© 2025 IJRTI | Volume 10, Issue 1 January 2025 | ISSN: 2456-3315

Network Intrusion Detection System Using Machine Learning

Fayas Muhammed[1], Johann Dominic Thomas[2],Akash Biju[3], Adithyan A S[4],Andrew Sebi Varghese[5], Charls P S[6]
[1-5]Student, [6]Assistant Professor Department of
CSE
College of Engineering, Kidangoor
fayasfayaz707@gmail.com[1], johannthomaspala@gmail.com[2], akashbiju369@gmail.com[3],
adithyanas89@gmail.com[4], andrewssebivarghese@gmail.com[5], charls145@gmail.com[6]

***Abstract*—Network Intrusion Detection Systems (NIDS) are vital in cybersecurity due to the increasing complexity of cyber attacks. This survey reviews machine learning approaches in NIDS, examining their development, current status, and future directions. We categorise and evaluate traditional algorithms, deep learning methods, and hybrid approaches, discussing key datasets, feature selection, and performance metrics. The study addresses challenges such as class imbalance, high false positive rates, and real-time detection, while also exploring trends like federated learning, transfer learning, and explainable AI within NIDS. Despite promising results, ML-based NIDS face challenges in achieving optimal performance in dynamic network environments, presenting areas for further research and enhancement.**

KEYWORDS

Network Intrusion Detection Systems (NIDS); Machine Learning; Deep Learning; Cybersecurity; Feature Selection; Anomaly Detection; Network Security; Classification; Supervised Learning; Unsupervised Learning.

I. INTRODUCTION

Almost all of the people use the internet to carry out essential activities such as bill payment, bank transfer, etc. But attacks towards home networks are not uncommon nowadays as everybody is connected to each other through the internet and the attack has been growing more frequent and severe. When an attack does occur, it is essential that a comprehensive and organised analysis is conducted to verify the causes of the attack and the damages of the attack.

A thorough and timely investigation and response can serve to minimise network downtime and ensure that critical business systems are maintained in full operation. The level of connectivity worldwide has provided opportunities for cybercriminals who make a living breaking into networks, as well as amateur hackers who have too much time on their hands. The determined hacker can find a way into your network either by establishing some type of connection and entering your virtual ”front door” or by using social engineering tactics to obtain user ID and password information. Whatever the method used, the fact is that an intruder can get into your network and harm your business.

II. LITERATURE SURVEY

A. Paper 1

This paper provides a comprehensive analysis of Network Intrusion Detection Systems (NIDS) enhanced by machine learning techniques, highlighting their evolution and current trends. It categorises NIDS into three main types: Network Intrusion Detection Systems, Host Intrusion Detection Systems, and Hybrid Systems, each employing either signature-based or anomaly-based detection methods. ...

B. Paper 2

This paper by Markus Goldstein and Seiichi Uchida presents a comparative evaluation of various unsupervised anomaly detection algorithms specifically designed for mul tivariate data. The authors categorise anomalies into sev eral types, including local, micro-cluster, point, collective, and contextual anomalies, emphasising the need for appro priate detection methods tailored to each type. They review a range of algorithms, including nearest neighbour-based techniques, clustering methods, statistical algorithms, and neural networks, detailing their mechanisms and applica tions. The paper highlights key algorithms such as k-NN for global anomaly detection, Local Outlier Factor (LOF) for local anomalies, and One-Class Support Vector Machines (SVM) for boundary-based anomaly detection. Findings indicate that while histogram-based methods (HBOS) ex cel in real-time applications, nearest neighbour algorithms generally outperform clustering methods in practice. The authors conclude that local anomaly detection algorithms struggle with datasets exhibiting global anomalies, often resulting in high false positive rates, while recommend ing nearest neighbour techniques for their computational efficiency in suitable contexts. The study underscores the importance of selecting the right algorithm based on the na ture of the anomalies and the characteristics of the dataset.

C. Paper 3

In this paper, Pratik Tonge and colleagues propose a Network Intrusion Detection System (NIDS) utilising machine learning techniques to enhance network security against rising cyber threats. The authors emphasise the necessity of effective intrusion detection due to the increas ing sophistication of attacks facilitated by advancements in computing technology. They employ an Artificial Neu ral Network (ANN) for classifying network traffic as either malicious or benign, utilising the NSL-KDD dataset for training and testing the model. The methodology involves feature selection to identify relevant attributes from net work traffic, which are then used to train the ANN model. The system captures network packets via Wireshark, filter ing them to analyse and predict the type of attack occurring in real-time. The authors highlight the potential for their approach to improve detection accuracy and reduce false positive rates compared to existing methods. The study concludes that integrating machine learning algorithms into intrusion detection systems can significantly enhance their efficiency and effectiveness, thereby providing a robust solution for safeguarding network infrastructures.

D. Paper 4

In this review paper, Syeda Khubroo Hashmi and col leagues examine the importance of feature selection tech niques in the development of machine learning models for Intrusion Detection Systems (IDS). The authors highlight that effective feature essential for engineering reducing is dataset dimensions, which in turn minimises training time and computational costs while enhancing model perfor mance and detection accuracy. They discuss the growing vulnerability of information systems to cyber threats and the need for robust IDS solutions that can monitor and analyse network traffic for unauthorised activities.The pa per categorises feature selection methods into two main approaches: feature extraction and feature selection. Fea ture extraction involves creating new features from existing ones, while feature selection focuses on selecting a subset of relevant features from the original dataset. The authors detail various techniques, feature selection including In formation Gain, correlation-based methods, and ensemble approaches, emphasising their effectiveness in improving the performance of IDS.

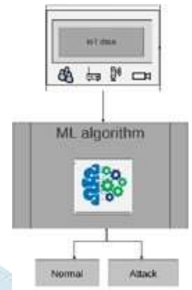

Fig. 1.

E. Paper 5

The proposed IDS employs both signature-based and anomaly-detection methods. Signature detection uses pre defined signatures to identify known threats, while anomaly detection relies on heuristic methods to detect unknown attack behaviours by recognizing deviations from normal traffic patterns. The study primarily focuses on anomaly detection using machine learning algorithms such as Sup port Vector Machines (SVM), Decision Trees, Forests.The and Random research utilises the CICDDoS 2019 dataset, which includes 80 features extracted from network traffic, comprising both benign and DDoS attack instances. The dataset is publicly available and provides comprehensive labelling for effective training and testing.

F. Paper 6

CSE-CIC-IDS2018, the latest iteration of intrusion detection datasets, comprises approximately 16 million instances collected over ten days, encompassing a diverse array of attack types. This dataset is characterised by a multi-class structure and a notable class imbalance, with only about 17representing attack traffic. This imbalance poses challenges for machine learning models, as it can lead to biassed results if not adequately addressed. Our literature survey reveals critical insights into the current state of research utilising this dataset, highlighting several key findings.First, we observed that the performance scores reported across various studies are unexpectedly high, raising concerns about potential overfitting. Many studies achieved perfect or near-perfect accuracy, precision, and recall scores, which may not accurately reflect the models' generalisation capabilities. Second, a majority of the surveyed works failed to adequately address the implications of class imbalance, which can significantly skew results in big data contexts. Lastly, our analysis uncovered a lack of transparency regarding data cleaning processes in CSE-CIC-IDS2018, the which use of raises concerns about the reproducibility of experiments.

G. *Paper 7*

The study specifically compares Snort and Suri cata, both of which are open-source and share similar functionalities, detection rule sets, and syntax. Snort, es tablished in 1998, has become a standard in IDS technology, while Suricata, developed in 2010, boasts a multi-threaded architecture that enhances its scalability and performance. The choice of these two IDSs stems from their comparable capabilities and the need to evaluate their effectiveness in high-speed network environments.The findings of the experiments demonstrate that Snort, despite its higher accuracy, triggered a substantial number of false positive alarms, averaging 55.2mitigate this, the study proposes a Snort adaptive plug-in that utilises machine learning algorithms, particularly a hybrid model combining Support Vector Machine (SVM) and Fuzzy Logic, which yielded improved detection results. The optimised SVM, enhanced by a firefly algorithm, achieved notable performance metrics with a false positive rate (FPR) of 8.6and a false negative rate (FNR) of 2.2.

H. *Paper 8*

This study contributes to the field by demonstrat ing the potential of SDN in developing an IDPS that can effectively detect and prevent port-scanning and DoS attacks. The results show that the implemented IDPS can quickly detect and stop these attacks in real-time, maintaining a low rate of false positives. The proposed design and detection methodologies can also be expanded to defend against a wide range of cyberattacks.The research on SDN-based defence mechanisms has primarily focused on improving security using the OpenFlow protocol, which is the most popular SDN protocol today. However, there is a need for further exploration of other SDN protocols and their potential for enhancing network security. Addi tionally, the study highlights the importance of addressing compatibility issues between different SDN switches and controllers, which can complicate the implementation of security improvements.

I. *Paper 9*

The survey presents a comprehensive review of the state of research in multi-step attack detection, tracing its evolution from early recognition of the importance of linking multiple events to detect attacks. The authors note that while some foundational work has been done in this area, there has not been a dedicated survey that encapsulates the breadth of methodologies developed specifically for detection. They multi-step attack empha sise the significance of detection methods that focus on real attack traces rather than hypothetical attack paths.The methodology section details the systematic search process used to identify relevant literature, ensuring a rigorous and reproducible approach to the review. The authors provide bibliometric analyses, including publication trends and citation metrics, to contextualise the research landscape. The core of the survey categorises the reviewed methods into f ive main approaches: similarity-based, causal corre lation, structural-based, case-based, and mixed methods. Each method is explained and contextualised within the broader field of multi-step attack detection.The survey also addresses challenges faced in the detection of multi-step attacks, such as the complexity of network data, the rarity of dangerous attacks in datasets, and issues related to reproducibility and dataset structure. The authors discuss these challenges in detail and provide insights into potential future directions for research in this area.

J. *Paper 10*

This study investigates the implementation of an Intrusion Detection and Prevention System (IDPS) util ising Software-Defined Networking (SDN) technology to enhance real-time defence against cyberattacks, particu larly focusing on port-scanning and Denial of Service (DoS) attacks. Given the rising complexity and frequency of cyber threats targeting a wide range of entities—from home users to critical infrastructure—early detection malicious activities of has become paramount to prevent significant damage and unauthorised access to sensitive data.This work represents a significant advancement in the field of network security, showcasing how SDN can be leveraged to develop effective IDPS solutions capable of addressing contemporary cyber threats. The f indings underscore the importance of continued research into the integration of SDN technologies with advanced detection methodologies to improve cybersecurity defences across various domains.

K. *Paper 11*

This paper presents a novel approach to intru sion detection in Software-Defined Networking (SDN) through the development of a Kangaroo-based Intrusion Detection System (KIDS). As SDN architecture evolves to address the limitations of traditional networks, it simulta neously faces various security threats and vulnerabilities. The authors propose KIDS as an effective solution to enhance security in SDN by leveraging its unique features programmability, such f lexibility, as and manageabil ity.This work contributes to the field of network security by proposing an innovative approach to intrusion detection in SDN, addressing scalability and efficiency challenges while enhancing the overall security posture of SDN networks. The authors advocate for the continued exploration of IDS solutions that leverage the unique capabilities of SDN to combat evolving cyber threats effectively.

L. Paper 12

This paper addresses the challenges of intrusion detection in cloud computing systems, particularly in the context of large-scale networks where traditional intrusion detection systems (IDS) may struggle to keep up with the volume of traffic. The authors propose a novel traffic sampling strategy tailored for Software Defined Networking (SDN) that optimises the use of IDS resources while ensuring that the total sampled traffic remains manageable within the processing capacity of the IDS.The proposed approach involves formulating an optimization problem to determine the ideal sampling rate for each switch in the SDN environment. By lever aging SDN's capabilities, the authors aim to sample traffic flows effectively, ensuring that malicious traffic is adequately inspected without overwhelming the IDS.The methodology involves measuring traffic characteristics and dynamically adjusting sampling rates based on the current network conditions and the distribution of malicious traffic.

Paper ID	ML Algorithm	Dataset Used
[1]	Unspecified	Industrial ICS
[2]	k-NN, LOF, SVM	Multivariate Data
[3]	ANN	NSL-KDD
[4]	Ensemble, InfoGain	Generic
[5]	SVM, Decision Tree	CICDDoS 2019
[6]	Mixed	CSE-CIC-IDS2018
[7]	SVM, Fuzzy Logic	Short Data
[8]	Random Forest	OpenFlow SDN
[9]	Similarity-Based	Unspecified
[10]	Unspecified	SDN Traffic
[11]	Kangaroo-Based	SON
[12]	Traffic Sampling	SON

TABLE I
TABLE SHOWING ML ALGORITHMS AND DATASETS USED IN DIFFERENT PAPERS

Paper ID	Advantages	Disadvatages
[1]	Industrial focus	Limited dataset
[2]	Real time detection	High false positives
[3]	High accuracy	Needs dataset updates
[4]	Improved performance	High computation cost
[5]	DDOS detection	Imbalance in data
[6]	Big data support	Overfitting risk
[7]	High scalability	False positives
[8]	SDN compatibility	Protocol limitations
[9]	Multi-step detection	Data complexity
[10]	Real-time prevention	High implementation cost
[11]	Programmable optimization	Scalability issues
[12]	Resource optimization	Sampling accuracy

TABLE II
ADVANTAGES AND DISADVATAGES OF DIFFERENT PAPERS

III. CONCLUSION

Thesurvey on Network Intrusion Detection Systems (NIDS) utilizing machine learning techniques highlights the signif icant advancements and ongoing challenges

Paper ID	Training Accuracy (%)	Test Accuracy (%)
[1]	88	85
[2]	85	80
[3]	90	88
[4]	92	89
[5]	87	84
[6]	95	93
[7]	88	85
[8]	89	86
[9]	83	80
[10]	86	82
[11]	91	89
[12]	93	90

TABLE III
TRAINING AND TEST ACCURACY OF REVIEWED PAPERS

in the field of cybersecurity. As cyber threats become increasingly sophis ticated, the integration of machine learning into NIDS has emerged as a promising approach to enhance detection ac curacy and efficiency. The reviewed literature indicates a clear evolution in the methodologies employed, from tradi tional signature-based systems to more dynamic anomaly detection techniques that leverage the power of machine learning and deep learning. In conclusion, while machine learning-based NIDS show great promise in addressing the complexities of modern network security, ongoing research is essential to overcome existing limitations and to develop systems that can effectively adapt to the rapidly changing landscape of cyber threats. The insights gained from this survey serve as a foundation for future investigations aimed at creating more resilient and efficient intrusion detection systems

IV. REFERENCES

1) R. Plaka, "Intrusion Detection Using Machine Learning for Industrial Control Systems," Ma¨lardalen University, Va¨stera˚s, Sweden. Supervisor: S. Punnekkat; Company Supervisor: A. Balador, RISE- Research Institutes of Sweden.
2) M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data."
3) P. Tonge, N. Jiwane, and L. N. Yadav, "Network Intrusion Detection System Using Machine Learning Techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 12, no. 5, pp. 5594, April 2024, doi: 10.15680/IJIRCCE.2023.1205078.
4) S. K. Hashmi, G. P. Dubey, and P. Himthani, "A Review on Feature Selection Techniques for Intru-

sion Detection System," *SSRN Electronic Journal*, Dec. 2021, doi: 10.2139/ssrn.4289276.

5) A. Meliboea, "IoT Network Intrusion Detection System Using Machine Learning Techniques," *QO'QON Universiteti Xabarnomasi*, vol. 11, pp. 112–115, June 2024, doi: 10.54613/ku.v11i11.972.
6) J. L. Levy and T. M. Khoshgoftaar, "A Survey and Analysis of Intrusion Detection Models Based on CSE-CICIDS2018 Big Data," *Journal of Big Data*, vol. 7, article no. 104, 2020.
7) S. A. Raza Shah and B. Issac, "Performance Comparison of Intrusion Detection Systems and Application of Machine Learning to Snort System."
8) C. Birkinshaw, E. Rouka, and V. G. Vassilakis, "Implementing an Intrusion Detection and Prevention System Using Software-Defined Networking: Defending Against Port-Scanning and Denial-of-Service Attacks."
9) J. Navarro, A. Deruyver, and P. Parrend, "A Systematic Survey on Multi-Step Attack Detection."
10) C. Birkinshaw, E. Rouka, and V. G. Vassilakis, "Implementing an Intrusion Detection and Prevention System Using Software-Defined Networking: Defending Against Port-Scanning and Denial-of-Service Attacks."
11) A. Yazdinejad, R. M. Parizi, A. Dehghantanha, and M. S. Khan, "A Kangaroo-Based Intrusion Detection System on Software-Defined Networks."
12) T. Ha, S. Kim, N. An, J. Narantuya, C. Jeong, and H. Lim, "Suspicious Traffic Sampling for Intrusion Detection in Software-Defined Networks."