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

*Note: Sub-titles are not captured in Xplore and should not be used

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 threat.However, they always consider single dimension feature to detect insider threat. Some detect through comparing behaviours across different individuals or some compare behaviours acorss different times for the same individual to detect. Both of them lead leak detection This paper focuses on detecting from different dimensions to improve the accuacy. We propose two anomaly detection methods.The first method seeks to identify anomalous behavior that blends within each information source.The second method identifies unusual changes in behavior over time using a Markov model approach. Finally, we present a fusion approach that integrates evidence from both methods to improve the accuracy and robustness of the insider threat detection system.**

***Index Terms*—anomaly detection, insider threat detection, information fusion, machine learning**

## I. Introduction

Insider threats are threats with malicious intent directed towards organizations by people internal to the organization. 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 [3] 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. Moreover, attacks by malicious insiders are also the costliest to fix ($145,000), followed by denial of service ($127,000) and Web-based attacks ($96,000), indicating the severity of this problem.

Existing literature focuses on two dimentions of the detection models: data driven detection [9] and behavior driven detection . The first model aims to find a normal portrait in all users data in order to detect insider threat that does not meet this normal portrait. For exampl as the "compare with their peers" in Figure1. After get off work, the general behavior is that everyone will go home to rest, and some of the few people will choose to continue working after the work, even in the early hours will still deal with some of the company's sensitive information services. This is one of the abnormal scenarios we have taken into account based on data-driven. However,it will lead some of the special nature of the staff to be the error detection anomalies. If someone strictly observes the regular schedule but has some unusual change in his daily work which can't be detected by the data-driven model. The second model regards the abnormal change in the behavior as the basis for detection, which will miss some of the true threat with abnormal behavior.For instance as the "compare with themselves"in figure1, an employee's long-term behavior is to click on the browser after booting, and view the email then reply to a series of messages. But one day, he connected the mobile device after booting , and a series of operations on the file-copy, which does not match his usual style of doing things. This series of operations is also sensitive to company's information.However, the model just work out based on unusual change of user behavior but anomaly behavior from others which suitable for data-driven model.

Insider threat research and surveys suggest this problem cannot be considered only as a data driven problem; it needs to be considered as data and behavior driven problem [9].Thus we propose a fusion method based on multi-dimensional features that combines models based on data-driven (ADAD)and behavior-driven(ATAD) to detect in a more robust and accurate manner. The multidimensional data from the enterprise network is formatted and fed separately into the ADAD and ATAD, which generates an abnormal score representing the users unusual behavior. Then we get the final abnormal score based on different dimensions by fusion the abnormal scores for abnormal detection.

**Contribution:** In this paper we have examined the problem of insider threat detection. Our main contribution is that a novel fusion approach for accuracy detection of anomalies is discussed. The two main components of our framework - ADAD and ATAD - improve anomaly detection prediction

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

accuracy by combining information from multiple domains and time-instances.As a result, these methods are able to determine anomalies that not only act inconsistent with peers but also has unusual change compared with their historic action with accuracy of 95%.

## II. Related work

Methods based data analytics include the work done by Mathew et al. [5] on account of user access patterns, the work of Eberle et al. [6] on using social graphs to detecting the abnormal.More recently, Eldardiry et al. [22] have also proposed a system for insider threat detection based on feature extraction from user activities. Michael Goldsmith use some methods based on fusion of multi-source information and user behavior [31]. There have also been various approaches based on specific insider threat scenarios to detect abnormal [7] [8]. However, they did not factor in the change of user behavior 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 papers based on behavioral models [10] [11] [12]. 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. Tabish Rashid [30] takes the change of user behavior over time to detect the anomaly and achieved some results ,however,it has to pay algorithm complexity.

Considering these two factors, we propose a method with high accuracy and low algorithm complexity, which is of great significance in industry and scientific research. We fusion two method based on data and behavior. The first one, we compare behaviours across different individuals,while in the latter,we compare behaviours acorss different times for the same individual.

The remainder of this article is structured as follows. Section 3 we detail our approach and discuss algorithms used and how we train our model to detect insider threats. Next, in Section 4 we detail our implementation and analyze the reault. Finally we conclude in Section 5, also presenting limitation of our work and outlining avenues for future research.

## III. Overview of proposed approach

We propose a fusion method based on multi-dimensional features for insider threat detection which is illustrated in Figure 2.First, we count and 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.

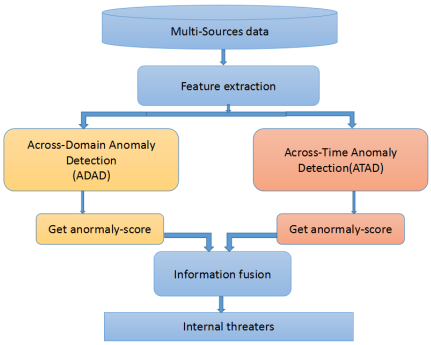



Fig. 2: Anomaly Detection Framework.

### 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 [32]. 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.

## IV. APPROACH

### 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 different from their peers. We believe that it is sufficient to consider all behaviors from every domain to characterize the user's behavior. The figure 2 is the structure of the approach.

*1) Feature Extraction:* **Individual Logon-Logoff Behavior.** This parameter can be used in identifying users abnormal logon/logoff activities as most disgruntled insiders tend to commit malicious activities after hours [33]. 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

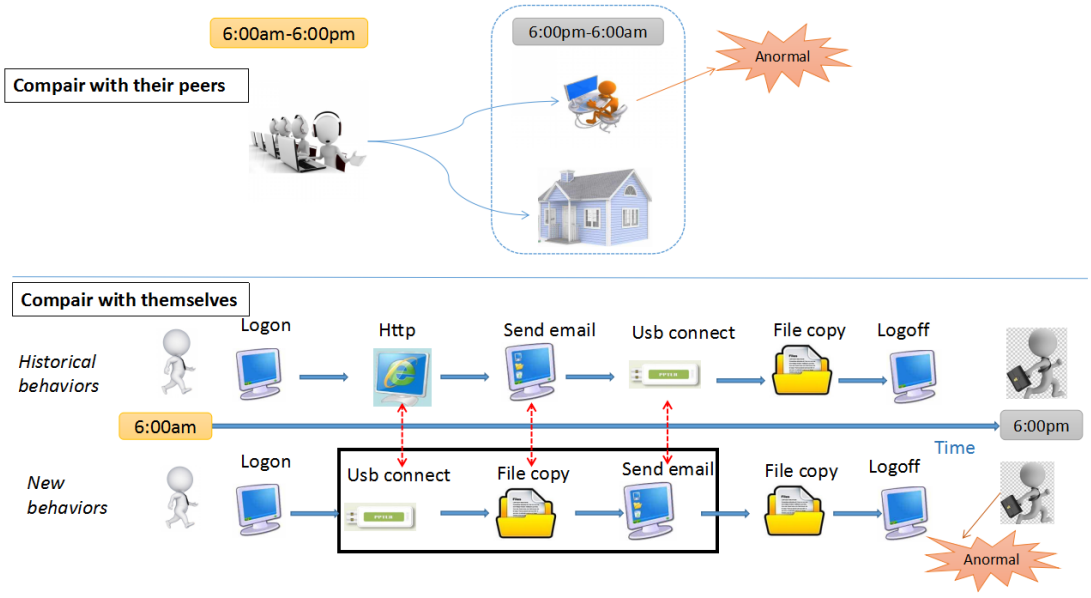



Fig. 1: Example of insider threat.

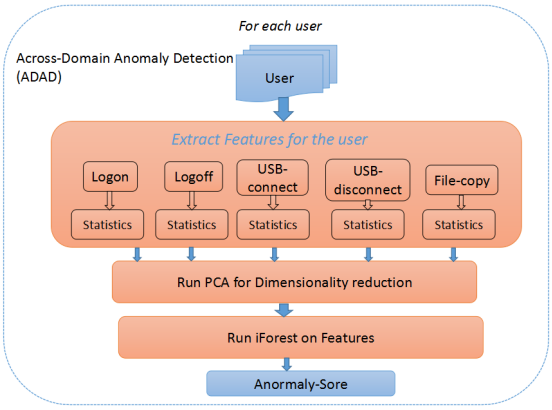
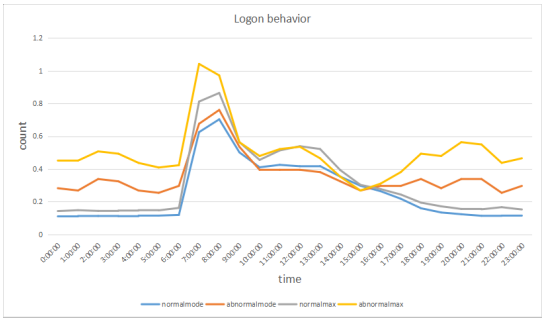
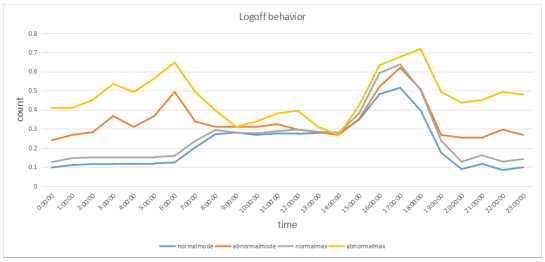



Fig. 3: An overview of the Across-Domain Anomaly Detection(ADAD).




(a) Logon behavior




(b) Logoff behavior

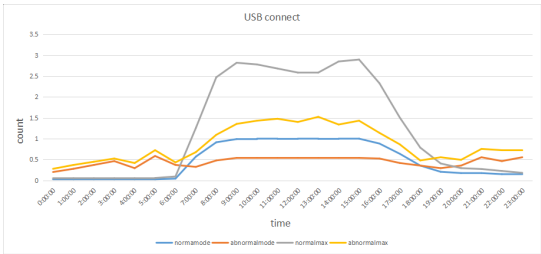
Fig. 4: Users logon and logoff behavior

[34]. 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. The average of the number of files copied per hour by normal and abnormal is also used in this analysis. Figure5 shows it.

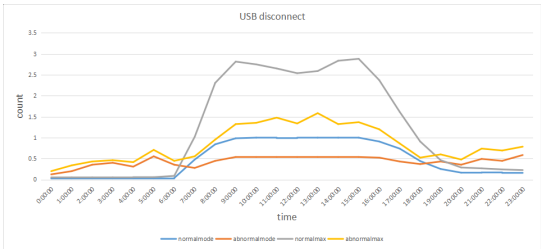
From above, we found that there is a a big difference in behavior between normal user and abnormal user at different times, so we decide to merge the times of behavior every 6 hours as the parameter to input to our ADAD model. Figure 5 is an illustration of the parameters.

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 [35] 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.

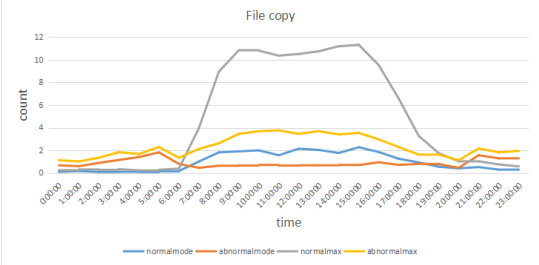
*2) Anomaly Detection:* Due to the complex nature of insider threat problem, it is extremely hard to pinpoint a user as a malicious insider. Therefore, the first step should be the identification of possible malicious insiders who are maxi-




(a) USB connect




(b) USB disconnect




(c) File copy

Fig. 5: 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<br>Maximum/Mode Disconnect counts |
| File copy events | Maximum/Mode Filecopy counts |

mally 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 [36].

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

In this section we use Markov chain model [38] to detects individuals with unusual changes in activity. 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.

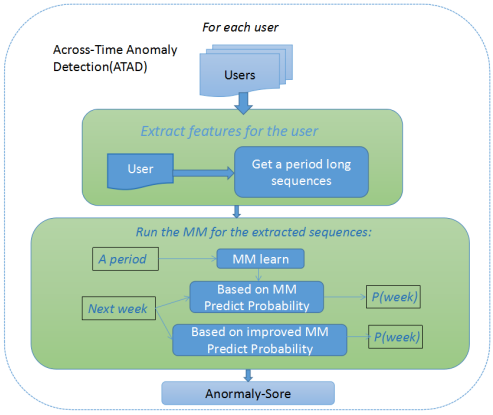



Fig. 6: An overview of the Across-Time Anomaly Detection(ATAD).

*1) Feature Extraction:* The CERT Dataset contains many log files which describe a particular kind of activity for all users (users being the same as employees), for instance http.csv contains logs pertaining to web browsing (website, date accessed, username, etc.). We load each file, and assign a symbol to each entry based on the set of features we are defined in ADAD. After this, we join all of these files, partition them by each user and then finally sort them based on the respective timestamps. We get a list of actions and the time at which they undertook them for each user. We then devide these actions into two parts, one is the action in a period, which gives us the final output for the feature extraction phase as the train data. The other is the next week as the test data

*2) Anomaly Detection:* Actions correspond to MM model states. Let $P_{ij}$ denote the probability that the user is in a state $j$ at time $t+1$ given the user is in state $t$ at time $t$. where $q_i$ is the probability that the system is in state $i$ at time 0, and The probability that a sequence of states $X_1, \cdots, X_T$ at time $1,\cdots, T$ occurs in the context of the stationary Markov chain is computed as follows:

$$P(X_1, ..., X_T) = q_{x1} \sum_{t=2}^{T} P_{X_{t-1}X_t} \tag{1}$$

Individuals are scored based on their total transition likelihood over time, and suspicious individuals with unusual transitions between temporal states are detected. We use three methods to compute the users score of the next week, then compare the result of detection between these methods.

**Baesd on MM:**

$$R_{mm} = P(X_1, ..., X_T) \tag{2}$$

**Based on improved MM:**

First,We treat the sequences of the week($X_1$,$X_2$,.....$X_T$) as a whole,the anomaly score $R_{point}$ for the user is calculated

by estimating the users transition likelihood over time. The anomaly probability score is computed as

$$R_{point} = q_{x1} \prod_{t=2}^{T} P_{X_{t-1}X_t}.. \tag{3}$$

Second,we regard the sequences of every day during this week ($X_{11}$,$X_{12}$,.....$X_{WT}$)as a whole and treat the average daily anomaly score as the anomaly score of the week named $R_{trend}$ computed as

$$R_{trend} = 1/w \sum_{t=1}^{W} q_{x1} \prod_{t=2}^{T} P_{X_{w(t-1)}X_w t} \tag{4}$$

### 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 is computed as

$$R_{combined} = W_1 * R_{ADAD} + W_2 * R_{ATAD} \tag{5}$$

$$r = \begin{cases} w_2/w_1 & \text{if } w_1 > 0 \\ 1 & \text{if } w_1 \equiv 0 \end{cases} \tag{6}$$

## V. Experiment

This section is dedicated to a comprehensive discussion of results obtained through our analysis.

### A. Across-Domain Anomaly Detection(ADAD)

Figure 7 shows the results of running our model ADAD with those parameters. We found that removable media domains is the most predictable (with the highest predictive accuracy), while Logon and file domains are harder to predict. It appears that users show great variation in their logon and file behavior, but are more uniform in device usage . Except that, the property of the mode is more effective than maximum to detect, which indicates that the mode represents the difference between normal and abnormal. Finally, we use *PCA* to decompose of the features to a 2-D space,and get a higher predictive accuracy.

### B. Across-Time Anomaly Detection(ATAD)

We apply our Across-Time Anomaly Detection(ATAD) detection method as follows. It's obvious that the improved MM method is more effective than the mathod based on MM. For improved MM,the impro_Point(iP) is the experience that the week is a whole. The impro_Trend(iT) is the experience that the day is a whole, based on which we regard the average of this week's daily anomaly score as the final anomaly score of the week. Compare with the iP, iT has a better result. This is because that for iP if a user has once unusual change during this week,which will has a important influence on the final result,maybe the change just is working overtime. This will has effect on the detection. However, it can avoid the miscarriage of justice brought about by this accidental abnormal changes to some extent for iT. In other words, iT is more likely to detect abnormal changes in behavioral trends over a period of time, which detect insider threat more accurately. We regard AAS as the result of ATAD to combine with ADAD in next section.

TABLE II: The experiment result of ADAD.

| iForest Input | | Accurate | 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% |
| All properties | dimension=2 | 87% | 79% | 35% |

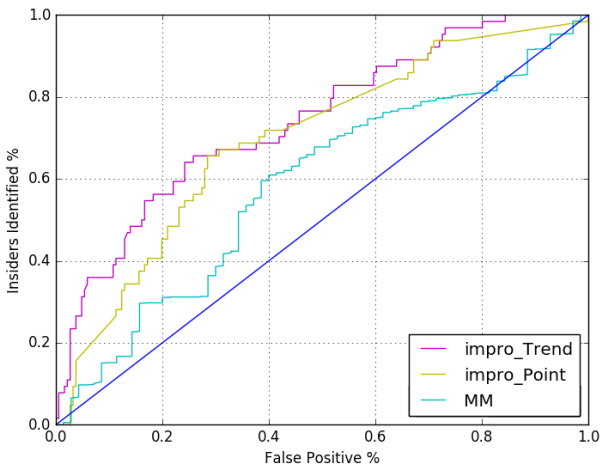



Fig. 7: ROC Curve showing the differences between MM algorithm and improved MM algorithm.

### C. Information fusion

We get the final anomaly score based on weight fusion as shown in Table 2. r represents the proportion of the weight of the two methods score. The scores are normalized before the weight fusion(combine).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 when r is 3/7, the accuracy and recall rate is the highest, where the combination of weights is the best. Indeed, it is remarkable that a suitable combination of the individual ATAD and ADAD scores in an appropriate fashion leads to significant improvement in

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% |

performance relative to any of the individual ATAD or ADAD sources.

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. Users belong to those points can be considered as anomalous users.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. It indicates that 95 percent of the insider threaters we detected are true insider threaters.

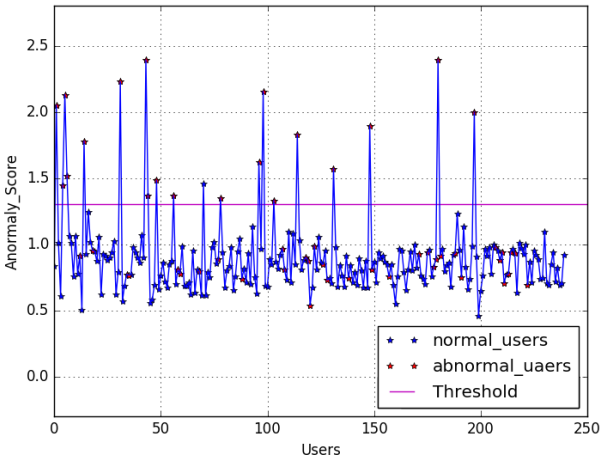



Fig. 8: Anomaly score distribution.

## VI. Conclusion And Future work

In this paper we have examined the problem of insider threat detection. Our main contribution is that a novel fusion approach for robust detection of anomalies is discussed. The two main components of our framework - ADAD and ATAD - improve anomaly detection prediction accuracy by combining information from multiple domains and time-instances.

**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., [42]). 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., [43]).


## 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] A. M. Dawn Cappelli Randall Trzeciak,Timothy J. Shimeall,Common Sense Guide to Prevention and Detection of Insider Threats , 3rd Edition, 2009.

[2] Gemalto. Breach level index—data breach database & risk assessment calculator, 2016. http://www.breachlevelindex.com/.

[3] 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

[4] Teresa F Lunt. A survey of intrusion detection techniques. Computers & Security, 12(4):405418, 1993.

[5] 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.

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

[7] Alex Memory, Henry G Goldberg, and E Ted. Context-aware insider threat detection. In Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence, 2013.

[8] Robert F Mills, Michael R Grimaila, Gilbert L Peterson, and Jonathan W Butts. A scenario-based approach to mitigating the insider threat. Technical report, DTIC Document, 2011.

[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] Miltiadis Kandias, Alexios Mylonas, Nikos Virvilis, Marianthi Theoharidou, and Dimitris Gritzalis. An insider threat prediction model. In Trust, privacy and security in digital business, pages 2637. Springer, 2010.

[11] Frank L Greitzer, Lars J Kangas, Christine F Noonan,and Angela C Dalton. Identifying at-risk employees: A behavioral model for predicting potential insider threats. Pacific Northwest National Laboratory Richland, WA, 2010

[12] GB Magklaras and SM Furnell. Insider threat prediction tool: Evaluating the probability of it misuse. Computers & Security, 21(1):6273, 2001.

[13] Hoda Eldardiry, Evgeniy Bart, Juan Liu, John Hanley, Bob Price, and Oliver Brdiczka. Multi-domain information fusion for insider threat detection. In Security and Privacy Workshops (SPW), 2013 IEEE, pages 4551. IEEE, 2013.

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

[15] Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams Aaron Tuor and Samuel Kaplan and Brian Hutchinson Western Washington University Bellingham, WA Nicole Nichols and Sean Robinson Pacific Northwest National Laboratory Seattle, WA

[16] 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.

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

[18] ] M. Bishop, S. Engle, S. Peisert, S. Whalen, and C. Gates, We have met the enemy and he is us, in Proc. NSPW, Lake Tahoe, CA, USA, Sep. 2008, pp. 112

[19] J. R. C. Nurse et al., Understanding insider threat: A framework for characterising attacks, in Proc. IEEE SPW, 2014, pp. 214228

[20] F. Kammueller and C. W. Probst, Invalidating policies using structural information, J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl., vol. 5, no. 2, pp. 5979, Jun. 2014.

[21] M. R. Ogiela and U. Ogiela, Linguistic protocols for secure information management and sharing, Comput. Math. Appl., vol. 63, no. 2, pp. 564572, Jan. 2012.

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

[23] Brdiczka et al., Proactive insider threat detection through graph learning and psychological context, in Proc. IEEE Symp. SPW, San Francisco, CA, USA, May 2012, pp. 142149.

[24] W. Eberle, J. Graves, and L. Holder, Insider threat detection using a graph-based approach, J. Appl. Security Res., vol. 6, no. 1, pp. 3281, Dec. 2010.

[25] B. Klimt and Y. Yang, The enron corpus: A new dataset for email classification research, in Machine Learning: ECML 2004, vol. 3201, Lecture Notes in Computer Science, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds. Berlin, Germany: Springer-Verlag, 2004, pp. 217226

[26] T. E. Senator et al., Detecting insider threats in a real corporate database of computer usage activity, in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2013, pp. 13931401.

[27] P. Parveen, J. Evans, B. Thuraisingham, K. W. Hamlen, and L. Khan, Insider threat detection using stream mining and graph mining, in Proc. IEEE 3rd Int. Conf. Social Comput. PASSAT, Oct. 2011, pp. 11021110.

[28] P. Parveen and B. Thuraisingham, Unsupervised incremental sequence learning for insider threat detection, in Proc. IEEE Int. Conf. ISI, Jun. 2012, pp. 141143.

[29] S. Greenberg, Using unix: Collected traces of 168 users, Univ. Calgary, Calgary, AB, Canada, Tech. Rep., 1988.

[30] A New Take on Detecting Insider Threats: Exploring the use of Hidden Markov Models

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

[32] CERT Insider Threat Data Set, Software Engineering Institute, Carnegie Mellon University, CERT Division and Exact Data LLC. [Online]. Available: https://www.cert.org/insiderthreat/tools/

[33] 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.

[34] 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.

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

[36] 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

[37] L. Sun, S. Versteeg, S. Boztas, and A. Rao, Detecting anomalous user behaviour using an extended Isolation Forest algorithm: An enterprise case study, Sept 2016, arXiv:1609.06676.

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

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

[40] T. M. Mitchell, Machine Learning. Boston, MA: McGraw-Hill, 1997.

[41] J. R. C. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. Wright, and M. Whitty. Understanding insider threat: A framework for characterising attacks. In IEEE Security and Privacy Workshops (SPW). IEEE, 2014. DOI: 10.1109/SPW.2014.38.

[42] 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.

[43] 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.