IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

THREAT HUNTING : THE PROACTIVE CYBER DEFENCE

[1]Nikhil G V, [2]Niruv Bhardwaj, [3]Neha Afreen, [4]Nisarga K N,[5]Veda N
[1]Undergraduate Student, CSE
[1]Sri Venkateshwara College of Engineering, Bengaluru, India

Abstract: This paper presentation explores the imperative practice of threat hunting in the realm of cybersecurity. In an era marked by escalating cyber threats, the paper delves into the methodologies and strategies employed in proactive threat detection and mitigation. Through a synthesis of realworld case studies and advanced threat intelligence analysis, the authors showcase the significance of threat hunting in identifying and neutralizing potential security risks before they manifest. The presentation emphasizes the proactive stance of threat hunting, aiming to provide insights, methodologies, and practical approaches for enhancing cyber resilience. The findings serve as a valuable resource for cybersecurity professionals offering a roadmap to navigate the evolving landscape of digital threats.
Keywords: Cybersecurity, Mitigation, Proactive stance, Neutralizing

INTRODUCTION:

Threat Hunting is a proactive cybersecurity measure aimed at identifying and mitigating threats that may have evaded traditional security measures. It involves continuously analyzing data to detect any abnormal behavior or potential security breaches within an organization's network. By actively searching for indicators of compromise, threat hunters can identify and neutralize threats before damage occurs. It requires a combination of advanced tools, techniques, and expertise to effectively uncover and respond to cyber threats.

Proactive Defence: Threat Hunting allows organizations to take a proactive stance against cyber threats, reducing the dwell time of potential threats and minimizing the impact of security breaches.
Risk Mitigation : By continuously hunting for threats, organizations can minimize the risks associated with undetected malicious activities, thereby safeguarding sensitive data and critical systems.
Enhanced Detection : It enhances the detection capabilities beyond what automated security measures can achieve, enabling the identification of sophisticated and evasive threats.

LITERATURE REVIEW:

According to the literature review done, there are various methods for the identification of threats. By synthesizing existing knowledge this review aims to contribute to a deeper understanding of the methodologies, strategies, challenges, and benefits associated with proactive threat detection.

Nataliia Lukova-Chuiko [1] The first paper discusses about the Order of the Threat Hunting Conduct, Which states that we can use four stages of hunting to counter and neutralize the threat which namely are Hypothesis Creation, Research with tools, Identification of tactics and Expansion of Analytics.

Zichen Wang [2] in this paper the threat hunting encountering has been conducted in various stages which are being discussed in brief here. The first step taken is Data identification which identifies the type of data which falls under the circumstances of being hunted. The second step is the Data Extraction process. Once the Data has been identified, it needs to be extracted. The last and the final step involves the Data Analysis which deals with the analyzation of data.
Mario Aragone Lozano [3] this paper deals with the threat hunting architecture using Machine Learning for Critical Protection. Firstly the data is collected and analyzed in which the trends of attacks is studied which is used as the Database for algorithmic generations. Using the algorithm the hypothesis are generated which undergoes ML filtration which eventually provides us with a solution for the hunt.

Gautam Srivastava [4] This paper deals strictly with threat hunting methodologies used for edge devices. The Relevant Single Classifiers uses Decision Tree, Supports Vector Machine, Logistic Regression and Random Forest. The formula used for Information Gain in DT Model is mentioned below :

IG (Mp,z) = I(Mp) - Li/Mp I(Lp) - Ri/Mp (I(Rp))
In the above stated formula Mp, Ri and Li represent the data stored in a given parent node, its right and its left child. Lp and Rp are the numbers samples present in the left and right child.

Israel Perez Llopez [5] in this paper they talk about Prototype Evaluation insight of Industrial Network, Critical Infrastructure Network, DMZ which consists of different servers namely Apache, NGINX, DNS and the Exchange Servers. The another step involves the Prototype Improvements which implements the developed algorithms using different languages such as Python and C++. The C++ libraries are mostly used for this purpose.

Abbas Yazdinejad [6] In this paper, an ensemble deep RNN model for cyber threat hunting in IoT that overcomes issues related to the imbalanced datasets, long term time dependency, and big data in existing approaches is presented. The cyber threat hunting in IoT is addressed in the proposed model with deep anomaly detection and deep anomaly prediction methods.

Fengyu Yang, Yanni Han, Ying Ding, Qian Tan, Zhen Xu [7] The article discusses the complexity and difficulty of detecting advanced threats in enterprise networks due to the variety of attacking means. It proposes to construct a knowledge graph based on kernel audit records to quickly and effectively find information related to attack events from massive data streams. The knowledge graph can be used for automatic relation calculation and complex relation analysis. The article also explores different ways to use the constructed knowledge graph for hunting actual threats in detail and introduces a LAN-wide cybersecurity hunting system. Attackers constantly change their attacking means, making it difficult for auto-sis. Raw kernel audit logs have fatal flaws, increasing the difficulty of automated threat detection. Many enterprises employ professional security teams to detect potential threats in their systems, but threat hunting becomes complicated and difficult.

Akinsola [8] The dynamic environment of cyber security requires organizations to stay ahead of new trends of cyber threats. Threat hunting has evolved with the present wave of cyber threat hunters. Machine learning (ML) is an important technique in the usage of data as well as huge technology for various fields. The aim of this study is to proffer solution to the problem of cyber attack and to reduce the rate by which cyber criminals gain.

Jiawei Li, Ru Zhang, Jianyi Liu and Gongshen Liu [9] in this paper The article provides a 2D visualization of mutual reachability distance and outliers indicate threat BPGs for checking mail behaviors. The clustering result is shown in Table 4, with Min Distance representing minimum distance from other clusters and Number of Graphs indicating the number of graphs. The behavior provenance graphs of two scenarios are similar, but some mail-related processes are divided into multiple clusters due to differences in subsequent operations caused by different attachment types. The attack scenarios in the malicious dataset include OceanLotus, APT28, and Kimsuky attacks, each with different key nodes and operations. The attacks involve phishing emails, process hollowing, process injection, and sending collected information over FTP to remote servers. The article also includes two homologous cyber weapons with no initial intrusion and delivery phases and sees what happens when logs are incomplete.

Z. Jadedi and Y. Lu [10] in this paper we discussed the need for a unified hunting solution for Industrial Control System (ICS) networks to prevent targeted attacks. The ICS Threat Hunting Framework (ICS-THF) is proposed to detect cyber threats against ICS devices in the earliest phases of the attack lifecycle. The framework consists of three stages and uses a combination of the MITRE ATT&CK Matrix and a Diamond model of intrusion analysis to generate a hunting hypothesis. The goal of the hunting process is to detect threat actors early in the cyber kill chain by searching for signs of an intrusion and then, providing detection strategies for future use. The article has been accepted for publication in a future issue of this journal, but content may change prior to final publication.

Initialization : Select the number of clusters (K) to divide the dataset into randomly initialized K cluster centroids Additionally, the sustainable aspect of IoT-based environmental monitoring lies in its ability to optimize resource utilization and reduce environmental impact. By continuously monitoring environmental parameters, organizations can identify inefficiencies and implement targeted interventions to optimize energy consumption, reduce waste, and minimize pollution. Furthermore, the data-driven insights obtained from IoT-enabled monitoring systems can inform policy decisions and drive initiatives aimed at promoting sustainable practices and environmental conservation. Through the integration of IoT technologies, environmental monitoring becomes not only more efficient and effective but also contributes to the broader goal of building resilient and sustainable communities in the face of environmental challenges.
Assignment: Assigns each data point to the nearest centroid based on a distance metric, usually Euclidean distance and form K clusters by grouping points assigned to the same centroid.
Update Centroids: Recalculate the centroids of each cluster by taking the mean of all points in the cluster. The centroids represent the center of each cluster.
Repetition.: The CNN's output feature maps are divided into a grid. Each cell in the grid predicts object confidence and bounding boxes for the anchor boxes that best fit objects in that cell..

APPLICATIONS :

Feature Selection : Features relevant to cybersecurity, such as network traffic patterns, user behavior, or system activities, are selected for analysis.
Data Preparation : Security logs, network traffic data, or other relevant data sources are preprocessed and transformed into a format suitable for clustering.
Cluster Interpretation : The resulting clusters represent groups of data points with similar characteristics. Security analysts examine these clusters to identify patterns and anomalies.

Anomaly Detection : Data points that do not conform to the patterns of their assigned clusters are considered anomalies. Analysts investigate these anomalies as potential security threats.

ADVANTAGES :

Unsupervised Learning : K-Means is an unsupervised learning algorithm, making it suitable for threat hunting where the types of threats are not predefined.
Scalability : K-Means can handle large datasets efficiently, allowing it to scale to the extensive logs and data generated in cybersecurity environments.
Interpretability : The clusters generated by K-Means provide interpretable results, aiding analysts in understanding the structure of the data.

CONSIDERATIONS :

Initial Centroid Sensitivity : Results may vary based on the initial random selection of centroids, and multiple runs with different initializations may be needed.
Assumption of Circular Clusters : K-Means assumes that clusters are spherical or circular, which may not always reflect the true nature of cybersecurity data.
Manual Inspection Required : Analysts need to manually inspect clusters and anomalies, requiring domain expertise for accurate threat identification. The K-Means Clustering Algorithm is just one of many algorithms used in threat hunting. Depending on the nature of the data and the specific threat scenarios, other algorithms such as DBSCAN, hierarchical clustering, or machine learning approaches may also be employed.

EXPERIMENT AND RESULTS :

In the stated experiment we are Identifying anomalous patterns in network traffic that may indicate a security threat.
Data :
Features : IP addresses, port numbers, data transfer volumes, connection durations, etc.
Dataset Size : Large volume of network traffic data collected over a specific time period.

Experiment Steps :

Data Preprocessing : Collect and preprocess network traffic data, extracting relevant features and handling missing or inconsistent values.
Feature Engineering : Standardize numerical features and encode categorical features if necessary. Normalize the data to ensure all features contribute equally to the clustering process.
Determining K : Experiment with different values of K to find the optimal number of clusters. Use Techniques like the elbow method or silhouette analysis to guide the selection.
Applying K-Means : Apply the K-Means Clustering Algorithm to the preprocessed network traffic data, iterate until convergence, assigning each data point to a cluster.
Cluster Interpretations : Analyze the resulting clusters to understand the patterns within the network traffic. Each cluster represents a group of network activities with similar characteristics.
Anomaly Detection : Identify clusters that deviate significantly from the expected normal patterns. Data points in these clusters may be considered anomalies.
Integration with Threat Intelligence : Correlate clusters with threat intelligence feeds to identify known malicious IP addresses, suspicious port numbers, or other indicators of compromise.
Behavioral Analysis : Understand the typical behavior represented by each cluster. Deviations from expected behavior may indicate potential security threats.
Incident Response : Investigate anomalies identified by the clustering algorithm. Initiate incident response actions for confirmed security threats.
Refinement and Iteration : Refine the analysis by adjusting parameters, experimenting with different features, or re-running the algorithm. Aim to improve the accuracy of threat detection and reduce false positives.
Documentation and Reporting : Document findings, actions taken, and lessons learned. Generate a report summarizing the identified threats, false positives, and recommendations for enhancing threat detection.

RESULTS :

Identified Clusters : Clear clusters representing normal network activities.
Anomalies : Anomalies detected in clusters with deviations from expected patterns.
Threat Correlation : Clusters correlated with known threats from threat intelligence.
Behavioral Insights : Improved understanding of typical network behaviors aiding in future threat hunting efforts.

REFERENCES :

[1] Natalia Lukova Chukka, Andriy Fesenko, Hanna Papirna, Sergiy Gnatyuk, Threat Hunting as a Method of Protection Against Cyber Threats, Taras Shevchenko National University
[2] Zichen Wang, A Systematic Literature Review on Cyber Threat Hunting, University of Guelph.
[3] Mario Aragones Lozano, Israel Perez Llopis and Manuel Esteve Domingo, Threat Hunting Architecture Using a Machine Learning Approach for Critical Infrastructures Protection
4] Abbas Yazdinejad, Behrouz Zolfaghari, Ali Dehghantanha, Hadis Karimipour, Gautam Srivastava, Reza M Parizi, Accurate Threat Hunting in Industrial Internet of Things Edge Devices
[5] Mario Aragones Lozano, Israel Perez Llopis and Manuel Esteve Domingo, Threat Hunting System for Protecting Critical Infrastructure Using a Machine Learning Approach
[6] Abbas Yazdinejad, Mostafa Kazem, Reza M Parizi, Ali Dehghantanha, Hadis Karimipour, An ensemble deep learning model for cyber threat hunting in industrial internet of things
[7] Fengyu Yang, Yanni Han, Ying Dong, Qian Tan and Zhen Xu, A flexible approach for cyber threat hunting based on kernel audit records
[8] Akinsola, Olajubu, Aderounmu, Development of Threat Hunting Model Using Machine Learning Algorithms for Cyber Attacks Mitigation, University Ile-Ife, Nigeria
[9] Jiawei Li, Ru Zhang, Jianyi Liu and Gongshen Liu, LogKernel : A Threat Hunting Approach Based on Behavior Provenance Graph and Graph Kernel Clustering, Beijing University, Beijing, China
[10] Z. Jaded and Y. Lu, Queensland University of Technology, Queensland, Australia