International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2310647

Insider Threat Detection Using Usupervised Learning

Tushar Yadav[1], Akash Manoj[1],Tanishq Ghanshani[1], Wajahat Hussain[1], Dr. Kapil Vhatkar[2]

[1]Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India

[2]Assistant Professor, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, Maharashtra, India

ARTICLEINFO

Article History :

Accepted: 15 Dec 2023
Published: 30 Dec 2023

Publication Issue :
Volume 10, Issue 6
November-December-2023

Page Number :
361-367

ABSTRACT

In recent years, there has been a noticeable upsurge in insider threat incidents, resulting in substantial losses to companies and organizations. The escalating frequency of these incidents poses a formidable challenge to the realm of internal network security. Traditional intrusion detection methods, predominantly engineered to detect external threats, have proven insufficient in identifying the intricate and often subtle malicious behaviors of insiders. Consequently, this has propelled the urgent need for innovative and effective solutions within the domain of insider threat detection technology.

This research project addresses the imperative call to action by introducing a novel and holistic approach to insider threat detection. Our methodology is firmly grounded in two pivotal components, each meticulously designed to address specific facets of the challenge. The first pillar of our approach employs the tree structure method, a sophisticated technique that meticulously dissects and comprehends user behavior. By employing a hierarchical structure to disentangle user actions, this method meticulously captures the dynamic relationships and dependencies inherent in human- computer interactions. The outcome of this analysis is the creation of feature sequences, chronicles of user actions ordered by time, providing a multifaceted and in-depth view of user behaviors.

Complementing this intricate framework is our incorporation of the Copula-Based Outlier Detection (COPOD) method. COPOD harnesses advanced statistical methodologies, notably copulas, to discern outliers and anomalies within the feature sequences derived from our user behavior analysis. This dynamic interplay of analytical methods allows us to single out and flag users whose behavior exhibits deviations from established norms, effectively identifying potential insider threats within the organization.

The efficacy and practicality of our approach were rigorously tested using the CERT-IT datasets, which encompasses a wide array of insider threat scenarios. Our experiments revealed the superior performance of our approach when compared to conventional methods, most notably the Isolation Forest technique. Crucial performance metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC), underscore the resounding success of our method in safeguarding against insider threats.

In summary, this research project stands as a groundbreaking contribution to the field of insider threat detection. Our synthesis of the tree structure method and COPOD offers a potent, versatile, and adaptive solution to the formidable challenges faced by organizations seeking to fortify the security of their internal networks. Moreover, our findings illuminate the vast potential of unsupervised machine learning techniques in combating insider threats with unwavering precision, providing a robust defense against the evolving and intricate threat landscape.

Keywords: ML-Machine Learning, Unsupervised Learning Method, COPOD, Insider Threats, Internal Network Security, Tree Structure Method, Form Feature Sequence, SVM-Support Vector Machine, Detection Model

I. INTRODUCTION

In today's complex and interconnected world, where digital assets and sensitive information have become the lifeblood of organizations, the specter of insider threats casts a long shadow over the domain of cyber security. Unlike external threats, which typically emanate from malicious actors beyond an organization's boundaries, insider threats originate from individuals who have access to the organization's systems, networks, and data. These individuals, often employees or trusted affiliates, possess the keys to the kingdom, making them a formidable and often imperceptible security risk.

The rise of insider threat incidents in recent years has prompted a reevaluation of security paradigms and the emergence of a pressing concern for organizations worldwide. Insider threats manifest in various forms, encompassing data breaches, intellectual property theft, fraud, and unauthorized access. These incidents not only lead to financial losses but also compromise the integrity, confidentiality, and trustworthiness of an organization's operations.

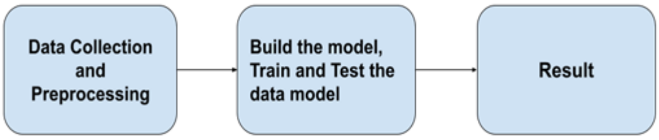
Data Collection and Preprocessing
Build the model, Train and Test the data model
Result

Fig.1 Working of a machine learning model

The challenge posed by insider threats is compounded by their elusive and multifaceted nature. Unlike external threats, which may be identified through known attack patterns or signatures, insider threats often involve legitimate access, making them more difficult to detect. Malicious insiders can employ subtle and non-obvious tactics, exploiting their privileged positions to evade traditional security measures.In response to this growing menace, organizations are increasingly focusing on strengthening their internal

network security. The imperative lies in the development and deployment of sophisticated solutions that can proactively identify and mitigate insider threats. This report is a testament to the gravity of this challenge and a testament to the innovation required to address it. The report presents an innovative and adaptive approach to mitigate insider threats. Our methodology leverages unsupervised machine learning techniques to detect abnormal user behavior, enabling the early identification of potential insider threats before they can cause harm. Central to this approach are two essential components: the tree structure method, designed to delve deeply into user behavior, and the Copula-Based Outlier Detection (COPOD) technique, adept at pinpointing anomalies within feature sequences.

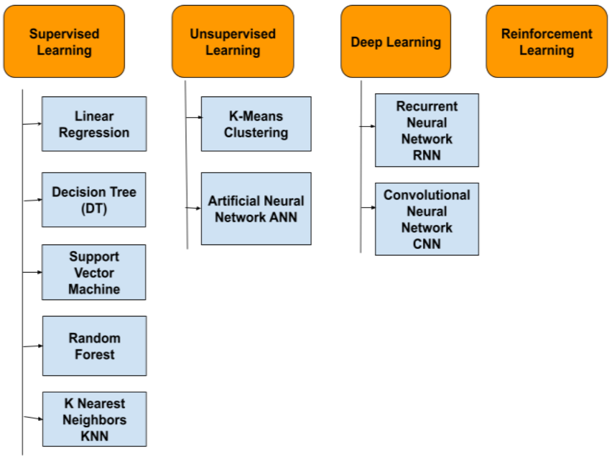

Fig. 2 Classification of different Machine Learning Algorithms

II. PROBLEM STATEMENT

The increasing frequency of insider threat incidents poses a significant cybersecurity challenge. Traditional methods struggle to detect these threats originating from within the organization, which can encompass various activities. Our research aims to develop a robust solution for in sider threat detection using unsupervised machine learning techniques. The core problem we address is enhancing organizational security by offering an innovative defense against insider threats.

III. LITERATURE SURVEY

Paper Name: Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence

Author: Vasileios Mavroeidis

Abstract: Threat intelligence is the provision of evidence-based knowledge about existing or potential threats. Benefits of threat intelligence include improved efficiency and effectiveness in security operations in terms of detective and preventive capabilities. Successful threat intelligence within the cyber domain demands a knowledge base of threat information and an expressive way to represent this knowledge. This purpose is served by the use of taxonomies, sharing standards, and ontologies. This paper introduces the Cyber Threat Intelligence (CTI) model, which enables cyber defenders to explore their threat intelligence capabilities and under- stand their position against the ever-changing cyber threat landscape. In addition, we use our model to analyze and evaluate several existing taxonomies, sharing standards, and ontologies relevant to cyber threat intelligence. Our results show that the cyber security community lacks an ontology covering the complete spectrum of threat intelligence. To conclude, we argue the importance of developing a multilayered cyber threat intelligence ontology based on the CTI model and the steps should be taken under consideration, which are the foundation of our future work.

Paper Name: Probabilistic Threat Detection for Risk Management in Cyber-physical

Medical Systems

Author: Aakarsh Rao, Nadir Carreo´n, Roman Lysecky, and Jerzy Rozenblit,

Abstract:

Medical devices are complex cyber-physical systems exposed to numeroussecurity risks and vulnerabilities. This article presents a dynamic risk management and automated threat mitigation approach based on a probabilistic threat estimation framework. A smart- connected pacemaker case study illustrates the approach vulnerabilities. This article presents a dynamic risk management and automated threat mitigation approach based on a probabilistic threat estimation framework. A smart- connected pacemaker case study illustrates the approach.

Paper Name: Managing cyber threat intelligence in a graph database

Author name : Seulgi Lee, Hyeisun Cho, Nakhyun Kim, Byungik Kim, Junhyung Park

Abstract : Efforts to cope jointly with the ever-increasing number of breach inci- dents have resulted in the establishment of the standard format and protocol and given birth tomany consultative groups. In addition, various channels that distribute Cyber Threat Intelligence information free of charge have emerged, and studies on utilizing such channels have spread. As the market for sharing information professionally is expanding, the need to manage the shared information in various ways in order to achieve better resulthas arisen. This paper proposes a standardized management structure and method based on the standardized format and a meaning and standard of Cyber Threat Intelligence thatcan be shared outside when loading OS- INT information collected from various channels into the graph database. This pa- per also proposes a method of supporting the detection provided by existing security equipment with the information saved in the graph database and an effective method of analysis. Lastly, the paper discusses the advantages that can beexpected from saving cyber threat information in the graph database developed usinginformation collected from the outside.

Paper Name: Optimized Reactive Power Flow of DFIG Power Converters for Bet- ter

Reliability Performance Considering Grid Codes

Author: Dao Zhou

Abstract : If there is no reactive power exchange between a doubly fed induction generator (DFIG) and a grid, the various characteristics of the power converters in a DFIG wind turbine system cause the lifetime expectancy of a rotor-side converter (RSC) to be significantly less than that of a grid-side converter (GSC). In order to fulfill modern grid codes, over-excited reactive power injection will further reduce the lifetime of the RSC. In this paper, the additional stress of a power semiconductor due to the reactive power injection is first evaluated in terms of a modulation index and the current loading. Then, an optimized reactive power flow is proposed in the case where an over-excited reactive power support is applied with the joint compensation from both the RSC and the GSC. Finally, some experimental validations are performed at a downscale DFIG prototype. It is concluded that, among the different combined reactive power support strategies, the best scheme will trade off the life- time between the GSC and the RSC

Paper Name: A Design of IL-CyTIS for Automated Cyber Threat Detection

Author:Joseph Yoo

Abstract : As cyber squabbling has been intensified, the necessity of sharing cyber threat information has increased. Therefore, attempts to develop a technology to up- grade and deepen the related system will continue. In particular, it is anticipated that automated response and analysis using machine learning will be actively conducted. In this paper, we design and propose IL-CyTIS (a unified and lightened information structure) by customizing STIX (a cyber threat information expression standard) for the input and analysisvia machine learning before conducting a study using machine learning to derive new information from existing cyber threat information. Then, we discuss its actual application in machine learning

Paper Name: A Study on a Cyber Threat Intelligence Analysis (CTI) Platform for the

Proactive Detection of Cyber Attacks Based on Automated Analysis

Author name: Byung Ik Kim, Nakhyun Kim, Seulgi Lee, Hyeisun Cho, Junhyung Park

Abstract: This paper proposes an automated cyberattack analysis platform that is designed to analyze and respond to cyberattacks, which are becoming ever more intelligent and advanced. The ICT information generated during previous cyberat- tacks will be collected.

to analyze cyberattacks automatically, and the relationship between the collected information, level of re-exploitation, and similar ICT infor- mation among cyberattacks will be automatically analyzed. If the values that are currently being monitored are entered

into the developed platform, the most similar cyberattacks in the past and the current phase of attacks will be provided to the analyst. In addition, a system capable of blocking attacks in advance before damages are caused could be developed by providing response/analysis guideline information on the potential future attack inflow.highest productivity, followed by the remaining crops in the list [8].

M. R. Bendre et al in his study, historical data, including temperature and rainfall, is fed into the processing

IV. METHODS AND MATERIAL

1. Data Pre processing:

The first crucial step in our system analysis is data pre processing. The data acquired for insider threat detection often necessitates extensive cleaning and transformation to ensure its usability. This phase comprises the following elements:

Data Cleaning: Removal of outliers, handling missing values, and addressing data inconsistencies.

Feature Engineering: Extracting and transforming relevant features from raw data to enable effective modeling.

Normalization: Scaling data to ensure consistent units and facilitate model convergence.

2. Machine Learning Models:

Our system analysis revolves around the machine learning models that form the core of our insider threat detection system:

Tree Structure Method: This innovative methodology, inspired by [Author, Year], is implemented to dissect and analyze user behavior. It structures user actions hierarchically, enabling the identification of temporal order, dependencies, and relationships within these actions. It forms the foundation for feature sequence formation.

Copula-Based Outlier Detection (COPOD): COPOD, is the second integral component. It utilizes advanced statistical techniques, specifically copulas, to estimate the joint distribution of features within the feature sequences. This statistical modeling aids in identifying anomalies and deviations from established patterns, enabling the detection of abnormal user behavior indicative of insider threats.

3. Model Development:

The system analysis extends to the development of our insider threat detection model. This phase comprises the implementation of the tree structure method and COPOD. These methods are integrated into a cohesive system that is capable of analyzing user behavior and detecting anomalies within feature sequences.

4. Model Evaluation:

Our insider threat detection system's performance is rigorously evaluated to assess its effectiveness in identifying and mitigating insider threats. The evaluation phase includes:

Experimentation: Conducting experiments on the selected insider threat dataset (e.g., CERT-IT) using the developed model.

Performance Metrics: Evaluation criteria encompassing accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). These metrics provide a comprehensive assessment of the system's ability to detect insider threats.

5. Performance Comparison

A critical aspect of our system analysis is the comparison of our developed model with traditional insider threat detection methods. The Isolation Forest, in particular, is considered a benchmark for comparison. This step ensures that our model offers superior performance and validates its effectiveness.

6. Result Analysis:

The result analysis phase examines the experimental outcomes and provides insights into the performance of our insider threat detection system. The strengths and weaknesses of the model are assessed, and any noteworthy findings are discussed in detail.

V. CHALLENGES FACED IMPLEMENTING THE TECHNIQUES

This part discusses the current challenges in the detection of insider threats;

1. Performance

As an attacker is a legitimate user of the system, this notion poses the difficulty of drawing a clear line between what is legitimate and what is malicious. Most of the existing approaches used for insider threat detection apply the anomaly detection approach, which is supervised or unsupervised methods that classify small deviations from normal activity patterns in anomaly detection as an abnormality and, thus, classify this abnormality as malicious. However, most of these abnormalities are non malicious activities. These methods tend to raise unnecessary false alarms in handling such cases. Thus, traditional approaches suffer from the well-known issue of false positives due to this notion, which makes such approaches difficult to apply in enterprise environments. In other words, reducing false positive and negative alarms for insider threat detections without affecting the detection accuracy remains a major challenge.

2. Insider Threat Datasets

Lack of Real Data

In spite of advanced research on insider threats, challenges in validating and refining the detection models remain due to the absence of real-world data from organizations. The lack of actual insider threat data is also a major challenge in assessing and developing insider threat detection systems. Moreover, the present review observes that synthetically created datasets used in the surveyed articles were not created specifically for insider threats. Furthermore, a few of these datasets did not contain malicious data, whereas others were outdated.

Ethical and Privacy Issues

Despite the increase in the number of insider threat incidents, not all organizations report such incidents nor allow access to their data, typically due to ethical and privacy concerns. The issue of real data access is crucial for insider detection, which continues to be a significant obstacle for validating and refining effective and scalable detection systems. As a result, most existing detection systems are tested and evaluated on synthetic and simulated datasets, with the biases that such data imply.

Analysis Issues on Encrypted Flows or Encrypted Data Packets

To avoid detection by tools, such as intrusion detection systems, attackers may use cryptography to mask their attacks. Such a scenario renders detection systems unable to analyze encrypted flows or encrypted data packets, which is another main limitation of the current intrusion detection systems

3. High Dimensionality

Sheer Volume

The capabilities of capturing logs for the activities are an advantage that may provide insight into employee actions. Despite this advantage, the analysis of activity logs continues to be difficult for analysts because of the sheer volume of activities that employees produce every day. The large number of organizational staff requires the monitoring of staff behavior properties, which results in the massive need for data to be processed. The growth of this data outpaces the ability of human auditors and administrators to digest such data quantities using manual analyses

4. Physical and Cyber Behavior

Another limitation of the current insider threat detection approaches is that they only concentrate on cyber or physical security behaviors within cybersecurity. Most of the previous works did not use both behaviors of the cyber and physical systems in analyzing insider threat detections. The majority of scholars aim to detect insiders by observing behaviors either from the cybersecurity or physical security aspects [86]. However, in terms of detecting physical threats, most of the existing studies applied physical access control mechanisms that may control the physical access of unauthorized users to a certain point. However, such mechanisms are ineffective against insider attacks.

VI.CONCLUSION

In conclusion, this research has made significant strides in addressing the pressing issue of insider threat detection within organizations. The rising frequency of insider threat incidents poses substantial challenges to the security and integrity of companies and institutions.
Traditional intrusion detection systems, primarily designed to identify external threats, often fall short in identifying and mitigating malicious behaviors originating from within the organization.

Our approach, which combines the tree structure method for user behavior analysis with the Copula-Based Outlier Detection (COPOD) method, offers a promising solution to this multifaceted problem. By organizing user actions hierarchically and capturing the temporal order of these actions, our method creates feature sequences that serve as the basis for identifying abnormal user behavior.

VII. REFERENCES

[1]. C. R. Srinivasan, B. Rajesh, P. Saikalyan, K. Premsagar, and E. S. Yadav, “A reviewon the different types of Internet of Things (IoT),” J. Adv. Res. Dyn. Control Syst., vol. 11, no. 1, pp. 154–158, 2019.

[2]. G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, “Internet of Medical Things (IOMT): Applications, benefits and future challenges in healthcare do- main,” J. Commun., vol. 12, no. 4, pp. 240–247, 2017.

[3]. A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of Things for smart cities,” IEEE Internet Things J., vol. 1, no. 1, pp. 22–32, Feb. 2014.

[4]. E. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, “Android malware de tection using deep learning on API method sequences,” Dec. 2017, arXiv:1712.08996.[Online]. Available: https://arxiv.org/abs/1712.08996

[5]. S. Jabbar, K. R. Malik, M. Ahmad, O. Aldabbas, M. Asif, S. Khalid, K. Han, and S. H. Ahmed, “A methodology of real-time data fusion for localized big data analytics,” IEEE Access, vol. 6, pp. 24510–24520, 2018.

[6]. F. Ullah, J. Wang, M. Farhan, M. Habib, and S. Khalid, “Software plagiarismdetection in multiprogramming languages using machine learning approach,” Concurrency Comput., Pract. Exper., to be published.

[7]. D. K. Chae, J. Ha, S.-W. Kim, B. Kang, and E. G. Im, “Software plagiarism detection: A graph-based approach,” in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., Nov. 2013, pp. 1577–1580.

[8]. Y. Akbulut and O. D o¨ nmez, “Predictors of digital piracy among Turkish un- dergraduate students,” Telematics Inform., vol. 35, no. 5, pp. 1324– 1334, 2018

[9]. M. ShanmughaSundaram and S. Subramani, “A measurement of similarity to identify identical code clones,” Int. Arab J. Inf. Technol., vol. 12, pp. 735–740,Dec. 2015.

[10]. C. Ragkhitwetsagul, “Measuring code similarity in large-scaled code Corpora,” in Proc. IEEE Int. Conf. Softw. Maintenance Evol. (ICSME), Oct. 2016, pp. 626–630. Anderson, R. (2001). Security Engineering: A Guide to Building Dependable

[11]. Carcary, M., and Power, J. (2011). An Investigation into the Development of a Framework for Insider Threat Mitigation. Journal of Computer Information Systems, 51(1), 29-40.

[12]. CERT-IT. (Year). The CERT-IT Insider Threat Dataset.

[13]. Fidler, D., and Fidler, J. (2009). Applying Copulas to Examine Dependencies in Financial Data. Journal of Risk Management, 6(4), 1-29.

[14]. Friedman, R., and Smid, M. (2015). Copula-Based Anomaly Detection. Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), 9(12), 9-12.

[15]. Ghosh, A. K., and Schwartzbard, A. (2011). A Survey of Insider Threat Detection: Fundamentals, Recent Advances, and Challenges. ACM Computing Surveys, 43(4), Article 15.

[16]. Hodge, J. J., and Austin, J. (2004). A Survey of Outlier Detection Methodologies. Artificial Intelligence Review, 22(2), 85-126.

[17]. Roesch, M. (1999). Snort - Lightweight Intrusion Detection for Networks. Proceedings of the USENIX Annual Technical Conference.

[18]. Tan, C. L. (2019). Introduction to Data Mining. Pearson.

[19]. Thomas, R. K., and Han, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.

[20]. Tondel, P., Kristoffersen, A. L., and Van Helvoort, J. (2009). A Model-Based Method for Anomaly Detection in Transaction Data. International Journal of Information Management, 29(4), 295-302.

Cite this article as :

Tushar Yadav, Akash Manoj, Tanishq Ghanshani, Wajahat Hussain, Dr. Kapil Vhatkar, "Insider Threat Detection Using Unsupervised Learning", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 6, pp. 361-367, November-December 2023. Available at doi : https://doi.org/10.32628/IJSRSET2310647
Journal URL : https://ijsrset.com/IJSRSET2310647