IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Hybrid Framework for Enhance Intrusion Detection based on Support Vector Machine and Ant Colony Optimization

Prashansa Shrivastava
M.Tech. Research Scholar, Dept. of CSE
LNCT, Bhopal

Dr. Megha Kamble
Asst. Prof., Dept. of CSE
LNCT, Bhopal

Abstract — currently, an integral component of network protection is the Intrusion Detection System (IDS) which can detect suspicious access data or attacks. Many organizations, with increasing strategies, are penetrating to protect their networks. Despite certain preventive measures such as encryption, authentication and firewall, attackers are still able to find a way for unauthorized access and network attacks. In this paper we compare and analysis various types of Intrusion Detection System and also compare the classification Accuracy of different approaches that are used earlier research, here we also suggested an multiclass classifier hybrid approach i.e. a combination of supervise learning and Ant Colony Optimization that is useful to increase the accuracy of the system.

Keywords: KDD, IDS, Dos, Probe, R2L, U2R.

I. INTRODUCTION

An intrusion detection system has a static base of malicious behavior identified. An IDS is a hardware device or software that tracks network traffic data on a machine or network. The hackers use various forms of attacks to obtain useful information. Many intrusion detection techniques, methods and algorithms help detect such attacks, it is still desirable to know what intrusions have occurred in order to understand the threats and risks to security. With the ability to monitor network traffic and identify incoming and ongoing network attacks, most network administrators have switched to IDS to help them detect network

traffic anomalies. The current trend in intrusion detection consists of combining both host-based and network-based information to develop more efficient hybrid systems. IDSs are divided into two broad categories: Host-based and Network-based (NIDS) categories. A host-based IDS requires the installation of small programs (or agents) on individual systems for oversight. The agents track the operating system and write down data for file logging and/or warning triggering [1][5]. A network-based intrusion detection system typically consists of a network application (or sensor) running in promiscuous mode with a Network Interface Card (NIC) and separate access management. Host based intrusion detection (HIDS) refers to the detection of intrusion occurring on a single host network. [13] The data is collected from a single host device. These host-based procedures are considered the passive component. To track and analyze network traffic, a network-based intrusion detection system (NIDS) is used to protect a system from network-based threats where data is network-wide traffic. The active portion is known to be those network-based procedures.

Supervised Learning trains a model with a dataset that also includes the correct response to a prediction called as a label. Data used in this study is the NSL-KDD Cup 99 dataset. The KDD-99 dataset is more than 15 years old, but it is still

commonly used for research purposes in the field of intrusion detection systems as the lack of datasets is freely accessible to the public.
According to analysis of the NSL-KDD CUP 99 Data Set, attacks at the network layer can be divided into four basic categories:

1. Probing Attack
2. Denial of Service Attack (DoS)
3. User to Root Attack (U2R)
4. Remote to Local Attack (R2L)

- **Existing Intrusion Detection Systems**

Snort: A free and open source network intrusion detection and prevention system was created by Martin Roesch and now developed by Source fire. In 2009, Snort entered InfoWorld's Open Source Hall of Fame as one of the "greatest open source software of all time". Through protocol analysis, content searching, and various pre-processors, Snort detects thousands of worms, vulnerability exploit attempts, port scans, and other suspicious behavior.

OSSEC: An open source host-based intrusion detection system performs log analysis, integrity checking, rootkit detection, time-based alerting and active response.

OSSIM: The goal of Open Source Security Information Management, OSSIM is to provide a comprehensive compilation of tools which, when working together, grant network/security administrators with a detailed view over each and every aspect of networks, hosts, physical access devices, and servers . OSSIM incorporates several other tools, including Nagios and OSSEC HIDS.

Suricata: An open source-based intrusion detection system was developed by the Open Information Security Foundation (OISF).

Bro: An open-source, Unix-based network intrusion detection system. Bro detects intrusions by first parsing network traffic to extract its application-level semantics and then executing event-oriented analyzers that compare the activity with patterns deemed troublesome.

Fragroute/Fragrouter: A network intrusion detection evasion toolkit. Fragrouter helps an attacker launch IP-based attacks while avoiding detection. It is part of the NIDS bench suite of tools by Dug Song.

BASE: The Basic Analysis and Security Engine, BASE is a PHP-based analysis engine to search and process a database of security events generated by various IDSs, firewalls and network monitoring tools.

II. LITERATURE SURVEY

In this paper [1] ANFIS based intrusion detection was a system proposed to detect attacks in networks. Because the ANFIS, the combination of fuzzy interference model and ANN which has more advantages over other techniques. Additionally, the crow search optimization CSO algorithm was used to optimize the ANFIS model to enhance its performance over the intrusion detection which is an advantage for the IDS system. The proposed model has been used to solve the issues of intrusion detection and the model is validated using the familiar NSL-KDD dataset. Other current techniques such as BPNN, FC-ANN, GA-ANFIS, and PSOANFIS are compared to the proposed model. The intrusion detection results based on the NSL-KDD dataset were stronger and more effective than those models, with a detection rate of 95.80 percent and a FAR of 3.45 percent.

In this paper [2] The current network traffic is massive, and network attacks come in a variety of forms. As a result, anomaly detection models combined with machine learning are rapidly evolving. Web Application Firewall (WAF) bypass attacks are common, and the redundancy of data characteristics in the Hypertext Transfer Protocol (HTTP) protocol makes extracting data characteristics difficult. A web intrusion detection system with feature analysis and support vector machine (SVM) optimization is proposed in this paper. The characteristics of popular Web attacks are analyzed using expert knowledge. The HTTP protocol analysis is used to select the relevant data characteristics. The mature and robust support vector machine algorithm is used for classification learning, and the grid search approach is used for parameter optimization. As a result, a better detection capability for Web attacks is possible. Experiments were conducted using the HTTP

DATASET CSIC 2010 data set to compare the detection capability of various kernel functions. The results show that the proposed device has strong detection capabilities and can effectively detect WAF bypass attacks.

In this study [3], From the large network dataset, the most important features for improving IDS performance and building a smaller dataset in order to minimize the execution time for detecting attacks are chosen. This study used an updated Cuckoo Search Algorithm (CSA) named Mutation Cuckoo Fuzzy (MCF) for feature selection and an Evolutionary Neural Network (ENN) for classification to create an anomaly-based detection system. To allow candidates to escape local minima, the proposed search algorithm employs mutation to more precisely examine the search space. Furthermore, the solution's worth is assessed using the objective function and the Fuzzy C Means (FCM) clustering tool, which is used to produce the best results for the overlapping dataset and construct the fuzzy membership search domain, which contains all possible compromise solutions. The NSL-KDD dataset was used to test a proposed model, which was then applied to the problem of intrusion detection. The results of the experiments show that reducing features by selecting and using the most appropriate features can reduce execution time while also improving IDS efficiency and performance.

This paper [4] Using a systematic mapping analysis, provides an overview of how ensemble learners are used in IDSs. From the current literature, we gathered and reviewed 124 notable publications. Years of print, publication venues, datasets used, ensemble approaches, and IDS techniques were all used to categories the selected publications. Furthermore, this research presents and analyses an empirical investigation of a new classifier ensemble method for anomaly-based IDS called stack of ensemble (SoE). The SoE is an ensemble classifier that uses parallel architecture to combine three individual ensemble learners in a homogeneous manner, namely random forest, gradient boosting machine, and extreme gradient boosting machine. The Matthews correlation coefficients, accuracies, false positive rates, and area under the ROC curve metrics are used to assess the output significance of classification algorithms. Our research fills a void in the current literature by providing an up-to-date systematic mapping analysis, as well as a comprehensive empirical review of recent developments in ensemble learning techniques for IDSs.

In this paper [5] the researchers explored how classifiers for Multilayer Perception (MLP), J48, and the Bayes Network can be used to detect network interference using machine learning methods. This work also shows how the dataset Knowledge Discovery in Databases (KDD) was pre-processed and how different machine learning techniques were used to search and evaluate it. According to their findings, the J48 classifier had the highest accuracy in detecting and classifying attacks on KDD datasets.

In this study [6], We look at the permission-induced danger, which starts with giving these Android apps excessive permissions. The research paper's experimental work involves the development of an efficient malware detection system that aids in determining and investigating the detective impact of a number of well-known and widely used malware detection features. We use ten different feature selection methods to choose the best features from our feature data collection. We have created the malware detection model using the LSSVM (Least Square Support Vector Machine) learning approach, which is made up of three different kernel functions: linear, radial basis, and polynomial. Experiments were performed by using 2,00,000 distinct Android apps. When compared to different anti-virus scanners, the model built using LSSVM with RBF (i.e., radial basis kernel function) called FSdroid is able to detect 98.8% of malware and also achieved a 3 percent higher detection rate when compared to different frameworks or approaches proposed in the literature.

Now a day's [7] , network traffic is increasing due to the exploding usage of smart devices and the Internet. The intrusion detection work centered on feature selection or decrease because few of the features are irrelevant and excess which results prolonged detection procedure and reduces the performance of an intrusion detection system (IDS). The NSL-KDD data set is a refined variant of its predecessor KDD‟99 data set. The intent of this work is to determine essential selected input features in building IDS that is computationally efficient and amazing. For this standard feature selection jaya optimization method is used. In this paper the NSL-KDD data set is analyzed and

applied Adaptive Jaya Technique for selecting best features to minimize low false alarm rate & maximize detection rate.

Girma et al. [8] They discuss the need to prevent DDoS attacks by identifying and demonstrating a hybrid detection model that uses an innovative and efficient approach to distinguish flood attacks from flush crowd attacks (legitimate access). Furthermore, this paper introduces and addresses the use of multivariate correlation between selected and graded features to dramatically reduce false alarm rates. This is one of the most serious flaws in the proposed approach.

Hebatallah Mostafa et al. [9] present a structure for selecting features for successful network anomaly detection using a variety of machine-learning classifiers. Using filter and wrapper selection techniques, the system employs various strategies. The aim of this architecture is to choose the fewest number of features possible while still achieving optimum precision. UNSW-NB15 data set is used in the experimental results to evaluate the frame. The results show that using 18 features of a filter the classification methods and apply J48 as a classifier, an accuracy of 88% is achieved.

Rohit Kumar Singh Gautam and Amit Doegar [13] organized three exams. Since the main inspection, the frames have finished with 41 strong points. The second study, in which we perform feature selection using entropy-based analysis as a filtering method to determine satisfactory factors (rank) instead of using 41, includes the study with Naive Bayes, Adaptive Boost and P ART (decision tree partial). The third analysis, in which we use the ensemble's approach using the information obtained to choose the fine components instead of using the 41, involves and implements the experiment with Naive Bayes, Adaptive Boost and PART effects.

S. NO	Author	Name of Algorithm	Name of Attack	Accuracy
1.	S. Manimurugan [1]	Crow Search Optimization algorithm with Adaptive Neuro-Fuzzy Inference System	Dos Probe R2L U2R	96.25 92.51 94.15 90.26
2.	Afreen Bhumgara [14]	J48-Decision Tree	Dos Probe R2L U2R	98.1 97.6 97.7 97 .5
3.	Afreen Bhumgara [14]	Support Vector Machine	Dos Probe R2L U2R	97.5 97.1 93.3 93.4
4.	Afreen Bhumgara [14]	Naïve Bayesian	Dos Probe R2L U2R	74.2 73.9 69.9 71.1
5.	Baojiang Cui [15]	NADAL	Dos Probe R2L U2R	85.44 90.55 84.42 85.55
6.	Baojiang Cui [15]	Incremental Naïve Bayees	Dos Probe R2L U2R	92.20 92.25 85.78 87.24

Table 1. Comparison Table of Various Mechanisms

III. PROBLEM DOMAIN

1. All types of attack are not well detected [1].
2. There are various algorithm based on data mining like: K-Means clustering, Naivs bayes, Neural network, Support vector machine is proposed but there is still need of improvement for detecting U2R and R2L Attacks. [10].
3. Data analysis phase the rule is fixed and if the rule generation is fixed the data

clustering which will be generated for pattern finding is fixed [11].

4. Low classification accuracy in previous technique [12].

IV. Propose Work

The proposed work focuses on the limitation faced in the traditional approaches as in this dissertation a hybrid framework based on Support Vector Machine (SVM) and Ant Colony Optimization (ACO) is proposed. Combing the properties of SVM and ant colony may provide better classification in comparison to the previous methodologies. This approach considered the dataset of NSL-KDD. It is a data set which does not include redundant record and test sets are reasonable, that contain four types of attacks i.e. Denial of Service (DoS), User to Root (U2R), Remote to User (R2L) and Probing (Probe) There are several parameters that can be evaluated to compute the performance of the proposed work

Performance Metrics: The proposed method will be evaluated using all the result parameters like accuracy, precision, recall, FAR, and detection rate the following are the result parameters used for the performance analysis.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}$$

$$DR = \frac{TP}{TP+FN}$$

$$FAR = \frac{FP}{TN+FP}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{(TP+FN)}$$

V. CONCLUSION

The presented information represents an important point to address research & development in the field of ids. Based on our study various limitations are addressed such as: high false alarm rate, difficulty to apply algorithm in massive datasets, time complexity in training and testing process, etc. After surveying the various ids technique we concluded that single technique is not able to provide accurate detection rate. For this, an efficient hybrid technique is suggested to achieve accurate detection rate. Also, helps to reduce the false prediction rate as well as decrease the time complexity.

VI. References

[1] S Manimurugan , Al-qdah Majdi , Mustaffa Mohmmed, C Narmatha , R Varatharajan "Intrusion Detection in Networks using Crow Search Optimization algorithm with Adaptive Neuro-Fuzzy Inference System", "Microprocessors and Microsystems" Elsevier 6 September 2020.

[2] Chao Liu, Jing Yang and Jinqiu Wu, “Web intrusion detection system combined with feature analysis and SVM optimization”, “EURASIP Journal on Wireless Communications and Networking”, Springer 2020.

[3] Samira Sarvari, Nor Fazlida Mohd Sani, Zurina Mohd Hanapi, Mohd Taufik Abdullah, “An Efficient Anomaly Intrusion Detection Method with Feature Selection and Evolutionary Neural Network” Creative Commons Attribution 4.0, IEEE Access 2020.

[4] Bayu Adhi Tama , Sunghoon Lim, “Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation” Computer Science Review 39, ELSEVIRE 2021.

[5] M. Alkasassbeh and M. Almseidin, “Machine Learning Methods for Network Intrusion Detection,” no. October, 2018.

[6] Arvind Mahindru, A.L. Sangal, “FSDroid:- A Feature Selection Technique to Detect Malware from Android using Machine Learning Techniques”, Multimedia Tools and Applications, Springer Nature 22 December 2020.

[7] Thupakula Bhaskar, Tryambak Hiwarkar, K. Ramanjaneyulu, "Adaptive Jaya Optimization Technique for Feature Selection in NSL-KDD Data Set of Intrusion Detection System" International Conference on Communication and Information Processing (ICCIP-2019).

[8] Anteneh Girma, Mosses Garuba, and Rajini Goel, "Advanced Machine Language Approach to Detect DDoS Attack Using DBSCAN Clustering Technology with Entropy", Information Technology - New Generations. Advances in Intelligent Systems and Computing, Vol. 558, 2018, pp 125-131, 2018.

[9] Hebatallah Mostafa Anwer et al., "A Framework for Efficient Network Anomaly Intrusion Detection with Features Selection", IEEE International Conference on Information and Communication Systems (ICICS), 2018, pp. 157 – 162, 2018.

[10] Amreen Sultana, M.A.Jabbar, " Intelligent Network Intrusion Detection System using Data Mining Techniques" in 2nd International Conference on Applied and T heoretical Computing and Communication Technology (ICATCCT), IEEE 2016.

[11] Poonam Pandey, Radhika Prabhakar, "An Analysis of Machine Learning Techniques (J48 & AdaBoost) – for Classification", IEEE Conference CNC – 16, 2016, PP 978 - 984.

[12] M.H. Ali, Bahaa Abbas Dawood Al Mohammed, A. Ismail and M.F. Zolkipli, "A New Intrusion Detection System Based on Fast Learning Network and Particle Swarm Optimization," IEEE Access, vol. 6, pp. 20255–20261, 2018.

[13] Rohit Kumar Singh Gautam and Amit Doegar, "An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms", IEEE International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2018,

[14] Afreen Bhumgara, Anand Pitale. "Detection of Network Intrusions using Hybrid Intelligent Systems", "International Conference on Advances in Information Technology" IEEE 2019.

[15] Baojiang Cui and Shanshan He "Anomaly detection model based on Hadoop platform and Weka interface", Innovative Mobile and Internet Services in Ubiquitous Computing, 2016, Pp 84-89.