International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Various Machine Learning Approach for Detecting Various Attacks

Alka, Prof. Sumit Sharma

Department of Computer Science Engineering
Vaishnavi Institute of Technology & Science, Bhopal

ABSTRACT

The intrusion detection system, or IDS, has shown to be a crucial component of system security, helping to detect several attacks with the goal of protecting systems from serious damage and identifying the system's intrusion threats. Thus, the primary role of the majority of intrusion detection systems is to reliably detect intrusions. Therefore, it is imperative that all organizations—private, public, or government keep information secure. Even though there are several IDSs, the primary problem is that they have low detection rates and a high number of false positives. Our research aims to address this problem by improving the rate of intrusion detection and reducing false alarms. The reason why current algorithms perform poorly is because of the raw dataset, which confuses the classifier and leads to erroneous identification because of extraneous characteristics. This review paper shows numerous intrusion detection strategies, tactics, and algorithms will fend off these attacks. This primary objective is to present a thorough analysis of the definition, background, life cycle, and methodologies of intrusion detection as well as many forms of attacks, tools and techniques, and application issues. A variety of studies were conducted utilizing the various machine learning technique to perform Multi-Class Classification.

Keywords: *Feature Selection, Attribute, Optimal Subset, Classification, Support Vector Machine, Convolutional Neural Network.*

1. INTRODUCTION

In day to day life the need for speed access of information through internet has increased. Hence the room for maintaining security in any organization either public or private system has become fundamental. Because -4- of increase in network connections and systems, unauthorized access and interruption of the data is triggered. As a result, it is indispensable to create a virtual access path. In general intruders have capacity to find out defect in systems or networks and can spawn vulnerabilities. Even though the access control points exist in network, they fail in providing scrupulous security to the systems. To identify intruders, developing Intrusion Detection Systems (IDSs) is the best solution to protect systems and networks. Therefore the task of IDS is not only to detect intruders but also to monitor the raid of intruders. An accurate system of protecting data and resources from illicit access, damaging and denial of use is to be built. For every system, the security perspective is to be planned based on the expected performance.

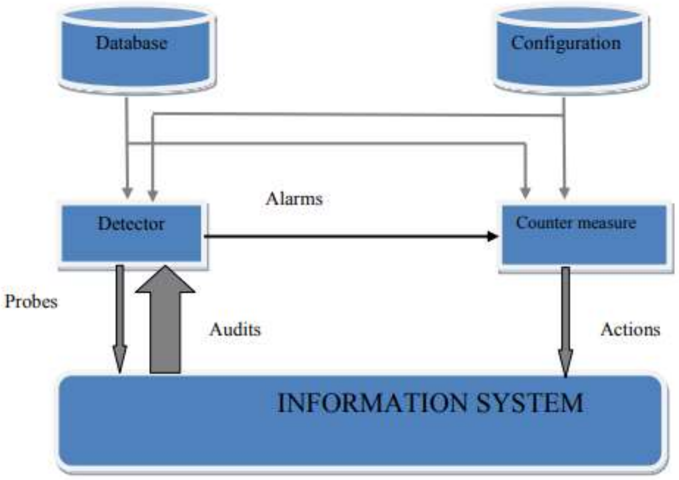

Figure 1: IDS Structure.

A masterful and accurate tool for real time intrusion discovery is the target of main experimenters in IDSs. There is a variety of Artificial Intelligence (AI) concepts were exploited for transforming intrusion discovery procedure, therefore human involvement is decreased. And also in common, the procedures which deal with IDS are utilizing machine learning. Basically Soft Computing techniques that were used in IDS implementation are Artificial Neural Networks (ANNs), Support Vector Machines(SVMs),Bayesian Networks, Fuzzy logic, Particle Swarm Optimization and Genetic Algorithms (GA).

Discovering intrusions merely with human eyes, will be tremendously intricate. Towards diminishing the crisis, system security scientists use prevailing data mining and artificial intelligence methodologies in exploring probable intrusions. Conversely, if the total set of features employed in network data is increased then classifying intrusions become complicated, since complex relationships exists between features [3]. There are complex relationships existing among features as well as intrusion classes. It will produce more processing costs and also delays in detecting intrusions. In view of the restrictions on humans and computers together, feature selection is accordingly essential such that burden in handling data and time required in noticing intrusions will be lessened [4]. In detecting intrusions, IDS defends a computer network from illicit persons, possibly insiders. The attack recognition task is considered as the model of classification expert in distinguishing „harmful‟ connections referred as intrusions or attacks, and „sympathetic‟ connections referred as normal. There are various categories of IDSs are prevailing that are based on structure and detection method.

2. REVIEW ON IDS METHODS

Shirani et al. (2015) constructed a model for web service intrusion detection based on the Autoregressive Integrated Moving Average (ARIMA). The ARIMA model is applied first on the training data. Second, it estimates its future behavior within a certain standard error. Third, it checks the training dataset; if any instance deviate significantly from the confidence level, the computer will warn the manager that there may be an irregularity. Real-world data is utilized in the studies to obtain the desired results.

An intrusion detection system based on a variety of attribute selection approaches, such as information gain, reciprocal connection, as well as feature closeness centrality, was proposed by *Cleetus et al. (2014).* The modifiable subset of features is handled using genetic algorithms. This model's accuracy of 87.54% shows how the information-based method for selecting features may boost the detection accuracy.

Mehmood et al. (2013) have presented an overview of the many cloud intrusions that occur. Following that, it evaluates many cloud-based IDS solutions in terms of their kind, placement, detection time, detection strategy, data source, and threats that they can identify. The assessment also analyzes the weaknesses of each technique to evaluate whether or not they match the cloud computing environment's security needs. It necessitates the implementation of an intrusion detection system (IDS), which utilizes a number of detection methods to handle cloud security challenges.

Kim et al. (2017) studied an Artificial Intelligence (AI) intrusion detection system (IDS) that leverages a Deep Neural Network (DNN) on the KDD Cup 99 dataset as a protection against changing network attacks. The initial steps in preparing data for the DNN model are data transformation and standardization. The whole KDD Cup 99 dataset is utilized for confirmation, and the DNN approach is applied to preprocessed and enhanced data to generate a learning model. Finally, the detection efficacy of the DNN model, that has been shown to generate positive results for intrusion detection, was evaluated using the accuracy, detection rate, and false alarm rate.

Santoso et al. (2016) investigated the performance of well-configured NIDS on an OpenStack private cloud. The purpose of this study is to assess the usefulness of NIDS and the accuracy of its attack classification. The results show that the model's output is dependable and accurate. Furthermore, NIDS real-time warning is effective at detecting attacks that are classified as occurring across the network.

Bombatkar et al. (2015) investigated a scalable and effective technique for performing intrusion detection and categorization. It focuses on adopting a distributed approach for vulnerability scanning, which improves both scalability and efficacy. This approach is divided into two categories detection and classification. Preprocessing follows the building of the attack graph. The attack graph method works by extracting and modeling properties from input flows. The classification of internet data leads to the detection of attacks. According to experimental tests, a system outperforms a conventional one substantially.

Zhao et al. (2016) employed a novel intrusion detection algorithm based on augmented K-means to address the security needs of cloud computing. A distributed intrusion detection method and a clustering algorithm are the results of this methodology. It can detect both known and novel threats in the cloud computing environment. The results of a software simulation indicate that this strategy may improve intrusion detection speed while decreasing the number of false positives and false negatives.

Sangve et al. (2015) proposed the algebraic strategy for Anomaly Network Intrusion Detection Systems (ANIDS) on vast size datasets using the sensors developed, demonstrating methods for machine learning employing a variety of datasets. Fewer ANIDS make use of the NSL-KDD dataset, a revamped version of the well-known KDD Cup 99 dataset. The KDD Cup 99 dataset has been utilized by many ANIDS. When matched to the KDD99 dataset, the NSL-KDD dataset clearly outperforms it.

Qazanfari et al. (2012) employed a hybrid anomaly-based intrusion detection system that depends on both signature-based as well as anomaly-based methodologies. To enhance the performance of this scheme, it employs the following strategies: first, it pulls optimal data from the KDD data set using a feature extraction technique based on feature entropy; secondly, it employs a unique methodology to integrate the result of these two learning-based techniques. Finally, the detection strategy uses the KDD dataset to illustrate the efficacy of these hybrid systems. The simulation results demonstrate the

KDD features that are suitable for discriminating between normal and abnormal traffic. This result indicates how successfully the method can detect DOS, Probe, U2R, and R2L attacks.

Zhu et al. (2012) created a unique intrusion detection approach (U-D methodology) that considers both received and published data. Using the new analytic approach, intrusion clues may well be recognized with greater efficacy and efficiency. The connections between these data points may give some intuitive clues for detecting significant incursions. Experiment results show that the approach is successful in terms of high detection.

According to *Chen et al. (2011)*, integrating rough sets with data mining to augment standard intrusion detection systems will boost detection effectiveness and reduce fake alerts. Data collection starts with categorization, then processing, which includes standardizing independent variables, and lastly discrete processing of nominal variables. The Pawlak feature weighted rough set technique may also be used to minimize the amount of variables in the result set by utilizing the property upward and downward approximated set's features. Association rules that fulfill a certain degree of confidence can be constructed using attribute reductions and then imported into the rule set. According to studies, the detection strategy that integrates data mining with rough sets improves detection efficacy by more than 20%. As the number of incursions rises, the detection rate approaches linearity.

Geng et al. (2009) emphasized the negatives selection method, which applies the maximum entropy model, as a method of reducing degeneration caused by repeated meaningless program execution. The method described above employs the negative selection technique to eliminate the wasteful computation predicted by the maximum entropy model. Experiment results show that the cost of computing is lowered by 50-80% while retaining the same detection rate.

In their network intrusion detection methodologies, *Jing et al. (2016)* employed a restricted Boltzmann machine learning algorithm as well as relevance deep learning to uncover the key factors of relevant machine learning. It evaluated the feasibility of utilizing deep learning in networking IDS, an element of intrusion detection for networks technology. Depth learning is relevantly utilized for network intrusion detection systems that can attain improved detection accuracy. Network intrusion detection systems that depend on relevance deep learning have greater average detection performance and average false detection rates for unidentified intrusions and attacks, according to simulated data. The results of the trial also indicate the new technology's dependability and efficacy.

Yuan et al. (2017) employed the Two Layers Multi-Class Detection (TLMD) approach, the C5.0 methodology, as well as the Naive Bayes algorithm to improve detection and false alarm rates in adaptable network intrusion detection. The TLMD approach also handles a variety of data mining difficulties, such as dealing with imbalanced datasets, managing with consecutive characteristics, and minimizing noise in the training dataset. The performance of the recently revealed TLMD strategy is compared to that of previous approaches using the detection rate, accuracy, as well as false alarm rate out from KDD Cup99 standard intrusion detection dataset. Experiment results reveal that now the unique TLMD technique has a lower false alarm rate and greater detection accuracy, based on the imbalanced dataset.

2.2.1 Research Gap

The summary of the works available in the literature is given in Table 2.2.

Table 2.2-Summary of Existing Researches in IDS

Detection	Type	Datasets	Methodology Employed	Attack Detection	Limitations
Anomaly	NIDS	KDD	Machine Learning, Data mining	DoS, U2R, R2L, Probing, and Normal are all options.	The method hasn't been tested in a smart grid setting based on the Internet of Things with programmable hardware like FPGA. It is being investigated to centralize the deployment of IDS.
Anomaly	NIDS	Dataset collected using Cooja simulator in the Contiki operating system	Distributed and collaborative	Routing and application-specific attacks	The results are not confined to one kind of attack. The dataset that was utilized to evaluate the recommended technique is not typical. Just the devices power is taken into consideration when comparing the results.
Anomaly	NIDS	Dataset collected from the simulation of the IoT network in Matlab.	Statistical model	Advisory present at the physical layer	The information utilized to evaluate does not come from typical datasets. A theoretical framework doesn't need any special training.
Anomaly	NIDS	Data collected using fuzzy clustering	Machine learning, data mining	High-risk data, Low-risk data	There is no information on how the data was gathered. It doesn't make any

					particular reference of any cyberattack.
Anomaly	Hybrid IDS	Data collected in Matlab simulation of IoT healthcare system	Protocol model	Sinkhole, selective forwarding, version number	Based on simulated data, validated results
Anomaly	NIDS	Data collected from IoT smart Home Testbed for four consecutive days	Machine Learning	DoS	The proposed method only applies to routing-related attacks.
Anomaly	NIDS	Data collected by simulation of IoT networks in Cooja under Contiki environment	Machine Learning	Blackhole	There is no set method for selecting features.
Anomaly	NIDS	KDD-CUP99, AWID	Machine Learning	DoS, U2R, R2L, Probing, Normal	There were no standard datasets used in the evaluation of the proposed method.
Anomaly	NIDS	Data collected from testbed created using a Raspberry Pi, Android phone, router	Signature model, the protocol model	Jam-attack, false-attack, reply- attack	On a standard dataset, the proposed method was tested. The data used is not from the actual world.
Anomaly, Specification	Hybrid IDS	Tested on WSN simulator created on .Net framework	Machine learning, signature model	Sinkhole, selective forwarding attack, wormhole	Other types of attack are not considered. The proposed solution is only relevant to attack routing. The prevalence of false positives.

3. Need of Intrusion Detection System (IDS)

The Intrusion Detection Expert System (IDES) was developed by the Stanford Research Institute (SRI) in the early 1980. IDES monitors user behavior and spots questionable behavior [27]. Dorothy Denning seminal research an Intrusion Detection Model, which was published in 1987, provided a methodological framework that inspired the work of other academics [28]. The framework includes rules for learning about this behavior through audit records, for identifying anomalous activity, and for profile, which records how people interact with things using metrics and mathematical analysis. It is possible to set up an ID to keep an eye out for attacks or other events that come from a company's network. This protects the company from any prospective legal actions brought forth by internal attacks.

Both an IDS and a firewall cannot replace each other since an IDS frequently lacks the firewalling capabilities of a firewall and a firewall lacks the intrusion detection capabilities of an IDS. The maximum level of security is often provided by combining the two technologies since they operate well together.

Having IDS in addition to a firewall may add an additional layer of security to a system such-

- Recognizing attacks that a firewall really permits to pass (such as http attacks against web servers).
- Detect efforts like port scans.
- Offers extra checks for purposefully or accidentally exposed ports or gaps in the firewall.

4. Problem Statement

False positives and false negatives are IDSs' primary shortcomings. False positives add to the noise that can significantly reduce IDS's efficacy. A high false positive rate might cause security professionals to become weary of responding to alerts, which leaves serious threats unreported and at risk of going undiscovered. With malware and attack strategies growing more complex, false negatives are a major issue. For instance, a novel form of malware might elude IDS detection by exhibiting no previously documented patterns of behavior. Feature selection is a combinatorial optimization problem where the objective is to extract certain characteristics from a collection. The existing algorithms have issues with error accumulation that can be seen with sequential techniques, and thus proposed algorithm chooses features based on more reliable data than can be obtained from individual models.

5. Conclusion

This investigation of feature selection algorithms for a sizable survey demonstrates that the feature selection technique regularly raises the classifier's accuracy. Each feature selection process has benefits and drawbacks of its own. The dataset with more characteristics uses wrapper techniques, which provide less accuracy gain. Accuracy is decreased when a greater characteristic is included. Since each method behaves differently, it is impossible to use a single approach across several datasets. The accuracy of the categorization of various datasets is determined by feature selection algorithms. The feature selection algorithm must choose the pertinent characteristics and exclude the unrelated and inconsistent features that reduce the categorization algorithms' accuracy.

REFERENCES

[1] Wang, G., Hao, J., Ma, J., & Huang, L. (2010). A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering. *Expert systems with applications*, *37*(9), 6225-6232.

[2] Beghdad, R. (2008). Critical study of neural networks in detecting intrusions. *Computers & security*, *27*(5-6), 168-175.

[3] Hlaing, T. (2012). Feature selection and fuzzy decision tree for network intrusion detection. *International Journal of Informatics and Communication Technology*, *1*(2), 109-118.

[4] Kim, D. S., Nguyen, H. N., & Park, J. S. (2005, March). Genetic algorithm to improve SVM based network intrusion detection system. In *19th International Conference on Advanced Information Networking and Applications (AINA'05) Volume 1 (AINA papers)* (Vol. 2, pp. 155-158). IEEE.

[5] Ghorbani, A. A., Lu, W., & Tavallaee, M. (2009). *Network intrusion detection and prevention: concepts and techniques* (Vol. 47). Springer Science & Business Media.

[6] Labib, K., & Vemuri, V. R. (2004, June). Detecting and visualizing denial-of-service and network probe attacks using principal component analysis. In *Third Conference on Security and Network Architectures, La Londe,(France)*.

[7] Rangarajan, L. (2010). Bi-level dimensionality reduction methods using feature selection and feature extraction. *International Journal of Computer Applications*, *4*(2), 33-38.

[8] Farid, D. M., & Rahman, M. Z. (2010). Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm. *J. Comput.*, *5*(1), 23-31.

[9] Zimek, A., Schubert, E., & Kriegel, H. P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *5*(5), 363-387.

[10] Yen, L. H., & Tsai, W. T. (2010). The room shortage problem of tree-based ZigBee/IEEE 802.15. 4 wireless networks. *Computer Communications*, *33*(4), 454-462.

[11] Haddara, M., & Staaby, A. (2018). RFID applications and adoptions in healthcare: a review on patient safety. *Procedia computer science*, *138*, 80-88.

[12] Li, C. T., Lee, C. C., Weng, C. Y., & Chen, C. M. (2018). Towards secure authenticating of cache in the reader for RFID-based IoT systems. *Peer-to-Peer Networking and Applications*, *11*(1), 198-208.

[13] Park, S. S. (2018, January). An IoT application service using mobile RFID technology. In *2018 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-4). IEEE.

[14] Lin, J. R., Talty, T., & Tonguz, O. K. (2015). On the potential of bluetooth low energy technology for vehicular applications. *IEEE Communications Magazine*, *53*(1), 267-275.

[15] Raza, S., Misra, P., He, Z., & Voigt, T. (2017). Building the Internet of Things with bluetooth smart. *Ad Hoc Networks*, *57*, 19-31.

[16] Collotta, M., Pau, G., Talty, T., & Tonguz, O. K. (2018). Bluetooth 5: A concrete step forward toward the IoT. *IEEE Communications Magazine*, *56*(7), 125-131.

[17] Fürst, J., Chen, K., Kim, H. S., & Bonnet, P. (2018, April). Evaluating Bluetooth low energy for IoT. In *2018 IEEE Workshop on Benchmarking Cyber-Physical Networks and Systems (CPSBench)* (pp. 1-6). IEEE.

[18] Hasan, K., Biswas, K., Ahmed, K., Nafi, N. S., & Islam, M. S. (2019). A comprehensive review of wireless body area network. *Journal of Network and Computer Applications*, *143*, 178-198.

[19] Nabila, A. (2019, April). A QoS based comparative analysis of the IEEE standards 802.15. 4 & 802.15. 6 in WBAN-based healthcare monitoring systems. In *2019 International conference on wireless technologies, embedded and intelligent systems (WITS)* (pp. 1-5). IEEE.

[20] Kim, T. (2018, April). A study of the Z-wave protocol: implementing your own smart home gateway. In *2018 3rd International Conference on Computer and Communication Systems (ICCCS)* (pp. 411-415). IEEE.

[21] Naidu, G. A., & Kumar, J. (2019). Wireless protocols: Wi-Fi son, Bluetooth, zigbee, z-wave, and Wi-Fi. In *Innovations in electronics and communication engineering* (pp. 229-239). Springer, Singapore.

[22] Lavric, A., & Petrariu, A. I. (2018, May). LoRaWAN communication protocol: The new era of IoT. In *2018 International Conference on Development and Application Systems (DAS)* (pp. 74-77). IEEE.

[23] Haxhibeqiri, J., De Poorter, E., Moerman, I., & Hoebeke, J. (2018). A survey of LoRaWAN for IoT: From technology to application. *Sensors*, *18*(11), 3995.

[24] Jalaian, B., Gregory, T., Suri, N., Russell, S., Sadler, L., & Lee, M. (2018, February). Evaluating LoRaWAN-based IoT devices for the tactical military environment. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)* (pp. 124-128). IEEE.

[25] Mekki, K., Bajic, E., Chaxel, F., & Meyer, F. (2018, March). Overview of cellular LPWAN technologies for IoT deployment: Sigfox, LoRaWAN, and NB-IoT. In *2018 ieee international conference on pervasive computing and communications workshops (percom workshops)* (pp. 197-202). IEEE.

[26] Ayoub, W., Samhat, A. E., Nouvel, F., Mroue, M., & Prévotet, J. C. (2018). Internet of mobile things: Overview of lorawan, dash7, and nb-iot in lpwans standards and supported mobility. *IEEE Communications Surveys & Tutorials*, *21*(2), 1561-1581.

[27] Lavric, A., Petrariu, A. I., & Popa, V. (2019, August). SigFox communication protocol: The new era of IoT?. In *2019 international conference on sensing and instrumentation in IoT Era (ISSI)* (pp. 1-4). IEEE.

[28] Mekki, K., Bajic, E., Chaxel, F., & Meyer, F. (2018, March). Overview of cellular LPWAN technologies for IoT deployment: Sigfox, LoRaWAN, and NB-IoT. In *2018 ieee international conference on pervasive computing and communications workshops (percom workshops)* (pp. 197-202). IEEE.

[29] Aldahdouh, K. A., Darabkh, K. A., & Al-Sit, W. (2019, March). A survey of 5G emerging wireless technologies featuring LoRaWAN, Sigfox, NB-IoT and LTE-M. In *2019 International conference on wireless communications signal processing and networking (WiSPNET)* (pp. 561-566). IEEE.

[30] Osman, N. I., & Abbas, E. B. (2018, August). Simulation and modelling of LoRa and Sigfox low power wide area network technologies. In *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1-5). IEEE.

[31] Chere, M., Ngqondi, T., & Bembe, M. (2019, January). Wireless Low Power Area Networks in the Internet of Things: A Glimpse on 6LoWPAN. In *2019 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-10). IEEE.

[32] Pai, V., & Shenoy, U. K. K. (2019). 6LoWPAN—Performance analysis on low power networks. In *International Conference on Computer Networks and Communication Technologies* (pp. 145-156). Springer, Singapore.

[33] Al-Kashoash, H. A., Kharrufa, H., Al-Nidawi, Y., & Kemp, A. H. (2019). Congestion control in wireless sensor and 6LoWPAN networks: toward the Internet of Things. *Wireless Networks*, *25*(8), 4493-4522.