SURVEY PAPER ON DETECTING CYBER THREATS USING RANDOM FOREST ALGORITHM

[1]Divya D P, [2]Divyashree S, [3]Ananya Vinayak Shanbhag, [4]Bhuvana H R, [5]Deepthi M

RNS Institute of Technology
Bengaluru.

***Abstract*- An incursion is any unapproved activity, access, or an effort to undermine the safety of a computer system, network, or application. Numerous malicious behaviors, such as denial-of-service attacks, malware infections, unauthorized access, and data breaches, can compromise the availability, confidentiality, or integrity of data systems. Intrusions pose a major risk to cybersecurity because they can lead to a multitude of unfavorable outcomes. Because the Random- Forest algorithm is an effective method for collective education that can handle a variety of complex and diverse datasets, we have chosen to employ it in this study.**

***Keywords:* Intrusion Detection-System(IDS),Random Forest,DataSet, Networks, Detection,Cyber security, Neuralnetwork, machine learning.**

I. INTRODUCTION

Intrusion detection-systems (IDS) are required to safeguard computer networks and systems from malicious activity, unauthorized access, and attacks. Given the existing situationcyber threats[8], it is critical to have robust and adaptable intrusion detection techniques. One such alternative is toapply machine-learning techniques; in this case, the Random-Forest method is a useful tool. One well- liked way of group learning that is well-known for its adaptability and effectiveness in handling complex data patterns is Random Forest.When in relation to intrusion detection, it offers the capacity to differentiate between typical and abnormal network behaviour by learning from different aspects and patterns within the data.. This fusion ofcybersecurity utilizing machine learning aims to decrease false positives, enhance proactive threat detection, and offer an adaptable and intelligent defense system against constantly changing cyberthreats. With machine learning's benefits, this study on "Intrusion System of Detection Employing Random Forest "[9] aims to strengthen network security by providing a proactive defence against the intricate and ever-evolving world of cyber threats.

II. SCOPE

The "Intrusion System of Detection Employing Random Forest Algorithm" project can be utilised in a range of sectors and industries to enhance cybersecurity measures. Among the potential applications are:

1. Large organizations and enterprises can deploy this system to safeguard their internal networks, servers, and private information from malevolent actions and illegal access.
2. Banks and financial institutions can employ the framework for protecting financial transactions, customer data, and safeguarding the integrity of their systems and protecting vital infrastructure from cyberattacks.
3. Healthcare organizations can use the framework for secure patient records, medical devices, and sensitive healthcare information against unauthorized access and data breaches.
4. Government agencies can put the framework for bolstering the security of their networks, protecting sensitive information, infrastructure, and ensuring the continuity of essential services.
5. Online businesses and e-commerce platforms candeploy the framework for secure customer data, financial transactions, and prevent fraudulent activities.
6. Universities and educational institutions can utilize the framework for protecting academic and research data, student records, and sensitive information from cyber threats.

II. OBJECTIVES

This project's goal is to develop and implement a put in place a collection of choices of trees for accurate and dependable using the Random-Forest for intrusion- detection technique.

III. PROPOSED SYSTEM

By means of numerical models and computational algorithms,[10] machine learning (ML) employing Random Forest to identify intrusions automatically identifying unusual activity or possible security risks within a computer network or system. While rule-based techniques are often thebackbone of conventional intrusion detection solutions, machine learning offers a more flexible and data- drivenparadigm. Forintrusion-detection, Random Forest is a well- liked and effective option due to several noteworthy advantages that address concerns about intrusion detection classification requirements and cybersecurity data characteristics.

IV. LITERATURE SURVEY

"Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network." [1]

Authors: Andrei-Grigore Mari, Daniel Zinca, Virgil Dobrota
Two among the most crucial problems in the infrastructure of system safety to solve are interruption detection and prevention. Intrusion-detection systems (IDSs) protect networks by identifying malicious traffic using patterns. A number of systems for detecting intrusions[1] using machine learning has been created in response to attempts by adversaries to use traffic simulations to alter the rules. In this work, we concentrated on among the models that was trainedand evaluated against the NSL-KDD dataset as a reference using a sequence of methodologies. We show how to generate malicious network traffic examples that could be used to avoid detection by an intrusion-detection system thatmakes use of machine learning. Additionally, by training on the current traffic, act can be enhanced against potentialdangers in the future.Therefore, a deep-learning algorithm- based architecture known as a generative adversarial network (GAN) that can produce generative models wasimplemented.

“Cyber Intrusion Detection Using Machine Learning”[2]

Authors: Hamed Alqahtani , Iqbal H. Sarker , Asra Kalim , Syed MD. Minhaz Hossain , Sheikh Ikhlaq , Sohrab Hossain.
As the alarming growth of connectivity of computers[2]
and the significant number of computer-related applications increase in recent years, the task of fulfilling cyber-security is increasing consistently.

It also requires a strong defense against various cyberattacks.Consequently cybersecurity may be improved by detecting anomalies and cyberattacks on a computer system and by creating intrusion finding systems. It has become possible to generate a practical data-driven intrusion-detection system with artificial intelligence, particularly with machine learning This study uses a number of well-known machine learning classification techniques to detect breaches since they provide insightful cyber-securityservices. These algorithms include Decision Tree, Naïve- Bayes classifier, Bayesian network, Random Decision Forest, Random Tree, Decision Table, and Artificial Neural Network. In conclusion, w assess the effectiveness of diverse experiments conducted on cyber-security datasets containing many kinds of cyber attacks and assess the accuracy, recall, f1-score, precision and efficiency of the performance indicators.

“Network intrusion-detection system: A machine learning approach”[3]

Authors: Mrutyunjaya Panda, Ajith Abraham, SwagatamDas, Manas Ranjan Patra.
These days, intrusion-detection systems, or IDSs, are receiving a lot of attention as an essential component ofsystem security. For the purpose systems for interruptiondetection and network security gather network circulation data from numerous locations within a computer system or network.. Thanks to machine learning methods, it's now possible to identify a network incursions (or attacks) and let network managers take preventative measures to stop invasions. This is becoming an increasingly important capability. The ten machine-learning approaches that we suggest applying in this study are Selective multinomial Naïve Bayes, Hybrid J48 utilising Rotation Forest, Decision Tree, Hybrid J48 with Decision Tree, Hybrid J48 with Lazy Locally Weighted Learning, and merging random-Forest with AdaBoost (AB) and Naïve Bayes in conjunction with NB and J48 classifiers are accustomed to effectively detect network breaches. We assess the machine learning techniques we have suggested for detecting network intrusions. Using the NSL-KDD dataset, that is a version of the popular KDD Cup 1999 intrusion- detection benchmark dataset. Lastly, the average cost of misclassification, false positive rate, as well as detection rate are shown for a 5-classclassification according to the results of the experiment[3]. These are employed to give researchers looking into network- intrusion detection a better understanding.

“Intrusion Detection Using Machine Learning: A Comparison Study” [4]

Author: Saroj Kr. Biswas
As the internet has grown over time, so too has the frequency of cyber attacks. To provide network security, an intrusion- detection system (IDS) must be strong. Thepurpose[4] of an intrusion-detection system (IDS) is to track all active processes on a system and look for any signs of possible anomalies. Although a great deal of study has been conducted in this field, a comprehensive and in-depth analysis is still lacking. This research proposes an IDS using machine- learning for networks with a good union of feature selection technique and classifier by looking at the combinations of most often used feature selection methods and classifiers. Using feature selection approaches, a subdivision of pertinent characteristics is chosen from the initial collection of features. The subsection of significant features is then used to train multiple classifiers, which builds the IDS. To get results, the NSL-KDD dataset is subjected to five folds cross validation. Ultimately, it is discovered that the K-NN classifier and the information gain ratio-based feature selection techniqueperform superior[4]to the others.

“PRACTICAL REAL-TIME INTRUSION DETECTION USING MACHINE LEARNING APPROACHES” [5]

AUTHORS: PPHURIVIT SANGKATSANEE,NARUEMON WATT ANAPONGSAKORN, CHALERMPOL CHARNSRIPINYO
The growing frequency of network outbreaks is one well- known problem that can compromise the accessibility, privacy, and honesty of crucial information for both people and enterprises. In this investigation, we offer a real-time intrusion-detection method using supervised machine learning. Our method is effective with a range of types of methods for machine learning and is simple to use. We made use of several well-known machine learning approaches to assess our IDS plan's effectiveness. According[5] to the outcomes of our experiments, the Decision Tree strategy is more potent than the alternative methods. Consequently, we developed a real-time intrusion-detection system (RT-IDS) that differentiates between attack and legitimate online system traffic using the Decision Tree technique. By employing the information gain as our feature selection criterion, we also discovered 12 significant network data pieces that are needed to establish network assault criteria.

Our RT-IDS has a spotting rate of more than 98% in less than two seconds and can discern between typical network activityand the two primary attack types (Probe and Denial of Service(DoS)). To ensure that lower the false-alarm rate and improve[5] the intrusion-detection system's dependability and detection accuracy, we also created a brand-new post- processing technique.

“Machine Learning Based Intrusion-Detection System”[6]Author: Anish Halimaa A. , K. Sundarakantham
Intrusion-detection systems are accustomed to investigate malicious actions that occur within a system or network. Software or hardware employed for intrusion-detection searches a system or network for suspicious activities. The increasing interconnectedness of computers has made intrusion detection essential to network security. Network protection Intrusion-Detection Systems possessed been developed using several types of machine-learning and statistical approaches. Accuracy is the primary determinant of intrusion-detection system performance. To lower false alarms and raise detection rates, intrusion detection accuracy must be enhanced. In recent works, [6]Numerous tactics havebeen used to improve performance standards. An intrusion [6]detection system's primary task is to analyze largeamounts of info on network traffic. To deal with this matter, awell- structured classification system is needed. This problem is approached in the suggested manner.

“A Review on Intrusion-Detection System using MachineLearning Techniques” [7]

Authors: Usman Shuaibu Musa, Sudeshna Chakraborty,Muhammad M. Abdullahi
The widespread utilisation of the internet exposes computer networks to cyber-related attacks; thus, numerous intrusion detection-systems (IDSs) have been proposed by various researchers. Identifying intrusions is among the principal study topics in network security. To ensure the network's security, it is helpful to identify instances of illegal use and attacks. Many methods[7] have been employed forth to determine which is most useful features and thereby raise the efficiency of systems for detecting intrusions. These techniques include Markov neural networks, machine learning-based (ML), Bayesian-based algorithms, swarmsmart algorithms, nature-inspired meta-heuristic techniques, and Markov learning. A variety of databases been employed to evaluate[7] the numerous initiatives that have worked on overtime.

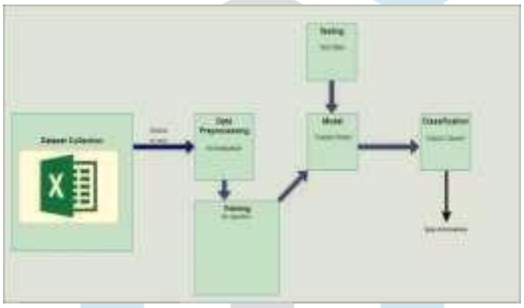
V. ADVANTAGES OF PROPOSED SYSTEM

- Random Forest able to adjust to diverse and complex patterns within network traffic, making it efficient in finding both known and novel threats. Its the ability to analysis a multitude of features allows it to identify minute departures from typical conduct.
- Unbalanced datasets can be handled using Random Forest, a common scenario in intrusion-detectionwhere normal traffic significantly outweighs malicious activities. The algorithm's ensemble nature mitigates biases and supports accurate detection belonging to a minority class instances.

VI. PROPOSED METHODOLOGY

1. Data Collection and Preprocessing: Collect a dataset that includes labelled examples of normal and malicious network traffic. By managing missing values, encoding categorical variables, and scaling or normalising numerical features, preprocessing the data is done.
2. Dataset Splitting: Divide the dataset into sets for testing and training. The Random-Forest model is trained on the training set, and its execution is assessed on the testing set.
3. Feature Selection: Determine pertinent characteristics for intrusion-detection. Random-Forest inherently provides feature importance scores, allowing you tochoose the most informative features for the model.
4. Model Training: Train the dataset used for training is utilised by the Random-Forest model. By selecting subsets of features for each tree at random and combining their results, the programme creates numerous decision trees.
5. Evaluate the the trained model's performance using the dataset used for testing.

VII. SYSTEM ARCHITECTURE



The widespread utilisation of the internet exposes computer networks to cyber-related attacks; thus, numerous Systems for detecting intrusions (IDSs) have been proposed by various researchers. Identifying intrusions is among the principal study topics in network security. To ensure the network's security, it is helpful to identify instances of illegal use and attacks. Numerous methods have been employed to ascertain the most useful features and thereby raise how successful intrusion detection issystems. These techniques include Markov neural networks, machine learning-based (ML), Bayesian-based algorithms, swarm smart algorithms, nature-inspired meta- heuristic techniques, and Markov learning.

XI. CONCLUSION

The installation of an intrusion-detection system (IDS) built on a random forest is a crucial step in strengthening the cybersecurity posture of a network or system. By utilizing ensemble learning, feature importance analysis, and flexibility to accommodate a broad variety of patterns, Random Forest is positioned as a reliable and effective way to identify dangers that are both known and unknown.The ability to handle unbalanced datasets, fend off overfitting, achieve highaccuracy, and produce few false positives. emphasizessignificant challenges in intrusion-detection.

REFERENCES:

1. Mari, Andrei-Grigore, Daniel Zinca, and Virgil Dobrota. "Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network." *Sensors* 23.3 (2023): 1315.
2. Alqahtani, Hamed, et al. "Cyber intrusion detection using machine learning classification techniques." *Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1*. Springer Singapore, 2020.
3. Panda, Mrutyunjaya, et al. "Network intrusion detection system: A machine learning approach." *Intelligent Decision Technologies* 5.4 (2011): 347-356.
4. Biswas, Saroj Kr. "Intrusion detection using machine learning: A comparison study." *International Journal of pure and applied mathematics* 118.19 (2018): 101-114.

5. Sangkatsanee, Phurivit, Naruemon Wattanapongsakorn, and Chalermpol Charnsripinyo. "Practical real-time intrusion detection using machine learning approaches." *Computer Communications* 34.18 (2011): 2227-2235.
6. Halimaa, Anish, and K. Sundarakantham. "Machine learning based intrusion detection system." *2019 3rd International conference on trends in electronics and informatics (ICOEI)*. IEEE, 2019.
7. Musa, Usman Shuaibu, et al. "A review on intrusion detection system using machine learning techniques." *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, 2021.
8. Buczak, A.L.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1153– 1176.
9. Mishra, P.; Varadharajan, V.; Tupakula, U.; Pilli, E.S. A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 686–728.
10. Dhanabal, L.; Shantharajah, S.P. A Study on NSL-KDD Dataset for Intrusion Detection system Based on Classification Algorithms. *Int. J. Adv. Res. Comput. Commun. Eng.* **2015**, *4*, 446–452.