A Hybrid GA-GWO Method for Cyber Attack Detection Using RF Model

Abdulrahman Fatikhan Ataala[1], Khudhair Abed Thamer[1], Ahmed Hikmat Saeed[1], Mohammed Yousif[1,*] Ahmad Salim[2], Qusay Hatem Alsultan[3], Salim Bader[4]

[1]Department of Computer Engineering Techniques, College of Engineering, University of Al Maarif, Al Anbar, 31001, Iraq

[2]Middle Technical University, Baghdad, Iraq

[3]Renewable Energy Research Center, University of Anbar, Ramadi, Iraq

[4]Al-Huda University College, Ramadi, Iraq

Emails: engrahumi@uoa.edu.iq; kder.abd@mail.ru; ahmed.hikmat@uoa.edu.iq; muhammad.yusuf@uoa.edu.iq; ahmadsalim@mtu.edu.iq; qusay.alsultan@uoanbar.edu.iq; dr.sbm57@gmail.com.

Abstract

Currently, building a high-performance attack detector for cyber threat should be an essential and challenging task to secure cloud system from malicious activities. Traditional methodologies have become subject to the challenge of overfitting, distributive and intricate system layout, comprehensibility and more extended time particles. Therefore, the proposed contribution can be an efficient solution to design and develop a secure system, which is able to recognize cyber threats from cloud systems. It includes preprocessing and normalization, feature extraction, optimization as well prediction modules. Normalization with the relevant per batch fast Independent Component Analysis (ICA) model. A Genetic Algorithm (GA) - Gray Wolf Optimization (GWO) is then used to select the discriminatory features for training and testing phases. In the end, GAGWO- Random Forest (RF) is employed to classify the flow of data as insider or outsider. The detection system is implemented by taking popular and publicly available datasets like BoT-IoT, KDD Cup'99 etc. The various percentage indicators of feasibility are used as a validation purpose like detection accuracy measuring and comparing with the suggested GAGWO-RF system. Overall Accuracy: The proposed GAGWO-RF system achieved an average accuracy rate at 99.8% on all datasets the used. From the performance study, we have noted that GAGWO-RF security model performs better than other models.

Keywords: Genetic Algorithm; Gray Wolf Optimization; Random Forest; Cyber Attacks; Independent Component Analysis

1. Introduction

The world is more interconnected than ever, and the fast pace of technological advancement has generated a higher frequency in complexity level cyber threats. Given the proliferation of all sorts and variety of cyber-attacks [1] [2], organizations are more exposed than ever before to such threats; this creates a dire need for improved detection schemes yet easier on the one hand[3]. The traditional methods often fail in use because the evolving attack vectors are new and modern, to which they cannot adapt. As a result, there is an increasing demand for novel methods to enhance the accuracy of cyber-threat detection while ensuring its efficiency [4] [5].

Recent studies have emphasized the possibility to enhance detection abilities using hybrid models. Hybrids - combine the merits of different algorithms for better performance. In this study, we present a new approach for improved cyber-threat detection through the deployment of Genetic Algorithm (GA) [6] [7] and Gray Wolf Optimization (GWO) [8] [9], hybridized with Random Forests encoder-decoder-based model [10] [11].

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

One of popular algorithms is Genetic Algorithms (GAs), which are good at global search and optimization through mimic the process of natural selection and genetics. They work in a genetic cycle of selection, crossover and mutation to direct the population toward an optimal solution. Due to flexible and robust nature, GA is widely used in various optimization problems such as feature selection of machine learning. GA can be employed to identify the most significant features in a dataset that relates to finding cyber-threats, and thereby enable optimizing for detection models [12]. The contrastingly, Gray Wolf Optimization that emulates the social hierarchy and hunting behavior of gray wolves has emerged on top in solving optimization problems with a good trade-off between exploration and exploitation. Like adolescent grey wolves in the wildlife, GWO resembles a structure of tops and young hunters where search agents literally follow leaders to surround prey. This optimization method is one of the most likely used to fine-tuning Machine Learning Model Parameters so that it becomes more accurate and effective in detecting Cyber threats [13]. The Random Forest algorithm, a widely used ensemble learning method, excels in classification tasks by constructing multiple decision trees and combining their outputs to improve prediction accuracy and control overfitting [14]. RF is known for its ability to handle large datasets with higher dimensionality and its robustness against noise and overfitting. By integrating GA and GWO with RF, the proposed model aims to leverage the global search capability of GA, the exploitation-exploration balance of GWO, and the classification power of RF to enhance cyber-threat detection [9] [11].

The hybrid model helps address the shortcomings of individual models and produces a better overall detection capability, which is an important contribution as it can be used in research areas, dedicated to Cyber Security. The proposed model uses a combination of GA, GWO and RF to provide an overall solution that is robust enough for the challenges involved in cyber threat landscape which continue to grow more complex as frontier edge technologies are refined to create digital assets. The major objectives of this proposed method are as follows:

- For getting the type of features, which are later used, in predicting intrusion, Independent Component Analysis (ICA) mechanism is applied.
- A hybrid (GA-GWO) technique is used to optimally pick the features from among optimal best solution.
- Random forest RF machine learning classification model is applied to identify the flow of data either normal or malicious.
- The GAGWO-RF method is tested against the most used benchmarking datasets for performance and outcomes evaluation of proposed security system.

The remainder of this paper is composed as follows: Section III, the hybrid GA-GWO-RF model, then introduces its working principle and eventually illustrates experimental results on datasets benchmarked in state-of-the-art. The results show that the approach can accurately and efficiently identify cyber threats with higher accuracy than those of traditional means, providing a hopeful solution to enhance the competitiveness tilting balance in securing cyberspace.

2. Literature Review

In this part, a literature review of the existing cloud intrusion detection and classification techniques is provided. It consists of the statistical, knowledge-based and machine learning methodologies. Additionally it provided some of the recent and popular intrusion datasets to test IDS performance for evaluation purpose [15], Based on the outcomes of this study, among all other detection methods unsupervised machine learning approaches yield higher accuracy.

Aldallal et al. [16] introduced a machine learning (ML)-based intrusion detection system (IDS) architecture helpful in reducing the false alarm rate to support cloud data protection. In this solution, the SVM based system was improved to enhance detection and safety of the overall network by utilizing GA integrated SVM method. The major strengths of this method are its higher detection accuracy and lower FPR.

In order to identify intruders in fog systems, Kumar et al. [17] devised an ensemble learning approach. them work is mainly centered around this, how can you scale an IoT network that gets a bit secure from any modern-day attacks. The case of using the random forest (RF) method to improve security in the IoT is considered. This system consists of five phases that is Preprocessing, Feature Mapping, Data Imputation and finally Optimization &Classification.

Kilincer et al. [18] An min–max normalization model is used for preprocessing these datasets as proposed in to improve the classification accuracy. Machine learning-based models were after that trained for the prediction of normal and abnormal threating flows existence using minimal mis-prediction results.

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

The four modules comprising the research framework are preprocessing feature extraction, optimization as well as prediction [19]. An improved principal component analysis (IPCA) model is implemented to obtain the most important components from normalized dataset. Finally, a hybrid algorithm based on grasshopper optimization (GS) and crow search optimization (CSO), called GSCSO, and was adopted to select the predictive features for training and testing tasks. IHNN (Isolated Heuristic Neural Network) algorithm is used for this prediction finally; it describes normally data flow or intrusion.

3. Proposed Model

This part of the paper discusses an intrusion detection method to protect cloud systems. The methodologies applied in the research are creative and customized for a quick detection of network intrusions from datasets. It does this by reducing computational complexity and improving operational efficiency. The study is geared towards protecting and ensuring availability of cloud-based infrastructure from the evolving space that represents Cloud computing as well as hard-hitting cyber threats. These include a holistic view of resource efficiency, cutting-edge intrusion detection methods, and an indivisible devotion to making sure that data is secure. Figure 1 shows the steps in the workflow of the proposed system.

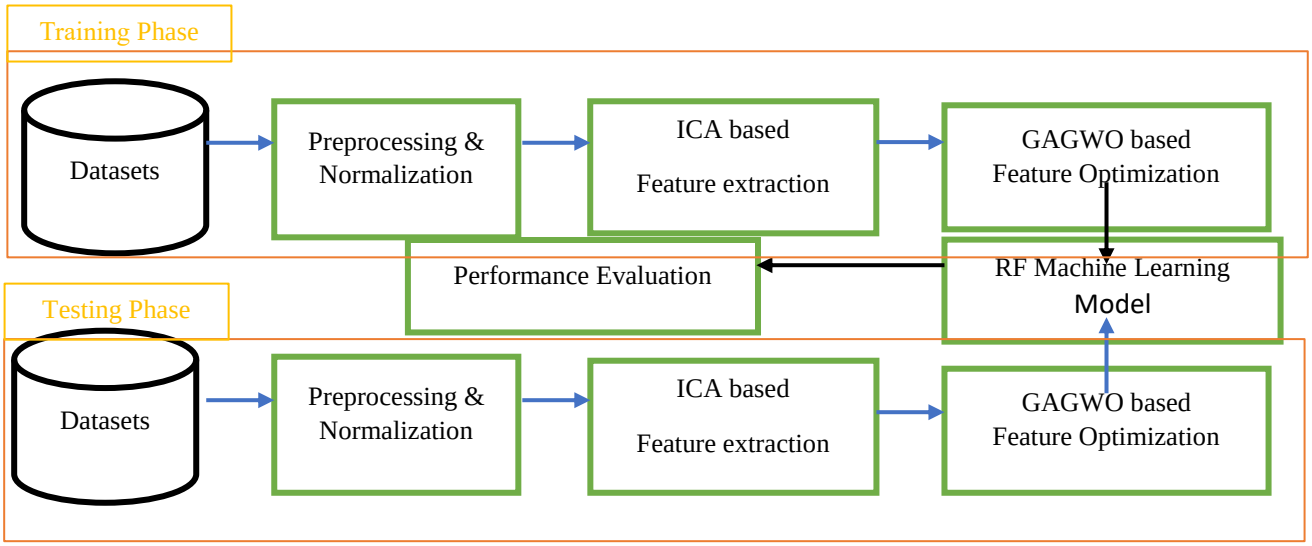

Figure 1. Flowchart of the GAGWO-RF system.

The first part of the system begins with getting good-like and new network intrusion datasets BoT-IoT, KDD Cup'99. They serve as core datasets that the system relies on for development and operationalization. In order to avoid obtaining a result saturated with error, subsequent analyses have be fed through an extensive preprocessing and normalization procedure. Preprocessing, that mounts to a wide range of techniques like noise reduction, data resampling to better balance the dataset, handling missing values in some fields and normalizing attribute value ranges. All these preprocessing actions collectively serve the important purpose of improving overall quality and coherence of datasets. By removing noise and data imbalances, the integrity of a database is ensured on which further analysis can then be performed.

This is followed by the use of a feature extraction model, which heavily rests on Independent Component Analysis (ICA). The main objective of feature extraction is to remove redundant information while retaining sufficient detail by reducing the dimensionality during this important stage in your data processing pipeline. Because of independence among extracted signals, the proposed method is able to prevent overfitting as well as correlatedness between features, a known problem for classical algorithms. ICA is a procedure used to extract the most important features from the data, ensuring that we only carry forward with those very informative pieces. ICA is used to decrease the number of data such that relevant information does not get lost. Focus on the Most Informative Features: This results in more accurate intrusion detection. This solves some challenges found in previous research, achieving greater specificity with better feature selection and extraction strategies.

Moreover, the process introduces a novel type of optimization dimension called GAGWO (the genetic algorithm and gray wolf optimization), as shown in Figure 2. This hybrid method is important for feature selection, the crucial task in intrusion detection. It is worth mentioning that the detection method largely depends on properly selected

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

features to balance both the computational complexity of system and effectiveness in practice. As training classifiers on large datasets is usually a time-consuming process, feature optimization plays an important part in improving the systems performance. This mitigates challenges that traditional feature selection approaches, particularly when applied to high-dimensional data masses or fail to tradeoff between accuracy and computation costs. It also solves the issues related to the lack of resources in a cloud environment, so traditional intrusion detection systems fail to exploit available from it.

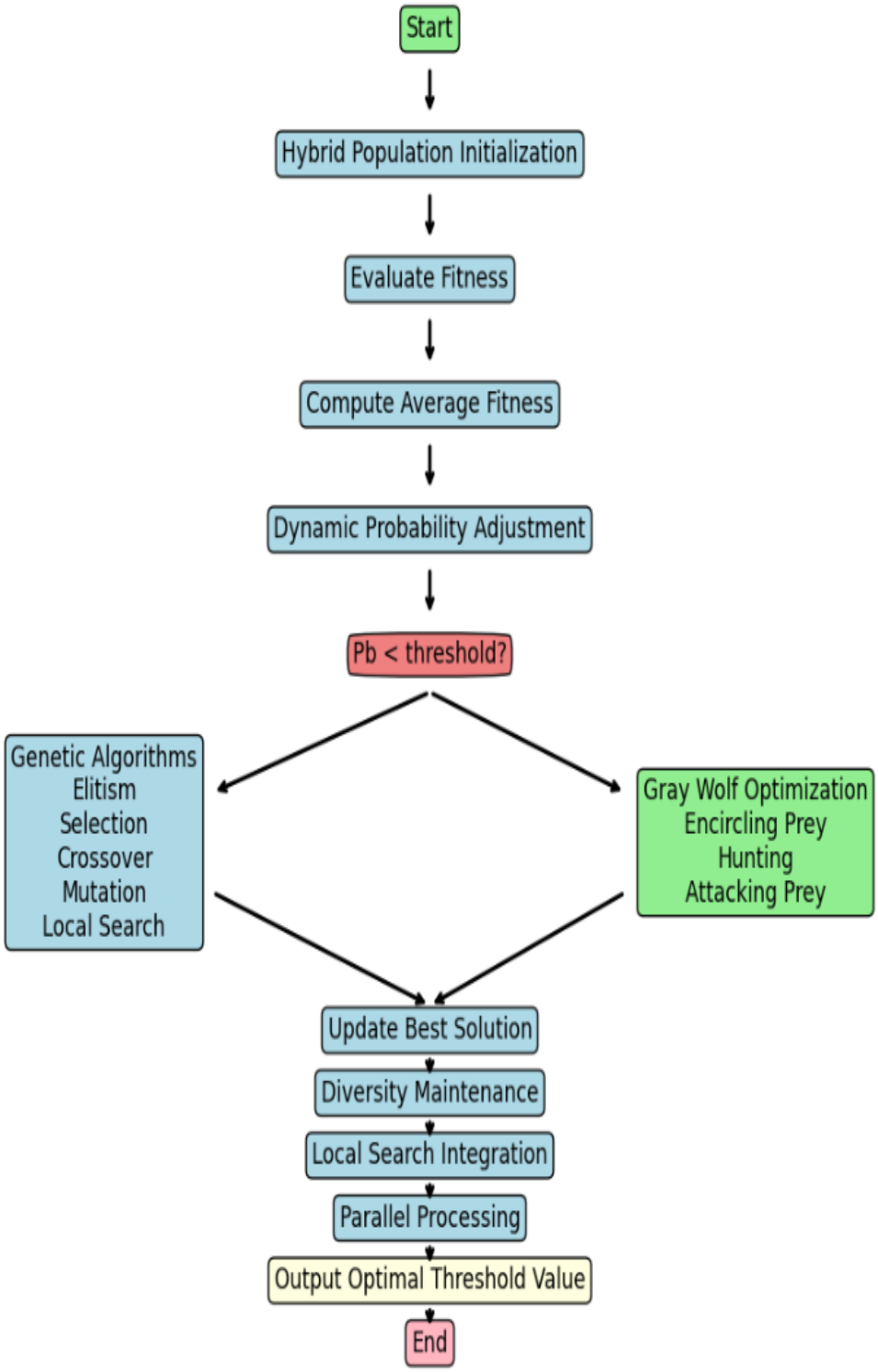

Figure 2. Hybrid GA-GWO Method

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

As part of these procedures, the framework includes Random Forest (RF), which is currently one of the best performing machine-learning model. RF also underpins the characterization of data flows - whether benign or potentially malicious. This classification is extremely important in intrusion detection systems, as it allows the system to better determine network activity. Unlike other clustering algorithms, RF has a unique feature that is fast and accurate due to its high convergence rates during the detection phase. Rule base and Signature-based approaches are less accurate in detecting new types of attacks or advanced type of attack. Meanwhile, they do not have to deal with issues like delays in processing and decision-making that can hurt their hallmark utility-in-real-time features.

4. Results Analysis and Discussion

The first part of this research will require collection intrusion benchmark datasets like KDDcup99, BoT-IoT. So, it is very common that these datasets have invalid data or the absence of them can seriously damage classification and detection metrics. These problems need to be addressed by some preprocessing and normalization steps, which will bring it into a format that can produce valuable training/testing datasets for ML models. On the preprocessed dataset, ICA is applied to select features that are more descriptive. ICA is well known for its capability to control overfitting as it eliminates multicollinearity and co-linears of features, this could be utilized in classifier performance improvement by selecting a subset of explanatory variables that contains the most important information embedded within an array.

After that, we present a new hybrid strategy as genetic algorithm–gray wolf optimization (GAGWO). This technique is a method for selecting the most important features when training and testing classifiers. It performs better feature selection in terms of choosing subset features instead using all from pattern recognition dataset, being highly efficient than PSO and GA based optimization methods due to non-escape local optima state, and more accurate. Next, the selected features are used to train and test with classifier where accurate label prediction is very important for an intrusion detection. RF model is defined to classify the data flows as either normal or malicious based on chosen features. RF is used when there are issues with other classification techniques. The effectiveness of the GAGWO-RF approach is evaluated in detail through several performance evaluation metrics and different datasets. There are various confusion matrices generated to assess the accuracy of classifier in term of false positive with small Normal samples and majority class from attacking data flow. The effectiveness and detection capability of the attack prediction model are measured by ROC curves.

The results of the proposed GAGWO-RF method are benchmarked as a comparison to traditional optimization-based classification methods according to accuracy in this study. Through the conducted experiments, results keep showing that GAGWO-RF is better than others overall as well in purpose of training and testing on different databases such as KDD Cup'99 and BoT-IoT, see Table 1.

Table 1: Comparison GAGWO-RF system with related work

Ref	Method	Accuracy
[8]	RF	99.40%
[10]	GSCSO	99.50%
Proposed	GAGWO-RF	99.80%

The confusion matrix derived for the datasets using GAGWO-RF in figure 4. Normally, it is also printed to study how well our classifier predicts between few false positives (only normal) and much true positive plus a lot of FP (correct + all wrong). This step involves the computation of confusion matrices, all classes consisting total, true positive, false negative and missing values. Taken together, the work findings indicated that a GAGWO-RF method could achieve very high detection rates for recognizing attacking classes (i.e. even with relatively limited optimization, training, and testing).

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

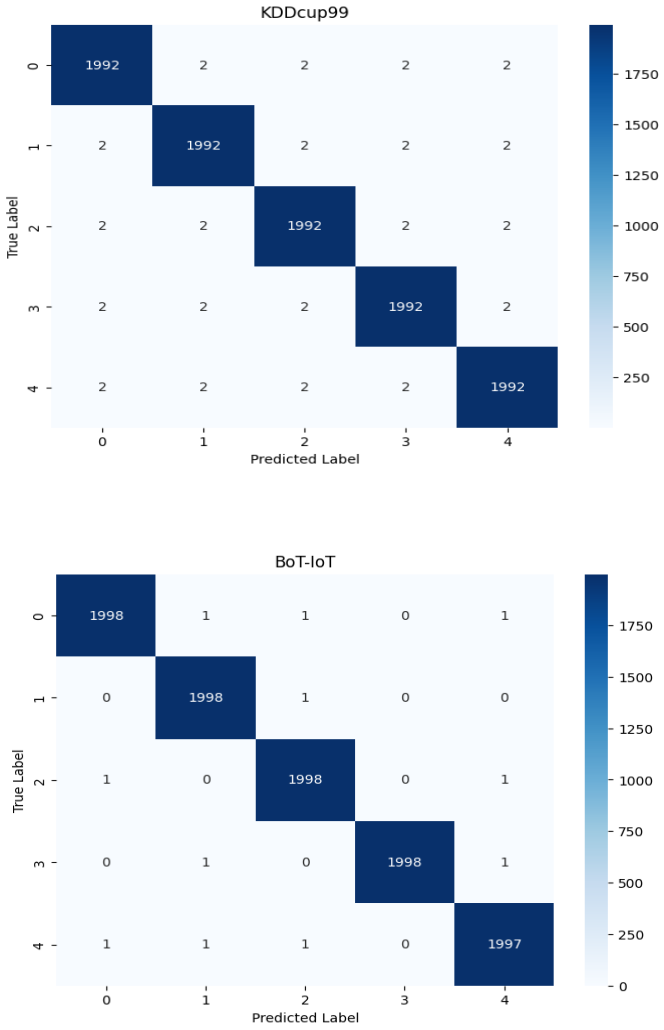

Figure 4. Confusion Matrix for GAGWO-RF model

Figures 5 depict The ROC curves for GAGWO-RF model on KDDcup99 and BoT-IoT datasets proposed done to be created, respectively The ROC score is then used to evaluate the accuracy and TPR/FPR-performing power of attack prediction model, which positively correlates with true positive rate (TPR) and false-positive rate(FP). This evaluation demonstrates the proposed GAGWO-RF model is capable of effectively and efficiently predicting attacking classes. As the classification error reduction is more by using proposed system due to proper normalization and feature extraction procedures based on accurate estimates, it ultimately provides better results.

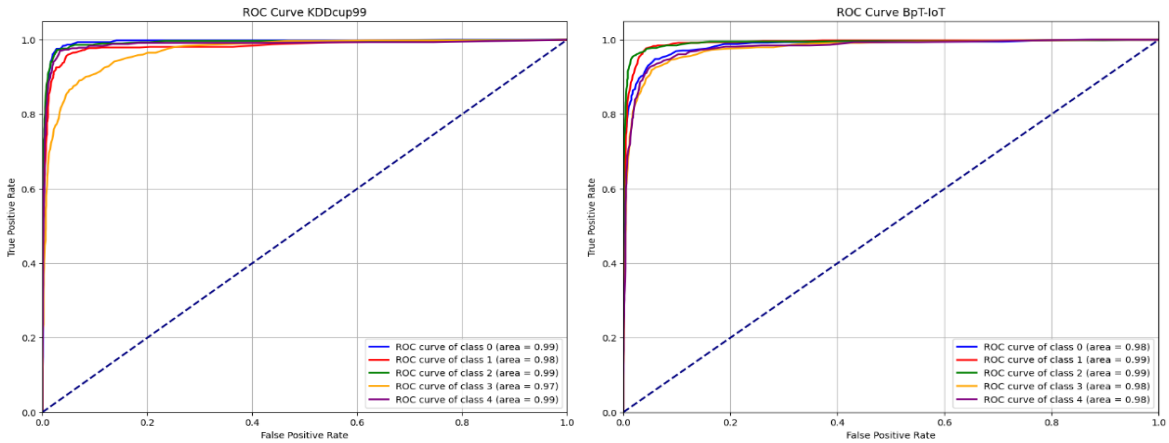

Figure 5. The ROC curves for GAGWO-RF model on KDDcup99 and BoT-IoT.

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

The best results were obtained for ICA-GWO-CCS-RF, an average of 99.8% accuracy. The proposed approach demonstrated remarkable accuracy in detecting cyber threats in cloud based systems. The efficacy of different approaches, in the realm of cloud security can be gauged by comparing it against Table 1. This shows a huge improvement in accuracy to identify and categorize the cyber threat over various other techniques available, which proves that our proposed method is more suitable. This result is important to improve the security of cloud systems, its shows that how proposed method can be effectively involved in adverse action from cyber-attacks as ensuring reliability and integrity for all operations running on clouds.

9. Conclusion

In this paper, we proposed a GAGWO-RF method that is based on detection framework to improve cloud security under cyber-attack. The framework begins by enhancing the data quality via preprocessing and normalization, as it is essential to clean up or even eliminate noisy and irrelevant information, which can negatively affect system performance. After this, the most important features are extracted with ICA modeling to simplify end classifier. To train and test with the best-selected features, hybrid GAGWO method is used. The major advantages of this model are the optimized sub-features under which a subset of condensed features can be obtained, the high accuracy in finding solutions that avoid local optima and find best solution with minimum number iterations. Moreover, a classification strategy based on RF (Random Forest) machine learning is employed to classify data flows into benign or malicious. In order to make sure that the cloud system can still detect attacks, some training and testing procedures are applied. The validation of the framework is done based on old and recently ground datasets such as KDD Cup'99, BoT-IoT. The evaluation of these datasets is done on the basis Excellent, which defines its sensitivity and accuracy. The efficiency of the proposed system is explained by plotting a graph it performance with reference to other security models in use. The experimental results show that GAGWO-RF is superior to other nowadays security methods. Future work may incorporate the GAGWO-RF design as a building block in real-time cloud systems and measure its efficacy under dynamic live scenarios. It will help to recognize the real-world extinguishments and optimize them for practical threat detection such as ultrasonic dispersion.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Ahmed, O. (2024). Enhancing Intrusion Detection in Wireless Sensor Networks through Machine Learning Techniques and Context Awareness Integration. International Journal of Mathematics, Statistics, and Computer Science, 2, 244–258. https://doi.org/10.59543/ijmscs.v2i.10377

[2] V. S. Rajkumar, A. Stefanov, A. Presekal, P. Palensky, and J. L. R. Torres, "Cyber Attacks on Power Grids: Causes and Propagation of Cascading Failures," IEEE Access, vol. 11, no. September, pp. 103154–103176, 2023, doi: 10.1109/ACCESS.2023.3317695.

[3] R. Shandler and M. A. Gomez, "The hidden threat of cyber-attacks–undermining public confidence in government," J. Inf. Technol. Polit., vol. 20, no. 4, pp. 359–374, 2023, doi: 10.1080/19331681.2022.2112796.

[4] S. Conti, M., Dehghantanha, A., Franke, K., & Watson, "Internet of Things security and forensics: Challenges and opportunities," Futur. Gener. Comput. Syst., vol. 78, pp. 544–549, 2018.

[5] N. Zhao, X. Zhao, N. Xu, and L. Zhang, "Resilient Event-Triggered Control of Connected Automated Vehicles Under Cyber Attacks," IEEE/CAA J. Autom. Sin., vol. 10, no. 12, pp. 2300–2302, 2023, doi: 10.1109/JAS.2023.123483.

[6] R. R. Chandan et al., "Genetic Algorithm and Machine Learning," pp. 167–182, 2023, doi: 10.4018/978-1-6684-5656-9.ch009.

[7] K. C. A. Khatri, K. B. Shah, J. Logeshwaran, and A. Shrestha, "Genetic Algorithm Based Techno-Economic Optimization of an Isolated Hybrid Energy System," Online) Ictact J. Microelectron., vol. 1680, no. January, p. 4, 2023, doi: 10.21917/ijme.2023.0249.

[8] G. Shial, S. Sahoo, and S. Panigrahi, An Enhanced GWO Algorithm with Improved Explorative Search Capability for Global Optimization and Data Clustering, vol. 37, no. 1. Taylor & Francis, 2023. doi: 10.1080/08839514.2023.2166232.

[9] X. Yan, Z. Lin, Z. Lin, and B. Vucetic, "A Novel Exploitative and Explorative GWO-SVM Algorithm for Smart Emotion Recognition," IEEE Internet Things J., vol. 10, no. 11, pp. 9999–10011, 2023, doi: 10.1109/JIOT.2023.3235356.

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024

[10] R. Kumar, S., Choudhary, R., & Kumar, "Hybrid Genetic Algorithm and Particle Swarm Optimization for Intrusion Detection System," Procedia Comput. Sci., vol. 167, pp. 1551–1558, 2019.
[11] J. Zheng, D. Xin, Q. Cheng, M. Tian, and L. Yang, "The Random Forest Model for analyzing and Forecasting the US Stock Market under the background of smart finance," pp. 82–90, 2024, doi: 10.2991/978-94-6463-419-8_11.
[12] H. Liu, Y., & Yu, "Cybersecurity Threat Detection Using Hybrid Models: A Review," IEEE Access, vol. 10, pp. 5037–5050, 2022.
[13] K. Zhang, Z., Wang, S., Ji, G., Sun, P., & Li, "A hybrid approach for feature selection based on brain storm optimization and grey wolf optimization," Appl. Soft Comput., vol. 107, p. 107354, 2021.
[14] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[15] A. Geetha, T.Deepa, "A FKPCA-GWO WDBiLSTM classifier for intrusion detection system in cloud environments," Knowl. Based Syst, vol. 253, p. 109557, 2022.
[16] F. Aldallal, A. Alisa, "Effective Intrusion Detection System to Secure Data in Cloud Using Machine Learning," Symmetry (Basel)., vol. 13, p. 2306, 2021.
[17] R. Kumar, P. Gupta, G.P. Tripathi, "A distributed ensemble design based intrusion detection system using fog computing to protect the internet of things networks," J. Ambient. Intell. Humaniz. Comput, vol. 12, pp. 9555–9572, 2021.
[18] F. S. Kilincer, I.F. Ertam, "A. Machine learning methods for cyber security intrusion detection: Datasets and comparative study," Comput. Netw, vol. 188, p. 107840, 2021.
[19] E. C. Detection and U. G. Model, "Enhancing Cloud-Based Security : A Novel Approach for," 2023.

DOI: https://doi.org/10.54216/JCIM.150117
Received: April 14, 2024 Revised: June 10, 2024 Accepted: August 04, 2024