OSE-IDS: Optimized Stacked Ensemble Intrusion Detection System using Automated Machine Learning Approach

Mohamed A. Salama, Radwa M. Tawfeek, Sara Hamdy, Omar Salim.

Electrical Engineering Department, Benha Faculty of Engineering, Benha University,Benha, Egypt.

mohamed.salama, radwa.tawfeek, sara.hamdy, omar.salim {@bhit.bu.edu.eg}

***Abstract*— As cybercrimes are increasingly evolving, the existence of an intelligent Network intrusion detection system (NIDS) is indispensable in the network infrastructure. In addition, there are many challenges face NIDS design based on Artificial Intelligence – technology such as irrelevant features in network traffic, rare examples of malicious traffic, and the efforts for Machine learning model selection and models’ hypermeters finetuning. This study proposes efficient NIDS concerned with these challenges to accurately detect malicious behaviors. First, A parallel hybrid feature selection approach filters the most important features. Second, to address data imbalance, we integrated a combined Random Under-sampling Strategy and Synthetic Minority Oversampling Technique—Edited Nearest Neighbors technique to ensure the balanced representation of minority attacks. Finally, the stacked ensemble classifier, comprising the four best base models selected through the Automated Machine Learning approach. Using the CICIDS2017 dataset, a comprehensive benchmark for intrusion detection research, our approach achieves an impressive detection rate of 99.76%, effectively identifying both majority and minority classes.**

**Index Terms—NIDS, Anomaly Detector, Optimal Feature Selection, Imbalance Dataset, SMOTE, Ensemble Classifier.**

# Introduction:

Cybersecurity has become a great concern in today's digital age, as the frequency and sophistication of cyber-attacks continue to rise [1]. With the rapid integration of technologies like cloud computing and the Internet of Things (IoT), networks are increasingly vulnerable to attacks that threaten confidentiality, integrity, and availability.

NIDS are essential for safeguarding network infrastructures by monitoring traffic and alerting security admins to threats. There are two main types: anomaly-based and signature-based NIDS [2]. While signature-based systems effectively detect known attack patterns, they struggle against new or sophisticated threats, emphasizing the necessity for more adaptive solutions to identify unseen and zero-day attacks in an evolving threat landscape.

To overcome traditional NIDS limitations, artificial intelligence (AI) offers adaptive detection mechanisms. AI-based NIDS utilizes machine learning (ML) algorithms to identify patterns in network traffic, enhancing the detection of known and unknown threats. Classical ML models, including Support Vector Machines[3], Random Forests, and deep learning techniques like deep neural networks [4], exhibit potential in intrusion detection, However, these complex architectures require high storage and computational cost.

979-8-3503-6777-5/24/$31.00 ©2024 IEEE

Another promising ML architecture can solve the problems of complexity and storage, Ensemble techniques[5, 6], which combine multiple estimators, have proven to be more effective than single classifiers in many scenarios. Common ensemble methods include bagging, boosting, and stacking. These techniques aggregate the predictions of multiple models to enhance overall accuracy, robustness, and generalization. Bagging reduces variance by training multiple models on different subsets of the data while boosting improves performance by focusing on misclassified instances. Stacking, on the other hand, combines the predictions of diverse base learners using a machine learning model, achieving superior results.

Despite the superior capabilities of ensemble classifiers, they still face challenges such as optimal model selection, and hyperparameter tuning. Without proper tuning and model selection, even the most advanced ensemble methods may fail to deliver optimal performance, necessitating an efficient approach to model optimization.

AutoML [6] has emerged as a transformative solution for addressing these challenges. AutoML automates the model selection and hyperparameter tuning processes, significantly reducing the time and effort required to develop an efficient NIDS. By leveraging AutoML, researchers can focus on more critical tasks such as data validation and model interpretability while ensuring that the most suitable model configurations are employed, thus enhancing the detection capabilities of the NIDS.

Popular AutoML frameworks like H2O [7] and AutoSklearn provide robust tools for constructing and fine-tuning ML estimators. These platforms enable rapid experimentation with various algorithms, model architectures, and hyperparameters, which significantly contribute to the development of reliable and precise NIDS. Additionally, they incorporate ensemble techniques, facilitating the creation of advanced models that can adapt to a wide range of attack patterns in real-time.