2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA) | 979-8-3503-4922-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIMLA59606.2024.10531322

A NOVEL APPROACH FOR DETECTING ADVANCED PERSISTENT THREATS

1st Ms.Geetha S
Department of Information Technology
K S Rangasamy College of Technology
Tiruchengode, Tamilnadu
geethas@ksrct.ac.in

2nd Kannan S
Department of Information Technology
K S Rangasamy College of Technology
Tiruchengode, Tamilnadu
kannanabd1718@gmail.com

3rd Karthicknathan S
Department of Information Technology
K S Rangasamy College of Technology
Tiruchengode, Tamilnadu
skarthicknathan2003@gmail.com

4th Mahendran A
Department of Information Technology
K S Rangasamy College of Technology
Tiruchengode, Tamilnadu
mahemano308@gmail.com

***Abstract*— An innovative solution called the Homomorphic Linear Authenticator (HLA) has been proposed to enhance the accuracy of detecting data loss, while also prioritizing data privacy and resource efficiency. By leveraging correlations among lost data, the HLA improves detection capabilities and ensures resistance against collusion. This architecture is particularly suitable for wireless devices with limited resources, as it minimizes the need for excessive communication and storage. Moreover, the system incorporates a data-block-based mechanism that allows for a balance between detection accuracy and computational complexity. Importantly, the system guarantees that auditing information cannot compromise the confidentiality of the data, thereby safeguarding data privacy. In summary, this comprehensive system offers a secure and robust solution for detecting data loss in various scenarios, including low-cost cloud sensors and resource-constrained wireless devices.**

Keywords—APT (Advanced Persistent Threats), Cyber Attacks,HLA (Homomorphic Linear Authenticator)

I. INTRODUCTION

The cybersecurity sector is currently confronted with a significant obstacle in identifying and countering Advanced Persistent Threats (APTs). These APTs are orchestrated by highly skilled and well-funded threat actors who aim to infiltrate and compromise targeted organizations. APTs are sophisticated and operate covertly, making them challenging to detect using conventional security measures. To tackle this issue, there is an urgent requirement for innovative and advanced methods that can identify APTs at an early stage and strengthen network defences. This research paper proposes a unique approach that combines state-of-the-art technologies, behavioural analysis, and the integration of threat intelligence to enhance the capability of detecting APTs. By harnessing the power of artificial intelligence and machine learning algorithms, this approach can analyse network traffic and system behaviours to identify abnormal patterns that indicate APT activities. With this proactive defence mechanism, organizations can minimize the potential damage caused by these malicious threats. APTs are differentiable from other cyber threats by their sustained persistence, sophisticated tactics, and targeted, high-value asset-specific targeting. To differentiate themselves from more frequent, transient threats, they deliberately target important assets within companies over protracted periods of time.

II. LITERATURE REVIEW

A. A REVIEW OF APT ATTACK DETECTION METHODS AND DEFENSE STRATEGIES

Kai Xing College of Compute et.al. has presented in this paper the proposal that since its establishment, Cyberspace has been continuously under threat from attacks. Technological solutions and new types of cyberattacks are continuously developing as a result of the Internet's and artificial intelligence's rapid advancements. Notably, advanced persistent threats (APTs) are increasing in intensity. The focus has shifted towards effectively preventing this type of attack, leading to significant advancements in attack detection and defense technology. This paper primarily focuses on the progress being made in science about APT attack detection and defense strategies, both nationally and globally. The use of machine learning in conjunction with traditional attack detection methods is examined. The defensive strategy, which ascertains the optimal defense plan given resource limits, mostly consists of cloud platforms, game theory, and dynamic information flow tracking. APT attacks have not gone unnoticed, even with the extensive integration of artificial intelligence technology across numerous businesses. The long-term goal of APT attacks is to stealthily penetrate valuable targets. The development of clever attack strategies by attackers through artificial intelligence technologies makes it more difficult to identify and fight against advanced persistent threats (APT) attacks. APT attacks exhibit heterogeneity, leading many scholars to propose diverse detection and defense approaches tailored to specific scenarios. Although the manifestations of these attacks are constantly changing, they still follow certain patterns. More creative and practical approaches are available thanks to the notable and quick advancements in game theory and machine learning. The primary crucial stage in an APT attack involves gaining entry. Attackers frequently gather a lot of information before launching an incursion by using social engineering and open source intelligence methods.

B. ADVANCED PERSISTENT THREATS CAMPAIGNS AND ATTRIBUTION

The aim of this study, proposed by Pedro Ramos Brandao et.al., is to conduct a comprehensive review of existing literature on Advanced Persistent Threats (A.P.T.) and A.P.T. Campaigns, with a specific focus on campaigns originating from China. The study examines the key A.P.T. campaigns from this region, utilizing various types of documentation, including gray literature from official and government agencies. The objective is to demonstrate the potential for Attribution in relation to specific Groups in China, who have targeted multiple western countries through APT. The challenge at hand is to Identify these Groups and determine the authors behind the APT attacks. The scope of the research is specifically centered around APT attacks and their potential

origin in China. As the use of digital documents continues to rise, cybersecurity has become increasingly vital in safeguarding information and systems against various threats, such as theft, terrorism, and cybercrime. Despite the implementation of antivirus software, firewalls, and intrusion detectors, incidents of theft and exposure of sensitive information persist, resulting in significant financial and reputational harm to banks, companies, and governments. The failure of security systems raises several questions: What are the underlying causes of their failure? How are adversaries managing to surpass them? It is now evident that attackers not only possess advanced expertise in this field but also employ highly sophisticated tools to accomplish their objectives. While information theft and unavailability are commonly discussed as the primary threats, this study emphasizes the importance of attacks targeting critical infrastructures.

C. *EFFICIENT DETECTION OF HACKER COMMUNITY BASED ON TWITTER DATA USING COMPLEX NETWORKS AND MACHINE LEARNING ALGORITHM*

Ahmed Al-Tarawneh and his colleagues have proposed a system that utilizes Twitter as a popular platform for sharing and posting ideas. However, this platform is also exploited by hackers and anonymous attackers for malicious purposes. To enhance the prediction of future attacks, machine-learning techniques are employed to gather and categorize hackers' tweets. Prior techniques for identifying compromised tweets were restricted to either enumerating the concealed text inside tweet lines or depending on human scrutiny. The objective of this study is to improve the effectiveness of Twitter hacker identification through the use of customized machine learning algorithms and intricate networks. Based on the terms that hackers frequently use in their tweets, the method entails creating a list of persons and their followers who are active on Twitter and have similar interests. Based on network centrality, proximity, and betweenness, a complex network is subsequently created to determine the relationships between people. It is possible to identify the most prominent members of the hacker community and then gather and categorize their tweets into good and bad groups.

D. *ANOMALY DETECTION IN LOG DATA USING GRAPH DATABASES AND MACHINE LEARNING TO DEFEND ADVANCED PERSISTENT THREATS*

According to Timo Schindler and colleagues, there is a serious risk to computer network cybersecurity from Advanced Persistent Threats (APTs). Based on their research, it was found that successful breaches in 2015 went unnoticed for an average of 146 days. The objective of our research is to present a workable and efficient method for examining actual log data to find breaches or attempted intrusions. We are able to create flexible attack profiles by combining well-known kill chain algorithms and making use of an abstracted graph technique along with a time series database. By examining the log data of a simulated computer network, we have been able to effectively show our capacity to identify simulated attacks using this method. Furthermore, our system can quickly undertake high-performance analysis of real-world data while taking into account different sources of log data. By leveraging the computational capabilities of the graph database, we can identify the attacker and detect other affected system components. We firmly believe that this approach will significantly reduce the time required to detect breaches and enable prompt response to new attack vectors. Over the past decade, Vukalovic et al. have observed a notable increase in APTs targeting companies across various industries. This claim is supported by the Verizon Data Breach Investigations Report, which states that nearly 100% of compromises occur within a day, with only about 10% being discovered within the same timeframe. Additionally, the Fire eye report highlights that security breaches in 2015 went undetected for an average of 146 days. Furthermore, the recent security breach at ThyssenKrupp, with a 45-day gap between breach and detection, further emphasizes the significant amount of time attackers have to operate without being detected within compromised systems.

E. *DETECTION AND CLASSIFICATION OF ADVANCED PERSISTENT THREATS AND ATTACKS USING THE SUPPORT VECTOR MACHINE*

Wen-Lin Chu et al. have created a system in response to the rapid advancement of network technology that aims to manage the dynamic traditional hacking and attack models. This system primarily targets advanced persistent threats (APTs), which are extremely proficient and targeted attack methods typically orchestrated by hacker groups. Prior to launching an attack on a specific target, extensive strategic planning and information search are carried out. The objective of the system presented in this research is to detect APT attacks at an early stage. The NSL-KDD database is used by the system for attack detection and verification in order to dothis. Through feature sampling, the principal component analysis (PCA) technique is the primary means of improvingdetection efficiency. The study also compares and analyzes the advantages and disadvantages of other classifiers, such as decision trees, neural networks, naive Bayes classification, and support vector machines, in order to determine the dataset. The support vector machine (SVM), which achieved 97.22% for the training subdata A, has the highest recognition rate, according to the experiment's results. The purpose of this research is to establish an early warning model system for APT assaults that can lessen their impact. With the recent surge in Internet technology development, there has been a corresponding increase in hacker attack methods, necessitating increased attention to information security from both industry and government sectors.

III. EXISTING SYSTEM

Because of the increased reliance on modern technology and systems, there has been a noticeable increase in the attention on cyber security recently. As a result, it has become crucial to secure these systems against cyber-attacks in today's world. Among the most complex and persistent types of cyberattacks is the advanced persistent threat. Malevolent actors entering a network without authority and remaining concealed for an extended period of time are the hallmarks of this particular assault type. Attacks by advanced persistent threats are becoming more common, and with them come increasing risks to enterprises. Machine learning is one technique that has been used to identify attacks from sophisticated persistent threats. But because there aren't enough thorough datasets covering every stage of an advanced persistent threat attack's life cycle, this strategy hasn't gotten much attention in earlier research. As a result, the objective of this study is to offer a fresh dataset that covers all stages of an advanced persistent threat attack, including data exfiltration, normal activities, reconnaissance, lateral movement, and first breach. The newly acquired dataset encompasses a range of tactics, procedures, guidelines, and signs of compromise, and

it is based on real advanced persistent threat assaults. Next, this dataset is used in a machine learning model that is suggested, which makes use of the analysis of variance feature selection approach and extreme gradient boosting.

IV. Proposed System

The growing complexity and endurance of cyberattacks, together with the shortcomings of conventional detection techniques in terms of accurately identifying and countering these threats, have spurred the development of novel methodologies for detecting Advanced Persistent Threats (APTs). The proposed system aims to improve the accuracy of data loss detection by utilizing correlations within lost data. It achieves secure and truthful auditing through the use of a Homomorphic Linear Authenticator (HLA) by Virtual Machines, while also preventing collusion. This system places a high priority on data privacy and resource efficiency, reducing communication and storage overheads. As a result, it is well-suited for resource-constrained wireless devices. The incorporation of a data-block-based mechanism provides flexibility in adjusting detection accuracy at the expense of computational simplicity. When doing calculations on encrypted data, the method uses the Homomorphic Linear Authenticator (HLA) to guarantee the authenticity and integrity of the data. The system can enhance privacy and detection capabilities in a cloud-based environment by securely analyzing sensitive data for APT detection without exposing it to unwanted access by utilizing HLA. Homomorphic Linear Authenticator (HLA) is a methodology used by this new approach to detect Advanced Persistent Threats (APTs) to determine whether data has been altered. Through the examination of system logs and network traffic, HLA can identify anomalous behavior that may indicate an APT attack and assist in thwarting it before damage is done.

V. Module Description

A. Network formation

In this Module, our network is furnished with a network controller. All sensor Virtual Machines are connected to the network controller. Within the network, there is an autonomous auditor known as Ad. Ad functions independently and is not associated with any Virtual Machine in PSD. Additionally, Ad doesn't know what information the different Virtual Machines are holding, like encryption keys. When asked, the auditor's main duty is to identify harmful virtual machines. More specifically, we suppose that D notifies S if it believes that an assault is being launched against the link.

B. Data transmission

A sequence of intermediary Virtual Machines n1,..., nK facilitates the transfer of data from the source Virtual Machine S to the destination Virtual Machine D. Where i is a number between 1 and K-1, each ni denotes the upstream Virtual Machine of ni-1. It is assumed that S has knowledge of the link PSD, similar to the Dynamic Source Routing (DSR) protocol. A trace link procedure can be used by S to locate the Virtual Machines in PSD in the event that DSR is not present. Our main focus is on Cloud Computing networks that are static or quasi-static, meaning that their link properties and network architecture don't change over an extended period of time. Utilizing a variety of data sources, including threat intelligence feeds, endpoint telemetry, and network traffic records, the method works. In order to provide a thorough picture of potential threats throughout the network, these sources are combined using a centralized platform that uses sophisticated correlation and analytics algorithms to identify patterns suggestive of APT activity.

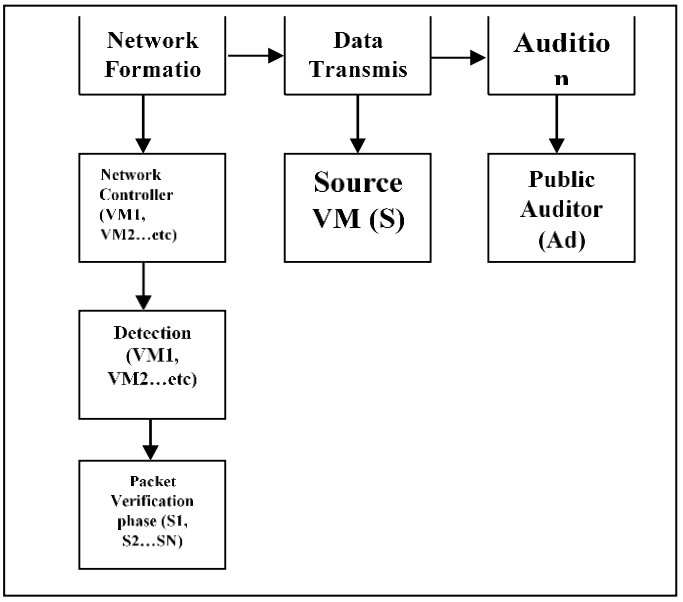
C. Audition

This process is initiated by public auditor Ad upon to receipt of an ADR communication from S. The sequence numbers from the latest M data transfer that S has sent, as well as the ADR message's sequence numbers for the portion of these M data that D received. The virtual machine ID on PSD that is downstream-ordered and written as n1;...; n K? It is imperative to stress that S and D are assuming that the information they have shared is accurate because detecting assaults is one of their top concerns.

D. Detection

In the detection stage, the public auditor (Ad) takes on significant duties, such as going over Virtual Machine responses on PSD, spotting overstated cases of data loss, making bitmaps of the data loss for each hop, assessing autocorrelation for data loss, and successfully spotting malicious activity, guaranteeing comprehensive detection of all data droppings. By maintaining data integrity and secrecy throughout the detection process, the method reduces the possibility of false positives and false negatives by using Homomorphic Linear Authenticator (HLA) to do safe computations on encrypted data. The system can reliably analyze patterns and anomalies related to APTs while lowering the possibility of false identifications by guaranteeing computations protect data privacy. When doing calculations on encrypted data, the method uses the Homomorphic Linear Authenticator (HLA) to guarantee the authenticity and integrity of the data. The system can enhance privacy and detection capabilities in a cloud-based environment by securely analyzing sensitive data for APT detection without exposing it to unwanted access by utilizing HLA.

System Flow Diagram



VI. Conclusion

The Homomorphic linear authenticator (HLA) based public auditing framework presents a practical approach to identifying malicious nodes in ad hoc networks. It successfully addresses the obstacles of improving detection accuracy, ensuring precise correlation calculation, and reducing computation and communication overheads. It can also be used to enhance the security of a variety of networks, including peer-to-peer, ad hoc, and sensor networks. This framework is adaptable enough to handle large networks. Currently, the framework is undergoing development and is

expected to be released in the near future.

VII. FUTURE WORK

More features, like node mobility patterns and dropped data packet kinds, can be added, and complex machine learning techniques can be developed to increase the system'sdetection accuracy. Furthermore, the computational burden can be reduced by enhancing homomorphic encryption schemes and optimizing the data-block-based mechanism. Detecting APTs and enabling secure computation on encrypted data while maintaining its confidentiality is achieved by integrating Homomorphic Linear Authenticator (HLA) software. The system's ability to securely analyze encrypted data streams allows it to adjust to changing APT tactics without jeopardizing sensitive data. This means that when attack techniques change, the system's detection skills remain intact. Future research may focus on developing real-time analytical tools and refining HLA algorithms, which may involve machine learning for improved detection. By enhancing threat intelligence integration, adapting to novel attack vectors, and promoting collaboration for data sharing and privacy-preserving safeguards, the approach may evolve to address emerging cybersecurity challenges.

REFERENCES

[1] A comprehensive analysis was conducted by K. Xing, A. Li, R. Jiang, and Y. Jia to examine the different techniques and strategies utilized for detecting and defending against Advanced Persistent Threat (APT) attacks. The results of their research were presented in the proceedings of the 2020 conference.

[2] In the year 2020, Steffens T. explored the attribution of Advanced Persistent Threats (APTs) and presented their findings.

[3] Al-Tarawneh and Al-Saraireh (2021) proposed an effective approach that leverages Twitter data, complex networks, and machine learning algorithms to identify hacker communities.

[4] T. Schindler conducted a study titled "Anomaly Detection in Log Data using Graph Databases and Machine Learning to Defend Advanced Persistent Threats," which was published in the Lecture Notes Informatics (LNI).

[5] Chu WL, Lin CJ, and Chang KN conducted a study that focused on the identification and categorization of Advanced Persistent Threats (APTs) and attacks using the Support Vector Machine (SVM) algorithm.

[6] Ahmed Y, Asyhari AT, Rahman MA. A Cyber Kill Chain Approach for Detecting Advanced Persistent Threats. Compute Mater. Contin. 2021.

[7] Alzahrani AO, Alenazi MJF. Designing a network intrusion detection system based on machine learning for software defined networks. Future. Internet 2021.

[8] Xuan CD, Duong D, Dau HX. HFFPNN classifier: a hybrid approach for intrusion detection based OPSO and hybridization of feed forward neural network (FFNN) and probabilistic neural network (PNN). IFS 2021.

[9] M. A. Umar and C. Zhanfang, ‘‘Effects of Feature Selection and Normalization on Network Intrusion Detection,” June 2020.

[10] Leevy JL, Hancock J, Zuech R, Khoshgoftaar TM. Detecting cybersecurity attacks across different network features and learners. J. Big Data 2021.