A Novel Machine Learning Framework for Advanced Attack Detection using SDN

Zakaria Abou El Houda[1,3], Abdelhakim Senhaji Hafid[1], and Lyes Khoukhi[2]

[1]NRL, Department of Computer Science and Operational Research, University of Montreal, Canada
[2]GREYC CNRS, ENSICAEN, Normandie University, France
[3]University of Technology of Troyes, France
zakaria.abou.el.houda@umontreal.ca, ahafid@iro.umontreal.ca
lyes.khoukhi@ensicaen.fr

***Abstract*—Recently, software defined networks (SDN) has emerged as novel technology that leverages network programmability to facilitate network management. SDN provides a global view of the network, through a logically centralized component, called SDN controller, to strengthen network security. SDN separates the control plane from the data plane, which allows for a more control over the network and brings new capabilities to cope with the new emerging security threats ($i.e.$, zero-day attacks). Existing attack detection schemes are facing obstacles due to high false positive rates, low detection performances, and high computational costs. To address these issues, we propose a multi-module Machine Learning (ML) framework that combines unsupervised ML techniques with a scalable feature collection and selection scheme to effectively/timely detect network security threats in the context of SDN. In particular, our proposed framework consists of: (1) a data flow collection module (DFC) to gather the features of network data in a scalable and efficient way using sFlow protocol; (2) an Information gain Feature Selection (IGF) module to select the most informative/relevant features to reduce training and testing time complexity; and (3) a novel unsupervised ML module that uses a novel outlier detection scheme, called Isolation Forest (ML-IF), to effectively/timely detect network security threats in SDN. The experimental results using the well-known public network security dataset UNSW-NB15, show that our proposed framework outperforms state-of-the-art contributions in terms of accuracy and detection rate while significantly reducing computational complexity; making it a promising framework to mitigate the new emerging network security threats in SDN.**

***Index Terms*—Machine Learning; Intrusion Detection System; Isolation Forest; SDN.**

I. Introduction

In recent years, there has been a huge increase in the number of cyber attacks on critical domains/infrastructures ($e.g.$, healthcare and Smart grids (SGs)); these attacks continue to grow at a rapid rate and cause huge damage, and financial losses for educational and business organizations, and may lead to large-scale blackouts [1]–[7]. The new emerging security threats have been increasing in strength/sophistication and are becoming more devastating/destructive; these attacks are predicted to cost huge financial losses of about $20 Billion (USD) by 2021 for both education organizations as well as enterprises [8]. To protect the network and end users from the new emerging security threats, Internet Service Providers (ISPs) deploy different security mechanisms such as access control schemes and stateful/stateless firewalls. However, it has been shown that these cyber-security mechanisms are not sufficient to protect the network as well as end users from zero-day attacks. To this end, Intrusion Detection Systems (IDSs) need to be carefully developed/designed to cope with the new emerging network security threats ranging from data leakage to phishing attacks. Traditional IDSs are limited by their need of new/up-to-date attack patterns, which make them vulnerable to zero-day attacks. The recent emergence of ML/DL techniques have revolutionized many fields and in particular the security field; since then, several IDSs have adopted ML/DL to detect network intrusions [9]–[17]. However, most of the existing ML/DL-based IDSs, such as density-based techniques ($e.g.$, DBSCAN [18]), suffer from high false positive rates since they consider any deviation from the normal behavior as an abnormal activity. Also, they are computationally expensive since they need to memorize a large number of normal data sample features.

To alleviate these issues, we propose a novel multi-module Machine Learning (ML) framework that combines an optimized feature selection scheme with advanced unsupervised ML technique to reduce computational complexity while effectively improving detection performance $i.e.$, accuracy and detection rate. In particular, our proposed framework consists of: (1) a novel data flow collection module (DFC) to gather the features of network data in a scalable and efficient way using sFlow protocol; (2) an Information gain Feature Selection (IGF) module to select the most informative/relevant features to reduce training and testing time complexity; and (3) a novel unsupervised ML module that uses a novel outlier detection scheme, called Isolation Forest (ML-IF), to effectively/timely detect network security threats in SDN. Our proposed framework isolates abnormal data samples instead of memorizing the profile of normal data samples; this makes it more lightweight than existing unsupervised ML/DL-based IDSs. Abnormal data samples are few data samples; this makes them susceptible to isolation with low computation. Our proposed framework not only enhances the detection performance $i.e.$, accuracy and detection rate, but also reduces the false positive

rate with minor computational complexity. To evaluate the effectiveness of our proposed framework, we conduct extensive experiments using variety of real-world network security threats. The experimental results, using the well-known public network security dataset, namely the UNSW-NB15 [19], [20], show that our proposed framework outperforms state-of-the-art contributions in terms of accuracy and detection rate while significantly reducing computational complexity; making it a promising framework to mitigate the new emerging network security threats in SDN.

The main contributions of this paper are summarized as follows:

- We propose a novel data flow collection module (DFC) to gather the features of network data in a scalable and efficient way using sFlow protocol.
- We design an Information gain Feature Selection (IGF) module to select the most informative/relevant features to reduce training and testing time complexity.
- We propose a novel unsupervised ML module that uses a novel outlier detection scheme, called Isolation Forest (ML-IF), to effectively/timely detect network security threats in SDN.
- We evaluate the performance of the proposed framework in terms of accuracy, detection rate, and computational complexity. We compare the performance of our proposed framework with state-of-the-art ML/DL-based IDSs. The experimental results show that our proposed framework achieves high detection performances while reducing computational complexity, making it a promising framework to cope with the new emerging security threats in SDN.

The remainder of this paper is organized as follows. In Section II, we present a review of related work. Section III presents the system design of our proposed framework. In Section IV, we evaluate the proposed framework. Finally, Section V concludes the paper.

II. Related Work

The new emerging attacks are becoming more devastating and destructive; several state-of-the-art contributions have integrated supervised and unsupervised ML/DL techniques to enhance detection performances of traditional IDSs. In the following, we overview the most representative ML/DL based IDSs and we discuss their security issues.

Ashfaq et al. [21] proposed a novel Semi Supervised Learning (SSL) technique that uses a single hidden layer feed-forward neural network (SLFN) along with a sample categorization scheme to detect network anomalies. The proposed scheme uses a fuzzy quantity scheme to categorize/classify data samples. The effectiveness of the proposed scheme was tested and evaluated using the NSL-KDD dataset. Singh et al. [22] proposed a scalable peer-to-peer (p2p) anomaly detection scheme that uses ML technique ($i.e.$, Random Forest (RF)) to detect network anomalies. The proposed scheme consists of: (1) a distributed framework that dynamic extracts network data features; and (2) a classification module that uses RF model to detect, in real-time, network attacks. The effectiveness of the proposed scheme was tested and evaluated using the CAIDA dataset. McDermott et al. [23] proposed FeedWSN, a novel IDS that uses a back propagation neural network and a support vector machine (SVM) to detect intrusions in Wireless Sensor Networks (WSNs). The authors have compared the detection performance of the two models for six cyber-attacks using the NSL-KDD dataset.

Tuan et al. [24] proposed a novel unsupervised learning method that uses Local outlier Factor (LoF) scheme to detect network attacks ($e.g.$, DDoS attacks) in SDN. The proposed method uses a local density scheme to measure the local deviation for a given data sample with respect to its neighbors. The proposed method was tested and evaluated using CAIDA dataset. Ali et al. [25] proposed a three-tier intrusion detection and prevention system (IDPS) to detect Distributed Denial of Service (DDoS) attacks in SDN. The proposed IDPS includes packet validation, user validation, and flow validation. The effectiveness of the proposed system was evaluated using the OMNeT++ emulator. Moustafa et al. [26] proposed a novel architecture that uses Outlier Gaussian Mixture (OGM) scheme to detect web attacks. The proposed architecture consists of: (1) a data pre-processing module that uses an Association Rule Mining (ARM) scheme to dynamically extract informative features; and (2) a classification module that uses OGM scheme to detect network anomalies. The authors evaluated the effectiveness of OGM using two well-known public network security datasets, the Web Attack and the UNSW-NB15. Nour et al. [27] proposed a new threat intelligence scheme that uses beta mixture-hidden Markov models (MHMMs) to detect network anomalies in the context of Industry 4.0. The proposed scheme consists of: (1) a novel smart management module to handle/manage heterogeneous network data flow includes data from actuators and IoT sensors; and (2) a novel threat intelligence module that uses MHMMs to detect network anomalies. The authors evaluated the effectiveness of MHMMs using two well-known public network security datasets, the CPS dataset and the UNSW-NB15.

Based on our analysis of these existing contributions [21]-[27], we found that a number of these solutions [24]–[26] are computationally expensive. Also, most of them suffer from high false positive rates since they consider any deviation from the normal state as an abnormal activity. To address the shortcomings of the existing solutions [21]- [27], we propose a multi-module ML framework that uses a novel outlier detection scheme $i.e.$, ML-IF to effectively/timely detect attacks in SDN. Our proposed framework isolates abnormal data samples instead of memorizing the profile of normal data samples; it does not use any computationally expensive method ($e.g.$, density measure) to detect abnormal data samples. Moreover, we introduce a novel data flow collection module and optimized feature selection module to gather the features of network data in a scalable and efficient way. Our proposed framework is much scalable and accurate in comparison with the ones that uses OF schemes [24], [25] and outperforms state-of-the-art contributions in terms of accuracy and detection rate while

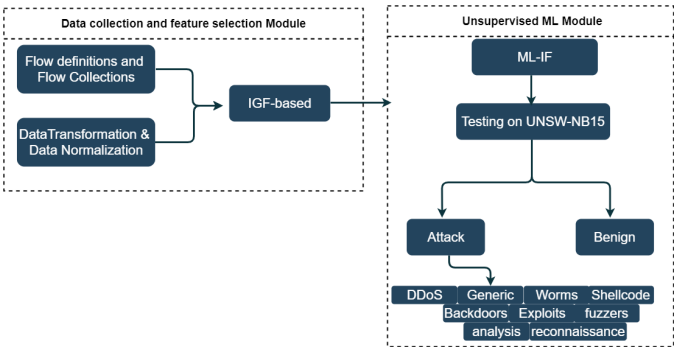


Fig. 1. Proposed Multi-Module ML-based IDS Framework

significantly reducing computational complexity.

III. System design

A. Overview

When designing our proposed framework, we did consider the following goals/objectives. First, our proposed framework should ensure/guarantee a full protection from the new emerging security threats. Unlike existing ML/DL based IDSs [21]-[27] that try to find a profile of similar data samples, then classify others as abnormal data samples ($i.e.$, intrusions). Since abnormal data samples are mostly few and rare, we aim to isolate them with low computation instead of profiling a large number of normal data samples. Second, these anomalies/intrusions should be timely/effectively detected, and the overall system has to be as secure as possible. Fig. 1 shows our proposed multi-module framework; it includes three modules: (1) a novel data flow collection module (DFC) to gather the features of network data in a scalable and efficient way using sFlow protocol; (2) an Information gain Feature Selection (IGF) module to select the most informative/relevant features to reduce training and testing time complexity; and (3) a novel unsupervised ML module that uses ML-IF to effectively/timely detect network security threats in SDN.

B. Data Flow Collection Module (DFC)

Network Data collection is the first step of security attack detection. In SDN, two commonly techniques are used to collect the features of network data flow. The first technique uses OF protocol while the second one uses flow sampling techniques such as Sflow protocol. In OF based technique, SDN controllers periodically sends state requests to forwarding elements. Each forwarding element responds with one or multiple flow statistics reply messages that contain network data flow statistics. To this end, each forwarding element needs to maintain/store a large number of network data flow entries. However, this will exhaust the resources of OF devices ($i.e.$, Ternary Content Addressable Memory (TCAM)). Also, a large size of flow statistics reply messages sent by forwarding elements to SDN controllers may congest the bandwidth between OF devices and SDN controllers and exhaust the OF channel by attack traffic. To alleviate the issues of OF-based technique, DFC makes use of flow sampling techniques to gather the features of network data flow in a scalable and efficient way using sFlow protocol. DFC is more scalable and does neither consume the bandwidth between OF devices and SDN controllers nor exhaust the OF channel by attack traffic. NDC defines the flow monitoring metrics such as flow aggregation attributes and thresholds, and uses an sFlow collector ($i.e.$, sFlow-RT [28]) to deploy these metrics in network data plane elements. Once the features of network data flow is collected using DFC, we encoded categorical features into a numeric values. Then, we re-scale their values using a standardization technique as follows:

$$F_{stand} = \frac{F_i - \mu}{\sigma} \tag{1}$$

where F_i denotes the input feature, μ and σ denote, respectively, the mean and standard deviation values for each input feature value F_i.

C. Information Gain Feature Selection Module (IGF)

Irrelevant features can slow down the process of training and prevent a ML model from making accurate decisions, especially when dealing with large scale data. In our framework, we investigate the use of information gain feature selection module (IGF) to select the most informative features. IGF ranks features based on mutual information and entropy calculation. For an input feature X and a class Y, IGF ranks a feature X based on the amount of knowledge that can be gained about Y. In fact, if X and Y are independent, their IGF score is almost zero. Thus, X does not contain any knowledge/information about Y and it is very likely an irrelevant feature that may prevent a ML model from making accurate decisions. IGF is defined as follows:

$$IGF(X;Y) = H(X) - H(X/Y) \tag{2}$$

where $H(X)$ is the entropy of an input feature X, $H(X|Y)$ is the conditional entropy of a feature X given a class Y. The formula for IGF can also be expressed as follows:

$$TIGFS(X;Y) = \int_X \int_Y p_{X,Y}(X,Y) log \frac{p_{X,Y}(X,Y)}{p_X(X) p_Y(Y)} dxdy \tag{3}$$

where $p_{X,Y}(X,Y)$ is the joint probability density function of a feature X and a class Y while $p_X(X)$ and $p_Y(Y)$ are marginal density functions of X and Y, respectively.

D. Unsupervised ML Module (ML-IF)

ML-IF constructs forests of trees using multiple decision trees. Each decision tree trains on a particular sub-set of the dataset. ML-IF uses two methods, namely Entropy Information Gain (EIG) and Gini index (GI), to select the best features that maximizes the information for a particular split a particular decision tree. EIG uses use entropy to measure the degree of randomness; the higher the value of the EIG, the more order of disorder. Thus, we aim to decrease the entropy from the top of the tree to the bottom. EIG is defined as follows:

$$\mathcal{EIG} = -\sum_{i=1}^{n} q_i * log(q_i) \quad (4)$$

where q_i is the class probability; it is the proportion of class i in the dataset.

Gini Index, called also Gini impurity, calculates the amount of probability of a given feature that is misclassified when selected randomly. GI is defined as follows:

$$\mathcal{GI} = 1 - \sum_{i=1}^{n} {p_i}^2 \quad (5)$$

where p_i is the probability of a misclassified feature.

Once ML-IF constructs the optimal decision tree, we compute an anomaly score, for each input feature f, as follows:

$$Anomaly\_score(f,t) = 2^{\frac{-E(\varphi(f))}{d(t)}} \quad (6)$$

where $\varphi(f)$ is the path length of input feature f; it is defined as the number of edges that input feature f traverses from the top to the bottom of the tree, t is the number of leaf nodes, $E(\varphi(f))$ is the average of $\varphi(f)$, and $d(t)$ is the average path length of unsuccessful searches in the binary tree, it is defined as follows:

$$d(t) = 2H(t-1) - (2\frac{t-1}{t}) \quad (7)$$

where $H(k)$ is an harmonic number that can be estimated by Euler's constant ($i.e.$, $ln(k) + 0.5772156649$).

Using the anomaly score $Anomaly\_score(f,t)$, we can classify each input features f as either normal or abnormal. The closer an input feature f to the root of the decision tree (score close to 1), the more likely this data sample is an anomaly/intrusion, while an score close to 0, indicates that f is very likely a normal observation. Abnormal data samples are mostly few and have different representations, which result in a shorter path $\varphi(f)$ in the tree. Thus, a anomaly score close to 1.

IV. Evaluation

In this section, we present the evaluation of our proposed framework. First, we introduce the experimental environment. Then, we evaluate the performance of our proposed framework.

A. Experimental environment

The implementation of our proposed framework is done using scikit-learn [29], an open source library that integrates a wide range of supervised and unsupervised machine learning techniques. DFC, IGF, and ML-IF are implemented in the application layer as REST applications. We use Mininet [30] to create a realistic virtual network; it consists of: (1) SDN controllers (i.e., Floodlight [31]); (2) network monitors (i.e., sFlow-RT [28]) to monitor/gather the features of network data flow in a scalable and efficient way; and (3) multiple OF switches and multiple hosts are simulated to act as legitimate users. For the attack traffic, we leverage the well-known public synthetic dataset from Cyber range Lab of the Australian Centre for cyber Security (ACCS); it contains 100 GB of raw network packets; it contains a variety of simulated attacks including DDoS attacks. UNSW-NB15 contains about three million connection records; it includes the following network attacks: analysis, fuzzers, DoS, backdoors, reconnaissance, generic, exploits, shellcode, and worms. We run our experiments on Google Colaboratory [32] using the Tesla T4 GPU on a PC with Intel Core i7-8750H-2.2 GHz, 16GB RAM, and GTX 1050 GPU.

B. Performance Evaluation

We evaluate the performance of our proposed framework in terms of Accuracy, Precision, True Positive Rate (TPR) called Detection Rate (DR), Area Under the ROC Curve (AUC), and F1-score. We define these metrics as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = DR = TPR = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

$$FPR = \frac{FP}{TN + FP} \quad (12)$$

where, TP (True Positives) represent abnormal data samples that are correctly classified as abnormal data samples, FN (False Negatives) represent abnormal data samples that are classified as normal data samples, FP (False Positives) represent normal data samples that are classified as abnormal data samples, while TN (True Negatives) represent normal data samples that are correctly classified as normal data samples.

First, we encoded the categorical features of the UNSWNB15 dataset ($i.e.$, 'proto', 'state', 'service', and 'attack_cat') into numeric values and we re-scale the features values according to Eq. (13).

$$X'_i = \frac{X_i - Mean(X_i)}{stdev(X_i)} \quad (13)$$

where X_i denotes the feature ($e.g.$, 'attack_cat'), $Mean(X_i)$ and $stdev(X_i)$ denote, respectively, the mean and standard deviation values for each feature.

Once the preprocessing phase is done, we used IGF to select the most promising features and remove the redundant ones. Fig. 2 shows IGF score of features for the UNSWNB15 dataset; it shows the highest features that have gained the highest amount of knowledge/information in a descending order. We observe that more than 75% of the 49 features of the UNSWNB15 dataset are not contributing to make accurate classifications. IGF can greatly exclude the redundant/irrelevant features that can slow down the process of

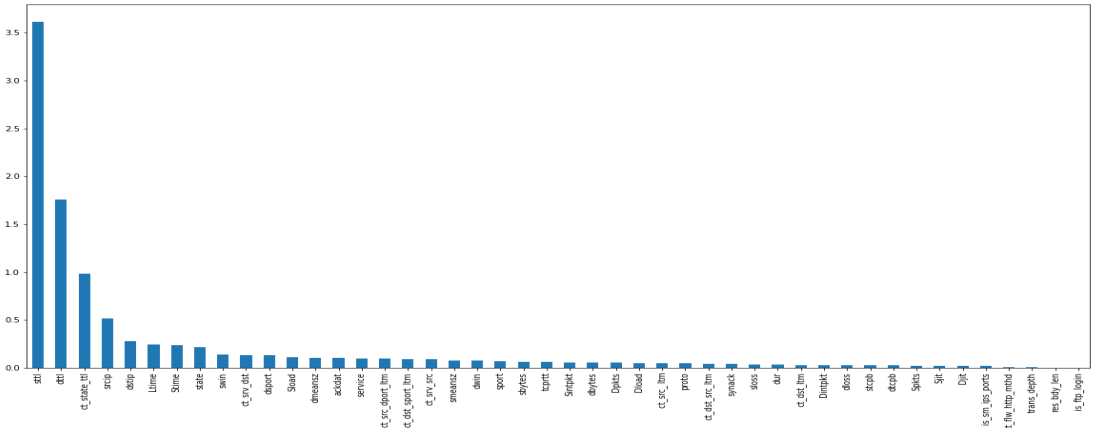
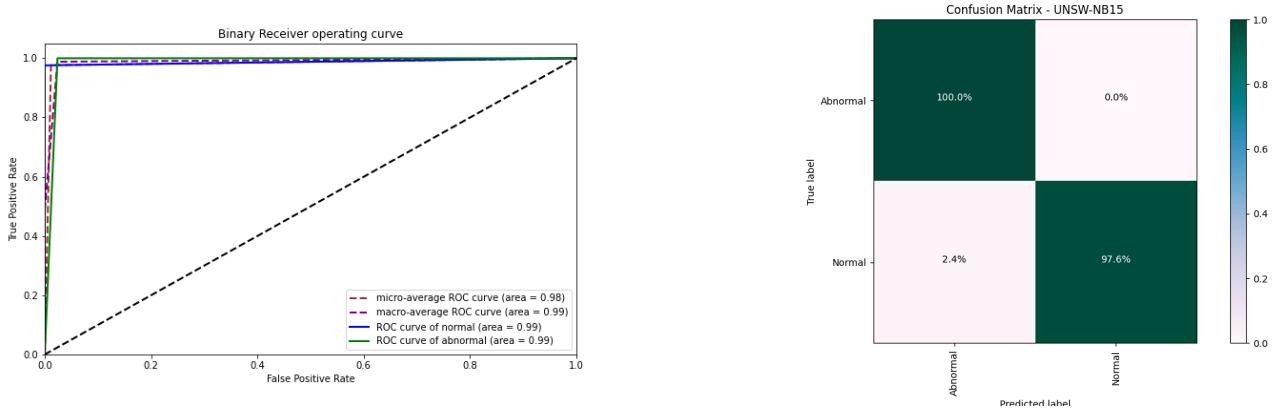


Fig. 2. IGF Score of features using the $UNSW-NB15$ dataset



(a) (b)

Fig. 3. Performance evaluation of the proposed framework using the $UNSW-NB15$ dataset in terms of: (a) ROC curve; and (b) Confusion matrix

training. Thus, IGF can significantly reduce computational complexity while maintaining/ensuring a high detection performance. Figs. 3(a) and 3(b) show the ROC curve and confusion matrix of our proposed framework using the UNSW-NB15 datasets, respectively. Our proposed framework achieves 97%, 96%, 96%, and 97% in accuracy, precision, Detection Rate, and F1 score, respectively, with only 38.33 seconds of training time. The ROC curves show the TPR according to the FPR. We obtain an AUC of 0.98 in the UNSW-NB15 dataset.

we compare the results obtained by our proposed framework with recent state-of-the-art ML/DL models using the UNSW-NB15: SSL [21], RF [22], FeedWSN [23], OGM [26], and MHMM [27]. Table 1 shows the values of the metrics of our proposed framework and the state-of-the-art ML/DL models. Our proposed framework achieves the highest accuracy of 97% and the highest detection rate of 96% with only 38.33 seconds of training time. The experimental results confirm that our proposed framework outperforms recent state-of-the-art contributions in terms of accuracy and detection rate while significantly reducing computational complexity. This makes it a promising cyber-security framework to detect the new emerging attacks while reducing the computational complexity.

TABLE I
PERFORMANCE METRICS OF OUR PROPOSED FRAMEWORK AND STATE-OF-THE-ART ML/DL MODELS BASED ON THE AVAILABLE MEASURES USING THE UNSW-NB15 DATASET

Methods	**Accuracy**	**DR**	**Time (second)**
SSL [21]	0.86	0.85	NA
RF [22]	0.93	0.92	NA
FeedWSN [23]	0.92	0.91	NA
OGM [26]	0.95	0.94	NA
MHMM [27]	0.96	0.95	NA
ML-IF	**0.97**	**0.96**	**38.33**

V. CONCLUSION

In this paper, we proposed a novel multi-module ML framework that combined unsupervised ML techniques with

a scalable feature collection and selection scheme to effectively/timely detect network security threats in the context of SDN. First, we introduced a novel data collection and feature selection module to reduce computational complexity while effectively improving detection performance. Then, we proposed a novel unsupervised ML module (ML-IF) to effectively/timely detect network security threats in SDN. The experimental results using the well-known public network security dataset UNSW-NB15, showed that our proposed framework outperforms state-of-the-art contributions in terms of accuracy and detection rate while significantly reducing computational complexity; making it a promising framework to mitigate the new emerging network security threats in SDN.

REFERENCES

[1] M. I. Oozeer and S. Haykin, "Cognitive risk control for mitigating cyber-attack in smart grid," *IEEE Access*, vol. 7, pp. 125 806–125 826, 2019.

[2] 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.

[3] S. Baker, S. Johnson, and P. Schneck, "Critical industries confront cyberattacks." Accessed: April. 1, 2021. [Online]. Available: https://csrc.nist.gov/CSRC/media/Events/ISPAB-JULY-2011-MEETING/documents/Jul14_CIP-CSIS-2011-ISPAB.pdf

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

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

[6] H. Amari, W. Louati, L. Khoukhi, and L. H. Belguith, "Securing software-defined vehicular network architecture against ddos attack," in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*, 2021, pp. 653–656.

[7] 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.

[8] S. Morgan, "Global ransomware damage costs predicted to reach $20 billion (usd) by 2021." Accessed: April. 1, 2021. [Online]. Available: https://cybersecurityventures.com/

[9] 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.

[10] 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.

[11] 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.

[12] G. Apruzzese, M. Andreolini, M. Marchetti, A. Venturi, and M. Colajanni, "Deep reinforcement adversarial learning against botnet evasion attacks," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 1975–1987, 2020.

[13] 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.

[14] F. O. Olowononi, D. B. Rawat, and C. Liu, "Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for cps," *IEEE Communications Surveys Tutorials*, vol. 23, no. 1, pp. 524–552, 2021.

[15] 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.

[16] J. A. Pérez-Díaz, I. A. Valdovinos, K.-K. R. Choo, and D. Zhu, "A flexible sdn-based architecture for identifying and mitigating low-rate ddos attacks using machine learning," *IEEE Access*, vol. 8, pp. 155 859–155 872, 2020.

[17] 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.

[18] S.-S. Li, "An improved dbscan algorithm based on the neighbor similarity and fast nearest neighbor query," *IEEE Access*, vol. 8, pp. 47 468–47 476, 2020.

[19] 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.

[20] N. Moustafa, "Unsw-nb15." Accessed: April. 1, 2021. [Online]. Available: www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-NB15-Datasets

[21] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484–497, 2017. Accessed: April. 1, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025516302547

[22] K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, "Big data analytics framework for peer-to-peer botnet detection using random forests," *Information Sciences*, vol. 278, pp. 488–497, 2014. Accessed: April. 1, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025514003570

[23] C. D. McDermott and A. Petrovski, "Investigation of computational intelligence techniques for intrusion detection in wireless sensor networks." *International journal of computer networks and communications*, vol. 9, pp. 45–56, 2017. Accessed: April. 1, 2021. [Online]. Available: http://hdl.handle.net/10059/2526

[24] N. N. Tuan, N. Danh Nghia, P. H. Hung, D. Khac Tuyen, N. M. Hieu, N. Tai Hung, and N. H. Thanh, "An abnormal network traffic detection scheme using local outlier factor in sdn," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, 2021, pp. 141–146.

[25] A. Ali and M. M. Yousaf, "Novel three-tier intrusion detection and prevention system in software defined network," *IEEE Access*, vol. 8, pp. 109 662–109 676, 2020.

[26] 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.

[27] N. Moustafa, E. Adi, B. Turnbull, and J. Hu, "A new threat intelligence scheme for safeguarding industry 4.0 systems," *IEEE Access*, vol. 6, pp. 32 910–32 924, 2018.

[28] "sflow-rt." Accessed: April. 1, 2021. [Online]. Available: http://www.sflow-rt.com

[29] "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 85, p. 2825–2830, 2011. Accessed: April. 1, 2021. [Online]. Available: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[30] "Mininet." Accessed: April. 1, 2021. [Online]. Available: http://mininet.org

[31] "Floodlight openflow controller." Accessed: April. 1, 2021. [Online]. Available: https://floodlight.atlassian.net/wiki/spaces/HOME/overview

[32] "Google colaboratory." Accessed: April. 1, 2021. [Online]. Available: https://colab.research.google.com/