Using machine learning techniques, Cyber Attack detection: A Review

Masrath Parveen[1], Dr. Saurabh Pal[2], Dr. Venkateswara Rao CH[3]
[1]Research Scholar, Dept of CSE, V.B.S.Purvanchal University, Jaunpur
[2] Department of CSE, V.B.S.Purvanchal University, Jaunpur
[3]Department of CSE, Siddhartha Institute of Engineering and Technology, Hyderabad

Abstract –
Data transmission and reception on all networks currently experience network congestion issues. Due to different cyber attacks, this data flow is moving slowly for a variety of reasons. The server system is harmed by the internal workings of the cyberattacks. Operating a system while a network is experiencing cyber attacks is particularly dangerous. The survey on cyberattack detection utilising machine learning techniques, the Jupiter simulation tool, and the WEKA simulation tool is presented in this work. The several types of cyberattacks mentioned in this study include DOS, TCP/P, flood, UDP, ICMP, U2R, R2L, DDOS, probing assaults, along with detection using novel approaches and machine learning techniques. While data mining techniques work with specialised data sets and the identification of cyberattacks, machine learning approaches work with generalised data sets. Along with the simulation method, the goals of cyberattacks detection and classifications are also elaborated.

Keywords: Network security, Cyberattacks, cyber serucity, machine learning, attacks
DOINumber:10.48047/nq.2022.20.22.NQ10377 **NeuroQuantology2022;20(22):3796-3803**

1. Introduction

The detection of cyberattacks is increasingly using machine learning techniques. Using these methods, malicious behavior can be automatically detected, and patterns that might point to an impending attack can be found. Machine learning algorithms are the best choice for spotting cyberattacks since they can quickly and accurately analyze vast amounts of data. Machine learning can be used to spot suspicious activity on a system or network, find malicious files, and spot anomalies in network traffic. Additionally, it is possible to train machine learning models to recognize well-known attack patterns and notify security personnel when they are discovered. Organisations may lessen the risk of data breaches and better defend their networks from assaults by utilizing machine learning techniques.

Enhanced crime is a bad habit that recalls the use of automated advancements for offence commission, helped by enrollment and communication headways. The most recent methods that are becoming more prevalent in the use of web activity attain making deceptive, deficient ways for conveying sensitive data to carry out an offence through criminal behaviour. The development includes things like hacking into ICDSS, theft, created images, online trade blackmail, web deal distortion, and more. It also includes plans for web harmful activities like disease, worms, and untouchable abuse like phishing, email scams, etc.[1] The entire framework of association, like the web, needs to stop engaging in criminal activity from one side of the world to the other and stop the criminal nature by defending illegal development by maintaining various firewalls placed under each nation's separate control to

monitor and thwart bad behaviour carried out online. To prevent developers from entering networks that combine firewalls, virtual private networks, and encryption computations, network security measures are utilised. [2]

2. Literature review

The 2000s saw the rise of cybercrime as a significant problem. Cybercriminals started using the internet to perform a wide range of crimes, such as fraud, online harrassment, and identity theft. In 2001, the first significant cybercrime to affect the United States took place when a hacker obtained access to the personal data of millions of Card Systems Solutions customers [1].

The 2010s saw a continual evolution and sophistication rise in cybercrime. Malware and ransomware were first used by cybercriminals to target people and companies for monetary gain. In order to acquire private data, thieves have also started employing social engineering strategies like phishing and spear phishing [2]. More than 200,000 machines in 150 countries were impacted by the WannaCry ransomware assault in 2017 [3].

Due to the COVID-19 pandemic, cybercrime has increased in the 2020s [4]. Cybercriminals have launched phishing efforts that target people and businesses with dangerous links or attachments in order to capitalise on people's anxieties about the virus [5]. Additionally, there have been more attacks on cloud-based services and instances of cryptojacking, in which criminals mine bitcoin on the computers of their victims without their knowledge or agreement [6].

One of the most frequent consequences of cyber crime is financial loss. Hackers can target the average person and steal their personal data, including bank account details and credit card numbers. Once this information is taken, it can be used to access accounts secretly and make unauthorised transactions or withdrawals of money. Victims may not be able to recoup from their severe financial losses as a result of this[8].

2.1 Review on Cyber Crimes

India has recently had a large number of cybercrimes. With more than 1.5 million incidents reported in 2019 alone, cybercrimes have dramatically increased in India over the last few years [9]. Phishing is one of the most prevalent forms of cybercrime, in which perpetrators send emails or messages that seem to come from reliable sources but are actually harmful attempts to steal people's money or personal information. Phishing assaults are getting more and more complex, and they can be hard to spot [10]. Identity theft is another sort of cybercrime, in which thieves use credit card numbers or other stolen personal information to perform fraud or other illicit actions. Financial losses and reputational harm can result from identity theft [11]. Another type of cybercrime is cyber extortion, in which offenders threaten to damage victims if they do not pay a ransom. Businesses, governments, and people are frequent targets of cyber extortionists who demand money from them [12]. In India, cyberbullying is a growing issue as well. Bullies target young people online by making harsh comments or spreading false information about them on social media platforms. Cyberbullying can have severe psychological repercussions on victims and, in some situations, can even cause them to consider suicide [13].

The primary example of cross-line misconduct is cybercrime. All countries are connected by PC networks, allowing scoundrels to cause any form of harm anywhere on the planet without leaving their home office. The potential harm is unexpectedly distributed, ranging from people being unable to access their PC for a short period of time or unintentionally discovering bigoted or lewd content online, to an organization's internal network being unavailable for a full day or losing proprietary information, to government websites being blocked or discovering state secrets online [14]. The loss of money can be anything from two or three hundred dollars due to coercion to exorbitant losses brought on by digital deception or digital injury [15]. As the internet becomes more and more integrated into daily life, cybercrime also entails the risk of psychological oppressor attacks taking down a

sizable portion of the internet and resulting in a global economic and social crisis [16].

2.2 Acts on cyber crime

1. Computer Fraud and Abuse Act: Violations of this federal law, which carries a maximum 10-year prison sentence, are punishable by accessing a computer without permission or exceeding that permission.
2. Identity Theft and Assumption Deterrence Act: According to this federal legislation, it is illegal to knowingly transfer or use someone else's identification information without their consent with the intention of breaking the law or helping them do so.
3. CAN-SPAM Act: This federal law outlines regulations for commercial messages, defines guidelines for commercial emails, allows recipients the option to stop receiving emails, and outlines severe consequences for noncompliance.
4. The Children's Online Privacy Protection Act (COPPA) is a federal regulation that mandates parental consent be obtained before collecting personal data from children under the age of 13 on websites. Additionally, it mandates that websites include privacy notices and give parents access to their children's personal data.
5. The Digital Millennium Copyright Act (DMCA), a federal law, forbids the development and distribution of tools, gadgets, or services designed to get around restrictions on access to works protected by copyright. Additionally, it offers safe harbours from punishment for online providers of services who follow particular guidelines when responding to user complaints of copyright infringement.

2.3 Machine Learning Tools and Techniques to handle Cyber Crimes

1. Anomaly Detection: is an approach to machine learning that makes use of statistical models to find out-of-the-ordinary data patterns that could point to a cyberattack. It can be utilised to spot bad behaviour including malware infestations, unauthorised access, and data espionage.
2. Network Intrusion Detection System (NIDS): An example of a security system, NIDS scans network traffic for any unusual activity and notifies administrators if an attack is found. It may be used to detect a variety of cyberattacks by using machine learning techniques to spot suspicious patterns in network data.
3. Malware analysis: is a method of examining the source code of malicious software to determine the intent and behavior of the programme. Security analysts can swiftly recognize and respond to threats thanks to machine learning algorithms that may be used to recognize the patterns and behaviors of malicious code.
4. Phishing Detection: is a form of social engineering assault in which attackers send phoney emails or links to victims in an effort to get them to divulge personal information or download dangerous software. By looking for suspicious patterns in emails or links, machine learning algorithms may be employed to spot phishing efforts.
5. User Behaviour Analysis: Using machine learning methods, user behaviour analysis tracks activity by users on networks or systems for any unusual activity that might point to an attempted attack or security breach. By keeping an eye on user activity for indications of malevolent intent or unauthorised access attempts, it is additionally possible to identify insider threats.

2.4 TCP, UDP, ICMP fragmenting attacks

Attacks that fragment TCP, UDP, or ICMP packets are examples of cybercrimes that can stop network traffic. The most popular internet protocol for transferring data is TCP (Transmission Control Protocol) [17]. It is a connection-oriented protocol that guarantees trustworthy data packet delivery between computers. User Datagram Protocol (UDP) [18] is a connectionless protocol that cannot ensure the delivery of data packets with any degree of reliability. An error message and a query message are sent between computers using the network layer protocol known as ICMP (Internet Control Message Protocol) [19].

In order to get beyond security measures or disrupt the network, fragmenting attacks divide huge data packets into smaller pieces. DDoS assaults, which saturate networks with traffic in an effort to overwhelm them and render them

unavailable, can be launched via fragmenting attacks. Attacks that fragment data can also be carried out maliciously to steal information or start malware attacks [20].

Organisations should put in place robust security measures like firewalls, intrusion detection systems, and antivirus software to protect themselves against these kinds of cybercrimes. Organisations should also routinely check their networks for suspicious behavior and take action to eliminate any dangers as soon as they arise [21].

2.5 DDOS, DOS, Probe attacks

Cyberattacks known as distributed denial of service (DDoS) aims to render a network or system inaccessible by flooding it with traffic from numerous sources. Typically, this kind of attack is used to obstruct services or bring down websites. Any system, including web servers, email servers, and even entire networks, is vulnerable to DDoS attacks [22].

DDoS attacks and denial of service (DoS) attacks both aim to render a network or system inaccessible by saturating it with traffic from numerous sources. While DDoS attacks originate from numerous sources, DoS attacks are normally conducted from a single source. Websites can be taken down or services can be interrupted using DoS assaults, but they are typically less successful than DDoS attacks [23].

Another kind of cyberattack that is used to get private data on a network or system is the probe attack. These kinds of attacks deliver nefarious data packets to the target system in an effort to obtain access and gather data. Probe attacks can be used for a variety of things, including obtaining credit card numbers, passwords, and other private information [24].

2.6 Malware attacks, Network attacks, WSN attacks

Attacks by malicious software that aim to disrupt, harm, or gain unauthorised access to a computer system are known as malware attacks. Malicious websites, downloads, and email attachments can all transmit malware. On a computer system, it can also be installed without the user's awareness. Malware is a tool that can be used to steal data, destroy files, or take over a computer system [25].

Network attacks are attempts to disrupt a network or its resources or gain unauthorised access to them. These assaults can be launched both from outside and inside the network. Denial-of-service (DoS), distributed denial-of-service (DDoS), man-in-the-middle (MITM), and spoofing assaults are examples of network attacks [2].

Attacks on WSNs: Wireless Sensor Networks (WSNs) are networks of tiny devices that use wireless connection to gather information from their surroundings and send it back to a central point for analysis. WSNs are susceptible to a number of attacks, including node capture, jamming, spoofing, and eavesdropping. These kinds of attacks can be used to obstruct network functionality or get unauthorised access to confidential information [7].

2.7 Issues with present cyber attacks detection system

1. Lack of automation: The majority of cyber attack detection systems are labour- and time-intensive manual processes.
2. False Positives: A huge proportion of cyber attack detection systems produce false positives, which can result in pointless alarms and the waste of resources.
3. Limited Visibility: Cyber attack detection systems frequently lack complete network coverage, creating blind spots that attackers can take advantage of.
4. Complexity: Setting up and maintaining cyber attack detection systems can be challenging and call for specialized skills and knowledge.
5. Price: Because cyber attack detection systems can be expensive to develop and operate, many organizations find them to be unaffordable.

3. Proposed methodology

The fracture in P/ICMP arises because of the massive traffic quantities, DDOS attacks, like previous DDOS attacks, will overwhelm the target assets. However, this DDOS attack will also force the target to use resources to try and reassemble the bundles, which frequently causes network devices and servers to break. In the end, it is difficult to determine which

parcels are safe to drop and which are not because the non-starting sections do not provide information about whose administration they have a place with.

- With the aid of the Jupiter Anaconda Navigator simulation tool and the Python code, machine learning methods like CNN and DEEP learning will be implemented.
- In the next step the process involves identifying different types of cyber-attacks using decision trees and J48 algorithms on the WEKA simulation tool for a given data set as a sampling for simulation analysis.
- Initially collection of data set network files that already includes information about different types of cyber-attacks will be converted into a CSV file and uploaded to simulation tools like the WEKA tool and the Jupiter anaconda navigator tool for different types of cyber-attacks detection, respectively.
- The results will be generated in the form of tables and graphs with respect to the confusion matrix, ROC curve, plotted graph, etc. and will compare machine learning methods for cyber-attacks detection as compared in percentile ratio.
- Adapting an intrusion detection system which monitors and identifies harmful activities on a network. It is capable of spotting unusual activity such as unauthorised access attempts and suspicious traffic.
- Adaption of Firewall which pre-establishes security rules in a system for monitoring and regulating incoming and outgoing network traffic. The network can be protected by using it to stop malicious traffic from entering or exiting.
- Usage of Antivirus software where the malicious software is detected, stopped, and removed from computers and networks using antivirus software. By checking for known malware signatures, it can also find unusual activities on the network.
- Imparting network monitoring entails which keeps an eye on network activities to look for harmful or suspicious behaviour for identifying anomalous traffic patterns, unauthorised access attempts, and other actions that can point to an ongoing or completed assault.
- Performing data loss prevention (DLP) which is a security mechanism that stops unauthorised individuals or systems from accessing or stealing sensitive data. It can be used to keep track of data transfers between networks and spot any odd behaviour that might point to an ongoing or completed attack.
- Verifying User Behaviour Analytics (UBA) is a security measure that analyses user behaviour using machine learning algorithms to find outlier behaviours that might point to an ongoing or completed assault.

4. Experimental Results

The digital assaults discovery is acted out in the WEKA TOOL SIMULATION TOOL game using a decision tree and j48 computation put combined execution per frame with regard to WEKA SIMULATION TOOL. In this, we use the KF model addressing and collect the number of tasks completed based on DOS attacks, PROBE attacks, R2L attacks, U2L attacks, and others kind of organisation attacks. The KF model is correctly executed on the WEKA recreation system as part of the suggested study effort, and then precision, efficiency, and all estimates boundaries are created. These attacks by a wide range of organisations can be classified as cybercrime due to the web and PC component of the crime. It is stated that IT 2000 is being demonstrated.

The Pycharm Simulation Tool provides a game simulates cross-prearranging attacks, which are a subset of cybercrime assaults amassed on online application space. As a result, we are illustrating in this research project how to count and anticipate cross prearranging assaults using Python and the PYCHARM reproduction device. In this research effort, these kinds of both executions were used. We are accustomed to finding two different types of assaults include:

1. Cyberattacks on the network infrastructure.
2. Cross-application pre-arrangements of attacks.

These are re-enactment devices that WEKA and PYCHARM are performing. The decision tree and J48 computations are done on the WEKA instrument in order to use execution. PYCHARM reenactment device's cross-prearranging attacks detection and counteraction. All of these sorts of attacks fall under the act IT 2000's cyber law categories.

5. Conclusion

This study suggests a cleverly disseminated IDS to detect and prevent actions like refusal management, testing, client to attach and remote to client attacks. In this study, we implement and evaluate an IDS using a Bayesian organisational structure layout demonstration technique. The suggested model combines a flexible teaching experience with an abnormality-based IDS framework. The KDD DAPRA dataset, which was designed for network IDS assessment, has been compared to the suggested IDS framework.

Four different Bayesian organisations are used in the exploration technique as arrangement models, and each of these classifier models is connected to the others and used to forecast incoming organisation traffic data. Every intended Bayesian organisational model is capable of detecting a large class of attack, such as denial of service (DoS). job and tests.

Execution supported the IDS model using the WEKA Java API and called for reenactment scenarios. The results of the tests support the suggested IDS frameworks. The tests demonstrate that the suggested framework is effective in differentiating attacks in the test dataset and is extremely precise in recognising all important attacks included in the DARPA dataset. Additionally, the proposed IDS framework's successful implementation can be utilised to plan and identify attacks in actual organisation traffic.

References

[1] T. Thomas, A. P. Vijayaraghavan, and S. Emmanuel, ‘‘Machine learning and cybersecurity,’’ in Machine Learning Approaches in Cyber Security Analytics. Singapore: Springer, 2020, pp. 37–47.

[2] M. A. Shaik, M. Varshith, S. SriVyshnavi, N. Sanjana and R. Sujith, “Laptop Price Prediction using Machine Learning Algorithms”, 2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS), Nagpur, India, 2022, pp. 226-231, doi: 10.1109/ICETEMS56252.2022.10093357.

[3] Mohammed Ali Shaik, Praveen Pappula, T Sampath Kumar, "Predicting Hypothyroid Disease using Ensemble Models through Machine Learning Approach", European Journal of Molecular & Clinical Medicine, 2022, Volume 9, Issue 7, Pages 6738-6745. https://ejmcm.com/article_21010.html

[4] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, (2022), "COVID-19 Detector Using Deep Learning”, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 443-449, doi: 10.1109/ICAAIC53929.2022.9792694.

[5] Y. Li, H. Voos, M. Darouach and C. Hua, "An application of linear algebra theory in networked control systems: stochastic cyber-attacks detection approach," in *IMA Journal of Mathematical Control and Information*, vol. 33, no. 4, pp. 1081-1102, Dec. 2016, doi: 10.1093/imamci/dnv026.

[6] Mohammed Ali Shaik and Dhanraj Verma, (2022), "Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020025-1–020025-9; https://doi.org/10.1063/5.0081719, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020025-1 to 020025-9.

[7] M. A. Shaik and Dhanraj Verma, (2022), "Predicting Present Day Mobile Phone Sales using Time Series based Hybrid Prediction Model", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020073-1–020073-9;

https://doi.org/10.1063/5.0081722, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020073-1 to 020073-9

[8] J. Zhou, B. Chen and L. Yu, "Intermediate-Variable-Based Estimation for FDI Attacks in Cyber-Physical Systems," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2762-2766, Nov. 2020, doi: 10.1109/TCSII.2020.2964650.

[9] Mohammed Ali Shaik, MD.Riyaz Ahmed, M. Sai Ram and G. Ranadheer Reddy, (2022), "Imposing Security in the Video Surveillance", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020012-1–020012-8; https://doi.org/10.1063/5.0081720, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020012-1 to 020012-8.

[10] M. A. Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, (2022), "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, 020072-1–020072-8; https://doi.org/10.1063/5.0081726, Published by AIP Publishing. 978-0-7354-4368-6, pp. 020072-1 to 020072-8.

[11] Mohammed Ali Shaik, Dhanraj Verma, (2021), Agent-MB-DivClues: Multi Agent Mean based Divisive Clustering, Ilkogretim Online - Elementary Education, Vol 20(5), pp. 5597-5603, doi:10.17051/ilkonline.2021.05.629

[12] Z. Liu, Q. Wang and Y. Tang, "Design of a Cosimulation Platform With Hardware-in-the-Loop for Cyber-Attacks on Cyber-Physical Power Systems," in *IEEE Access*, vol. 8, pp. 95997-96005, 2020, doi: 10.1109/ACCESS.2020.2995743.

[13] Mohammed Ali Shaik and Dhanraj Verma, (2020), Enhanced ANN training model to smooth and time series forecast, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022038, doi.org/10.1088/1757-899X/981/2/022038

[14] M. A. Shaik, Dhanraj Verma, P Praveen, K Ranganath and Bonthala Prabhanjan Yadav, (2020), RNN based prediction of spatiotemporal data mining, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022027, doi.org/10.1088/1757-899X/981/2/022027

[15] A. N. Jahromi, H. Karimipour, A. Dehghantanha and K. -K. R. Choo, "Toward Detection and Attribution of Cyber-Attacks in IoT-Enabled Cyber–Physical Systems," in IEEE Internet of Things Journal, vol. 8, no. 17, pp. 13712-13722, 1 Sept.1, 2021, doi: 10.1109/JIOT.2021.3067667.

[16] Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041

[17] Mohammed Ali Shaik, “Time Series Forecasting using Vector quantization”, International Journal of Advanced Science and Technology (IJAST), ISSN:2005-4238,Volume-29,Issue-4 (2020), Pp.169-175.

[18] R. V. Yohanandhan, R. M. Elavarasan, P. Manoharan and L. Mihet-Popa, "Cyber-Physical Power System (CPPS): A Review on Modeling, Simulation, and Analysis With Cyber Security Applications," in *IEEE Access*, vol. 8, pp. 151019-151064, 2020, doi: 10.1109/ACCESS.2020.3016826.

[19] Mohammed Ali Shaik, “A Survey on Text Classification methods through Machine Learning Methods”, International Journalof Control and Automation (IJCA), ISSN:2005-4297,Volume-12,Issue-6 (2019), Pp.390-396.

[20] R. Sager, R. Javari, and C. Borrego, “Applications in security and evasions

in machine learning: A survey", Electronics, vol. 9, no. 1, p. 97, Jan. 2020.

[21] Mohammed Ali Shaik, P. Praveen, T. Sampath Kumar, "Integration and application of Fog, IoT and Edge Computing", Fog Computing: Concepts, Frameworks, and Applications (FCCFA) Aug-2022, CRC Press, ISBN: 9781003188230.

[22] Praveen, P, Mohammed Ali Shaik, T. Sampath Kumar, Choudhury T, "Smart Farming: Securing Farmers Using Block Chain Technology and IOT", Aug-2021, EAI/Springer Innovations in Communication and Computing, ISBN: 978-3-030-65690-4

[23] M. A. Shaik, "Protecting Agents from Malicious Hosts using Trusted Platform Modules (TPM)," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 559-564, doi: 10.1109/ICICCT.2018.8473278.

[24] Y. Guan and X. Ge, "Distributed Attack Detection and Secure Estimation of Networked Cyber-Physical Systems Against False Data Injection Attacks and Jamming Attacks," in *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 1, pp. 48-59, March 2018, doi: 10.1109/TSIPN.2017.2749959.

[25] A. Ashok, M. Govindarasu and J. Wang, "Cyber-Physical Attack-Resilient Wide-Area Monitoring, Protection, and Control for the Power Grid," in *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1389-1407, July 2017, doi: 10.1109/JPROC.2017.2686394.