Towards Insider Threat Detection Using Psychophysiological Signals

Yessir Hashem, Hassan Takabi, Mohammad GhasemiGol, Ram Dantu
Department of Computer Science and Engineering
University of North Texas, Denton, TX, USA
YassirHashem@my.unt.edu, {Takabi, Mohammad.ghasemigol,Ram.Dantu}@unt.edu

ABSTRACT
Insider threat is one of the greatest concerns for the information security system that could cause greater financial losses and damages than any other attacks. Recently many studies have been proposed to monitor and detect the insider attacks. However, implementing an effective detection system is a very challenging task. In this paper, we investigate the usability of human bio-signals to detect the malicious insiders in real time. We present an insider threat monitoring and detection framework based on the electroencephalography (EEG) signals to distinguish between normal and malicious activities. We describe the framework and its components. We then evaluate the proposed framework using several real world scenarios. The results show that the detection accuracy of the malicious activities is up to 90% and demonstrate that electroencephalography (EEG) can reveal valuable knowledge about the user behaviors and could be a very effective solution for detecting insider threats.

Categories and Subject Descriptors
• K.6.5 Security and Protection - *Unauthorized access*

General Terms
Security, Experimentation, Human Factors

Keywords
Insider threat detection, electroencephalograph, brain computer interface, physiological indicators.

1. INTRODUCTION
Threats from the inside of an organization's perimeters are a significant problem since it is difficult to distinguish them from benign activities. The insider threat has long been a prime security concern for government and industry organizations. It is also considered the most difficult problem to deal with because insiders often have information and capabilities not known to external attackers, and as a consequence, can cause serious harm [1]. Little is known about the insider threat, and the threat of insider activities continues to be of paramount concern in both the public and private sectors [2]. However, many surveys have been done that demonstrate the severity of insider threat. For example, the Cybercrime report by PwC states that the most serious fraud cases were committed by insiders [3]. A report recently produced by the security management company AlgoSec found that a significant proportion of security professionals view insider threat as their greatest organizational risk [4]. In 2015, a Federal cybersecurity survey of 200 federal IT managers shows that 76% of the participants are concerned about leaks from insider threats (untrained and malicious insiders) [5].

MIST'15, October 12-16, 2015, Denver, CO, USA

DOI: http://dx.doi.org/10.1145/2808783.2808792

The insider threat has been subject of extensive study and many approaches from technical perspective to behavioral perspective and psychological perspective have proposed to detect or mitigate insider threat [6]. However, implementing effective insider threat detection or monitoring system is a very challenging task. There have been several detection approaches that aim to detect insiders by monitoring the network activities and using organization resources [7]. However, it is recognized that solutions to insider threat are mainly user centric and several psychological and psychosocial models have been proposed [8]. With the growing interest in psychological aspects of cyber security, researchers are concerned with identifying predictors of these behaviors. However, most of these models rely on humans to recognize the signs and record the behaviors to detect insider threat. Further, these models tend to rely on detecting insider's voluntary behaviors that could potentially fail to detect individuals who are capable of feigning normal behavior. Other behavioral approaches aim to assess cybersecurity violations through changes in user's activities and require the user to interact with computer peripherals (e.g., mouse; keyboard). Behavioral data, when used in isolation, are highly susceptible to influences outside the user's own targeted attitudes. On the other hand, psychophysiological metrics such as electroencephalography (EEG), electrocardiogram (ECG), galvanic skin response (GSR), etc. provide a number of advantages over behavioral assessments alone. The psychophysiological signals are continuously available, very difficult if not impossible to mimic or change as they are generated involuntarily, and can also be measured automatically.

In this paper, we explore the use of the human bio-signals in building a framework to monitor and detect the malicious insider threats. More specifically, we focus on using electroencephalogram (EEG) signals that arise from the user's brain activities to measure the changes in his normal bio-signals pattern. In other words, we propose an insider threat monitoring system based on the EEG signals pattern. The proposed system analyzes the user's bio-signals, extracts features, and classifies them in order to detect the malicious activities. Our experiments involve human subjects to collect the signal samples and evaluate the proposed monitoring framework.

The rest of the paper is organized as follows. In section 2, we present the system design including the signal pre-processing and

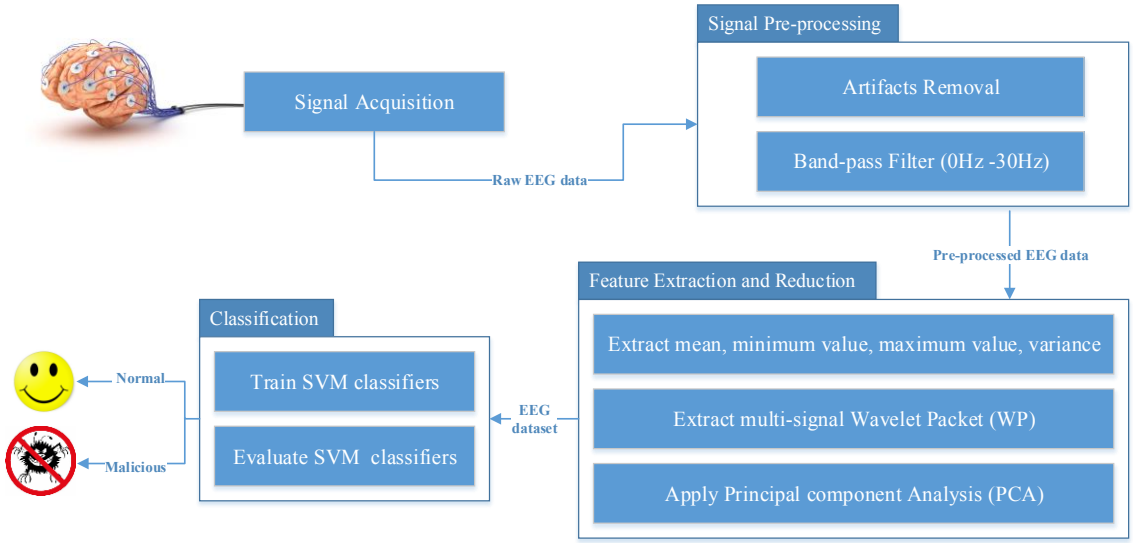

Figure 1. The proposed framework to detect insider attacks.

feature extraction components. In section 3, we describe the experiment setup, evaluation and results. Finally section 4 discusses the future work, and concludes the paper.

2. SYSTEM DESIGN

At the high level, the proposed framework design consists of four major components as shown in Figure 1. The first component is the signal acquisition unit and records users' EEG signals in real-time. The signal filtering unit is responsible of pre-processing the recorded EEG signals and removing any noise or artifacts to prepare them for further process. It also samples the recorded signals into specific time frames which represent the time window we monitor the users brainwave signals. The feature extraction component applies various algorithms to each sample to extract features that will be used for detection of the malicious activities. In our experiments, we use two activity tasks: the regular and the malicious activities, the regular activities recorded the subject EEG signals while doing regular office job activities and the malicious activities recorded the subject EEG signals while doing malicious tasks represent an insider threat attack. The labeled features vector is then feed to the classifier component as training set. We use support vector machines (SVMs) [11] to distinguish between the regular activity and the malicious activity. Our trained classifier is able to detect any insider threat on the system before an incident is executed and send the results to the monitoring alarm system. In the followings, we describe each component in detail.

a. EEG Signal Acquisition

There are a number of relatively low cost and low risk EEG devices that could be used to acquire brainwave signals. In this study, we use a consumer-grade BCI device developed by Emotive [9]. Emotiv EPOC is one of the most widely studied of the inexpensive off-the-shelf EEG systems. It is a compact, wireless headset that requires comparatively little effort to set up, and equipped with 16 sensors out of which 14 are used to record brainwave signals from different parts of the brain.

b. Signal pre-processing

The EEG signals locate in very low frequencies (between 0.1 Hz to 50 Hz) and very low electrical activity power measured by microvolt. When we record the brainwave signals, typically there are some noises due to unintentional movements that can cause unwanted data in the EEG signals. In order to detect clear EEG signals, we have to filter the signals from the high frequencies and remove any artifacts. First, we removing the higher frequencies using band-pass filter and selecting our frequency range (0.1-30 Hz). We then remove the electrooculogram (EOG) artifacts that could be included in the signals during the recording.

c. Feature extraction

The features extraction component is the most important part of the framework, where we extract useful information from the EEG signals. We use time domain and frequency domain to extract the features by applying linear and nonlinear measures. We use wavelet packet decomposition (WPD) method to decompose the EEG frequencies into different frequency subsets and extract features from each subset. This is done when the wavelet packet breaks down the signal to a number of sub-bands using the wavelet function. This could be presented as a sub-band tree where each level of the tree is computed by passing the previous approximation coefficients over high and low pass filters. In our case, we decompose our filtered EEG signals with the frequency of 30 Hz to smaller sub-bands and get the energy for each band. To do this, we use three levels of decomposition that give us six different bands. In the time domain, the features are extracted from all the recording electrodes (channels) for the same time frame and added to one feature vector representing that specific time frame. We also extract other features such as Microvolts (µV) mean value, maximum µV, minimum µV, number of peaks, the distance between the high and low µV, from each five-second time frame. Since the number of extracted features is high and not all of them are relevant in the analysis process, we choose only the features that have the most impact on the final results. We apply principal component analysis (PCA) [10] to reduce the number of features, decrease redundancy and consequently improve the system performance.

d. Classification

Once the features are extracted and reduced, we apply classification algorithms to these features to distinguish between regular activities and malicious activities. In this work, we apply support vector machines (SVMs), one of the most well-known and powerful classifiers used widely in machine learning. Before feeding our extracted features vector to the classifier, we create the

training set by labeling each feature vector by its activity task based on the experiment explained in the next section.

3. EXPERIMENTS AND EVALUATION

The goal of our experiments is to establish an approach to distinguish between normal activities and malicious activities for the purpose of insider threat detection. To do this, we recorded EEG signals of numbers of participants while performing three different tasks. Each experiment emulates a real-life scenario and we tried to choose scenarios are very close to a normal work environment and as realistic as possible. The EEG data was collected from participants who were recruited from the university community. We had a total of 10 subjects of which 5 were male and the other five were female. All the subjects were between the age of 18 and 33 years old.

The experiments were done for each participant separately and at different times during the day. Upon arriving at the testing lab, the participants were briefed on the objectives of the study and given a written informed consent explaining the experiment procedure and their right on participating to read and sign. Once the consent was obtained, the participants were seated in a comfortable chair, the Emotive EEG headset was positioned on the participant's head. The examiner verified impedances in connections between each electrode and the participant's scalp. The experiment was divided into 10 minutes activity sessions for three sessions: one session as a regular activity task and two sessions as a malicious activity tasks. In the first scenario, the participants perform regular office job activities such as browsing the internet, using computer applications or using the email account. This scenario observes the brain reaction to the regular daily activities done by most of the employees in any organization. In the second and third scenarios, the participants were asked to perform malicious activities by trying to access to information they were not authorized for. We used two realistic scenarios when employee uses the remote control access or the network to access to unauthorized information. We recorded EEG signals generated by the brain in reaction to these malicious activities during each experiment. The EEG signals recorded during these experiments are used by our framework to extract the features and train the classifier. In the following, we describe each task in more detail.

Regular activity task: this task requires the participants to do the following activities for a period of 10 minutes:

- Browse the internet and find answers for a set of questions provided to him/her by the investigator.
- Use Microsoft Word to type the results.
- Attach the results and send to the investigator by email.

These tasks cover three normal office job activities, namely browsing the internet, using computer applications such as Microsoft office, and sending an email.

Malicious activity task I: participants have to connect remotely to two computers in the same room and do the following:

- Login to the computers remotely and try to find a folder containing student's grades excel file. Search the file to find the information for some specific students (list of students names are given to the subject during the explanation part).
- Copy the information to a new file in the experiment computer desktop. Close the files and exit without leaving any trace and making the account owner notice the access.

The participants were asked to perform these tasks without the computer account owners noticing and as an incentive, they received extra reward if they could complete these tasks without living any trace. To make the situation as realistic as possible, the account owners were physically present in the same room but away from their computer monitors.

Malicious activity task II: participants were told that a project's source code was located somewhere on the network and they were going to make a copy of the code without anyone noticing. They needed to perform the following computer activities for a period of 10 minutes:

- Search the lab network and find the project's code folder, access the folder, and copy the files from the folder to their computer's hard drive. In order to avoid raising suspicious of the network administrator, they were told that no file larger than 1 MB should be transferred. If the file was larger than 1 MB, the participants had to split the file into smaller files and transfer them separately.
- By the end of the task, the participant needed to close the files and the folder and exit without leaving any trace.

The participants were asked to perform the task without alarming the network administrator who was in the same room. The participants needed to be careful and copy a small amount of data for each transaction (less than 1MB). The participant were told that a successful transfer of the files without leaving any visible trace will get them an extra reward.

3.1 EXPERIMENTAL RESULTS

After the experiments were performed and the EEG data was collected, we organized the recorded signals into three groups: regular activities group, malicious I activities group, and malicious II activities group. Each group contains ten signals samples represented by each participate in the experiments. We then analyzed and pre-processed the signals by removing the high frequency noise and the electrooculogram (EOG) artifacts. We applied our feature extraction algorithm for each five second timeframe and labeled the result features vector with its represented group type. Then we applied the SVM using k-fold cross validation (k=5) to evaluate the accuracy of the SVM to classify normal and malicious activity. We use several well-known and widely used metrics to evaluate the SVM classifier. All of these evaluation metrics can be derived from four combinations of a classifier result: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A TP occurs when a classifier correctly classifies an insider intrusion, whereas a FP occurs when a normal action is misclassified as being an intrusion. Likewise, a TN is generated whenever a normal action is correctly classified, while a FN occurs when an intrusion is not detected by the classifier. According to these variables, we define Accuracy, Precision, Recall, F-measure, and Error-rate.

Table 1 shows the results in details for the EEG dataset containing the first malicious scenario. The results show that our approach is able to detect the malicious activities with 84% accuracy. Additionally, the results show that our framework recognizes regular activities with 84% accuracy. The results also show around 84% precision and about 85% recall in detecting malicious activities. Similar to the first test, we applied the SVM on EEG dataset containing the second malicious activities scenario. The results show 90% accuracy in detecting both regular and malicious activities for EEG dataset containing the second malicious activity scenario as shown in Table 2. In the third test, we applied the SVM

on both malicious scenarios to evaluate the detection accuracy for normal and malicious EEG signals in general. Figure 2 shows that the best results are achieved when sigma (kernel function parameter) is set to 5 and the classification accuracy around 86% in detecting both regular and malicious activities as show in Table 3. In general, our results shows detecting accuracy above 84% for all the scenarios that demonstrates electroencephalography (EEG) can reveal valuable knowledge about the user attacks behaviors and could be a good solution for the continues insider threat monitoring.

Table 1. SVM results for EEG dataset containing the first malicious scenario.

	Classification Accuracy	precision	recall	F-measure	Error Rate
Normal Class	0.8445	0.8383	0.8231	0.8513	0.1555
Malicious Class	0.8390	0.8430	0.8557	0.8320	0.1610

Table 2. SVM results for EEG dataset containing the second malicious scenario.

	Classification Accuracy	precision	recall	F-measure	Error Rate
Normal Class	0.8995	0.8874	0.8929	0.8969	0.1005
Malicious Class	0.8960	0.9090	0.9045	0.8985	0.1040

Table 3. SVM results for EEG dataset containing both malicious scenarios.

	Classification Accuracy	precision	recall	F-measure	Error Rate
Normal Class	0.8501	0.8865	0.7734	0.8876	0.1499
Malicious Class	0.8555	0.7797	0.8877	0.7856	0.1445

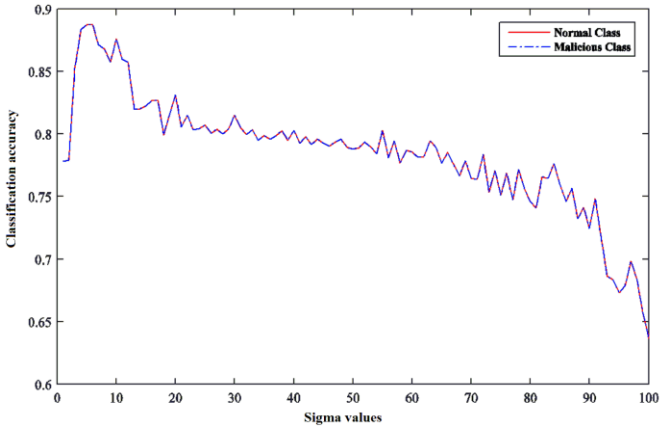

Figure 2. Containing both malicious scenarios. Best results tuning the RBF kernel parameter in EEG dataset are occurred for Sigma=5.

4. CONCLUSION AND FUTURE WORK

Insider threat is one of the greatest concerns for the information security domain. However, implementing an effective insider threat monitoring and detection system is a very challenging task. In this paper, we proposed a real-time insider threat monitoring framework based on the EEG signals. The system analyzes users' bio-signals, extracts features, and classifies them in order to detect if what the users do is normal activities or malicious activities. We evaluated our framework by conducting an experiment including ten subjects in three different scenarios that indicate regular and malicious activities. The results show that our framework is able to detect the malicious insider with average detection accuracy 86%. As for future work, we plan to combine the EEG signals with other biological signals such as the electrocardiogram (ECG) signals or the Electromyography (EMG) signals and evaluate the proposed framework with different insider threat scenarios. The ultimate goal is to implement a bio-signals based insider threat monitoring and detection system that is able to detect the insider malicious activities with excellent accuracy and reliability.

REFERENCES

[1] Bertino, E., & Ghinita, G. (2011). Towards mechanisms for detection and prevention of data exfiltration by insiders: keynote talk paper. InProceedings of the 6th ACM Symposium on Information, Computer and Communications Security (pp. 10-19). ACM.

[2] Glasser, J., & Lindauer, B. (2013). Bridging the gap: A pragmatic approach to generating insider threat data. In Security and Privacy Workshops (SPW), 2013 IEEE (pp. 98-104). IEEE

[3] Cybercrime: Protecting against the growing threat - Events and Trends, vol. 256, 2012.

[4] AlgoSec. The State of Network Security 2013: Attitudes and Opinions. AlgoSec, Inc., 2013. http://www.algosec.com/resources/files/Specials/Survey%20files/State%20of%20Network%20Security%202013_Final%20Report.pdf

[5] New Market Research - SolarWinds Survey Investigates Insider Threats to Federal Cybersecurity." Federal and Government Discussions. N.p., n.d. Web. 25 Mar. 2015. https://thwack.solarwinds.com

[6] Salem, M. B., Hershkop, S., & Stolfo, S. J. (2008). A survey of insider attack detection research. In Insider Attack and Cyber Security (pp. 69-90). Springer US.

[7] Hunker, J., & Probst, C. W. (2011). Insiders and insider threats an overview of definitions and mitigation techniques. Journal of Wireless Mobile Network ,Ubiquitous Computing, and Dependable Applications, 2(1), 4-27.

[8] Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2012, January). Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 2392-2401). IEEE.

[9] "Emotiv Systems," https://emotiv.com/ [last accessed: July 20, 2015].

[10] Jolliffe, I. (2002). Principal component analysis. John Wiley & Sons, Ltd.

[11] Vapnik, V. (2013). The nature of statistical learning theory. Springer Science & Business Media.