2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

An Ensemble Learning to Detect Decision-based Adversarial Attacks In Industrial Control Systems

Narges Babadi, Hadis Karimipour, and Anik Islam
Schulich School of Engineering, Department of Electrical & Software Engineering,
University of Calgary, Calgary, Canada
Email: {narges.babadi1, hadis.karimipour anik.islam}@ucalgary.ca

***Abstract*—An increasing number of Intrusion Detection Systems (IDSs) rely on Artificial Intelligence (AI), specifically Machine Learning (ML) algorithms, to distinguish between benign and malicious data and detect cyber attacks. However, using ML algorithms exposes IDSs to Adversarial Machine Learning (AML) attacks during the training and test phase. These AML attacks aim to deceive ML algorithms by misclassifying data, posing significant disruptions to the system and its users. Two critical categories of AML attacks are White-box and Black-box attacks, with Black-box attacks being more practical and representative of real-world scenarios. This paper investigates the impact of adversarial examples on supervised ML models in IDSs and proposes an ensemble learning-based detection approach. The study uses a power system dataset and employs Random Forest, AdaBoost, and Decision Tree classifiers to achieve this. During the test phase, adversarial examples are generated using the decision boundary and HopSkipJump attacks, two types of Black-box decision-based attacks. The research applies a deep neural network to the dataset containing the generated adversarial examples to detect these AML attacks, achieving an accuracy of 98% to 99%.**

***Index Terms*—Adversarial Machine Learning attacks, Decision-based attacks, Industrial Control Systems, and Intrusion Detection systems.**

I. Introduction

Industrial Control Systems (ICSs) play a critical role in infrastructures like transportation, healthcare, power grids, and water plants. Given their interconnectedness through networks, ICS assets are susceptible to cyber-attacks, posing risks to operations and consumers [1]. To address this, Intrusion Detection Systems (IDSs) have been integrated into ICSs to detect anomalies and malicious events. Leveraging Artificial Intelligence (AI) techniques like machine learning (ML) algorithms can enhance IDSs' ability to distinguish malicious behavior. However, ML algorithms are vulnerable to Adversarial Machine Learning (AML) attacks, where attackers perturb datasets during training or testing to evade detection and cause misclassification of critical events like cyber attacks in IDSs. Thus, designing robust IDSs that can withstand AML attacks is essential [2], [3].

AML attacks can be categorized based on their characteristics [4]. They fall into targeted and untargeted settings, depending on whether the objective is to misclassify data points to a specific class or any classes except the original target class, respectively. Another classification is based on the attackers' accessibility to the model parameters, leading to two main groups: black-box attacks (only access to model outputs) and white-box attacks (complete knowledge of the model). These attacks can also be further divided into targeted and untargeted settings.

Black-box attacks are regarded as more sophisticated and representative of real-world scenarios, making their detection a critical concern. Identifying and mitigating such attacks is of paramount importance in the field of machine learning security.

Motivated by the threat of such sophisticated attacks, this paper aims to evaluate machine learning algorithms' robustness in the presence of decision-based AML attacks. Then, apply ensemble adversarial training to defend against such attacks in this context. The contributions of this paper are as follows:

- The paper implements two classes of decision-based attacks, namely the Decision Boundary Attack and HopSkipJump attack, during the test phase to evaluate the performance of the classifiers under consideration.
- A specially tuned deep learning neural network as an ensemble learning algorithm has been proposed. This ensemble approach effectively handles newly labeled datasets and detects AML attacks more accurately.
- The proposed algorithm is trained by various scenarios of generated adversarial attacks to enhance its robustness against this class of AML attacks.

The paper is structured as follows: Section II presents a brief review of related works. Section III provides an in-depth introduction to decision-based attacks. Section IV explains the methodology used in this study. In Section V the results of applying decision-based attacks to various ML algorithms. The paper concludes in Section VI with a summary and mentions potential future research to enhance this study.

II. Literature Review

AML attacks are classified into two main groups: black-box and white-box attacks. In the white-box attack scenario, also known as gradient-based attacks, adversaries exploit the gradient of the loss with respect to the model input. Conversely, black-box attacks, called gradient-free attacks, do not rely on gradient information. An overview of existing AML attacks is provided in Table I.

In the black-box category, attacks are divided into score-based and decision-based attacks. Score-based attacks lever-

TABLE I
ADVERSARIAL MACHINE LEARNING ATTACKS.

	White-box	**Black-box**	
		Score-based	Decision-based
Targeted	L-BFGS-B [5] JSMA [6]	Zoo [7]	Decision Boundary [8] HopSkipJump [9]
Untargeted	FGSM [10] DeepFool [11]	Local Search [12]	

age access to the model's prediction probabilities. Examples include the Zoo attack [7] in the targeted setting and Local Search [12] in the untargeted setting.

On the other hand, decision-based attacks involve adversaries who only know the model's predicted labels (i.e., the final decision). Some previous works have attempted to design black-box attacks similar to decision-based attacks, but these often required knowledge of data distribution or involved substantial perturbations [13], [14]. Addressing this limitation, the decision boundary attack was introduced in [8], which is a pure decision-based attack. This iterative algorithm rejects samples initialized in the target class and generates perturbations from the required distribution to minimize the distance from the original input. However, this attack requires numerous model queries. To enhance query efficiency, researchers in [9] proposed the HopSkipJump attack, a query-efficient decision-based attack that optimizes the minimum l_2 norm distance without relying on additional transferability assumptions.

Several approaches have been introduced regarding defense against AML attacks, such as adversarial training, ensemble adversarial training, defensive distillation, and stateful detection [10], [16]–[18], which are commonly used to defend against white-box attacks. However, it is essential to note that these defenses have less impact on black-box attacks, particularly decision-based ones. Defensive distillation, for example, was examined against the decision boundary attack in [8] and found ineffective. This paper uses ensemble learning [16] as a defense mechanism against decision-based adversarial attacks.

In the context of Industrial Control Systems (ICSs), significant research has been conducted to empower Intrusion Detection Systems (IDS) with AI tools. Common classifiers used in ICSs include Random Forest [19]–[24], Decision Tree [19]–[23], Recurrent Neural Networks (RNNs) such as Long Short Term Memory (LSTM) [25]–[27], Naïve Bayes [21]–[23], Adaboost [21], and Support Vector Machines (SVM) [21], [22], [28].

While AML attacks have been extensively studied in image classification, voice recognition, and e-mail spam classification, their application in IDS for ICS has received limited attention in the literature. Some notable works include [3], where a class of white-box AML attacks was applied to an ICS dataset. The perturbations in this work were manually generated to evaluate the classifier's robustness. Additionally, [29] focused on the Fast Gradient Sign Method (FGSM) and Jacobian-based Saliency Map Attack (JSMA) to assess Random Forest and J48 classifiers, attempting to improve accuracy through adversarial training as a defense mechanism. Furthermore, a model evasion attack against the Multi-Layer Perceptron (MLP) network was implemented in [29], demonstrating a significant decrease in classifier accuracy in the presence of this attack. This attack aimed to misclassify malicious network traffic in network-based IDS datasets. In [31], FGSM, DeepFool, and Jacobian-based saliency map attacks were generated to evaluate the performance of deep neural networks in Cyber-Physical Systems (CPS). To minimize the influence of FGSM attacks on network-based IDS, a min-max optimization problem was introduced in [32]. Moreover, a robust method was proposed in [33] to defend against generated adversarial examples in an IoT network intrusion detection system.

It is essential to highlight that many of these works primarily focused on white-box attacks, which are relatively simplistic and may not accurately represent real-world attack scenarios. Consequently, evaluating ML algorithms in the presence of decision-based attacks on ICS datasets remains relatively unexplored and represents a significant research gap in the field of AML for IDS in ICS.

III. DECISION-BASED ATTACKS

AML attacks are categorized based on complexity and knowledge about the learning model [6]. This study focuses on black-box attacks, specifically decision boundary attacks and HopSkipJump attacks, which are sophisticated methods for misclassifying ML models [8], [9].

Throughout the paper, the following notations and equations are used: x represents the original input, $y = F(x)$ denotes the probability of the model's final prediction, $y_{max} = argmaxF(x)$ is the predicted label, x' is the perturbed input, and x'_k refers to the perturbed input at the k-th step of the attack algorithm. The distance between the original and perturbed inputs, $d(x, x') = x - x'_2$, is defined as the l_2 norm. An adversarial region is the targeted region where the original data point will be placed. In a targeted setting, the attacker chooses this region knowingly, while in an untargeted setting, it can be any region except the original one.

Decision boundary attack: The decision boundary attack algorithm starts with selecting an initial point on a sphere around the original point in the adversarial region (Figure 1). It then generates a random walk toward the original output target, satisfying two conditions: staying in the adversarial region and reducing the distance toward the target label. The algorithm's primary goal is to find the slightest adversarial perturbations according to a given adversarial criterion $C(.)$, achieved by generating rejection sampling with a suitable proposal distribution.

In each step, the adversarial input will be updated by $x^{'}_k = x^{'}_{k-1} + \eta_k$ where η is a random perturbation drawn from the proposal distribution s.t. $\eta_k \sim P(x^{'}_{k-1})$. It will continue if $x^{'}_{k-1} + \eta_k$ is still adversarial. The proposal distribution plays a critical role in the efficiency of the boundary attack and depends on the input domain and the model. To draw the

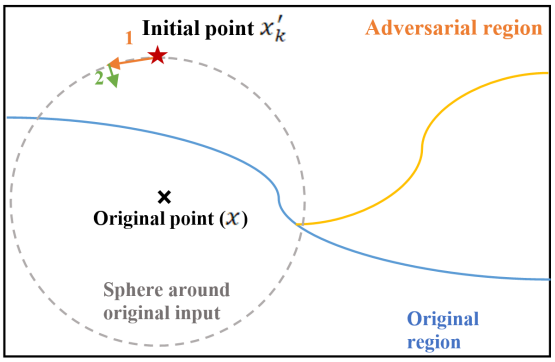

Fig. 1. Decision boundary attack.

perturbation from a maximum entropy distribution following constraints are essential [8]:

1) The perturbed sample $x'_{k-1} + \eta_k$ should lie within the input domain.
2) The perturbation has a size of δ such that:$\|\eta_k\|_2 = \delta.d(x, x'_{k-1})$.
3) The perturbation reduces the distance of the perturbed input towards the original input w.r.t such that: $d(x, x'_{k-1}) - d(x, x'_{k-1} + \eta_k) = \varepsilon.d(x, x'_{k-1})$.

However, in practice, it is difficult to sample from the distribution as mentioned above. Hence, it is suggested to sample from an IID Gaussian distribution $\eta_k \sim N(0, 1)$ and as a result, (1) and (2) will be held [8].

AML attacks aim to misclassify inputs into different classes from the original. These attacks adjust two key hyperparameters: the total perturbation length (δ) and the step size towards the original input (ε). These parameters are dynamically adjusted in each step. The attack process involves checking the adversarial nature of the orthogonal perturbation, and the step size ε is modified accordingly. The attack continues until the perturbation successfully misclassifies the input or converges to zero ε [8].

HopSkipJump attack: This advanced version of a boundary attack utilizes an iterative algorithm that requires access to gradient information, similar to a boundary attack. To reduce queries, [9] introduced additional steps to optimize the attack's model. In each step, the attack generates a gradient from the decision boundary and projects the point back to the boundary (Figure 2). This iterative approach involves an optimization problem. In the untargeted setting, the goal of this attack is to change the class of a given input c^* to any other class x^*, while in the targeted setting, it aims to change the decision to a predefined class $c \in \{m|m \neq c^*\}$. This is achieved by defining a function that proposes the difference between the original and adversary classes of the perturbed input [9].

$$S_{x^*} \triangleq \begin{cases} \max F_c(x') - F_{c^*}(x') & \text{Untargeted} \\ F_{c'}(x') - \max F_{c^*}(x') & \text{Targeted} \end{cases} \quad (1)$$

The perturbed input x' is a successfully attacked data point if and only if $S_{x^*} > 0$. This means that the label of the given input has been successfully changed. By redefining the l_2 norm $d(x, x^*) = \| x - x^* \|_2$ and giving access to the gradient ∇S_x , one can write [9]:

$$x_{k+1} = \alpha_k x^* + (1 - \alpha_k)\{x_k + \zeta_k \frac{\nabla S_{x^*}(x_k)}{\|\nabla S_{x^*}(x_k)\|_2}\} \quad (2)$$

Where ζ_k is a positive step size and α_k is a line search parameter s.t. $\alpha_k \in [0, 1]$. Equation (2) explains how to get access to the next iterate x_{k+1} that lies on the decision boundary [9].

The next iteration starts by re-projecting the perturbed sample to the decision boundary. This iteration continues until the perturbation becomes successful.

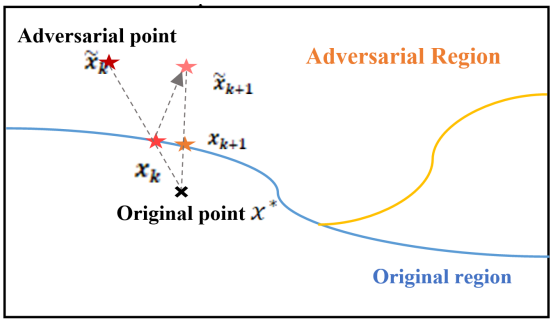

Fig. 2. Initiating the HopSkipJump attack.

A binary search is performed to find the boundary, then $\tilde{x}_K \rightarrow x_K$. The gradient of the boundary point x_K will be estimated. Then it is moving forward and updates $x_K \rightarrow \tilde{x}_{K+1}$. Again, a binary search would be performed, and then $\tilde{x}_{K+1} \rightarrow x_{K+1}$ will be updated [9].

IV. Proposed Approach

To detect AML attacks, we propose an ensemble approach utilizing Convolutional Neural Networks (CNNs). This ensemble of models is trained on generated adversarial examples to detect AML attacks effectively. This research involves three main steps, as Figure 3 illustrates.

1) **Generating adversarial examples:** The first step involves the generation of adversarial examples. Adversarial examples are crafted by perturbing the input data to mislead the model's predictions. Since attackers may not directly access the target model's internal information, they often use substitute models to generate adversarial examples. A substitute model is an alternative model trained to mimic the behavior of the target model based on the available query access. We employ Algorithm 1 to generate decision-based adversarial examples, which iteratively perturbs the input data to craft adversarial samples. The details of the "AttackGenerator" for each specific attack are elaborated in Section III.
2) **Training the ensemble:** The second step involves training the ensemble of models. In ensemble adversarial training, multiple models are trained using clean data and the generated adversarial examples. In the context of decision-based attacks, we employ the same substitute

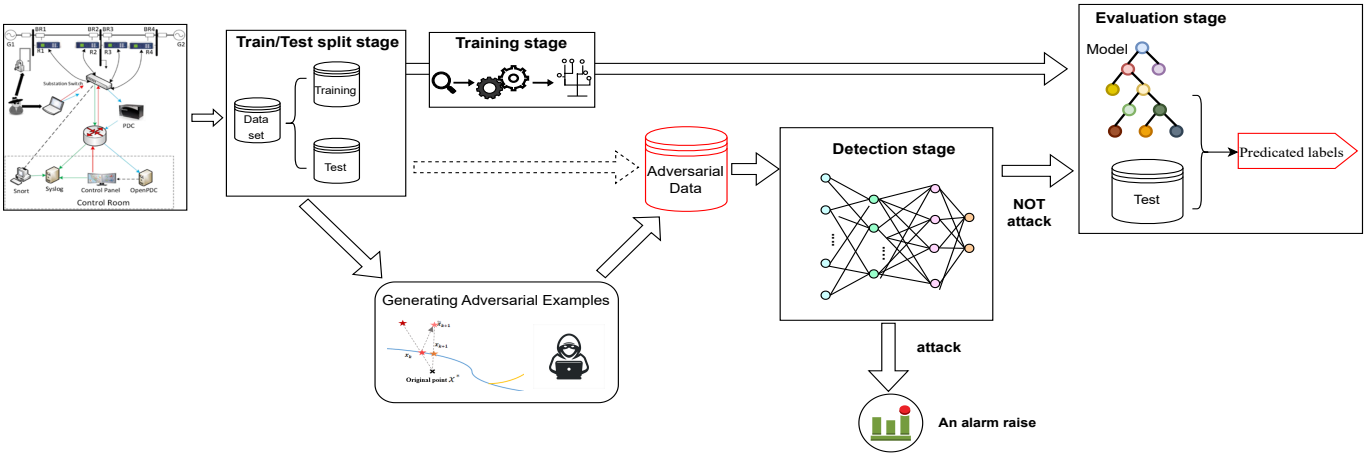

Fig. 3. Process of detection of Decision-based AML attacks using Ensemble learning.

Algorithm 1: Generating Adversarial Attack.

Data: X_{test} (clean data), Y_{test} (true labels)
Result: X_{adv} (adversarial examples)
$steps = 20$ (number of iterations for generating adversarial examples);
for *each iteration* i *in range(steps):* **do**
  $X_{adv} \leftarrow \text{AttackGenerator}(X_{adv\text{-}initial} = X_{test})$

model the attacker uses to generate adversarial examples for training. The ensemble models learn to defend against these attacks by exposing them to clean and adversarial data. Algorithm 2 provides an overview of the labeling process for AML attacks.

Algorithm 2: Labeling AML Attacks.

Data: X_{adv} (adversarial examples), Y_{test} (true labels)
Result: Y_{new} (labels indicating the presence of AML attacks)
$Y_{\text{prediction}} \leftarrow F(X_{test})$ (predictions on clean data);
$Y_{\text{adv}} \leftarrow F(X_{adv})$ (predictions on adversarial examples);
for *each data point* i *in the test set* **do**
  if $Y_{prediction}(i) = Y_{adv}$ **then**
    $Y_{\text{new}}(i) \leftarrow$ "No AML Attack"(clean data prediction);
  else
    $Y_{\text{new}}(i) \leftarrow$ "AML Attack"; (AML attack successfully changed the label)

3) **Ensemble prediction:** During the inference phase, the ensemble of models is utilized to make predictions on new input data. The predictions from each model are combined to arrive at the final prediction. This ensemble-based approach enhances robustness against decision-based attacks because individual models may exhibit different vulnerabilities and defense mechanisms, thus reducing the effectiveness of the attacker's strategies.

By implementing this ensemble of CNNs trained on generated adversarial examples, our proposed methodology enables accurate detection of AML attacks and enhances the overall security and reliability of the model in real-world scenarios. The robustness gained from the ensemble approach makes our method particularly suitable for defending against decision-based attacks, which have become increasingly sophisticated and challenging to detect.

V. Experiments and Results

The power system framework (Figure 4) implemented by Mississippi State University, a small representation of a larger power system, is used as the case study. The framework consists of power generators (G1 and G2), Intelligent Electronic Devices (IEDs) (R1, R2, R3, R4), and corresponding breakers (BR1, BR2, BR3, BR4) responsible for protecting electrical circuits from excess current. The supervisory control system allows manual commands to be sent to the IEDs, and additional network monitoring devices (SNORT and Syslog servers) are connected to the testbed. The power system comprises cyber and physical layers, both requiring secure measures. The framework also utilizes Synchrophasor or Phasor Measurement Units (PMUs) technology to provide real-time data to the energy management system (EMS) [35]. The power system datasets used in the evaluation include 15 datasets, each consisting of 37 power system event scenarios. These scenarios are categorized into No Events (1), Natural Events (8), and Attack Events (28). The scenarios include various events such as short-circuit faults, line maintenance, remote tripping command injection (cyber-attack event), relay setting change (cyber-attack event), and data injection (cyber-attack scenario). The datasets encompass 128 features, including measurements from synchrophasors (116 features), Snort logs, control panel logs, and relay logs (12 features). Pre-processing of the datasets revealed that using approximately 40 features from the available 128 can achieve comparable classification

TABLE II
PERFORMANCE RESULTS OF MACHINE LEARNING ALGORITHMS.

		Recall			Precision			Accuracy			F1 score		
		Random Forest	Decision Tree	Adaboost	Random Forest	Decision Tree	Adaboost	Random Forest	Decision Tree	Adaboost	Random Forest	Decision Tree	Adaboost
Normal operation		0.87	0.83	0.83	0.87	0.83	0.83	0.92	0.86	0.86	0.90	0.83	0.83
Decision Boundary	Targeted	0.49	0.45	0.45	0.48	0.45	0.45	0.78	0.72	0.72	0.48	0.45	0.45
	Untargeted	0.47	0.67	0.66	0.63	0.72	0.72	0.50	0.78	0.78	0.50	0.67	0.67
HopSkipJump	Targeted	0.49	0.45	0.55	0.48	0.45	0.53	0.78	0.71	0.79	0.48	0.45	0.54
	Untargeted	0.48	0.44	0.53	0.84	0.73	0.75	0.84	0.79	0.83	0.82	0.74	0.77

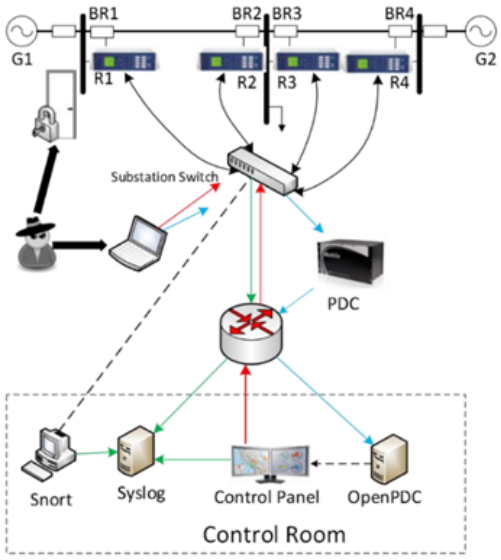

Fig. 4. The power system framework [35].

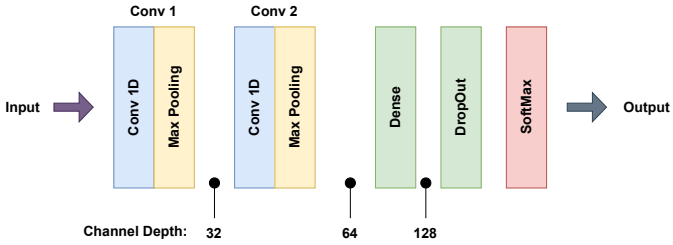

Fig. 5. Proposed CNN Model for Identifying AML Attacks.

performance. Researchers selected 52 features with the highest information gain based on the information gain rank. These features include Apparent Impedance measurements for each relay, Voltage Phase Angles, Current Phase Angles, and Voltage and Current Magnitudes.

The evaluation of the proposed methodology comprises three main aspects: regular operation and performance under the presence of two distinct decision-based attack scenarios. For this purpose, three classifiers, namely Decision Tree, Random Forest, and Adaboost, were trained on the respective datasets. The evaluation involved generating adversarial attacks, which were subsequently applied to the test data to assess the classifiers' performance. The datasets used in the evaluation are comprised of three distinct classes: attack events, natural events, and no events. In the targeted setting, the deceptive scenario involved relabeling data points such that instances initially classified as attack events were changed to no events or natural events. Similarly, data points initially labeled as natural events or no events were altered to be classified as attack events.

Machine Learning Algorithms Performance: Evaluation measures of each classifier for each scenario have been calculated (Table II). The F1 score describes the classification performance for both recall and precision scores. Adaboost and Random Forest outperformed better under HopSkipJump and Decision boundary Attacks in targeted and untargeted scenarios.

Detection of AML attacks using a CNN: The CNN model, depicted in Figure 5, was employed to detect AML attacks using carefully crafted adversarial examples. In this approach, if an adversarial example could successfully mislead any of the classifiers, it was categorized as an AML attack; otherwise, it was deemed a No AML attack. To ensure robustness and achieve optimal accuracy, the model underwent training with diverse scenarios of generated AML attacks, encompassing both targeted and untargeted settings for each attack type.

Following the rigorous training, the CNN model demonstrated impressive performance, achieving an accuracy of 99% for HopSkipJump attacks and 98% for Decision Boundary attacks.

VI. CONCLUSIONS AND FUTURE WORKS

This study delves into the impact of adversarial examples on supervised ML models within IDSs. The study employs a CNN to detect these attacks. The ML models are trained on a power system dataset consisting of three different classes, utilizing Random Forest, AdaBoost, and a variant of the Decision Tree classifier. During the testing phase, adversarial examples are generated using two types of decision-based attacks: the decision boundary attack and the HopSkipJump attack. The results demonstrate the CNN model's high accuracy of 99% for HopSkipJump attacks and 98% for Decision Boundary attacks, underscoring the potential effectiveness of ML models in detecting AML attacks within IDSs. The study emphasizes the significance of further research to develop more robust and efficient techniques for safeguarding AI-based IDSs against AML attacks.

REFERENCES

[1] Sakhnini, J., Karimipour, H., Dehghantanha, A., Yazdinejad, A., Gadekallu, T. R., Victor, N., and Islam, A., "A Generalizable Deep Neural Network Method for Detecting Attacks in Industrial Cyber-Physical Systems," in IEEE Systems Journal, doi: 10.1109/JSYST.2023.3286375.

[2] Hink, R.C.B., Beaver, J.M., Buckner, M.A., Morris, T., Adhikari, U., and Pan, S., 'Machine Learning for Power System Disturbance and Cyber-Attack Discrimination', in, 2014 7th International symposium on resilient control systems (ISRCS), (IEEE, 2014).

[3] Zizzo, G., Hankin, C., Maffeis, S., and Jones, K., 'Adversarial Machine Learning Beyond the Image Domain', in, 2019 56th ACM/IEEE Design Automation Conference (DAC), (IEEE, 2019)

[4] Qiu, S., Liu, Q., Zhou, S., and Wu, C., 'Review of Artificial Intelligence Adversarial Attack and Defense Technologies', Applied Sciences, 2019, 9, (5), p. 909.

[5] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R., 'Intriguing Properties of Neural Networks', arXiv preprint arXiv:1312.6199, 2013.

[6] Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.B., and Swami, A., 'The Limitations of Deep Learning in Adversarial Settings', in, 2016 IEEE European symposium on security and privacy (EuroS & P), (IEEE, 2016)

[7] Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., and Hsieh, C.-J., 'Zoo: Zeroth Order Optimization Based Black-Box Attacks to Deep Neural Networks without Training Substitute Models', in, Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, (2017)

[8] Brendel, W., Rauber, J., and Bethge, M., 'Decision-Based Adversarial Attacks: Reliable Attacks against Black-Box Machine Learning Models', arXiv preprint arXiv:1712.04248, 2017.

[9] Chen, J., Jordan, M.I., and Wainwright, M.J., 'Hopskipjumpattack: A Query-Efficient Decision-Based Attack', in, 2020 ieee symposium on security and privacy (sp), (IEEE, 2020)

[10] Goodfellow, I.J., Shlens, J., and Szegedy, C., 'Explaining and Harnessing Adversarial Examples', arXiv preprint arXiv:1412.6572, 2014.

[11] Moosavi-Dezfooli, S.-M., Fawzi, A., and Frossard, P., 'Deepfool: A Simple and Accurate Method to Fool Deep Neural Networks', in, Proceedings of the IEEE conference on computer vision and pattern recognition, (2016)

[12] Narodytska, N. and Kasiviswanathan, S., 'Simple Black-Box Adversarial Attacks on Deep Neural Networks', in, 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (IEEE, 2017)

[13] Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G., and Roli, F., 'Evasion Attacks against Machine Learning at Test Time', in, Joint European conference on machine learning and knowledge discovery in databases, (Springer, 2013)

[14] Ren, K., Zheng, T., Qin, Z., and Liu, X., 'Adversarial Attacks and Defenses in Deep Learning', Engineering, 2020.

[15] Brunner, T., Diehl, F., Le, M.T., and Knoll, A., 'Guessing Smart: Biased Sampling for Efficient Black-Box Adversarial Attacks', in Proceedings of the IEEE International Conference on Computer Vision, (2019)

[16] Tramèr, F., Kurakin, A., Papernot, N., Goodfellow, I., Boneh, D., and McDaniel, P., 'Ensemble Adversarial Training: Attacks and Defenses', arXiv preprint arXiv:1705.07204, 2017.

[17] Papernot, N., McDaniel, P., Wu, X., Jha, S., and Swami, A., 'Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks', in, 2016 IEEE Symposium on Security and Privacy (SP), (IEEE, 2016)

[18] Chen, S., Carlini, N., and Wagner, D., 'Stateful Detection of Black-Box Adversarial Attacks', in Proceedings of the 1st ACM Workshop on Security and Privacy on Artificial Intelligence, (2020)

[19] Siddavatam, I.A., Satish, S., Mahesh, W., and Kazi, F., 'An Ensemble Learning for Anomaly Identification in Scada System', in, 2017 7th International Conference on Power Systems (ICPS), (IEEE, 2017)

[20] Wang, D., Wang, X., Zhang, Y., and Jin, L., 'Detection of Power Grid Disturbances and Cyber-Attacks Based on Machine Learning', Journal of Information Security and Applications, 2019, 46, pp. 42-52.

[21] Morris, T.H., Thornton, Z., and Turnipseed, I., 'Industrial Control System Simulation and Data Logging for Intrusion Detection System Research', 7th annual Southeastern Cyber security summit, 2015, pp. 3-4.

[22] Yeckle, J. and Abdelwahed, S., 'An Evaluation of Selection Method in the Classification of Scada Datasets Based on the Characteristics of the Data and Priority of Performance', in Proceedings of the International Conference on Compute and Data Analysis, (2017)

[23] Teixeira, M.A., Salman, T., Zolanvari, M., Jain, R., Meskin, N., and Samaka, M., 'Scada System Testbed for Cybersecurity Research Using Machine Learning Approach', Future In-ternet, 2018, 10, (8), p. 76.

[24] Hammad, M., Hewahi, N., and Elmedany, W., 'T-Snerf: A Novel High Accuracy Machine Learning Approach for Intrusion Detection Systems', IET Information Security, 2021.

[25] Feng, C., Li, T., and Chana, D., 'Multi-Level Anomaly Detection in Industrial Control Systems Via Package Signatures and Networks', in, 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), (IEEE, 2017)

[26] Gao, J., Gan, L., Buschendorf, F., Zhang, L., Liu, H., Li, P., Dong, X., and Lu, T.,' for Scada Intrusion Detection', in, 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), (IEEE, 2019)

[27] Goh, J., Adepu, S., Tan, M., and Lee, Z.S., 'Anomaly Detection in Cyber-Physical Systems Using Recurrent Neural Networks', in, 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), (IEEE, 2017)

[28] Han, W., Xue, J., and Yan, H., 'Detecting Anomalous Traffic in the Controlled Network Based on Cross Entropy and Support Vector Machine', IET Information Security, 2019, 13, (2), pp. 109-116.

[29] Anthi, E., Williams, L., Rhode, M., Burnap, P., and Wedgbury, A., 'Adversarial Attacks on Machine Learning Cybersecurity Defences in Industrial Control Systems', Journal of Information Security and Applications, 2021, 58, p. 102717.

[30] Ayub, M.A., Johnson, W.A., Talbert, D.A., and Siraj, A., 'Model Evasion Attack on Intrusion Detection Systems Using Adversarial Machine Learning', in, 2020 54th Annual Conference on Information Sciences and Systems (CISS), (IEEE, 2020)

[31] H. Figueroa, Y. Wang and G. C. Giakos, "Adversarial Attacks in Industrial Control Cyber Physical Systems," 2022 IEEE International Conference on Imaging Systems and Techniques (IST), Kaohsiung, Taiwan, 2022, pp. 1-6, doi: 10.1109/IST55454.2022.9827763.

[32] Abou Khamis, Rana, M. Omair Shafiq, and Ashraf Matrawy. "Investigating Resistance of Deep Learning-based IDS against Adversaries using min-max Optimization." ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.

[33] Jiang, Hongling, Jinzhi Lin, and Haiyan Kang. "FGMD: A robust detector against adversarial attacks in the IoT network." Future Generation Computer Systems 132 (2022): 194-210.

[34] Strauss, T., Hanselmann, M., Junginger, A., and Ulmer, H., 'Ensemble Methods as a Defense to Adversarial Perturbations against Deep Neural Networks', arXiv preprint arXiv:1709.03423, 2017.

[35] Powersystem Dataset Readme.Pdf', 2020[Accessed 18 March 2020].