Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper 82

Comparative Analysis of Intrusion Detection System and Machine Learning Approches

Krupali Gosai, Harsh Mehta
Department of Computer Engineering
Marwadi University
Rajkot, India
krupaligosai17@gmail.com, harsh.mehta@marwadieducation.edu.in

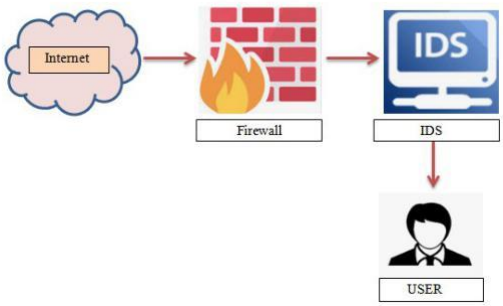
***Abstract*— In the IT industry, cyber security is becoming increasingly important. Due to the rise of the infinite communication paradigm and the extended spectrum of communication technologies, there has been a rising worry about cyber security in recent years, which aims to secure either the data or the system's communication technology. Digital gadgets that are connected to one other. A important component of network security is the Intrusion Detection System (IDS). Intrusion detection systems based on Machine Learning techniques have lately piqued the interest of researchers. IDS is a software or hardware device that collects and analyses security symptoms from a number of system and network sources in order to identify and respond to assaults. This research looks at Intrusion Detection Systems in general, as well as the different datasets and machine learning approaches that are commonly utilized to create IDS systems. It is flourishing results are obtained in a number of the studies given within the literature.**

***Keywords*— IDS, Machine Learning, Network Security, Cyber Security**

I. INTRODUCTION

Network security is a must-have condition in today's environment. An Intrusion Detection System (IDS) is a countermeasure for detecting a sequence of intrusions that threaten data sources familiarity, availability, and integrity. Which monitors the network for unusual behavior and issues an alert if it is identified? An intrusion detection system (IDS) is a piece of software that analyses network data in order to defend it against assault or infiltration. The Intrusion Detection System keeps track of the system's operations and looks for any suspicious activity.

Fig. 1.Structure of IDS



Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper 82

are passing from firewall and IDS. So if any malicious traffic received from user side then IDS system gives an alert and it continuous monitor that system throughout the communications.

II. IDS

IDS system continuous monitor the activity of network and host activity to identified normal traffic as well as suspicious activity.
User received filter packets from firewall and IDS system .If any suspicious activity found than it gives alert to the host and administrators. It gathered all traffic data centrally. Many techniques and algorithms are used to evaluate that data because that data is in large amount. IDS system simplify by its component and functionality .That things are followed below:
Components of IDS

IDS having mainly three components they are as following:

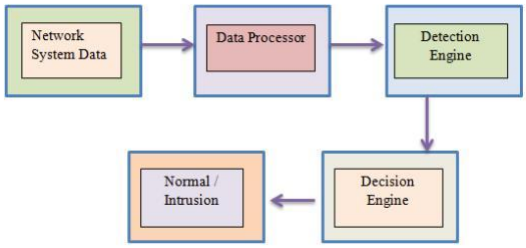


Fig. 2.Components of IDS

1. Data Pre-processor: In this part capture network and system data are processing and generate processed data.

2. Detection Engine: Captured data analyzed by detection model combine in IDS.

3. Decision Engine: Based on decision table it responds alerts for captured data.

Mainly in any system connecting with Internet, it facing both type of traffics like normal and malicious. In Fig.1.Clearly shows that user connecting with Internet that packets

Fig. 3.Function of IDS

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper 82

Data Collection: This part gathers data by keeping an eye on the system and network. It maintained track of the actions of both the host and the network. All of the protocol and parameter information can be found here.

Feature Selection: It requires feature selection since it works with massive volumes of network and host data. Data may be filtered out and processed in a simple and effective manner based on feature selection. Network and host system characteristics such as protocols, Source IP, and Destination IP are used to choose features.

Analysis: Data must be analysed in order to discover infiltration from the collected data. Rule-based IDS relies on specified rules to monitor traffic patterns and behaviour. Another anomaly-based IDS is constructed to identify deceptive behaviour.

Action: IDS was primarily concerned with determining if the situation was normal or abnormal. After analysing problematic activities, it is required to take action.

III. BACKGROUND

In the early 1960s, the financial system required an audit procedure for financial security, in which data was examined to ensure its integrity. Based on this concept, after evaluating the digital world in 1980, a technique to monitor network and system activity is required. Dorothy Denning and Peter Neumann created the first prototype model for an intrusion detection expert system between 1984 and 1986. It operates on the basis of a hypothesis with a matching pattern. In 1980, James P. Anderson et al. [1] pioneered automated IDS in the field of information security. Following this study the first intrusion detection model was created. Following the expansion of digital devices and network area, IDS requires certain enhancement. Computer networks have blossomed since the introduction of the Internet, and the original IDS system only alerted on known assaults. This IDS system is unable to detect zero-day attacks. In 1990, the IDS system was enhanced to identify a growing number of network threats. Threat stack plays a significant part in the IDS community's evolution, and they're focusing on cloud advancements and the finest approaches for dealing with threats in big and dispersed environments. Complete analyst training, strategic hardware deployment, and a comprehensive security plan are all necessary to attain this aim. The tools we need to reach this goal every day. Data is acquired from routers, the hostcomputer, firewalls, virus scanners, and an Intrusion Detection System (IDS) intended specifically to identify known attacks.

IV. LITERATURE SURVEY

An intrusion detection system guards data and computer networks against hostile assaults by monitoring a huge volume of network traffic data. To categories predictable and suspicious acts, a rapid and effective classification algorithm is necessary. We employed the Naive Bayes, Random Forest, and J48 classification algorithms with results for TPR(True Positive Rate), F-measure, accuracy FPR(False Positive Rate), and recall parameters according to D. Kapil et al. [2]. As a consequence, it was observed that Random Forest outperforms in terms of accuracy.

In R. Kumar et al. [3] completed three phases of work, the first of which was normalization, for which they used 41 features from the KDD- 99 dataset. They use PART(Partial Decision Tree), Naive Bayes(NB), and Adaptive Boost classifiers in the second stage to do feature selection utilising Entropy-

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7- Issue 1, January 2024
Paper 82

based analysis as a filter method to decide satisfactory factors. They also apply Ensemble Approach in the subsequent analysis. In the third study, instead of utilising all 41 features and performing the experiment with Naive Bayes, PART, and Adaptive Boost and reviewing the results, the Ensemble technique is employed by voting to choose some of the best components. Ensemble approaches, according to the data, outperform other classifiers on average.

Machine Learning and Neural Network techniques were used to identify DoS/DDoS attacks using the CIC IDS 2017 dataset. S. Wankhede et al. [4] offered work that largely employed Random Forest and Multi Layer Perception to identify these attacks, with the Random Forest approach being proven to be more accurate than the Multi Layer Perception methodology. The Multi-Layer Perception has a greater accuracy of 98.89 % and a % training record, according to the data. The RF algorithm, on the other hand, has a 99.96 % and only requires 80 percent training.

Ali H.Mirza et al. [5] have proposed that they employ three distinct types of classifiers to improve overall performance: neural networks, decision trees, and logistic regression. Following that, we may use ensemble learning to improve the overall performance of the intrusion detection algorithm. A series of studies using the KDD Cup 99 data set and a weighted majority voting mechanism indicated a considerable boost in accuracy for an ensemble learning approach for computer network intrusion detection. And the study found that Ensemble Learning performed well in datasets with a lot of anomalies.

Implementing the system with an anomaly detection learning system [6] is essential to boost the system's adaptability. Deep learning, according to G. Karatas et al. [6], is especially useful for dealing with Big data since it requires minimal training time and gives high accuracy. It also looked at how deep learning techniques are employed in IDS.

Many out-of-date databases only find known attacks. To detect previously unknown attack patterns, both supervisedand unsupervised machine learning algorithms are used. The Kyoto dataset was utilized as a result. F. Salo et al. [7] suggested work on clustering enabled categorization and a classification model built on that cluster to identify density regions that are either normal or anomalous. The findings of the trials in this suggested study demonstrated that clustering was successful in identifying hidden threats.

K. Rani et al. [8] argued that it is vital to adopt an efficient strategy and assess network data features as well as importance-based selection in order to improve IDS performance. For feature selection, Random Forest classifiers were employed, followed by four different types of classifiers for attribute selection, and the results were examined on NSL-KDD using machine learning classifiers such as Naive Bayes, Decision Trees, k-NN, and Logistic Regression. Improved prediction was achieved by using an efficient feature selection strategy before submitting it to any machine learning algorithm.

Balan, S et al. proposed in [9] that they used Random Forest Regression on the CICIDS2017 dataset to evaluate the patterns of network attack flows from the SSH and FTP server. And, by employing realistic traffic characteristics, locate outliers in the detection and deliver excellent accuracy.

Intrusion detection based on network flow data is known as flow-based intrusion detection. S. Zwane et al. [10] developed a flow-based IDS for ensemble categorization of network flow data by analysing machine learning methods. Using the flow-based IDS evaluation CIDDS-001 datasets and the ensemble techniques adaptive boosting, bootstrap aggregation, random forests, and majority voting, the performance of the ensemble of decision tree, probabilistic, and non-probabilistic classification methods

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7- Issue 1, January 2024
Paper 82

was examined. The findings imply that Decision Tree- based approaches for flow-based intrusion detection systems can outperform other methods.

To translate requests into vectors and subsequently train models, machine learning based on ensembles and natural language processing has recently been applied. S. Das et al.

[11] proposed an approach that was tested on datset mentioned in table. And the data suggest that NLPIDS excelled with a 99.96 % detection rate.

Many supervised machine learning approaches can only detect known assaults and are unable to recognise patterns that are unknown. To identify abnormalities, M. Verkerken et al. [12] proposed employing flow-based characteristics. On the basis of complexity and classification performance, four unsupervised techniques were evaluated: Principal Components Analysis, One-Class SVM, Isolation Forest, and Auto encoder. Finally, it was determined that unsupervised algorithms are utilised in situations where all of the available ways are successful in identifying assaults that are undetectable, which is challenging with supervised procedures.

M. S. Abirami et al. [13] introduced a unique IDS that employs a heuristic method, followed by an ensemble approach, and ultimately voting techniques on datasets to correctly and efficiently identify a variety of threats. The experimental results for the NSL-KDD dataset are promising, with classification accuracy of 99.81 percent, 99.8% Detection Rate, and 0.08 percent False Alarm Rate with a subset of 10 features, and the obtained results for the AWID dataset provide accuracy of 99.52 percent and 0.15 percent FAR with only 8 features.

X. Shi et al. [14] propose that, in order to improve accuracy, feature selection can be used to optimise processing efficiency. Accuracy and flexibility should be increased as well. Bagging is used to improve the extreme trees model and maximise the advantages of the upgraded extreme trees and Quadratic Discriminant Analysis in order to acquire the learning outcomes. When extreme trees and ensemble learning are combined, the model has a higher accuracy rate and takes less time to train and evaluate.

S. Seth et al. [15] offered stream-oriented learning for adopting the idea Drift for real-world intrusion detection. The CIC –IDS 2018 dataset is employed with the aid of the Adaptive Random Forest classifier, which offers an accuracy result of 99.5 % and a recall rate of 99.8 %.

The most important aspect is identifying unknown assaults. B. S. Bhati et al. [16] suggested ensemble-based IDS utilising XGBoost based on tree boosting machine learning algorithm using KDDCup99 datasets, with 99.95 percent accuracy.

V. COMPARATIVE ANALYSIS

The analysis of research publications on the IDS is summarised in Table 1.

TABLE I. COMPARATIVE ANALYSIS OF INTRUSION DETECTION TECHNIQUES

No.	Dataset	Algorithms	Result (Accuracy)
[2]	NSL-KDD	Naïve Bayes,	Naïve Bayes

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper 82

		Random Forest, J48 Decision tree	= 81% Random Forest =98.7% J48 = 98.5%
[3]	KDD-99	Ensemble Approach, Naïve Bayes, Adaptive Boost, PART	Naïve Bayes = 91.98% PART = 99.96% Adaptive Boost = 97.86% Ensemble Approach = 99.97%
[4]	CIC IDS 2017	RF, MLP	With 80% training records , MLP = 98.89% RF = 99.96%
[5]	KDD Cup 99	Logistic Regression, Decision Trees(DT), Neural Networks, Ensemble Learning Classifiers	Ensemble Learning= 97.53% DT = 92.08% Neural Network = 90.67% Logistic Regression = 96.66%
[6]	KDD Cup99, NSL-KDD, CIC IDS 2017, CSE-CIC-IDS2018, MCFP Bot Traffic Merged with Benign	DL algorithm	-
[7]	Kyoto	Quadratic discriminant analysis(QDA) , SVM, *k*-NN, RF, Clustering Method	SVM = 98.24% k-NN(k=5) = 98.56% RF = 99.98% QDA = 93.49%
[8]	NSL-KDD	Decision Tree, Naïve Bayes, Random Forest , k-Nearest Neighbour , Logistic Regression,	Naïve Bayes = 95.58% Decision Tree = 99.95% k-NN = 99.76% Logistic Regression = 97.30%
[9]	CICDS2017	Random Forest Regression	RF Regression =99.9%
[10]	CIDDS-001	Naive Bayes, Decision Tree, SVM	Decision Tree = 99.09% Naïve Bayes = 60.56% SVM = 62.9%
[11]	HTTP DATASET CSIC 2010	SVM, Logistic Regression, Naive Bayes with Gaussian	SVM = 99.92%, Ens SVM=99.96%

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 7- Issue 1, January 2024
Paper 82

		function (NB), Neural Networks, Decision Tree (DT)	NB = 98.55% Ens NB = 99.89%
[12]	CIC-IDS-2017	Principal Components Analysis, Isolation-Forest, One-Class Support Vector Machine, Auto-Encoder	The precision of PCA is 93.7%, Isolation Forest is 95.84%, One-Class SVM is 91.04%, Auto-Encoder is 94.59%
[13]	NSL-KDD, AWID, and CIC-IDS2017	C4.5, RF, Forest PA	95%
[14]	KDD CUP 99 dataset, UNSW-NB15	Quadratic discriminant analysis , Extra-Trees	New model accuracy on KDD is 92.88% and on UNSW-NB15 having 92.45%
[15]	CIC-IDS 2018	Drift Detection with Adaptive Random Forest and Adwin(Adaptive Windowing)	Adaptive Random Forest (ADWIN) = 99.5%
[16]	KDDCup99	XGBoost	XGBoost = 99.95%

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper 82

CONCLUSION

We have completed our research on IDS in this article. This study provides a comprehensive overview of IDS and their functions, as well as an examination of several machine learning methodologies. With new machine learning and deep learning concepts, it can assist improve existing IDS systems. The number of computer networks and network applications is continually growing. It is necessary to enhance their safety. Intelligent and efficient IDS are required due to security concerns. We examined several IDS approaches utilizing a variety of works of literature. Diverse datasets are merged with various machine learning algorithms to increase IDS accuracy and cope with real-time IDS data. Where we discovered that the Ensemble technique outperforms other classifiers in terms of accuracy. We also looked at how ancient IDS datasets like KDDCup99 and NSL-KDD provide low-performance IDS, and how proposed solutions sometimes failed to handle hidden attack traffic in real-time. By making some recommendations and assessing research for future ensemble learning directions and enhancing IDS performance, we want to shed some light on this topic.

REFERENCES

[1] Anderson, "Computer security threat monitoring and surveillance," *Tech. Rep. James P Anderson Co Fort Washingt. Pa*, 1980, doi: citeulike-article-id:592588.

[2] R. Kumar Singh Gautam and E. A. Doegar, "An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms," 2018,doi: 10.1109/CONFLUENCE.2018.8442693.

[3] D. Kapil, N. Mehra, A. Gupta, S. Maurya, and A. Sharma, "Network Security: Threat Model, Attacks, and IDS Using Machine Learning," 2021, doi: 10.1109/ICAIS50930.2021.9395884.

[4] S. Wankhede and D. Kshirsagar, "DoS Attack Detection Using Machine Learning and Neural Network," 2018, doi: 10.1109/ICCUBEA.2018.8697702.

[5] A. H. Mirza, "Computer network intrusion detection using various classifiers and ensemble learning," 2018, doi: 10.1109/SIU.2018.8404704.

[6] G. Karatas, O. Demir, and O. K. Sahingoz, "Deep Learning in Intrusion Detection Systems," 2019, doi: 10.1109/IBIGDELFT.2018.8625278.

[7] F. Salo, M. N. Injadat, A. Moubayed, A. B. Nassif, and A. Essex, "Clustering Enabled Classification using Ensemble Feature Selection for Intrusion Detection," 2019, doi: 10.1109/ICCNC.2019.8685636.

[8] K. Rani, H. Roopa, and V. Vani, "Prediction of network intrusion using an efficient feature selection method," 2019, doi: 10.1109/ICCS45141.2019.9065313.

[9] Balan, S., & Howell, P." A Machine Learning Approach for Network Traffic Analysis using Random Forest Regression"2019. *ACET Journal of Computer Education & Research*, *13*(1).

[10] S. Zwane, P. Tarwireyi, and M. Adigun, “Ensemble Learning Approach for Flow-based Intrusion Detection System,” 2019, doi: 10.1109/AFRICON46755.2019.9133979.

[11] S. Das, M. Ashrafuzzaman, F. T. Sheldon, and S. Shiva, “Network Intrusion Detection using Natural Language Processing and Ensemble Machine Learning,” 2020, doi: 10.1109/SSCI47803.2020.9308268.

[12] M. Verkerken, L. D’Hooge, T. Wauters, B. Volckaert, and F. De Turck, “Unsupervised Machine Learning Techniques for Network Intrusion Detection on Modern Data,” 2020, doi: 10.1109/CSNet50428.2020.9265461.

[13] M. S. Abirami, U. Yash, and S. Singh, “Building an Ensemble Learning Based Algorithm for Improving Intrusion Detection System,” 2020, doi: 10.1007/978-981-15-0199-9_55.

[14] Y. Shi, Y. Gai, “2021, doi: 10.1109/AEECA49918.2020.9213695.detection framework based on ensemble learning,” and Y. Yang, “Hoeffding trees network intrusion