [1]. Often, the information that is needed in anomaly detection methods to reduce the uncertainty is not available due to ethical and privacy concerns [6]. Additionally, the problem is unique due to the importance of the role of the non-technical aspect of the threat; an aspect that is challenging to incorporate in detection solutions [5, 7, 8].

1.1. Paper Contributions

This paper has five key contributions: three detection contributions, one visual analytics dashboard and one feasibility

Email addresses: kholood.tabash@aramco.com (Kholood Al Tabash), jassim.happa@cs.ox.ac.uk (Jassim Happa)

study. In the effort to decrease false positive and false negative rates, while increasing true positive rates, we: 1) investigate how **Gaussian Mixture Models (GMMs)** can be used to conduct insider-threat detection and present detection-rate results, 2) propose that the **human analyst's knowledge and experience** is a rich resource to be capitalized upon during data analysis [9] by incorporating **non-technical indicators of insider threat** that unlike other methods does not require analysis of sensitive data on employees and 3) provide the capability of adding **contextual information during analysis** of detected anomalies. Finally, we 4) present the design and implementation of a **visualization dashboard** to communicate GMM detection and 5) present an in-depth **feasibility study** in which we let cybersecurity analysts use the tool, provide feedback and test the tool's usability.

1.2. Organisation of The Paper

This paper is organised as follows: Section 2 reviews related work in insider-threat detection and visualization. Then, Section 3 provides background on Gaussian Mixture Models. Background is followed by a discussion of the proposed approach in Section 4. Then, Sections 5 and 6 present the testing of the system and results followed by Section 7 that discusses the observed results. Finally, Section 8 concludes the work presented and discusses future work.

2. Related Work

Insider threat poses its own set of detection challenges from external attacks because of the significant role of the non-technical factors of the threat, in addition to the legitimate access insiders have within an organisation and their knowledge of the organisation and its resources [2, 10, 8, 11, 12]. Cappelli et al. wrote in their book, which is the result of one of the most recognised research on insider threat with more than 700 documented cases of insider threats, stating that *"insider threat cannot be prevented and detected with technology alone"* [2, p. 14]. McCormac et al. [8] also highlight that technical indicators are not sufficient in insider-threat detection. This is a key motivator of our work. Therefore, insider-threat detection methods differ from those of external attacks. They are either signature-based, or anomaly-detection methods [1].

2.1. Signature-based

Signature-based methods detect known real-world insider attacks. They are limited in detecting attacks that have implemented policies which when violated, alarms are triggered. For instance, Agrafiotis et al. [13] develop a tripwire grammar to detect actions that are indicators of insider threat based on designed policies on alarming behaviours, and attack-patterns. IBM [14] uses a similar approach as part of the IBM QRadar SIEM solution [15] through the implementation of offences. Offences are designed to detect threats in general, and may be used for detecting steps of known insider attacks. Bishop et al. [16] take a different approach by developing a solution based on the targets of insider attacks. The authors propose using process modelling to identify how a process may be attacked and build countermeasures accordingly. This approach is highly dependent on the successful design of a process model that identifies the vulnerabilities of the process and possible attack targets. It is also limited to detecting attacks on the proposed targets.

2.2. Anomaly Detection

Anomaly detection is capable of detecting unknown and new types of attacks under the assumption that a malicious behaviour deviates from normal behaviour. It is implemented by checking against a normal behaviour and generating alarms when deviations occur. Some detection-solutions consider non-technical indicators of insider threat. Non-technical indicators, such as the psychological state of the insider are of crucial value to insider-threat detection [2, 12]. Therefore, work has been done to incorporate their analysis in insider-threat detection systems. For example, Brdiczka et al. [17] use structural anomaly detection relying on graph analysis and additional techniques to learn the normal behaviour of employees. They also use psychological profiling to take into consideration an insider's intention, which is a non-technical indicator, with the aim of reducing false positives generated by monitoring technical indicators. Chen et al. [5] also consider the intention of an insider as an indicator and propose a detection system that uses probabilistic modelling. Their solution is designed to predict the success of an attack by conducting behavioural analysis using probabilistic model checking. Prediction is done after a potential insider has been identified through intentional analysis using Bayesian networks. Both Brdiczka et al. [17] and Chen et al. [5] apply automated analysis of non-technical indicators of insider threat requiring the collection of sensitive data, such as the contents of email communications to be used for sentiment analysis.

Moreover, some anomaly-detection methods are developed to detect a certain type of insider threat. For example, Zhang et al. [18] propose a solution to analyse document-access behaviour to classify users based on the contents of accessed documents. Each user is identified by the type of documents they usually access. Anomaly detection checks for deviations from historical and current behaviours of the user, and the behaviour of the community using the Naive Bayes algorithm and correlation matrices. This approach is limited to monitoring a single indicator, which is accessed files, of a specific type of insider threat, which is information leakage. Other detection methods aim at detecting threats to a specific resource in an organisation. For example, Senator et al. [19] develop a solution to detect threats to a database based on database-access behaviour. Their solution is an example approach that is limited to protecting a specific resource which is the corporate database, but which addresses multiple indicators by implementing various types of anomaly-detection algorithms to tackle the low signal-to-noise ratio challenge in insider threat.

Finally, some detection methods are designed to learn a normal behaviour of employees from their online activities. For instance, a more relevant work to the approach taken in this paper in insider-threat detection includes the work of Legg et al. [20]. They developed an automated detection system that uses PCA to detect anomalies. They compute hourly feature vec-

tors on the activities of employees and build a 24-hour matrix of activities. Then, PCA is applied to project the multivariate vectors into a 2D space based on the maximum variance exhibited by features. Anomaly detection then measures the distance of points in the projected space from the origin. The chosen anomaly-detection method is difficult to interpret limiting security analysts' capability of gaining insight on the decision-making process of the method while investigating generated alarms.

Rashid et al. [21] is the work most similar to the proposed solution presented in this paper in recognizing the importance of anomaly explanation. They use Hidden Markov Models to learn the normal behaviour of employees and analyse deviations from the learned behaviour to detect insider threat. The authors highlight that their model offers the advantages of learning parameters from the dataset that describe an employee's behaviour. Their model is also advantageous in learning from data that is sequential in nature. However, the computational cost of training the models increases as the number of states captured increases, while the effectiveness of the method in detecting insider threat is highly impacted by the number of states. Moreover, Song et al. [22] use Gaussian Mixture Models for modelling the behaviour of users for insider threat and masquerade detection. They compare their Gaussian Mixture Models to several other machine learning methods and find it superior in achieving higher accuracy values. However, their model is applied on system-level events, such as process creation, intended for a biometric identification of a user as opposed to insider threat detection specifically, which this paper investigates.

2.3. Visual Analytics

Visual Analytics is defined as *"the science of analytical reasoning facilitated by interactive visual interfaces"* [23, p. 4]. It allows people to gain insight from large amounts of data that is otherwise hard to grasp. The visual analytics process is a constant interaction between automated data analysis, and human knowledge and skill through the usage of visualizations as a communication tool [24]. The process starts with data which may be from multiple sources and in different formats that must be transformed before being processed. Then, data may be visualized directly or processed through machine learning algorithms, for example, before visualization. Throughout the process, analysts are a crucial element that interact with the system through the usage of visualizations to provide and gain knowledge. The knowledge of analysts may be fed back into the process to influence the parameters of automated data analysis models or the data itself, by removing uninformative elements, for example.

Insider threat is a classical problem faced in the era of Big Data and information overload where massive amounts of data must be analysed to make decisions. Although automated data analysis tools provide the means to handle information extraction and automated decision making, analysts are often faced with the challenge of *"analysing [the resulting] analyses"* [24]. They are only exposed to the results provided by automated methods and often without being able to benefit from the knowledge gained by those tools during data analysis. Therefore, additional means of analysis are needed which allow the communication of information in the large amounts of data to analysts while providing them with expedited data analysis and decision making which machine learning and data analysis tools are valued for. Visual analytics is a solution which capitalises on the strengths of both humans and machines during data analysis [24].

Limited work has been done in insider-threat visualization. Raffael Marty [25] devotes a chapter in the book *Applied Security Visualization* to insider threat providing example visualizations that include link graphs and treemaps for showing user activity. However, the suggested approach relies on scoring employees based on a set of precursors making it unsuitable for detecting new attacks. Link graphs are not scalable and limited to the analysis of a small number of employees. Nance and Marty [26] use bipartite graphs for the detection and analysis of insider threat. Their visualization method assigns a group of nodes to activities and another group to job roles. Each edge from one group to another represents an activity done by an employee. Anomalous behaviours are indicated by coloured edges. Due to scalability limitations of the chosen visualization method, their solution is limited to small organisations with a small number of employees and job roles. Moreover, Philip Legg [27] proposes a visual analytics solution for insider-threat detection implemented as an interactive, multi-view dashboard. The dashboard builds on the results of the automated anomaly-detection component. However, the chosen anomaly-detection method, which is PCA, is difficult to interpret especially when applied to multi-dimensional data adding a challenge when analysing the results of anomaly detection.

3. Preliminaries

3.1. Mixture Models

Mixture Models (MM) are a type of Latent Variable Models (LVM) with the latent variables being the assignment of data points to the model's components [28, p. 339]. MMs provide the capability of modelling *"complex probability distributions"* by using a mixture of distributions, which are the components of a model [29, p. 423]. It results in clustering observations in a dataset based on their shared probability distribution, and models the entire dataset as a linear combination of the different probability distributions. MMs are implemented using a probabilistic approach which, unlike other clustering techniques such as K-means, is distinguished by the soft-assignment of observations to clusters enabling the model to assign a degree of uncertainty during clustering [29, 30]. This approach is advantageous in not only providing a degree of anomaliness but also in explaining why observations are classified as anomalous [30]. MMs are unexplored in the insider threat space.

Each observation in the dataset is assigned a probability that is a linear combination of the probabilities computed by each component in the model. Given a dataset $X = \{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_N}\}$, where N is the number of observations, and $\mathbf{x}$ is a D-dimensional vector, the probability of an observation is computed as follows:

$$p(\mathbf{x}) = \sum_{k=1}^{K} p(k)\, p(\mathbf{x} \mid k) \tag{1}$$

Each probability distribution, $p(\mathbf{x} \mid k)$, is a component, and K is the number of components [29, 31]. $p(k)$, or π_k, is the mixing coefficient which is the *prior* probability of the k^{th}.

The *posterior* probability of the component is, then, calculated using Bayes rule as shown in Equation 2. The *posterior* probability refers to the uncertainty in the parameters of the model after observing the data, whereas the *prior* probability models the assumptions made about the parameters before observing the data [29, p. 22]. Moreover, the model then calculates the probability of an observation $\mathbf{x_i}$ being generated by a component k. This enables the model to achieve a soft-assignment of observations to components based on the resulting probabilities of each component. The *posterior* probability is referred to as the *responsibility* of the component since it indicates the responsibility of the parameters of the component for generating the input $\mathbf{x}$ [29, 31]. It is normalized to achieve a value between 0 and 1 as shown in Equation 2.

$$p(k \mid \mathbf{x}) = \frac{p(k)\, p(\mathbf{x} \mid k)}{\sum_{j=1}^{K} p(j)\, p(\mathbf{x} \mid j)} \tag{2}$$

Observations are clustered based on the resulting probabilities $p(k \mid \mathbf{x})$. The component with the largest $p(k \mid \mathbf{x})$ is the cluster to which $\mathbf{x}$ belongs.

3.1.1. Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is a MM with each component being a Gaussian distribution [28, p. 341]. More generally, a GMM with K components is described as a superposition of K Gaussian distributions:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k\, \mathcal{N}(\mathbf{x} \mid \mu_\mathbf{k}, \mathbf{\Sigma_k}) \tag{3}$$

$\mu_\mathbf{k}, \mathbf{\Sigma_k}$ are the mean and covariance of the component k. The joint probability density function of a D-dimensional vector $\mathbf{x}$ is computed as follows [28, p. 46]:

$$\mathcal{N}(\mathbf{x} \mid \mu, \mathbf{\Sigma}) = \frac{1}{(2\,\pi)^{\frac{D}{2}}} \frac{1}{|\mathbf{\Sigma}|^{\frac{1}{2}}} \exp^{\frac{-1}{2}\,(\mathbf{x}-\mu)^T\,\mathbf{\Sigma}^{-1}\,(\mathbf{x}-\mu)} \tag{4}$$

Where μ is the D-dimensional mean vector, $\mathbf{\Sigma}$ is the $D \times D$ covariance matrix. Therefore, the responsibility of the component k is calculated as follows:

$$p(k \mid \mathbf{x}) = \frac{\pi_k\, \mathcal{N}(\mathbf{x} \mid \mu_\mathbf{k}, \mathbf{\Sigma_k})}{\sum_{j=1}^{K} \pi_j\, \mathcal{N}(\mathbf{x} \mid \mu_\mathbf{j}, \mathbf{\Sigma_j})} \tag{5}$$

The parameters of the model π_k, $\mu_\mathbf{k}$, $\mathbf{\Sigma_k}$ $\forall k$ are learned using the Expectation Maximization (EM) algorithm by maximizing the log likelihood of the dataset (Equation 7).

$$log\; p(X \mid \theta) = \sum_{n=1}^{N} log \sum_{k=1}^{K} p(k)\, p(\mathbf{x} \mid \theta_k) \tag{6}$$

$$log\; p(X \mid \pi, \mu, \mathbf{\Sigma}) = \sum_{n=1}^{N} log \sum_{k=1}^{K} \pi_k\, \mathcal{N}(\mathbf{x_n} \mid \mu_\mathbf{k}, \mathbf{\Sigma_k}) \tag{7}$$

EM is used because a MM has latent variables and because maximum likelihood does not have a closed-form solution due to the summation term inside the log [29]. EM algorithm is applied in an iterative manner alternating between an Expectation (E) step and Maximization (M) step. It is summarized as follows:

1. The parameters $\pi, \mu, \mathbf{\Sigma}$ are initialized.
2. Alternating between two updates:
   (a) Log likelihood is calculated using Equation 7
   (b) E step: using the current values of π, μ, $\mathbf{\Sigma}$, the responsibility of each component is calculated using Equation 5. This is the soft-assignment of each observation to a component k based on the resulting values of responsibilities
   (c) M step: using the responsibilities calculated in the E step, the model's parameters are updated:

$$N_k = \sum_{n=1}^{N} p(k \mid \mathbf{x_n}) \tag{8}$$

$$\pi_k = \frac{N_k}{N} \tag{9}$$

$$\mu_\mathbf{k} = \frac{1}{N_k} \sum_{n=1}^{N} p(k \mid \mathbf{x_n})\, \mathbf{x_n} \tag{10}$$

$$\mathbf{\Sigma_k} = \frac{1}{N_k} \sum_{n=1}^{N} p(k \mid \mathbf{x_n})\, (\mathbf{x_n} - \mu_\mathbf{k})(\mathbf{x_n} - \mu_\mathbf{k})^T \tag{11}$$

   (d) Log likelihood is calculated and the E-step is repeated.
3. The algorithm terminates when increases in the log likelihood stop or when they fall under a chosen threshold.

3.1.2. Z-Score

Z-score is a metric for scoring a sample from a Gaussian distribution with respect to the remaining samples in the dataset. It describes *"how many standard deviations above or below the mean"* a sample is [32, p. 192]. Once a Z-score is computed, one can get an approximation of how many other samples in the dataset have higher or lower scores. The Z-score is computed as follows [32, p. 193]:

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{12}$$

where μ is the mean of the dataset and σ is the standard deviation. The Z-score can be used to determine if a sample is anomalous by measuring how far it is from the mean of the dataset. A Z-score greater than |3| indicates that the sample has a probability less than 1% based on the remaining samples in the dataset.

This paper introduces a novel approach to insider-threat detection addressing current challenges in the field. The proposed detection solution is based on incorporating the knowledge of security analysts as an integral component of the system to lower the number of false positives that are common in insider-threat detection. The solution also addresses another challenge in insider-threat research proposing a novel approach for including non-technical indicators of insider threat as critical elements of the system. The proposed solution requires the following functionalities:

1. The ability to compute a vector representation of employees' activities
2. The ability for automated anomaly detection
3. The ability to communicate information to security analysts for analysis of detected anomalies
4. The ability to provide analysts with the capability of classifying detected anomalies
5. The ability to include non-technical indicators of insider threat as part of the detection system

Visual analytics is used for building into the detection system the knowledge of security analysts. In order to communicate to analysts why each observation is classified as anomalous, the information learned by the automated anomaly detection method is provided through visualisations. Furthermore, analysts' classification of observations is a feedback loop from the knowledge of analysts to the detection system. This enables the detection system to learn from analysts directly since the input is labelled observations, thereby, increasing the amount of information available when learning the normal behaviour of employees. Finally, each employee is represented by a feature vector computed from activity logs; each feature is a count of activity and computed as follows:

$$\mathbf{x_e}[index_{feature}] = f\,(activity\,logs, feature) \tag{13}$$

$$= count\,(events_{feature}) \tag{14}$$

Where $\mathbf{x_e}$ is the vector representation of employee e, $\mathbf{x_e}[i]$ is the value of the i^{th} dimension of the vector, and $events_{feature}$ are the events used for computing the value of the feature.

3.2. Anomaly Detection

Anomaly detection is done in two phases: *training* to learn the normal behaviour of each employee, and *detection* to apply anomaly detection for each new input. The proposed automated anomaly detection method introduces the usage of a GMM for modelling the normal behaviour of an employee. A GMM is implemented with a probabilistic approach which allows the model to explain why observations are classified as anomalous [29, 30]. In addition to the simplicity of GMMs, the parameters of the models and the resulting predictions provide analysts with insight on the method's decision-making process.

A GMM provides the capability of modelling a dataset of a complex probability distribution. GMMs have been used in other work for modelling a dataset on a behaviour that varies over time, such as modelling the patterns of flight operations, and for the biometric identification of user behaviour [33, 22]. Since the behaviour of employees varies throughout time, and each employee has a characteristic normal behaviour, a mixture of probability distributions is used to model the behaviour of each employee [11, 20, 21]. Each employee is represented by a GMM that is trained during a period assumed to represent the employee's normal behaviour [22, 21]. A GMM is trained using the EM algorithm with the objective of maximising the likelihood of the dataset.

The metrics used for measuring deviation from the norm are the likelihood of the input (Equation 3), and the Z-score of each feature (Equation 12). The Z-score determines how sensitive the system is towards false positives [34, 35]. During anomaly detection, the automated anomaly detection method computes the likelihood of an observation, $p(x)$, and predicts the component responsible for generating it. Using the parameters of the predicted component, k, which are the mean and covariance, the Z-score of each feature is computed. An observation is classified as anomalous when:

1. The likelihood of the observation is less than the likelihood threshold ϵ_l:

$$p(x) < \epsilon_l \tag{15}$$

2. The absolute value of the Z-score of any of the features is greater than the threshold of the corresponding feature ϵ_f:

$$|Z_f| > \epsilon_f \tag{16}$$

3.3. Non-Technical Indicators

Prior works found in the related work section established that non-technical indicators are of crucial value in detecting insider threat. However, incorporating them in detection methodologies is still a challenge. This paper introduces a method for recognising non-technical indicators of insider threat. It is a novel approach to threshold control called *Sensitivity Profiles*.

Sensitivity profiles may be applied before or after analysing anomalies generated by automated anomaly detection. For instance, if an employee is reported by a supervisor to exhibit an unusual behaviour, analysts can choose a sensitivity profile associated with the observed behaviour which updates the thresholds of the employee's model before applying anomaly detection. Their application determines the sensitivity of the system to employees' behaviour since parameter-refinement is dependent on the chosen profile.

4. The Insider-Threat Detection System

The Insider-Threat Detection System is composed of five components: data processing, anomaly detection, visualization, classification of detected anomalies, and threshold control. The design of the Insider-Threat Detection System is shown in Figure 1. The tool runs in two phases: training and detection. During training, the behaviour of each employee is modelled by a GMM. Then, the system is used for insider-threat detection during the detection phase. The system's components are divided

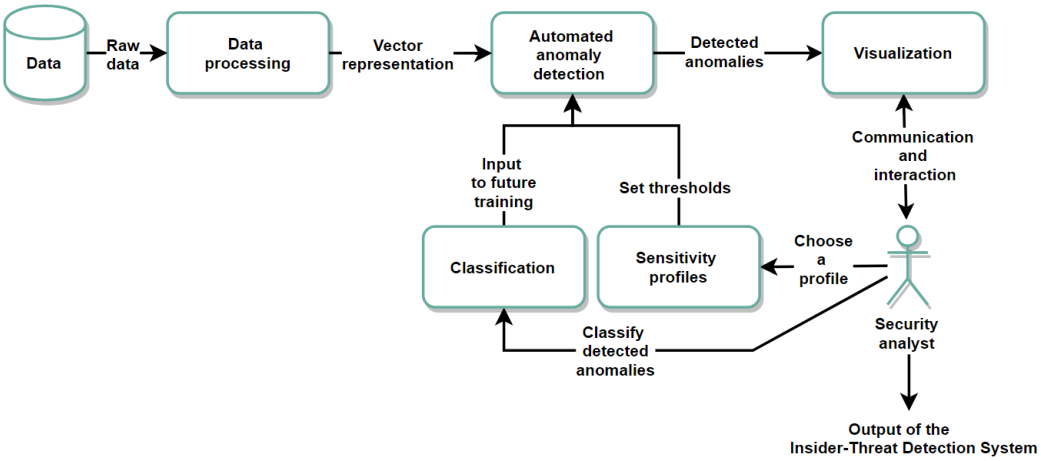

Figure 1: The components of the Insider-Threat Detection System

to back-end and front-end components. The front-end of the system refers to the components that are developed as part of a web-based Automated Anomaly Detection Dashboard which are *Visualization*, *Classification*, and *Sensitivity Profiles*. The back-end components are the *Profile Generator*, *Feature Calculator*, and *Automated Anomaly Detection*.

As shown in Figure 1, the input to the system is data on the activities of employees. The data is parsed by the *Profile Generator*. It creates an hourly profile from activity logs for each employee for computing a feature vector (although the duration of the profile can be straightforwardly set to any arbitrary duration). Then, the *Feature Calculator* receives as input current and history profiles. The module computes for each employee a vector representation of their activities. The *Automated Anomaly Detection* component is used for training the GMMs that model the normal behaviour of employees which are then used for anomaly detection.

The front-end of the system is a web-based dashboard. It is the user interface of the system used by security analysts for analysing detected anomalies, classifying detected anomalies, and applying sensitivity profiles. *Visualization* is implemented as part of the Automated Anomaly Detection Dashboard for communicating information from the automated anomaly detection component to security analysts. The next component of the system is the security analyst who contributes with knowledge by interacting with the three front-end components shown in Figure 1. The final two components are *Classification* which corresponds to analysts' classification of anomalies, and *Sensitivity Profiles* which is the threshold-control component that is used when a non-technical indicator of insider threat is observed, or to adjust the parameters of an employee's model to lower the number of false positives.

4.1. Data Processing

The *Profile Generator* is the parser of the system. For each employee, there is a current and a history profile. The current profile has information on the employee's behaviour at the time the system is run. History profiles summarise some information on employees' behaviour through time, such as the list of PCs used by each employee. For each employee, data is collected from different activity logs and parsed to develop a profile that has information relevant to insider-threat detection. The history profile is updated after the *Feature Calculator* computes the vector representation of employees' activities. The *Feature Calculator* computes a vector representation of an employee's activities at the input hour t. For each feature, it retrieves the relevant information from the profile and computes the value of the feature which is a count of events.

4.2. Automated Anomaly Detection

The *Automated Anomaly Detection* component is used during the training phase of the system for training the GMMs that model employees' normal behaviours, and during the detection phase for applying anomaly detection using the trained models. In both phases, the input to the *Automated Anomaly Detection* component is a set of feature vectors. During training, the input to the component is the system's training dataset which is the set of computed feature vectors for all employees. Therefore, the system requires the availability of historical data on the activities of employees of an organisation where employees exhibit a normal behaviour. Consideration of cases where historical data is not available or where employees are not known to exhibit an expected normal behaviour is outside the scope of this paper but it is considered in future work. For each employee, a training dataset of N vectors, where $t \in N$ is an hour of activity, is retrieved. The training of each GMM starts with the initialisation of parameters. Then, the EM algorithm is applied for learning the parameters of the model that maximize the log likelihood of the training set. After the GMMs are trained, the Insider-Threat Detection System can be used in the detection phase.

The *Automated Anomaly Detection* component is used in the detection phase to detect anomalous behaviour that is an indicator of a potential insider threat. For each employee, it takes as input the vector representation of the employee's activities in hour t. Then, the employee's model is retrieved to compute the likelihood of the input, $p(x)$. The component checks if the value of $p(x)$ falls below the likelihood threshold. If it does, it

is added to the component's output as a type of anomaly. Then, the model predicts the component with the highest *responsibility*, $p(k \mid x)$. The parameters of the predicted component are then retrieved, which are the mean vector and covariance matrix, to compute the Z-score of each feature to check for a feature anomaly. If an anomaly is detected, it is appended to the output as a type of anomaly. The parameters of the predicted component are also added to the output to be used during data analysis.

Therefore, in addition to the list of detected anomalies, the output of the *Automated Anomaly Detection* component, which is communicated using visualizations, for each detected anomaly is:

1. **Anomaly type(s)**: to provide analysts with the reason(s) behind classifying the observation as anomalous.

2. **Feature vector**: to provide analysts with the raw data, which is the input data to the *Automated Anomaly Detection* component.

3. **Likelihood**: to provide analysts with information on how likely the observation is as predicted by the employee's model.

4. **Parameters of the predicted component**: since the Z-score is used for detecting anomalies, the value of the mean and variance used for computing the Z-score are provided to analysts to explain the reasoning of the *Automated Anomaly Detection* method.

5. **Historical behaviour**: to provide analysts with information on the historical behaviour of each element which are the features, likelihood of the observation, and parameters of the predicted component. Analysts can compare the historical behaviour against the current value and detect patterns. They can also asses the significance of deviations by comparing values observed at the same hour on previous days.

4.3. Automated Anomaly Detection Dashboard

The Automated Anomaly Detection Dashboard is designed as a web-based visual analytics dashboard. The dashboard is designed to take as input the detected anomalies by the *Automated Anomaly Detection* component. Input data include the IDs of employees and roles, the vectors that are classified as anomalous, and the likelihood of each vector. Then, the interfaces for interacting with the three components discussed previously are built. There are two views for each component of the dashboard which are *employees* or *roles*. Depending on the chosen view, either data on employees is uploaded or data on roles. After a view is selected, the dashboard can be used for insider-threat detection. The dashboard also provides a functionality for classifying observed anomalies. Analysts assign classifications to observations for future training of GMMs.

4.3.1. Visualization

The Automated Anomaly Detection Dashboard is the front-end of the system which uses multiple visualisations for communications different types of information and for visualization of different data types. A Parallel Coordinates Plot can visualize multivariate data providing the ability to detect outliers, relationships between variables, and retrieve values. Therefore, it is used for visualizing feature vectors, the role of each employee, and the likelihood where each is assigned to an axis. The parallel coordiantes plot may reveal outliers. This may be the case when no employees but the outlier are copying files, for instance. A Scatter Plot is used for visualizing predicted anomaly types since it can visualize categorical data providing the ability to retrieve values and observe patterns; there may be more than one anomaly type per employee. Analysts use the scatter plot to retrieve information on why an employee is classified as anomalous. Furthermore, for analysing the behaviour of a certain employee, a histogram and line charts are used. The histogram is used for visualizing the historical behaviour of each anomaly type, which are the likelihood and features. A line chart is added on the same plot for visualizing the predicted parameters over time. When the analysis of the behaviour of a selected anomaly type is complete, analysts may go back to the scatter plot to choose another anomaly type and continue the analysis of the selected employee. Otherwise, they may go back to the parallel coordinates plot and continue the analysis of the employee's behaviour by observing the remaining features, or continue the analysis of the remaining employees. Analysts may also filter out an employee, classify their behaviour, or apply a sensitivity profile.

4.3.2. Sensitivity Profiles

We relied on the literature to recognise non-technical indicators of insider threat and common causes of false positives to develop sensitivity profiles [2, 11, 12, 8]. We used heuristics to develop the values of the parameters of sensitivity profiles. In our study, we are proposing the feasibility of further development. The proposed solution is a proof of concept that uses the existing literature to heuristically come up with values. These values are subject to refinement, once we have conducted further studies to provide empirical evidence to suggest how external factors can influence employee behaviour in organisations (which in turn help us identify how to compensate for it).

Sensitivity Profiles is accessed through the Automated Anomaly Detection dashboard. The profiles are created only once during the initial implementation of the system. After analysts provide a selection, the GMM of the selected employee and the parameters of the selected profile are retrieved. Afterwards, the thresholds of the employee's model are updated as determined by the parameters of the profile. Each sensitivity profile is associated with a set of features and a corresponding set of threshold-control values, and it is applied as follows:

$$\mathbf{v}_{modelThresholds} = f\,(SensitivityProfile,\, model) \tag{17}$$

$$= \mathbf{v}_{modelThresholds} \pm \mathbf{p}_{profiles} \tag{18}$$

Where $\mathbf{p}_{profiles}$ is an n-dimensional vector that represents the

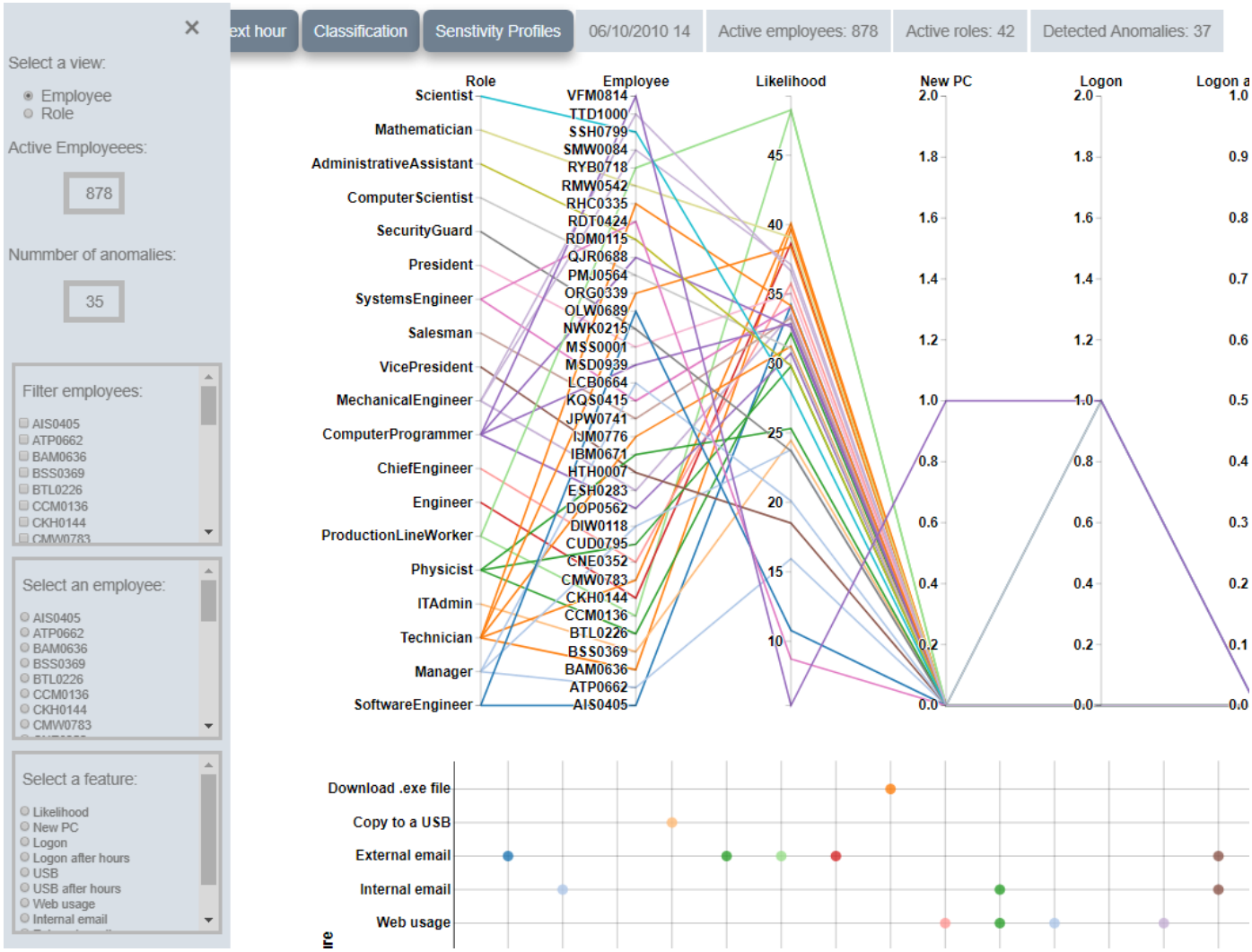

Figure 2: A screenshot of the Automated Anomaly Detection dashboard showing the visualization selection menu, the inforamtion header, the parallel coordinates plot, and the scatter plot.

parameters of the chosen profile, n is the number of parameters of the profile, the increase or decrease, $\pm$, of thresholds is also determined by the chosen profile, and $\mathbf{v}_{modelThresholds}$ are the thresholds of the employee's model. Then, anomaly detection is applied again on the selected employee's observation using the updated thresholds. The result is provided back to the Automated Anomaly Detection Dashboard to update the visualizations.

Moreover, in our concept demonstrator there are six sensitivity profiles where each is associated with a set a features and a corresponding set of threshold-control values. Control values may increase or decrease a threshold depending on the level of sensitivity required (more profiles can straightforwardly be created). In this paper, our focus is to demonstrate their feasibility as well as their usability with cybersecurity researchers trialing the concept. Profiles are categorised as either non-technical indicators of insider threat, or as causes of false positives. A sensitivity profile that is designed as a non-technical indicator of insider threat makes the system more sensitive to changes in the employee's behaviour. It is an approach for monitoring an employee's actions more closely. On the other hand, a sensitivity profile that is designed as a cause of false positivess makes the system less sensitive to changes in the behaviour of an employee. It is an approach to acknowledging legitimate causes of changes in employees' behaviours. Further work will be necessary to determine how accurate and precise the use of sensitivity profiles are and can be. For the purpose of this study, we determined the specific use cases and values using heuristics. The profiles are informally described as follows:

- **Negative organisation change**: negative changes in organisations impacting the jobs of employees, such as layoffs and reduction in wages, are a common catalyst for insider attacks [2, 11, 12, 8]. Dissatisfaction with the changes may cause an insider with certain technical skills to carry out an IT sabotage attack, job insecurity may cause an insider to seek a job elsewhere and carry out theft of intellectual property, and fear of financial instability may cause an insider to commit fraud [2, 12, 8].

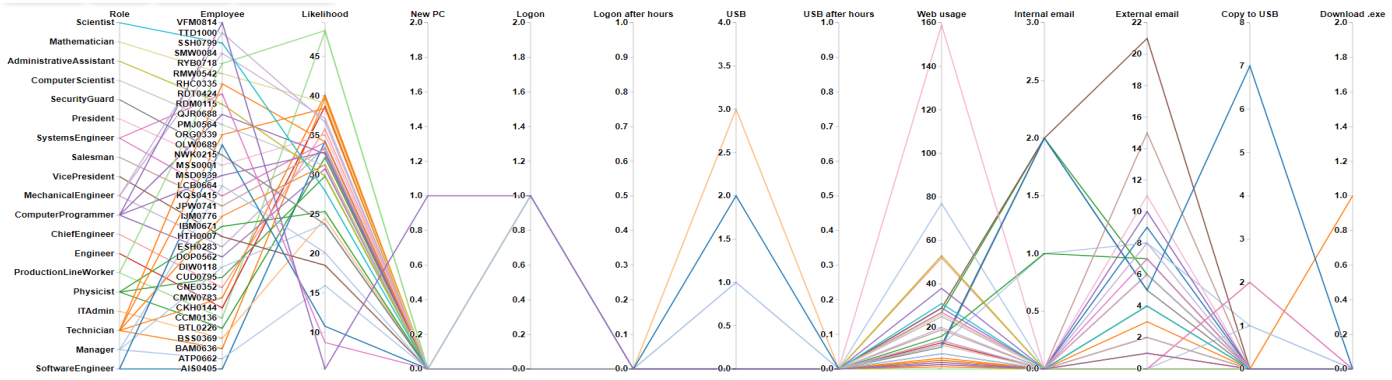

Figure 3: Parallel coordinates plot on the raw data. The axes plot the employee ID, employee's role, value of likelihood, and the value of each feature of the employee's vector representation.

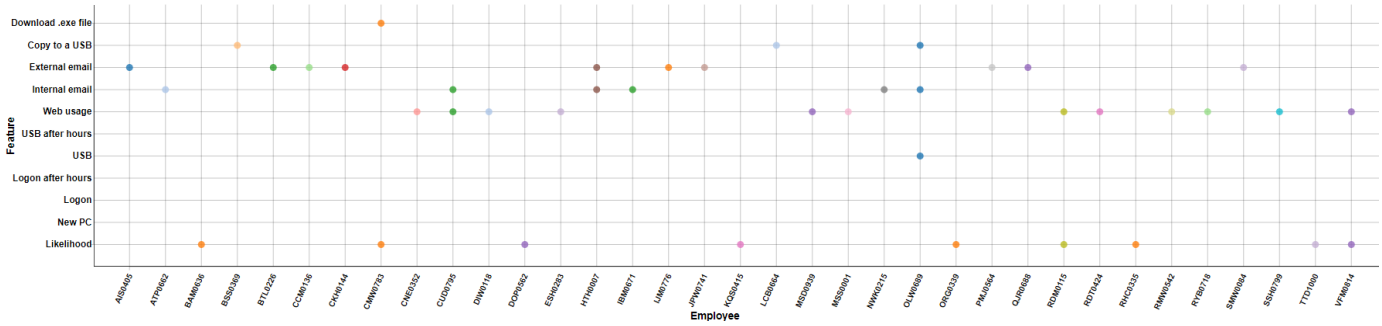

Figure 4: Scatter plot on the predicted anomalies for each employee. The x-axis maps the employee ID and the y-axis maps to the anomaly types which are the likelihood and features. Each dot corresponds to a detected anomaly type which is the cause of classifying the employee's observation as anomalous.

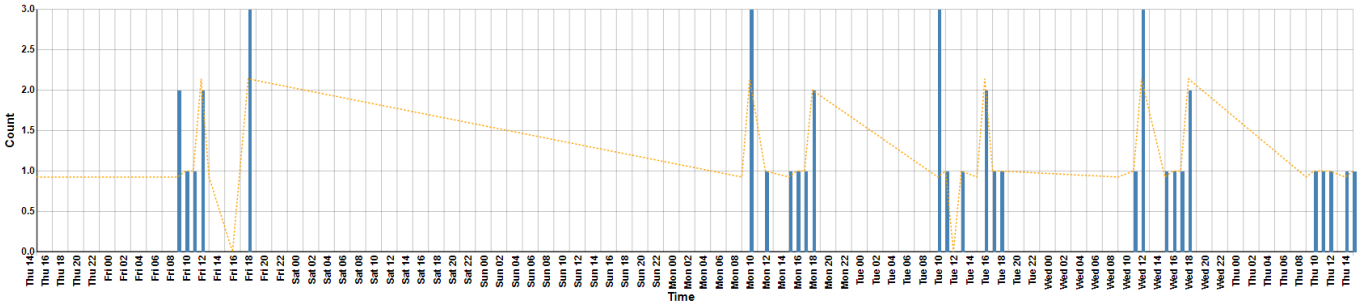

Figure 5: Visualizing the historical behaviour of web usage. The histogram plots the value of the anomaly type. Interval is an hour. The yellow line is a representation of the "current" average as predicted by the employee's GMM.

Therefore, this sensitivity profile makes the system more sensitive to technical indicators related to all three types of insider threat. The proposed features of this profile are email communication, activities during non-working hours, USB connections, file copying and download, and web-based activities.

- **Unusual behaviour**: unusual employee behaviour that has not been observed before or not expected, such as an aggressive attitude towards colleagues or lack of performance, is another common indicator of insider threat [2, 11, 12, 8]. It is an indicator of theft of intellectual property and IT sabotage in most cases. In few cases where unusual behaviour is due to personal and financial struggles, it is an indicator of insider fraud. Therefore, this sensitivity profile makes the system more sensitive. The proposed features of this profile are file copying and download, external email communication, and USB connections.
- **Financial problems**: financial problems may be indicators of insider fraud. Only in few cases of theft of intellectual property the insider was financially motivated [2]. Therefore, this sensitivity profile makes the system more sensitive to indicators of insider fraud and theft of intellectual-property. The proposed features of this profile are file copying, email communications, web-based activities, and logins during non-working hours.
- **Termination**: termination, such as job dismissal and res-

ignation, is an indicator of insider threat. Most insider-threat cases of types IT sabotage and theft of intellectual property occurred either right before or after insiders left the organisation [2]. Employees with a sense of entitlement to their work, and with certain job roles that have access to confidential data, such as scientists, may engage in theft of intellectual property either to secure a job, or start a business [11, 2, 8]. On the other hand, employees with certain personality characteristics and psychological state that have the technical skills for carrying out IT sabotage, may be triggered to attack due to such a catalyst. Therefore, this sensitivity profile makes the system more sensitive to all of the employee's activities.

- **Acting for another employee**: this profile refers to an employee that is assigned to work in place of another employee which is most likely going to trigger false positives for the change in behaviour may correspond to the change in work activities. For example, a certain job role may not require the employee to exchange external emails, but the new position may involve exchanging external emails. In such a case, the automated anomaly detection component will trigger false alarms on the employee's email behaviour. This sensitivity profile makes the system less sensitive to changes in the employee's behaviour. The suggested approach is to have the employee's model parameters be temporarily adopted from the other employee's model.
- **New project**: a new project refers to any activity that requires the employee to work more, and carry out anomalous activities compared to their normal behaviour. This sensitivity profile makes the system less sensitive to changes in the employee's behaviour. The proposed features of this profile are activities during non-working hours, web-based activities, email communications, and file copying.

Finally, analysts may use the component before or after analysing detected anomalies. The choice of when to apply a profile is determined by analysts. For example, analysts may apply a profile when an employee is reported for exhibiting an alarming behaviour. Otherwise, analysts may apply a profile after analysing detected anomalies. It may be applied after enquiring on why an employee was working during the weekend when they have never done so before and discovering that the employee was busy working on a new project.

4.4. Implementation

Each employee is represented as a 10-dimensional vector. The ten selected features that describe an employee's behaviour were derived from the literature. We relied on work done in the field of insider threat which uses a similar type of dataset on the activities of employees and with the same objective of detecting different types of threats to identify common elements describing an employee's behaviour [36, 21, 11].

The features constituting a vector representation are usage of a new computer (PC), login, login during non-working hours, usage of a removable storage device (USB), usage of a USB during non-working hours, web browsing, exchange of internal emails, exchange of external emails, file copying to a USB, and download of executable files. Internal emails are exchanged with colleagues using the organisation email. External emails are exchanged with members outside of the organisation. The vector representation of an employee is shown in Equation 19.

$$\mathbf{x}_e = [newPC \quad logon \quad logonAH \quad connect \quad connectAH \quad http \quad email \quad emailE \quad copy \quad exe] \tag{19}$$

Furthermore, to apply a sensitivity profile on an employee's observation, a vector representing a profile is either added, to decrease sensitivity, or subtracted, to increase sensitivity, from the employee's learned thresholds. The proposed parameters of each profile discussed in Section 4.3.2 are shown in Table 1. The dimension of each value corresponds to the dimension assigned to the feature in the vector representation of employees shown in Equation 19

Moreover, the Python [37] programming language is used for the back-end implementation of the Insider-Threat Detection System. For the implementation of the web-based dashboard, the Data-Driven Documents (D3) JavaScript library is used for implementing visualizations [38]. A screenshot of the Automated Anomaly Detection Dashboard is shown in Figure 2. As can be seen in the figure, there is an information header and control buttons. There is a **Next Hour** and a **Previous Hour** buttons for updating data. An hour increment is chosen since the date update duration is an hour. An example parallel coordinates plot is shown in Figure 3, and an example scatter plot is shown in Figure 4. Furthermore, when an employee is selected for analysis of historical behaviour, the employee's data is retrieved and a histogram and line chart are generated as shown in Figure 5.

5. Detection Performance and Results

This section presents the results of the novel automated anomaly detection method. The dataset used for testing the system is the CERT Insider-Threat dataset [39] version *r4.2*. The dataset has logs on the activities of 1000 employees and it is rich with insider-threat incidents.

The automated anomaly detection method is evaluated by assessing how well GMMs capture the normal behaviour of employees during training, and its capability in detecting insider threat during anomaly detection. Therefore, in each phase, different evaluation metrics are used. There are several metrics for evaluating a model's classification performance. Among them are sensitivity, precision, and the false positive rate [28, p. 183]. Sensitivity, which is also referred to as recall and the true positive rate, indicates the ratio of detected positives (Equation 20). Precision indicates the ratio of correctly classified positives (Equation 21). The false positive rate indicates the ratio of misclassified negatives (Equation 22).

Table 1: Parameters of sensitivity profiles. $v_{thresholds}$ is the learned vector of thresholds where each dimension corresponds to a feature as shown in equation 19. The vector gets updated upon application of a sensitivity profile.

Profile	**Equation**
Negative organisation change	$\mathbf{v}_{thresholds} = \mathbf{v}_{thresholds} - [0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$
Unusual behaviour	$\mathbf{v}_{thresholds} = \mathbf{v}_{thresholds} - [0\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 1]$
Financial problems	$\mathbf{v}_{thresholds} = \mathbf{v}_{thresholds} - [0\ 0\ 1\ 0\ 0\ 1\ 1\ 1\ 1\ 0]$
Termination	$\mathbf{v}_{thresholds} = \mathbf{v}_{thresholds} - [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]$
Acting for another employee	The suggested approach is to have the employee's model parameters be temporarily adopted from the other employee's model.
New project	$\mathbf{v}_{thresholds} = \mathbf{v}_{thresholds} + [0\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 0]$

$$R = \frac{TP}{TP + FN} \tag{20}$$

$$P = \frac{TP}{TP + FP} \tag{21}$$

$$FPR = \frac{FP}{FP + TN} \tag{22}$$

For a model that requires both high recall and high precision, the F1 score can be used. It can offer a balance by maximizing both recall and precision as it combines them in one measure which is computed as follows [28, p. 185]:

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \tag{23}$$

Since the objective during training is to model the normal behaviour of employees, recall is used to evaluate the models' effectiveness in correctly detecting positives, where a positive label is a normal behaviour. A false negative during training indicates the misclassification of a normal observation as anomalous. Taking into consideration the number of false negatives during training provides insight into how well the model is capable of recognising a normal behaviour. When testing the automated anomaly detection method in its ability to detect anomalous behaviour, the F1 score and false positive rate are used as evaluation metrics. During anomaly detection, a positive label is a malicious act.

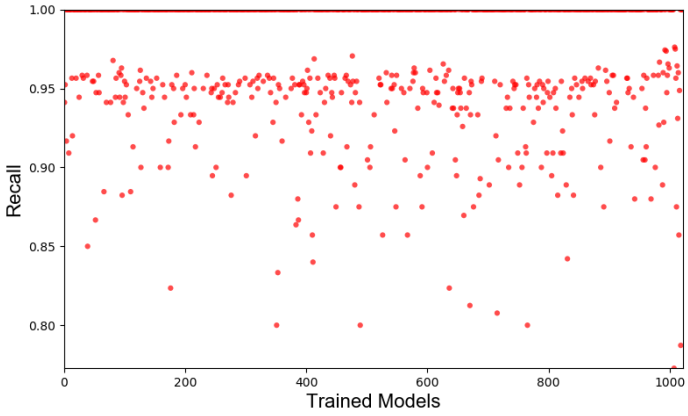

Figure 6: Performance of each trained GMM. Please note: the y-axis starts at 0.77 in order to zoom in on the results. This means that at no point does the recall results go below 0.77

GMMs are trained using a set of vectors computed from the logged activities of employees during a month. The results of testing GMMs in capturing the normal behaviour of employees is shown in Figure 6. The figure shows high recall values for all models where only a few resulted in a ratio less than 0.8. The results indicate that GMMs are successful in distinguishing normal observations, thereby, modelling the normal behaviour of employees.

To evaluate the performance of the automated anomaly detection method in detecting insider threat, the F1 score and false positive rate are used. A positive label includes an indicator of insider threat or an execution of an attack. There are two attacks that took place in the test set at different times of the month where two insiders of the role IT Admin carry out the same attack scenario. The insiders first download a keylogger, and copy it to a USB. They then log in to their supervisor's PC connecting a USB to download the keylogger for stealing the supervisor's password. The following day, they log in to their supervisor's machine using the supervisor's user ID and stolen password, and send a mass email that causes panic in the organisation [39]. In addition, there are four insiders in the test set that only show indicators of a future attack. They search online for job applications and exchange emails with a competitor organisation enquiring about employment [39].

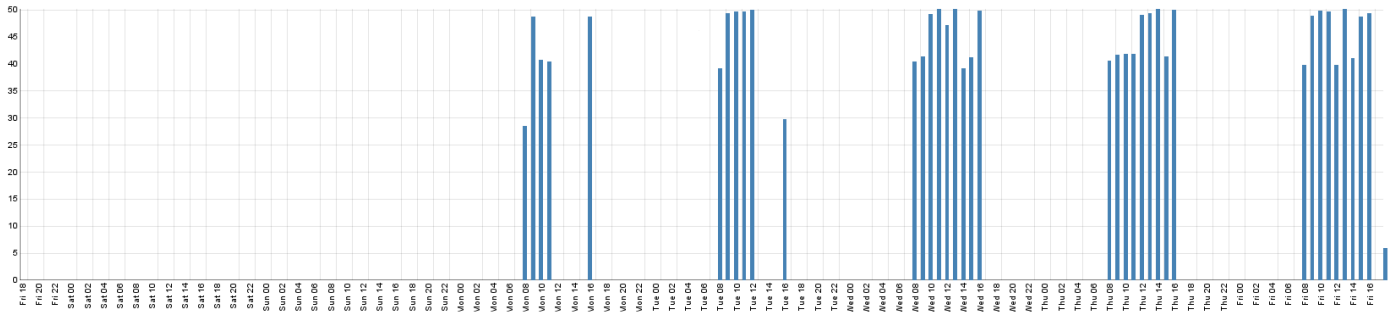

Figure 7: The plot shows the computed likelihood of the supervisor for the duration of a week. The last observation, at hour 17:00, indicates an anomaly as the likelihood value is very low compared to the history of computed likelihood.

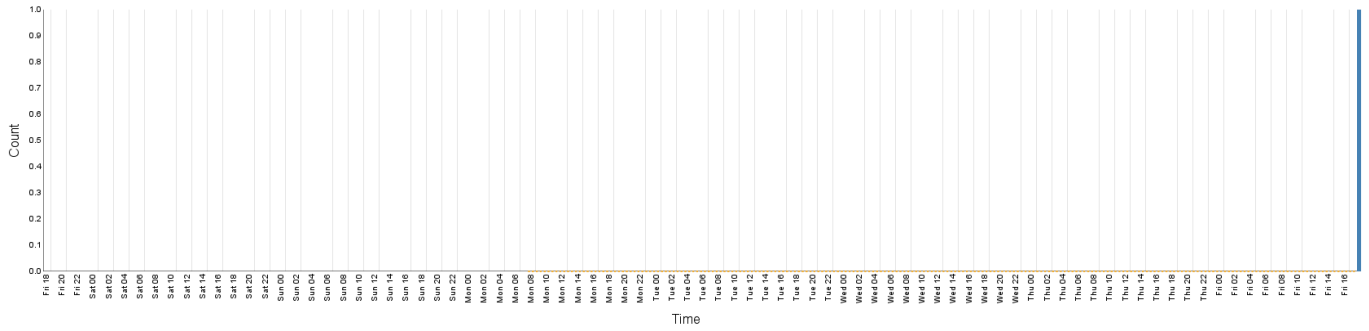

Figure 8: The plot shows the logon behaviour of the supervisor during working hours for the duration of a week. It shows an unusual logon time as all previous observations have a value of 0, and the predicted average shown by the yellow line is also 0. This means that the supervisor usually logs in during non-working hours.

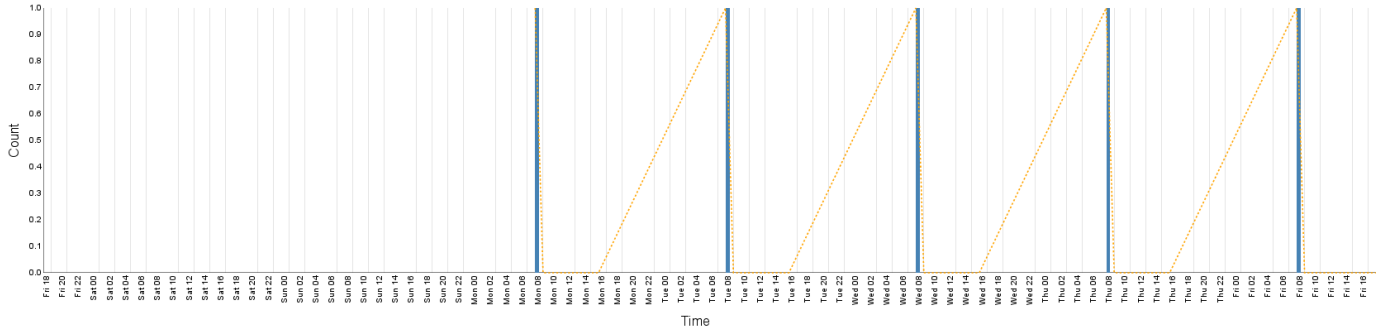

Figure 9: The behaviour of the logon activity during non-working hours of the supervisor. The plot shows the logon-time pattern of the supervisor where the expected logon time as shown by both the count values and predicted average is at 7:00; thereby indicating an anomaly in the logon behaviour of the supervisor.

The two attacks executed in the test set are successfully detected as anomalies with a *likelihood* anomaly type and a *logon* anomaly. The *likelihood* anomaly expresses the low likelihood of the observed behaviour of the supervisor when the mass email is sent as shown in Figure 7. Analysing the logon behaviour of the supervisor also shows an unusual pattern as shown in Figures 8 and 9. As seen in the figures, 17:00 is an unusual logon time for the supervisor where the usual logon time is at 7:00.

Moreover, the automated anomaly detection method was successful in detecting indicators of insider threat. Indicators refer to hourly activities that involve an action done in preparation for an attack. The indicators that reflect a change in behaviour as captured by the selected features are successfully detected either as an anomaly in the count of activities related to the indicator or as an overall unusual behaviour. However, indicators that reflect a change not captured by the selected features, such as the type of websites accessed by an insider, are not detected. For example, one insider in the test set prepares for a theft of intellectual property contacting a competitor organisation for a job opportunity. At one instance, the insider exchanged 4 external emails discussing a job opportunity. However, the predicted average number of external emails exchanged in an hour for that employee is 0.15 with a predicted variance of around 0.12 resulting in a successful detection of the indicator as an anomaly of type *external email*. On the other hand, indicators

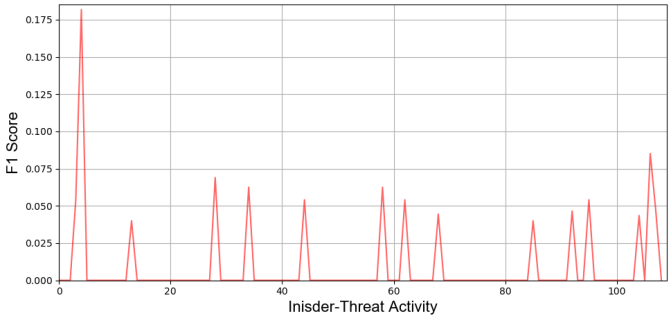

Figure 10: The F1 score computed on the test dataset. The results indicate a successful detection of attacks and indicators of attacks. A lower F1 score corresponds to a higher number of false positives. The number of activities includes every hour with an insider acting.

Figure 11: The false positive rate computed on the test dataset. The plot shows a low false positive rate where most system runs resulted in a false positive rate below 20%.

are detected in some cases as an overall unusual behaviour. For example, another employee who is involved in the same scenario visits a large number of websites over time searching for jobs. Several of the employee's observations are classified as anomalous with anomaly type *USB connect*. On inspection of the predicted component over time for the employee, a high number of web visits is unusual to take place with a USB connect activity. Therefore, on observing a high number of web visits, the employee's GMM predicted a component that is responsible for modelling this activity resulting in predicting an anomaly of type *USB connect*.

For the overall performance of the method, the F1 score is computed for every hour with insider-threat activities as shown in Figure 10. The F1 scores of value 0 indicate the instances where insiders' activities are not detected. A low F1 score corresponds to a successful detection of a malicious behaviour with a relatively high number of false positives compared to the number of true positives. To analyse the performance of the automated anomaly detection method in terms of the number of false positives, the false positive rate is computed for every run of the system as shown in Figure 11. The figure demonstrates a low false positive rate. A lower false positive rate of the automated anomaly detection method contributes to a better performance of the Insider-Threat Detection System when security experts analyse detected anomalies using the Automated Anomaly Detection dashboard. The average number of false positives is 15 and the standard deviation is 20.8.

The results demonstrate the success of the automated anomaly detection method in detecting an insider-attack and indicators of attacks.

6. Feasibility Study

6.1. Study Design

The ability of the proposed approach in incorporating the knowledge of security experts is evaluated with a feasibility study. Target participants are experts in threat detection for their knowledge is a crucial element. The study aims at evaluating the usability of the solution. Detection performance is evaluated using the false positive rate since only data on detected anomalies is provided to analysts. The study is designed using a mixed-method approach collecting both quantitative and qualitative data. The data-collection tools are voice recording, screen recording, questionnaires, and notes on observations made during the study. The study is designed as follows:

1. **Introduction**: participants are given the information sheet, consent forms, and a demographic questionnaire.
2. **Training**: participants are provided with a description and a live demo of The Insider-Threat Detection System and its components. They are, then, provided with training tasks. The training tasks are presented in Appendix A.
3. **Scenario**: a description of the scenario is provided to participants. Then, the solution is tested by having participants analyse detected anomalies during an hour where there is an insider preparing for an attack.
4. **Reflection**: an interview is conducted to collect feedback on the solution and the set-up of the study. The interview and feedback questions are presented in Appendix A.

A live demo of the user interface and a verbal description of the solution makes it easier for participants to ask questions. Additionally, participants are trained before testing to avoid misinterpretation of results since performance during testing is impacted by the ability of participants to use the system's functionalities and interpret the information communicated through visualizations. Participants are trained on using each functionality that is made available during testing. In addition, feedback is collected during training as participants are only focusing on how to use the solution. After training, participants are provided with a scenario for using the solution in a fictitious organisation. A scenario is used to provide them with context and background knowledge on the data they are analysing just as they would have when working with a real organisation. During the scenario, a non-technical indicator of insider threat is introduced to test the usability of sensitivity profiles. The objective of the feasibility study is to evaluate the usability of the solution and observe how participants use it. Therefore, participants are not provided with tasks to guide them through the analysis, but are left to decide how to conduct the analysis of detected anomalies. This enables the collection of data on when, and how analysts use their knowledge with the information communicated through visualizations during the analysis of detected

anomalies, and whether they use the solution as designed. Data is also collected on the successful detection of the threat and false positives.

The data collected in the feasibility study is organised as follows:

1. Usability of the components of the Automated Anomaly Detection dashboard
2. Number of classified observations
3. Number of observations detected as false positives
4. Number of applied sensitivity profiles
5. Effectiveness of visualizations in communicating the decision-making process of the automated anomaly detection method
6. Success in incorporating the knowledge of participants into the system during analysis
7. The process of participants' analysis using the dashboard's components

Data collected on each area listed above is presented in the following sections. Data is collected on five participants that have a background in cybersecurity and threat detection, specifically. The reason there are few participants is that we considered the study to be a feasibility study. We wanted to go more in-depth on a per user basis and get more feedback, but which may not be generalisable at the moment, and obtain more data that can be investigated qualitatively to improve key aspects of the usability of the system before any future larger scale study. The invited participants are researchers with experience in threat detection. Three out the five participants have a background in visualization or computer graphics. Only one participant considers visualization as not very valuable for their work, while three consider it somewhat valuable and one considers it as very valuable for their work. None of the participants are colour-blind. Participants are referred to in this section by p# where # refers to an assigned participation ID. The screen recording for $p2$ was lost and data presented on the participant is from a backup data-collection mechanism. The scenario phase of the study took on average 27 minutes. During the scenario, there are 36 anomalous employees and 2 roles. There is only one insider acting during the scenario.

6.2. Detection Capabilities

Table 2: Number of detected False Positives (FPs) and the False Positive Rate (FPR) for each participant. Improvement is measured from the false positive rate of the automated anomaly detection method which is 0.04

Participant	Anomalies	FPs	FPR	Improvement
p1	4	3	0.036	9.30%
p2	3	2	0.037	6.57%
p3	6	5	0.034	14.77%
p4	5	4	0.035	12.03%
p5	7	6	0.033	17.52%

The number of detected false positives in the scenario is shown in Table 2. The results show the ability of analysts to detect false positives during analysis demonstrating the contribution of their knowledge to the performance of the system. The false positive rate of the automated anomaly detection method on employees' data at the time of the scenario is 0.04. The improvement in performance is shown in Table 2 highlighting the contribution of analysts. In addition, all participants used the *Classification* and *Sensitivity Profiles* functionalities as expected. All participants applied a profile and classified the observation of the detected insider.

6.3. Interview Findings

Feedback is collected during the interview on both the *Classification* and *Sensitivity Profiles* functionalities. The comments of participants are summarised in Figure 12. All participants found *Classification* to be useful. $p2$, $p4$, and $p5$ commented that additional feedback after classification is required, such as greying out observations from the parallel coordinates plot and scatter plot, as $p5$ suggested, to signal the end of the sequence of actions. Participants also suggested the addition of more classification tags to help feeding back more knowledge during the training of GMMs. $p1$ commented positively on the consistency in the design of interfaces. As for *Sensitivity Profiles*, all participants had positive comments on the functionality but the majority said they wanted to see the effect of the application of a profile on an employee more clearly. Some suggestions included the visualization of an observation before and after a profile is applied. $p2$ suggested, in reply to how *Sensitivity Profiles* can be improved, providing analysts with the ability to control the mapping of a profile to features, and the ability to create new profiles. $p1$ also offered a suggestion which is adding on the dashboard the list of features associated with each profile.

Moreover, participants used visualizations as expected and the results during training and scenario showed that the visualizations were successful in communicating the decision-making process of the automated anomaly detection method. Participants used the parallel coordinates plot to analyse raw data and get an overview on the activities of employees. The scatter plot was used for analysing the cause of anomaly. The histogram and line chart were successfully used for communicating information on the historical behaviour of employees. The overall feedback on the dashboard was also collected during the interview and shown in Figure 12. The ratings provided by each participant during the interview are presented in Table 3. $p4$'s reply to Q3 is Not Applicable (NA) because $p4$ is currently not working in the insider-threat field but $p4$ added that if assigned to do insider-threat detection, he would be interested. Feedback on Q1 is positive and participants commented on the visualizations used in the dashboard as shown in Figure 12. $p3$ highlighted throughout the interview that the scatter plot is very useful in developing scenarios and providing a clear view on who is potentially involved in a malicious activity. $p4$, on the other hand, signalled out the parallel coordinates plot on being helpful during analysis because it provides information on all of the activities in one plot making analysis easier.

Feedback on Q2 was also positive. $p3$ was able to detect the insider immediately as it was the first observation analysed, and $p5$ found the insider after analysing two observations only. The remaining participants were successful in detecting the threat after a non-technical indicator is injected during the scenario. Moreover, all participants, except $p4$ as discussed previously, said they are interested in using the solution. However, the

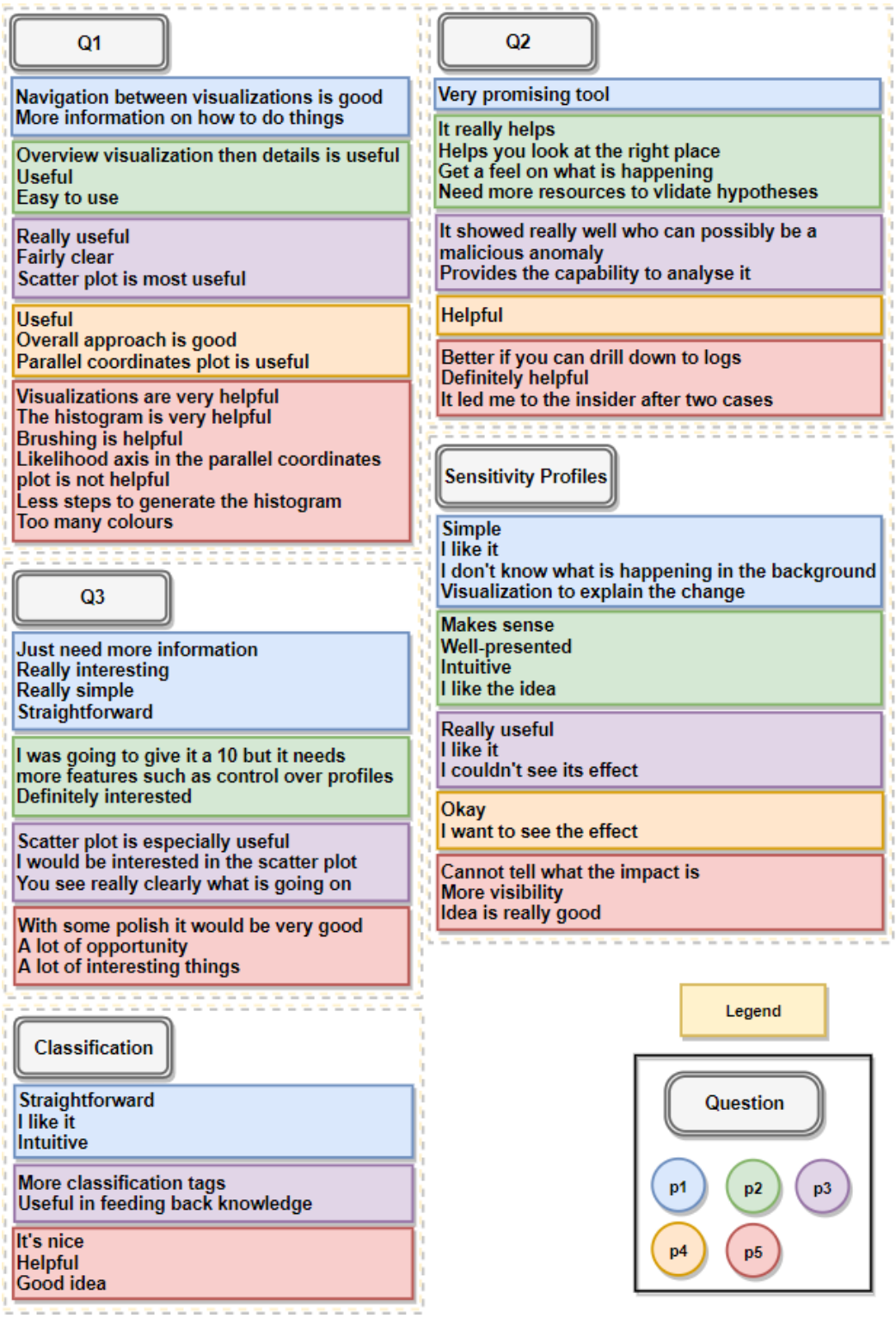

Figure 12: Key feedback collected during the interview. Each participant is represented by a colour and answers are sectioned based on the organisation of questions presented in Appendix A for the interview questions.

Table 3: Ratings of participants collected during the interview. The interview questions are listed in Table A.6.

Question	p1	p2	p3	p4	p5	Avg.
Q1: How useful did you find the Automated Anomaly Detection dashboard with respect to analysis of detected anomalies?	9	8	9	7	7	8
Q2: How well do you think the system is able to detect insider threat?	9	8	9	7	7	8
Q3: Overall, how interested are you in using such a tool for detecting and analysing insider threat?	10	9	9	NA	8	9

reason why $p2$, $p3$, and $p5$ did not provide a rating of 10 is the polishing of the interface, such as clearer axis labels, easier data retrieval from the paralllel coordinates plot, and less steps for generating the histogram.

All participants used their background knowledge to determine which combination of activities is more alarming. Participants also used their knowledge in developing scenarios on the behaviour of employees when analysing raw data, and, more importantly, when analysing anomaly types. Throughout the following discussion, the following definitions on how the knowledge of participants is incorporated during analysis are used:

- **Defining an unusual behaviour**: to determine which employees to analyse first using the parallel coordinates plot by developing a story on the behaviour of employees and deciding the activities that define an alarming behaviour.
- **Defining an alarming behaviour**: to determine which anomaly types or combination of anomaly types are most alarming using the scatter plot and also to develop scenarios based on the activities of the employee and anomaly types.
- **Defining a non-malicious behaviour**: to determine which behaviour is not malicious, thereby, detecting false positives.
- **Correlating the job role**: to correlate the roles of employees with their behaviour to determine if it is an expected or unusual behavior.
- **Assigning a weight to a feature**: to prioritise the selection of employees that engaged in the activity of the feature for further analysis.

Each participant used background knowledge during the analysis for:

- $p1$: defining an unusual behaviour, defining an alarming behaviour, and determining when to contact Human Resources (HR) to enquire on non-technical indicators observed for the employee under analysis.
- $p2$: defining an alarming behaviour. $p2$ also prioritised the analysis of employees with the highest number of anomalies. Background knowledge was also used for correlating the job role and $p2$ detected false positives using this approach using only the parallel coordinates plot and scatter plot. Knowledge was also used for defining a non-malicious behaviour using only the parallel coordinates plot and scatter plot, and assigning a weight to the external-emails feature.
- $p3$: defining an unusual behaviour, defining an alarming behaviour, correlating the job role, and assigning a weight to the external-emails feature.
- $p4$: defining an unusual behaviour, defining an alarming behaviour, assigning a weight to the following features in order:

  1. USB connection
  2. File copying to a USB
  3. External emails

  and for correlating the job role. $p4$ detected false positives using this approach.
- $p5$: defining an unusual behaviour, and defining an alarming behaviour. $p5$ also prioritised the analysis of employees with the highest number of anomalies. Knowledge was also used for assigning a weight to the following features in order:

  1. File copying to a USB
  2. External emails
  3. Web usage
  4. Download of executable files

  $p5$ also used background knowledge for correlating the job role and detected false positives using this approach. Finally, knowledge was used for deciding to analyse the history of computed likelihood values for each employee under analysis.

Furthermore, it is expected that the parallel coordinates plot is first used to get an overview on the activities of employees and analyse raw data. Then, the scatter plot is used to analyse anomaly causes. Then, the historical behaviour of a selected employee is analysed. Finally, classification of an observation and, if applicable, a sensitivity profile are applied. All participants except $p3$ used the dashboard's functionalities as expected. $p3$, unlike the remaining participants who started the analysis with the parallel coordinates plot, used the scatter plot as the first step in analysis and relied on it more heavily during analysis. The results presented demonstrate the success of the novel approach to insider-threat detection in incorporating the knowledge of security experts into the solution and improving the anomaly detection performance.

The results also show that it is possible to use visualizations to communicate the information learned by the automated anomaly detection method. Participants used visualizations as expected and after familiarity with the purpose of each visualization, they relied more heavily on the visualizations that communicated the decision-making process of the detection

method, which are the scatter plot, histogram, and line chart. In addition, all participants used sensitivity profiles during the scenario demonstrating that non-technical indicators of insider-threat can be incorporated into the detection system using this methodology. Some participants' comments on their interest in being able to create new profiles and control the associated selection of threshold controls, demonstrate the potential of the novel threshold-control mechanism.

7. Discussion

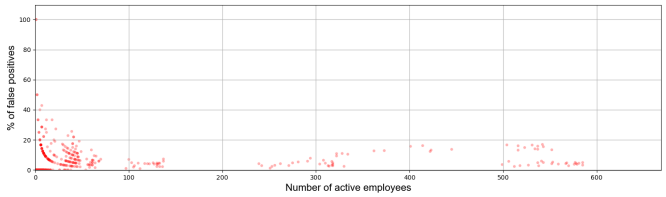

Figure 13: Percentage of false positives per active employees which is computed from the number of false anomalies and the total number of observations for the employee in the test dataset. The plot shows that some employees generated a higher number of false positives which affects the performance of the system. This may be due to a change in behaviour, role, or workload in the test dataset compared to the training dataset.

The results presented in Section 5 on the performance of the novel automated anomaly detection method for insider-threat detection show promising results. The method was able to detect insider-attacks and indicators of attacks. The results also demonstrate the method's performance in terms of the number of false positives which contributes to the performance of the Insider-Threat Detection System by reducing the noise when security experts analyse detected anomalies. Figure 13 shows the percentage of false positives for each run of the system ordered by the number of active employees. The figure shows the contribution of the automated anomaly detection method where security experts analyse less than 20% of active employees. Moreover, to improve the performance of the automated anomaly detection method, more features that provide the method with more information must be explored. As mentioned in Section 5, indicators that are not captured by the features, such as the type of content of the email or website, are also not captured by the detection method.

Furthermore, a pilot study was conducted to test the design and setup of the feasibility study. Prior to the pilot study, feedback was not collected during the training phase. After the pilot study, realizing that participants during training focus more on using the dashboard than analysis, the design of the study was changed to collect such data during training. In addition, prior to the pilot study, the study was designed to have participants read a description of the Insider-Threat Detection System and its components, in addition to a description of the scenario. However, realizing that burdens participants with reading a lot of information, the study was redesigned to have the study coordinator verbally describe the contents of the documents briefly to also encourage participants to ask questions. Finally, because it is a feasibility study, we pseudo-randomly selected two hours where there is an insider either preparing for, or executing an attack; one is assigned for training and the other for the scenario.

Moreover, the results presented in Section 6.3 show how participants used their experience in threat detection during analysis in several ways. For example, they used their knowledge in weighting features during the analysis of raw data, and to prioritize the investigation of certain anomalies over others based on their activities and predicted anomaly type(s). They started the analysis with, and dedicated more attention to employees that engaged in the activities that they gave more weight to. The results, therefore, not only show the success of the system in building on the knowledge of security experts but also highlight the potential of the proposed approach as more methods are introduced for incorporating their knowledge further. As discussed in Section 6.3, participants commented during the interview on the need for visualizing the inner-workings of sensitivity profiles. Therefore, as more tools are provided to explain the automated elements of the system, the system is more able to capitalise on the knowledge of security experts. In addition, the ratings provided in Table 3 only relate to first-impressions and further investigation will be necessary.

The limitation of the work presented is in evaluating it using a synthetic dataset. Unlike data collected from real-world organisations, synthetic data is organised and generated to provide information relevant to an insider-threat detection solution. However, validating solutions to insider threat is a major challenge in the field due to the ethical and privacy concerns related to revealing such data [6]. Future work includes validation of the solution on a real dataset and collecting more data from analysts' usage of the Insider-Threat Detection Solution. Doing so also enables us to consider the scalability of the solution. By exploring other visualization methods and collecting data as analysts use the various visualization methods, we can investigate approaches to analysing larger amounts of data that correspond to larger numbers of employees without impacting the detection performance of the system. In addition, future work includes exploring methods for improving the performance of the automated anomaly detection method, such as adding more features, automating feature selection, and exploring additional anomaly-detection metrics. Furthermore, *Sensitivity profiles* is a rudimentary approach to test a novel idea to incorporate soft knowledge, through heuristics. It remains to be validated with real data. In order to determine whether the profiles' parameters are actually representative would require significantly more research into psychology and threat detection. Before applying any sensitivity profile, it is necessary to be critical of its use and only use them if and when appropriate. We propose that policies or standard operation procedures should be in place facilitating their appropriate usage. This investigation is subject to future work. Finally, other methods for building into the solution the knowledge of security experts can be explored, such as adding a signature-based method for detecting patterns as determined by policies set by analysts. Exploring the addition of signature-based methods to the remaining components of the solution is especially of value when considering detection analysis in organisations where historical data on employ-

ees is not available or where employees are not known to exhibit an expected normal behaviour, such as in project-based work. Another approach is defining deviation in the anomaly detection method to be from the behaviour of employees of the same role/project instead of employees' history.

8. Conclusions

The Insider-Threat Detection System was developed as an integration of various methods designed to incorporate the knowledge of security experts into a detection solution. The system included a component for processing activity logs to compute vector representations of employees' activities to be analysed by an automated anomaly detection component, and by analysts. The paper introduced a new anomaly detection method for insider-threat detection that uses GMMs to model the normal behaviour of employees and uses likelihood and Z-score as the anomaly-detection metrics. The system also includes a component for feeding the knowledge of security experts further into the system by having analysts classify detected anomalies. The resulting classifications are used for future training of GMMs. Furthermore, visual analytics is utilised as the main method for incorporating the knowledge of security experts using visualizations to communicate to analysts the information learned by the automated anomaly detection method. Finally, the paper introduced a new method for incorporating non-technical indicators of insider threat into a detection solution which is sensitivity profiles. Sensitivity profiles are threshold-control templates used for increasing or decreasing the sensitivity of the automated anomaly detection method to employees' changing behaviour based on observing a non-technical of insider threat or a legitimate cause of change in behaviour.

The results of the feasibility study presented in Section 6 demonstrate that an insider-threat detection solution can be designed to incorporate the knowledge of security experts to improve its performance. During the feasibility study, participants, who are experts in threat detection, demonstrated the success of the proposed approach by using their knowledge throughout the analysis in various ways, and improving the false positive rate of the automated anomaly detection method. Visualizations were also shown during the study in being successful in communicating the decision-making process of the automated anomaly detection. Therefore, the paper's contribution to insider-threat detection is in introducing a novel method for designing detection solutions. Furthermore, the novel automated anomaly detection method introduced achieved excellent detection rates when placed side by side to the state of the art. It was able to detect the two insider attacks in the test dataset, in addition to attack indicators. The results also show the contribution of the novel method to the field of insider-threat detection in resulting in a low false positive rate, in addition to a successful detection of attacks. Finally, the results of the feasibility study demonstrate the project's additional contribution to the field of insider threat in introducing sensitivity profiles and their potential in being used for improving the performance of a detection solution by including non-technical indicators of insider threat and legitimate causes of false positives.

References

[1] I. A. Gheyas and A. E. Abdallah, "Detection and Prediction of Insider Threats to Cyber Security: A Systematic Literature Review and Meta-analysis," *Big Data Analytics*, vol. 1, p. 6, August 2016.

[2] D. Cappelli, A. Moore, and R. Trzeciak, *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*. Addison-Wesley Professional, 2012.

[3] Ponemon Institute Research Report, "2016 Cost of Cyber Crime Study and the Risk of Business Innovation." [Online] Ponemon Institute LLC. Retrieved from: `http://www.ponemon.org/local/upload/file/2016%20HPE%20CCC%20GLOBAL%20REPORT%20FINAL%203.pdf`., October 2016. [Accessed 2017-05-01].

[4] I. Agrafiotis, J. R. C. Nurse, O. Buckley, P. Legg, S. Creese, and M. Goldsmith, "Identifying Attack Patterns for Insider Threat Detection," *Computer Fraud and Security*, pp. 9–17, July 2015.

[5] T. Chen, F. Kammuller, I. Nemli, and C. W. Probst, "A Probabilistic Analysis Framework for Malicious Insider Threats," in *Proceedings of the Third International Conference on Human Aspects of Information Security, Privacy, and Trust - Volume 9190*, (New York, NY, USA), pp. 178–189, Springer-Verlag New York, Inc., August 2015.

[6] 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)*, pp. 286–295, May 2016.

[7] United States Computer Emergency Readiness Team (US-CERT), "Combating the Insider Threat." [Online] Department of Homeland Security. Retrieved from: `https://www.us-cert.gov/sites/default/files/publications/Combating%20the%20Insider%20Threat_0.pdf`, May 2014. [Accessed 2017-04-27].

[8] A. McCormac, K. Parsons, and M. Butavicius, "Preventing and Profiling Malicious Insider Attacks," *Australian Government Department of Defense, Defense Science and Technology Organization*, April 2012.

[9] G. K. L. Tam, V. Kothari, and M. Chen, "An Analysis of Machine- and Human-Analytics in Classification," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 71–80, January 2017.

[10] I. Mann, *Hacking the Human*. UK Book Publishing, 2016.

[11] P. A. Legg, N. Moffat, J. R. Nurse, J. Happa, I. Agrafiotis, M. Goldsmith, and S. Creese, "Towards a Conceptual Model and Reasoning Structure for Insider Threat Detection," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 4, no. 4, pp. 20–37, 2013.

[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 *2014 IEEE Security and Privacy Workshops*, pp. 214–228, May 2014.

[13] I. Agrafiotis, A. Erola, M. Goldsmith, and S. Creese, "A Tripwire Grammar for Insider Threat Detection," in *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*, MIST '16, (New York, NY, USA), pp. 105–108, ACM, October 2016.

[14] "IBM." [Online] IBM Official Website. Retrieved from: `https://www.ibm.com/us-en/?lnk=m`. [Accessed 2017-07-11].

[15] "IBM QRadar SIEM." [Online] IBM Official Website. Retrieved from: `https://www.ibm.com/ms-en/marketplace/ibm-qradar-siem/details#product-header-top`. [Accessed 2017-07-11].

[16] M. Bishop, H. M. Conboy, H. Phan, B. I. Simidchieva, G. S. Avrunin, L. A. Clarke, L. J. Osterweil, and S. Peisert, "Insider Threat Identification by Process Analysis," in *2014 IEEE Security and Privacy Workshops*, pp. 251–264, May 2014.

[17] 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 *2012 IEEE Symposium on Security and Privacy Workshops*, pp. 142–149, May 2012.

[18] R. Zhang, X. Chen, J. Shi, F. Xu, and Y. Pu, "Detecting Insider Threat Based on Document Access Behavior Analysis," in *Asia-Pacific Web Conference*, pp. 376–387, Springer, September 2014.

[19] T. E. Senator, H. G. Goldberg, A. Memory, W. T. Young, B. Rees, R. Pierce, D. Huang, M. Reardon, D. A. Bader, E. Chow, I. Essa, J. Jones, V. Bettadapura, D. H. Chau, O. Green, O. Kaya, A. Zakrzewska, E. Briscoe, R. I. L. Mappus, R. McColl, L. Weiss, T. G. Dietterich, A. Fern, W.-K. Wong, S. Das, A. Emmott, J. Irvine, J.-Y. Lee, D. Koutra, C. Faloutsos, D. Corkill, L. Friedland, A. Gentzel, and D. Jensen, "De-

tecting 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*, KDD '13, (New York, NY, USA), pp. 1393–1401, ACM, August 2013.
[20] 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*, vol. PP, pp. 1–10, June 2015.
[21] T. Rashid, I. Agrafiotis, and J. R. Nurse, "A New Take on Detecting Insider Threats: Exploring the Use of Hidden Markov Models," in *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*, MIST '16, (New York, NY, USA), pp. 47–56, ACM, October 2016.
[22] Y. Song, M. B. Salem, S. Hershkop, and S. J. Stolfo, "System Level User Behavior Biometrics using Fisher Features and Gaussian Mixture Models," in *2013 IEEE Security and Privacy Workshops*, pp. 52–59, May 2013.
[23] J. Thomas, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.
[24] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, *Visual Analytics: Definition, Process, and Challenges*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
[25] R. Marty, *Applied Security Visualization*. Addison-Wesley Professional, 1 ed., 2008.
[26] K. Nance and R. Marty, "Identifying and Visualizing the Malicious Insider Threat Using Bipartite Graphs," in *2011 44th Hawaii International Conference on System Sciences*, pp. 1–9, January 2011.
[27] P. A. Legg, "Visualizing the Insider Threat: Challenges and Tools for Identifying Malicious User Activity," in *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–7, October 2015.
[28] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
[29] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
[30] I. Davidson, "Anomaly Detection, Explanation and Visualization," *SGI Technical Report*, 2007.
[31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data mining, Inference and Prediction*. Springer, 2 ed., 2009.
[32] R. Hanneman, A. Kposowa, and M. Riddle, *Basic Statistics for Social Research*. Jossey-Bass, 2013. Ebook Central.
[33] L. Li, R. J. Hansman, R. Palacios, and R. Welsch, "Anomaly Detection via a Gaussian Mixture Model for Flight Operation and Safety Monitoring," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 45 – 57, March 2016.
[34] T. Iwata and K. Saito, "Visualization of Anomalies Using Mixture Models," *Journal of Intelligent Manufacturing*, vol. 16, pp. 635–643, December 2005.
[35] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
[36] H. Eldardiry, E. Bart, J. Liu, J. Hanley, B. Price, and O. Brdiczka, "Multi-Domain Information Fusion for Insider Threat Detection," in *2013 IEEE Security and Privacy Workshops*, pp. 45–51, May 2013.
[37] G. Rossum, "Python Reference Manual," tech. rep., Amsterdam, The Netherlands, The Netherlands, 1995.
[38] M. Bostock, V. Ogievetsky, and J. Heer, "D3 Data-Driven Documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, pp. 2301–2309, December 2011.
[39] CERT Division, "Insider Threat Tools." [Online] Software Engineering Institute at Carnegie Mellon University. Retrieved from: `https://www.cert.org/insider-threat/tools/.`

Appendix A. Questionnaires

This section presents the questionnaires discussed previously that are used in the user study. Table A.4 presents the demographic questions provided to participants at the beginning of the study. Table A.5 presents the training tasks provided to participants after the live demo on the Automated Anomaly Detection dashboard. Table A.6 lists the interview and feedback questions.

Table A.4: Demographic Questionnaire. White = Multiple Choice, Orange = Short Answer

Question no.	Question Description	Answer Choices
1	Do you have a background in cybersecurity?	[yes, no]
1.1	If yes, which aspect (mainly) of cybersecurity?	[Technical aspects - e.g. focused on the implementation side of cybersecurity, Social aspects - e.g. primarily focused on the social science]
1.2	If yes, do you have a background (work or education) in threat detection?	[yes, no]
1.2.1	If yes, for how long?	[0-5 years, 5-10 years, 10-15 years, 15-20 years, 20+ years]
1.2.2	If yes, do you have a background in insider threat?	[yes, no]
2	Do you have a background (work or education) in graphics or visualization?	[yes, no]
2.1	If yes, which aspect (mainly)?	[Computer graphics technical – e.g. implementation of graphic algorithms, Computer graphics design – e.g. designing media based on existing tools, Visualization technical – e.g. worked with visual analytics at a greater capacity than at high school]
3	How important is the use of visualization for your work?	[N/A, Not at all, Not very, Somewhat, Very]
4	Do you generally prefer static dashboard visualisations or interactive ones to explore data?	[Static, Interactive, I don't know]
5	Have you used a tool or system for insider threat detection before?	[yes, no]
6	Are you colour-blind?	[yes, no, I don't know]
7	Do you wear glasses or contact lenses to correct for eyesight?	[yes, no]

Table A.5: Training Tasks provided to participants after the live demo on the usage of the Automated Anomaly Detection dashboard. White = Instruction, Orange = Question

Task no.	Task Description
1	Click on the Menu button and choose the Employee visualization view. Close the side panel.
2	Using the parallel coordinates plot, which employee has the lowest likelihood? List all if there is more than one.
3	Using the brushing functionality, limit the range of the Likelihood axis to the observed value in the previous step.
4	What is the value of the employee's likelihood?
5	What is the role of the employee? List all if there is more than one employee.
6	If there is more than one employee listed in previous steps, choose one. What activities was the employee engaged in? List the value of each activity.
7	Using the axis swapping functionality, move the Copy to USB and Download .exe axes next the USB Connect axis.
8	Click on the Employee visualization view again to remove the brushing effect.
9	Select the Role view.
10	How many role anomalies are detected?
11	If there is more than one role, choose one.
12	Using the scatter plot, what anomaly type is predicted for the role retrieved in the previous step?
13	Go back to the Employee visualization view.
14	Using the filter box, filter out the employee(s) you observed in the previous steps.
15	Using the scatter plot, retrieve the employee of role Production line worker with only an external email anomaly.
16	What other activities was the employee engaged in? You can use the brushing functionality to get a more focused view.
17	Using the employee selection box, select the employee's user ID obtained in step 15.
18	Using the feature selection box, choose the *Likelihood* option.
19	What is value of likelihood at the same time for each day of the week?
20	For the same employee, select the *External email* feature.
21	What patterns do you observe?
22	Approximately, what is value of this feature at the same time for each day of the week?
23	Using the activity logs in the *data* folder, open the file *06˙24˙16˙email* and check the emails that the employee exchanged at this time.
24	What type of emails are exchanged?
25	Using the activity logs in the *data* folder, open the file *06˙24˙16˙http* and check the websites that the employee visited at this time.
26	What type of websites are visited?
27	Using the *psychometric* file, what are the employee's measured values?
28	Go back to the Automated Anomaly Detection dashboard. Click the ***Previous Hour*** button.
29	How many employees are active?
30	How many roles are active?
31	How many anomalies were detected?
32	Click on the ***Next Hour*** button
33	Open the *Employee* visualization view.
34	Click on the ***Classification*** button and select the Role view.
35	Select any role and classify it as *Normal*.
36	Close the panel.
37	Click on the ***Sensitivity Profiles*** button.
38	Select the *Employee* view
39	Check the anomaly type predicted for employee **AKC0924**.
40	From the employee selection box in the *Sensitivity Profiles* menu, choose the employee **AKC0924**.
41	Apply the sensitivity profile *New project (busy)*.
42	Observe the resulting change in the visualizations.

Table A.6: Interview and feedback questionnaire provided to participants after the scenario. White = Rating Question, Orange = Interview Question

Question no.	Question Description
1	How useful did you find the Automated Anomaly Detection dashboard with respect to analysis of detected anomalies? Please rate from 1=not at all to 10 = very.
2	What is your opinion about the Automated Anomaly Detection dashboard?
3	What do you think of the classification functionality?
4	What do you think of the Sensitivity Profiles functionality?
5	In what ways do you think Sensitivity Profiles can be improved?
6	How well do you think the system is able to detect insider threat? Please rate from 1=not at all to 10 = very.
7	Any comments on the rating?
8	Overall, how interested are you in using such a tool for detecting and analysing insider threat? Please rate from 1=not at all to 10 = very.
9	Any comments on the rating?
10	Do you have any concerns or comments with regards to the study itself?