Ensemble Learning for Intrusion Detection in SDN-Based Zero Touch Smart Grid Systems

Zakaria Abou El Houda[1], Bouziane Brik[2], and Lyes Khoukhi[3]

[1]L@bISEN, ISEN Yncréa Ouest, Carquefou, France
[2]DRIVE EA1859, university of Bourgogne Franche-Comté, France
[3] GREYC CNRS, ENSICAEN, Normandie University, France
zakaria.abou.el.houda@umontreal.ca, bouziane.brik@u-bourgogne.fr, lyes.khoukhi@ensicaen.fr

***Abstract*—Software-defined network (SDN) is widely deployed on Smart Grid (SG) systems. It consists in decoupling control and data planes, to automate the monitoring and management of the communication network, and thus enabling zero touch management of SG systems. However, SDN-based SG is prone to several security threats and varios type of new attacks. To alleviate these issues, various Machine/Deep learning (ML/DL)-based intrusion detection systems (IDS) were designed to improve the detection accuracy of conventional IDS. However, they suffer from high variance and/or bias, which may lead to an inaccurate security threat detection. In this context, ensemble learning is an emerging ML technique that aims at combining several ML models; the objective is to generate less data-sensitive ($i.e.$, less variance) and more flexible ($i.e.$, less bias) machine learning models. In this paper, we design a novel framework, called BoostIDS, that leverages ensemble learning to efficiently detect and mitigate security threats in SDN-based SG system. BoostIDS comprises two main modules: (1) A data monitoring and feature selection module that makes use of an efficient Boosting Feature Selection Algorithm to select the best/relevant SG-based features; and (2) An ensemble learning-based threats detection moel that implements a Lightweight Boosting Algorithm (LBA) to timely and effectively detects SG-based attacks in a SDN environment. We conduct extensive experiments to validate BoostIDS on top of multiple real attacks; the obtained results using NSL-KDD and UNSW-NB15 datasets, confirm that BoostIDS can effectively detect/mitigate security threats in SDN-based SG systems, while optimizing training/test time complexity.**

***Index Terms*—Smart Grid; Software-defined network; Intrusion detection systems; Ensemble Learning.**

I. Introduction

Smart Grid (SG) is rapidly growing as the future of power systems [1]. SG introduces and deploys a large number of intelligent equipment, exchanging and processing both real-time critical information and huge amounts of data. SG is expected to deal with the dynamic availability of power as well as dynamic users' demands, that need uninterrupted availability of the network communication [2]. Hence, the design of novel platforms optimizing both network and power management for the SG system is more than required.

In this context, Software-defined network (SDN) emerged as promising solution for dynamically monitoring, managing, and configuring the communication networks of SG. SDN decouples the control plane from the data plane, enabling the network control to become directly programmable. Moreover, advanced Machine/Deep learning (ML/DL) algorithms are leveraging the huge amount of generated data, in SDN-based SG, to enable zero touch management of the SG systems [3]. For instance, ML/DL algorithms may analyze power data from SDN-based SG system to study electricity consumption behavior of users and improve then power equipment management accordingly.

SDN-based SG is prone to several security threats and may bring two main risks [4]. First, the deployed software in its control plane may contain vulnerabilities. Second, the SDN controllers are subject to distributed denial of services (DDoS) attacks and single-point failures, $e.g.$, compromised SDN switches can be manipulated to flood SDN controllers' resources. Although the security of the SDN has been widely investigated in different network contexts [5]–[29]. However, it is also essential to study its security on top of specific requirements of SG system. For instance, malicious redirection of control data flow through a high-latency path may be valid for some delay-tolerant networks, but may degrade the operational quality of a SG system. To deal with these issues, various Machine/Deep learning (ML/DL)-based intrusion detection systems (IDS) were designed, to optimize/enhance the accuracy of conventional IDS. However, ML/DL-based IDSs are mostly based on highly complex models with a large number of features. Thus, they can suffer from high variance and/or bias, which may lead to an inaccurate/inconsistent detection of new emerging security threats. In this context, ensemble learning is an emerging machine learning technique which aims at combining several learning; the objective is to generate less data-sensitive ($i.e.$, less variance) and more flexible ($i.e.$, less bias) machine learning models [30].

In this work, we develop a new framework, called BoostIDS, that leverages ensemble learning to efficiently detect and mitigate security threats in SDN-based SG system. BoostIDS comprises two main modules: (1) Data monitoring and feature selection module to gather data and select the most important/informative features, in order to optimize training/test time complexity; and (2) Ensemble learning-based threats detection module that implements a Lightweight Boosting Algorithm (LBA). We conduct extensive experiments to validate

BoostIDS on top of multiple real-world security threats; the obtained results, using NSL-KDD [31] and UNSW-NB15 [32], [33], confirm that BoostIDS can effectively detect/mitigate security threats in SDN-based SG systems, while optimizing training/test time complexity.

The remainder of this paper is organized as follows. Section II provides a comprehensive review of the related work. Section III presents the design and specification of the BoostIDS. Section IV presents the performance evaluation of BoostIDS. At last, section V concludes the paper.

II. Related Work

The new cyber-attacks ($i.e.$, zero-day attacks) continue to grow at a rapid pace and are becoming increasingly devastating. In the following, we provide an overview of the most prominent state-of-the-art solutions. Moudoud et al. [34] proposed a new Artificial Intelligence (AI)-based attack detection process that uses a hidden Markov model (HMM) to detect false data injection (FDI) attacks in the Internet of Things (IoT) system. The proposed process covers both detection and prediction of FDI attacks; it also includes a reputation and punishment-based trust management scheme to establish trust among IoT devices. The authors have shown that their proposed scheme outperforms the state-of-the-art contributions in terms of detection accuracy while reducing latency. Marir et al. [35] developed a distributed approach that uses multilayer support vector machines (SVMs) to detect anomalous behaviors in the network. Their proposed approach includes two steps. The first step uses a deep belief network to reduce the input features dimensionality; the objective is to select the best features. Once this selection is done, the second stage uses the best features to train a multi-layer SVM to perform efficient detection of abnormal behaviors in the network. The authors demonstrated the efficiency of their approach using four well-known datasets, namely, KDD'99, NSL-KDD, UNSW-NB15, and CICIDS2017. Tufan et al. [36] developed a network anomaly detection scheme based on two ML models, namely a convolutional neural network (CNN) and an ensemble learning model. The authors demonstrated the efficiency of their approach using an institutional dataset, namely UNSW-NB15. Nour et al. [37] developed a novel attack detection approach based on three common ML models, namely Decision Tree (DT), Naive Bayes (NB), and AI neural network to detect cyber attacks ($i.e.$, malicious/abnormal events) in IoT networks, including Domain Name System (DNS) and Message Queue Telemetry Transport (MQTT) based attacks. In particular, the authors developed a new complete learning model, called AdaBoost, based on the three ML models to evaluate the impact of such features and effectively detect malicious/abnormal events. The authors demonstrated the efficiency of their approach using two UNSW-NB15 and NIMS botnet datasets. Upadhyay et al. [38] developed a novel attack detection framework called Recursive Feature Elimination-eXtreme Gradient Boosting (RFE-XGBoost) suitable for SCADA (Supervisory Control and Data Acquisition) systems. RFE-XGBoost has two steps; first, it uses a Weighted Feature Importance (WFI) scheme to identify the features that are most useful. Afterwards, it gives these important features to a Majority Vote Ensemble Algorithm that consists of three bagging methods along with ANN, NB, and k-nearest neighbors (KNN) for end-to-end final intrusion detection. The authors evaluated the effectiveness of RFE-XGBoost using Receiver Operating Characteristic (ROC) curves, Precision, and Recall metrics.

Tama et al. [39] proposed a novel ML scheme that uses Rotation Forest along with Bagging techniques to detect malicious activities in the network. First, the authors use a hybrid feature selection scheme based on swarm optimization, genetic algorithm, and ant colony scheme to identify the features that are most useful; the objective is to reduce the error pruning tree (REPT). The authors demonstrated the feasibility of their approach using UNSW-NB15 and NSL-KDD datasets. Alkadi et al. [40] proposed an efficient deep blockchain framework (DBF) that uses a blockchain smart contract and a bidirectional long-term memory deep learning (BiLSTM) algorithm to ensure IoT network security. DBF uses the Ethereum smart contract to preserve privacy in a distributed IDS. The authors demonstrated the feasibility of their approach using two well-known datasets, namely UNSW-NB15 and BoT-IoT. Gao et al. [41] developed a novel framework based on EL to detect network attacks on a cloud-based robotic system. First, the authors construct an EL system using labeled data ($i.e.$, NSL-KDD). Then, they used a fuzzy-based scheme to efficiently utilize the unlabeled data. Thus, the proposed system includes both supervised and unsupervised techniques. The authors demonstrated the effectiveness of their approach using the well-known dataset, namely NSL-KDD. Seth et al. [42] developed an effective attack detection framework that combines multiple ML models to detect Network attacks in a multi-attack classification environment. To address the issue of imbalanced classes; the authors use a hybrid approach involving SMOTE and under-sampling techniques. The authors demonstrated the effectiveness of their approach using the well-known dataset, namely CIC-IDS 2018. Li et al. [43] designed a novel sustainable EL scheme based on an incremental learning process for multi-class regression models to detect abnormal/malicious behavior in the network. The authors demonstrated the feasibility/effectiveness of their approach using the well-known dataset, namely NSL-KDD. Al-Abassi et al. [44] have proposed a new framework suitable for industrial control systems (ICS) that uses a deep neural network (DNN) and a decision tree (DT) to detect cyber attacks. The authors demonstrated the effectiveness of their approach based on 10-fold cross-validation using two well-known ICS datasets. Gao et al. [45] designed an adaptive EL framework that leverages common ML models, including, Logical Regression (LR), and DNN to detect network attacks/anomalies. The authors demonstrated the feasibility/effectiveness of their approach using NSL-KDD dataset.

According to our analysis of these contributions, we have noticed that some of these solutions is based on a single learner, such as SVM and RF, to detect cyber attacks in the

network. First, these systems suffer from a problem of poor generalization, as they fail to generalize to unseen attacks such as "zero day" attacks. Moreover, it is challenging for only a single learner to handle the task to effectively detect all types of attacks, especially with the large amount of data generated, which my lead of over-fitting issues. To overcome the weaknesses of such existing solutions, we propose a novel framework that utilizes advanced boosting techniques to build more robust and less data-dependent ML/DL based IDSs ($i.e.$, less bias and less variance).

III. Lightweight adaptive Boosting Algorithm: Ensemble Learning-based Framework for Intrusion detection in SDN-based Smart Grid

In this section, we describe our BoostIDS framework along with its main modules. First, we briefly describe our proposed BoostIDS architecture. Then, we present our data monitoring and feature selection module. Finally, we highlight our Ensemble learning-based threats detection module.

A. System Architecture

Our BoostIDS framework is built on top of an architecture, comprising three main planes(see Fig. 1):

1) *End-users plane* in terms of power devices that are managed by the smart grid to provide their required energy in real-time. It is clear that the generated data at this level can be leveraged to enable several ML/DL-based applications at application plane.
2) *SDN plane* including control plane that is composed of SDN controllers, or the brain of the SDN plane. SDN controllers are in charge of determining the potential data-paths, based on the SDN application requirements. Besides, SDN plane contains also a data plane in charge of forwarding data flows based on already defined and configured rules at the SDN control plane. Both control and data planes can communicate with each other through an hypervisor, which is in charge of translating control flows to data-paths.
3) *Application plane* to provide several applications for managing effectively power production, distribution, and consumption. Various applications may be implemented at this level, including how the end-users behave in terms of power consumption. This helps also to optimize the energy production and distribution as well as the management of the provider power devices. Noting that these applications may also leverage ML/DL algorithms and exploit generated data at the SDN plan to build efficient Data-driven models.

On top of this architecture, we aim to build a new ensemble learning-based intrusion detection framework. Our framework can be deployed as an application at the application plane, in order to secure the whole SDN-based smart grid system against various threat, especially those related to the centralized SDN plane.

B. Data Collection and Feature Selection

When designing our framework ($i.e.$, BoostIDS), we have taken into consideration the following goals/objectives. First, BoostIDS must provide a complete protection of the SDN-based smart grid system against new emerging attacks. Second, these attacks need to be quickly and accurately recognized. Finally, the entire SDN-based smart grid system should be as robust and secure as possible. BoostIDS includes two modules. The first module enables efficient data collection from electrical devices managed by the smart grid; it makes use of an efficient Boosting Feature Selection Algorithm to select the best/relevant ($i.e.$, most informative) SG-based features. The second module includes a Lightweight Boosting Algorithm (LBA) to timely and effectively detects SG-based attacks in a SDN environment.

Data collection: In a SDN-based Smart Grid environment, Open-Flow protocol (OF) is used collect data from a large number of electrical devices ($i.e.$, End-users) managed by the smart grid; the objective is collect data ($e.g.$, number of flows per ingress port) for analyzing and detecting threats. OF-based technique depletes the memory of the end devices; also it exhausts the OF channel between the control plane and the data plane with attack traffic; which makes this technique unable to effectively detect high-throughput attacks. To address the challenges of OF-based technique, we use a flow sampling technique from a large number of electrical devices in a scalable and effective manner; our data collection method used sFlow and will not consume any bandwidth between data plane and SDN plane and does not exhaust the OF channel with attack traffic and it monitors high-speed traffic and has the ability to monitor networks at 100 Gbps and beyond. Once the SG-based data is collected/gathered, we encode the non-numeric/categorical data into numerical values using two encoding techniques, namely label and one hot encoding techniques. Once done, we re-scale the selected feature values using a normalization technique as follows:

$$\hat{F} = \frac{F_j - \mu}{\sigma} \tag{1}$$

where F_j denotes the collected feature, while σ and μ are the standard deviation and mean value of the collected feature, respectively.

Description of Datasets: In our study, we use NSW-NB15 and NSL-KDD datasets; these datasets inculdes the main real SG-based attacks, including fuzzers (24246 Attack data samples), DDoS (16353 Attack data samples), analysis (2677 Attack data samples), reconnaissance (13987 Attack data samples), backdoors (2329 Attack data samples), and others; NSL-KDD, an enhancement of the conventional KDD'99 dataset, contains the main real-world SG-based attacks $e.g.$,, Probe (Probing) and Distributed Denial of Service (DDoS).

Boosting Feature Selection: This module makes use of an efficient Boosting Feature Selection Algorithm to select the best/relevant ($i.e.$, most informative) SG-based features; it ranks the features according to an importance score that indicates the relevance of a feature in classifying attacks. The

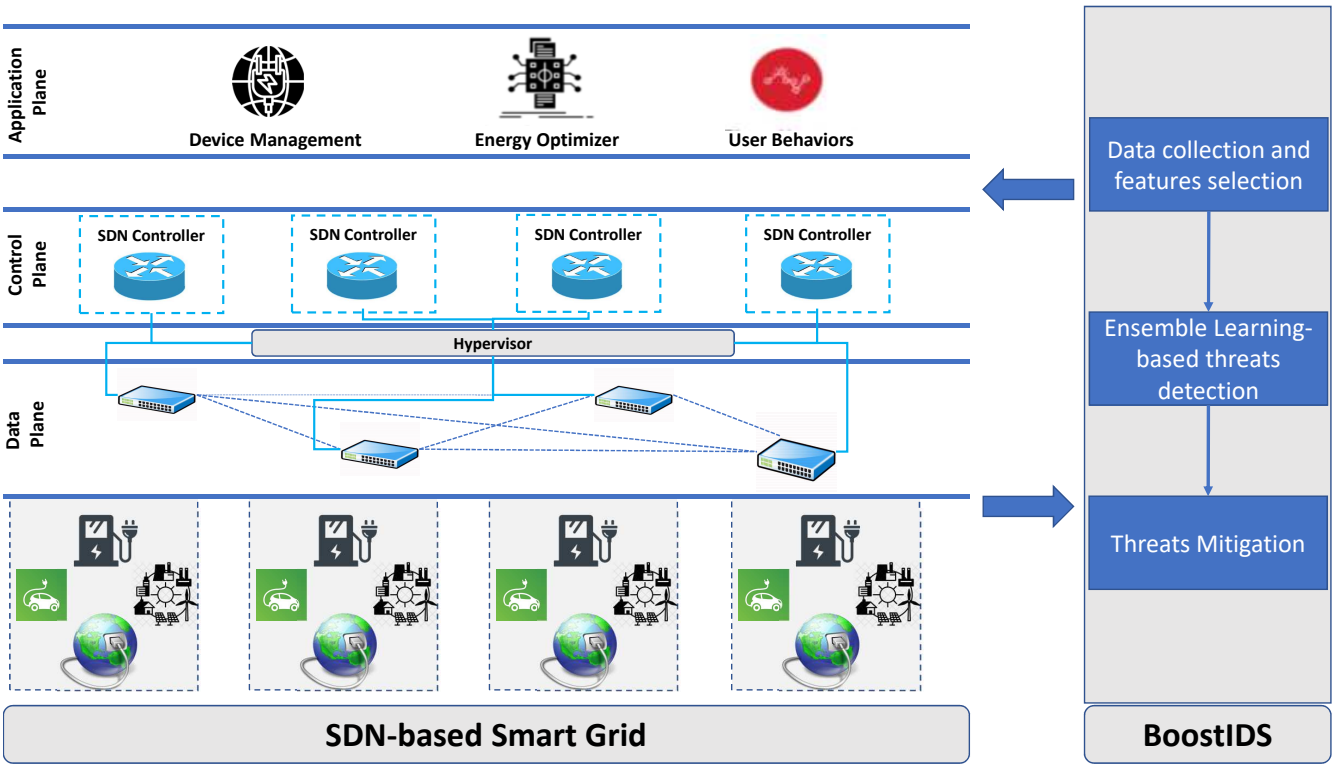


Fig. 1. Overview of our BoostIDS Architecture.

more valuable a feature is for building an optimal tree, the more valuable its score is.

For a given DT, a relevance measurement for the i^{th} feature f_i is calculated as follows:

$$BFS_i(DT) = \sum_{k=1}^{K-1} \hat{i}_t^2 I(f_k = i) \tag{2}$$

where i is the number of internal nodes and $\hat{i^2}$ is an estimation of the loss error function.

Then, the average significance of features is computed over all the DT for boosting as follows:

$$BFS_i = \frac{1}{N} \sum_{j=1}^{J} BFS_i(T_j) \tag{3}$$

where J is the number of trees.

C. Ensemble Learning-based Threats Detection

Our Ensemble Learning-based Threats Detection is a novel variant of boosting techniques that uses gradient descent process to find optimal values ($\hat{z}$) minimizing a loss function, defined as follows:

$$\mathcal{L} = -\frac{1}{K} \sum_{k=1}^{K} z_k * log(\hat{z_k}) \tag{4}$$

where z_k is the actual value of the k^{th} output class, $\hat{z_k}$ is the estimated value for k^{th} class, and K is the number of the samples of the data.

We start by setting the model with a constant value as follows:

$$M_0(s) = \arg\min_{\Theta} \sum_{k=1}^{K} \mathcal{L}(z_k, \Theta) \tag{5}$$

where K is the number of the samples of the data ($i.e., \{(s_k, z_k)\}_{k=1}^{K}$), while Θ is the estimated value.

Afterwards, we calculate pseudo-residuals for k^{th} data point at iteration i as follows:

$$r_{ki} = -[\frac{\partial \mathcal{L}(z_k, M(s_k))}{\partial (M(s_k))}]_{M(s)=M_{i-1}(s)} \tag{6}$$

Then, we fit the weak learner $w_i(s)$ to pseudo-residuals and we compute the output value at iteration i as follows:

$$\Theta_i = \arg\min_{\Theta} \sum_{k=1}^{K} \mathcal{L}(z_k, M_{i-1}(s_k) + \Theta w_i(s_k)) \tag{7}$$

To find the optimal/best value of Θ that optimize the loss function, we use a second-order Taylor polynomial approxi-

mation as follows:

$$\mathcal{L}^i \simeq \sum_{k=1}^{K}[\mathcal{L}(z_k, M_{i-1}(s_k)) + g_k w_i(s_k) + \frac{1}{2}h_k w_i(s_k)^2] + \frac{1}{2}\Lambda \sum_{n=1}^{N} \Theta^2 j$$

where $g_k = \partial\mathcal{L}(z_k, M(s_k))$ and $h_k = \partial^2\mathcal{L}(z_k, M(s_k))$ are first and second order gradient of the loss function, where Λ is a regularization parameter, and N is the total number of leaves in the tree.

Then, we calculate the value of Θ at iteration i as follows:

$$\Theta_i = -\frac{\sum_{k=1}^{K} g_k w_i(s_k)}{\sum_{k=1}^{K} h_k w_i(s_k) + \Lambda} \tag{8}$$

Lastly, we then update the resulting model at the i iteration as follows:

$$M_i(s) = M_{i-1}(s) + \Theta i w_i(s_k) \tag{9}$$

The following algorithm summarizes the main process of our ensemble learning-based threat detection ($i.e.$, Lightweight Boosting Algorithm (LBA)).

Algorithm 1 *Lightweight Boosting Algorithm (LBA)*

Input: Sequence of K data points $\{(s_k, z_k)\}$, $k = 1, ..., K$
A loss function $\mathcal{L}(z, M(s))$
Init the model with $M_0(s) = \arg\min_\Theta \sum_{k=1}^{K} \mathcal{L}(z_k, \Theta)$
for $i \leftarrow$*1 to T* **do**
  Calculate $r_{ki} = -[\frac{\partial\mathcal{L}(z_k, M(s_k))}{\partial(M(s_k))}]_{M(s)=M_{i-1}(s)}$
  Fit a weak learner $w_i(s)$ *to pseudo-residuals* $\{(s_k, r_{ki})\}_{k=1}^{K}$
  Calculate optimal Θ_i *by solving:*
  $\Theta_i = \arg\min_\Theta \sum_{k=1}^{K} \mathcal{L}(z_k, M_{i-1}(s_k) + \Theta w_i(s_k))$
  Update the model as follows:
  $M_i(s) = M_{i-1}(s) + \Theta_i w_i(s_k)$
end
Output the final Model $M_T(s)$ *for final attack detection*

IV. Performance Evaluation

A. Parameter Settings

To test the effectiveness of our proposed framework, BoostIDS, we emulate a real realistic/real-network environment by using a popular SDN emulator tool, called Mininet [46]. Mininet uses virtual OF switches and containers to create a realistic/real-network virtual environment. We use Floodlight controller [47] to implement our Lightweight Boosting Algorithm (LBA). Finally, we use sFlow-RT [48]) to collect/gather the features of SG-based network in a efficient/scalable manner.

Figs. 2(a) and 2(b) show the most relevant/important features on UNSW-NB15 and NSL-KDD datasets, respectively; it shows the features with the highest scores in a descending order. We notice that higher than 80% and 85% of the collected input features of the UNSW-NB15 and NSL-KDD datasets, respectively, do not contribute to accurate decisions ($i.e.$, SG-based attack classification). Thus, our proposed Boosting Feature Selection scheme can largely eliminate unnecessary features that may delay the learning process, making it an effective scheme to keep only the relevant features that can assist the model in making accurate decisions. Figs. 3(a) and 3(b) show the Log Loss of BoostIDS for UNSW-NB15 and NSLKDD datasets, respectively. We notice that both the training as well testing losses values decrease, in UNSW-NB15 and NSLKDD datasets, until they reach almost zero for the UNSW-NB15 dataset and a point of stability for the NSLKDD dataset; this indicates that the model is learning efficiently.

TABLE I
Performance metrics of BoostIDS and state-of-the-art ML/DL models using UNSW-NB15 dataset

Methods	Accuracy	Precision	Recall	F1	Time (second)
ENS-SVM [35]	0.98	0.9	0.97	0.93	N/A
CNN [36]	0.99	0.9	0.97	0.95	N/A
ENS-STA [37]	0.98	0.9	0.97	N/A	145
TSE-IDS [39]	0.85	0.87	0.88	N/A	N/A
OGM [49]	0.95	0.94	N/A	N/A	N/A
BoostIDS	**0.99**	**0.99**	**0.99**	**0.99**	**60**

TABLE II
Performance metrics of BoostIDS and state-of-the-art ML/DL models using NSLKDD dataset

Methods	Accuracy	Precision	Recall	F1	Time (second)
CharCNN-IDS [50]	0.85	0.91	0.81	0.86	N/A
ResNet50 [51]	0.79	0.91	0.69	0.79	N/A
GoogleNet [51]	0.77	0.91	0.65	0.76	N/A
Adaboost [52]	0.85	0.86	0.85	0.84	N/A
BoostIDS	**0.86**	**0.96**	**0.87**	**0.87**	**14**

B. Results and Discussions

To test the effectiveness of BoostIDS, we use several metrics, including, detection rate (DR), Accuracy, Precision, and F1-score. Also, we study the performance of BoostIDS using ROC curves and confusion matrices. Figs. 5 and 6 show the confusion matrices and the ROC curves of BoostIDS on UNSW-NB15 and NSLKDD datasets, respectively. Tables 1 and 2 and shows the detailed performance of BoostIDS and state-of-the-art contributions for UNSW-NB15 and NSLKDD datasets, respectively. BoostIDS performs best in both datasets and outperforms the Machine/Deep learning (ML/DL)-based intrusion detection systems/models in both datasets. In UNSW-NB15 dataset, BoostIDS achieves 99% in Accuracy, detection rate, F1-score, and Precision with only 60s of training; while

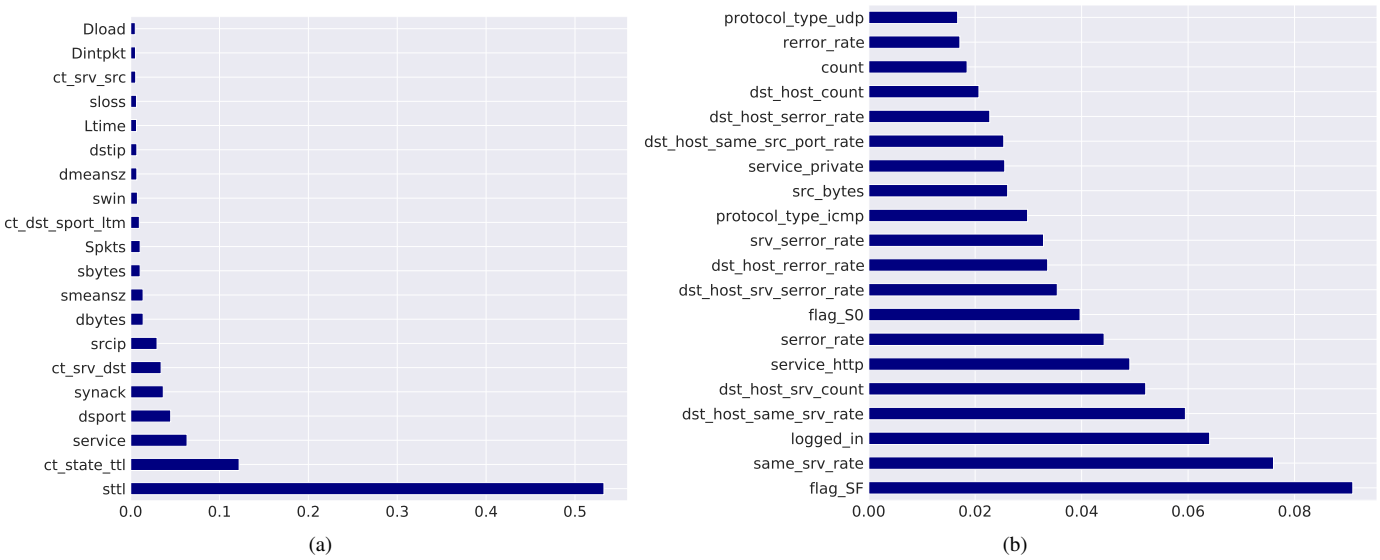


Fig. 2. Feature importance scores on: (a) UNSW-NB15 and (b) NSL-KDD.

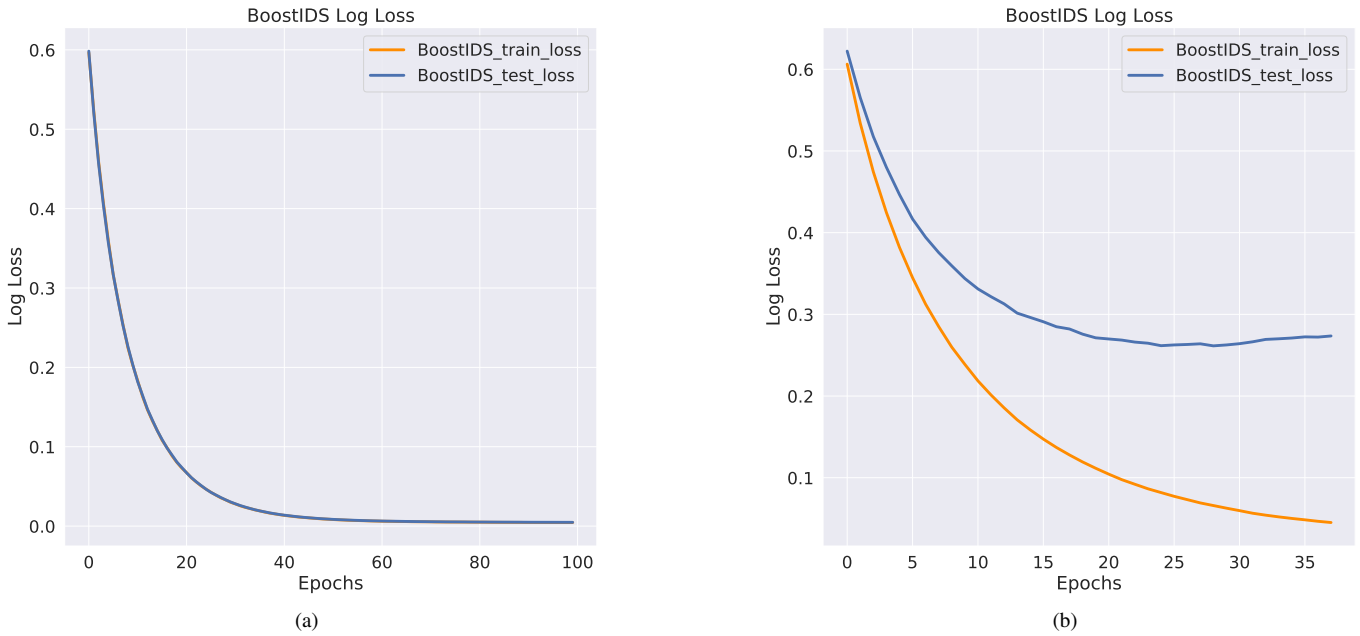
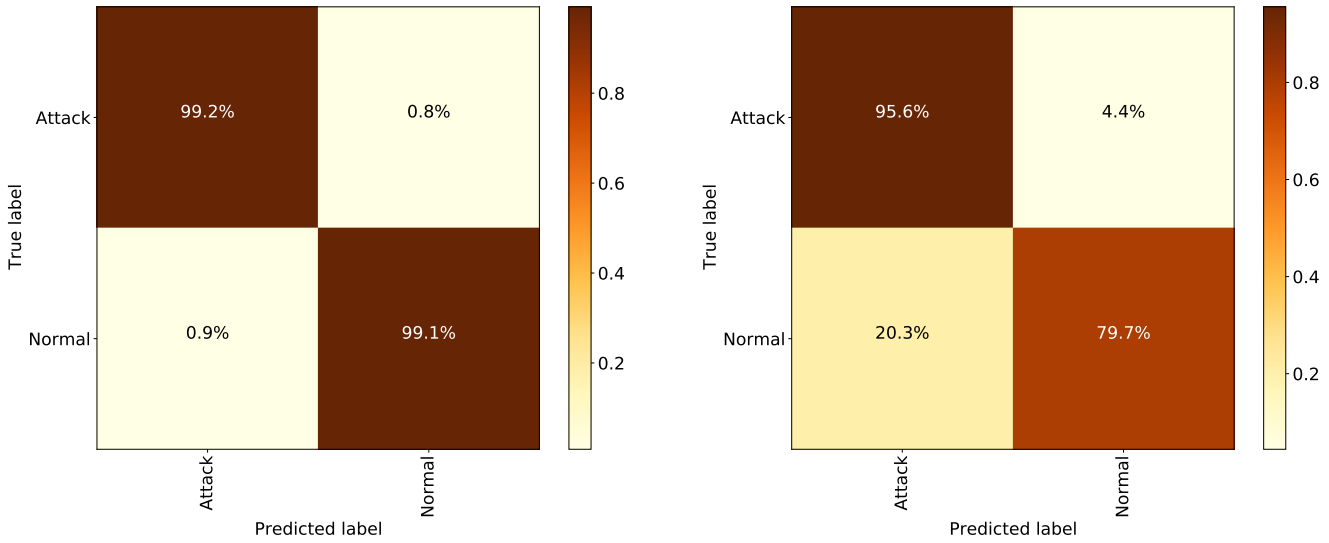


Fig. 3. Model loss of BoostIDS for: (a) UNSW-NB15 and (b) NSL-KDD.

BoostIDS achieves 86% in Accuracy, 87% in detection rate, 96% in Precision, and 87% in F1-score with only 14s of training for NSLKDD dataset. The experimental results confirm that BoostIDS has better Precision, Accuracy, F1-score, and detection rate than existing Machine/Deep learning (ML/DL)-based intrusion detection systems.

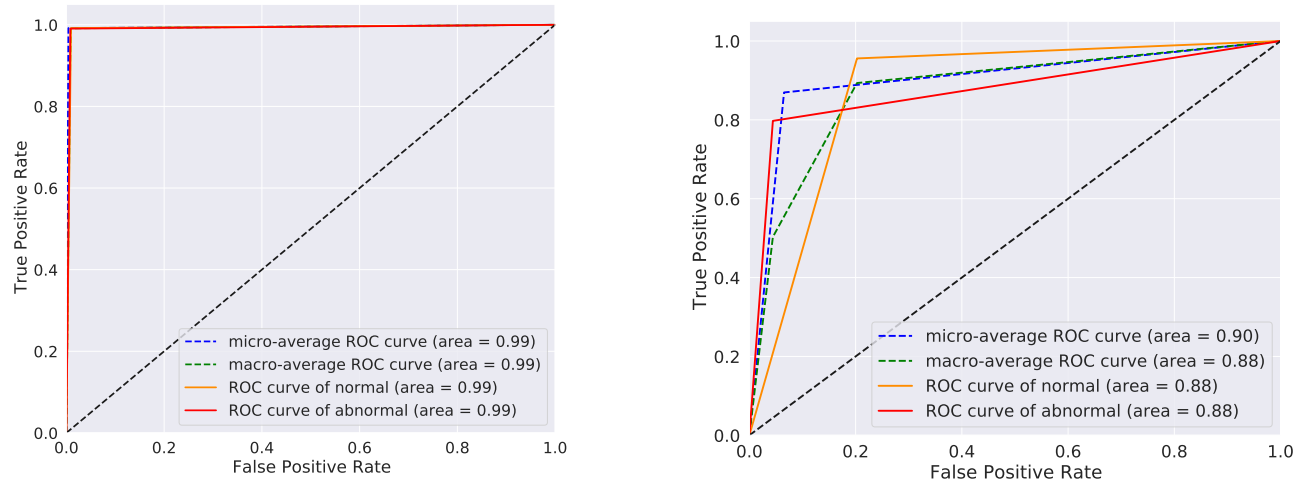
V. Conclusion

In this paper, we designed a novel framework, called BoostIDS, that leverages ensemble learning to efficiently detect and mitigate security threats in SDN-based SG system. BoostIDS comprises two main modules: (1) Data monitoring and feature selection that makes use of an efficient Boosting Feature Selection Algorithm to select the best/relevant ($i.e.$, most informative) SG-based features; and (2) An ensemble learning-based threats detection that implements a Lightweight Boosting Algorithm (LBA) to timely and effectively detects SG-based attacks in a SDN environment; the obtained results, using NSL-KDD and UNSW-NB15 datasets, confirmed that BoostIDS can effectively detect/mitigate security threats in SDN-based SG systems, while optimizing training/test time complexity. This makes BoostIDS a very prominent cybersecurity framework to mitigate more advanced and sophisticated



(a) (b)

Fig. 4. Confusion matrices of BoostIDS for: (a) UNSW-NB15 and (b) NSL-KDD.



(a) (b)

Fig. 5. ROC curves of BoostIDS for: (a) UNSW-NB15 and (b) NSL-KDD.

attacks.

REFERENCES

[1] M. A. Benblidia, B. Brik, M. Esseghir, and L. Merghem-Boulahia, "Power dispatching in cloud data centers using smart microgrids: A game theory approach," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[2] M. A. benblidia, B. Brik, M. Esseghir, and L. Merghem-Boulahia, "A game based power allocation in cloud computing data centers," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2018, pp. 1–7.

[3] G. Huang, F. Wu, and C. Guo, "Smart grid dispatch powered by deep learning: a survey," *Frontiers of Information Technology Electronic Engineering*, pp. 2095–9230, 2022.

[4] H. Maziku, S. Shetty, and D. M. Nicol, "Security risk assessment for sdn-enabled smart grids," *Computer Communications*, vol. 133, pp. 1–11, 2019.

[5] H. Moudoud, S. Cherkaoui, and L. Khoukhi, "An iot blockchain architecture using oracles and smart contracts: the use-case of a food supply chain," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–6.

[6] Z. Abou El Houda, "Security enforcement through software defined networks (sdn)," Ph.D. dissertation, Troyes, 2021.

[7] H. Moudoud, L. Khoukhi, and S. Cherkaoui, "Prediction and detection of fdia and ddos attacks in 5g enabled iot," *IEEE Network*, vol. 35, no. 2, pp. 194–201, 2021.

[8] Z. A. El Houda, A. Hafid, and L. Khoukhi, "Co-iot: A collaborative ddos mitigation scheme in iot environment based on blockchain using sdn," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[9] H. Moudoud, S. Cherkaoui, and L. Khoukhi, "Towards a scalable and trustworthy blockchain: Iot use case," in *ICC 2021 - 2021 IEEE International Conference on Communications (ICC)*, 2021, pp. 1–6.

[10] Z. A. El Houda, L. Khoukhi, and A. Hafid, "Chainsecure - a scalable and proactive solution for protecting blockchain applications using sdn," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[11] H. Moudoud, S. Cherkaoui, and L. Khoukhi, *An Overview of Blockchain and 5G Networks*. Cham: Springer International Publishing, 2022, pp. 1–20.

[12] Z. A. E. Houda, A. Hafid, and L. Khoukhi, "Blockchain meets ami: Towards secure advanced metering infrastructures," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[13] Z. A. E. Houda, B. Brik, and L. Khoukhi, ""why should i trust your ids?": An explainable deep learning framework for intrusion detection systems in internet of things networks," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2022.

[14] Z. A. E. Houda, A. Hafid, and L. Khoukhi, "Blockchain-based reverse auction for v2v charging in smart grid environment," in *ICC 2021 - 2021 IEEE International Conference on Communications (ICC)*, 2021, pp. 1–6.

[15] H. Moudoud, S. Cherkaoui, and L. Khoukhi, "Towards a secure and reliable federated learning using blockchain," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 01–06.

[16] Z. A. E. Houda, A. Hafid, and L. Khoukhi, "Brainchain - a machine learning approach for protecting blockchain applications using sdn," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[17] B. Brik and A. Ksentini, "Toward optimal mec resource dimensioning for a vehicle collision avoidance system: A deep learning approach," *IEEE Network*, vol. 35, no. 3, pp. 74–80, 2021.

[18] Z. Abou El Houda, B. Brik, A. Ksentini, L. Khoukhi, and M. Guizani, "When federated learning meets game theory: A cooperative framework to secure iiot applications on edge computing," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2022.

[19] B. Brik and A. Ksentini, "On predicting service-oriented network slices performances in 5g: A federated learning approach," in *2020 IEEE 45th Conference on Local Computer Networks (LCN)*, 2020, pp. 164–171.

[20] Z. Abou El Houda, A. Senhaji Hafid, and L. Khoukhi, *A Novel Unsupervised Learning Method for Intrusion Detection in Software-Defined Networks*. Cham: Springer International Publishing, 2022, pp. 103–117.

[21] Z. Abou El Houda, "Renforcement de la sécurité à travers les réseaux programmables," Ph.D. dissertation, Université de Montréal, 2021.

[22] Z. A. El Houda, A. S. Hafid, and L. Khoukhi, "A novel machine learning framework for advanced attack detection using sdn," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[23] R. Khatoun, P. Gut, R. Doulami, L. Khoukhi, and A. Serhrouchni, "A reputation system for detection of black hole attack in vehicular networking," in *2015 International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC)*, 2015, pp. 1–5.

[24] Z. Abou El Houda, A. S. Hafid, and L. Khoukhi, "Cochain-sc: An intra- and inter-domain ddos mitigation scheme based on blockchain using sdn and smart contract," *IEEE Access*, vol. 7, pp. 98 893–98 907, 2019.

[25] A. Nabet, R. Khatoun, L. Khoukhi, J. Dromard, and D. Gaïti, "Towards secure route discovery protocol in manet," in *Global Information Infrastructure Symposium - GIIS 2011*, 2011, pp. 1–8.

[26] Z. Abou El Houda, L. Khoukhi, and A. Senhaji Hafid, "Bringing intelligence to software defined networks: Mitigating ddos attacks," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2523–2535, 2020.

[27] B. Brik, A. Ksentini, and M. Bouaziz, "Federated learning for uavs-enabled wireless networks: Use cases, challenges, and open problems," *IEEE Access*, vol. 8, pp. 53 841–53 849, 2020.

[28] Z. Abou El Houda, B. Brik, and S. Sidi-Mohammed, "A novel iot-based explainable deep learning framework for intrusion detection systems," *IEEE Internet of Things Magazine*, 2022.

[29] Z. Abou El Houda, A. S. Hafid, and L. Khoukhi, "Cofed: A privacy preserving collaborative ddos mitigation framework based on federated learning using sdn and blockchain," *IEEE Transactions on Network Science and Engineering*, 2021.

[30] T. N. Rincy and R. Gupta, "Ensemble learning techniques and its efficiency in machine learning: A survey," in *2nd International Conference on Data, Engineering and Applications (IDEA)*, 2020, pp. 1–6.

[31] "Nsl-kdd dataset." [Online]. Available: https://www.unb.ca/cic/datasets/nsl.html

[32] N. Moustafa and J. Slay, "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 Military Communications and Information Systems Conference (MilCIS)*, 2015, pp. 1–6.

[33] N. Moustafa, "Unsw-nb15." [Online]. Available: www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets

[34] H. Moudoud, Z. Mlika, L. Khoukhi, and S. Cherkaoui, "Detection and prediction of fdi attacks in iot systems via hidden markov model," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2022.

[35] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," *IEEE Access*, vol. 6, pp. 59 657–59 671, 2018.

[36] E. Tufan, C. Tezcan, and C. Acartürk, "Anomaly-based intrusion detection by machine learning: A case study on probing attacks to an institutional network," *IEEE Access*, vol. 9, pp. 50 078–50 092, 2021.

[37] N. Moustafa, B. Turnbull, and K.-K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815–4830, 2019.

[38] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Intrusion detection in scada based power grids: Recursive feature elimination model with majority vote ensemble algorithm," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2559–2574, 2021.

[39] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "Tse-ids: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," *IEEE Access*, vol. 7, pp. 94 497–94 507, 2019.

[40] O. Alkadi, N. Moustafa, B. Turnbull, and K.-K. R. Choo, "A deep blockchain framework-enabled collaborative intrusion detection for protecting iot and cloud networks," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9463–9472, 2021.

[41] Y. Gao, Y. Liu, Y. Jin, J. Chen, and H. Wu, "A novel semi-supervised learning approach for network intrusion detection on cloud-based robotic system," *IEEE Access*, vol. 6, pp. 50 927–50 938, 2018.

[42] S. Seth, K. K. Chahal, and G. Singh, "A novel ensemble framework for an intelligent intrusion detection system," *IEEE Access*, vol. 9, pp. 138 451–138 467, 2021.

[43] X. Li, M. Zhu, L. T. Yang, M. Xu, Z. Ma, C. Zhong, H. Li, and Y. Xiang, "Sustainable ensemble learning driving intrusion detection model," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 4, pp. 1591–1604, 2021.

[44] A. Al-Abassi, H. Karimipour, A. Dehghantanha, and R. M. Parizi, "An ensemble deep learning-based cyber-attack detection in industrial control system," *IEEE Access*, vol. 8, pp. 83 965–83 973, 2020.

[45] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82 512–82 521, 2019.

[46] "Mininet." [Online]. Available: http://mininet.org

[47] "Floodlight openflow controller." [Online]. Available: https://floodlight.atlassian.net/wiki/spaces/HOME/overview

[48] "sflow-rt." [Online]. Available: http://www.sflow-rt.com

[49] N. Moustafa, G. Misra, and J. Slay, "Generalized outlier gaussian mixture technique based on automated association features for simulating and detecting web application attacks," *IEEE Transactions on Sustainable Computing*, pp. 1–1, 2018.

[50] S. Z. Lin, Y. Shi, and Z. Xue, "Character-level intrusion detection based on convolutional neural networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8.

[51] Z. Li, Z. Qin, K. Huang, X. Yang, and S. Ye, "Intrusion detection using convolutional neural networks for representation learning," in *Neural Information Processing*. Springer International Publishing, 2017, pp. 858–866.

[52] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82 512–82 521, 2019.