J. Sustain. Wireless Syst., vol. 03, no. 2, pp. 118-127
https://doi.org/10.36548/jsws.2021.2.006

Modified Gray Wolf Feature Selection and Machine Learning Classification for Wireless Sensor Network Intrusion Detection

Subarna Shakya
Professor,
Department of Electronics and Computer Engineering,
Central Campus, Institute of Engineering,
Tribhuvan University,
Pulchowk, Lalitpur, Nepal.
drss@ioe.edu.np

Abstract: The ability of wireless sensor networks (WSN) and their functions are degraded or eliminated by means of intrusion. To overcome this issue, this paper presents a combination of machine learning and modified grey wolf optimization (MLGWO) algorithm for developing an improved intrusion detection system (IDS). The best number of wolves are found by running tests with multiple wolves in the model. In the WSN environment, the false alarm rates are reduced along with the reduction in processing time while improving the rate of detection and the accuracy of intrusion detection with a decrease in the number of resultant features. In order to evaluate the performance of the proposed model and to compare it with the existing techniques, the NSL KDD'99 dataset is used. In terms of detection rate, false alarm rate, execution time, total features and accuracy the evaluation and comparison is performed. From the evaluation results, it is evident that higher the number of wolves, the performance of the MLGWO model is enhanced.

Keywords: Machine learning, Support vector machine, Intrusion detection, Wireless sensor networks, Grey wolf optimization algorithm

1. Introduction

General purpose computation elements, sensors and small actuators are available in the heterogeneous wireless sensor network (WSN) system [1]. Self-organizing, low-power, low-cost wireless nodes are available in hundreds or thousands in a WSN for monitoring and controlling the corresponding environment. Security, robustness, scalability, reliability and

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

self-healing characteristics must be taken into account while developing a WSN system. Military applications, earthquake monitoring, monitoring performance in manufacturing machines, ocean monitoring and several such applications make use of WSNs [2-3]. Monitoring water quality, building security, highway traffic, pollution and other futuristic applications can make use of WSN architecture and principles. The raw information may be grouped and aggregated efficiently with the help of WSN. The base station is a centralized control unit of a WSN. A WSN may have one or more base station units. The base station has multiple functionalities. It offers an access point for human interface, enables data processing, acts as a storage unit and a gateway to another network. Data may be extracted from the network and the control information may be disseminated using the base station [4]. It is also referred to the sink. A routing forest is built by a combination of the sensor nodes. The base station contains the root of each tree. When compared to the sensor nodes, the power and storage capacity of the base station is high. Communication outside the WSN facilitated by strong processors with better speed and performance, storage memory for saving cryptographic keys and battery power to sustain throughout the lifetime of the sensor nodes are facilitated by the base station [5].

A robust and accurate IDS model is created by selecting the most pertinent features using an appropriate feature selection (FS) scheme [6]. From a given dataset, certain attributes and redundant features are removed to identify the relevant features using FS. Improving the detection performance and data dimensionality reduction are the significant objectives of FS. Along with redundant features, the data presentation makes use of several features in real time applications [7]. Segregation of extra features is made possible as the role certain of features will be taken by other specific features [8]. The behavior of the dataset and other essential data are available at the relevant features while directly impacting the output of the system [9]. Conventionally, in high dimensional space, obtaining the best set of features using a comprehensive search has been impractical. A combinatorial optimization problem is modelled using the FS by several researchers where an optimal feature space is obtained from the set of given features [10].

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

2. Related Works

Over any type of network, in order to enhance its security an intrusion detection system (IDS) is essential [11]. The threats imposed on the network in terms of intrusion or by the hosts are detected and prevented by providing a high level of security using this system. Detection of new attacks and ensuring adaptability is the primary goal of these system. Anomaly IDS and misuse IDS are the major classifications of IDS [12]. New attacks are identified by employing signatures in misuse IDS, while intelligent methods and statistical patterns are used to identify if the behavior is healthy or not using anomaly IDS. Anomaly detection based IDS is introduced using several schemes that may classify the abnormality and normality using artificial intelligence algorithms and other smart classification schemes [13]. The attack detection is ensured by the IDS classifier and the detection process is enhanced with intelligent computation. Accuracy, false alarm rate, detection rate and pattern are the factors used for characterization of each classifier. Along with the computational cost and the classifier model complexity based issues, high false alarm rate, low detection rate and accuracy are the drawbacks of most classifiers [14]. The detection process and its performance are affected by massive storage capacity, long processing time, high complexity and high classification level [15]. The IDS mainly faces issues relating to reduction of processing cost and enhancement of classifier performance which has to be improved significantly. The challenges involved in overhead classification may be overcome by improving the efficiency of the detection process by reducing the dataset dimensionality. Such problems may be overcome largely by using feature selection process [16].

In order to secure the infrastructure and data, the WSN based applications must ensure high security level. Intrusion and abnormal behavior may be detected by the IDS. Distributed sensors in a WSN gather data and transmit it to the node at the base station [17]. This information is not completely protected from external attackers using cryptographic security schemes. IDS acts may as a second level defense mechanism in this scenario for continuous monitoring of network traffic. When the sensors show any malicious activity, alert is sent to the base station. Various problems in FS has been solved efficiently using The Grey Wolf Optimization (GWO) algorithm. High accuracy and performance may be obtained by selection of essential features with maximum relevance and minimum redundancy thereby reducing the data dimensionality using feature selection [18]. The efficient utilization of classification

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

process is ensured by optimal selection of features while addressing the over-fitting issues and providing appropriate data for noise recognition using the feature selection schemes. Selection of prominent features without irrelevance or redundancy while reducing the dataset dimensionality may be performed using the preprocessing optimization process called feature selection. No over-fitting problem, a data out noise, data storage capacity, minimum processing time is ensured while avoiding the classification overhead and enhancing the accuracy as well as classification performance is obtained using the features that are selected resembling the optimal subset [19].

3. Proposed Work

The main objective of this analysis is achieved by the novel intrusion detection scheme proposed in this paper. A valid and verified NSL-KDD dataset commonly used for intrusion detection testing is used for experimentation. As shown in figure 1, data acquisition is the initial stage of the quantitative scheme. With this dataset, the WSN based intrusion detection model and the steps involved is represented in figure 1. The dataset is categorized into training and testing datasets comprising of 80% and 20% of the overall dataset respectively. Feature selection, classification and evaluation are the three major stages in the data processing as shown in figure 1.

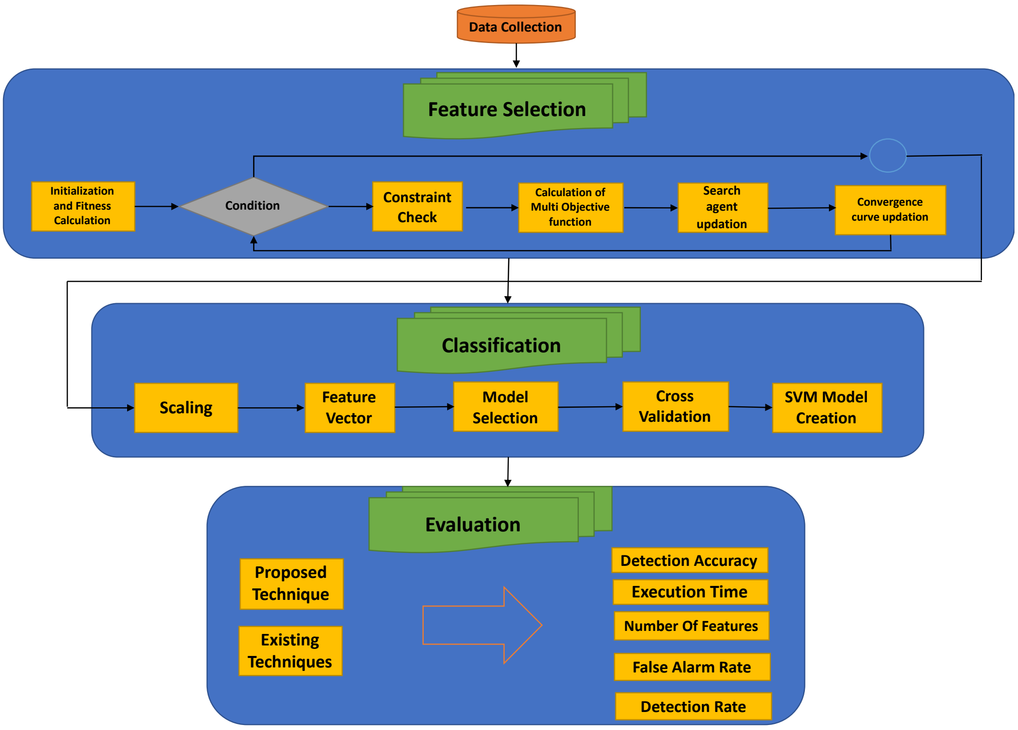

Fig. 1. Proposed IDS architecture

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

NSL-KDD'99 dataset is used for evaluation of the proposed model. Various IDS have been tested using this dataset for analyzing the performance in terms of global security. The lack of redundant records is a major advantage leading to unbiased classification using this dataset. When compared to existing IDS techniques, the proposed model offers improved intrusion detection rate. User to root (U2R) and remote to local (R2L) attacks, DoS attacks and probe attacks are included in the dataset. The dataset is normalized making it more appropriate for deployment in classification and feature selection schemes. With lesser requirements of hardware processing, faster processing is achieved by dataset size reduction. During data pre-processing, the non-numerical values are encoded and numerical values are normalized. Further, normalization is performed on the encoded non-numerical values whereby scaling of the content is done.

A modified grey wolf optimization algorithm is used for feature selection enabling optimal feature set identification through attribute selection. The hunting and leadership mechanism of grey wolves in real world forms the basis of the meta-heuristic feature selection algorithm termed as GWO. It consists of different types of wolves inclusive of alpha, beta, delta and omega wolves. Wolves often reside in packs. Each pack size may range between five and twelve. The first place is occupied by the alpha wolves while the second and third places are taken by beta and delta wolves and the least expected solution is provided by the omega wolves. In general, optimization is performed by the first three wolves while tracking is performed by the fourth. In the modified GWO algorithm we determine the fitness of a subset with respect to the core goal using a multi-objective fitness function and increase the number of wolves. The reduced dataset with lesser features obtained from the algorithm are fed as an input the SVM classifier. The classes are categorized on identifying the hyper-plane using this classifier.

4. Results and Discussion

The proposed technique is implemented on a MATLAB environment for experimental purpose. Various systems and applications are modelled and analyzed using the MATLAB tool. Massive dataset volumes can be analyzed while providing optimal solution using this environment. Integration of MATLAB codes with those written in other programming languages is made possible for deployment of algorithms in enterprises and web platforms.

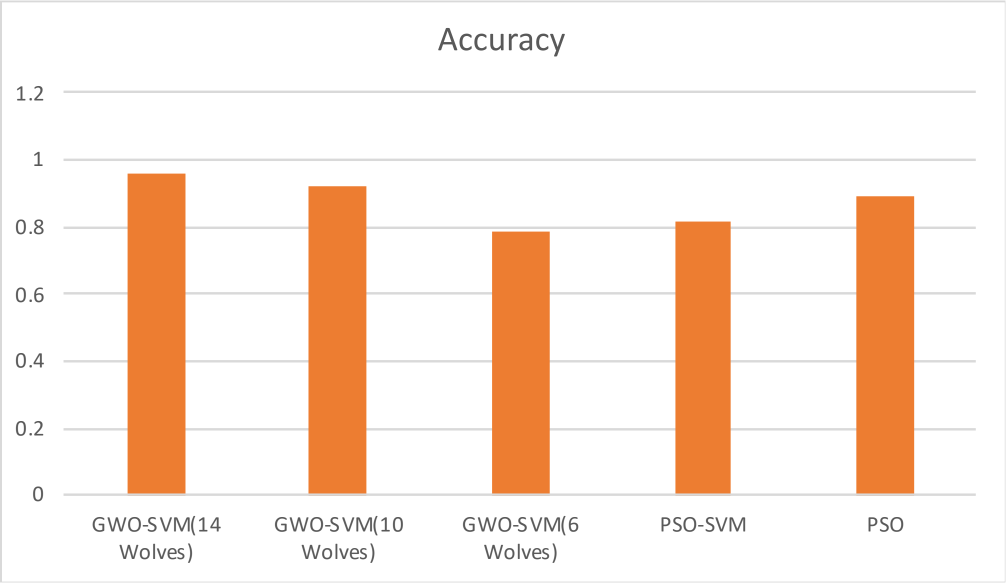

Fig. 2. Accuracy based evaluation and performance comparison

Detailed evaluation of the proposed model and comparison with existing PSO and conventional GWO schemes are performed and represented graphically. The model with the best configuration of wolves is analyzed by varying the number of wolves. From the results of evaluation, it is evident that the system is more optimized with an increased number of wolves. This helps in overcoming the drawbacks like slow search, solution diversity and premature conference. The optimization process is affected by the variation in wolf size. The tradeoff between exploitation and exploration searches has to be improved by selecting the most appropriate number of wolves. Figure 2 represents the evaluation and comparison of performance based on accuracy when the number of wolves are set at 6, 10 and 14 along for the proposed model along with the existing PSO and PSO-SVM models. The classification accuracy of the proposed multi-objective function increases correspondingly as accuracy is given high weight compared to the other features.

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

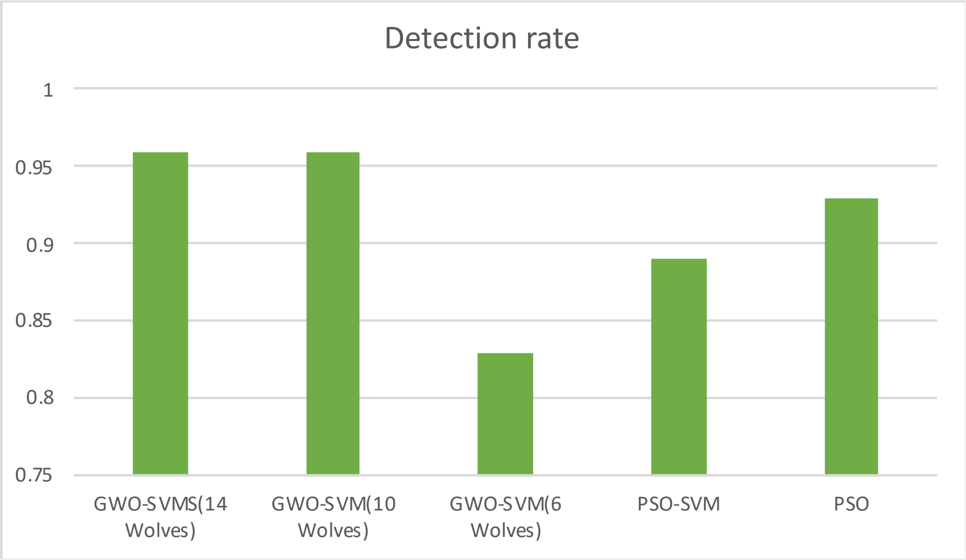

Fig. 3. Detection rate based evaluation and performance comparison

The proposed technique also outperforms the other models in terms of false alarm rate. The total features as well as false alarm rate is reduced in the WSN environment with this IDS. Figure 3 provides the comparison of performance evaluation with respect to detection rate. It is observed that the overall detection rate decreases with the increase in the complexity of the model. However, with the increase in the number of wolves, we achieve a decreased feature count in the WSN based IDS environment. The execution time also is reduced by the proposed model when compared to the existing schemes. The overall metrices and their enhancement percentage is shown in figure 4. Detection rate, false alarm, execution time, total features and accuracy are estimated and represented graphically. It is evident from these results that the increase in the number of wolves enhances the performance of the model. The execution time is also reduced with the increased number of wolves. Table 1 represents the estimated values obtained from the results.

Table 1. Comparison of the proposed model with existing schemes

Technique	Total Features	Execution time	False alarm	Detection rate	Accuracy
GWO-SVM 14 wolves	8	58	0.02	0.97	0.97
GWO-SVM 10 wolves	10	70	0.03	0.96	0.96
GWO-SVM 6 wolves	12	75	0.09	0.95	0.92
GWO	28	85	0.2	0.85	0.80
PSO-SVM	20	130	0.3	0.83	0.90
PSO	25	135	0.38	0.81	0.95

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

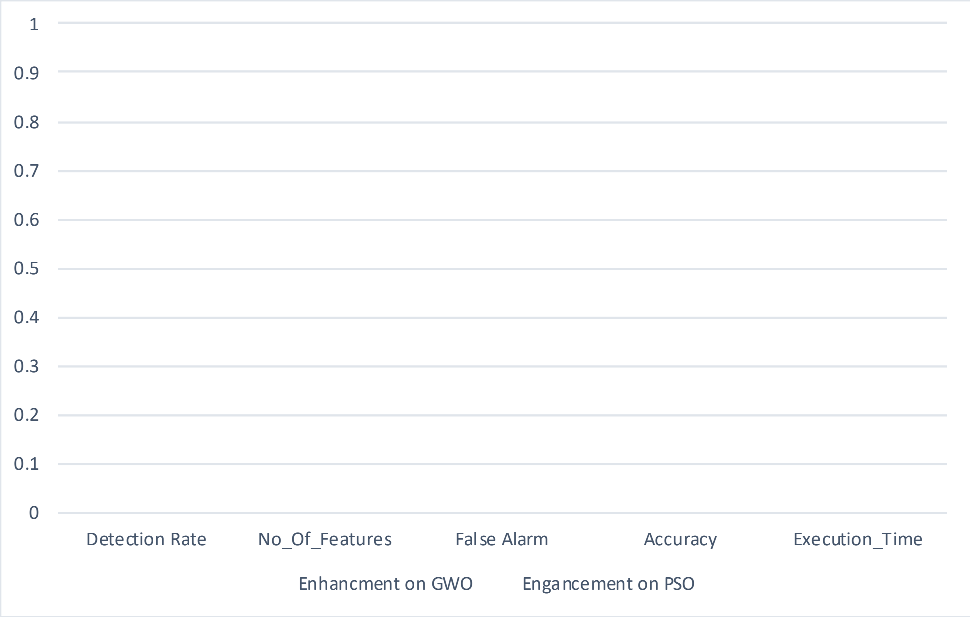

Fig. 4. Comparison of performance based on rate of enhancement

5. Conclusion

IDS and various other fields commonly use the robust and widespread GWO feature selection algorithm. The rate of intrusion detection and classification accuracy may be improved largely by selecting the most significant features using this algorithm. The performance enhancement of the GWO based IDS scheme has been proposed by several researchers. However, in terms of execution time, total features selected and accuracy, several shortcomings are faced that affects the effectiveness of the algorithm. In this paper, the total wolves is increased in combination with a machine learning SVM based classifier that contributes towards performance enhancement of GWO based IDS scheme in the WSN environment. The prediction system and its overall performance is enhanced using a multi-objective function. The execution speed, total features selected and accuracy factors are enhanced based on the its efficiency in prediction of unknown classes. Training and testing of the proposed model is performed using the NSL KDD'99 dataset and comparison of the results with existing IDS techniques using PSO and GWO is performed. The comparison results show that the performance of the proposed model is better in terms of execution time, total features selected, false alarm rate, detection rate and accuracy when compared with the existing models. Future work is focused on enhancing the classification performance with other deep learning schemes.

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

References

[1] RM, S. P., Maddikunta, P. K. R., Parimala, M., Koppu, S., Gadekallu, T. R., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. Computer Communications, 160, 139-149.

[2] Mugunthan, S. R., & Vijayakumar, T. (2021). Design of Improved Version of Sigmoidal Function with Biases for Classification Task in ELM Domain. Journal of Soft Computing Paradigm (JSCP), 3(02), 70-82.

[3] Dutta, S., & Banerjee, A. (2020). Highly Precise Modified Blue Whale Method Framed by Blending Bat and Local Search Algorithm for the Optimality of Image Fusion Algorithm. Journal of Soft Computing Paradigm (JSCP), 2(04), 195-208.

[4] Wilson, A. J., & Giriprasad, S. (2020). A Feature Selection Algorithm for Intrusion Detection System Based On New Meta-Heuristic Optimization. Journal of Soft Computing and Engineering Applications, 1(1).

[5] Smys, S., Basar, A., & Wang, H. (2020). Hybrid intrusion detection system for internet of Things (IoT). Journal of ISMAC, 2(04), 190-199.

[6] Baraneetharan, E. (2020). Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey. Journal of Information Technology, 2(03), 161-173.

[7] Çavuşoğlu, Ü. (2019). A new hybrid approach for intrusion detection using machine learning methods. Applied Intelligence, 49(7), 2735-2761.

[8] Shakya, Subarna. "Process mining error detection for securing the IoT system." Journal of ISMAC 2, no. 03 (2020): 147-153.

[9] Kunhare, N., Tiwari, R., & Dhar, J. (2020). Particle swarm optimization and feature selection for intrusion detection system. Sādhanā, 45, 1-14.

[10] Bashar, Abul. "Sensor Cloud Based Architecture with Efficient Data Computation and Security Implantation for Internet of Things Application." Journal of ISMAC 2, no. 02 (2020): 96-105

[11] Tubishat, M., Idris, N., Shuib, L., Abushariah, M. A., & Mirjalili, S. (2020). Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. Expert Systems with Applications, 145, 113122.

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021

[12] Jacob, I. J., & Darney, P. E. (2021). Artificial Bee Colony Optimization Algorithm for Enhancing Routing in Wireless Networks. Journal of Artificial Intelligence, 3(01), 62-71.

[13] Davahli, A., Shamsi, M., & Abaei, G. (2020). Hybridizing genetic algorithm and grey wolf optimizer to advance an intelligent and lightweight intrusion detection system for IoT wireless networks. Journal of Ambient Intelligence and Humanized Computing, 11(11), 5581-5609.

[14] Rahman, M. A., Asyhari, A. T., Wen, O. W., Ajra, H., Ahmed, Y., & Anwar, F. (2021). Effective combining of feature selection techniques for machine learning-enabled IoT intrusion detection. Multimedia Tools and Applications, 1-19.

[15] Chen, D. J. I. Z., & Lai, K. L. (2020). Internet of Things (IoT) Authentication and Access Control by Hybrid Deep Learning Method-A Study. Journal of Soft Computing Paradigm (JSCP), 2(04), 236-245.

[16] Mugunthan, S. R. (2020). Decision Tree Based Interference Recognition for Fog Enabled IOT Architecture. Journal of trends in Computer Science and Smart technology (TCSST), 2(01), 15-25.

[17] Zhou, Y., Cheng, G., Jiang, S., & Dai, M. (2020). Building an efficient intrusion detection system based on feature selection and ensemble classifier. Computer Networks, 174, 107247.

[18] Shakya, S. (2020). Analysis of artificial intelligence based image classification techniques. Journal of Innovative Image Processing (JIIP), 2(01), 44-54.

[19] Abdulhammed, R., Musafer, H., Alessa, A., Faezipour, M., & Abuzneid, A. (2019). Features dimensionality reduction approaches for machine learning based network intrusion detection. Electronics, 8(3), 322.

ISSN: 2582-3167
Submitted: 05.05.2021
Revised: 21.05.2021
Accepted: 7.06.2021
Published: 14.06.2021