ISSN 2347 - 3983
Volume 8. No. 9, September 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter09892020.pdf
https://doi.org/10.30534/ijeter/2020/09892020

Improving the Performance of Intrusion Detection System using Machine Learning based Approaches

Ghanshyam Prasad Dubey[1] Dr. Rakesh Kumar Bhujade [2]
[1] Ph.D. Research Scholar, [2] Ph.D. Supervisor, Department of Computer Science & Engineering, Mandsaur University, Mandsaur (M.P.), INDIA, ghanshyam_dubey2@yahoo.com

ABSTRACT

The purpose of artificial intelligence is to make machine intelligence and machine learning enables them to acquire knowledge. Machine learning (ML) a branch of artificial intelligence, make machine self-learner. This self-learning ability will help to solve many complex problems. Using a Machine learning Intrusion detection system can make it more efficient and capable to detect new attack patterns by self-learning or acquiring knowledge. IDS are the first line of defense that obtain information or knowledge from a network and analyze it to determine elements that are responsible for violating the security policies of computer and networks. In this paper importance of machine learning is discussed because of the betterment of the intrusion detection system.

Key words: Intrusion Detection System, Machine Learning, Learning Techniques, True Positive, False Negative, KDD-99

1. INTRODUCTION

The communication networks and information systems are much more important in the field of economic, social development, and various facets of our daily lives. Due to the enhancements in Internet services, networking and communication technologies, the number of attacks to the information systems are increased significantly. A system that is used to prevent from unauthorized access, alteration of information and render a system unreliable, timely monitor traffic of network; such systems are comes under the Intrusion Detection System (IDS). Prevention of such actions by IDS is done by raising an alarm or generating an alert or by some other specific approach. Most IDS suffers from the bottleneck of detecting new or unknown attack patterns; sometimes a false alarm is generated unnecessarily and sometimes a malicious activity remains undetected by the system [1]. An IDS is a suitable solution, a tool or resource for identifying, assessing, and claiming the unauthorized or malicious network activities.

1.1 Signature Based Detection

The signature-based methodology (Figure 1) compares the observed signatures to the signatures on the database. The repository that contains a list of known attack signatures called database. If a pattern built by extracting the features from the packets in a monitored environment is similar to one of the signatures on the database, the system will detect this activity as an attack or a violation to security policy and takes appropriate action to handle that attack. It is also known as misuse detection; here the characteristics of existing attacks forms the basis of detecting new or unknown attacks. An activity or event is detected as an attack or as suspicious, if it is already known or occurred in the past [2]. If a suspicious activity is known to the system, then only it can recognize the attack and generate an alarm. Signature-based detection approach fails to detect new or unknown attacks [3].

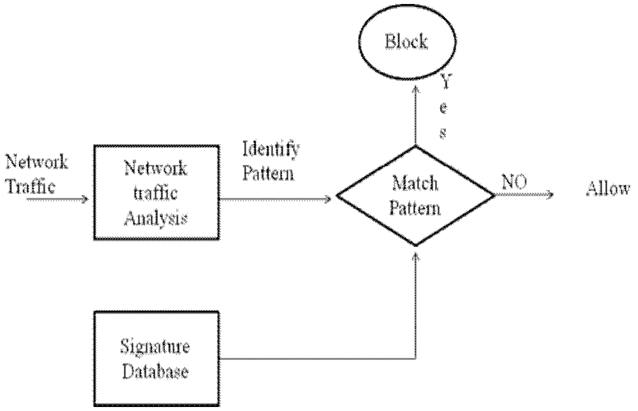

Figure 1: Working of signature based IDS.

1.2 Anomaly Detection

In Anomaly detection or behavior- based detection (Figure 2), the behavior of the network, users, and computer systems is modeled and an alarm is raised to justify an attack whenever components misbehave. Anomaly detection is capable to presuppose and to be able to identify a new attack type or kind of new prospective attacks. The primary working principle of anomaly detection is distance measurement through rebuilding profiles representing normal usage and then comparing it with their current behavior to find out a likely mismatch. Detection of any significant deviation from the normal behavior by the system will be specified as an attack and will be dealt accordingly. Anomaly detection is capable of detecting attack symptoms without specifying attack models; they can easily detect unknown or new attacks but are high sensitive to false alarms [4].

Based on the location of installation, IDS can also be classified as Host IDS, Network IDS and Hybrid IDS (combination of Host IDS and Network IDS incorporating the features of both).

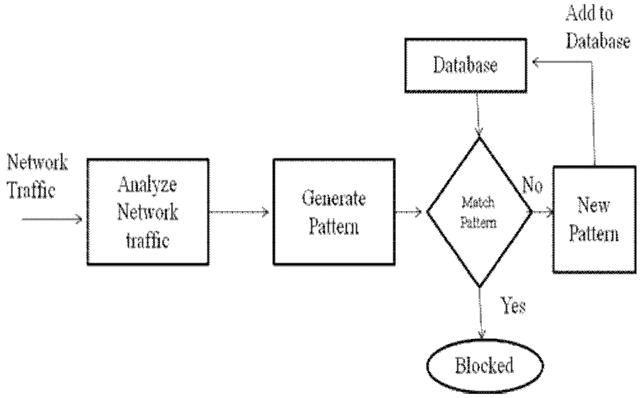

Figure 2: Working of Anomaly based IDS.

2. MACHINE LEARNING TECHNIQUES

Machine Learning is a branch of artificial intelligence to develop learning ability in the machine according to the situation and further applied this knowledge to solve the upcoming problems. Anomaly-based IDSs using machine learning techniques can implement a system that can learn from data (experience) and can classify or predict the behavior as normal or malicious. Some common machine learning techniques used to classify intrusive and non-intrusive behaviour include Neural Networks, Support Vector Machines (SVM), Genetic Algorithms, Fuzzy Logic, Bayesian Networks, Decision Tree, etc.

3. LITERATURE SURVEY

J. Jabez and B. Muthukumar [4] proposed an Anomaly-based Intrusion Detection approach using Outliers. K Nearest Neighbour Classifier is used to classify the Samples. They propose to develop the Anomaly-based IDS that is Precise and can sustain small variations in Patterns and Low False Alarms and at the same time is Adaptive and Real-Time. The proposed system is tested on the standard KDD [5] Data Set. Results show that the proposed ML-based IDS is better than other existing ML-based IDS and can detect almost all Anomaly Data. The Performance and Efficiency of the proposed system are also better than the existing approaches [4]. Javaid et al. [6] proposed an approach to trained Intrusion detection system using deep learning mechanisms. Self-taught Learning (STL) and deep learning-based technique used to improve the performance of the network intrusion detection system. Self-taught Learning (STL) is a kind of deep learning method consists of two classification stages. In the first step deal with the good classification of features from large unlabeled dataset (Unsupervised Feature Learning) and second step applied to leveled dataset for classification. They implemented sparse auto-encoder and soft-max regression-based NIDS using the NSL-KDD [7] dataset to calculate the accuracy of anomaly detection [6].

Adel Sabry Eesa et al (2015) used the Cuttlefish optimization algorithm for the features selection and intrusion detection system. The investigated and evaluated the performance of their proposed hybrid IDS model based on CFA and Decision Tree through feature selection on the benchmark KDD Cup 99 [5] intrusion data. In the first step, they modified the cuttlefish algorithm to obtained feature selection. In the second step, they used discrete transform classifiers as measurements on generated features. Results justify that the detection rate and accuracy of their hybrid model is very high, when the dataset has 20 or less features. In general, the reduction in number of features leads to increase in detection rate and accuracy. When all 41 features of the dataset are incorporated, still the cuttlefish feature-selection model gives much better performance in all cases when compared with the results of other state of the art approaches [8].

L. Koc et al (2012) proposed an Extended Hybrid Feature Selection Method based on 3 leading Filters along with Naive Baye’s Classifier. They extended the use of Naive Baye’s Classifier along with the Filters like Correlation-based Filters and Consistency based Filters. The three Filters used are Correlation Based Filter, Consistency Based Filter, and INTERACT Feature Selection Filter. Their approach is a kind of Extended Hybrid Naive Baye’s Classifier for Intrusion Detection. Results show that the approach improves the Accuracy of the System and at the same time, reduces the Resource Requirements of the Intrusion Detection System. Error Rate and Misclassification are also greatly reduced and detection of Denial of Service Attacks has been enhanced specifying that the Overall Performance of the System is enhanced [9].

R. A. R. Ashfaq et al (2016) proposed a Fuzzy based Semi-Supervised Learning Approach for Intrusion Detection. They proposed a Single Hidden Layer Feed Forward Neural Network with a Fast Learning mechanism, Random Weights and Fuzziness for Intrusion Detection. Their approach is based on the principle of Divide and Conquer Algorithm Design technique. The fuzziness of each Sample is evaluated and classified into 3 categories as High Fuzziness Samples, Low Fuzziness Samples, and Mid Fuzziness Samples. Samples with High and Low Fuzziness are used to retrain the System. Neural Network with Random Weights has shown an Excellent Learning Performance and is Computationally Efficient. The proposed system has shown a High Accuracy Rate for Samples with High and Low Fuzziness but it has shown Low Accuracy Rate with Samples having Mid Fuzziness [10].

A. A. Halimaa and K. Sundarakantham (2019) proposed Machine Learning based Intrusion Detection System. Their approach is based on Naïve Baye's Classifier and Support Vector Machines. There are 3 main phases in the operation of the proposed model as Data Set Pre Processing, Classification, and Result Evaluation. In Data Set Pre Processing, all Non-Numeric and Symbolic Features are removed or exchanged as they don't have any major part to play in the process of Intrusion Detection. Classification on the Pre Processed Data Set is performed using Naïve Baye's Classifier and Support Vector Machines separately. Results show that the SVM outperforms the NB Classifier to a great extent. SVM attains the Accuracy of more than 97% while NB Classifier has Accuracy below 70% and the Misclassification Rate of NB Classifier is also high as compared to SVM. Even after Normalizing the Data Sets and Feature Reduction, the average Accuracy of SVM is more than 96% and that of NB Classifier is around 68%. Results clearly show that SVM is a superior approach in all cases than NB Classifier for implementing the Intrusion Detection System [11].

Shadi Aljawarneh et al (2018) [12] suggested a hybrid model to approximate the intrusion scope threshold degree, derived from the network transaction data's optimal features, made available for training. The hybrid classification-based intrusion detection model used to find an efficient detection system and a feature selection is used to reduce the NSL-KDD data set's dimensions. Dimensionality Reduction increases the speed of execution and accuracy of system in detecting the attacks. The decision tree approach simply divides the classification problem into multiple sub-problems. A decision tree will be created and then utilized for developing a model that can perform the task of classification. The neural networks are responsible for approximating or estimating functions which normally relies on a huge training dataset. Results show that the dimensionality reduction in their model leads to an improvement in the accuracy rate and reduction intthe detection time. The analysis on the NSL-KDD dataset shows that the inherent drawbacks of the TCP protocol are responsible for causing majority of the attacks [12].

Sharmila Kishor Wagh and Satish R. Kolhe (2014) presented an effective semi-supervised self-learning approach to reduced false alarm rates and uplift the attack detection. This scheme also capable of labeled the features of intrusion data with flexible training and adaptation. The first step in this approach is to refine the KDD99 dataset and this newly generated refined dataset will be used to train the model. The model is trained using the refined labeled data, while the testing of model is being carried out using the unlabeled data. Then the datasets with high prediction values from testing phase's output are selected and added in the labeled data. This formulation of learned set plays an important role in eliminating the redundancies in labeled data with restricting the size. As far as result analysis is concerned this semi-supervised approach offers accuracy of 99.516 % and is best suited for DoS attack with respect to other attacks, as its obtained false positive rate is 0.102% [13].

Praveen Kumar Kollu and R. Satya Prasad [14] proposed a Gated Recurrent Unit (GRU) based recurrent neural network that leverages the attention mechanism. This model focuses on focus on the most optimal features for detecting subtle changes in the network traffic. Pavithra G et al [15] compared various machine learning techniques that are used for effective and efficient intrusion detection.

4. PROPOSED ALGORITHM

The recommended algorithm (Figure 3) is:

```
Step 1: Initialization of the KDD-99 data set
Step 2: Using principles of Correlation and Information Gain, reduce the size of Dataset (Feature Extraction)
Step 3: Train the Neural Network (EBPA) on the reduced data set [Supervised Learning / Training]
    REPEAT:
        • Initialize the Weights of the different Links connecting the various Nodes of Layer K with Layer (K + 1)
        • Compute the Weight Sum of Input [NET]
        • Apply Activation Function [OUT]
        • Compute ERROR (//ERROR is the difference between Desired Output and Actual Output)
        • If (ERROR is large)
            ✓ Apply the Weight Change of Links
            ✓ GO TO REPEAT
        • Else EXIT
Step 4: Initialize the Network and IDS
Step 5: Apply ACO algorithm [16]
Step 6: After completion of process, generate the results.
Step 7: Store the results in data base for future references.
Step 8: END.
```

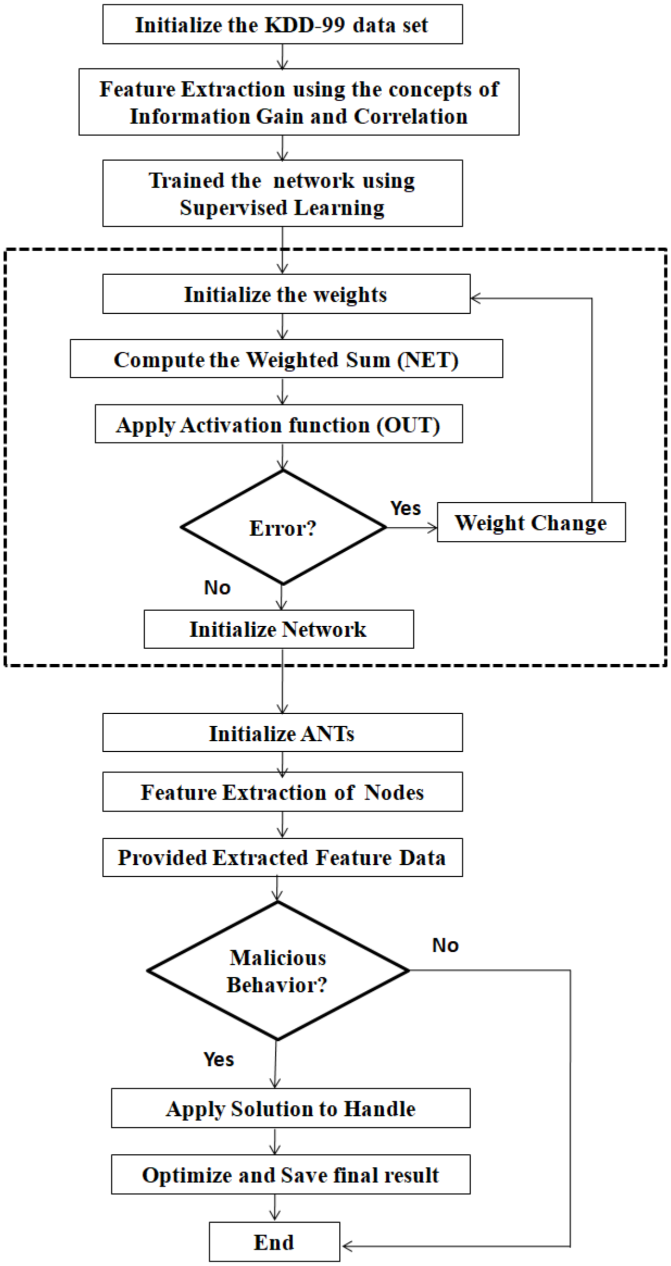
Initialize the KDD-99 data set
Feature Extraction using the concepts of Information Gain and Correlation
Trained the network using Supervised Learning
Initialize the weights
Compute the Weighted Sum (NET)
Apply Activation function (OUT)
Error?
Yes
Weight Change
No
Initialize Network
Initialize ANTs
Feature Extraction of Nodes
Provided Extracted Feature Data
Malicious Behavior?
No
Yes
Apply Solution to Handle
Optimize and Save final result
End

Figure 3: Proposed Algorithm for effective IDS.

5. EXPECTED OUTCOMES & COMPARATIVE ANALYSIS

An excellent IDS must possess improved true detection rate along with low false alarm rate and ability to detect new or unknown attacks. By this recommended approach, the author planning to improve the performance of the IDS to a certain extent. The expectations with the proposed approach is minimizing the processing time, reducing the training time, enabling self-attack detection, and applying the ACO based specifiers for enhancing the effectiveness in detecting intrusion. A system with machine learning ability can predict new attacks and enhance system security.

The table (Table 1) below shows the Comparative Analysis of the Expected Outcomes of the proposed approach with the Outcomes of the other state-of-the-art approaches.

Table 1: Comparative Analysis of Proposed Approach with Other Approaches

Approach	Type of IDS	Outcomes
1. Jabez [4]	Anomaly based IDS using KNN Classifier	~ Precise, Adaptive and Real Time System ~ Better Performance and Efficiency due to Low FALSE Alarms
2. Javaid [6]	Network IDS using Self Taught Learning and Deep Learning	~ Better Precision ~ Better Recall Rate ~ Better f – Measure
3. Eesa [8]	Cuttlefish Optimization based IDS	~ Improved Detection Rate ~ Improved Accuracy ~ Improved Performance
4. Koc [9]	Naïve Baye's Classifier based IDS	~ Improved Accuracy ~ Reduced Resource Requirements ~ Reduced Error Rate ~ Reduced Misclassification Rate ~ Detection of Denial of Service Attacks ~ Improved Overall System's Performance
5. Ashfaq [10]	Fuzzy based Semi Supervised IDS using Divide and Conquer approach	~ Excellent Learning Performance ~ Computationally Efficient ~ High Accuracy Rate for Samples with High and Low Fuzziness ~ Low Accuracy Rate for Samples having Mid Fuzziness
6. Halimaa [11]	ML based IDS using Naïve Baye's Classifier and SVM	~ Low Misclassification Rate ~ High Accuracy of approx. 97 %
7. Aljawarneh [12]	ML based Hybrid Network IDS	~ Fast ~ Efficient ~ Accurate

8. Kolhe [13]	Semi Supervised Self Learning based IDS	~ Accuracy of more than 99.5 % ~ Low False Positive Rate below 0.15 % ~ Best suited for Denial of Service Attack
9. Proposed Approach	IDS using enhanced EBPA and ACO.	Expected Outcomes ~ Improved TRUE Detection ~Minimize FALSE Alarm Rate ~ Reduced Training Time ~ Fast Processing ~ Optimal and Effective than other approaches

6. CONCLUSION

In this work, we presented an Optimal Intrusion Detection System (IDS) with Ant Colony Optimization (ACO) based on Error Back Propagation (EBPA) Neural Network. Machine Learning based Systems are highly Accurate and Adaptive; at the same time, are Self Learning in nature. Optimization Techniques are proved to be compromising for enhancing the Performance of the models to a greater extent. The main objective of this work is to develop an Intrusion Detection System that has better Performance than other approaches with High Sensitivity and Low Specificity. It also aims at reducing the Training Time and increasing the speed of operation.

REFERENCES

1. Rais, Helmi Md, and Tahir Mehmood. **Dynamic Ant Colony System with Three Level Update Feature Selection for Intrusion Detection.** International Journal of Network Security, vol. 20, no. 1, pp. 184-192, Jan. 2018.
2. Wu, Han-Ching, and Shou-Hsuan Stephen Huang. **Neural networks-based detection of stepping-stone intrusion**. Expert Systems with Applications, vol. 37, no. 2, pp. 1431-1437, Mar. 2010.
3. Lin, S.W., Ying, K.C., Lee, C.Y. and Lee, Z.J. **An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection**. Applied Soft Computing, vol. 12, no. 10, pp. 3285-3290, Oct. 2012.
4. Jabez, J., & Muthukumar, B. **Intrusion Detection System (IDS): Anomaly detection using outlier detection approach**. Procedia Computer Science, vol. 48, pp. 338-346, Jan. 2015.
5. KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, accessed on march 2020.
6. Javaid, Ahmad, Quamar Niyaz, Weiqing Sun, and Mansoor Alam, **A deep learning approach for network intrusion detection system**. Proc. 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), pp. 21-26, May 2016.
7. NSL-KDD dataset. https://www.unb.ca/cic/datasets/nsl.html. Accessed 15 August 2020.
8. Eesa, Adel Sabry, Zeynep Orman, and Adnan Mohsin Abdulazeez Brifcani. **A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems.** Expert Systems with Applications, vol. 42, no. 5, pp. 2670-2679, Apr. 2015.
9. Koc, Levent, Thomas A. Mazzuchi, and Shahram Sarkani, **A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier**, Expert Systems with Applications, vol. 39, no. 18, 13492-13500, Dec. 2012.
10. Ashfaq, Rana Aamir Raza, Xi-Zhao Wang, Joshua Zhexue Huang, Haider Abbas, and Yu-Lin He. **Fuzziness based semi-supervised learning approach for intrusion detection system**, Information Sciences, 378, pp. 484-497, Feb. 2017.
11. Halimaa, Anish, and K. Sundarakantham. **Machine Learning Based Intrusion Detection System**, Proc. 2019 IEEE 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 916-920, Apr. 2019.
12. Aljawarneh, Shadi, Monther Aldwairi, and Muneer Bani Yassein, **Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model**, Journal of Computational Science, 25, pp. 152-160, Mar. 2018.
13. Wagh, Sharmila Kishor, and Satish R. Kolhe, **Effective intrusion detection system using semi-supervised learning**, Proc. 2014 IEEE International Conference on Data Mining and Intelligent Computing (ICDMIC), pp. 1-5, Sep. 2014.
14. Kollu, Praveen, and R. S. Prasad. **Intrusion Detection System Using Recurrent Neural Networks and Attention Mechanism.** International Journal of Emerging Trends in Engineering Research, 7, pp: 178-182, 2019.
15. Pavithra G , Abirami P , Bhuvaneshwari S , Dharani S , Haridharani B. **A Survey on Intrusion Detection Mechanism using Machine Learning Algorithms.** International Journal of Emerging Trends in Engineering Research, Volume 8, No. 4, pp: 945-949, April 2020.
16. Dubey, Ghanshyam Prasad, and Rakesh Kumar Bhujade. **Impact of Ant Colony Optimization on the Performance of Network Based Intrusion Detection Systems: A Review**, International Journal of Scientific & Technology Research (IJSTR), vol.8, no. 9, pp. 1830-1834, Sep. 2019.