INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 08 ISSUE: 05 | MAY - 2024 SJIF RATING: 8.448 ISSN: 2582-3930

Network Intrusion Detection and Anomaly Classification

Swayam Manori(Student),
Shreya Rawat(Student),
Mrs. Preeti Chaudhary(Mentor)
Graphic Era Hill University
Dehradun,Uttrakhand, India

Abstract

A software application known as an intrusion detection system (IDS) is utilized to monitor networks for any unauthorized or malicious activities that may violate security policies related to system confidentiality, integrity, and availability. Our thesis involved conducting extensive literature reviews on various types of IDS, anomaly detection methods, and machine learning algorithms that can be utilized for detection and classification. IDS are crucial for safeguarding advanced communication networks and were initially designed to identify specific patterns, signatures, and rule violations. In recent years, machine learning and deep learning approaches have been employed in the field of network intrusion detection, providing promising alternatives. Our proposed system involved identifying noisy features through causal intervention, preserving only those features that had a causality with cyberattacks. The ML algorithm was then utilized to make a preliminary classification to select the most relevant types of cyberattacks, ultimately leading to the detection of unique labeled cyberattacks through the counterfactual detection algorithm.

Keywords—Intrusion Detection System (IDS), anomaly detection, machine learning algorithms, deep learning, causal intervention, cyberattack classification, and counterfactual detection.

I. INTRODUCTION

A network intrusion is any illegal activity carried out on a digital network. Network incursions frequently entail the theft of valuable network resources and virtually always compromise a network security and/or data security. Organizations and their cybersecurity teams must have a comprehensive understanding of how network intrusions operate and implement network intrusion, detection, and response systems that are designed with attack techniques and cover-up methods in mind in order to detect and respond proactively to network intrusions.

With the surge in internet usage, the occurrence of cyber-attacks has also increased. These attacks often involve new methods that necessitate the use of intelligent systems for detection. community intrusion is an illegal intrusion into the digital belongings of a business network. it is performed with the intention of destroying or stealing private facts. attempts are made by using malicious parties to achieve access to the inner structures.

community intrusions include DDoS (distributed denial of service), square injection, and Man-in -the middle(MitM)etc.

A. The important dangers of network intrusion maybe indexed as follows:

1. Corruption of statistics: A massive wide variety of requests or unlawful requests would possibly corrupt the business enterprise's or customers' critical facts. The repute of orders and workflows can also shift, and patron bills may additionally end up behind schedule. during audits, ledgers and tainted economic statistics can exacerbate problems for a enterprise. it is vital for agencies to have a statistics backup.
2. Economic Loss for the organization: to be able to gather the agree with in their customers and stakeholders, a business can also want to offer rewards and incentives. relying at the severity of the attack, they'll also want to coordinate with 0.33-celebration businesses with a view to take care of and mitigate the attack on their behalf. it's also probably that the company gets taxed primarily based at the variety of requests, with the intention to best make subjects worse. If an attack occurs at some stage in the season or throughout income, possible orders are also misplaced, resulting in in addition financial losses. Repairing the broken assets is an extra cost.
3. Robbery of records: one of the maximum preferred belongings for attackers is the personal records of consumers. Their address/location, phone numbers, electronic mail addresses, and even payment information can be exploited through social engineering and other method. In fact, companies with access to cameras and contacts might pose a ways extra risks to their consumers.
4. Operational Disruption: in an effort to recover from the assault, the corporation can also choose to suspend operations and sports until it regains its health, inflicting a substantial postpone in the workflows of operations.
5. lack of popularity: reputation loss may be disastrous for a corporation. lack of customers, an opening for

rivals, a rise in liquidity chance, and the effect in the marketplace and stocks will only make it extra tough for the corporation to recover.

The employer should put into effect middle safety features, teach personnel, assemble firewalls, enable proper authentication and get entry to manipulate, manage passwords, and have records backups.

B. Network Intrusion Detection system(NIDS)

An intrusion detection system (IDS) is employed to analyze network traffic and identify any malicious activity. IDS can generally be categorized into two main types: misuse-based IDS and anomaly-based IDS, which will be discussed in detail in the following section. In short, misuse-based detection looks for known attacks and compares incoming traffic with these known patterns. If a match is found, an alarm is triggered. On the other hand, anomaly-based detection examines incoming traffic for any deviations from normal behavior and flags them as anomalies. To effectively detect new attacks, a substantial amount of data is required to build a model that defines what is considered normal and what is deemed anomalous. This has led to the adoption of supervised machine learning algorithms to efficiently analyze the data and construct predictive models with high accuracy rates for identifying new attacks. However, the high dimensionality of the data poses a challenge, as an increased number of features with a relatively small dataset can result in the "curse of dimensionality" problem, which can impact classification outcomes. Consequently, there has been a growing interest in the application of feature selection and reduction techniques to improve classification performance. Intruders can take the form of actual users or software that infiltrate an organization's resources. Unauthorized logins or illicit acquisition of access rights are common strategies employed by intruders, while software-based intruders can manifest as viruses, worms, or ransomware. Numerous other types of attacks also exist. Undetected intrusions can have severe consequences for governments and businesses. They can compromise national security, lead to financial losses and data breaches, and tarnish the reputation of organizations.

II. BACKGROUND

This paper primarily concentrates on the utilization of supervised learning, which is one of the two types of ML algorithms, the other being unsupervised learning.The computer acquires knowledge from existing input and output data in order to construct a model (program) capable of predicting future output based on fresh input. Deep neural networks are composed of multiple layers that extract features simultaneously. This approach has become increasingly popular and has had a significant impact on the field of science. The existing research in network IDS has shown enhanced detection accuracy across various datasets.

IDSs, similar to antivirus software, firewalls, and access control schemes, serve as security solutions aimed at enhancing the security of information and communication systems. The emergence of IDSs was a direct response to the limitations of conventional security approaches. Research papers on IDS machine learning typically start by using the original DARPA datasets to kickstart machine learning research. Initially, shallow neural networks like Decision Trees, Support Vector Machines, and Random Forests were employed, resulting in impressive accuracy. However, these models often struggled when it came to underrepresented attack types.

In this paper the types of attack we will be focusing on are:

1. Unauthorized Access: Intruders may try using insecure passwords, software or hardware flaws, and weaknesses in networks' security mechanism.
2. Phishing attack: "Phishing" refers to an attempt to steal sensitive information, typically in the form of usernames, passwords, credit card numbers, bank account information or other important data in order to utilize or sell the stolen information.
3. DOS(Denial of service)attack: A Denial-of-Service (DoS) attack is a malicious attempt to disrupt the normal functioning of a targeted server, service, or network by overwhelming it with a flood of internet traffic. The goal of a DoS attack is to make the targeted system or network unavailable to its intended users, causing a denial of service.
   DDOS(Distributed denial of service)attack: In a DDoS attack, multiple compromised computers (often a botnet) are used to launch a coordinated attack against a target. This makes it more challenging to mitigate the attack, as the traffic comes from various sources.
4. MITM(Man-in-the-middle)attack: A Man-in-the-Middle (MitM) attack is a type of cyberattack where an unauthorized entity intercepts and potentially alters the communication between two parties without their knowledge. The attacker positions themselves between the communicating parties, allowing them to eavesdrop on or manipulate the data being exchanged.

III. LITERATURE SURVEY

It is worthwhile to look into training feature sizing optimization. Crucially, irrelevant features in a dataset have the potential to reduce model accuracy and lengthen the amount of training time needed to establish a model. In order to ascertain the ideal training size, a multitude of investigations have been carried out. Reducing the time and space complexity of model development has been accomplished by feature selection, a procedure that involves choosing the most pertinent features manually or using algorithms.

An intrusion detection wrapper feature selection algorithm was presented by Hadeel et al. The feature selection is implemented using a dove-inspired optimizer in this technique, and the binarizing algorithm of the suggested cosine similarity method demonstrated a greater accuracy and a faster convergence speed than the sigmoid method. An additional study created a feature selection model by fusing the BEES and ID3 classifier algorithms. The intended feature subset in this model was generated using the BEES method. In the studies that are currently available, accurate cyberattack classification is just as important as training feature size. All of the current ML- or feature-based NIDS algorithms rely on the correlation between cyberattacks and features to achieve classification. Because there are several erroneous correlations, this association results in multiple incorrect classifications. Causal reasoning is widely used to overcome the false correlations in order to tackle this problem. this paper starts from the decoupling of the correlation of features and the classification of types of cyberattacks under counterfactual scenarios to achieve a high accuracy in the detection of cyberattacks. The counterfactual model is based on the BN, which can model relationships among hundreds of cyberattacks and features.

This chapter will cover the many IDS models that are generated through the use of ensemble-based, feature selection, and machine learning techniques. The intrusion detection system is constructed using a variety of machine learning algorithms. A portion are under supervision, while others are not. Many techniques for detecting intrusions that rely on supervised machine learning algorithms, including Support Vector Machines, Random Forest, Gaussian Naive Bayes, and Logistic Regression (LR). Data are pre-processed to divide it into training and testing sets. A variety of machine learning algorithms are used here, including RBF SVM, Linear SVM, K-Nearest Neighbor, and Logistic Regression. Out of all the techniques, RBF SVM produced the best accuracy. The model's ability to produce a low false alert rate was its primary benefit.

Using the KDD data set, Mr. Subhash Waskle et al. built a model using the PCA and Random Forest techniques. The model was compared with Support Vector Machine, Naive Bayes, and Decision Tree. In comparison to the previous methods, this model produced greater accuracy. This model's low error rate is one of its advantages. Decision Tree and K-Nearest Neighbor algorithm based intrusion detection method was utilized to locate the necessary features.After comparing the two methods, it was discovered that the Decision Tree outperformed K-Nearest Neighbor. Additionally, machine learning techniques are shown to perform against a variety of attacks.Sumaiya Thaseen et al. suggested a model in which the features most dependent on the target class are chosen using the chi-square feature selection method.

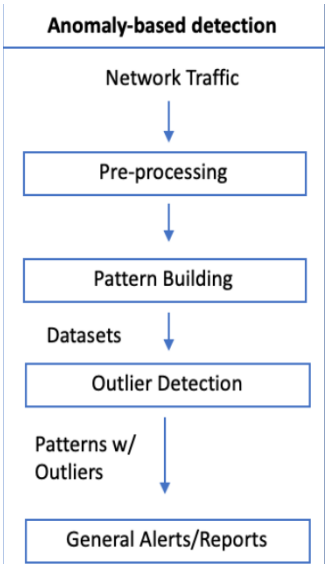
A threshold value of 0.55 was determined. The over-fitting problem was removed from the dataset using a four-step of data pre-processing method. Out of these four techniques Random Forest achieved the highest accuracy but the computational time is higher than J48 and Naive Bayes. M. C. Belavagi et al. They used wrapper- and filter-based techniques for feature selection. There are 41 features in the NSL-KDD dataset. The experiment was run independently on 10 and 41 features from the dataset. While the Bayesian Network outperformed the Decision Tree in the Probe and U2R attacks, the Decision Tree outperformed the R2L and DOS attacks. The model's high detection rate is one of its strengths, but its lengthy computation time is one of its weaknesses.

IV. PROPOSED METHODLOGY

Network intrusion detection system is a program that searches a computer, network, or port for signs of hacker activity, denial-of-service attacks, or port scans. Via its monitoring of network traffic, the NIDS finds unusual patterns in incoming packets, which aids in the detection of these hostile activities. Local, outgoing, and incoming traffic can all be observed by the NIDS. Both incoming and outgoing traffic inspections can provide important information about malicious activity. Certain assaults may start inside the network and remain there, or they may be orchestrated there with an external objective. Additionally, the NIDS collaborates with other systems, such as a firewall, to strengthen defenses against known attack sources. Network intrusion detection systems, or NDIS, are essential tools for spotting and neutralizing security risks in computer networks. Network intrusion detection is accomplished through a variety of strategies and tactics.

1. Signature- based detection –
   This method, often called knowledge-based technique, searches for particular signatures, or combinations of bytes, which almost always indicate bad news. Read: packets delivered by malware in an effort to cause or take advantage of a security breech, or the infection itself. Because the search criterion is so specific, these solutions produce less false positives than anomalous solutions, but they also only cover signatures that are already in the search database, meaning that truly new attacks have a decent chance of succeeding.

2. Anomaly- based detection-
   These solutions, also known as behavior-based ones, monitor activities inside the designated scope in an attempt to detect instances of malicious behavior, at least according to their definition of it. This is a challenging task that occasionally results in false positives. For example, outgoing URLs of Web activity may be taken into account, and websites including specific domains or URL length/contents may be blocked automatically, even when the user is attempting to access the website legitimately for business purposes and the request is being made by a human, not malware.



3. Heuristic- based detection-
Using rules or heuristics to find patterns or behaviors suggestive of possible security risks is known as heuristic-based detection in a Network Intrusion Detection System (NIDS). Heuristic-based detection, in contrast to signature-based detection, which depends on well-known attack patterns, enables a more adaptable and dynamic method of spotting suspicious activity.

4. Protocol analysis-Using rules or heuristics to find patterns or behaviors suggestive of possible security risks is known as heuristic-based detection in a Network Intrusion Detection System (NIDS). Heuristic-based detection, in contrast to signature-based detection, which depends on well-known attack patterns, enables a more adaptable and dynamic method of spotting suspicious activity.

5. Signature-less detection-
Network intrusion detection systems (NIDS) use a technique called signature-less detection that finds possible threats without the use of pre-defined signatures or patterns. Rather, this method concentrates on identifying patterns linked to malicious activity in the network traffic, anomalies, or departures from typical behavior.

6. Flow -based analysis-
In Network Intrusion Detection Systems (NIDS), flow-based analysis looks for possible security risks by analyzing the patterns and features of network traffic. The packet sequence that two devices exchange over a predetermined length of time is represented by a network flow. Understanding the communication patterns and spotting abnormalities or malicious activity are made easier by analyzing these flows.

7. Session tracking-
In Network Intrusion Detection Systems (NIDS), session tracking is the process of keeping an eye on and evaluating the status of active communication sessions between network hosts. This approach facilitates the detection and handling of security risks involving several network interactions.

8. Hybrid approach-
In order to improve intrusion detection's overall efficacy, hybrid approaches in Network Intrusion Detection Systems (NIDS) combine many detection methods or techniques. The goal of this combination is to overcome the shortcomings of individual techniques while utilizing the advantages of several strategies.

V. CONCLUSION

The increasing spike in internet content has led to an increase in cybercrime. The first step in identifying and reporting such assaults is the usage of intrusion detection systems (IDS). Finding distinct attacks is necessary for identifying anomalies, and it's a challenging undertaking. Academics from all around the world are now interested in learning more about this topic and specifically how to use supervised learning algorithms for intrusion detection to get around these challenges.In many cases, selecting the right features is essential to improving performance. Moreover, sampling techniques can assist in resolving data imbalance, which may be a cause for concern. Lastly, deep learning is required for big intrusion detection data sets to get good performance.

Even if machine learning (ML) seeks to make anomaly detection easier, it's crucial to first comprehend the process of detection and specify exactly what our algorithms' intended results are.It is challenging to produce a solid forecast when conventional machine learning algorithms are unable to separate causality from correlation.First, the noisy features could be eliminated and the minimum quantity of training features could be ascertained by creating a causal relationship between events and features by causal intervention. then, to determine the unique label, the ML and counterfactual detection technique were applied.

The severity of new cybersecurity threats is rising right now, and they cannot be categorized using the current schemes. Therefore, a new line of inquiry may be how to efficiently integrate unsupervised learning and causal machine learning to create new NIDs to identify emerging cybersecurity dangers. An additional line of defense against security attacks is an intrusion detection system (NIDS), which scans network data for unusual activity. NDSs must advance because attackers are

always creating new attack techniques. Businesses are always looking for ways to make their NDSs better because assaults can have negative effects on their brand and create financial losses.

As previously mentioned, hybrid approaches show how well integrating several detection techniques can result in a more reliable and strong network intrusion detection system. Organizations can improve their capacity to identify a wide variety of threats, including both known and unknown assaults, by combining signature-based and anomaly-based detection or by utilizing the advantages of heuristics and machine learning. Moreover, a comprehensive picture of network activity is provided by the integration of behavioral analysis, protocol analysis, and flow-based analysis, which enables more precise threat identification.

We used four supervised machine learning methods to develop a model that can identify and classify current and real-world threats in order to address this issue. These supervised algorithms included logistic regression (LR), random forest (RF), linear support vector machine (LSVM), and Gaussian Naive Bayes (GNB). The difficulties NIDS encounters—such as managing encrypted communications, labor-intensive procedures, and the requirement for constant adaptation—highlight the intricacy of the cybersecurity environment. It is essential for enterprises to invest in cutting-edge technology and techniques that can keep up with new risks as they continue to embrace digital transformation. Furthermore, by supplying up-to-date knowledge on recognized harmful actors and their strategies, the incorporation of threat intelligence feeds adds another line of protection.

An NIDS examines network data to look for unusual activity; it offers Using the recall macro-measure is one of the finest ways to assess the dependability of a model built from a dataset with several classes. Recall shows how often a damaging assault is identified by the model, and the macro guarantees that all classes are assessed identically, regardless of size. NIDS must adapt to the ongoing evolution of cyber threats by integrating cutting-edge technologies, advances in machine learning, and flexible methodology. An intrusion detection system's (NIDS) performance is contingent upon its ability to identify and neutralize threats as well as reduce false positives, maximize resource usage, and expedite incident response. As a first line of defense against a variety of cyberthreats, the Network Intrusion Detection System continues to be a mainstay in the cybersecurity environment. The study article presents insights and approaches that enhance the comprehension of Network Intrusion Detection Systems (NIDS) and lay the groundwork for future advancements and enhancements in network security. The ongoing development and modification of NIDS will be essential to guaranteeing the availability, integrity, and security of sensitive data in the linked world as enterprises traverse the always shifting digital landscape.

Examining the effectiveness of deep learning algorithms is one possible future path. Finally, it is vital to think about utilizing the best model when developing an NIDS system. Then, by implementing and evaluating this NIDS system in a network, its predictions can be verified.

REFERENCES

1 Scarfone, K.; Mell, P. Guide to intrusion detection and prevention systems (IDPS). NIST Spec. Publ. 2007, 800, 94

2 Saranya, T.; Sridevi, S.; Deisy, C.; Chung, T.D.; Khan, M.A. Performance analysis of machine learning algorithms in intrusion detection system: A review. Procedia Computer. Sci. 2020, 171, 1251–1260.

3 Sharafaldin , I.; Lashkari, A.H.; Ghorbani, A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp 2018, 1, 108–116.

4 Kinam Park; Youngrok Song; Yun-Gyung Cheong, 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigData Service), Classification of Attack Types for Intrusion Detection Systems Using a Machine Learning Algorithm.

5 Anish Halimaa A, Dr K.Sundarakantham: Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) 978-1-5386-9439-8/19/$31.00 ©2019 IEEE "MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM.

6 1D. Srinivas, 2Prabhakar Kumar Thakur, 3Pramod Kumar, 4K. Sindhu Sri Vani, 5Mr. S. Nirmal Sam © 2023 IJRTI | Volume 8, Issue 4 | ISSN: 2456-3315.

7 Vinayakumar, R.; Alazab, M.; Soman, K.; Poornachandran, P.; Al-Nemrat, A.; Venkatraman, S. Deep learning approach for intelligent intrusion detection system. IEEE Access 2019, 7, 41525–41550.

8 Ansam Khraisat, Iqbal Gondal, Peter Vamplew, and Joarder Kamruzzaman. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity, 2(1):20, 2019.

9 Nevrus Kaja, Adnan Shaout, and Di Ma. An intelligent intrusion detection system. Applied Intelligence, 49(9):3235–3247, 2019.

10 E. Tsukerman, Designing a MACHINE LEARNING Intrusion Detection System: Defend Your Network from Cybersecurity AreatsApress, NY, USA, 2020.

11 M. A. Ferrag, L. Maglaras, S. Moschoyiannis et al., "Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study," Journal of Information Security and Applications, vol. 50, Article ID 102419, 2020.

12 J. Zhang, L. Yu, F. Xingbing, X. Yang, X. Gang, and Z. Rui, "Model of the intrusion detection system based on the integration of spatial-temporal features," Computers & Security, vol. 89, 2020.

13 Hammad, M., El-medany, W., & Ismail, Y. (2020, December). Intrusion Detection System using Feature Selection With Clustering and Classification Machine Learning Algorithms on the UNSW-NB15 dataset. In 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT) (pp. 1-6). IEEE.

14 Malhotra, H., & Sharma, P. (2019). Intrusion Detection using Machine Learning and Feature Selection. International Journal of Computer Network & Information Security, 11(4).

15 Uma, M., & Padmavathi, G. (2013). A Survey on Various Cyber Attacks and their Classification. IJ Network Security, 15(5), 390-396