JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

CYBERSECURITY RESEARCH USING MACHINE LEARNING METHODS

Dr. Pankaj Kawadkar[1], Ajinkya S. Yadav[2]

[1]Asso. Prof., Sri. Satya Sai University of Technology and Medical Science, Sehore MP.

[2]Research Scholar , Sri. Satya Sai University of Technology and Medical Science, Sehore MP.

kawadkarpankaj@gmail.com, yadav.ajinkya008@gmail.com

Abstract: It is nearly impossible to quantify or justify the reasons why cyber security has such an outsized impact in the constantly expanding and quickly expanding field of cyber security. Allowing malicious threats to operate anywhere, at any time, or in any situation is far from acceptable and may result in serious harm. It specifically applies to the complex web of internet users, business information, and consumer data that cyber security organizations are struggling to protect and contain. For individuals and families, businesses, governments, and academic institutions that operate within the parameters of the global network or internet, cyber security may be an important consideration. We will improve the state of cyber security by using machine learning. Infrastructure systems that are currently in use are included together in this. The high-tech infrastructure of today, which includes network and cyber security systems, collects a vast quantity of data and performs analytics on nearly all the important components of mission-critical systems. Machine learning and artificial intelligence (AI) are gaining speed and gathering immense momentum in most of the areas of today's systems, whether it's positioned on premises or within the cyber security house, while people still provide the key operational oversight and insightful insights into the infrastructure of today.

Keywords: Machine Learning, cyber security, k-means, Random Forest, SVM etc.

I. INTRODUCTION

The cyber world is experiencing an increase in cyber attacks. To reduce or prevent the number of cyber attacks, enhanced security measures should be implemented. D-Dos assaults, Man in the Middle, information espionage, PROBE, User-To-Root, and Remote-To-Local attacks are only a few examples of the many types of attacks. Hackers or intruders use these assaults to gain illegal access to any non-public network, websites, data, or maybe our personal systems. Therefore, to protect the sensitive information, information, and financial data, outside or inside hackers utilize cutting-edge approaches or find ways to irritate or breach any defense systems. Intelligent intrusion weaponry should strive to manage or block a variety of inventive attacks that hackers have designed or coded.

To protect networks, devices, programmes, and information from assaults, damage, or illegal access, networks, devices, procedures, and practices are referred to as being "cybersecurity." The year 2016 saw numerous improvements in machine learning approaches, including self-driving cars, linguistic communication, the health sector, and sensible virtual assistants. Cybersecurity can also be referred to as it's security. They must be employed to locate useful information from various audit datasets that is used to the intrusion detection topic[8].

Submitted: 24/11/2022 **Accepted**: 09/12/2022

We will apply these concepts to cyber security using machine learning technologies to strengthen the defenses built into the intrusion detection system. We need to first feed the data into the machine learning model. The dataset sample trains the model, resulting in a trained model. The next step is to employ and apply the machine-learning formula after feeding the dataset sample.

This intrusion detection system's increased protective measures are greatly aided by the machine learning algorithm. The two categories of ML algorithms are supervised learning and unsupervised learning. By the information (i.e., input) they choose, they can be distinguished from one another. Algorithms that are given a set of labelled training data with the objective of determining what sets the labels apart are said to be learning under supervision. Unsupervised learning describes techniques where algorithms are given unlabeled training data and left to deduce the classes on their own. Most of the time, labelled data is extremely rare, or even labelling the data itself is a laborious operation, and we may not be able to tell if labels are indeed present.

II. RELATED WORK

In the early stages of developing intrusion detection systems, ML/DM (Machine learning/Data mining)-based cyber analytics support was investigated. Anomaly-based techniques simulate the behavior of the system and the conventional network, making it easier to identify anomalies as deviations from the norm. Its advantage is that traditional activity profiles are built specifically for each system, application, or network, making it difficult for attackers to understand the kinds of actions that can be carried out covertly and without detection. They appear appealing because of their unique capacity to detect zero-day assaults.

Anomaly detection and misuse are combined in hybrid approaches. They are used to lower the frequency of unidentified attacks and increase the detection rates of acknowledged intrusions. Once more, the creation of intelligent intrusion detection systems depends on the availability of a solid data set. An information set with a lot of high-quality data and one that simulates real time will only make it easier to train an intrusion detection system's associated check.

A Comprehensive Cybersecurity Audit Model

Public institutions and private businesses must to contend with ongoing, sophisticated cyber threats and cyber attacks. In order to protect themselves from cybercriminals, organizations should create and grow a cybersecurity culture and awareness. To deal with cyber threats, cyber risks, and cyber attacks that develop in a competitive cyber landscape, data technology audits like IT and data security audits like infosec, which were cost-effective in the past, attempt to converge into cybersecurity audits[1]. The complexity of the cyber threat landscape and the increase in the number and quality of assaults, however, are posing a challenge to the current cybersecurity audit models and providing justification for a new cybersecurity audit model. The simplest techniques and procedures used by global experts in the field of cybersecurity assurance and audit are reviewed in this text. In order to create a strong and coherent synthesis, the genuine scope, strengths, and shortcomings of these approaches and their theoretical foundation are highlighted through examination. In order to undertake cybersecurity audits in enterprises and Nation States, this work proposes an innovative and comprehensive

cybersecurity audit model. For all structure-useful areas, the CyberSecurity Audit Model (CSAM) examines and certifies audit, preventive, rhetorical, and detective controls [2]. CSAM has undergone testing, enforcement, and validation alongside cybersecurity. Each model is being validated by a research case study, and as a result, the results are made public.

Feature selection for machine learning techniques to detect botnets

It is granted a wholly original method to try to offer options to sight botnets at their portion of Command and control (C&C). A significant drawback is that although researchers have suggested solutions based on their research, there is no way to evaluate these solutions because some of them might have a lower detection rate than alternatives. In order to achieve the current goal, we identify the feature set that supports links between botnets at their C&C section and maximizes the rate at which those botnets are detected. Genetic formula (GA) was selected as the option with the highest detection rate since it is familiar to users. We frequently employ the machine learning formula C4.5, which distinguished between connections that belonged to a botnet and those that did not.

The datasets used in this work were taken from the ISOT and ISCX repositories [3]. A few experiments were conducted to introduce the GA's most basic characteristics and the C4.5 formula. We typically conduct trials simultaneously in order to obtain the most straightforward set of alternatives for each analysed botnet in particular, as well as for each type of botnet in general. The findings, which include a significant decrease in features and a greater detection rate than the related study conferred, are presented at the conclusion of the publication.

Intrusion Detection using Deep Belief Network

The issues with neural network-based intrusion detection, such as redundant data, a lot of data, and lengthy training, are easily solved at the local optimum. The use of deep belief networks (DBN) and probabilistic neural networks (PNN) is presented as an intrusion detection method. First, using the nonlinear intelligence of DBN, the raw data is converted to low-dimensional data while retaining the key features of the data. Second, the number of hidden-layer nodes per layer is optimized using the particle swarm optimization method to get the simplest learning performance. PNN is then used to categories the low-dimensional data [4]. Finally, the KDD CUP 1999 dataset is used to evaluate the effectiveness of the strategies stated before. The experiment result shows that the tactic performs higher than the standard PNN, PCA-PNN and non-optimized DBN-PNN.

Machine learning has moved from the lab to the forefront of operational systems during the past few of years. Machine learning is used frequently by Amazon, Google, and Facebook to improve customer experiences, guide purchases, link people socially with new apps, and enable personal interactions.

The strong capability of machine learning is also present in cybersecurity. Machine learning can be used by cybersecurity to improve malware detection, organize events, recognize breaches, and notify businesses of security issues. Machine learning may be used to identify sophisticated threats and targeting, including organization identification, infrastructure vulnerabilities, and potential win-win vulnerabilities and exploits. The cybersecurity environment will change significantly as a result of machine learning [5].

In an hour, malware alone may represent three million new samples. Malware analysis and detection techniques from the past are unable to keep up with modern attacks and variants. Cyber attacks are being delivered at frightening rates thanks to sophisticated malware and new attacks that are ready to evade network and endpoint detection. To address the growing malware problem, new methods like machine learning should be used. This claim discusses how machine learning can be used by cyber defense analysts to find and highlight sophisticated malware. The findings of our preliminary investigation are presented, along with a discussion of potential follow-up research to improve machine learning.

Comparison of Machine Learning Techniques' Effects on Intrusion Detection Systems (IDS)

Secure and dependable networks are essential given the rapid expansion of laptop networks and user content consumption. Because it has been established that there are more and more different types of network assaults, it is essential to develop a supply of reliable automated solutions for attack detection. One of the attack systems that finds incursions coming back from the internet is the intrusion detection system. The literature has identified a number of methods for network intrusion detection. To depict intrusion detection in the recent past, mining approaches were popular [6].

By employing the thoroughly mined data over the information provided within the network, the characteristics of incoming intrusions were known. When an identical object is discovered inside the parameters of the well-mined data, it is deemed to be an incursion. As a result of the current analysis's development of several intrusion detection models in support of this criterion, accuracy has increased. A brief examination of the quicker approaches is done. Information preprocessing procedures and detection approaches make up the complete strategy. Additionally, there are two types of information preparation approaches: feature extraction and transformation models, which support operational methods over the alternatives. The detection methods are categorized similarly as machine learning and organic process methods.

Improving Cybersecurity Assurance Model

When a group of auditors conducts an IT audit, a data security audit, or a compliance audit, there are recurring phases like designing, defining objectives and scope, defining terms of engagement, conducting the audit, gathering supporting evidence, assessing risks, reporting the audit findings, and scheduling follow-up tasks. A cybersecurity audit can be designed in a manner similar to other audits. However, given the high quality of the various cybersecurity fields, this will require a lot of work.

However, the scope of the internal audits does not include reviewing most cyber capabilities. This particular framework addresses the need for assurance, which will be attained through management reviews, cyber risk assessments, information management and protection, risk analytics, and crisis management, as well as the development life cycle, security programme, third-party management, information/asset management, access management, threat/vulnerability management, and security programme. Additionally, Deloitte's framework is compatible with industry frameworks like those of the Committee of Sponsoring Organizations of the Tread way Commission (COSO), National Institute of Standards and

Technology (NIST), Data Technology Infrastructure Library (ITIL), and the International Organization for Standardization (ISO).

Additionally, since live cybersecurity audits don't have any measurements, the subject is difficult to understand because it is constantly changing. Khan believes that in order to create a cybersecurity audit that has a meaningful scope, the auditors must cover all relevant areas of any organization, including client operations, finance, human resources, IT systems and applications, legal, purchasing, regulatory affairs, physical security, and all relevant third parties that interact with the company.

System for Detecting Database Intruders Using Octraplets and Machine Learning.

For host systems and networks, several intrusion detection solutions are created. There are, however, just a select few notable papers on information intrusion detection. A method by Chung et al. that presents a misuse detection strategy for information intrusion detection was one of the most recent publications to be made public. Here, common informational patterns are well-mined and continue to exist as conventional profiles. Its primary drawback is that role profiles are not generated. Users carry out completely distinct tasks in accordance with their jobs. User profiles cannot be the sole determining factor. Users can take acts that are supported by their positions and that will be flagged as malevolent. Time signatures were supported by the true time intrusion detection system Lee et al. developed. Systems for real-time information use temporal information objects, whose values change over time.

As a result, anytime it is upgraded, a device dealing is produced. Over a period of time, the temporal information is updated. If when a transaction tries to change temporal information that has already been changed by that much, an alarm is generated. The drawback of this approach is that it just considers updates rather than role profiles. Hu Panda creates user profiles using log files. Information that is frequently retrieved, tables, and hold on for comparison. The issue with this strategy is that when the size of the material is simply too large and the number of users concurrently increases dynamically, maintaining the knowledge becomes extremely difficult.

The variety and speed of today's cyber security threats make them unsuitable for manual protection. Additionally, machine learning gives you the strength and quickness to combat massive numbers of strikes with numerous permutations. However, using AI in conjunction with human intelligence, a combination of strength, speed, abilities, and judgment, is the $64,000 key to investing in cyber protection. Machine learning and artificial intelligence are frequently quite lovely and useful in sleuthing out cyber security breaches. The job that humans must accomplish is frequently completed at a much faster rate and with higher accuracy with the help of machine learning. Utilizing a variety of machine learning techniques can help the US identify cyber security attacks.

III. DISCUSSION AND FINDINGS

To "learn" what to watch out for and how to respond to various things on networks, machine learning security companies often use training methods on large data sets. However, machine

learning's technique should naturally fit antivirus defense and malware scanning because it is considerably more powerful than its name suggests:
The use of machine learning techniques can help the US detect malware, block the leakage of sensitive information, and monitor corporate executives' intrusions.
• Machine learning is increasingly enabling businesses to operate with greater effectiveness, precision, agility, and intelligence today.
• This strategy might make it easier to solve the safety issues.
• Any new, unusual pattern or behavior that might be caused by intrusions from the outside will be found and determined using machine learning rules.
• Security flaws brought on by permissions granted to users, programmers, or directors need to be handled with great care.

IV. CONCLUSION

In this research, machine learning and deep learning unit approaches for network security are reviewed. The literature paper introduces the most recent ML and DL unit applications in the field of intrusion detection systems, with a focus mostly on the last four years. Unfortunately, the most effective strategy for detecting intrusions has not yet been identified, thus investigation is still ongoing. It is clear from the discussion that was undertaken to make comparisons between the various approaches that each method for creating an intrusion detection system has its own advantages and disadvantages. As a result, choosing one implementation approach for an intrusion detection system over the others can be difficult. Network intrusion detection datasets are crucial tools for training and testing systems. Without representative data, machine learning and deep learning algorithms are useless, yet obtaining such a dataset is difficult and time-consuming. However, the currently available public dataset has a number of problems, such as inconsistent or outdated information, and therefore the problems are comparable. Most of these problems have limited the event of analysis in this specific space. Rapid network information updates present a bigger difficulty for the ML and DL model coaching. The Model needs to be retrained fast and on a semi-permanent/long-term basis. The future study of this topic will therefore have a long-term focus on progressive learning and extended learning.

REFERENCES

1. J. Cano, "Cyberattacks-The Instability of Security and Control Knowledge", ISACA Journal, vol. 5, pp. 1-5, 2016.
2. C. Hollingsworth, "Auditing from FISMA and HIPAA: Lessons Learned Performing an In-House Cybersecurity Audit", ISACA Journal, vol. 5, pp. 1-6, 2016.
3. Li X, Wang J, Zhang X, “Botnet Detection Technology Based on DNS”, J. Future Internet, 2017.
4. Y J Hu, Z H Ling, "DBN-based Spectral Feature Representation for Statistical Parametric Speech Synthesis", IEEE Signal Processing Letters, vol. 23, no. 3, pp. 21-325, 2016.

5. Dinil Mon Divakaran et al., "Evidence gathering for network security and forensics", Digital Investigation, pp. 56-65, 2017.
6. S Fong, R Wong, A V Vasilakos, "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data", IEEE Transactions on Services Computing, vol. 9, no. 1, pp. 33-45, 2016.
7. M. Khan, "Managing Data Protection and CybersecurityAudit's Role", ISACA Journal, vol. 1, pp. 1-3, 201
8. Ram Kumar, Sarvesh Kumar, Kolte V. S.," A Model for Intrusion Detection Based on Undefined Distance", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011