Prediction and Detection of Malicious Insiders' Motivation based on Sentiment Profile on Webpages and Emails

Jianguo Jiang[1], Jiuming Chen[1,2], Kim-Kwang Raymond Choo[3], Kunying Liu[1], Chao Liu[1], Min Yu[1,2*], Prasant Mohapatra[4]

[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
Email: {chenjiuming, jiangjianguo, liukunying, liuchao, yumin}@iie.ac.cn

[3]Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249, USA
Email: raymond.choo@fulbrightmail.org

[4]Department of Computer Science, University of California, Davis, USA
Email: pmohapatra@ucdavis.edu

***Abstract*—Recent high profile data breaches have highlighted the importance of insider threat detection research for cyber security. Anomaly based insider detection approaches are generally associated with high false positives; thus, there has been increased focus on including prediction of user psychology and attack motivations. However, data relating to psychological profile and personality trait of employees are challenging to collect, and do not generally adequately capture attack motivations such as disgruntlement (e.g. towards certain behavior). Therefore, in this paper, we demonstrate how one can build a user psychological profile based on the sentiment analysis of their network browsing and email content. We then evaluate our approach using real-world datasets, and the findings suggest that our approach can proactively and accurately detect malicious insiders with extreme or negative emotional tendencies. This is the first work to build user profile and predict insider threats using sentiment analysis of their browsing and email content.**

Keywords—Insider Threat Detection; Insider Attack Motivation; Sentiment Analysis; User Psychological Profile

I. INTRODUCTION

On November 5, 2009, Major Nidal Hasan [16] opened fire at the Soldier Readiness Center and killed thirteen people. The investigation found that he had came into contact with radical Islamist elements and frequently browsed websites about suicide bombers from his computer and email account. It could be speculated that the tragedy may have been avoided if such behavior could be proactively predicted and detected. The increasing number and frequency of insider-related incidents, such as information theft, sabotage and fraud, reinforce the risk of malicious insiders to the organization they work for.

To detect and mitigate insider threats, the research community and many organizations have put forward a (large) number of models and systems to characterize and detect insider threats, such as anomaly-based or features-based methods [3, 20], scenario-based methods [4], graph-based methods [5] and game theory-based methods [6]. These methods generally focus on analyzing the behaviors of users. However, insider threats cannot be simply detected by only considering the anomalous behaviors of users, particularly in the constantly evolving threat landscape.

A recent trend is to integrate information such as psychological features and other subjective factors to detect malicious insiders [7, 8]. Such approaches focus on extracting and analyzing of the indicators associated with user psychology and emotion such as user pressure, user satisfaction and user emotions. Although these indicators can be useful in predicting malicious insiders, the research in this field is still in its preliminary stage due to challenges in obtaining relevant user data (e.g. data indicating user emotion and psychology tendencies). In addition, such user data are subjective and can be easily forged or distorted.

Social network data generally include users' network browsing, messages and other user generated content can be a good source of indicators of user psychology and personnel traits [9]. In addition, a user who became disillusioned, angry, and so on, generally displays obvious behavioral changes that may be reflected in their network browsing and email content [10, 11]. Therefore, it is no surprise that many insider threat detection schemes are based on analyzing user network browsing history and email records. Findings from the analysis are then used to build the users' behavior profiles, and any deviation from the profiles will be flagged as (potentially) malicious. However, these approaches mostly focus on the anomaly in communication patterns between websites and emails, without considering their content. Although there have been efforts to characterize the relationship between users' network content with their personality traits, say using the OCEAN model [12], such approaches do not consider attack motivations such as negative emotions and extreme psychology tendencies.

A report from FBI [13], for example, explained that a negative / toxic workplace can often fuel feelings of revenge or disgruntlement, particularly if that particular individual had been wronged, and this and financial-related issues are the most frequently reported motivations in insider threats. Gavai et al. [11] explained that malicious insiders with such motivations would have obvious emotional tendencies in their web browsing content. In addition, Brown et.al [21] also showed that there are obvious emotional tendencies in insiders' language usage of their web browsing and emails. The authors [11] posited the potential of applying sentiment analysis to predict insiders' motivation, but no evaluation was presented.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

In this paper, we propose a malicious insider prediction scheme based on user sentiment profile (built from user's network browsing and email content). Given a large collection of network browsing and email activity log data, the scheme computes the users' daily and weekly threat profile using sentiment analysis. Deviations from the history profiles and peer profiles would be identified using anomaly detection method, and relevant stakeholders will then be alerted. It was found that the scheme is effective in detecting users who display (extreme) emotion. The contributions and innovations of this paper are summarized as follows:

- The proposed scheme can predict malicious insiders' attack motivation and proactively detect malicious insiders, unlike other competing detection methods that are generally post-incident (i.e. after a user has 'turned bad / malicious' and came to the attention of the authorities).

- The proposed scheme is the first work to build a sentiment analysis based profile, based on user network behaviors. The scheme can also be extended in the future to integrate subjective analysis and objective analysis methods, which may improve insider threat detection accuracy.

The rest of this paper is organized as follows. Section II presents the proposed approach. Then in Section III, we describe the process of constructing effective dataset and the experiments to evaluate the proposed scheme. Related work will be discussed in Section IV. Finally, we discuss our scheme and conclude this paper with some future research agenda.

II. Sentiment Profile Scheme

The architecture of the detection system is detailed in Fig.1. We aim to build insiders' sentiment profile based on their browsing websites and email content. The sentiment profile can be used to display the quantitative changes in the content of insiders' network browsing and emails. Malicious insiders' motivations, such as dissatisfaction and revenge, could be predicted by the deviation of their sentiment profile.

To identify changes of users' psychology and emotion quantitatively and proactively detect malicious insiders, we first build daily and weekly sentiment profiles for each user based on their web browsing content and email content using sentiment analysis and malicious URL detection. Then, we define and compute a threat value for each user. Next, we compute and compare the changes of these sentiment profiles using anomaly detection algorithm. Users whose deviation exceeds the threshold would be flagged as potentially malicious insiders for further investigation.

In the remaining of this section, we will explain the key components of the system. We begin with the indicators that we use for predicting insiders' attack motivation. Then, we describe the data pro-processing module that extracts text content from raw webpages and emails. After pro-processing, the sentiment analysis module is used for classifying the text content into whether the content users browsed or emailed is negative, and the malicious URL detection module is used to compute the probability of the website being malicious or not. Then, we describe the process to build a sentiment profile indicative of the user's attack motivation. Finally, we introduce the anomaly detection method we use to compute the changes of user's psychology profiles.

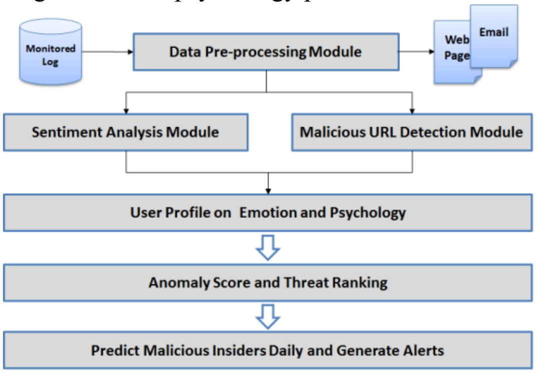

Fig. 1. Architecture of the sentiment profile for insider threat prediction

A. Indicators

Our approach is partly inspired by the findings of FBI [13], in the sense that disgust and revenge are two of the most frequently motivations for insider threats, and these negative motivations can be 'observed' from their browsing habits [10] and the language they used [9] in their network activities. Now, we describe the following indicators that are used to quantify the malicious insiders' motivation and the elements of our sentiment profile.

- Web browsing content—malicious insiders having feelings of disgust or revenge may frequently browse websites, whose content have extreme emotional tendencies (e.g. drugs, violence, and extremism). Frequently browsing these websites, particularly if such behavior deviates from their norm, may be an indicator before they carry out actual malicious activities such as sabotage.

- Malicious URL—the possibility of some insider threats, such as information leakage and sabotage, will be significantly increased when users frequently browse malicious websites (e.g. pornographic websites or underground trading forums). These websites generally contain negative or extreme information to attract users to click and may contain malicious payload (e.g. malware).

- Email content—certain language usage in emails is a good indication of user emotion and psychology. The emails sent by malicious insiders, for example, may contain negative words or sentences (e.g. derogatory or racist) to express their dissatisfaction and negative emotion.

B. Data Pro-Processing

The input data for the proposed detection system include monitored user browsing history and emails. For network browsing logs, the pro-processing module translates each log into a record that consists of user's identification, time, URL link, page text content, whether file is uploaded, etc. For email logs, the pro-processing module translates each email into a record that consists of user's identification, email content, sent time, file attachments, etc.

Table 1. The Sentiment Dictionary for Sentiment Analysis Module

bullshit	drug	tortures	Ranting	Petrified	Bleak	Terrorist objectives	beating	forgery	treason
catastrophic	drugs	torturing	Raving	Phobic	Death-seeking	Terrorist group	mayhem	genocide	trespassing
damn	cannabis	whore	Seething	Shocked	Devastated	Terrorist goals	mugging	hijacking	vandalism
damned	drugs4you	ass	Spiteful	Terrorized	Doomed	Terrorist	onslaught	hit and run	voyeurism
dick	abraxas	asshole	Vengeful	Avaricious	Gutted	Terrorism	thumping	homicide	voyeur
dickhead	simurgh	catastrophic	Vicious	Gluttonous	Nihilistic	Terror tactics	violence	hooliganism	vandal
fraud	super haze	sperturbed	Vindictive	Grasping	Numbed	Radiological operation	antagonisms	kidnapping	trespasser
fraudster	haze	nuisance	Belittled	Greedy	Reckless	Nuclear weapon	abduction	libel	traitor
fraudsters	stimulant	detest	Degraded	Green with Envy	Self-destructive	Nation-state	arson	looting	smuggler
fraudulent	dirtyharry	inquietude	Demeaned	Persistently Jealous	Suicidal	Narco-terrorism	assassination	lynching	robber
fuck	hackboy	grieved	Disgraced	Possessive Resentful	Tormented	Insurgency	assault	manslaughter	rioter
fucked	Clandestine Drug	melancholy	Guilt-ridden	Anguished	Tortured	Guerilla warfare	bigamy	mugging	rapist
fuckers	bastard	Appalled	Guilt-stricken	Bereaved	Thug	Designated foreign	blackmail	murder	murderer
fucking	bastards	Belligerent	Humiliated	Bleak	Ghetto	terrorist organization	bombing	perjury	mugger
hell	bitch	Bitter	Mortified	Depressed	Ratchet	Counter-terrorism	bribery	pickpocketing	looter
jackass	bitches	Contemptuous	Ostracized	Despairing	Inner City	Chemicla agent	burglary	pilfering	kidnapper
piss	cock	Disgusted	Self-condemning	Despondent	Articulate	Chemical weapon	child abuse	rape	hooligan
pissed	cocksucker	Furious	Self-flagellating	Grief-stricken	Exotic	Biological weapon	corruption	riot	hijacker
rape	cunt	Hateful	Shamefaced	Heartbroken	Ethnic	Assest	crime	robbery	embezzler
shrew	motherfucker	Hostile	Stigmatized	Hopeless	Sketchy	acrimony	cybercrime	shoplifting	child abuser
shit	motherfucking	Irate	Filled with Dread	Inconsolable	Illegal alien	animosity	domestic violence	slander	burglar
shithead	nigger	Livid	Horrified	Morose	racism	annoyance	drunk driving	smuggling	bomber
torture	prick	Menacing	Panicked	Agonized	Racist	antagonism	embezzlement	terrorism	bigamist
tortured	slut	Outraged	Paralyzed	Anguished	Wmd	assault	espionage	trafficking	assailant
assassin	vengeance					attack			
arsonist	revenge								

C. Sentiment Analysis

We then use sentiment analysis method to analyze the text content of the user's network browsing and email records extracted from the data pro-processing module. The sentiment analysis for predicting insiders' motivation differs from traditional sentiment analysis task, since existing sentiment classification models such as LSTM based model are not appropriate. To improve the accuracy of sentiment classification, we implement this module using a dictionary based method. The implementation is explained as follows:

- Collect Sentiment Corpus - we collect from sources such as public sentiment dictionary, public sentiment analysis dataset, and webpages crawled by some extremism websites such as Dark Net Market [17].

- Construct Sentiment Dictionary - for webpages, we build the sentiment dictionary containing 243 words extracted from the high scored negative words, violence related vocabulary, drug related vocabulary and extreme religious related vocabulary, as displayed in Table 1. We use a public dictionary with labels and intensity of words' polarity for emails.

- Build Classification Model - we design a threshold based classification method for sentiment analysis on webpages and emails.

The sentiment dictionary based sentiment analysis module can be used to achieve high accuracy in determining whether the insiders' browsing webpages and emails have negative sentiment tendencies.

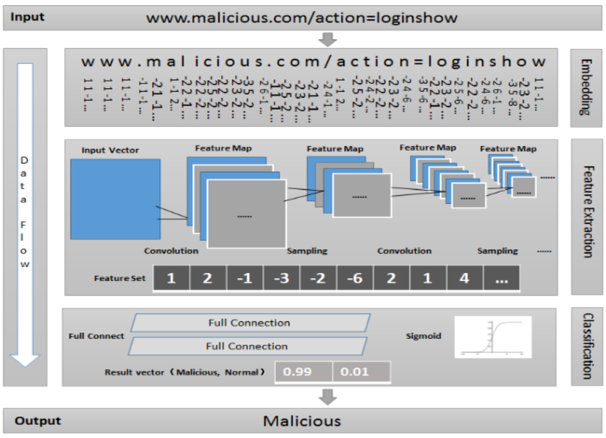

Fig. 2. Framework of Malicious URL Detection Module

D. Malicious URL Detection

Deep learning method methods can automatically extract hidden features within the URL strings and resist obfuscation attacks. In this paper, we design and implement the malicious URL detection module with the convolutional neural network (CNN) based classification model. The framework is displayed in Figure 2. Specifically, we use the character-level word embedding method to map the URL strings into vector as the input for CNN network. We also use our previous work [18] as the malicious URL detection model. We combine the deep learning framework with threat intelligence, such as blacklist, to improve the accuracy.

E. User Profiling based on Sentiment Analysis and URL Detection

We will now describe the components and the calculation method of the sentiment profile. For a specific user, we compute a psychology and emotion threat value for an active day. The threat value T (Daily) is computed using Formula (1). In this formula, p_l represents the sum of negative webpages browsed by the users during a specific day, p_u represents the sum of malicious websites clicked and browsed by the users during a specific day, and p_e represents the sum of negative writing emails sent by the user during a specific day. We let w_l, w_u, and w_e denote the weights on the threat value of insiders' psychology and emotion negative tendencies. The weekly threat value T (Week) is computed using Formula (2). In this formula, T (Week) is summed by each daily threat value during the specific week.

$$T(Daily) = w_l * p_l + w_u * p_u + w_e * p_e \qquad (1)$$

$$T(Week) = \sum_{\text{Day } \varepsilon \text{ last week}} T(Daily) \qquad (2)$$

F. Anomaly Detection and Threat Ranking

To express the changes of user's profiles during a time period for comparison, we implement the anomaly detection module using classical methods such as mean and variance of the threat values. We compute the mean threat value, M (User), during a week for a specific user and the variance, V (User), during the week. The computation for the insider's general threat values is as follows:

$$Mean(User) = T(Week)/7 \qquad (3)$$

$$Variance(User) = \sqrt{\sum_{Day\ \varepsilon\ Last\ Week}(T(Daily) - Mean(User))^2} \quad (4)$$

To predict the insiders' attack motivation based on their daily and weekly sentiment profiles, the proposed detection system defines an anomaly score for each user and updates the value daily. The anomaly score is determined by the parameters computed below.

- T (Daily)—the parameter can predict malicious insiders in early days of negative emotions and attack motivation with a high threat value on a specific day.
- Mean (User)—the parameter helps to predict the malicious insiders who have very negative emotion during the last week before the calculating time point. This provides a historical overview on the insiders.
- Variance (User)—the parameter helps to predict the malicious insiders who have significant changes in their emotion and psychology during the last week.

$$Anomaly\ Score(User) = T(Daily) + Mean(User) + Variance(User) \quad (5)$$

The system computes and updates the anomaly score for each user every day and ranks the malicious insiders based on the value. The detection system could also display the three parameters of the predicted users to facilitate the analysts in learning about the user's real-time and historical psychology and emotion tendencies. In addition, the detection system sets two threat threshold values to generate an alert for the analysts.

G. Evaluation

After the anomaly detection module has executed, the detection system will propose the top-10 insiders and users whose threat values exceed the threshold. We will evaluate whether the detection system can predict the malicious insiders and whether the users predicted by the system are truly users of interest.

III. Experiment

To evaluate the performance of the proposed detection system, we use the CMU-CERT dataset v4.2 [16] to construct the experimentation scenarios. The dataset provides the monitored email, http, file, device, and logon logs of users in a simulated organization from 2009 to 2011. There are also five threat scenarios and labeled insiders in this dataset, which were inserted using the Red Team model. The scenarios included in the dataset mainly focus on behavior anomaly, and only a few focus on insiders' attack motivations. In the dataset, there is a threat scenario related to insiders' motivation as follows.

The scenario described a system administrator who became disgruntled. Then he downloaded a key logger and transfer it to his supervisor's machine, then he send out some emails causing panic in the organization using the collected key logs to logon his supervisor's account.

In this threat scenario, the insider becomes disgruntled before he performed malicious threat activities. However, the scenarios did not consider abnormal browsing behavior and language usage on http and email. Thus, we expand the threat scenario to include users visiting malicious websites and sending emails using words to express their dissatisfaction, anger, etc. In this paper, we use the CMU-CERT dataset v4.2 [16] as the base dataset. We also extend the dataset using the collected negative samples.

A. Dataset

The http and email data of the CMU-CERT dataset are used as the normal monitoring data. Separately from the normal monitoring data, we use the Red Team model with the threat scenarios described above and insert the negative samples into the dataset.

For collecting websites with negative information, we collect these websites from the Dark Net Market (DNM) archives [18], mirror and scrape all existing English dark net market websites from 2013 to 2015. We code a small script using Python to extract the text from these webpages. For the malicious URL samples, we download these data from malicious URL sharing websites, such as Virus Total. For the negative samples of emails, we collect these data from the negative samples of some public Twitter emotion corpus.

The second threat scenario of the dataset describes a financially-motivated insider who frequently browsed job websites and contacted competitors via emails before he implemented threat activities, such as stealing organizational data. To extent the threat scenario, we insert negative web content, email, and URL samples to the labeled users' logs as the red team did to the second threat scenario. We develop the scenarios and insert the negative webpages, malicious URL links, and negative emails to their daily activities (see Table 2).

Table 2: Threat Scenarios Development and Negative Samples Inserted

User ID	Start Date	Threat Date	Samples Inserted
BBS0039	2010-7-8	2010-8-13	132
BSS0369	2010-8-27	2010-10-1	152
CCA0046	2010-9-18	2010-10-15	143
CSC0217	2010-5-14	2010-6-11	106
GTD0219	2010-5-1	2010-6-18	133
JGT0221	2010-6-12	2010-7-16	130
JLM0364	2011-3-22	2011-4-29	133
JTM0223	2010-6-20	2010-7-22	123
MPM0220	2010-10-3	2010-11-5	144
MSO0222	2010-11-12	2010-12-10	158

B. Results

In this section, we present the results of our experiments to evaluate the performance of proposed system. The sentiment analysis and malicious URL detection module are implemented as described in Section 2, and we achieve a sentiment classification accuracy of 100% and 96% for http and email content, respectively.

(1) Expanded CMU-CERT Dataset

According to the frequency of websites and emails, we set the weight of the motivation indicators w_l, w_u, w_e (described in section) as 0.3, 0.5, 1. Then, we build insiders' daily sentiment profile based on Formula (1). Figure 3 displays the malicious insiders' daily threat value and anomaly threat value during the last 30 days prior to the threat action date. Then, we compute the mean threat value and variance threat value according to the insiders' daily sentiment profile. The general anomaly score was computed using Formula (5) for each insider (see also Figure 3).

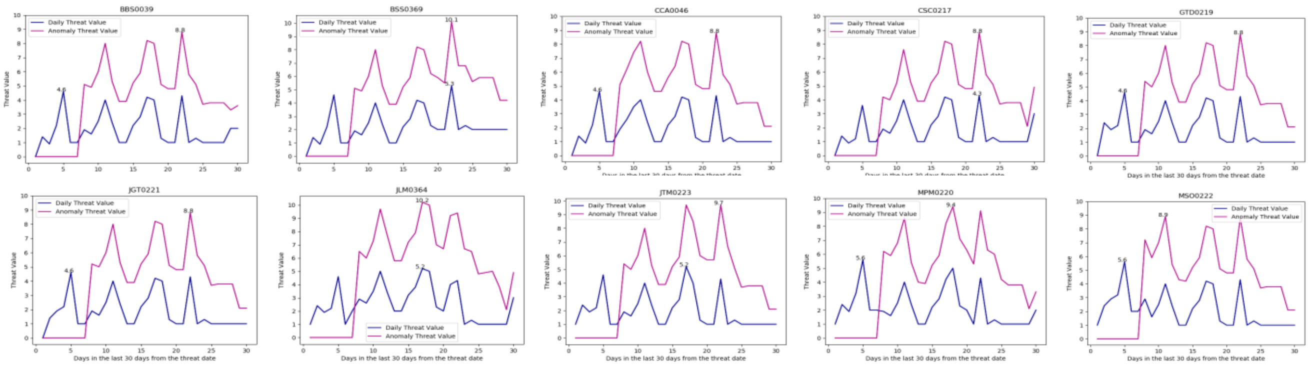

Fig. 3. Daily Threat Value and general anomaly value of the Malicious Insiders

The proposed system ranks the insiders with their daily threat value and final anomaly score. In addition, the proposed system would generate alerts when the insider's daily threat value and general anomaly score exceeds the threshold. Table 3 displays the predicted date and the threat date of malicious insiders. In our experiment, if we set the daily threat value threshold and general anomaly threat value as 4.5 and 9, then the recall of the system is 90% in the scenario and the insiders' with these motivations could be alerted about 25 days before they implement threat actions on average. In other words, the proposed system can proactively and accurately predict the malicious insiders based on their attack motivation indicators.

(2) Enron Email Dataset

We also evaluate the proposed scheme using Enron Email dataset. The dataset provided the users' emails of Enron company before it went bankrupt. We aim to find some hidden insiders who had negative emotion on the company and had the motivation to carry out some threat actions. Table 3 displays the identified insiders and the corresponding date, peak value date. From these alerts, we could predict the hidden insiders. For example, several insiders' sentiment became negative before Enron collapsed, which may be the motivation or indicators of the hidden insiders' threat actions such as information stealing and fraud. The peak value of alerted users is centralized around May 2001 and November 2001, which were just before the date that Enron company was fined by U.S. Securities and Exchange Commission and the date when the company declared bankruptcy, respectively.

The proposed scheme could predict hidden insiders who have negative emotion and psychological tendencies. Such tendencies could be the motivation and indicators prior to some threat activities. The investigators could further examine the alerted insiders' http, email, file and other activities to avoid misjudgment.

Table 3 Predict Hidden Insiders on Enron Dataset

User ID	Alerted Date	Peak Value Date
Blair-l	2001-6-27	2001-6-27
Causholli-m	2001-10-18	2001-10-18
Dean-c	2010.10.5-2001.10.29	2001.10.15
Guzman-m	2001-4-16,2001-4-30	2001.4.30
Linder-e	2001.4.2, 4.16, 4.22, 4.30	2001.4.30
Merriss-s	2001.4.16-5.1, 2002.1.16-23	2001.4.30

From the experiments, the proposed scheme could effectively improve recall and accuracy of insider threat detection framework. In addition, the scheme can also find potential evidence or clues missed by anomaly based schemes.

C. Discussion and Comparison

There are many works on detecting insider threat based on user profile compiled from their Internet activities or their psychology traits. For comparison, we use two recent representative approaches.

- PRODIGAL based Detection Framework [19], funded by ADAMS (Anomaly Detection at Multiple Scales) initiative. PRODIGAL combines graph analysis and multiple anomaly detection algorithms, and builds a comprehensive profile of user activities.

- LIWC-OCEAN based Detection Framework [12] explores how Internet browsing activity could be used to predict users' psychological characteristics. The approach extracts the content and maps the website LIWC keywords to OCEAN personality traits.

We compare our proposed scheme with the above two detection frameworks in terms of the following four aspects:

- Time required to predict the malicious insiders.
- Threat scenarios detection accuracy.
- Evaluated using real-world dataset (or not).
- Capable of predicting users' attack motivation (or not).

The comparative summary is displayed in Table 4. PRODIGAL focuses on characterizing users' behavioral profiles from the extracted quantitative and frequency features, whilst our proposed scheme and LIWC-OCEAN extract features from the content of these websites and emails. Based on these features, our proposed scheme and LIWC-OCEAN build user psychology profile to predict malicious insiders before a threat is executed. However, PRODIGAL can only detect malicious users when or after they have carried out the activities. Anomaly activities are not equivalent to threats and some insider threat may have not an obvious abnormality in behavior only, so the PRODIGAL scheme may lead to high false positive. Only PRODIGAL and our proposed scheme are evaluated using real-world dataset. Finally, although OCEAN model can characterize users' personality traits, it cannot accurately depict user attack motivations; thus, resulting in low recall. We also remark that the sentiment profile of users can be used to predict the users' attack motivation, such as negative emotion and extremely psychology tendencies. This could result in a higher recall.

Table 4 Comparison with PRODIGAL and LIWC-OCEAN

Method	Data source	Detection Time	Recall	Real dataset verified	Attack Motivation Prediction
PRODIGAL	Email, Internet Browsing	When or after a threat is implemented	High	Yes	No
LIWC-OCEAN	Internet Browsing	Before a threat is implemented	Low	No	No
Sentiment Profile	Internet Browsing, Email	Before a threat is implemented	High	Yes	Yes

IV. Related Work

Language features generated by users online and their communication patterns could be used to predict user behavioral intent. Ho et al. [10] assumed that language-action may change significantly when users attempt to deceive in group interactions. They tested their hypothesis in an online game environment. Many researches have also proposed profile models by analyzing the websites users browsed or emails sent by the users. For example, Christopher et al, [21] proposed a profile model to predict employees' loyalty and neurotic behavior. The system could build users' personal trait model by analyzing their email content and mapping these texts to neuroticism and agreeableness with LIWC.

The research closest to our research is [12]. Their contribution is the exploration of the relationship between malicious insiders' Internet browsing activity and psychological characteristics. However, their approach does not allow one to determine the attack motivation. This is the gap we addressed in this paper, i.e. our proposed sentiment analysis based profile of insiders can be used to quantify users' negative emotions and extreme psychological tendencies. Our work also differs from prior work by incorporating psychological and behavioral analysis based on network data using sentiment analysis methods.

V. Conclusion and Future work

In this paper, we proposed an effective approach for proactively insider threat detection, based on sentiment analysis. From the monitored webpages and emails, the proposed system builds a sentiment profile for each insider. The sentiment profile includes daily threat value and weekly threat value of insiders computed by the sentiment analysis module and the malicious URL detection module. Finally, the proposed system computes a daily anomaly score for each user for predicting malicious insiders and ranking threats. We evaluated the performance of the approach using CMU-CERT dataset and Enron Email dataset. The result demonstrated that the proposed detection system can proactively and accurately detect the malicious insiders before they execute risky / malicious activities. Our proposed approach also addresses the limitation of a lack of dataset required for psychological analysis. The approach does not include behavioral factors of insiders (e.g. file access, and device usage), and such an integrated mechanism can be explored as part of future research. In addition, we can also consider deep mining the patterns within websites and emails in order to build a more comprehensive attack profile.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No.61601459, 61402476).

References

[1] WikiLeaks. WikiLeaks publishes 'biggest ever leak of secret CIA documents'[EB/OL].

[2] Verizon. Verizon's 2017 Data Breach Investigation Report [EB/OL].

[3] Legg, Philip A., et al. "Automated insider threat detection system using user and role-based profile assessment." IEEE Systems Journal 11.2 (2017): 503-512

[4] Young, William T., et al. "Detecting Unknown Insider Threat Scenarios." Security and Privacy Workshops IEEE, 2014:277-288

[5] Brdiczka, Oliver, et al. "Proactive insider threat detection through graph learning and psychological context." Security and Privacy Workshops (SPW), 2012 IEEE Symposium on. IEEE, 2012

[6] Feng, Xiaotao, et al. "Stealthy attacks with insider information: A game theoretic model with asymmetric feedback." Military Communications Conference, MILCOM 2016-2016 IEEE. IEEE, 2016

[7] Hashem, Yassir, et al. "Towards Insider Threat Detection Using Psychophysiological Signals." ACM CCS International Workshop on Managing Insider Security Threats ACM, 2015:71-74

[8] Kandias, Miltiadis, et al. "Stress level detection via OSN usage pattern and chronicity analysis: An OSINT threat intelligence module." Computers & Security 69 (2017): 3-17

[9] Gamachchi A, Boztaş S. Web access patterns reveal insiders behavior[C]// International Workshop on Signal Design & ITS Applications in Communications. IEEE, 2016:70-74.

[10] Ho S M, Fu H, Timmarajus S S, et al. Insider Threat: Language-action Cues in Group Dynamics[J]. 2016:2729-2738

[11] Gavai G, Sricharan K, Gunning D, et al. Supervised and Unsupervised methods to detect Insider Threat from Enterprise Social and Online Activity Data[J]. JoWUA, 2015, 6(4): 47-63.

[12] Alahmadi B A, Legg P A, Nurse J R C. Using Internet Activity Profiling for Insider-threat Detection[C]//ICEIS (2). 2015: 709-720.

[13] Keeney M, Kowalski E, Moore A, et al. Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors [J]. 2005.

[14] Claycomb W R, Huth C L, Flynn L, et al. Chronological Examination of Insider Threat Sabotage: Preliminary Observations [J]. 2012.

[15] Agency, D., & Information, P. (2018). Anomaly Detection at Multiple Scales (ADAMS). Darpa.mil. Retrieved 13 April 2018, from https://www.darpa.mil/program/anomaly-detection-at-multiple-scales.

[16] Glasser, Joshua, and Brian Lindauer. "Bridging the gap: A pragmatic approach to generating insider threat data." Security and Privacy Workshops (SPW), 2013 IEEE. IEEE, 2013.

[17] Gwern Branwen, Nicolas Christin, David Décary-Hétu, Rasmus Munksgaard Andersen, StExo, El Presidente, Anonymous, Daryl Lau, Sohhlz, Delyan Kratunov, Vince Cakic, Van Buskirk, Whom, Michael McKenna, Sigi Goode. Dark Net Market archives, 2011-2015, 12 July 2015. Web

[18] Jiang J, Chen J, Choo K K R, et al. A Deep Learning Based Online Malicious URL and DNS Detection Scheme[C]// International Conference on Security and Privacy in Communication Systems. Springer, Cham, 2017:438-448.

[19] Goldberg H, Young W, Reardon M, et al. Insider Threat Detection in PRODIGAL[C]// Hawaii International Conference on System Sciences. 2017

[20] Li M, Liu Y, Yu M, et al. FEPDF: A Robust Feature Extractor for Malicious PDF Detection[C]// Trustcom/bigdatase/icess. IEEE, 2017:218-224

[21] Brown, Christopher R., A. Watkins, and F. L. Greitzer. "Predicting Insider Threat Risks through Linguistic Analysis of Electronic Communication." Hawaii International Conference on System Sciences IEEE Computer Society, 2013:1849-1858.