An Adaptive Hybrid Intrusion Detection Framework Using GoogleNet, Gradient Boosting, SVM, and Ant Colony Optimization for DDoS and Botnet Attack Detection

Dr. Rengarajan A[1, 2], Dr. Jyoti Sekhar Banerjee[3, 4]

[1] Postdoctoral Scholar, Lincoln University College, Malaysia
[2] Professor, School of Computer Science and IT, Jain (Deemed-to-be) University, Bangalore, India
[3] Lincoln University College, Malaysia
[4] Techno Bengal Institute of Technology, Kolkata, India
Email ID: [1] pdf.rengarajan@lincoln.edu.my / [2] a.rengarajan@jainuniversity.ac.in / [3] pdfsv.jsbanerjee@lincoln.edu.my / [4] jyotisekhar.banerjee@bitcollege.in

Abstract
With the growth of cloud computing, Internet-of-Things (IoT) and distributed network architectures, myriad distributed cyberattacks including Distributed Denial-of-Service (DDoS) and Intrusion via Botnets have been given a turbocharge. Because of the limitations on scalability, feature representation and adaptability, traditional intrusion detection systems (IDS) don't necessarily capture all the time the new and subtle attacks. In this context this research is motivated towards designing Adaptive Hybrid Intrusion Detection Framework - Deep learning (DL) and machine learning (ML) schemes with optimization based feature selection scheme to accurately and timely detect the attacks.
The proposed scheme is with hybrid approach with different types of classifiers (GoogleNet, Gradient Boosting (GB) and Support Vector Machine (SVM)) used in a single framework. The Ant Colony Optimization (ACO) is used as the main feature selection and optimization technique to determine which are the most relevant traffic attributes to be considered and reduce computation complexity. Using GoogleNet, deep hierarchical feature extraction is performed on the network traffic representations and the classifiers utilized are Gradient Boosting along with SVM for efficient classification of malicious traffic and benign traffic. It is then proposed to adopt the ensemble decision mechanism to improve its robustness and generalization in the various attack scenarios.
Two benchmark intrusion detection datasets (CICIDS2017 & Bot-IoT) are used for evaluation. Experimental test results show that the proposed hybrid model has a significantly higher accuracy, precision, recall and FPR when compared to individual classifiers. At the same time, within the hybrid system, the accuracy of detection is guaranteed at >98%, including effective low-rate DDoS and stealthy traffic from botnets. Furthermore, the use of explainable artificial intelligence methods enhances the interpretability and trust in operations.
It is noted that, optimized feature selection methods along with hybrid DCNN and CNN can be an apt solution, which is scalable and adaptive for deploying IDS of modern era in cloud and edge computing environment for optimized feature selection.

Keywords
Intrusion Detection System, DDoS Attack, Botnet Detection, GoogleNet, Gradient Boosting, Support Vector Machine, Ant Colony Optimization, Cybersecurity

1. Introduction

The digital communication technology, cloud computational system and the Internet-of-Things (IoT) ecosystems have significantly evolved leading toward an internetworked world [1] [2]. All these

developments have made for connectivity, scalability and data access, but introduced significant cyber security challenges. In the list of cyber threats, Distributed Denial-of-Service (DDoS) attacks and botnet-based attacks continue to be some of the most hazardous and disruptive attacks that impact the modern network. The objective of this attack is to deny the network access, suck the organization and critical infrastructures, resources, and throw them tremendous operational or financial losses [3] [4].
The DDoS attack has evolved from the conventional flooding attack into multifaceted attacks, which are able to evade old fashioned security solutions [5]. With the sophistication of the attackers, the utilization of application-layer DDoS attacks and slow-rate attacks and encrypted malicious traffic aimed at evading detection are growing. Similarly, botnets employ networks of compromised computers and use them to 'botnets' other computers, all while in the guise of being a legitimate network traffic, making this quite difficult to detect. Besides the number of IoT devices has also vastly grown creating opportunities for attackers to build botnets [6] [7] using a large number of vulnerable IoT devices.

2. Literature Survey

Intrusion Detection Systems (IDS) are extremely critical in understanding the network traffic and identifying suspicious activities. IDS System can be very broadly divided into Signature-Based IDS and Anomaly-based IDS. Signature Based IDS detection is based on a list of pattern and signatures of known attacks. These systems are successful when targeting known attacks, but are unable to protect against zero-day and polymorphic attacks. On the other hand, anomaly-based IDS attempt to detect anything in the network which deviates from the normal behaviors. These systems however tend to have high false positive rates and will have difficulty in differentiating between legitimate anomalies and actual attacks.
The machine learning (ML) approaches are becoming more promising intelligent intrusion detection solutions in that they can directly learn detection pattern from network traffic. The algorithms such as Support Vector Machines (SVM), Gradient Boosting (GB) have been worked out so well and the classification performance is very well in cyber security applications. SVM are especially successful in creating the nonlinear decision boundary for binary classification, as well as multiclass classification problems and the Gradient Boosting provides performance boost by reducing errors step by step and combining different models over time [8] [9].
although all the above advantages can be obtained when using ML based models, a large number of handcrafted features are required and scalability is hard to achieve due to the complex hierarchical structure of traffic pattern relations which are not easily extractable from non-ML based traffic analysis methods. To get over this problem, deep learning techniques have been employed that can gain automatically symbolic representations of underlying features from large amounts of network data traffic. Convolutional Deep Neural Network (CDNN) is used as an excellent feature extraction and pattern recognition. Since its inception modules are capable to deep hierarchical learning and very easy to compute, GoogleNet has become very popular among deep learning architectures. GoogleNet is able to extract the local as well as global traffic features making it well suitable to the intrusion detection task entails complex attack behavior [10].
Another important issue in the process of IDS are feature selection. Redundancy, irrelevancy and noise result from the huge amount of data gathered from a massive network traffic, and make data detection accuracy and computation efficiency difficult. To overcome this challenge, more and more, it is employing an optimization algorithm to solve this challenge. Ant Colony Optimization (ACO) is an efficient search strategy for metaheuristic optimization that can find the best subset of feature applied as ants foraging behavior in finding the optimal solution by using a pheromone-based search method. Another method for finding the optimal performance of the model is using the most discriminatory traffic features while reducing the dimensionality and calculation complexity [11].
In recent years, hybrid (i.e., combination) IDS design which fuses deep learning, machine learning and optimization techniques at the end of boosting the detection accuracy has been a point of research

interest. But, many of the existing systems are only working on one scale or unable to work with multiple attacks scenarios. Moreover, IDS mechanisms will be able to be implemented in the field with high detection rate and a low false alarm rate and complexity.

In order to overcome these drawbacks, this research has come up with a GoogleNet - Gradient Boosting - SVM - Ant Colony Optimization Adaptive Hybrid Intrusion Detection Framework. To improve detection rate of the DDoS attacks and botnets, the proposed system optimizes the features, and hybrid classification of methods are used. For this, two evaluation benchmark datasets (namely, CICIDS2017 and Bot-IoT) are used, which aims to improve the robustness and generalization of the framework in different attack scenarios.

3. Methodology

The purpose of this proposed Adaptive Hybrid Intrusion Detection Framework is to detect the Distributed Denial-of-Service (DDoS) and botnet attacks in an accurate, scalable and real-time manner in the modern network environment. The design combines various components such as deep learning, machine learning, and optimization, all working together to efficiently process large-scale network traffic data. It is comprised of several inter-related phases: data acquisition, data preprocessing, feature optimization, hybrid classification and decision analysis. Each phase is accountable to accomplish a certain job and additionally contributes to the framework becoming solid and discernable.

The introduced framework aims to counter the limitation of the existing IDS framework which is incapable to detect subtle and stealthy cyber-attack. Traditional systems are usually signature based and/or anomaly based, and inflexible with lots of false positives. To solve the above issues, the proposed architecture includes selection of features, dimensionality reduction by Ant Colony Optimization, classification of attacks by Gradient Boosting and Support Vector Machine classifiers and deep hierarchical feature extraction with GoogleNet. These techniques synergize to let the framework accurately identify complex traffic patterns, with minimal computational burden. Overall proposed system can boost the performance of the Intrusion Detection with optimized feature selection and Hybrid method of Classification, thus eliminating unnecessary computation overhead.

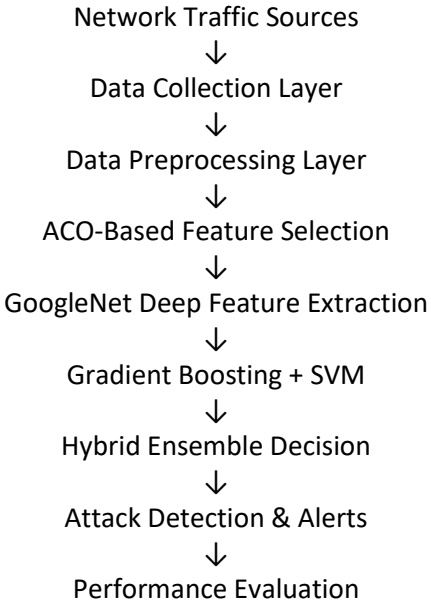

Figure 1: Adaptive Hybrid Intrusion Detection Framework for DDoS and Botnet Attack Detection Using ACO, GoogleNet, Gradient Boosting, and SVM

4. Results and Analysis

4.1 Results on CICIDS2017 Dataset

The result achieved for CICIDS2017 data set are displayed here.

The performance of the proposed Hybrid framework and individual classifiers while using CICIDS2017 dataset is shown in Table 1.

Table 1: Performance Evaluation on CICIDS2017 Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	95.4	94.6	94.9	94.7
Gradient Boosting	97.2	96.8	97.0	96.9
GoogleNet	98.1	97.6	97.9	97.7
Proposed Hybrid Model	98.8	98.4	98.6	98.5

The results show that GoogleNet can outperform the other traditional ML models because of its great performance to deep hierarchical feature extraction. A higher overall accuracy and F1-score is obtained, though, thanks to the proposed hybrid model enjoyed the ensemble based decision fusion strategy.

4.2 Results on Bot-IoT Dataset

Based on the experimental results presented in the Table 2: The Bot-IoT dataset has been used to obtain the experimental results.

Table 2: Performance Evaluation on Bot-IoT Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	94.8	94.0	94.3	94.1
Gradient Boosting	96.9	96.4	96.6	96.5
GoogleNet	97.8	97.2	97.5	97.3
Proposed Hybrid Model	98.5	98.1	98.3	98.2

Comparing the performance of the hybrid architecture with that of the standalone classifiers for Bot-IoT dataset shows that the hybrid architecture consistently outperforms the standalone classifiers. The results demonstrate the framework's usefulness in identification of IoT Botnets and complicated DDoS traffic patterns.

4.3 False Positive Rate Analysis

Table 3: False Positive Rate Comparison

Model	CICIDS2017 FPR (%)	Bot-IoT FPR (%)
SVM	4.2	4.5
Gradient Boosting	2.7	3.0
GoogleNet	2.1	2.4
Proposed Hybrid Model	1.3	1.5

The hybrid-based scheme is compared to the other results, and, as far as the false positive rate is concerned, the hybrid-based scheme provides the best results in both the datasets. While using this ACO based feature optimization along with ensemble classification technique the benign and malicious traffic classification gets improved.

5. Conclusion

As part of this research, we presented an Adaptive Hybrid Intrusion Detection Framework using GoogleNet, Gradient Boosting, Support Vector Machine (SVM), Ant Colony Optimization (ACO) for effectively detecting the DDoS and Botnet attacks. This framework solved the crucial problems with traditional IDSs, including high false positive rate, lack of adaptability with the changing trend of cyber-attacks and limited the scaling issues.

The computational efficiency and classification performance has been promoted, as the intelligent feature selection and dimensionality reduction has been accomplished by using Ant Colony Optimization with good success. Deep hierarchy feature sets were achieved through the use of GoogleNet from the network

traffic data, and Gradient Boosting and SVM were used to enhance the accuracy of attacks classification and stability of the decision making. The detection results with the ensemble-based detection mechanism were also good with a robust approach at various attack scene situations.

The experimental analysis using CICIDS2017 and Bot-IoT datasets demonstrated that the proposed hybrid system outperformed the individual classifiers in comparison in terms of the accuracy, precision, recall, F1-score, and False positive rates. It was capable of detecting the high volume attacks of the DDoS and identification of stealthy Bot traffic with a detection accuracy of over 98%.

The outcomes validate that a fusion of Deep Learning, ML and Optimization techniques present scalable and smart solution for modern IDSs. Future work then will include integration of federated learning and encrypted traffic analysis, real-time deployment in the cloud-edge networks, to enhance the security performance in upcoming networks.

References

1. S. Shafiq, M. S. Farooq, and M. A. Shah, “Comparative analysis of classification algorithms for DDoS detection in cloud environment,” *IEEE Access*, vol. 8, pp. 134527–134539, 2020.
2. M. Adil, N. Javaid, and Z. Rehman, “Feature selection using statistical methods for botnet detection,” *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 1, pp. 105–116, Mar. 2021.
3. I. D. Thaseen and C. A. Kumar, “An efficient feature selection algorithm for network intrusion detection using PCA,” *Comput. Electr. Eng.*, vol. 89, p. 106886, 2021.
4. R. Vinayakumar, K. P. Soman, and P. Poornachandran, “Applying deep learning approaches for network traffic classification and intrusion detection,” *ProcediaComput. Sci.*, vol. 132, pp. 1668–1677, 2020.
5. S. Aljawarneh, M. Aldwairi, and M. B. Yassein, “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model,” *J. Comput. Sci.*, vol. 25, pp. 152–160, 2020.
6. H. He, Y. Zeng, and J. Wang, “Real-time botnet detection using LSTM and traffic flow features,” *IEEE Access*, vol. 8, pp. 217905–217915, 2020.
7. B. Sharma, M. Gupta, and A. Saxena, “DDoS attack detection using CNN in SDN,” *IEEE Trans. Netw. Serv. Manag.*, vol. 17, no. 3, pp. 1475–1485, Sep. 2020.
8. Y. Wang, Y. Li, and X. Guo, “A transformer-based approach for network intrusion detection,” *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3276–3285, Mar. 2021.
9. A. Hafeez, M. A. Jan, and Y. Cao, “Autoencoder-based hybrid intrusion detection system in IoT,” *IEEE Access*, vol. 8, pp. 119821–119829, 2020.
10. S. Alharbi, M. Anbar, and K. Khowaja, “Scalable cloud-based IDS for dynamic environments,” *IEEE Cloud Comput.*, vol. 7, no. 4, pp. 62–71, Jul.–Aug. 2020.
11. L. Zhang, J. Sun, and J. Zhang, “Edge intelligence for real-time DDoS detection,” *IEEE Netw.*, vol. 34, no. 6, pp. 24–29, Nov.–Dec. 2020.