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

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address

***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 dentify 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 improved Markov model. Finally, our proposed hybrid model present a fusion method to integrate the evidences 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 [2]. 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 [4] 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 [5] [6] and behavior driven detection models [3] [7]:

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 [11]. Figure 1 shows an example of insider threats which can be detected by this kind of model [11]. 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 a 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 inside threat detection by comparing behaviors of themselves in different time periods. Figure 1 also show an example to depict a malicious insiders detection based on the behavior-driven method [9]. 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 recognise situations where a user systematically attacks an organisation over a long-term period. [3].

Thus, Insider threat surveys [10] 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 a effective model needs to be proposed.

Thus, 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 a insider threater if

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

the anomaly score exceeds the threshold. After a wide range experiments, it is verified that the hybrid model can detect 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 and combine temporal features from multi-source data 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 predictor of Markov model [11], 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 abtain 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 about the research. Then, section 3 details our approach. Next, in section 4, we detail our implementation of the models and analyze the experiments reaults. 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 [12] [13]. 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.

With regard to the data-driven methods, Mathew et al. [14] detected inside threat on account of user access patterns, Eberle et al. [15] used social graphs to detect the abnormal. More recently, Eldardiry et al. [16] have also proposed a system in managing insider attacks which compared users behaviour based on peer baselines. Michael Goldsmith applied a layered architecture by fusing across multiple levels information to detect anomalies from heterogeneous data [17]. Hoda et al. [14] 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 [8] [9]. 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. So some methods based on behavior-driven are proposed to solve the problems.

There are several literatures based on behavior-driven models [11] [12] [13]. Tabish Rashid [**?**] takes the change of user behavior over time to detect the anomaly and achieved some results. However, these behavior-driven models just work out based on unusual change of user behavior, but will miss recognising situations where a user systematically attacks an organisation over an extended time-framework.

III. Overview of proposed approach

We propose a hybrid model based on multi-dimensional features for insider threat detection which is illustrated in Figure2. First, we analyze the characteristics of each domain(e.g., logons) from multi-source information as parameters into Across-Domain Anomaly Detection(ADAD) to get anomaly scores. Second, we extract the features of time-series behaviors of users from multi-source information, then we identify changes in activities of a user compared to that users past activities through Across-Time Anomaly Detection(ATAD) to get anomaly scores. Finally, we combine the anomaly scores from the two components to detect insider threat. Next, we will introduce the experimental data.

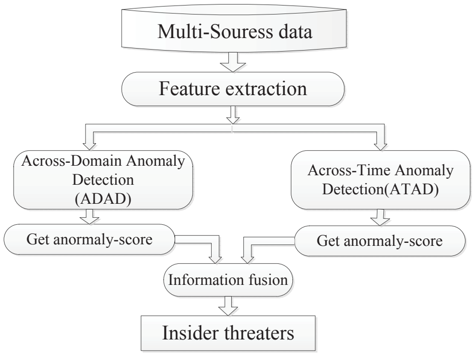


Fig. 2: Anomaly Detection Framework.

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 [**?**]. The dataset R4.2.tar.bz has been used for this analysis. According to the dataset owners, this is a dense needle dataset with a fair amount of red team scenarios. 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. Logon activity corresponds to either a user login event or a screen unlock event, while the Logoff event corresponds to user logoff event. The third data file device.csv is a collection of data records of removable media usage. It 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

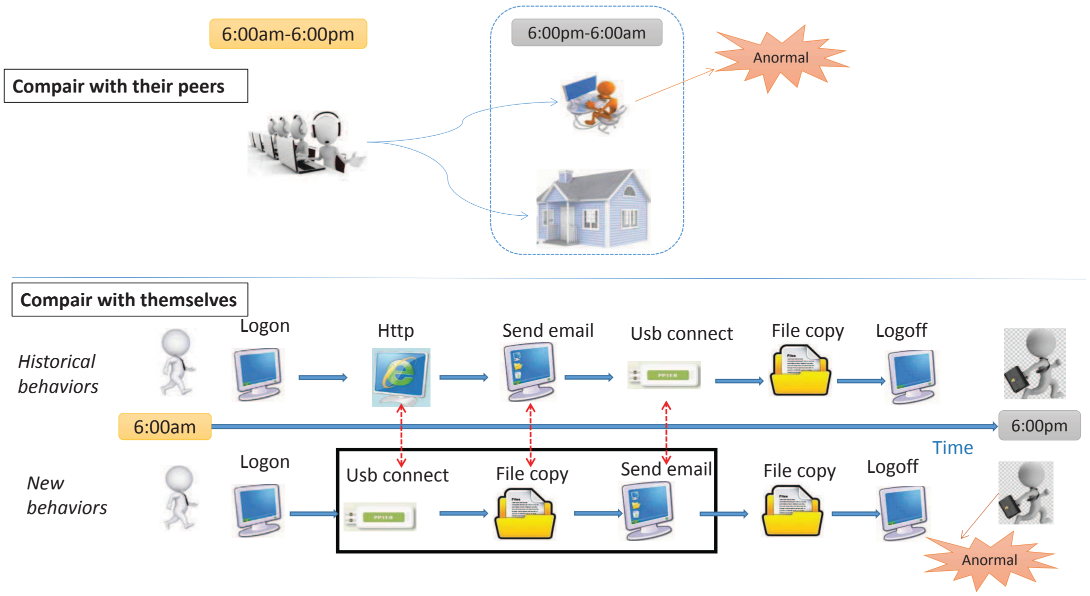


Fig. 1: Example of insider threat.

they are acting maliciously or not), which allows us to monitor the success or failure of our experiment.

IV. KEY METHODOLOGIES

In this section, we describe our across-domain anomaly detection(ADAD) model in detail at first. We next describe our proposed improved MM based on across-time to detect insider threats. Finally, we introduce the fusion method combining ADAD and ATAD.

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

This framework will utilize multidimensional inputs, such as user interactions with hardware assets, logon records and operation on file, to identify anormal users who behave differently from their peers. The figure 2 reports the structure of the approach.

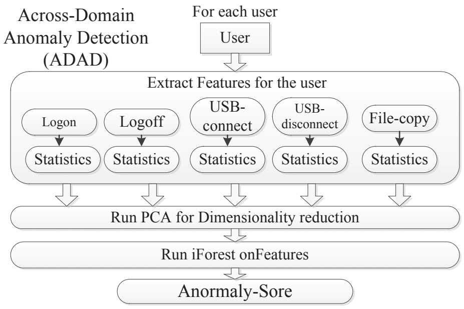


Fig. 3: An overview of the Across-Domain Anomaly Detection(ADAD).

1) Feature Extraction:

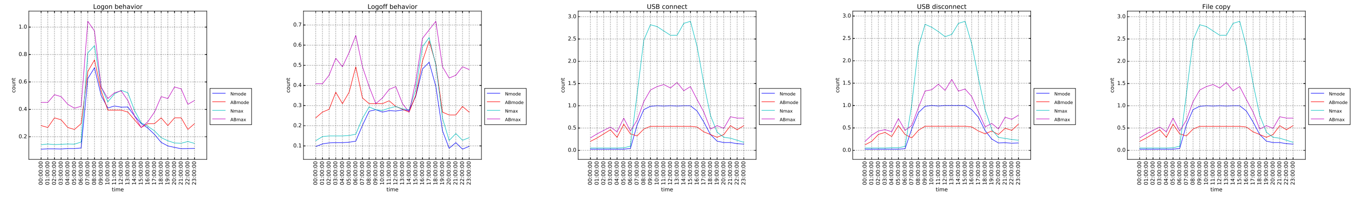
Individual Logon-Logoff Behaviors. This parameters can be used in identifying users abnormal logon/logoff activities, as most disgruntled insiders tend to commit malicious activities after hours [**?**]. Identifying users baseline behavior on system/device access is an essential part of malicious insider threat detection problem. For normal users and abnormal users, two parameters (the average of their maximum and mode) logon and logoff values have been calculated for every hour shown as Figure4.

Removable media usage. Removable media is among the most popular method used in theft of Intellectual Property (IP) in extracting confidential information from organizations [**?**]. Tracking the use of removable media can be an excellent information source for identifying suspicious events by trusted insiders . Baseline behavior of removable media usage is captured by the average of their maximum and mode time of Insert and Remove activities as in the logon/logoff event analysis. Figure4 shows it.

File copy Behaviors. The behaviors of file copying has some differences bwtween normal users and abnormal users. The average of the number of files copied per hour by normal and abnormal is shown in the figure4.

From the figure4, we found that there is a a big difference in behavior between normal user and abnormal user at different times, so we decide to merge the times of behavior every 6 hours as the parameter to input to our ADAD model. Figure 5 is an illustration of the parameters.

After the experiment we found that the features we extracted had noise effects (which in detail in the experimental part), in order to achieve higher accuracy, we use PCA [**?**] for denoising. All feature columns are normalized before the PCA decomposition is performed. By default, we consider a decomposition of the features to a 2-D space.



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

Fig. 4: Removable media usage behavior

TABLE I: Selected parameter set.

Module	Parameter (00:00-06:00)(06:00-12:00) (12:00-18:00)(18:00-24:00)
Logon events	Maximum/Mode Logon counts
Logoff events	Maximum/Mode Logoff counts
Removable Media	Maximum/Mode Connect counts Maximum/Mode Disconnect counts
File copy events	Maximum/Mode Filecopy counts

2) Anomaly Detection: Due to the complex nature of insider threat problem, it is extremely hard to pinpoint a user as a malicious insider. Therefore, the first step should be the identification of possible malicious insiders who are maximally deviating from peers as well as their normal behavior. Therefore, as the second stage of our analysis, we will focus on implementing an anomaly detection algorithm based on the the properties identified at the previous stage of this analysis. The anomaly detection algorithm adopted in this analysis is the "Isolation forest" algorithm, which stands out in effectively separating anomalous events from the rest of the instances [**?**].

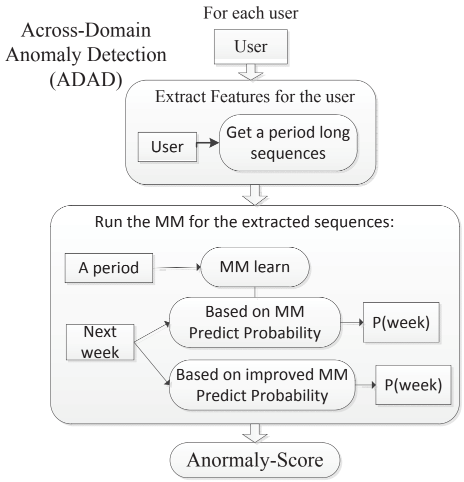

Across-Domain
Anomaly Detection
(ADAD)
For each user
User
Extract Features for the user
User
Get a period long sequences
Run the MM for the extracted sequences:
A period
MM learn
Next week
Based on MM Predict Probability
P(week)
Based on improved MM Predict Probability
P(week)
Anormaly-Score

Fig. 5: An overview of the Across-Time Anomaly Detection(ATAD).

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

Markov Model (MM) [**?**] 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) due to its high detection rate [**?**]. MM has also been used in the general area of intrusion detection by some notable works ([**?**]). Since we model users behaviors as a temporal sequence of observable query anomaly scores in our work in this section, and proposed an improved MM model to detect insider threats. Figure 6 shows a graphical overview of the processing approach and pipeline we use, and the remainder of this section is dedicated to describing each stage in it.

1) Model building: This detection model is implemented based on Markov Models taking historical behaviors features into account. Users have defferent behaviors on computers every day. When a user finish a behavior and begin the next behavior, a new record is created. The historical behaviors may be 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*.

Consider a user whose behaviors history is $B = a_1 a_2 ... a_n$. Let substring $B(i,j) = a_i a_{i+1} \ldots a_j$ for any $1<i<j<n$. We think of the users behaviors as a random variable *X*. Let $X(i,j)$ be a string $X_i X_{i+1} \ldots X_j$ representing the sequence of random variates $X_i, X_{i+1}, \ldots X_j$ for any $1<i<j<n$. Define the context $c = B(n-k+1, n)$. Let *A* be the set of all possible behaviors. For all $a \in A$ and $i \in \{1, 2, ..., n\}$, we let the notation $P(X_i = a_i | \ldots)$ denotes the probability that X_i takes the value a_i. These probabilities that can be represented by a *transition probability matrix M*. Depending on the assumption of Markov, both the rows and columns of *M* are indexed by *length-k* strings from A^k so that

$$P(X_{n+1} = a | X(1,n) = B(1,n)) = M(s, s'), \qquad (1)$$

where $s = B(n-k+1, n)$ is the current context and $s' = B(n-k+2, n)a$ is the next context. .In that case, knowing *M* would immediately provide the probability for each possible next symbol of *B*. In this study, we let *k* equal to 0 for more historical information for building users profile.

Since we do not know *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

$$P_k(a) = \hat{P}(X_{n+1} = a | B)) = \frac{N(ca, B)}{N(c, B)}, \qquad (2)$$

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

Given this estimate, we can easily define the behavior of the Markov model. It predicts the symbol $a \in A$ with the maximum probability $\hat{P}(X_{n+1} = a|B)$; that is, the symbol that most frequently followed the current context *c* in prior occurrences in the history. We will introduce the anomaly detection in the next section.

2) Anomaly Detection: In this section, an improves Markvo method(IM) proposed is to detect users' behaviors. For intrusion detection, we wanted to build a longterm norm profile of temporal behavior, and to compare the temporal behavior in the recent past to the long-term norm profile for detecting a significant difference. Using formulae (2), we trained and built a stationary Markov chain model of temporal behavior as the long-term norm profile by learning the transition probability matrix emphM from a stream of users behaviors that was observed during a period of time.

We defined the temporal behaviors in the recent past by opening up *W* continuous observation windows of size *N* on the continuous steam of users behaviors to view the users behaviors from the current time *DN*: $a_{11}, a_{12}, \ldots, a_{1N}, a_{21}, a_{22}, \ldots, a_{2N}, \ldots, a_{DN}$, where a stands for behavior. First, using formula (2), we compute the probability that the sequence of states $a_{11}, a_{12}, \ldots, a_{1N}, a_{21}, a_{22}, \ldots, a_{2N}, \ldots, a_{DN}$ occurs in the context *c* of length *k*.

Second, given the *M* matrix produced in the first step, the anomaly score R_{trend} for the user is calculated as

$$R_{trend} = \frac{\sum_{d=1}^{D} \prod_{t=1}^{N} P_k(a_{dn})}{D}, \tag{3}$$

In this study, we let *N* equal to a week. Because a week can build a complete profile for a user as a week cycle.

Third, 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 improved MM 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

Our goal is to combine suspicion/anomaly scores that have been generated from each of the aforementioned methods to detect anomalies. Therefore, we developed a technique based on weight fusion to combine acorss-domain and across-time of evidence from multiple domains. In a broad sense, each source of information provides a suspicion score and the goal is to combine these scores by wight(*W*) in order to identify anomalies with greater accuracy. The combined suspicion score $R_combine$ is computed as

$$R_{combined} = W_1 * R_{ADAD} + W_2 * R_{ATAD} \tag{4}$$

$$r = \begin{cases} \frac{W_2}{W_1} & \text{if } W_1{>}0 \\ 1 & \text{if } W_1 = 0, \end{cases} \tag{5}$$

we can then set a threshold T_{com}, which we use to classify users as anomalous or not. If the suspicion score of the user is below the threshold, we classify it as anomalous.

TABLE II: The experiment result of ADAD.

Domian	iForest Input	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 properties	MAX	82%	68%	34%
	MODE	79%	52%	31%
PCA	PCA	87%	79%	35%

V. Experiment

This section is dedicated to a comprehensive discussion of results obtained through our analysis. In the next sections, we will introduce the performance metrics for evaluating the three detection models separately. We then provide a comparative assessment of our proposed models with other existing detecting insider threat methods.

A. Across-Domain Anomaly Detection(ADAD)

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

1) Evaluation Metrics and Baselines: Because the isolation forest can give an anomaly score for each user, we consider the metrics *accuracy, Precision* and *recall* to quantitatively evaluate the models. *Precision* is the fraction of the data entries labelled 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 [**?**].

2) Comparative Evaluation: Table II shows the results of running our model ADAD with different parameters. We found that removable media domains is the most detectable (with the highest detective accuracy), while logon and file domains are harder to detect insider threats. It appears that users show great variations in their logon and file behaviors, but are more uniform in the behavior of device usages . Except that, the property of the mode is more effective than maximum to detect, which 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* (Principal Component Analysis) to decompose of the features to a 2-D space achieves a higher predictive 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: We need to set a threshold T to classify users as anomalous or not. So we consider the metrics detection Receiver Operating Characteristic curves (or *ROC* curves) curve to quantitatively evaluate the models. These curves are used to plot the true-positive (correct) rate against the false-positive rate for the different possible points in a diagnostic test. In order to produce the ROC curves, we use a process of threshold varying. After obtaining the anomaly score for each user using our model, we set a value for the threshold, check each user one-by-one, and classify them as anomalous if $score > threshold$. We then increase the threshold and repeat until no users are classified as anomalous.

Comparison Methods. The improved Markov (IM)is respectively compared with existing detection Markov:

a) Markov: when caculating the anomaly score, this model regard the behavior sequences of the week as 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.** We apply our Across-Time Anomaly Detection(ATAD) detection result as figure6. 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 has a 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 insder threats.

In other words, IM is more likely to detect abnormal changes in behavioral trends over a period of time. So it detects insider threat more robustly.

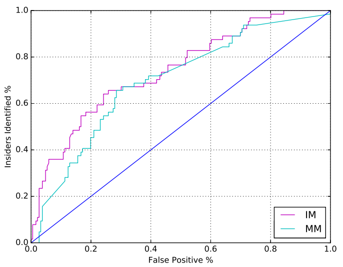


Fig. 6: ROC Curve showing the differences between MM algorithm and improved MM algorithm.

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 is to determine the proportion of each component. So we let r which demotes the proportion of the two model increase from 0 to 1. Then we use the metrics *accuracy, Precision* and *recall* to quantitatively evaluate the models.

2) Comparative Evaluation: The method based on weight fuses the ATAD and the ADAD, then get the final anomaly score. The result is shown as Table 2. Parameter r represents the proportion of the two methods' anomaly score. 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 that only the ADAD is better than only the ATAD, and fusingh two model is better than the ADAD. When r values varies between 1/9 and 3/7, we can clearly see that the precision has a greater improvement. The ADAD model is based on data driven detection, 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 to detect insider threats. So, it is remarkable that a suitable combination of the individual ATAD and ADAD scores in an appropriate fashion leads to significant improvement in performance relative to any of the individual ATAD or ADAD sources. Espeicially, when r is 3/7, the precision and recall rate is the highest, and the combination of weights is the best.

To get a better understanding and visualization of results based on our approach, we mark the positive sample with red and negative sample with blue. Figure 8 is an indication of how the anomaly scores are distributed when r is 3/7 in this analysis. The graph indicated few points above the red color horizontal line which is equivalent to an anomaly score of 1.3, and the threshold is determine by *ROC* curves. Users whose naomaly scores exceed the threshold can be considered as anomalous users. It reached 95% of the accuracy. 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.

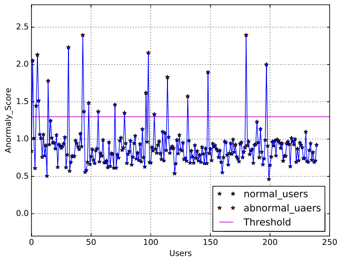


Fig. 7: Anomaly score distribution.

VI. Conclusion And Future work

In this paper we have examined the problem of insider threat detection. We propose a hybrid model that combines a data

TABLE III: The experiment result of information fusion.

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%

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 a insider threater if the anomaly score exceeds the threshold. After a wide range experiment, it is verified that the hybrid model can detect insider threats with the accuracy of 95

Future work. We use a combined method based on weight fusion to integrate the ADAD and the ATAD, which has a high precision but a pessimistic recall. To solve this problem, we could apply on some complex and effective fusion scheme to combine their information in order to improve the recall of anomaly detection(e.g., [?]). In addition, we could take user role into account in generates user Normal portrait that can describe the full extent of activities that users perform within the organization based on role to improve the recall of anomaly detection(e.g., [?]).

Acknowledgment

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

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

[3] 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.

[4] 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

[5] 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

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

[7] 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.

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

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

[10] 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

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

[12] 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.

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

[14] 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.

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

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

[17] Automated Insider Threat Detection System Using User and Role-Based Profile Assessment Philip A. Legg, Oliver Buckley, Michael Goldsmith, and Sadie Creese