A Hybrid Model based on Multi-Dimensional Features for Insider Threat Detection

Bin Lv [1,2], Dan Wang[1,2], Yan Wang[1*], Dan Lu[1,2], Qiujian Lv[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
Email:{lvbin, wangdan3, wangyan, ludan, lvqiujian}@iie.ac.cn
[*]Corresponding author

***Abstract*—Insider threats have shown their power by hugely affecting national security, financial stability, and the privacy of many thousands of people. A number of techniques have been proposed to detect insider threats either by comparing behaviors among different individuals or by comparing the behaviors across different time periods of the same individual. However, due to the fact that the behaviors of insider threats are always complex and diverse, both of them always fail to identify certain kinds of inside threats. This paper focuses on proposing a hybrid model to detect insider threats based on multi-dimensional features. First, based on the isolation Forest algorithm, an Across-Domain Anomaly Detection (ADAD) model is proposed to identify anomalous behaviors that deviates from the behaviors of their peers by using multi-source information. Second, we propose an Across-Time Anomaly Detection (ATAD) model to measure the degree of unusual changes of a user's behavior by implementing an improvement model based on Markov. Finally, our proposed hybrid model presents a fusion method to integrate the evidence from the above two models. With the data lasting 17 months, we evaluate our proposed models comprehensively. The results demonstrate the robust performance of the ADAD and ATAD models and the hybrid model is demonstrated to outperform the two separate models obviously.**

***Index Terms*—insider threat detection, information fusion, hybrid model, Isolation Forest, Markov Model**

I. Introduction

Insider threats are threats with malicious intent directed towards organizations by people internal to the organization [1]. These include physical sabotage activities, theft of confidential data and business secrets, and fraud. Financial loss and reputation damage caused by this "known unknow" cybersecurity threat far outweighs that caused by external attacks. One of the most recent articles from CSO magazine [2] compared the cost between external and internal attacks and noted that while it takes about 50 days to fix a data breach caused by an internal attack, it only takes 2 to 5 days in the case of external attacks. Nowadays, researchers have proposed different models to prevent or detect the presence of attacks.

Existing literature focuses on two types of insider threat detection models: data-driven detection models [3], [4] and behavior driven detection models [5], [6]:

1) The first model aims to find a normal portrait in all users data in order to detect insider threat that deviates from this normal portrait by comparing with their peers [7]. Figure 1 shows an example of insider threats which can be detected by this kind of model [7]. After getting off work, the general behavior is going home to have a rest. By contrast, few of users secretly deal with the company's confidential documents in the workplace when others are sunk in sleep at midnight, and are more likely to be an insider threat. However, merely comparing the behaviors among peers, this kind of models fails to detect a malicious insider who tries to behave like a normal user to cover up his evil.

2) The second model regards the abnormal changes in the behavior as the basis for insider threat detection by comparing behaviors of themselves in different time periods. Figure 1 also shows an example to depict a malicious insiders detection based on the behavior-driven method [8]. Over a long period in the past, the regular behavior of the user was to click on the browser after booting, view the email, and then reply. But one day, he acted abnormally: he connected the mobile device after booting, and had a series of operations on the file-copy to steal the company's confidential documents. These abnormal behaviors don't match his usual styles of doing things. It is worth to note that although the malicious insider may try to behave like a normal user, the threat can also be detected based on the deviation from his regular behavior routines. However, the model is unable to recognize situations where a user systematically attacks an organization over a long-term period [3].

Thus, Insider threat surveys [9] suggest the problem of insider threat detection cannot be defined only as a data or behavior driven problem. Therefore, it is necessary to model the problem as data and behavior driven problems, based on which an effective model needs to be proposed.

Hence, based on multidimensional features, this paper proposes a hybrid model that combines a data-driven model with a behavior driven model to detect insider threat in a more robust and accurate manner. First, the multi-dimensional features extracted from data collected from an enterprise network is formatted and fed separately into the two separate models. Second, each model generates an abnormal score to represent the degree of users' unusual behaviors. Finally, the abnormal scores of two models are fused to be the final abnormal score of each user, and a user is identified to be an insider threat if the anomaly score exceeds the threshold. After a wide range experiments, it is verified that the hybrid model can detect

Identify applicable funding agency here. If none, delete this.

insider threats in a more robust and accurate manner. Overall, the contributions of this paper can be summarized as follows:

1) We apply the isolation Forest to detect behavioral inconsistencies among the behaviors of users. In this model, we extract temporal features from multi-source data and combine them comprehensively when constructing the user's portrait.
2) We propose an improved model based on Markov to identify users' unusual changes. The proposed model considers all historical behaviors of users as the contextual information. As compared with the existing detection methods of Markov model [7], our proposed approach show higher efficiency.
3) We propose a hybrid model based on multi-dimensional features for insider threat detection. The model is able to determine malicious insiders that not only act inconsistent with their peers but also have unusual changes compared with their historic behaviors. Through extensive experiments, we obtain the accuracy of 95% through the proposed model, which is of great significance in industry and scientific research.

The remainder of this article is structured as follows. Section 2 presents the related work. Then, section 3 details our approach. Next, in section 4, we detail our implementation of the models and analyze the results of experiments. Finally, we conclude in section 5, also presenting limitation of our work.

II. Related work

The topic of insider threat has recently received much attention in the literature. Researchers have proposed different models aimed at preventing or detecting the presence of attacks [11], [12]. To elicit the state of art, the work presented here is focused on the approaches of detecting insider threat based on data-driven methods and behavior-driven methods, respectively.

With regard to the data-driven methods, Mathew et al. [13] detected inside threat on account of user access patterns, Eberle et al. [14] used social graphs to detect the abnormal. More recently, Eldardiry et al. [15] have also proposed a system in managing insider attacks compared users behavior based on peer baselines. Michael Goldsmith applied a layered architecture by fusing across multiple levels information to detect anomalies from heterogeneous data [16]. Hoda et al. [13] detect peer groups of users and modeling user behavior with respect to these peer groups, and subsequently detect insider activity by identifying users who deviate from their peers with respect to the user behavior models. There have also been various approaches based on data-driven to detect abnormal [10], [8]. However, they did not factor in the changes of user behaviors over time. We note that while a common activity not be suspicious, a rare change of the order common activity can be.

There are several literatures based on behavior-driven models [7], [11], [12]. Tabish Rashid [17] takes the change of the user's behavior over time to detect the anomaly and achieved some results. M. Bishop [11] examine the application of process modeling and subsequent analyses to the insider problem. However, these behavior-driven models just work out based on unusual change of user behaviors, but will miss recognizing situations where a user systematically attacks an organization over an extended time-framework.

III. Overview of proposed approach

To detect insider threat that deviates from this normal portrait by comparing with their peers and who tries to behave like a normal user to cover up his evil, we propose a hybrid model based on multi-dimensional features for insider threat detection which is illustrated in Figure 2. The hybrid model includes two components, one is named "Across-Domain Anomaly Detection (ADAD)", and the other is "Across-Time Anomaly Detection (ATAD)". We get anomaly scores from the two components, then fuse them as the basis for insider threat detection. The multi-source information contains activities such as logging on/off, sending and receiving emails, accessing external devices or files, and accessing web sites. We refer to different categories of data as domains, e.g., logon domain and email domain. The dataset used in this paper is described in the following section.

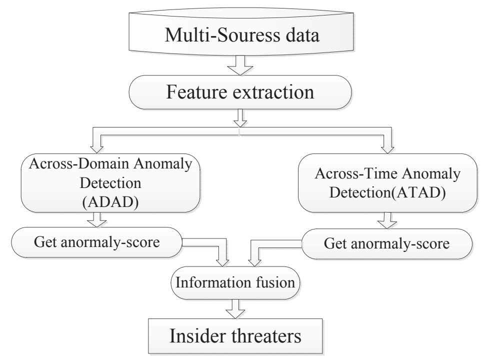


Fig. 2: Anomaly Detection Framework.

A. The Data Set

Due to the lack of availability of proper insider threat datasets, we have utilized the insider threat dataset published by CERT Carnegie Mellon University for this research [17]. The dataset R4.2.tar.bz lasting 17 months has been used for this analysis. This dataset consists of six broad types of data records (HTTP, logon, device, file, email and psychometric) of 1000 employees over a 17 months period. All HTTP records contain user, PC, URL and web page content with time stamps. Logon.csv consists of user logon/logoff activities with the corresponding PC with timestamps. The data file device.csv indicates insert/remove actions with the relevant user, PC, and timestamp. Details of file copies are stored in file.csv file with date, user, PC, filename, and content. We should note that the CERT Dataset contains the ground truth for each user (when they are acting maliciously or not), which allows us to monitor the success or failure of our experiment.

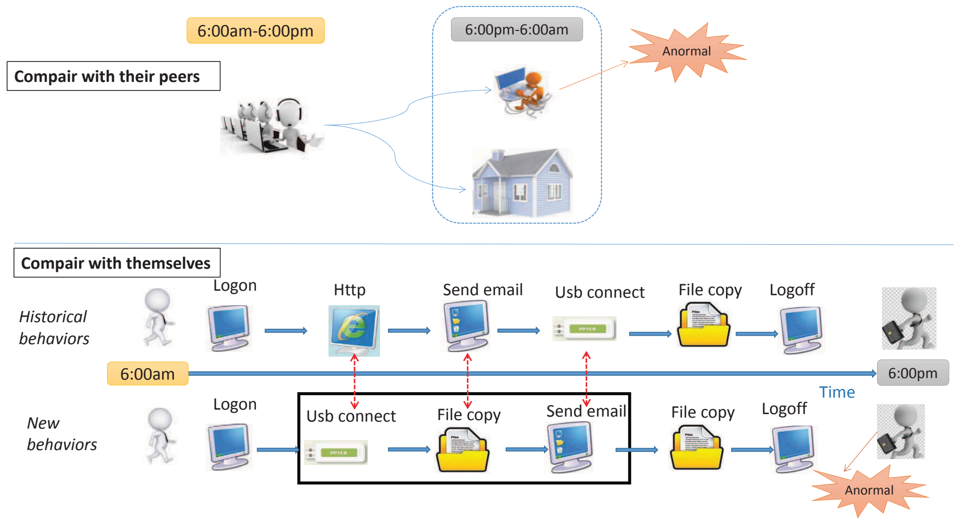


Fig. 1: Example of insider threat.

TABLE I: Selected feature set.

Module	Features (per 6 hours)	The number of features
Logon events	Max/Mode Logon counts	4
Logoff events	Max/Mode Logoff counts	4
Removable Media	Max/Mode Connect counts	4
	Max/Mode Disconnect counts	4
File copy events	Max/Mode Filecopy counts	4

IV. KEY METHODOLOGIES

In this section, we describe our Across-Domain anomaly detection (ADAD) model in detail at first. We then detail the Across-Time model (ATAD) to detect insider threats. Finally, we introduce the fusion method by combining ADAD and ATAD.

A. Approach 1:Across-Domain Anomaly Detection (ADAD)

This framework will utilize multi-domain information inputs, such as logon records and operation on file, to identify abnormal users who behave differently from their peers.

1) Feature Extraction: **Logon/Logoff Behaviors.** As most disgruntled insiders tend to commit malicious logon or logoff activities after hours, These behaviors are applied to identify malicious insiders.

Removable media usage. Removable media is among the most popular method used in theft of Intellectual Property (IP) in extracting confidential information from organizations [18]. Tracking the use of removable media can be an excellent information source for identifying suspicious events by trusted insiders.

File copy Behaviors. File copying is an easy method to steal confidential information from organizations. So the number of occurrences of the behavior can give some useful information to detect insider threats.

To model users' multi-domain behaviors, we extract several temporal features. First, we count the times of each behavior for every hour, and then calculate the average of the maximum counts and mode counts of each behavior. The maximum counts indicate the change range of user behaviors, and the mode counts present the general behavior of most users.

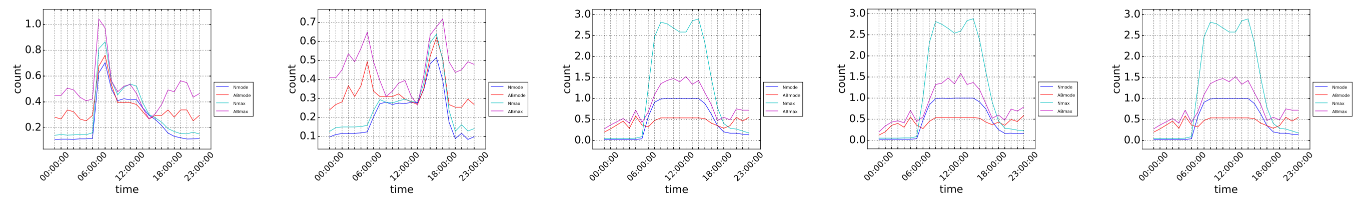
In order to find out whether the number of occurrences of these behaviors is different between normal users and abnormal users, figure 3 investigates their distributions. We found that there is a big difference in behavior between normal users and abnormal users at different times in a day. These abnormal users have more frequent operations in midnight comparing with normal users. So, we divide the 24 hours of a day into 4 time segments: (0:00-06:00),(0:60-12:00),(12:00-18:00),(18:00-23:00). The maximum or mode counts of user behaviors with regard to the five domains are calculated for the 4 segments respectively. Table 1 is an illustration of the feature set.

In addition, in order to discover features interfere with the ability to correctly identify insider threats, we use PCA [19] for denoising to achieves a higher detection accuracy.

2) Anomaly Detection: Due to the complex nature of insider threat problem, it is extremely hard to pinpoint a user as a malicious insider. This section will focus on implementing an anomaly detection algorithm based on the properties identified at the first step of feature extraction. The anomaly detection algorithm adopted in this analysis is the ”Isolation forest“ algorithm [20], which stands out in effectively separating anomalous events from the rest of the instances. In order to select domain features that have big impacts on user's behavior, we apply the principal component analysis to give a score value to each feature.

B. Approach 2: Across-Time Anomaly Detection (ATAD)

In this section, a new method (IM) based on Markov proposed is to detect insider threats. For insider threats detection, we first build a model to represent the regular behaviors. Then,



(a) Logon behavior (b) Logoff behavior (c) USB connect (d) USB disconnect (e) File copy

Fig. 3: "Nmode/Nma" presents the mode/max number of normal users behaviors, and the "ABmode/ABmax" presents the mode/max number of abnormal users behaviors.

we will compare the temporal behavior in the recent past with the regular behaviors for detecting unusual changes of user behaviors.

1) Model building: When building an improved model (IM), we take users behaviors as a temporal sequence and model users behaviors as Markov model. Markov Model (MM) [21] is an extremely powerful tool to model temporal sequence information. It has been widely used in temporal pattern recognition problems (e.g., speech recognition, bioinformatics, gesture recognition) [23]. MM has also been used in the general area of intrusion detection by some notable works ([21]).

Users have different behaviors on computers every day. When a user finishes a behavior and begins the next behavior, a new record is created. The historical behaviors of a user can be represented as a sequence of observations $B = (a_1, a_2..., a_n)$, in which a_i is the behavior that user is served by at time *t*.

Let substring $B(i, j) = a_i a_{i+1} \dots a_j$ for any $1<i<j<n$. Define the context $c = B(n-k+1, n)$. Let *A* be the set of all possible behaviors. We denote the users behaviors as a random variable *X*. $X(i, j)$ represents the sequence of random variates $X_i, X_{i+1}, \dots X_j$ for any $1<i<j<n$. For all $a \in A$ and $i \in \{1, 2, ..., n\}$, the notation $P(X_i = a_i | \dots)$ denotes the probability that X_i takes the value a_i. These probabilities that can be represented by a *transition probability matrix M*. In order to abtain *M*, we can generate an estimate $\hat{P}$ from the current history *B* using the current context *c* of length *k*. The probability for the next symbol to be *a* is

$$P_k(a) = \hat{P}(X_{n+1} = a|B)) = \frac{N(ca, B)}{N(c, B)}, \tag{1}$$

where *N(s',s)* denotes the number of times the substring *s'* occurs in the string *s*.

2) Anomaly Detection: We define the temporal behaviors in the recent past by opening up continuous observation windows of size *N* on the users behaviors. Then we treat the users behaviors in the *i-th* observation window as B_i=$a_{i1}, a_{i2}, \dots, a_{in}$, where *n* denotes the *n-th* behavior in the window. Given the *M* matrix produced in the first step, the anomaly score R_{trend} for the user is calculated as

$$R_{trend} = \frac{\sum_{d=1}^{D} \prod_{t=1}^{N} P_k(a_{dn})}{D}, \tag{2}$$

After obtaining the anomaly score for all users, we can then set a threshold *T*, which we use to classify users as anomalous or not. If the anomaly score of the user is below the threshold, we classify it as anomalous. This threshold *T* is a critical parameter of our model which must be set carefully. However, after running the IM Algorithm, we can save the anomaly scores generated and then experiment with many values of *T*. One could also imagine a human security analyst increasing T from 0 in order to be presented with more instances which the user deems anomalous.

C. Information fusion

in this section, our goal is to combine anomaly scores that have been generated from the two models proposed above to detect anomalies. Therefore, we developed a technique based on weight fusion to combine evidence from ADAD and ATAD. The combined anomaly score $R_{combine}$ is computed as

$$R_{combined} = W_1 * R_{ADAD} + W_2 * R_{ATAD} \tag{3}$$

$$r = \begin{cases} \frac{W_2}{W_1} & \text{if } W_1>0 \\ 1 & \text{if } W_1 = 0, \end{cases} \tag{4}$$

We can then set a threshold T_{com}, which is used to classify users as anomalous or not. If the anomaly score of the user is below the threshold, we regard the user as anomalous.

V. Experiment

This section is dedicated to a comprehensive discussion of the results of the experiment. We will first introduce the performance metrics for evaluating the three detection models, separately. Then, we provide a comparative assessment of our proposed models with other existing insider threat detection methods.

A. Across-Domain Anomaly Detection (ADAD)

In this section, we first introduce the performance metrics for evaluating the detection model. Then, selected features are evaluated by the isolation forest to detect insider threats.

1) Evaluation Metrics and Baselines: Because the isolation forest can identify the anomalies, we consider the metrics *accuracy, Precision* and *recall* to quantitatively evaluate the performance of this model. *Precision* is the fraction of the data entries labeled malicious that are truly malicious; *recall*, also known as sensitivity or true positive rate, is the fraction of malicious entries that are classified correctly; *accuracy* is the fraction of all entries that are classified correctly [22].

TABLE II: The experiment result of ADAD.

Domian	Features	Accuracy	Precision	Recall
Logon	MAX	89%	32%	47%
	MODE	87%	22%	34%
Logoff	MAX	88%	23%	34%
	MODE	85%	14%	21%
Connet	MAX	78%	65%	27%
	MODE	79%	70%	28%
Disconnect	MAX	80%	73%	30%
	MODE	79%	69%	28%
Filecopy	MAX	80%	60%	28%
	MODE	77%	44%	20%
All domains	MAX	82%	68%	34%
	MODE	79%	52%	31%
All domains combined with PCA method		87%	79%	35%

2) Comparative Evaluation: Table II shows the results of running our model ADAD with different parameters. We found that removable media domain is the most detectable (with the highest detective accuracy), while logon and file domains are weak in detecting insider threats. It appears that users show great variations in their behavior of device usages, but are more uniform in the logon and file behaviors. Except that, the mode counts are more effective than maximum to detect insider threats. The result indicates that the mode represents the more significant difference between normal and abnormal. It is obvious that some features interfere with the ability to correctly identify insider threats. In particular, using the *PCA* to decompose the features to a 2-D space. After denoising by *PCA*, the model can achieve a higher detection accuracy.

B. Across-Time Anomaly Detection (ATAD)

In this section, we first introduce the performance metrics for evaluating the detection model. We then provide a comparative assessment of our proposed models with existing MM detection methods.

1) Evaluation Metrics and Baselines: For ATAD, a threshold T is essential to classify users as anomalous or not. So we consider the metrics Receiver Operating Characteristic curves (or *ROC* curves) curve to quantitatively evaluate the models [17].

Comparison Methods. MM has also been used in the general area of intrusion detection by some notable works ([21]). The new model (IM) based on Markov is respectively compared with existing detection Markov:

Markov [21]: when calculating the anomaly score, this model regards the behavior sequences of the week as the whole and treat the evaluation result of the sequences as the final anomaly score for the week.

2) Comparative Evaluation: **Comparison between IM and MM.** If the training set is too long, the user's regular behaviors may have a change, which has an effect on building the normal model. So, Sixty days from the 16th month to the 17th month of dataset R4.2.tar.bz lasting 17 months has been used for this analysis. Specifically, the first 53 days in the 60 days ie treated as the training set to build models and the rest is the testing set. The testing set is consist of records of a week. Because a week can build a complete work cycle profile for a user. IM mode let the observation window N equal a day. We apply our IM model detection result as figure 4. It's obvious that the IM model is more effective than MM. For MM, if a user has once unusual change during this week, the change will have an important influence on the detection result. However, this effect can be avoided by IM. When scoring the user's behavior, IM uses the week's average anomaly score to evaluate a user's behavior, which will reduce the impact of accidental unusual change on detecting insider threats. In other words, IM tends to detect abnormal changes in behavioral trends over a period of time. Hence, it detects insider threats more robustly.

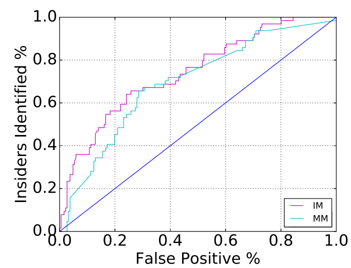


Fig. 4: ROC Curve showing the differences between MM algorithm and IM algorithm.

C. Information fusion

In this section, we first introduce the performance metrics for evaluating the detection model. We then provide a comparative assessment of our fusion method with the above two models.

1) Evaluation Metrics and Baselines: For the fusion method, the most important issue is to determine the proportion of each component. So we let r, the proportion of anomaly scores generated from the two models, increase from 0 to 1. Then we use the metrics *accuracy, Precision* and *recall* to quantitatively evaluate the models.

2) Comparative Evaluation: We compare the fusion method with the ADAD and the ATAD, the comparison results of detecting insider threats are shown in Table 3. To remove amplitude variation and only focus on the underlying distribution shape on data, the scores are normalized before the weight fusion. When $r = 0$, it is the result of individual ADAD method. When $r = 1$, it is the result of individual ATAD method. We can see the ADAD is better than the ATAD, and fusing two model achieves better performance. When r values vary between 1/9 and 3/7, the precision of the hybrid model has greater improvements. The ADAD model is a data-driven model, and it fails to detect a malicious insider who tries to behave like a normal user to cover up his evil. However, the ATAD can make up for this deficiency by comparing behaviors of users in different time periods. So, it is remarkable to build a hybrid model to combine the individual ATAD and ADAD scores. The hybrid model leads to significant improvement in performance relative to any of the individual ATAD or ADAD sources.

TABLE III: The result of the information fusion experiment.

r	0(ADAD)	**1/9**	**2/8**	**3/7**	4/6	5/5	6/4	7/3	8/2	9/1	1(ATAD)
Precision	79.17%	**90%**	**94.44%**	**95%**	74.73%	90%	85%	82.35%	66.67%	61.9%	60%
Recall	35.19	**35.18%**	**31.48%**	**35.19%**	33.33%	33.33%	31.48%	25.92%	25.92%	24.07%	27.78%

To get a better understanding and visualization of results based on our approach, we mark the positive sample with the red and negative sample with blue. Figure 5 is an indication of how the anomaly scores are distributed when *r* is 3/7 in this analysis. The graph indicates few points above the red color horizontal line which is equivalent to an anomaly score of 1.3. We chose the best threshold by *ROC* curves in this scene. Users whose anomaly scores exceed the threshold can be considered as anomalous users. It reached 95% of the accuracy when *r* is 3/7. The model exists some limitations, which causes some insider threats not to be detected. We will present the limitation of our work in the next section.

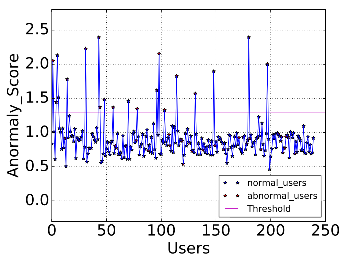


Fig. 5: Anomaly score distribution.

VI. Conclusion

In this paper, we propose a hybrid model that combines a data driven-model with a behavior-driven model to detect insider threat in a more robust and accurate manner. First, the multi-dimensional features extracted from data collected from the enterprise network is formatted and fed separately into the two separate models. Second, each model generates an abnormal score to represent the degree of users' unusual behaviors. Finally, the abnormal scores of two models are fused as the final abnormal scores for each user, and a user will be detected an insider threat if the anomaly score exceeds the threshold. After a wide range experiment, it is verified that the hybrid model can detect insider threats with a high accuracy of 95%, which is of great significance in industry and scientific research.

The hybrid model proposed has a high precision but a pessimistic recall. To solve this problem, we will take the users job role into account to improve the recall of anomaly detection(e.g., [16]).

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No.61372062).

References

[1] Gavai, Gaurang, et al. "Detecting Insider Threat from Enterprise Social and Online Activity Data." ACM CCS International Workshop on Managing Insider Security Threats ACM, 2015:13-20.

[2] By the numbers: Cyber attack costs compared, 2016, accessed on 31/05/2016. [Online]. Available: http://www.csoonline.com/article/3074826/security/bythe-numbers-cyber-attack-costs-compared.html

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

[4] Young, William T, et al. Use of Domain Knowledge to Detect Insider Threats in Computer Activities. 2013.

[5] Rashid, Tabish, I. Agrafiotis, and J. R. C. Nurse. "A New Take on Detecting Insider Threats: Exploring the Use of Hidden Markov Models." International Workshop 2016:47-56.

[6] Eldardiry H, Sricharan K, Liu J, et al. Multi-source fusion for anomaly detection: using across-domain and across-time peer-group consistency checks[J]. Computing & Informatics, 2014, 31(3):575-606.

[7] W. L. Winston, Operations Research: Applications and Algorithms. Belmont, CA: Duxbury Press, 1994.

[8] Sherali Zeadally, et al. "Detecting Insider Threats: Solutions and Trends." Information Security Journal A Global Perspective 21.4(2012):183-192.

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

[10] Gamachchi, Anagi, L. Sun, and S. Boztas. "A Graph Based Framework for Malicious Insider Threat Detection." Hawaii International Conference on System Sciences 2017.

[11] P. A. Legg et al., Towards a conceptual model and reasoning structure for insider threat detection, J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl., vol. 4, no. 4, pp. 2037, Dec. 2013.

[12] M. Bishop et al., Insider threat detection by process analysis, in Proc. IEEE SPW, 2014, pp. 251264.

[13] Sunu Mathew, Michalis Petropoulos, Hung Q Ngo, and Shambhu Upadhyaya. A data-centric approach to insider attack detection in database systems. In Recent Advances in Intrusion Detection, pages 382401. Springer, 2010.

[14] William Eberle, Jeffrey Graves, and Lawrence Holder. Insider threat detection using a graph-based approach. Journal of Applied Security Research, 6(1):3281, 2010.

[15] H. Eldardiry et al., Multi-domain information fusion for insider threat detection, in Proc. IEEE SPW, May 2013, pp. 4551.

[16] Legg P A, Buckley O, Goldsmith M, et al. Automated Insider Threat Detection System Using User and Role-Based Profile Assessment[J]. IEEE Systems Journal, 2017, 11(2):503-512.

[17] Rashid, Tabish, I. Agrafiotis, and J. R. C. Nurse. "A New Take on Detecting Insider Threats: Exploring the Use of Hidden Markov Models." International Workshop 2016:47-56.

[18] D. Cappelli, A. Moore, and R. Trzeciak, The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond

[19] I. Jolliffe, Principal Component Analysis. Hoboken, NJ, USA: Wiley,2005.

[20] F. T. Liu, K. M. Ting, and Z. H. Zhou, Isolation forest, in 2008 Eighth IEEE International Conference on Data Mining, Dec 2008, pp. 413422

[21] Ye, Nong. "A Markov Chain Model of Temporal Behavior for Anomaly Detection." 2000:171–174.

[22] Li, Ling Ko, et al. "Insider threat detection and its future directions." International Journal of Security & Networks 12.3(2017):168.

[23] Lv, Qiujian, et al. "Big Data Driven Hidden Markov Model Based Individual Mobility Prediction at Points of Interest." IEEE Transactions on Vehicular Technology PP.99(2016):1-1.