IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON INTRUSION DETECTION SYSTEMS FOR DETECTING NETWORKING ATTACK

Mr. Bilal Khan[1], Prof. Chetan Gupta[2]

M.Tech. Scholar, Dept. of CSE[1], Asst. Prof., Dept. of CSE[2]

SIRTS Bhopal India[1], SIRT Bhopal India[2]

Abstract: It is becoming more difficult to detect vulnerabilities with any degree of dependability as cyber-attacks become more complicated. Security services like data secrecy, integrity, and availability may start to lose credibility if the leaks are not halted. Intrusion detection has become more important and difficult as a result of the Internet and networking technologies. Machine learning and deep learning have been used to develop a number of techniques for detecting intrusions. This study provides a classification of contemporary IDS, a list of noteworthy recent papers, and an examination of the data sets that were employed in the assessment. Additionally, it describes the evasion strategies used by attackers to avoid detection and examines future problems that need to be resolved in order to improve computer network security. There has been a visual analysis of different machine learning techniques.

IDS, NSL-KDD, HIDS, NIDS, MACHINE LEARNING, DOS.

I. INTRODUCTION

The influence of networks on everyday living is greater, and cyber security is a major area of study. The main tools used in internet security are routers, IDSs, and antiviral software. This protects networks from dangers both inside and outside the network. One of these is an IDS, which by monitoring the software and hardware on the network is crucial for guaranteeing online security. They unveiled the first intrusion monitoring system in 1980 [1]. Since then, several IDS products have evolved. However, a lot of IDSs still have a high false alarm rate that causes them to issue a lot of warnings for low-anti situations, which adds to the workload for security specialists or might even make it possible for serious damaging assaults to go unnoticed. A lot of study has been done as a consequence to develop IDSs with better tracking skills or lower false alarm rates. The lack of unidentified attacks is another issue with current IDSs. New threat kinds and attacks appear as a result of the constantly shifting network configurations. IDSs that can identify puzzling attacks must be developed as a result. To address the aforementioned issues, researchers began to concentrate on creating IDSs using ML techniques. ML is a type of artificial intelligence technology that can autonomously extract information from huge databases [2]. Machine-based IDSs can perform satisfactorily when more data is provided and ML models can be expanded enough to recognize attack variants and novel threats. Additionally, IDSs that emphasis machine learning does not heavily depend on domain expertise, making them easy to design and create.

Researchers can gain illegal access to data surrounding computer systems through intrusion or negatively affect system performance when used in the context of IDS. An IDS, a tool for computer security, can detect a number of security breaches, starting with attempts to interfere with exterior entities, system assaults, and internal abuse [3]. Monitoring servers and networks, examining computer system behavior, sending alerts, and reacting to unusual activity are among IDS's primary duties. IDSs are typically implemented close to protected network sites due to the surveillance of comparable domains or networks. (e.g. transitioning in key network segments). There are two types of IDS classification techniques in use: a method based on identity and a method based on data source. IDSs can be divided into two categories among detection-based approaches: abuse detection and anomalous detection. Among the data source techniques, IDSs can be divided into host-based and network-based methods [4]. These two IDS classification techniques are combined in this study, which also considers the data source and the monitoring system as additional classification criteria.

Classification of IDS:

A. Host-based IDS (HIDS)

HIDS and other monitoring tools installed on the server. The functioning system maintains track, logs the data, and generates alerts. The stations where the workers are located are the only ones that can be seen. Utilizing host-based IDS services, it is possible to monitor critical server attack attempts. The host-based IDS checks for signs of a local system assault. An audit document serves as the basis for the host-based system's confidence. IDS can recognize complex misuse patterns that are concealed at higher levels of abstraction thanks to the details.

B. Network-based IDS (NIDS)

An NIDS typically consists of a network computer (or sensor) with a promiscuous Network Interface Card (NIC) and a distinct management interface. The IDS is located next to a network or border and controls all data in that area. Instead of gathering data from each individual computer, these systems gather data from the entire network.

As messages move through the network, the NIDS examines the network threats. The transparency of displays removes the possibility that an adversary will discover the display and instantly disable its features. On every host that is protected in the network, Network Node IDS (NNIDS) sensors are deployed.

C. Application-based IDS

IDS based on device are a unique subgroup of HIDS that examines actions taking place inside a software programmer. The most typical data source for application-based IDS is the transaction log file of the program.

II. LITERATURE REVIEW

Yu, Y et. Al. (2020) Other techniques used 20% of the KDDTrain+ dataset for training, including J48, Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), Recurrent Neural Network (RNN), and Channel boosted and residual learning based deep convolutional neural network (CBRCNN). While using less than 1% of the NSL-KDD KDDTrain+ dataset for training, our suggested intrusion detection method achieved high accuracy of 92.34% for K. Results from the experiment on the UNSW-NB15 sample were comparable. This method also increases the detection rates for Dos, U2R, R2L, and U2R. In particular, the detection rates for U2R and R2L, which only make up a tiny portion of the dataset, are increased from 13% to 81.50% and 44.41% to 75.93%, respectively[1].

L. Hakim et al. [2019] the method used by an IDS determines both its effectiveness and the training data it uses. IDS detection effectiveness and precision can be decreased by significant training data features. This thesis would look into how feature selection affected the intruder detection system. Knowledge gain, Gains Ration, Chi-squared, and Relief techniques of impact selection will be investigated in J48 along with Random Tree, Naive Bayes, and KNN algorithms. The findings demonstrate that, despite a small decrease in accuracy, the variety of characteristics can significantly increase IDS effectiveness [6].

K. A. Taher et al. [2019] this study demonstrates that the wrapper feature selection Artificial Neural Network (ANN)-based machine learning method outperforms the vector machine support technique when categorizing network data. (SVM). NSL-KDD evaluates performance by detecting network traffic using machine learning techniques, which are supervised by the SVM and ANN methodology. The proposed model has a greater success rate for intrusion detection than other current models, according to the comparative research [7].

M. M. Sakr et al. [2019] this paper presents several feature selection techniques to improve the efficacy of NIDS. The types of selected techniques that were used as sensors or wrappers are Information Gain (IG), Principal Component Analysis (PCA), and Correlation Function Collection (CFS) (instead of Genetic Algorithm (GA), Artificial Bee Colony (ABC), and Particle Swarm Optimization (PSO)). SVM for defining a network connection. The NIDS creation and testing process uses the NSL-KDD network communication standard. The evaluation's findings showed that wrapper approaches for NIDS produced superior categorised accuracy, recognition rates, true positive rates, and low false-positive rates when compared to filter methods. When combined with other connected NIDS, their ABC-NIDS has been regarded as having the highest utility in their system. [8].

S. Sun et al. [2018] for the wrapper feature selection system, a lightning attachment procedure optimization algorithm (LAPO) and SVM for intruder detection are suggested in the article. LAPO is a recently proposed, adaptable programmer that draws inspiration from nature. To evaluate the effectiveness of the suggested method, the well-known KDD Cup 99 dataset is used. According to experimental findings, searching for the optimal function subgroup is more accurate and efficient than using the GA or PSO. [9].

H. M. Answer et al. [2018] this study contains a feature selection technique with a broad range of classification devices for effective network anomaly detection. The method uses both filtering and wrapper filtering ways to integrate a number of strategies. The goal of this frame is to select the fewest number of features while still guaranteeing the highest level of precision. The data gathering UNSW-NB15 for assessing the proposed structure is included in the experimental findings. The outcomes demonstrated that a precision of 88 percent could be achieved using 18 traits based on one of the filters and J48 as a classifier [10].

H. Li et al. [2018] This study proposes a hybrid FS model based on random forest or particle swarm optimization, employing both an autonomous computation and an educational method. It employs an independent measure to choose the best subsets for a particular cardinality and a learning algorithm to choose the best subset overall from among the best subsets in cardinalities. The 1999 KDD dataset was used as a tested for the suggested model's TPR and FPR, and to compare them to the CFS and SVM methods. [11].

J. Ku et al. [2017] in security study, ID is currently a viable field. IDSs have developed into a crucial component and a key network security technology, and they can now identify anyone instantly who is not authorized to use the present computer system. Research in ELM is focused on spotting possible dangers or intrusions. In this work, they propose a better learning method called self-adaptive differential evolution ELM (SADE-ELM) for the categorization and identification of incursions. In order to evaluate our methods, we use the ELM and DE-ELM processes. According to the recommended SADE-ELM technique, the proposed method has a lower identifying accuracy for categorization instances. [12].

1) Differential evolution feature selection

Differential evolution (DE), a type of evolutionary algorithm, models feature search on an ant population. DE offers the advantages of being simple yet successful, just like other optimization methods [12, 13]. Different DE compensations exist. 1) The ability to handle non-differentiable, non-linear, and multimodal cost functions; 2) The ability to handle analytical cost functions in tandem; and 3) The ease of use; DE and GA both employ the same mutation, crossover, or selection variables. DE's effectiveness depends on how the goal vector or the vector used to acquire a test vector is handled during the search procedure.

2) Whale Optimization Algorithm (WOA)

In 2016 34, Lewis or Mirjalili created the WOA. First-stage circular or first spiral updating is one of the algorithm's two main stages. In the second phase, a haphazard hunt for an objective (exploration stage) will be conducted. Whales are initially provided arbitrary answers, and the min or max value of an objective function is taken into consideration to handle the best value based on the circumstances. The search agent for each goal function is then chosen. Based on the optimal answer or random search agent for each iteration, each search agent modifies its location.

3) Extreme Learning Machines (ELM)

ELM is used to train a single hidden layer NN. (SLFNs). In ELM, hidden nodes are initiated at random or in a fixed manner without gradual adjustment. ELM's hidden nodes, however, don't even have to mimic neurons. The only unrestricted component required for learning are the links (or weights) between the output layer and the hidden layer. The ELM is therefore built as a linear model in a parameter to solve a linear system. ELM aims to reach the world optimum and is noticeably more effective than conventional FNN methods of instruction. It has been shown that ELM maintains the SLFNs' capacity for universal approximation even when dealing with hidden nodes that are created arbitrarily.

4) Simulated Annealing (SA)

SA is a met heuristic solution that was motivated by the metallurgy technique. It is a simple way to increase material heating and cooling in order to enhance the crystal scale. By using room temperature, the energy required to remove flaws from metal structures is decreased. The SA technique makes use of its temperature progress as a controlling element and an internal energy feature. A original S solution and an updated S′ solution are used to start the simulated rectification. The solution for this method is produced if the fitness function F(S*) values are less than F(S).

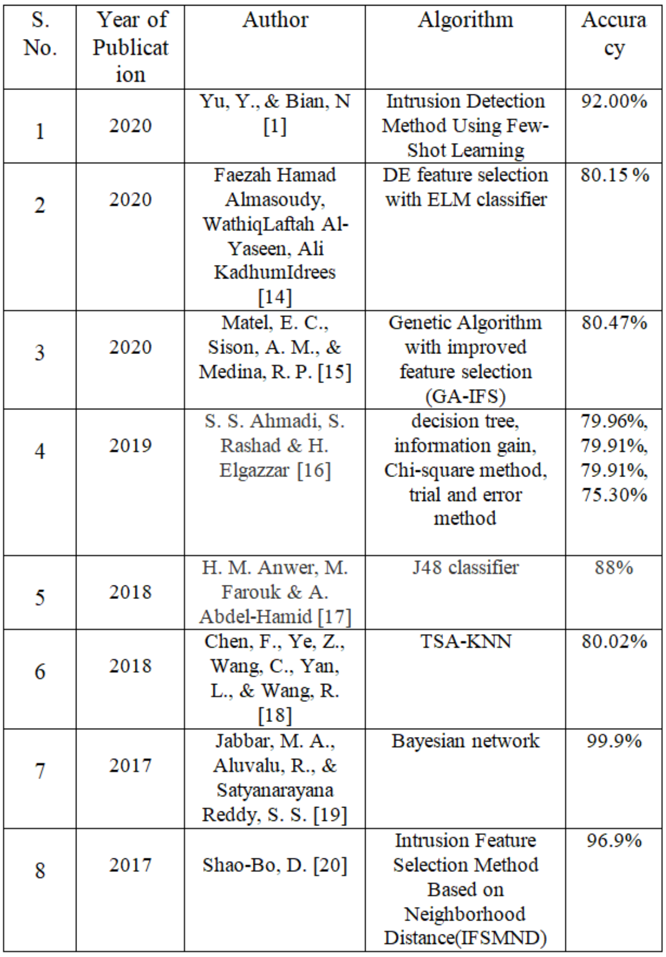
$$P_b = \exp\left(\frac{-(f(S^*) - f(S))}{T_m}\right)(1)$$

According to Eq. [6], the greater S* fitness value is recognised. With the help of this strategy, it is possible to stop the search process from interacting with local optima. The fitness function for the current solution is F(S), while F(S*) is the fitness function for the neighbouring solution. Temperature Tm serves to specify the control parameter. The order of movements is what creates equilibrium, and the cooling rate is what determines the temperature control parameter. The effectiveness of global search is impacted by TM parameter management. If the temperature starts out high, the likelihood of the simulated anneal process increases. If no improvements are made after several temperature decreases, the SA process will end. There is less chance of finding global solutions if the initial temperature is low and the calculation time is brief.

$$T_m = \delta^k + T_o + T_{fn} \qquad (2)$$

The number of stints given by the neighbouring solution is k in situations where k is the decreasing Tm, 0 to 1, To is the starting temperature value, and Tfn is the final temperature value. The process for SA is explained in the following formula.

Table 1: Literature Survey of IDS



S. No.	Year of Publication	Author	Algorithm	Accuracy
1	2020	Yu, Y., & Bian, N [1]	Intrusion Detection Method Using Few-Shot Learning	92.00%
2	2020	Faezah Hamad Almasoudy, WathiqLaftah Al-Yaseen, Ali KadhumIdrees [14]	DE feature selection with ELM classifier	80.15 %
3	2020	Matel, E. C., Sison, A. M., & Medina, R. P. [15]	Genetic Algorithm with improved feature selection (GA-IFS)	80.47%
4	2019	S. S. Ahmadi, S. Rashad & H. Elgazzar [16]	decision tree, information gain, Chi-square method, trial and error method	79.96%, 79.91%, 79.91%, 75.30%
5	2018	H. M. Anwer, M. Farouk & A. Abdel-Hamid [17]	J48 classifier	88%
6	2018	Chen, F., Ye, Z., Wang, C., Yan, L., & Wang, R. [18]	TSA-KNN	80.02%
7	2017	Jabbar, M. A., Aluvalu, R., & Satyanarayana Reddy, S. S. [19]	Bayesian network	99.9%
8	2017	Shao-Bo, D. [20]	Intrusion Feature Selection Method Based on Neighborhood Distance(IFSMND)	96.9%

Data Description

There are a number of intrusion detection datasets, but this study is built on the KDD'99 dataset in particular. The NSL-KDD data set, which is an updated variation of the original KDDCup'99 dataset, was suggested in 2009. On the one hand, NSL-KDD has continued to retain the advantages and challenges of KDD-Cup 99. The research didn't address the drawbacks carried over from the initial findings until it reduced redundancies, rationalized the number of cases, or preserved the variability of the chosen samples. Make certain that the NSL-KDD dataset is assembled to maximize the predictive complexity that makes it exceptional. The records from the first dataset were divided into five complexity ranges using numerous benchmark classifiers, or each instance was annotated with a variety of predictions [12]. For each complex level group, the quantity of selected data is inversely proportional to the record percentages of the original KDDCup99 dataset.

Every record is listed as usual or abnormal if the abnormal one is 22 attacks in the training set & 39 attacks in the test set. [21]:

- **DOS:** Resources have been allocated by more device demands to avoid users' availability.
- **Probe:** Check by network scanning for information about the target host.
- **User to Root (U2R):** Request unauthorized access to the controlling account by a devaluation of the device details by the password.
- **Remote to User (R2U):** Legal user access to the device.

III. PROBLEM DOMAIN

The various problem domains that I will attempt to address and improve in my study are listed below based on previous studies and the relevant research paper.

- Less Accuracy
- Detection Rate is low
- Precision percentage is low
- F-Score is less
- False Alarm Rate

IV. PROPOSE WORK

I'm attempting to suggest a hybrid intrusion detection system that will be based on machine learning approach and use either supervised learning or unsupervised learning techniques after reading various research papers and research problems. For data categorization, I'll also use a data mining algorithm. Additionally, compared to earlier work, this hybrid combination will yield better and enhanced results. and I will try to evaluate the proposed work by checking the following parameters:

- Accuracy:
- Detection Rate
- Precision
- F-Score
- False Alarm Rate

V. CONCLUSION

Cybercriminals target computer users by using cutting-edge techniques and social networking strategies. Some hackers are more skilled and driven than others. Cybercriminals have demonstrated their ability to conceal their identities, their communications, and their illicit earnings, as well as their use of robust infrastructure. Therefore, it becomes more and more important to protect computers with cutting-edge IDSs capable of detecting contemporary malware. For creating or creating such IDS systems, it is crucial to have a thorough understanding of the advantages and disadvantages of contemporary IDS study. We also provided a thorough analysis of the methods, varieties, or technologies for intrusion detection systems, along with their advantages and disadvantages.

REFERENCES

1. Yu, Y., & Bian, N.. (2020). An Intrusion Detection Method Using Few-Shot Learning. 8. https://doi.org/10.1109/ACCESS.2020.2980136.
2. Li, Y., Zhang, C., & Yang, L.. (2016). The Research of AMI Intrusion Detection Method using ELM in Smart Grid. 10(5). https://doi.org/10.14257/IJSIA.2016.10.5.27.
3. Juanchaiyaphum, J., Arch-int, N., Arch-int, S., & Saiyod, S.. (2015). A Novel Lightweight Hybrid Intrusion Detection Method Using a Combination of Data Mining Techniques. 9(4). https://doi.org/10.14257/IJSIA.2015.9.4.10
4. Wu, Z., Wang, J., Hu, L., Zhang, Z., & Wu, H.. (2020). A network intrusion detection method based on semantic Re-encoding and deep learning. 164. https://doi.org/10.1016/J.JNCA.2020.102688.
5. Gao, B., Bu, B., Zhang, W., & Li, X.. (2021). An Intrusion Detection Method Based on Machine Learning and State Observer for Train-Ground Communication Systems. https://doi.org/10.1109/TITS.2021.3058553.
6. L. Hakim, R. Fatma, and Novriandi, "Influence Analysis of Feature Selection to Network Intrusion Detection System Performance Using NSL-KDD Dataset," 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 2019, pp. 217-220, DOI: 10.1109/ICOMITEE.2019.8920961.
7. K. A. Taher, B. Mohammed Yasin Jisan and M. M. Rahman, "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection," 2019 International Conference on

Robotics,Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 2019, pp. 643-646, DOI: 10.1109/ICREST.2019.8644161.

8. M. M. Sakr, M. A. Tawfeeq and A. B. El-Sisi, "Filter Versus Wrapper Feature Selection for Network Intrusion Detection System," 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 2019, pp. 209-214, DOI: 10.1109/ICICIS46948.2019.9014797.
9. S. Sun, Z. Ye, L. Yan, J. Su, and R. Wang, "Wrapper Feature Selection Based on Lightning Attachment Procedure Optimization and Support Vector Machine for Intrusion Detection," 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Lviv, 2018, pp. 41-46, DOI: 10.1109/IDAACS-SWS.2018.8525742.
10. H. M. Anwer, M. Farouk, and A. Abdel-Hamid, "A framework for efficient network anomaly intrusion detection with features selection," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 157-162, DOI: 10.1109/IACS.2018.8355459.
11. H. Li, W. Guo, G. Wu, and Y. Li, "An RF-PSO Based Hybrid Feature Selection Model in Intrusion Detection System," 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, 2018, pp. 795-802, DOI: 10.1109/DSC.2018.00128.
12. J. Ku, B. Zheng, and D. Yun, "Intrusion Detection Based on Self-Adaptive Differential Evolutionary Extreme Learning Machine," 2017 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, 2017, pp. 94-100, DOI: 10.1109/ICCNEA.2017.57.
13. Sugianela, Y., & Ahmad, T. (2020). Pearson Correlation Attribute Evaluation-based Feature Selection for Intrusion Detection System. 2020 International Conference on Smart Technology and Applications (ICoSTA).
14. Almasoudy, F. H., Al-Yaseen, W. L., & Idrees, A. K. (2020). Differential Evolution Wrapper Feature Selection for Intrusion Detection System. Procedia Computer Science, 167, 1230–1239.
15. Elmer C., Ariel M. Sison Ruji P. Medina Matel “Implementation of GA-IFS-based Network Intrusion Detection System: A comparative analysis” ICSET'20: 2020 The 4th International Conference on E-Society, E-Education and E-Technology.
16. S. S. Ahmadi, S. Rashad and H. Elgazzar, "Efficient Feature Selection for Intrusion Detection Systems," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York City, NY, USA, 2019, pp. 1029-1034, DOI: 10.1109/UEMCON47517.2019.8992960.
17. H. M. Anwer, M. Farouk, and A. Abdel-Hamid, "A framework for efficient network anomaly intrusion detection with features selection," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 157-162, DOI: 10.1109/IACS.2018.8355459.
18. F. Chen, Z. Ye, C. Wang, L. Yan, and R. Wang, "A Feature Selection Approach for Network Intrusion Detection Based on Tree-Seed Algorithm and K-Nearest Neighbor," 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Lviv, 2018, pp. 68-72.
19. M. A. Jabbar, R. Aluvalu and S. S. Satyanarayana Reddy, "Intrusion Detection System Using Bayesian Network and Feature Subset Selection," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1-5.
20. D. Shao-Bo, "Intrusion Feature Selection Method Based on Neighborhood Distance," 2017 International Conference on Computer Systems, Electronics, and Control (ICCSEC), Dalian, 2017, pp. 748-751, DOI: 10.1109/ICCSEC.2017.8446849.
21. X. Zhang, P. Zhu, J. Tian, and J. Zhang, "An effective semi-supervised model for intrusion detection using feature selection based LapSVM," 2017 International Conference on Computer, Information and Telecommunication Systems (CITS), Dalian, 2017, pp. 283-286.