Machine Learning Mechanisms for Cyber Security and Botnet Networks

Mr. Sagar Sharad Bhuite[1]; Mr. Pravin P Kalyankar[2]

1 Assistant Professor, Department of Computer Science & Engineering, Brahmdevdada Mane Institute of Technology Solapur, University of PAH Solapur, Solapur, India

2 Assistant Professor, Department of Computer Science & Engineering,Brahmdevdada Mane Institute of Technology, Solapur University of PAH Solapur, Solapur, India

Email: bhuite.sagar@gmail.com, rajkamal875a@gmail.com

Abstract—Machine learning is, At the forefront of technological advancement, where it evolves with minimal human intervention, new machine learning models play a vital role in safeguarding data against manipulation by hackers.. In this paper, a brief introduction and importance of ML and ML security is given to grasp, and the role it plays in safeguarding data against manipulation is crucial hackers in botnet network. We introduce a Machine learning-based approach for cyber threat hunting at the endpoint level

Keywords: Botnet Network, CyberSecurity, Machine Learning,DDoS,Intrusion Detection.
,

I. INTRODUCTION

A network of compromised computers is known as a Botnet, Zombies, or Bots. The primary motivation behind botnets is rooted in the darker corners of the internet, giving rise to a new form of crime known as cybercrime. Efficient botnet detection involves the rapid collection and management of data to build detection mechanisms. However, there are limitations, such as the slow pace of information analysis..Machine learning approach-based botnet detection randomly specifies the packet's number per flow, and the byte's number per packet thus cannot be detected.Botnet is Networks of malware infected computers which cybercriminals use to perform tasks online without the user's permissionmachine learning algorithms are used to detect threats and identify malicious activity.Botnets represent a significant menace to internet security, standing alongside malicious codes. In addition to distributed denial of service (DDoS) attacks, click fraud, phishing, malware distribution, spam emails, and the illegitimate exchange of information or materials, botnets support a wide range of criminal activities.

The objective of this research is to create an advanced machine learning model for detecting botnets, incorporating the latest emerging methodologies and examining current and historical research trends. This study introduces a thematic taxonomy to categorize botnet detection techniques and evaluates their implications and key components. Machine learning techniques are pivotal across various cybersecurity applications, enabling early detection and prediction of diverse attacks including spam classification, fraud detection, malware identification, phishing, dark web or deep web site detection, and intrusion detection. These techniques help alleviate the shortage of skilled personnel with expertise in these specialized cybercrime detection technologies

II.HOW BOTNETS PERFORM?

Botnets have emerged as one of the predominant methods for deploying malware over the last decade, infecting hundreds of millions of computers. With the expansion of botnets into new technologies such as Internet of Things (IoT) devices in various environments including homes, public spaces, and secure areas, compromised systems can endanger even more unsuspecting users. Now that we have an understanding of what a botnet is, it's essential to delve deeper into how botnets operate. Here are the steps involved in initiating a botnet attack:

1.Prepping the Botnet Army: The initial phase of creating a botnet involves infecting as many connected devices as possible to amass a sufficient number of bots for executing the attack. By leveraging the computing power of the infected devices for clandestine tasks, botnets avoid detection by device owners. However, the bandwidth fraction obtained from a single machine isn't adequate. Therefore, botnets aggregate millions of devices to orchestrate large-scale attacks. This is typically achieved by exploiting security vulnerabilities in software or websites, or through phishing emails. Botnets are frequently deployed through trojan horse viruses.

2. Establishing the Connection: Once a device is compromised, it is infected with specific malware that establishes a connection back to the central botnet server, as outlined in the previous step. This process interconnects all the devices within the botnet network, preparing them to execute the attack. A bot herder utilizes command programming to control the actions of the bot.

3.Launching the Attack: Upon infection, a bot provides access to administrative-level operations, enabling various malicious activities such as gathering and stealing user data, manipulating system data, monitoring user behavior, conducting Distributed Denial of Service (DDoS) attacks, sending spam, initiating brute force attacks, engaging in crypto mining, and more.

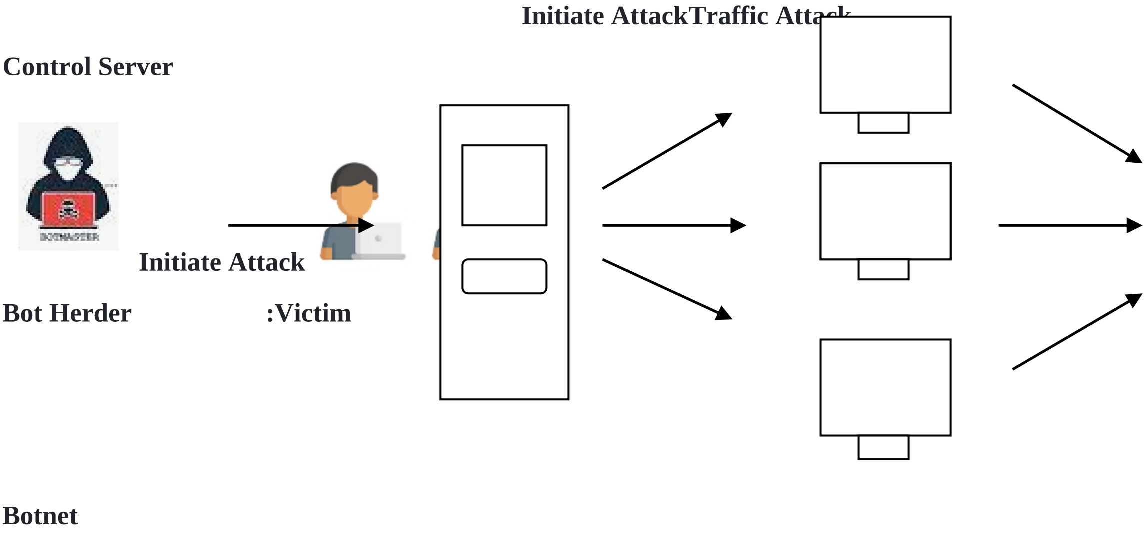


Figure1. Botnet Working

As we seen in figure1, A bot herder initiates the attack by infecting numerous devices with malicious code, effectively turning them into a botnet. Subsequently, these compromised devices orchestrate the final cyber-attack. Consequently, even if one attempts to trace the cyber-attack back to its source in such a scenario, identifying the bot herder proves challenging.

WHY CYBERCRIMINALS USE BOTNET ATTACKS:

To Steal Financial and Personal Information: Botnets are utilized by hackers to disseminate spam, phishing, or other scams with the intention of deceiving individuals into divulging sensitive financial or personal information. Additionally, they may harvest data from bot-infected devices to perpetrate identity theft and incur fraudulent charges in the victim's name.
To Attack Legitimate Web Services: Cybercriminals deploy botnets to execute Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks, inundating legitimate services or networks with an overwhelming volume of traffic. This onslaught can severely impair the service's functionality or completely incapacitate it.
To Extort Money from Victims: DoS attacks can generate revenue through extortion, wherein victims are coerced into paying to prevent their services from being disrupted or shut down entirely. Additionally, certain groups, including hacktivists and foreign military or intelligence organizations, may utilize such attacks for political or strategic objectives.
To Profit from Zombie and Botnet Systems: Cybercriminals may lease their botnets to other nefarious actors seeking to engage in activities such as spam dissemination, scam operations, phishing schemes, identity theft, and attacks on legitimate websites and networks.

In the machine learning process, training data is inputted to initiate the learning process. Machine learning algorithms utilize this training data, which may consist of known or unknown information. The efficacy of the method is influenced by the nature of the training data input, a concept that will be further elucidated shortly.

IV.ML MECHANISMS FOR CYBER SECURITY:

Technologies like machine learning and artificial intelligence allow systems to learn from data and make wise decisions without the need for explicit programming. Their use in OT and IT security is revolutionising the ways in which we identify, stop, and handle cyber threats.

1. Analysing and detecting threats: Predicting the Unpredictable: Machine learning and artificial intelligence algorithms are very good at finding patterns and irregularities in large datasets. These technologies can quickly examine user behaviour, system activity, and network traffic in the context of IT and OT security to spot possible threats or anomalies from regular operations. Rapid identification of questionable activity that conventional rule-based system might miss, allowing for faster reaction times and lowering the possibility of data breaches or interruptions to operations

2. Behaviour-Based Authentication: Revealing Undisclosed Access –Conventional authentication techniques frequently depend on set parameters such as tokens or passwords. Dynamic authentication is made possible by AI and machine learning. It

responds to user behaviour by examining interaction and access patterns to confirm identity. Enhanced security via adaptive authentication, which even when using valid credentials, recognises anomalous activity like unwanted login attempts

3. Predictive analysis, or staying ahead of the threat, is a proactive approach that enables organisations to take preventative measures against potential security breaches. It does this by mitigating risks by addressing vulnerabilities before they are exploited and by improving incident response planning based on predicted attack scenarios. AI and machine learning algorithms are capable of predicting future threats by analysing historical data, ongoing trends, and emerging attack vectors

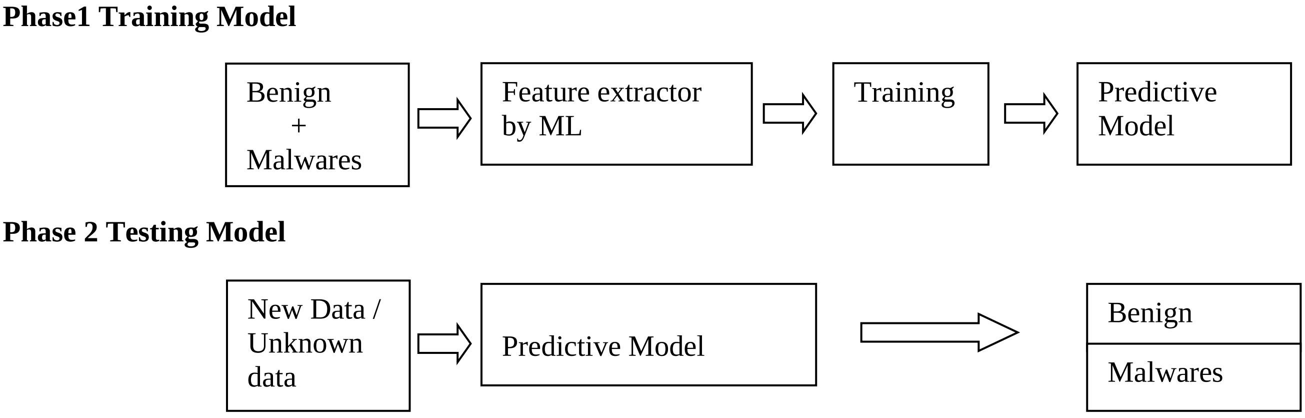
4. Automated Incident Response: Speed and Precision- AI and machine learning can automate some incident response tasks in the event of a security incident. This includes determining the extent of an attack, putting impacted systems under quarantine, and starting the cleanup process. Quicker reaction times and less need for manual intervention both of which can be vital in stopping a cyber incident from getting worse.

5. Vulnerability Management: Finding Weaknesses –By examining system configurations, code, and network traffic, AI and machine learning can assist enterprises in locating possible vulnerabilities within their IT and OT environments. Improved vulnerability identification accuracy, decreased false positives and false negatives, and allowed for more efficient security effort prioritisation.

6. Adaptive Threat Intelligence: Instantaneous Awareness- Large volumes of threat intelligence data can be processed in real-time by AI and machine learning, which enables businesses to keep up with new threats and modify their security protocols accordingly. Timely information about the changing threat landscape that enables businesses to modify their plans and defences in response to novel attack methods.

- **STRATEGY**

Data is analysed by machine learning in cyber security to determine the probability of an event. The system works by picking up knowledge from a dataset that is task-specific. The best way to complete a task is what it is supposed to do. ML will make an effort to identify the only answer that is feasible given the data at hand.Machine learning technologies excel in handling monotonous tasks. They are able to recognise and validate patterns in data. On the other hand, people must interpret the data. ML also assists in transforming the data into a comprehensible and analytical format.



Predictive Data

Figure 2 - A ML based predictive data by Phase 1 and Phase 2

Figure 2 illustrates the ML outcome of the Predictive data from Benign and Malwares. The predictive model of phase one is applied for new and unknown data in phase two. This is a mathematical process used to predict future events or outcomes by analysing patterns in a given set of input data. It is a crucial component of predictive analytics, a type of data analytics which uses new data to forecast activity, behaviour and trends.

V.ML APPROACH FOR BOTNET NETWORKS

The industry can benefit greatly from machine learning. Additionally, it helps them prepare for and face the future. However, machine learning in cyber security is not without its drawbacks. Big datasets can be quickly analysed by machine learning to identify trends and patterns. It also identifies the causal connections between various occurrences. Automation confers a significant advantage to machine learning. Additionally, very little to no human interaction is required. We enable machines to learn, and with learning, we also endow them with prediction capabilities. On their own account, they can also enhance the algorithms. Additionally, machine learning algorithms excel in handling data that is multi-dimensional and multi-variety. Even in environments that are uncertain or dynamic, this is feasible. Numerous applications also find use for machine learning. ML is applicable to cyber security and healthcare as well.

Generally speaking, machine learning is a methodological technique that uses data and algorithms to replicate human learning while progressively increasing accuracy. It automates the creation of analytical models. The loss function is a crucial element in the creation of machine learning algorithms and the improvement of their functionality. Figure 2 depicts the general architecture of a machine learning-based prediction model. The model is trained on past security data that includes both malware and benign content, and the output is produced for brand-new test data in the Testing phase. We first investigate rule-based modelling, security data clustering, and classification and regression analysis within the broad field of machine learning. In this section, we have also looked at deep learning approaches and how well they can address practical cybersecurity problems.

VI.CONCLUSIONS

Cyber security has become a matter of concern globally in achieving enhancements in security measures to detect and react against cyber attacksMachine learning techniques play a crucial role in various applications within cyber security systems.. This paper briefly presents the applications of machine learning models in the field of cyber security. There are peculiarities of each cyber threat that make it difficult even for the state-of-the-art ML model in dealing with such cyber-attacks. Accuracy of an ML model should be considered while selecting a particular model to detect a cyber-attack. We have described the basics of cyber security such as the classification of cyber-attacks on mobile deviceand computer networks. Due to the significance of ML, we have also described the foundations of machine learning, subtypes, and significant techniques for a beginner to get a better insight into this area. We are unaware of any work that discusses the applications of ML techniques in cyber security domain both on mobile device and computer networks in one paper. We have depicted a graphical summary of the attacks threatened to cyberspace and existing ML techniques to fight against these cybercrimes. We have presented an overview of several popular ML tools. We have also given the evaluation metrics to evaluate the working of any classifier. Dataset is very crucial for the training and testing of ML models.

VII.REFERENCES

1.T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, ``Machine learning and cyber security," in Machine Learning Approaches in Cyber SecurityAnalytics. Singapore: Springer, 2020, pp. 37_47.

2. Kamran Shaukat ,Suhuai Luo , Vijay Varadharajan, Ibrahim A. Hameed , And Min Xu ," A Survey on Machine Learning Techniques for Cyber Security in the Last Decade" ,IEEE Access, PP. 222310 -222355 ,VOLUME 8, 2020

3. M. Kadoguchi, S. Hayashi, M. Hashimoto, and A. Otsuka, ``Exploring the dark Web for cyber threat intelligence using machine leaning," in Proc.IEEE Int. Conf. Intell. Secur. Informat. (ISI), Jul. 2019, pp. 200_202.

4. J. Singh and M. J. Nene, ``A survey on machine learning techniques for intrusion detection systems," Int. J. Adv. Res. Comput. Commun. Eng., vol. 2, no. 11, pp. 4349_4355, 2013.

5. A. Feizollah, N. B. Anuar, R. Salleh, F. Amalina, and S. Shamshirband,"A study of machine learning classifiers for anomaly-based mobile botnetdetection," Malaysian J. Comput. Sci., vol. 26, no. 4, pp. 251_265, Dec. 2013.

6. E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid,A. O. Adetunmbi, and O. E. Ajibuwa, ``Machine learning foremail spam filtering: Review, approaches and open research problems," Heliyon, vol. 5, no. 6, Jun. 2019, Art. no. e01802

7. V. Labayen, E. Magaña, D. Morató, and M. Izal, ''Online classification of user activities using machine learning on network traffic,'' Comput. Netw., vol. 181, Nov. 2020, Art. no. 107557

8. R. Ahsan, W. Shi, and J. Corriveau, ''Network intrusion detection using machine learning approaches: Addressing data imbalance,'' IET Cyber-Phys. Syst., Theory Appl., vol. 7, no. 1, pp. 30–39, Mar. 2022

9. V. Ambalavanan, ``Cyber threats detection and mitigation using machine learning," in Handbook of Research on Machine and Deep Learning Applications for Cyber Security. Hershey, PA, USA: IGI Global, 2020, pp. 132_149

10. A. L. Buczak and E. Guven, ``A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Commun.Surveys Tuts., vol. 18, no. 2, pp. 1153_1176, 2nd Quart., 2016.

11. P. Ganapathi, ``A review of machine learning methods applied for handling zero-day attacks in the cloud environment," in Handbook ofResearch on Machine and Deep Learning Applications for Cyber Security. Hershey, PA, USA: IGI Global, 2020, pp. 364_387.
12. V. Narayan and D. Shanmugapriya, ``Big data analytics with machine learning and deep learning methods for detection of anomalies in networktraf_c," in Handbook of Research on Machine and Deep Learning Applications for Cyber Security. Hershey, PA, USA: IGI Global, 2020,pp. 317_346.