Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MODELLING AN INSIDER THREATS DETECTION METHOD AGAINST EMAIL CONTENT

AMEERA NATASHA MOHAMMAD[1], WARUSIA YASSIN*, RABIAH AHMAD[2], SUGANYA DEVI[3]

*,1,2Universiti Teknikal Malaysia Melaka, Faculty of Information & Communication Technology, Malaysia

[3]National Institute of Technology Silchar Assam, India

E-mail: *s.m.warusia@utem.edu.my

ABSTRACT

An insider threats has become one of the most challenging malicious activity in cybersecurity defense compared system recently. Since an insider threats usually expand and spread internally, no one could predict what, when and how exactly malicious insider launched their attacks. This is in a view of fact that an email becomes one of the primary targets of an internal threats as this medium widely used by everyone to communicate, share, and exchange confidential information. Therefore, it is extremely important to understand the nature of insider threats behavior beforehand and construct an accurate detection model. Furthermore, every single keyword used in an email can reflect the behavior of an individual and can be used to determine their intentions i.e., have motive to launch an insider threat or not. Henceforth, an innovative approach is proposed in modelling an insider threats detection in this work. In addition, various statistical analysis i.e., scoring, Friedman, linear regression (R2) and correlation coefficient applied to analyze an insider threat relationship among historical insider threats behavior and relevant extracted keywords from an email content. The Friedman statistical used to determine the minimum differences between each extracted insider threats keywords that represent different insider threats factor (motive, opportunity, capability). Besides, linear regression applied to estimate the relationship of an insider threats from training keywords and testing keyword with allocate anomaly score. Finally, the correlation coefficient approach used to determine how strong a relationship is between extracted insider threats keywords and insider threats behavior in this research. The proposed modelling approach has been evaluated using the benchmark dataset known as CERT that comprises malicious email file. Throughout the experiment, the proposed insider threats detection approach has achieved higher attack detection rate as well as minimized undetectable insider threats behavior as compared to previous researcher works.

Keywords: *Insider threats, Email Content, Insider Threats Keywords, Malicious Behavior, Statistical Based*

1. INTRODUCTION

In recent years, protecting an intellectual property and confidential information become more challenging as malicious activity coming from internal resources which known as insider threats. An insider threats defined as malicious activity that targeted from inside an organization which usually launched by people within an organization itself such as employees, former employees, contractors, or business partner. In addition, an insider usually be knowledgeable of an organization computing resources including its vulnerabilities or weakness. An IP theft, fraud and sabotage are types of insider threats that launched by malicious insider against targeted organization with the aim to create trust issues, financial issues, and loss of reputation. Today an email platform widely used by entire organizations, as this communication medium offers easier way in sharing and exchanging private information. Unfortunately, an email system is also more vulnerable to an insider threats as compromised email account using valid and legitimate credentials undetectable and easy to perform insider attacks. Consequently, detecting an insider threats behavior remain as challenging task. Therefore, it is essential to have knowledge on the behavior of an insider threats first at foremost, then it will be much easier to analyze and detecting such threats more accurately. Furthermore, an email content keywords also can reflect people expression or bad feelings i.e., dissatisfaction, anger, revenge and many more. Hence, we proposed a novel modelling approach for the detection of an insider

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

threats using suspicious keywords that comprises inside an email. Theoretically, three different insider threats factors have been chosen from the suggestion of previous study such as motive, capability and opportunity. These factors applied to categorize and construct an insider attacks scores using the entire historical insider keywords. This is due to the fact that each keyword commonly can reflect the behavior of sender either they are contributing to threats or wise versa. The historical insider keywords from an email content are necessary to understand insider behavior. Furthermore, the relationship of an insider threats further analyzed using three different statistical analysis such as Friedman, linear regression and correlation coefficient. The remainder of the content is presented as follows: Section 2 focused on the literature discussion of related study while the proposed modelling approach presented in section 3. Section 4 presents the discussion of the results and finally section 5 comprises the conclusion and future work directions.

2. RELATED WORKS

3.1 Insider Threats

An insider threat continuously getting much attention by cybersecurity community as an important threat that must be mitigated earlier before it spreads and causes huge damages. Furthermore, this threats always targeting private information and assets that are essential for an effective operation of an organization in various as critical infrastructure sectors such as information technology, manufacturing, and communications. The situation will worsen if an organization link their critical infrastructure, IT system that comprises vulnerabilities. For instance, an insider who also an employee or person who know on such vulnerabilities and have bad intention towards an organization, might use this opportunity to launch threat against an organization IT system via legitimate credentials. This fact is further reinforced through numerous previous researcher studies while stating an insider threats usually composed with the aim to cause fraud, sabotage and IP Theft such as from (Julian Jang-Jaccard 2014) (UK 2015) (Heneke et al. 2016) (Ekran 2016) (Wood et al. 2016) (Scott & Spaniel 2017) (Huff et al. 2019) (Biringer et al. 2019) (Georgia Lykou 2019) (Robert Ambrogi 2019). Moreover, for an effective insider threat to be launch, malicious insider must require an ability, motive, and opportunity to causes threats (Sarkar 2010), (Pfleeger et al. 2010), (Xiaojun et al. 2013).

2.1.1 Impact of Insider Threats

In this subsection, the negative impact of an insider threats discussed for better understanding based on several previous studies. Ho et al. (2018) described the impact of an insider threats could be huger as contrast to outsider as the insiders themselves has knowledge on the vulnerability of organization's network, from where can collect confidential information, know on how the configurations made and used within the organization. Moreover, Elmrabit, Yang, & Yang, n.d. (2015) also states that the impact of insider threats affect more on the principle element of information security such as confidentiality, integrity, availability. Furthermore, based on previous study, the impacts of insider threats is divided into trust issues, loss of intellectual and physical assets, financial expenses and disruption of organization (Christian W. Probst, Jeffrey Hunker, Dieter Gollmann 2010), (Majeed 2016), (Isotopes et al. 2017), (Ho et al. 2017), (Ho et al. 2018).

3.1 Characteristic of Insider Threats

The distinct characteristic of an insider threats has been highlighted by various researchers and some propose a model of solutions to combat these threats. For instance, according to (Greitzer et al. 2012), insider threats can be described as an action to harm an organization assets by trusted individual. Besides, Kramer & Crawford (2005) relates the human psychology and motivation factors as the major contributors that influence a person to do this kind of threat and the prediction of such activity remains challenges. Moreover, Shaw & Fischer (2005) reported in his studied case that severe employment crises can contribute to dissatisfaction and serious personnel problems and will subsequently turn someone to cause threat.

Consequently, various previous researcher has proposed framework, method, system, and approaches as a solution to insider threats. For instance, Greitzer et al. (2011) proposed a framework that able to predict an insider threats by integrating various set of data sources. The author considered psychological factors such as disgruntlement, anger management issues and disregard for authority contributed to psychosocial risk and turn employee to cause a threat. Beside, Greitzer et al. (2012) proposed a system that correlated human expert resources as an estimation to assess an employee's behavior that relate to risk of insider threats via prototype psychological model. In addition, this author categorized factors such as personality predispositions, concerning behaviours and demographics (age, gender and

cultural aspects) in psychosocial and demographic factors. In addition, Claycomb & Nicoll, n.d. (2012) has relate three types of insider threats which are cloud provider admin, an employee that exploit the vulnerabilities for unauthorized access and malicious insider who use cloud resources to launch threats against organization's IT security. Under an author perspective, an attacker usually will search for any possible opportunity to gain access and get any sensitive data for different purposes. For example, to sell it to outsider (also known as fraud), future employment opportunities (IP theft) and IT sabotage. Moreover, the author also emphasized that such insider threats is very difficult to be detectable. Due to this, subsequently, Alhanahnah et al. (2016) employed feature extraction method to extract and analyze an important features that related to an insider threats. The related features are from an online content, online activity, social data, and email communication patterns. Moreover, the author has classified an insider threats into five dimensions i.e., deployment of cloud, an insider threat source, an impact of insider threats, insider threat mitigation and suspicious cloud services. Besides, May et al. (2017) focuses on unintentional insider threats, in which the proposed technique used to illustrate workflow between categories, insider threat structural taxonomy that subsidizes orthogonal taxonomy and possible cyber security defense solutions against an insider threats. Moreover, Homoliak et al. (2018) proposed analysis taxonomies that contribute to an insider threats disambiguation generalize knowledge about an insider threats. The author also updates an insider threats theory in different perspective from various previous literature review.

TABLE 1 CHARACTERISTIC OF INSIDER THREATS

Author	Category of Features	Example of Features
Greitzer et al. 2012	Employee's behaviour	Disgruntlement, anger issues, disregard, stress, personal problems and dependability lack
May et al. 2017	Unintentional insider threats	Fatigue or sleepiness, lack of knowledge, workload, stress
Homoliak et al. 2018	Insider threat incidents	Political and individual gain, ignorant, revenge, lack of training, excessive workload, personal problems.
Maasberg et al. 2015	User personality	socially mean character, self-promotion, emotion sensitivity, betrayal, and aggressiveness due to deficit of guilt, understanding and intention to harm others.
Claycomb & Nicoll, n.d. 2012	Threats from cloud resources	Fraud or sell sensitive data, IP theft
Elmrabit, Yang, & Yang, n.d. 2015	Insider in information security	Revenge, personal and financial gain, lack of security awareness, pressure and stress
Gavai, Sricharan, Gunning, Rolleston, et al. 2015	Detection of employee abnormal behaviour from online activity data	Espionage, illegal contract agreement for financial gain and make a copy of the system's password and network file.

The characteristic of an insider threats is summarized in Table 1 based on previous work. Mitigation of an insider threats are very challenging, henceforth understanding on an insider threat characteristic is necessary. Throughout the study, we manage to verified the existence of an insider threats sentiment keyword from psychological state of malicious insider such as disgruntlement, anger management problems, disregard and revenge (Greitzer et al. 2012) (Nurse et al. 2014) (Alhanahnah et al. 2016). Moreover, an insider capable to launch threat from inside organization via online communication and activity such as an email. Therefore, the following section discusses the sentiment keyword analysis of an insider threat and objective of the sentiment keyword-based study.

3.1 Keyword-based Sentiment Study

The keyword-based is defined as a method of processing natural language, linguistics computational and text with the aim to determine and extract relevant information from multiple content that representing characteristic and behavior (Baumgarten et al. 2013) (Kim et al. 2018). Furthermore, the sentiment analysis defined as an understanding of a person opinion or emotion towards given context, product and so forth that representing an expression of a person feelings (Baumgarten et al. 2013) (Kharde 2016). Usually, the keyword-based sentiment analysis is the content of communication based on emotional state of a person. In addition, an example of sentiment keywords commonly derived from emotional and psychological state of an insider such as depressed, revenge, stress, boredom and annoyed which can be

found through online activity i.e., in an email communication. Besides understanding the common characteristic of an insider, keyword analysis also should be given an extra attention particularly in identifying the significant parameter which could help in improving an insider threat detection (Yoshinaga et al. 2010), (Mohammad 2016), (Jiang et al. 2018), (Tan 2019). Hence, keyword-based sentiment analysis of previous studies has been reviewed and summarized in Table 2.

TABLE 2 KEYWORD-BASED SENTIMENT STUDIES

Author	Objectives	Parameter of Keyword
Yoshinaga et al. 2010	To extract the different type of keywords based on the behaviour of user within email communication	Affiliates, terrorist, Bloomberg, attacks,
Greitzer et al. 206)	Assess behavioural indicators and used by insider threat using email	Disregard for authority, disgruntlement, anger management issues, stress, lack of dependability, personal issues
Jiang et al. 2018	To construct user profiling and insider threats activity prediction used sentiment analysis in email content	Depressed, revenge, nuisance, outraged, pissed
Tan 2019	Develop and demonstrate framework of insider's written communications and psychological and emotional states	Classification of emotion – anger, joy and none (no emotion)
Mohammad 2016	To highlight the important key in word analysis based on sentiment analysis	Boredom, disgust, rage, annoyance

According to (Yoshinaga et al. 2010), (Mohammad 2016), (Jiang et al. 2018), (Tan 2019) depressed, revenge, stress, boredom and annoyed are represent sentiment keyword from emotional and psychological state of insider which can be found through online activity especially email communication. Therefore, besides fundamentally categorized common characteristic of an insider, keyword analysis also should be given an extra attention in order to improve insider threat detection method.

3.1 Previous Insider Threats Through Email Content

The revolutionary of technology nowadays changed the way of people and organizations conducting their business and communication activity. This also due to increases of reliability, efficiency and productivity via utilization of an email system (Nkosi et al. 2014) (Wang 2018). Furthermore, current email communication system capable to carries rich of confidential information between sender and receiver (Wang 2018). Consequently, an email system widely accepted and adopted for performing critical and trusted transactions such as facilitating the setup of payments, sensitive data transfers, contracts negotiations, intellectual property, research and development of business planning and strategy. Unfortunately, today, an email system also become the targetable platform by an insider as they can launch malicious intent easily. According to Verma et al. (2012) a malicious person capable to become an insider once he/she gain access inside an organization in any way including via email phishing. This is because malicious insider can have control over confidential assets as he/she wanted to and perform malicious activity to steal sensitive information and used it for their own benefits. These malicious insider threats is one of the limitation of such systems as it is very difficult to be identified and an improved solutions need to be discover (Nkosi et al. 2014) (Xie et al. 2015) (Wang 2018) (LaRosa et al 2019). A number of previous researchers has proposed a detection method or approaches as a solution against these threats as highlighted in Table 3.

Nkosi et al. (2014) proposed model for the detection of an insider threat by sensing and analyzing suspicious activities of an email content. The proposed model used consecutive rule mining algorithm to compare incoming events against constructed historical user profiles. The entire incoming activities via an email will be examined by comparing it with user historical profiles to identify any existence of an abnormal pattern. Subsequently, an activity that does not match with historical user profiles are considered as an insider threats activity. The major component in this model is the application of an historical user profiles in prior to facilitate the detection of possible insider threats activity. The result shows a significant achievement in detecting malicious insiders based on the behavior pattern approach. Moreover, Form

et al. (2015) proposed an emails phishing detection using features of an email content, URL and malicious behavior. The applied features include sender's domain, number of keywords that blacklisted in subject and content of email, URL, exclusive sender and domain, uneven pattern, and arrival route. Throughout this detection model, first, features from incoming email extracted to produce feature scores. Later, the features scores used to train SVM classifier in classifying an email either as malicious or legitimate email. Unfortunately, the result shows degradation, as the related keywords that been blacklist in various subject and content is not updated regularly. On the other hand, To (2015) has proposed an analytic analysis in an email content with the aim to detect an insider threat based on data exfiltration method. This method able to identify any unusual removal of sensitive data that leaving the organization communication systems such as an email and removable media. In addition, for the prediction purpose, an indicator such as disregard, inability to adapt with stress, suspicious email communication and negative workplace events been considered and analyzed. The author emphasized that malicious insider usually gain access from legitimate employees account through email by make them believe and respond to that email which containing phishing activity. According to the author, once the phishing link activated by legitimate employee unintentionally, considered an employee has provide a platform for insider to launch threats.

Therefore, the author focused on unusual email content activity for the detection of malicious threat. Similarly, Jiang et al. (2018) also proposed an insider threat behavior detection system based on an email content. The author highlights as an email become necessities in our daily life, certain keywords used in an email content can be reflect as an indicator of emotion or sentiment and psychology. For instance, an email contents send by malicious insiders usually comprises bad keywords that expressed their dissatisfaction and negative emotion. Furthermore, Lu & Wong (2019) has claimed that an insider threats activity can occur when the insider gain access into targeted employee computer and forward the entire private files to public website or to outsider email. This incident known as data leakages and could directly causes huge damages to an organization. Therefore, the author proposed an insider threat detection approach using machine learning techniques that able to capture the patterns of normal and abnormal user behaviors. The proposed detection approach evaluated using well known CERT datasets which comprises historical data of an insider threats activity. Using the detection approach, once the abnormal pattern is detected, the related email further analyzed for possible insider threats activity. Conversely, Kim et al. (2019a) proposed an anomaly detection algorithm for the prediction of an insider threats behaviors on an email content dataset. In the proposed model, user behavior is modelled for the purpose to transform heterogeneous user's behaviors into structured data as an input for anomaly detection. Subsequently, any unusual sending or receiving information via an email will be analyzed and determined either email content has comprised an insider threat activity or not. The experimental results show the proposed detection algorithm able to detect the behavior of an insider threats.

TABLE 3 PREVIOUS INSIDER THREATS APPROACH THROUGH EMAIL CONTENT

Author	Detection Approach	Description
Nkosi et al. 2014	Insider Threat Detection Model via email content	Examine user behaviour patterns by analysis of email content w ith sequential rule mining algorithm if there any abnormal pattern performs by malicious employee which could be insider threats
Form et al. 2015	Detection of insider threats behaviour via email phishing scenario	Examine insider threats behavior with hybrid features include sender's domain, blacklist words in subject and email content, URL, unique sender, unique domain, uneven hyperlink and return path which selected based on email phishing scenario
To 2015	Analytic analysis of email content for detecting insider threat	Email content exfiltration method in unusual removal of sensitive data leaving the organizational systems via email
Jiang et al. 2018	architecture detection system	Display the insider threats behavior based on email content.

Musuva et al. 2018	Theoretical model for determining suspicious email content	Understanding the factor of responding to email phishing
Vidyapeetham 2018	Detection of malicious email content	Analyze the suspicious email content sent using TF-IDF, Decision tree, Logistic Regression, Random forest, Naïve Bayes, KNN, AdaBoost and SVM
LaRosa et al 2019	System for detecting fraudulent insider threats through suspicious email content	Detection of insider threats through suspicious email content with a combination of social network graph communications analysis with algorithms designed
Lu & Wong 2019	Insider threat detection method	Indicate unusual email content and performed anomaly score detection based on historical attack data
Huff et al. 2019	Insider threat detection method	Detect insider threats using NATO qualitative methods to examine email content sent by malicious person to legitimate user in organization
Kim et al. 2019	Insider threat detection via email content analysis	Employ anomaly detection algorithms to detect insider threat behaviors via examine malicious email content

Hence, after reviewed a number of previous work benefits that focuses on an insider threats and more specifically using an email contents and keyword analysis, in this study, we have identified that there is certain approach that much more important to be considered in modelling an insider detection model i.e., identification of the features that influences an insider threats activity, a profiling and scoring method, as well as further analysis against scored data. However, there is lack on existing work as most of researchers only adopted several approaches in theirs studies and the degradation of the detection rate could be improved if all above-mentioned approaches are considered in modelling an insider threat detection. Therefore, in this research, we have proposed a novel insider threats detection modelling approach that has briefly explained in next section

3. RESEARCH DESIGN AND IMPLEMENTATION

3.1 The Proposed Modelling Insider Threats Detection Approach

The proposed novel insider threats detection modelling approach consists of two stages namely as training stage and testing stage as illustrated in Figure 1. The training stage only involved a process of computation for anomaly scores generation based on historical behavior while the testing stage involve a process of matching keywords, getting scores from the profile and further statistical analysis. The process of extracting an insider threats keywords is conducted based on previous studies indicators recommendation, due to the fact that previous work has focused on what and how the keywords that used in an email content could reflect an insider threats behavior. The entire bad keywords which represent an insider threats behavior that match with insider threat indicator criteria is extracted from an email contents before the related keywords divide into two different sets i.e., training dataset and testing dataset as in stage 1.1, stage 2.1, and stage 3.1, respectively. Training stage contained several components such as historical or known keywords of insider threats behavior (step 2.1), attack profile (stage 2.2) and anomaly score (step 2.3). In contrast, testing stage comprises an extracted known and unknown keywords of an insider keywords (step 3.1), function to obtained anomaly score allocation (step 3.2), a sequences of statistical analysis methods such as Friedman test (step 3.3), linear regression (step 3.4), correlation coefficient (step 3.5) and the detection output or detection file in Step 3.6.

Furthermore, in training stage, the filtered insider threats keywords (step 2.1) are further employed to form an attack profile as shown in step 2.2. The attack profile generated using relative percentage equation (1) and the entire keywords are categorized into three different factors based on three significant factors i.e., motive, capability and opportunity that contribute to the behavior of insider threats. Moreover, such malicious keyword behavior is difficult to be detectable and validate on massive email content by human expert and thus pattern of such behavior needs to be further analyzed using scoring approach. Due to this

reason, an anomaly scores are calculated based on the extracted known keywords and above-mentioned factors. In addition, scoring approach will facilitate in valuing the existence of an insider threats behavior in percentage for better understanding. The anomaly score is illustrated in Table 4 and calculated based on following equation:

$$\text{Anomaly Score}=\sum_{k=1}^{n} \ln\left(\frac{R_k}{N_k}\right) * 100 \,,$$

$$k = 1, 2, 3 \ldots . n \qquad (1)$$

Table 4 Anomaly Score

Factor indicator	R	N	Anomaly Score
Motive	21	50	0.284
Capability	7	50	0.382
Opportunity	12	50	0.334
Total			1

Based on Table 4, each factor indicator has different value of anomaly score such as motive is 0.284, capability is 0.382 and opportunity is 0.334. Once, these three factors indicator scores are combined, total anomaly score become 1. The value of 1 indicated the existence of an insider behavior keywords in that email is 100% (1*100). On the other hand, for the case of the total value of anomaly score is less than 1 the result could be not much significant. For example, let assume the total value of anomaly score is 0.666 (motive = 0.284 + capability = 0.328), indicated there is approximately only 66% of insider threats behavior exist in that specific email. Therefore, an additional approach is needed to discover and examine further the existence of insider threats behavior such as Friedman statistical, linear regression, correlation coefficient.

3.2 Friedman Statistical

The Friedman analysis employed in testing phase against known and unknown keywords to determine the null H_0 and alternative hypothesis H_1 as in Figure 1 step 3.3. One of the benefits of scoring methods is allow the string keyword available in score values and further analyzed statistically. Using Friedman analysis, the H_0 is defined when there is no different between each examined keyword (historical keywords from training phase equal to keywords from testing phase) and H_1 is defined when there is different between each keyword (historical keywords from training phase not equal to keywords from testing phase). Alternatively, in this research, H_0 is considered as a potential suspicious insider threat behavior is existed in extracted insider threats keywords while H_1 considered as any suspicious insider threats behavior do not take place. The equation for H_0 and H_1 is as follows:

$$H_0 : M_1 = M_2 = \ldots = M_k$$
$$H_1 : M_1 \neq M_2 \neq \ldots \neq M_k \qquad (2)$$

Therefore, defining null and alternative hypothesis H_0 or H_1 using Friedman statistical is important in proposed insider threats modelling approach. Besides, this analysis helpful in identifying either the extracted keywords reflect as an insider threats behavior or not. Figure 2 illustrate the flow chart of the employed Friedman statistical test.

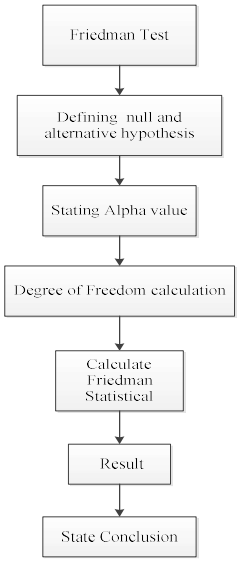

Figure 2 Flowchart of Friedman Statistical

Consequently, in this research five different critical value from chi-square distribution selected to state conclusion of null hypothesis such as 0.99, 0.95, 0.90, 0.75 and 0.50. Each one of alpha value represent the percentages of insider threats behavior portrayed in extracted insider threats keywords. For instance, 0.99 has the highest tendency of insider threats behavior while 0.50 has the lowest tendency of insider threats behavior in extracted insider threats keywords. The calculation of Friedman statistical is defined as:

Journal of Theoretical and Applied Information Technology
31st August 2021. Vol.99. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

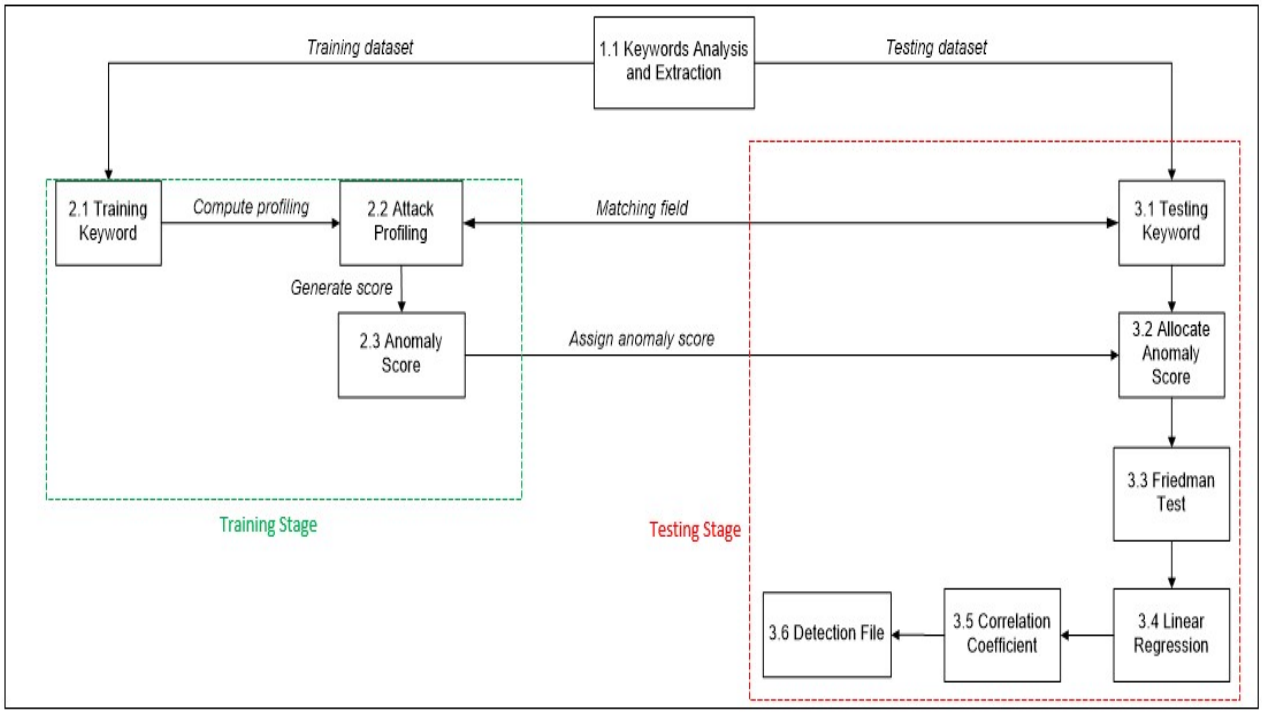

Figure 1 The Framework of Proposed Modelling Insider Threats Detection Approach

$$x_r^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1) \qquad (3)$$

Moreover, if H_0 is rejected, means there is no any insider behaviour comprises in testing sets or match with an malicious insider threats keywords behavior in training set. Conversely, if H_1 is rejected, means there is an insider behavior comprises in testing sets or match with a malicious insider threats keywords behavior in training set.

3.3 Linear Regression (R^2)

Linear regression (R^2) usually applies as it has ability to perform prediction analysis. The concept of linear regression applied in this research as to estimate the relationship of either an insider threats keywords with generated anomaly score in training stage (represent as the independent variable (x-axis)) and insider threats testing keyword with allocated anomaly score in testing stage (represent the dependent variable (y-axis)), are significantly indicate the behavior of insider threats or vice versa. The applied equation linear regression (R^2) is as follows:

$$y = c + b * x \qquad (4)$$

The relationship between each variable i.e., independent variable and dependent variable that represent insider threats behavior is graphically presented and analyzed. This is to identify which data points become constraint and containing outliers that will degrade the accuracy. Hence, after discovering such data's, correlation coefficient is further implemented to determine the relationship of above-mentioned variables i.e., strong or loosely coupled.

3.4 Correlation Coefficient

The correlation coefficient approach (as in step 3.5) in figure 1 is used to determine how strong a relationship between extracted insider threats keywords and insider threats behavior. Generally, using correlation coefficient, the function will return a number between -1 and 1. In this research, 1 indicated there is strong positive relationship between the extracted keywords and behavior of insider threats while number -1 indicated there is strong negative relationship and 0 between them indicated that there is no relationship ever exist. Comprehensively, if an applied correlation coefficient resulting in number 1, that means there is positive increased in variable (refer to insider threats keywords), while the fixed proportion in other variables also has positive increased (refer to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

insider threats behavior). For instance, when extracted insider threats keyword is strong enough to represent more than one of insider threats behavior from different factor indicator (motive or capability or opportunity), then the relationship between each is considered as strong positive relationship (1). Moreover, if the result is -1 means that for each one positive increased in one variable, there is negative decreased on other variable. For example, when the extracted insider threats keyword is relevant and represent not more than one element of insider threats behavior, then the relationship between each is considered as strong negative relationship (-1). In addition, number zero means that for each increased element, there is no positive or negative increased exist, for both related variables of insider threats. This will indicate that the extracted insider threats keyword is relevant to represent any insider threats behavior from different factor indicator (motive or capability or opportunity). The formula to determine the relationship of each email content presented as:

$$r = \frac{\sum((x-\bar{x})-(y-\bar{y}))}{\sqrt{\sum(x-\bar{x})^2 \quad \sum(y-\bar{y})^2}} \qquad (5)$$

In this research, the objective to apply correlation coefficient is for determine the estimation of strength relationship between variables, either it is strong or not. Therefore, whether the result of correlation coefficient is strong positive or strong negative, the relationship considered exists between the variables.

4. EXPERIMENT AND RESULT

4.1 CERT Dataset

CERT Insider Threat dataset is a collection of insider threats data that comprises synthetic background of a malicious insiders. The data supported through collaboration between ExactData, LLC and DARPA. The CERT dataset is analyzed, and the malicious email content is further filtered out as this research only focuses on email information. Subsequently, related bad keywords that has relation with a malicious activity is extracted based on previous study. Upon extracted all the keywords, the data divided into two sets i.e., training and testing. In training sets, malicious insider keywords used to profile an insider behavior into 3 categories of factor that contribute to an insider attack such as motive, capability and opportunity. Thereafter, an anomaly scores is generated for each of these factors using relative percentage ratio approach whereas later the similar scores is used in scoring testing sets insider keywords.

4.2 Evaluation of Proposed Modelling Insider Threats Detection Approach

The proposed insider threats detection modelling approach has been evaluated in a series of assessment as different variant of statistical analysis employed, i.e., Friedman, Linear Regression (R^2) and Correlation Coefficient. More specifically, for null hypothesis analysis using Friedman, five different threshold value has been considered as each contributes different result. The threshold value has been selected from chi-square distribution such as 0.99, 0.95, 0.90, 0.75 and 0.50 and this is due to fact to find the most suitable threshold that also been considered in previous literature (Xue & Sun 2015), (Chumachenko & Technology 2017), (Gibert et al. 2020). In addition, as the proposed modelling insider threats detection approach also focused on depth understanding and determination of malicious behavior of insider threats, the proposed approach is further measured using parameter of an attack detection rate. The attack detection rate parameter widely used in IDS field by various researcher such as ((Hilmi et al. 2017), (Salo et al. 2018), (Dwivedi et al. 2019), (Alsharafi et al. 2020), (Dwivedi et al. 2020)) as this parameter able to assess the capability or performance of the detection method in term percentage of detected attack and undetected attack.

4.3 Performance Evaluation

The evaluation of each factor indicator (motive, capability, opportunity) against Friedman statistical, linear regression R2 and linear regression R2 + correlation coefficient is illustrated in Figure 3. Each of the insider threats relationship between insider threats behavior and extracted insider threats keywords are represented and discussed. Based on analysis, the possibility of some relevant extracted insider threats keywords that represent insider threats behavior might not significantly or sensitive to be detected. This is because some of extracted insider threat keywords is not existing in training phase and exist during testing phase for analysis and prediction purpose (also known as unknown keywords). Therefore, this lead the single analysis approaches such as Friedman and linear regression (R^2) is less effective in detecting an insider threat behavior compared to the combined approach linear regression + correlation coefficient for each factor indicator.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

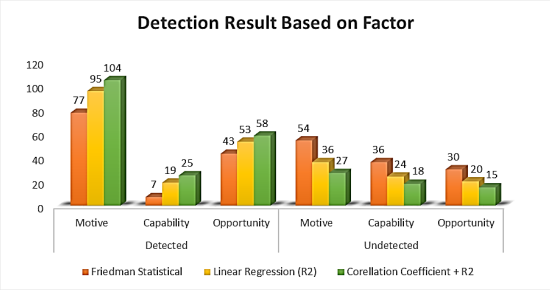

Figure 3 Detection Result Based on Factor for Each Statistical Approach

Based on analysis, using Friedman analysis, for the case of insider keywords that does not exist during training phase and appear during testing phase, every obtained critical values x^2 of null hypotheses, i.e., 0.02, 0.103, 0.201 and 0.575 for degree-of-freedom 0.99, 0.95, 0.90 and 0.75 respectively, the approach failed to differentiate between an historical insider threads keywords used in training phase and keywords represented during testing phase. Alternatively, it indicates there is no any insider threats behavior exist on that particular email in testing phase. Therefore, further analysis must be considered as this analysis have limitation in identifying unknown keywords. The second experiments are conducted using linear regression R^2 as this approach allow the relationship of factor indicators and malicious insider keywords presented graphically as illustrated in Figure 4. Based on analysis, as in figure 4, the unknown keywords have been detected as an outlier's data point. Besides, the relationship among historic or training insider malicious keywords scores and testing keywords scores are loosely or slightly lower as indicate by R^2 value 0.65. detect the seen and unseen malicious keywords behavior more correctly as compared to other approaches such as Friedman and linear regression R^2 individually.

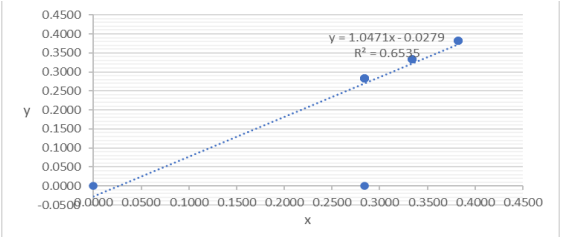

Figure 4 Linear Regression (R^2) Detection Illustration

Thus, there is needed for correlation coefficient as this approach have an ability to determine the strength of relationship specifically for overcome the limitation of unseen behaviour identification. Based on findings, the combination of R^2 and correlation efficient empower capability to identify seen and unseen behavior more accurately as illustrated in Figure 4.

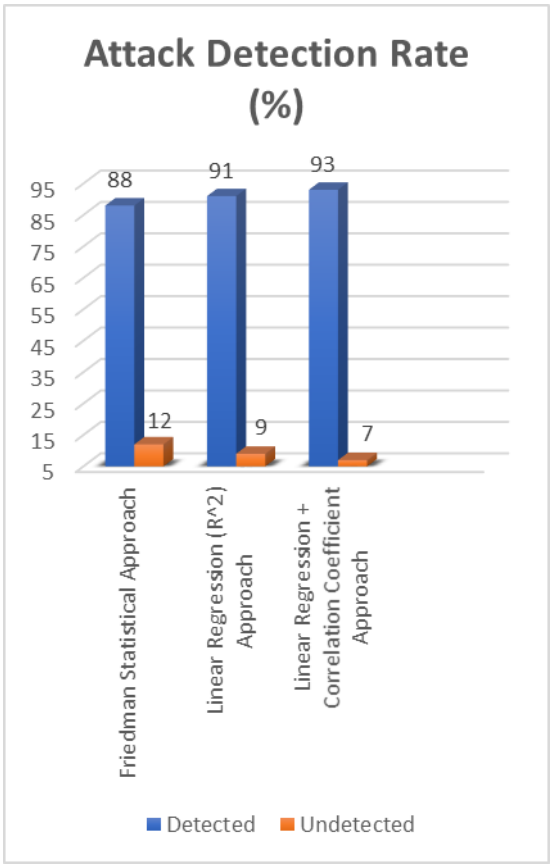

Figure 5 Attack Detection Rate (Correctly and Falsely Detected) of Proposed Modelling Insider Threats Detection Approach

The attack detection rate is calculated for every approach applied in proposed insider threats detection model to evaluate the effectiveness of their performance metrics in detecting the relationship between insider threats behavior and relevant extracted insider threats keywords. As shown in Figure 5, there are correctly detected and falsely detected of attack detection rate for each applied approach in proposed insider threats detection model. For example, the correctly attack detection rate for Friedman statistical approach is 88% and 12% of malicious relationship between insider behavior is falsely detected. Moreover, the linear regression approach results at 91% for cases of correctly detected while 9% falsely detected. Surprisingly, when correlation coefficient is applied together with linear regression approach the attack detection rate increased at 93% as correctly detected and 7% as falsely detected. The more

accurate approach is contributing by a combinational approach that is R^2 and correlation coefficient.

4.4 Performance Comparison Against Previous Approaches

TABLE 5 COMPARISON OF PREVIOUS INSIDER THREATS DETECTION METHODS AND PROPOSED METHOD

Author	Detection Method		Detection rate (%)
	Scoring	Statistical Approaches	
Goldberg et al. 2017	√	Statistical-outliner Technique, Pearson R Correlation	90
Gamachchi et al. 2018	√	-	85
Kim et al. 2019b	√	-	70
Le & Zincir-Heywood 2019	√	Linear Statistical Model	90
Jang et al. 2020	√	-	Not stated
Proposed method	√	Friedman Statistical, Linear Regression, Correlation Coefficient	93

Numerous previous researchers have proposed an insider threats detection that comprises a scoring and statistical approaches using similar CERT dataset as in Table 5. For instance, author Goldberg et al. (2017) has employed a scoring method and statistical approaches such as correlation as a threshold. The aim is to detect an insider threats behavior and the proposed solution has managed to obtain 90% as a detection rate. Moreover, Gamachchi et al. (2018) and Kim et al. (2019b) also employed a scoring approaches as a detection method in their works. The scoring approaches that been introduced by Gamachchi et al, (2018) has recorded 85% as the detection rate while Kim et al., (2019) achieved 70%. Furthermore, in 2019, Le & Zincir-Heywood has applied a scoring method with linear statistical model in identifying an insider threats data. This combination has contributed to 90% of detection rate. Even though, various researcher has proposed some solution in identifying an insider threats relation lately, but the achievement is not satisfactory and still have room for improvement. Besides, detecting an insider threats relationship need continuous relevant approach in order to recognized and acknowledge such malicious relationship accurately. Hence, the proposed detection modelling is implemented in variant of statistical analysis and scoring approach. The proposed modelling insider threats detection approach has achieved higher attack detection rate compared to other previous method which is at 93%. The applied scoring method against extracted keywords into three major factor indicators i.e., motive, capability and opportunity has contributed to value the entire keywords in scores in prior. Subsequently, the scored keywords have facilitated the statistical analysis to discover deeply into the behavior of an insider. Thus, the detection is more accurate and result with higher attack detection rate.

5. CONCLUSION AND FUTURE WORK

An insider threats known as a person that has a legitimate access towards legal information and assets of an organization, whereas he/she could easily bypass the traditional malicious detection method and performed malicious activity. In addition, as malicious insider has technical skills and knowledge about the vulnerability of on such valuable and confidential assets and information, they can launch threats with greater damages against an organization. Moreover, various insider threats attempt could be launch using various platform and the one that is more potential is via an email. Usually, based on previous literature study, an email content contained various keywords and wrote by many people including malicious insider to express their emotion or feelings and each content could reflect an individual behavior. Therefore, an email contents keywords can be analyzed to understand the behavior of malicious insider and detect it before he/she could launch harmful intentions towards valuable assets. Hence, a novel insider threats detection modelling that able to analyze and detect an insider threats behavior based on analysis of email content keywords using scoring and statistical approaches has been proposed. Throughout these various approaches i.e., scoring method, Friedman statistical, linear regression (R^2) and correlation coefficient, the insider threats behavior has been detected more accurately, in which 93% as an attack detection rate as compared to the previous recommended methods that is below 90%. For example with applied scoring method, proposed insider threats detection method able to detect insider threats relationship between insider threats behavior and insider threats keywords which represent by score. With score generated, each

insider threats existed could be represent, respectively. In addition with Friedman statistical, the insider threats detection could be statistically be represent as followed from scoring based before where behavior insider represent. The applied Friedman statistical able to statistically shows existed insider threats relationship between insider behavior and insider threats keywords through null hypothesis. In short, helpful in identifying either the extracted keywords reflect as an insider threats behavior or not statistically. Besides that, applied correlation coefficient able to detect the strength of insider threats relationship included the undetectable statistically by Friedman. The strength of insider threats is important to be detected as although previous method unseen the insider threat relationship, the strength of such malicious relationship could still be detected with correlation coefficient. These applied methods in proposed insider threats detection method shows the increasing percentage of insider threats detection where each applied method played important roles in recognized insider threats. The proposed modelling insider threats detection approach could be use as reference model to mitigate insider threats activities in this digitalization era as an email become high priority communication channel todays. In future, the proposed insider threats detection modelling approach could be implemented as an automated detection with big data and deep learning concerns. This solution could benefit an organization in developing an efficient cyber defense system against an insider threat. However, there are some limitation of this research where there is lack of studies on insider threats issues as most of the research focuses on outsider threats detection, lack of studies on detecting insider threats relationship analyzed from relevant insider threats keywords and difficult to detect specific behavior of malicious insiders as reflected in relevant extracted insider threats keywords accurately.

ACKNOWLEDGEMENT

This publication has been supported by Center of Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES:

[1] Alhanahnah, M.J., Jhumka, A. & Alouneh, S., 2016. A multidimension taxonomy of insider threats in cloud computing. *Computer Journal*, 59(11), pp.1612–1622.

[2] Alsharafi, W.M. et al., 2020. Normal Profile Updating Method for Enhanced Packet Header Anomaly Detection Normal Pro fi le Updating Method for Enhanced Packet Header Anomaly Detection. , (January).

[3] Baumgarten, M. et al., 2013. Keyword-Based Sentiment Mining using Twitter. , 5(June), pp.2–5.

[4] Biringer, B.E., Vugrin, E.D. & Warren, D.E., 2019. *Critical Infrastructure Resilience to Insider Risk*,

[5] Christian W. Probst, Jeffrey Hunker, Dieter Gollmann, and M.B., 2010. Aspects of Insider Threats. *Journal of Clinical Sleep Medicine*, 5(3), pp.277–279.

[6] Chumachenko, K. & Technology, I., 2017. Machine Learning Methods For Malware Detection.

[7] Claycomb, W.R. & Nicoll, A., 2012. Insider Threats to Cloud Computing : Directions for New Research Challenges.

[8] Dwivedi, S. et al., 2019. Implementation of adaptive scheme in evolutionary technique for anomaly - based intrusion detection. *Evolutionary Intelligence*, (0123456789). Available at: https://doi.org/10.1007/s12065-019-00293-8.

[9] Dwivedi, S., Vardhan, M. & Tripathi, S., 2020. An effect of chaos grasshopper optimization algorithm for protection of network infrastructure. , 176(April).

[10] Ekran, S., 2016. Insider Threats In It Infrastructure : Landscape , Sources , And Trends.

[11] Elmrabit, N., Yang, S. & Yang, L., 2015. Insider Threats in Information Security.

[12] Form, L.M. et al., 2015. Phishing Email Detection Technique by using Hybrid Features.

[13] Gavai, G., Sricharan, K., Gunning, D., Rolleston, R., et al., 2015. Detecting Insider Threat from Enterprise Social and Online Activity Data. *Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats - MIST '15*, pp.13–20. Available at: http://dl.acm.org/citation.cfm?doid=2808783.2808784.

[14] Gavai, G., Sricharan, K., Gunning, D., Hanley, J., et al., 2015. Detecting Insider Threat from Enterprise Social and Online Activity Data ∗. , pp.13–20.

[15] Georgia Lykou, A.A. and D.G., 2019. Smart Airport Cybersecurity : Threat Mitigation and Cyber Resilience Controls †.

[16] Gibert, D., Mateu, C. & Planes, J., 2020. Journal of Network and Computer Applications The rise of machine learning for detection and classification of malware : Research developments , trends and challenges. *Journal of Network and Computer Applications*, 153(January), p.102526. Available at: https://doi.org/10.1016/j.jnca.2019.102526.

[17] Greitzer, F.L. et al., 2012. Identifying At-risk Employees : Modeling Psychosocial Precursors of Potential Insider Threats.

[18] Greitzer, F.L. et al., 2016. Psychosocial Modeling of Insider Threat Risk Based on Behavioral and Word Use Analysis 1. , 9(1), pp.106–138.

[19] Greitzer, F.L., Ph, D. & Hohimer, R.E., 2011. Modeling Human Behavior to Anticipate Insider Attacks Modeling Human Behavior to Anticipate Insider Attacks. , 4(2), pp.25–48.

[20] Heneke, D., Ophoff, J. & Stander, A., 2016. The Threats that Insiders Pose to Critical Infrastructure – A South African Perspective. , (Haisa), pp.279–289.

[21] Hilmi, M. et al., 2017. Original citation : A New Unified Intrusion Anomaly Detection in Identifying Unseen Web Attacks. , pp.1–18.

[22] Ho, S.M., Hancock, J.T. & Booth, C., 2017. Ethical dilemma: Deception dynamics in computer-mediated group communication. *Journal of the Association for Information Science and Technology*, 68(12), pp.2729–2742.

[23] Ho, S.M., Kaarst-Brown, M. & Benbasat, I., 2018. Trustworthiness attribution: Inquiry into insider threat detection. *Journal of the Association for Information Science and Technology*, 69(2), pp.271–280.

[24] Homoliak, I. et al., 2018. Insight into Insiders: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures. Available at: http://arxiv.org/abs/1805.01612.

[25] Huff, J. et al., 2019. NATO Human View Executable Architectures for Critical Infrastructure Analysis. *Engineering Management Journal*, 00(00), pp.1–22. Available at: https://doi.org/10.1080/10429247.2019.1656594.

[26] Isotopes, E. et al., 2017. The Enemy Within The Connection Between Insider Threat and Terrorism. *Food Chemistry*, pp.1–42.

[27] Jiang, J. et al., 2018. Prediction and Detection of Malicious Insiders ' Motivation based on Sentiment Profile on Webpages and Emails. *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, pp.1–6.

[28] Julian Jang-Jaccard, S.N., 2014. A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, 80(5), pp.973–993. Available at: http://dx.doi.org/10.1016/j.jcss.2014.02.005.

[29] Kharde, V.A., 2016. Sentiment Analysis of Twitter Data : A Survey of Techniques. , 139(11), pp.5–15.

[30] Kim, J. et al., 2019. applied sciences Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms.

[31] Kim, J. et al., 2018. Patent Keyword Extraction for Sustainable Technology Management. , pp.1–18.

[32] Kramer, L.A. & Crawford, K.S., 2005. Technological , Social , and Economic Trends That Are Increasing U . S . Vulnerability to Insider Espionage. , (May).

[33] LaRosa et al, 2019. System For Detecting Fraudulent Electronic Communications Impersonation , Insider Threats And Attacks. , 1.

[34] Lu, J. & Wong, R.K., 2019. Insider Threat Detection with Long Short-Term Memory.

[35] Maasberg, M., Warren, J. & Beebe, N.L., 2015. The Dark Side of the Insider : Detecting the Insider Threat Through Examination of Dark Triad Personality Traits.

[36] Majeed, A., 2016. Internet of Everything (IoE) Exploiting Organisational Inside Threats : Global Network of Smart Devices (GNSD).

[37] May, C.R. et al., 2017. Insight into Insiders: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures.

[38] Mohammad, S.M., 2016. *9. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text*, Elsevier Ltd. Available at: http://dx.doi.org/10.1016/B978-0-08-100508-8.00009-6.

[39] Musuva, P.M.W., Getao, K.W. & Chepken, C.K., 2018. *A New Approach to Modelling the Effects of Cognitive Processing and Threat Detection On Phishing*, Elsevier B.V.

[40] Nkosi, L., Tarwireyi, P. & Adigun, M., 2014. Insider Threat Detection Model for the Cloud. , 062500001.

[41] Nurse, J.R.C. et al., 2014. Understanding Insider Threat : A Framework for Characterising Attacks.

[42] Nurse, J.R.C. et al., 2014. Understanding insider threat: A framework for characterising attacks. *Proceedings - IEEE Symposium on Security and Privacy*, 2014–Janua, pp.214–228.

[43] Permissions, F., 2016. A Multidimension Taxonomy of Insider Threats in Cloud Computing.

[44] Pfleeger, S.L. et al., 2010. Insiders behaving badly: Addressing bad actors and their actions. *IEEE Transactions on Information Forensics and Security*, 5(1), pp.169–179.

[45] Probst, C.W. et al., 2008. Countering Insider Threats. , pp.1–18.

[46] Robert Ambrogi, 2019. State of Insider Threats in the Digital Workplace.

[47] Salo, F., Nassif, A.B. & Essex, A., 2018. Dimensionality Reduction with IG-PCA and Ensemble Classifier for Network Intrusion Detection. *Computer Networks*. Available at: https://doi.org/10.1016/j.comnet.2018.11.010.

[48] Sarkar, K.R., 2010. Assessing insider threats to information security using technical , behavioural and organisational measures. *Information Security Technical Report*, 15(3), pp.112–133. Available at: http://dx.doi.org/10.1016/j.istr.2010.11.002.

[49] Scott, J. & Spaniel, D., 2017. In 2017 , The Insider Threat Epidemic Begins.

[50] Seaman, M.J. & San Jose, M., 2015. Processing Of Email Based On Semantic Relationship Of Sender To Recipient. , 2(12).

[51] Shaw, E.D. & Fischer, L.F., 2005. Ten Tales of Betrayal : The Threat to Corporate Infrastructures by Information Technology Insiders Analysis and Observations. , (September).

[52] Tan, S., 2019. Unified Psycholinguistic Framework : An Unobtrusive Psychological Analysis Approach Towards Insider Threat Prevention and Detection. , 7(1), pp.52–71.

[53] To, P., 2015. Analytic Approaches To Detect Insider Threats.

[54] UK, G.B.G.C.H.C.E.R.T., 2015. Common Cyber Attacks : Reducing The Impact.

[55] Vidyapeetham, A.V., 2018. A Machine Learning approach towards Phishing Email Detection.

[56] Wang, H., 2018. Edge Content Enhanced Network Embedding. *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp.900–907.

[57] Wood, P. et al., 2016. Internet Security Threat Report. , 21(April).

[58] Xiaojun, C. et al., 2013. An Intent-Driven Masquerader Detection Framework Based on Data Fusion. , pp.450–457. Available at: http://link.springer.com/10.1007/978-3-642-35795-4_57.

[59] Xie, L., Liu, Y. & Chen, G., 2015. A Forensic Analysis Solution of the Email Network Based on Email Contents. , pp.1613–1619.

[60] Xue, L. & Sun, G., 2015. Design and implementation of a malware detection system based on network behavior. , (June 2014), pp.459–470.

[61] Yoshinaga, N. et al., 2010. Content Propagation Analysis of Email Communications.