IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Countering Evasion Attacks for Smart Grid Reinforcement Learning-based Detectors

AHMED T. EL-TOUKHY[1,2], MOHAMED M. E. A. MAHMOUD[1], (Senior Member, IEEE), ATEF H. BONDOK[1], MOSTAFA M. FOUDA[3, 4], (Senior Member, IEEE), and MAAZEN ALSABAAN[5]

[1]Department of Electrical and Computer Engineering, Tennessee Tech University, Cookeville, TN 38505, USA.
[2]Department of Electrical Engineering, College of Engineering, Al-Azhar University, Cairo 11884, Egypt.
[3]Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID 83209, USA.
[4]Center for Advanced Energy Studies (CAES), Idaho Falls, ID 83401, USA.
[5]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia.

Corresponding author: Mohamed Mahmoud (e-mail: mmahmoud@tntech.edu).

ABSTRACT

Fraudulent customers in smart power grids employ cyber-attacks by manipulating their smart meters and reporting false consumption readings to reduce their bills. To combat these attacks and mitigate financial losses, various machine learning-based electricity theft detectors have been proposed. Unfortunately, these detectors are vulnerable to serious cyber-attacks, specifically evasion attacks. The objective of this paper is to investigate the robustness of deep reinforcement learning (DRL)-based detectors against our proposed evasion attacks through a series of experiments. Firstly, we introduce DRL-based electricity theft detectors implemented using the double deep Q networks (DDQN) algorithm. Secondly, we propose a DRL-based attack model to generate adversarial evasion attacks in a black box attack scenario. These evasion samples are generated by modifying malicious reading samples to deceive the detectors and make them appear as benign samples. We leverage the attractive features of reinforcement learning (RL) to determine optimal actions for modifying the malicious samples. Our DRL-based evasion attack model is compared with an FGSM-based evasion attack model. The experimental results reveal a significant degradation in detector performance due to the DRL-based evasion attack, achieving an attack success rate (ASR) ranging from 92.92% to 99.96%. Thirdly, to counter these attacks and enhance detection robustness, we propose hardened DRL-based defense detectors using an adversarial training process. This process involves retraining the DRL-based detectors on the generated evasion samples. The proposed defense model achieves outstanding detection performance, with a degradation in ASR ranging from 1.80% to 9.20%. Finally, we address the challenge of whether the DRL-based hardened defense model, which has been adversarially trained on DRL-based evasion samples, is capable of defending against FGSM-based evasion samples, and vice versa. We conduct extensive experiments to validate the effectiveness of our proposed attack and defense models.

INDEX TERMS

Security, electricity theft, evasion attacks, reinforcement learning, adversarial training, and smart power grids.

I. INTRODUCTION

With the rapid advancements in communication and power control systems, significant improvements and developments have been made in power grids. These changes have brought about revolutionary progress in conventional power grids, enabling the smart grid (SG) to assume important roles in enhancing the reliability, efficiency, resiliency, and sustainability of the power system [1], [2]. Alongside facilitating the reliable delivery of electricity, optimizing and regulating grid operation, and monitoring the performance of the system, the SG plays a vital role in achieving these objectives. The core structure of the SG consists of several main components, including advanced metering infrastructure (AMI), system operator (SO), and electricity production stations, along with transmission and distribution systems [3], [4]. The AMI network is a major component of the SG as it provides bidirectional communication between the SO, installed at the electric utility side, and smart meters (SMs) deployed at customers' homes. SMs regularly monitor electricity usage and periodically transmit detailed consumption readings to the SO through the AMI network. As a result, the SO can utilize these periodic consumption readings for various purposes, including load monitoring, implementing dynamic pricing for calculating consumption bills, and efficiently managing power resources [5], [6].

Despite the progress achieved in smart grid technology, the issue of electricity theft continues to pose a significant challenge. Dishonest individuals employ various fraudulent practices, such as tampering with mechanical meters in traditional power grids. Similarly, in smart power grids, the vulnerability lies in SMs, which are software-driven embedded systems. These SMs are target for cyber-attacks orchestrated by malicious consumers aiming to manipulate electricity consumption readings and unlawfully reduce their bills [7], [8]. In the context of the SG, the problem of electricity theft through cyber-attacks becomes a heightened and significant concern compared to traditional power grids. This is primarily attributed to the potential for significant financial losses and disruptions in the smooth operation of the power grid [9], [10]. The increased concern arises from the crucial role played by the consumption data reported by SMs in enabling efficient grid management [11], [12].

Consequently, experts in the domains of cyber-security and artificial intelligence (AI) are increasingly directing their attention towards the detection of electricity theft [13], [14]. The existing literature showcases a range of supervised and unsupervised machine learning (ML) models, including deep learning (DL) and shallow models, proposed to detect electricity theft cyber-attacks [5], [12], [15], [16]. Current shallow ML detectors have limitations in effectively detecting electricity theft, primarily due to their inability to capture the intricate patterns and temporal dynamics present in electricity consumption readings. Hence, the primary emphasis in the literature is placed on DL-based models due to their ability to achieve higher detection accuracy when compared to shallow classifiers.

The existing ML-based electricity theft detectors in the literature can be divided into two main categories: global models and customized models [1], [13], [17], [18]. A global model is trained on the electricity consumption data of a diverse range of customers with varying consumption patterns. It can be employed to detect theft across all customers. On the other hand, a customized model is trained on the electricity consumption data of an individual customer and is specific to that customer. However, the practicality of customized models is limited as they require a substantial amount of historical electricity consumption data for training the detector [1]. Furthermore, customized detection models are vulnerable to data contamination attacks. A malicious customer can submit false readings from the beginning, allowing future electricity theft to go undetected since the detector would be trained on the false readings [19]. Moreover, training a separate detector for each customer would impose a substantial computational burden on the power utility. As a result, the current literature predominantly advocates for constructing a global electricity theft detector rather than employing customized detectors [1], [13], [14], [17], [18], [20].

However, the existing models in the literature are not without limitations. They often rely on fixed datasets, making them susceptible to overfitting and learning specific patterns and features rather than more generalized ones. Furthermore, these models exhibit limited adaptability to changes in consumption patterns and emerging cyber-attacks, necessitating the time-consuming and computationally intensive process of retraining the models using both existing and new data, especially when dealing with large datasets.

Reinforcement learning (RL) has recently emerged as a noteworthy branch of machine learning in the realm of cyber-security. Its growing popularity can be attributed to its capability to interact with and adapt to the surrounding environment, enabling it to tackle dynamic decision-making challenges [21], [22]. RL is specifically designed to address such issues and has the capacity to learn optimal decision-making even with limited initial knowledge of the environment. Furthermore, RL models excel in finding the right balance between exploration and exploitation, which is a crucial aspect in cyber-security where attackers are constantly evolving and changing their strategies [23]–[25]. Additionally, RL enables the integration of human expertise into the decision-making process [26]. Experts can provide feedback and guidance to the RL agent, enhancing its performance. This human-in-the-loop approach enhances the accuracy and effectiveness of cyber-security attack detection, leveraging the strengths of both human expertise and machine learning [27]–[29]. In conclusion, RL offers unique advantages in handling dynamic decision-making challenges in the context of cyber-security attack detection, making it a promising approach to complement deep learning in this field.

In the research presented in [30], we conducted the first exploration of utilizing deep reinforcement learning (DRL) for the purpose of electricity theft detection. While this study yielded promising outcomes, its primary focus was on detecting false readings arising from electricity theft attacks, and it did not encompass the investigation of adversarial evasion attacks. In contrast, this paper specifically focuses on addressing more advanced cyberattacks, particularly adversarial evasion attacks, which occur during the testing phase. These sophisticated attacks are purposefully crafted to evade detection and deceive the detector, leading to a degradation in overall detection performance. Consequently, effectively countering such attacks poses a significant challenge for existing electricity theft detection models. Specifically, this paper presents a defense model that utilizes DRL to counter adversarial evasion attacks. The proposed solution encompasses three main phases. In the first phase, a global detection model based on DRL is introduced, employing Double Deep Q-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3312376

Network (DDQN) with various neural network architectures such as convolutional neural network (CNN), gated recurrent unit neural network (GRU), and feedforward neural network (FFNN). Moving to the second phase, an attack model is developed to generate adversarial evasion samples by using malicious electricity consumption readings. This is done under the assumption of a black-box attack scenario, where the attacker lacks knowledge about the detection model. Two techniques are developed for generating evasion samples: a DRL-based DDQN model that incorporates a substitute model on the attacker's side, and the Fast Gradient Sign Method (FGSM). Lastly, in the third phase, a defense model is introduced, aiming to strengthen the detection model through an adversarial training process using the generated evasion samples.

Despite RL's inherent adaptability and attractive properties, it has not received adequate attention in the field of cyberattacks pertaining to electricity theft. Therefore, in comparison to existing literature, *our study takes the lead by introducing the utilization of an RL-based attack model for generating evasion attacks* against another RL-based detection model. Additionally, we are the first to evaluate the ability of the hardened defense model, which is trained adversarially on evasion samples generated by one attack, to defend against evasion samples generated by a different attack method. The main contributions of this paper are summarized as follows.

- Developing DRL-based DDQN and FGSM-based attack models for generating adversarial evasion samples of SMs electricity consumption readings in the context of a black-box attack scenario.
- Investigating the effectiveness of a DRL-based adversarial training defense model in defending against both DRL and FGSM-based adversarial evasion samples.
- Addressing the challenge of whether the DRL-based hardened defense model, which has been adversarially trained on DRL-based evasion samples, is capable of defending against FGSM-based evasion samples and vice versa.

The remaining sections of this paper are structured as follows. Section II provides an overview of the relevant literature that has examined adversarial evasion attacks in the smart grid context. Section III introduces the fundamental concept of RL. In Section IV, we describe the dataset preparation process for training our attack and defense models. The proposed DRL-based and FGSM-based attack models, as well as the proposed adversarial training-based defense mechanism are discussed in Section V. In Section VI, we evaluate and analyze our proposed attack and defense models. Finally, in Section VII, we conclude the paper.

II. RELATED WORKS

In this section, we begin by examining the prior research conducted on the detection of electricity theft. Subsequently, our focus shifts towards an overview of the existing approaches used for launching evasion attacks, as well as the countermeasures proposed to address them. Lastly, we delve into the limitations present in the literature and identify the areas that require further research.

A. ELECTRICITY THEFT DETECTION METHODS

In the context of electricity theft detection, various methods have been developed to address and mitigate the impacts of this issue. These methods can be broadly categorized into three main types: hardware-based, statistical and analytical-based, and machine learning-based methods. In this section, we provide a brief overview of these methods, outlining their key characteristics and functionalities.

1) Hardware-based methods

One method to address electricity theft attacks involves the integration of hardware tamper-proof modules into smart meters, which act as a deterrent against unauthorized modifications and the transmission of falsified data [31]. However, it is important to recognize the limitations associated with this method. The implementation of such modules can be costly, and their effectiveness depends on a level of trust that may not always be guaranteed in real-world situations. Consequently, within the existing literature, there is an increasing preference for statistical-based and ML-based methods as they have the potential to overcome these limitations and offer more effective countermeasures against electricity theft [13].

2) Statistical and analytical-based methods

Various statistical and analytical techniques have been proposed as countermeasures against electricity theft attacks. These methodologies utilize various approaches such as metaheuristic methods [32], [33], game theory [34]–[36], data mining, state estimation [37], clustering, principal component analysis (PCA), and local outlier factor (LOF). For instance, Singh *et al.* [38] propose an innovative approach that employs PCA to detect anomalies by calculating an anomaly score and comparing it to a predefined threshold. Additionally, Peng *et al.* [39] utilize the robust k-means clustering algorithm to group customers based on their electricity consumption readings, enabling the identification of outlier candidates whose consumption readings significantly deviate from the centers of their respective clusters. To further enhance the anomaly detection, LOF technique is used to compute an anomaly score for each identified outlier candidate, providing a comprehensive assessment of their anomalous behavior. However, it should be noted that statistical and analytical methods often suffer from limitations in capturing the temporal dynamics and intricate patterns present in the data, which may degrade their accuracy [40].

3) Machine learning-based methods

In the literature, researchers have proposed multiple ML-based detectors to address the issue of detecting false power consumption readings reported by malicious SMs. These detectors can be categorized into two main groups. The first group comprises detectors that utilize shallow ML detection algorithms such as decision trees (DTs), logistic regression (LR), Naïve Bayes, autoregressive integrated moving average (ARIMA), and support vector machines (SVM) [13], [14], [17], [33], [41], [42]. The second group consists of detectors that leverage DL detection algorithms [18], [20], [31]. DL algorithms, unlike shallow ML algorithms, possess the advantage of automatic feature extraction without the need for explicit feature engineering. Numerous studies [18], [20], [31], [33], [43]–[47] consistently demonstrate the promising potential of DL, consistently showcasing its superiority over shallow ML algorithms. This superiority is reflected in the higher detection accuracy achieved by DL models, making them a crucial and effective approach for detecting malicious power consumption readings.

Jokar *et al.* [13] introduced a methodology for electricity theft detection utilizing custom detectors trained on real benign readings obtained from the Irish dataset [48]. Since the Irish dataset solely consisted of benign data samples, they devised a set of attacks to generate malicious data samples by manipulating the original benign data. The study conducted two distinct experiments. In the first experiment, a single-class SVM-based detector was exclusively trained using benign data samples from the Irish dataset. In the second experiment, a multi-class SVM-based detector for each customer was trained using both benign and malicious data samples. The experimental results indicate that the SVM-based detector trained on both types of data samples outperforms the one trained solely on benign data samples.

Furthermore, Buzau *et al.* [17] have introduced a detection approach utilizing Extreme Gradient Boosted Trees (XGBoost) as a global detector. To improve the accuracy of detecting electricity theft cyber-attacks, the detector was trained not only on consumers' electricity consumption readings but also incorporated additional information such as geographical locations and technical characteristics of their SMs. The experimental results indicate that the proposed detector surpasses other detectors based on K-nearest neighbors (KNNs), SVM, and logistic regression in terms of accuracy.

Bhat *et al.* [18] introduced multiple DL-based global detectors employing various architectures such as CNN, LSTM, and stacked autoencoder. Through a comprehensive comparison with shallow-based detectors, the results demonstrate the superior accuracy of DL-based detectors. Furthermore, Li *et al.* [49] proposed a hybrid CNN-RF detector that combines the CNN's ability to capture consumption reading features with the RF's classification power for identifying electricity theft. The experimental results validate the superiority of the hybrid model over other shallow detection models, including GBDT, RF, LR, and SVM.

Similarly, in the same context of DL-based detectors, Zheng *et al.* [31] devised a DL-based electricity theft detector that utilized CNN and MLP to analyze weekly consumption readings and identify fraudulent behaviors. The experimental results showcased the superior performance of the proposed detector compared to other models, including LR, RF, SVM, and CNN. Furthermore, a hybrid CNN-LSTM model for electricity theft detection was proposed in [50], showing promising performance with a classification accuracy of 89%. However, ML-based models proposed in the literature exhibit certain vulnerabilities. They may not efficiently adapt to changes in consumption patterns and cyber-attacks, necessitating retraining on new datasets, which can be time-consuming and computationally intensive, especially for large datasets. Additionally, ML models typically lack a built-in exploration mechanism, relying solely on the training data to minimize a loss function. As a result, DL models are less flexible and may struggle to handle novel situations effectively [30].

Therefore, RL offers a flexible solution for handling the dynamic nature of electricity theft attacks and consumption patterns. Our recent study in [30] made the first attempt to explore the application of RL in detecting such attacks. We proposed a DRL approach that encompassed four different scenarios. In the first scenario, we developed a global detection model using DQN and DDQN algorithms, employing architectures such as FFNN, CNN, GRU, and a hybrid CNN-GRU model. The second scenario involved constructing customized detection models for new customers based on the global detector, aiming to achieve high accuracy and prevent zero-day attacks. In the third scenario, we addressed changes in consumption patterns among existing customers, ensuring adaptability to evolving scenarios. Lastly, the fourth scenario tackled the challenges of defending against newly launched cyber-attacks. The experimental results showcased the ability of the proposed detectors to enhance the detection of electricity theft cyber-attacks. Moreover, our approach demonstrated efficient learning of new consumption patterns, adaptability to changes in existing customers' consumption patterns, and effective defense against newly launched cyber-attacks.

B. ADVERSARIAL EVASION ATTACKS AND COUNTERMEASURES

The previous subsection discussed works that primarily focused on training accurate models for detecting electricity theft attacks. However, these works did not address the security of the models against adversarial evasion attacks. Evasion attacks employ advanced techniques to make minimal alterations to malicious samples, causing the detector to incorrectly classify them as benign. In this subsection, our attention shifts to these specific attacks and the countermeasures proposed to mitigate them.

Numerous research studies have delved into the generation of adversarial evasion samples and their impact on machine learning-based detectors. The pioneering work by Szegedy et al. [51] explored the effects of evasion attacks on neural networks. Subsequently, Goodfellow et al. [52], Moosavi-Dezfoli et al. [53], and Rozsa et al. [54] proposed various techniques, including the fast gradient sign method (FGSM), DeepFool, and fast gradient value (FGV), respectively, to compute evasion samples that could deceive detection models. In the domain of electricity power, Badr et al. [1] introduced a novel form of evasion attacks that targeted global electricity theft detectors. They employed a generative adversarial network (GAN) trained on real data to generate deceptive low-consumption readings capable of evading the global detector.

In addition, Li et al. [49] demonstrated the susceptibility of DL-based electricity theft detectors to evasion attacks. They introduced the SearchFromFree algorithm, which leverages gradients to create evasion samples, enabling malicious samples to evade DL-based detectors while yielding financial benefits. To address and mitigate adversarial evasion attacks, the technique of adversarial training defense has emerged as a promising approach to enhance detector resilience. Adversarial training, initially proposed by Szegedy et al. [51], involves subjecting a trained detector to evasion attacks in order to generate adversarial samples that can evade detection. The detector is then retrained using these generated adversarial samples to improve its robustness.

C. LIMITATIONS AND RESEARCH GAPS

In this subsection, we discuss the main limitations present in the existing literature which constitute the motivations for this work, as follows.

- *Lack of generalization*. Many existing evasion attacks exploit specific vulnerabilities or weaknesses in a particular machine learning model. These attacks often rely on knowledge of the target model's architecture, parameters, or training data. As a result, they may not generalize well to different models or datasets. This limitation restricts their applicability in different applications where the attacker may not have complete knowledge of the target model, limiting the effectiveness of the attacks.
- *High computational cost*. Some evasion attack methods utilize computationally expensive techniques such as optimization algorithms or brute-force search. These approaches explore a large search space to find the optimal perturbations that can fool the target model. However, these techniques are time-consuming and resource-intensive, especially for large datasets. The high computational cost makes these evasion attacks impractical for large-scale applications where efficiency is crucial.
- *Limited transferability*. Many existing works assume that evasion samples generated by a substitute model will also evade the target model, but this assumption is not guaranteed practically. The underlying reasons for limited transferability include differences in model architectures or decision boundaries.

To address these limitations and fill the existing research gap, we offer the following rationale for investigating RL as a viable approach to generate evasion samples:

- *Model-agnostic attacks*. RL-based evasion attacks have the potential to generate evasion samples that are less dependent on the architecture and type of the target model or the training dataset. By learning from interactions with the target model, RL attack agent can adapt its attack strategy, making it more versatile and transferable. This adaptability allows for more successful evasion attacks in scenarios where the attacker has limited knowledge of the target model compared to the traditional approaches.
- *Optimization efficiency*. RL techniques leverage exploration and exploitation techniques to compute evasion samples more efficiently. Instead of relying on exhaustive search or computationally expensive optimization algorithms, RL agents can learn to navigate the evasion sample space in efficient manner. This improves the practicality and scalability of the evasion attack.
- *Stealth and imperceptibility*. RL agents can be trained to generate evasion samples that are indistinguishable from the benign samples. By incorporating appropriate reward values, RL agents can learn to minimize the perturbations in a way that makes it hard to detect the attack. This ability to generate stealthy and imperceptible evasion samples increases the effectiveness of the attacks in different applications or scenarios where the goal is to evade the detection model without raising suspicion.

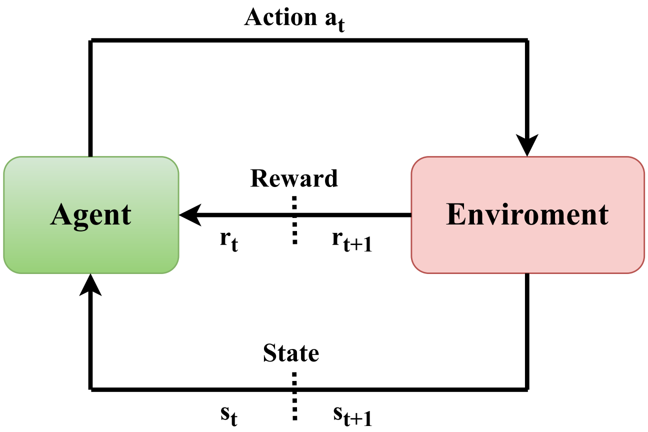

FIGURE 1: The main structure of an RL scheme.

III. PRELIMINARIES

A. REINFORCEMENT LEARNING (RL)

RL distinguishes itself as a unique branch of machine learning, setting it apart from popular methods like supervised learning. Unlike those methods, RL empowers autonomous agents to shape their own learning experiences through direct interaction with the environment and feedback. The fundamental structure of an RL model comprises two key components: the environment and the agent, as depicted in FIGURE 1. Initially, the agent possesses limited or no prior knowledge of the environment. To address RL problems, a Markov decision process incorporates four distinct components: state (s), action (a), reward (r), and policy (π). RL follows a trial-and-error approach, where the agent takes action a_t at each time step t, causing a transition from the current state s_t to a new state s_{t+1}, and receiving a reward or penalty from

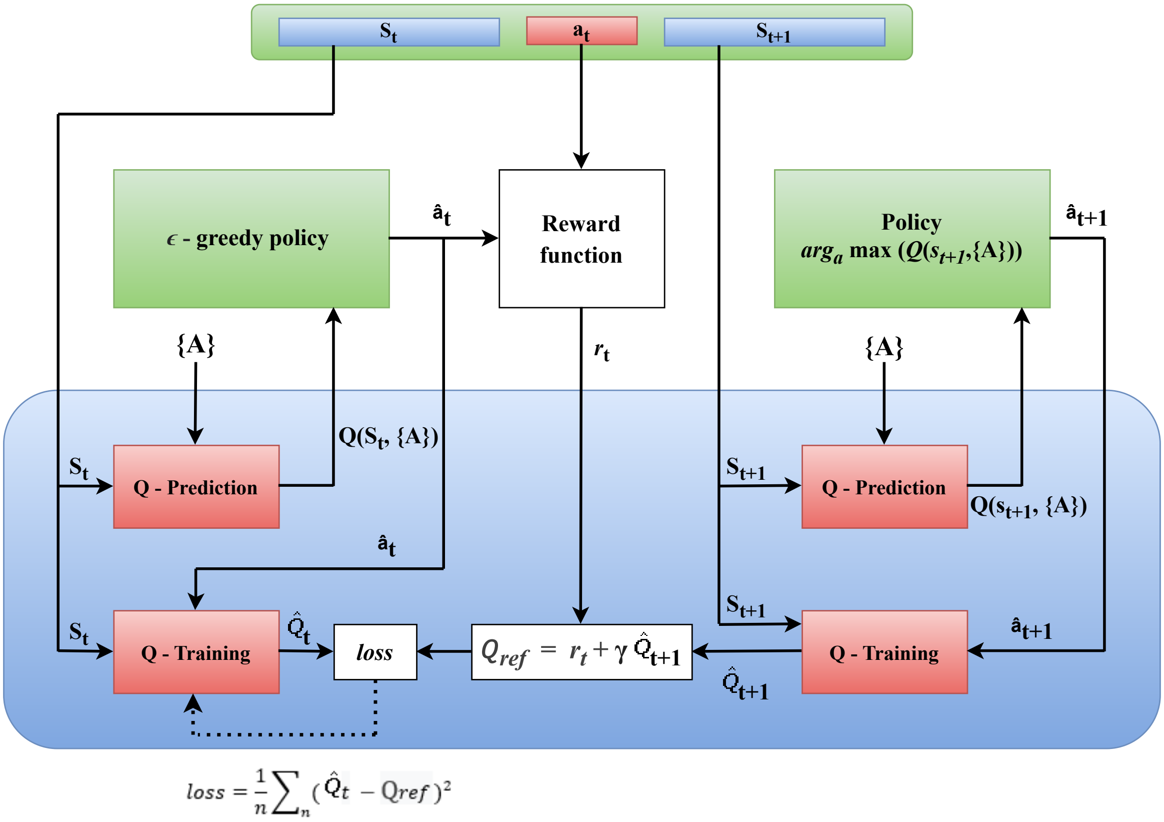

FIGURE 2: The training strategy of the DQN algorithm [30].

the environment. The reward function provides the agent with insights into the desirability of an action within a given state. Over time, the agent learns to make better decisions and avoids suboptimal choices based on the accumulated rewards. The primary objective of RL is to maximize the total accumulated reward and develop a policy that maps states to actions. This cumulative reward represents the aggregation of rewards obtained by the agent over time and is expressed by Eq. 1.

$$R_t = r_{t+1} + \gamma.r_{t+2} + \gamma.r_{t+3} + ... = \sum_{l=0}^{\infty} \gamma^l r_{t+l+l}, \qquad (1)$$

In order to deepen our understanding of the RL model's capacity to determine the optimal policy, it is crucial to explore and comprehend fundamental concepts such as exploration and exploitation. Exploration involves evaluating and investigating various predefined actions to identify the most suitable course of actions for the upcoming states, while exploitation focuses on utilizing current knowledge to adjust the action selection policy and maximize overall rewards. These concepts are mathematically examined through the ϵ-greedy policy. At each state, the agent has the capability to explore an action with an exploration rate ϵ from a predefined set of actions randomly or exploit a specific action with the maximum Q value using an exploitation rate of $1 - \epsilon$. The exploration rate, $\epsilon \in$ [0, 1], is initially set to its maximum value of 1, and gradually decreases as the learning process progresses. Ultimately, as the training model evolves, the agent relies solely on the exploitation mechanism and its accumulated knowledge to determine the optimal action to execute. The concept of Q-learning algorithm, introduced in [55], empowers the agent to learn and make optimal decisions through sequential exploration of different actions. The goal of this approach is to maximize the overall accumulated reward by leveraging the Bellman equation, as expressed in Eq. 2.

$$\begin{aligned} Q^{new}(s_t, a_t) &\leftarrow (1-\alpha)Q(s_t, a_t) \\ &+\alpha\left(r_t + \gamma \max_a Q(s_{t+1}, a_{t+1})\right), \qquad (2) \end{aligned}$$

Where $\alpha \in [0, 1]$, the learning rate plays a crucial role in determining the degree to which the updated Q-value supersedes the previous Q-value. When α equals 0, the agent relies solely on prior knowledge, disregarding new information gained from recent interactions. Conversely, when α equals 1, the agent discards the previously acquired knowledge and focuses on exploring available actions to acquire new insights.

The Q-learning algorithm utilizes a Q function to calculate the Q values for a given state, with the primary objective of maximizing rewards. To store the Q-values for state-action pairs, a Q-table is employed, where rows represent states and columns represent available actions. However, as the number of states and actions increases, the state-action space grows

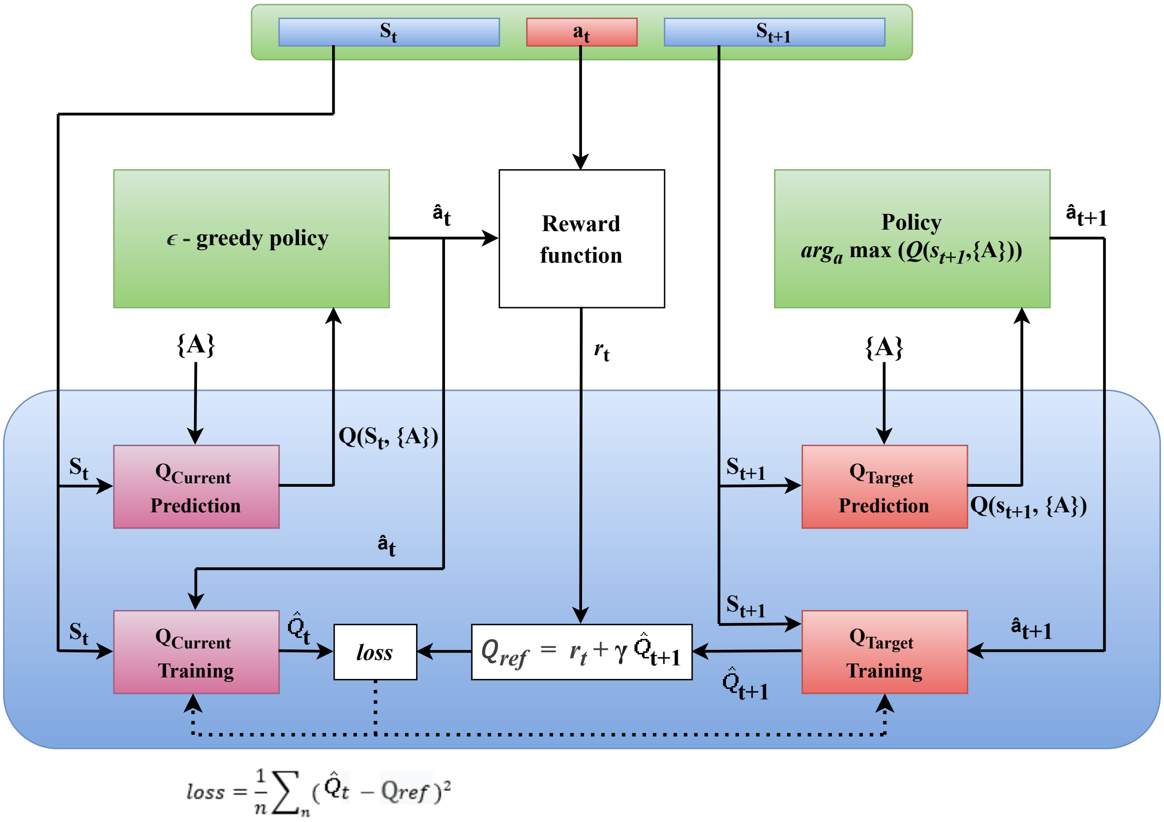

FIGURE 3: The training strategy of the DDQN algorithm [30].

exponentially, making the use of a Q-table impractical. In order to address this challenge, DL plays a crucial role in the integration with RL, leading to the development of the deep Q network (DQN). This integration leverages the remarkable DL capabilities of DQN to overcome the exponential growth in the state-action space. Thanks to DL, the DQN enables efficient and effective representation of Q-values without the need for an exhaustive Q-table.

B. DEEP Q NETWORK (DQN)

The Q-learning algorithm utilizes a Q function to calculate the Q value for a given state s_t and action a_t, which helps in formulating the policy for action selection. The optimal policy π^* is determined by selecting the action with the maximum Q value for each state-action pair. In the training of DQN model, individual samples are presented in a triple format (s_t, a_t, s_{t+1}), where s_t, a_t, and s_{t+1} represent the current state, the true label, and the next state, respectively as shown in FIGURE 2. During the training process, the reward value r_t for a particular state s_t depends on whether the predicted label $\hat{a}_t$ matches the ground truth label a_t. A reward of one is earned if the prediction is correct, while a reward of zero is given if it is incorrect, as illustrated in Eq. 3.

$$r_t = \begin{cases} 1 & \hat{a}_t = a_t. \\ 0 & \hat{a}_t \neq a_t. \end{cases} \qquad (3)$$

Furthermore, in the prediction phase of the DQN model illustrated in FIGURE 2, a set of Q-value combinations $Q(s_t, A) = [Q(s_t, a_0), Q(s_t, a_1), ..Q(s_t, a_b)]$ is computed for each given current state s_t, taking into account the available set of labels A. Here, b represents the total number of available labels. Subsequently, a selection process using ϵ-greedy algorithm is performed to determine an action from the computed set of combinations. This selection process employs either the exploitation concept with a probability of ϵ or the exploration concept with a probability of $1-\epsilon$. Likewise, the Q-value of the next state s_{t+1} is computed using the $arg_a\ max(.)$ policy, where $\hat{Q}_{t+1} = max_a Q(s_{t+1}, A)$. Upon successful completion of the training phase of the DQN model, the proposed DQN model is utilized for action prediction by selecting the action associated with the highest value of the Q function.

C. DOUBLE DEEP Q NETWORK (DDQN)

The double deep Q network (DDQN) is another variant of the DQN that shares similar fundamental structure with it. However, there is a notable difference in the way the next state prediction process is performed. DDQN incorporates two deep neural networks, namely the current network and the target network [30], [56]. The former predicts the Q function of the current state $\hat{Q}_t$, while the latter predicts the Q function of the next state $\hat{Q}_{t+1}$. Despite having the same architecture, the target network is updated with a time delay

Algorithm 1: Data preprocessing algorithm

```
Input: SMs' benign readings of $C = 130$ randomly
    selected customers from the Irish dataset, number of
    benign readings slots per customer $T = 48$, and the
    reduction factor $\beta \, \epsilon \, [0, 0.4]$.
Output: Two subsets includes benign and malicious
    readings samples.
Initiate the attack function $f(x_i(t)) = \beta x_i(t)$ to
    generate the malicious samples.
for $i = 0, 1, 2, ..., C$ do
    for each benign reading $x$ at time slot $t$ in $T$. do
        Input the benign reading $x_i(t)$ in the attack
            function to obtain the malicious reading of the
            sample.
        Concatenate the benign and malicious samples.
    end for
    Repeat until getting to epoch $C$.
end for
Divide the concatenated samples into two subsets, each
    consisting of benign and malicious reading samples.
```

synchronization to prevent the moving target effect during gradient descent calculation over $(\hat{Q}_t - Q_{ref})^2$. The target network's parameters are updated regularly with those of the current network. Apart from this difference, the training and prediction phases of the DDQN model are similar to those of the DQN model. Additionally, After the training phase of the DDQN model has been successfully concluded, the model is then employed for action prediction by selecting the action associated with the maximum value within the Q function. The typical architecture and training of DDQN scheme is shown in FIGURE 3.

IV. DATASET PREPARATION

In this section, we describe the procedure for preparing datasets that consist of real-time electricity consumption readings obtained from consumers' SMs. These datasets are utilized to train, evaluate, and analyze machine learning-based models for electricity theft detection, as well as for constructing adversarial attack models.

A. BENIGN SAMPLES

In this research paper, we employ a real electricity consumption dataset obtained from the Irish Smart Energy Trials [48] to create two distinct datasets. The first dataset is used for training and evaluating the performance of electricity theft detectors, while the second dataset is utilized to construct an attack model for generating adversarial evasion samples. The Irish dataset, publicly available since January 2012 by the Electric Ireland and Sustainable Energy, consists of half-hourly electricity consumption readings from $3,639$ residential consumers. The dataset spans a duration of 536 days, covering readings collected between 2009 and 2010. For our study, we randomly select readings from the smart meters of 130 customers, resulting in a total of $69,680$ benign samples. The electricity theft detection process focuses on a one-day period, where the detector analyzes a set of electricity consumption readings (48 readings) for a specific consumer to determine if electricity theft is occurring. Our attack model is designed to generate 48 adversarial evasion samples, mimicking the characteristics of real consumption patterns.

B. MALICIOUS SAMPLES

To ensure the effectiveness of an electricity theft detector in accurately distinguishing between benign and malicious electricity consumption samples, it is crucial to train the detector using both types of samples. However, in the case of the Irish dataset, only benign samples are available, and there is a lack of publicly available malicious samples. To address this limitation, we adopt an electricity theft attack methodology proposed in a previous study [13]. This attack involves generating malicious samples by modifying the benign samples. The attack model is represented as a function, denoted as $f(x_i(t)) = \beta x_i(t)$, where $x_i(t)$ represents the true value of electricity consumption readings for a specific consumer i at a given time step t. The function $f(x_i(t))$ reduces the electricity consumption value by a random factor β, where $0 < \beta < 1$. The objective of this attack is to decrease the true consumption value by a random reduction factor β, simulating malicious behavior in the samples.

C. DATASET PREPROCESSING

To generate malicious samples, we need to determine the parameter β for the attack function mentioned earlier. This parameter follows a uniform distribution ranging from 0.1 to 0.4 within the attack function $f(.)$. The benign samples are then subjected to this attack function, resulting in a malicious dataset. Each customer contributes 536 daily samples to the malicious dataset, combining both benign and malicious samples. This leads to a total of $1,072$ samples. Considering the 130 customers over 536 days, the overall dataset size is $139,360$ samples. The dataset is divided into two subsets, each containing 69,680 samples. The first subset is used for training and evaluating the performance of the electricity theft detectors, while the second subset is used to construct the attack model for generating adversarial evasion samples. Both subsets are further divided into training and testing subsets in a 2:1 ratio. The training subset consists of $46,453$ samples, and the testing subset contains $23,227$ samples. The entire dataset preprocessing is illustrated and annotated through steps (1 to 4) within the proposed framework, as depicted in FIGURE 4. Additionally, Algorithm 1 provides a detailed explanation of the preprocessing steps.

V. ATTACK AND DEFENSE MODELS

In this section, we first discuss evasion attacks that are used to attack electricity theft detectors, and then we discuss a countermeasure.

A. ATTACK MODEL

In this section, we introduce two attack models designed to generate adversarial evasion samples using malicious elec-

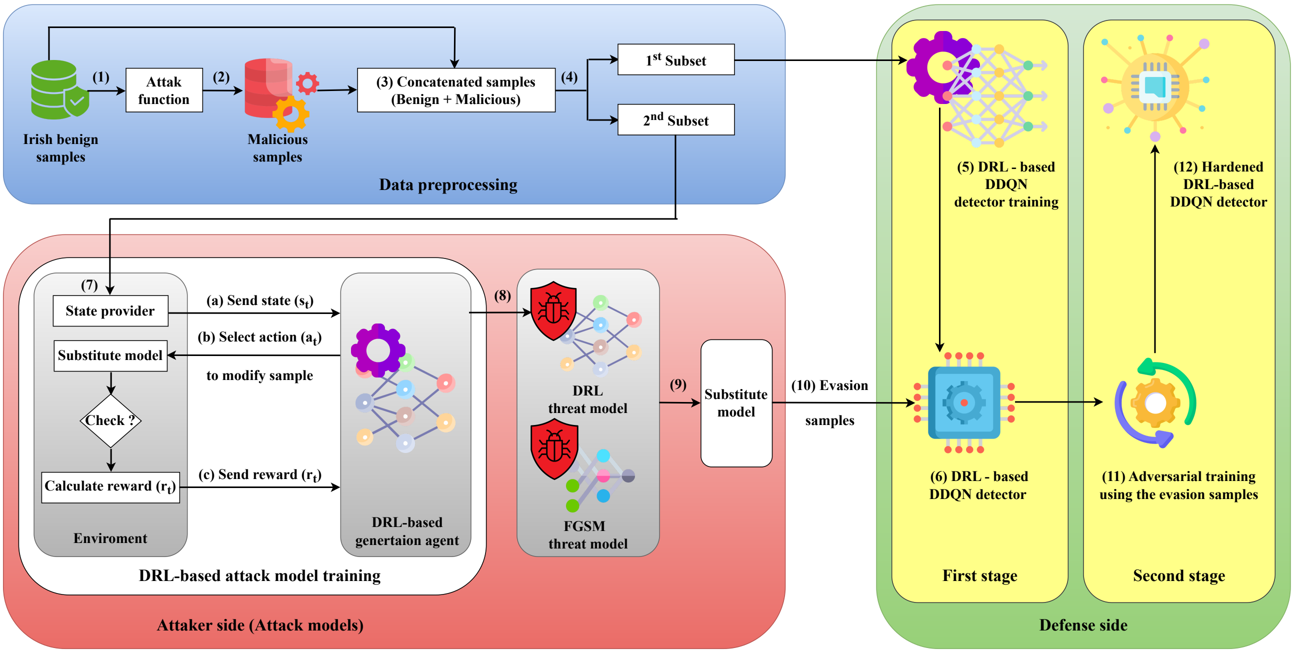

FIGURE 4: The proposed attack and defense models framework.

tricity consumption readings. The first model is the DRL-based DDQN model, and the second model is the FGSM-based model. TABLE 1 presents a comprehensive comparison between the DRL-based and FGSM-based evasion models, shedding light on their differences, advantages, and limitations.

1) Overview

Previously, electricity theft attackers resorted to use simple attack functions, as proposed in [13], to engage in electricity theft. These approaches involved manipulating and reporting fraudulent electricity consumption values to the utility. While these approaches were once effective in facilitating successful electricity theft and inflicting financial harm on the utility, recent advancements in detection technology have rendered them detectable by electricity theft detectors. However, attackers have actively been exploring new techniques to evade these detection systems. In this study, we propose an evasion attack model that empowers attackers to steal electricity by generating artificially low consumption readings. This evasion attack model operates under the assumption of a black-box attack scenario, where the attacker has no knowledge about the RL-based electricity theft detector employed by the utility. Additionally, the attack model and the electricity theft detector employ different neural network architectures and are trained on different datasets. Specifically, the RL-based electricity theft detector is trained on a combination of benign and malicious reading samples from the first subset of data presented in Section IV-B. On the other side, the attacker utilizes the malicious samples from the second subset of data, as outlined in Section IV-B, to generate low-consumption adversarial evasion samples. The objective of these evasion samples is to deceive the electricity theft detector and be classified as benign, thereby stealing electricity while evading detection.

2) DRL-based attack model

The proposed approach employs DRL to develop a generation agent that can autonomously generate adversarial evasion samples to bypass the electricity theft detector. The attacker creates a substitute model to verify if the generated evasion samples can avoid detection. The attacker assumes that the generated samples that evade the substitute model is also able to pass the electricity theft detector of the utility.

To train the generation agent, the environment provides states to the agent through a sample provider that extracts malicious samples from the second subset of data. The agent employs a trial-and-error approach, leveraging the aforementioned DDQN algorithm and its mechanism for selecting the optimal action, as explained earlier in Section III-C. This action space consists of perturbation values that the generation agent can apply to a malicious sample through a multiplication process, thereby modifying it and generating an adversarial evasion sample. Subsequently, the generated evasion sample undergoes testing to determine its capability to evade the substitute model. The substitute model is trained on the second subset of the dataset, containing both benign and malicious samples. The agent's reward is contingent upon the output of the substitute model when evaluating the generated sample. If the generated evasion sample successfully evades the substitute model, the generation agent receives a reward of 1; otherwise, the reward is 0.

In this setting, the substitute model is implemented using the DRL-based DDQN algorithm, employing three distinct

TABLE 1: Comparison between the DRL-based and FGSM-based evasion attack models.

Model	DRL-based evasion attack model	FGSM-based evasion attack model
Differences	**Methodology:** DRL involves training an agent that interacts with the environment and learns from feedback (rewards) to maximize cumulative rewards, and updates its policy to make optimal decisions. **Model Complexity:** DRL models often involve complex neural network architectures, such as Deep Q Networks (DQN), to capture intricate patterns in data and make informed decisions.	**Methodology:** FGSM is a gradient-based method that perturbs input data using gradients of the loss function with respect to the input features. **Model Interaction:** FGSM is typically applied to a pre-trained model without active interaction with an environment.
Advantages	**Adaptability:** DRL is highly adaptable to varying and dynamic environments. This adaptability is crucial for addressing various attack scenarios. **Learning:** DRL agents learn from the consequences of their actions, allowing them to adjust their strategies over time based on accumulated experience. **Versatility:** DRL can be applied to a wide range of tasks beyond evasion attacks due to its reinforcement learning foundation.	**Simplicity:** FGSM is relatively simple to implement, requiring minimal modification to an existing model. **Speed:** The generation of evasion samples using FGSM is usually faster compared to training DRL-based agents. **Interpretability:** The perturbations added by FGSM can provide insights into how input features impact model predictions.
Limitations	**Training Complexity:** DRL models require extensive training, which can be computationally intensive and time-consuming. **Resource-Intensive:** The training process often demands significant computational resources and a substantial amount of data. **Data Requirements:** Effective DRL training often requires large amounts of high-quality data, which might be challenging to obtain for some attack scenarios.	**Lack of Adaptability:** FGSM's static nature makes it less adaptable to dynamic and evolving attack scenarios. **Limited Exploration:** FGSM might generate evasion samples within a limited space around the original data point, restricting the diversity of strategies. **Insensitive to Sequential Effects:** FGSM doesn't consider the sequential nature of interactions and decision-making dynamically. **Transferability Concerns:** FGSM-generated samples might not transfer well to different models or settings, limiting the attack's broader applicability.

neural network architectures: CNN, GRU, and FFNN. On the other hand, the generation agent of the attack model is implemented using the CNN architecture within the DRL-based DDQN algorithm. The DRL-based DDQN attack model is visually depicted and annotated in steps (7 [a, b, c], 8) in FIGURE 4. The training phase and the training accuracy of the DRL-based DDQN attack model are both outlined in Algorithm 2 and FIGURE 5, respectively. This figure visually represents the training convergence as accuracy improves with the progressive increase in the number of training batches.

3) FGSM-based attack model

Additionally, alongside the DRL-based DDQN attack model, we leverage an FGSM-based attack model to generate adversarial evasion samples. The FGSM technique is widely employed in the literature to create such samples by introducing carefully calculated perturbations to the input samples, aiming to cause misclassification of these samples [52], [57]–[59]. FGSM presents several distinct advantages. It operates by employing gradients of the loss function, often yielding impactful perturbations that lead to misclassification. Furthermore, it can achieve misclassification with minimal modifications to the input samples. Additionally, FGSM exhibits a robust transferability property. To ensure that these added perturbations remain undetectable to the electricity theft detector, their magnitude must be within an acceptable limit. Thus, mathematically, these added perturbations can be described as follows:

$$\begin{array}{ll} \min_{\delta\vec{x}} & |\delta\vec{x}| \\ \text{s.t.} & \hat{f}(\vec{x}+\delta\vec{x}) \neq \hat{f}(\vec{x}) \end{array} \tag{4}$$

Where $\delta\vec{x}$ represents the adversarial perturbation applied to the input malicious sample $\vec{x}$, the FGSM method utilizes the sign of the gradient of the cost function $C_{\hat{f}}(\theta, x, y)$ with respect to the model $\hat{f}$. This gradient is evaluated for the input malicious sample, and its sign is employed to generate the adversarial perturbations described in Eq. 5. The objective is to maximize the value of the cost function to the greatest extent.

$$\delta_{\vec{x}} = \lambda\, \text{sign}\left(\nabla_{\vec{x}} C_{\hat{f}}(\theta, \vec{x}, y)\right), \tag{5}$$

Here, θ, $\vec{x}$, and y represent the model weights, the input malicious sample and the true label corresponding to $\vec{x}$, respectively. Meanwhile, λ is the parameter that is tuned so that the label produced by the model for the perturbed input data, i.e., $(\vec{x}+\delta\vec{x})$ changes from the malicious label to benign label and deceive the detector.

B. DEFENSE MODEL

This section focuses on defense mechanisms against the attacks discussed earlier. These mechanisms are divided into two stages.

In the first stage, a DRL-based DDQN detector is employed. The detector is constructed using various neural network architectures, including FFNN, CNN, and GRU. It is trained using the initial subset of the dataset discussed in Section IV-C, allowing it to learn and identify patterns indicative of electricity theft cyber-attacks. It serves as a defense mechanism against evasion attacks, considering that the attacker might exploit adversarial evasion samples generated by either DRL-based or FGSM-based attack models to launch evasion attacks against the detector. The first stage of defense is described by steps 5 and 6 in FIGURE 4.

Algorithm 2: DRL-based DDQN attack model training algorithm

```
Input: Exploration rate $\epsilon$, learning rate $\alpha$, discount factor $\gamma$,
    batch size $H$, and training epochs $G$.
Output: The optimal action $a^*$. and the adversarial evasion
    sample.
Initiate the action value function $Q(s, a)$ arbitrarily.
Initiate the state $s$ using state generator in recognizable
    format by the agent.
for $i = 0, 1, 2, ..., G$ do
    for each state $s$ in $i$. do
        Input the state $s_t$ and the actions set $A$ in the
            current network in order to predict $Q(s, A)$ for all
            actions.
        Use the $\epsilon$-greedy policy to select the action $\hat{a}_t$.
        Given $s_t$ and $\hat{a}_t$, obtain $Q(s_t, \hat{a}_t)$.
        Generate the adversarial evasion sample through
            multiplying the state sample $s_t$ with $\hat{a}_t$.
        Check whether the generated adversarial evasion
            sample can evade the substitute model.
        Calculate the reward $r_t$.
        Input the next state $s_{t+1}$ and the actions set $A$ in
            the target network in order to predict $Q(s_{t+1}, A)$
            for all actions.
        Use $\arg\max_a Q(s_{t+1}, A)$ policy to select $\hat{a}_{t+1}$.
        Given $s_{t+1}$ and $\hat{a}_{t+1}$, obtain $\hat{Q}_{t+1}(s_{t+1}, \hat{a}_{t+1})$.
        Using $\hat{Q}_{t+1}$, $r_t$, and $\gamma$, Obtain $Q_{ref}$.
        Calculate the loss function.
        Update the Q-value $Q(s_t, a_t)$.
        Repeat until $s_{t+1}$ is terminal.
    end for
    Repeat until getting to epoch $G$.
end for
Compute the optimal policy $\pi^*$ and optimal action $a^*$.
Execute the optimal action $a_t^*$ at current time slot $t$ and
    get the adversarial evasion sample.
```

Algorithm 3: Defense model algorithm

```
Input: The first subset of the dataset, and the adversarial
    evasion samples.
Output: The hardened DRL-based DDQN detector.
    The first stage of defense
Use the first subset of the dataset to train the DRL-based
    DDQN model using different architectures of neural
    network.
Obtain the trained DRL-based DDQN detector.
Utilize the generated adversarial evasion samples to
    attack the proposed detector.
    The second stage of defense
Record the evasion samples.
Use these recorded evasion samples to conduct the
    adversarial training process for the proposed detector.
Obtain the hardened DRL-based DDQN detector.
```

In the second stage of defense, evasion samples play a crucial role in the subsequent process known as adversarial training [60], [61]. This training significantly boosts a model's resilience against evasion attacks by enhancing its ability to detect alterations in input data, resulting in improved differentiation between benign and malicious samples. This heightened sensitivity empowers models to make accurate judgments even when facing adversarial variations, thereby reducing the effectiveness of evasion attacks. The primary goal of this training process is to 'harden' and reinforce the detector's defenses, making it more robust and resilient against future attacks. Leveraging the recorded evasion samples, the detector undergoes retraining, reinforcing its ability to identify and respond effectively to adversarial evasion attempts. Within the context of reinforcement learning (RL), adversarial training drives the RL agent to explore and adapt its ploicy to accommodate unexpected consumption patterns or attack tactics. Through intentional exposure to adversarial influences during training, the agent acquires the ability to make decisions that extend beyond routine situations, displaying resilience against variations and disruptions. Consequently, the agent enhances its capability to consider a wide range of scenarios and actions, enabling well-informed decisions even when confronting perturbed or adversarial observations.

The parameters guiding the adversarial training of the DRL-based DDQN model are outlined in TABLE 2. To gain a comprehensive understanding of the defense procedures, this defense is visually represented by steps 11 and 12 in FIGURE 4. These steps significantly contribute to safeguarding against adversarial evasion samples and enhancing the detector's capabilities. Additionally, Algorithm 3 provides a detailed explanation of the procedures of this defense, offering insights into their implementation and functionality.

VI. EVALUATIONS

In this section, we first discuss the experimental setup and the evaluation metrics used to assess the performance and effectiveness of our proposals. Subsequently, we present the experimental results of four conducted experiments to evaluate the severity of the attack models and the effectiveness of the defense models. In the first experiment, we train a global DRL detector that utilizes DDQN to detect electricity theft cyber-attacks. This detector, which serves as the first stage of defense, is constructed using diverse neural network architectures, including FFNN, CNN, and GRU. In the second experiment, we train DRL-based DDQN and FGSM-based attack models to generate adversarial evasion samples and evaluate their effectiveness in attacking the global electricity theft detector. Meanwhile, the third experiment focuses on evaluating the effectiveness of adversarial training in strengthening the DRL-based DDQN detector against both DRL and FGSM-based adversarial evasion samples. This hardened detector serves as the second stage of defense,

TABLE 2: Parameters of DRL-DDQN based attack and defense models.

Parameter	Value
Exploration rate (ϵ)	0.6
Learning rate (α)	0.00001
Discount factor (γ)	0.001
Batch size (H)	128
No. of training epochs (G)	10

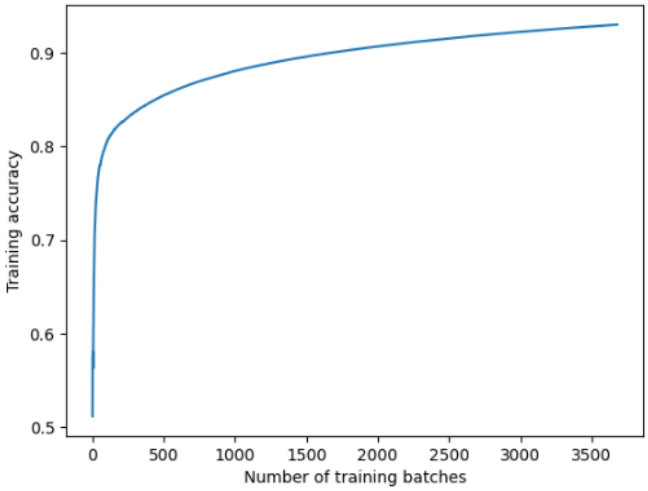

FIGURE 5: Training accuracy of DRL-based attack model.

TABLE 3: The hyperparameters for the neural network architectures utilized in the proposed models.

Architecture	Parameters		
	Layer	Number of units	AF
FFNN	Input	48	Linear
	Dense	512	Relu
	Dense	700	Relu
	Dense	850	Relu
	Dense	1024	Relu
	Dense	512	Relu
	Dense	256	Relu
	Dense	200	Relu
	Dense	50	Relu
	Output	2	Softmax
CNN	Input	48	Linear
	Conv1D	32	Relu
	Conv1D	64	Relu
	Conv1D	128	Relu
	Dense	64	Relu
	Dense	128	Relu
	Dense	256	Relu
	Dense	256	Relu
	Dense	512	Relu
	Dense	2	Softmax
GRU	Input	48	Linear
	GRU	64	Sigmoid
	GRU	64	Tanh
	Dense	64	Relu
	Dense	128	Relu
	Dense	2	Softmax

aiming to enhance the resilience against potential attacks. Lastly, in the fourth experiment, we investigate whether the DRL evasion samples-based defense model can defend against FGSM-based adversarial evasion samples, and vice versa.

The configuration parameters of the proposed DRL-DDQN based attack and defense models are specified and listed in TABLE 2. Additionally, the hyperparameters of the neural network architectures utilized in these models can be found in TABLE 3. These parameters have been fine-tuned through iterative experimentation. For our experiments, we utilized various Python 3 libraries, such as Scikit-learn, Pandas, Keras, Numpy, TensorFlow, and Matplotlib. It is worth mentioning that all experiments were conducted on the Google Colab platform, which offers the convenience of writing and executing Python code directly within a browser.

A. METRICS

The evaluations of the proposed attack and defense models include the analysis of multiple metrics, such as accuracy, precision, recall, false alarm, false negative rate, highest difference, F-1 score, evasion rate (EVR), attack success rate (ASR), and transfer-ability rate (TR). These metrics rely on the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP and TN represent correctly classified malicious and benign samples, respectively, while FP and FN denote misclassified malicious and benign samples, respectively. The computation of these evaluation metrics is as follows:

1) Accuracy (ACC)

It represents the proportion of correctly classified test samples by the detector out of the total number of samples in the test dataset, *which includes both benign and malicious samples.* Mathematically, it is computed using the following equation:

$$ACC(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100. \tag{6}$$

2) Adversarial accuracy (ACC_{adv})

It represents the proportion of correctly classified test samples by the detector out of the total number of samples in the test dataset, *which includes both benign and evasion samples.*

TABLE 4: Comparison between the performance of the different architectures of DRL-based DDQN detectors.

Metrics	DRL-based DDQN detector		
	FFNN	CNN	GRU
Accuracy (%)	85.739	93.302	87.666
Precision (%)	85.971	93.341	87.870
Recall (%)	85.739	93.302	87.666
FA (%)	17.68	5.50	15.46
HD (%)	68.06	87.80	72.21
F1 (%)	85.855	93.322	87.768

3) Overall accuracy (ACC_{all})

It represents the proportion of correctly classified test samples by the detector out of the total number of samples in the test dataset, *which includes benign, malicious, and evasion samples.*

4) Precision

It represents the proportion of true positive samples to the total number of samples classified as positive by the detector. Mathematically, it is computed using the following equation:

$$Precision(\%) = \frac{TP}{TP + FP} \times 100. \tag{7}$$

5) Recall

It represents the proportion of correctly identified positive samples to the total number of positive samples in the test dataset. Mathematically, it is computed using the following equation:

$$Recall(\%) = \frac{TP}{TP + FN} \times 100. \tag{8}$$

6) False alarm (FA)

It represents the proportion of false positive samples to the total number of negative samples in the test dataset. Mathematically, it is computed using the following equation:

$$FA(\%) = \frac{FP}{FP + TN} \times 100. \tag{9}$$

7) Highest difference (HD)

It is the difference between recall and false alarm (FA). Mathematically, it is computed using the following equation:

$$HD(\%) = Recall(\%) - FA(\%). \tag{10}$$

8) F-1 score (F1)

It is the harmonic mean between precision and recall. Mathematically, it is computed using the following equation:

$$F1(\%) = \frac{2 * Precision * Recall}{Precision + Recall} \times 100. \tag{11}$$

9) Evasion rate (EVR)

It represents the proportion of evasion samples that are misclassified as benign by *the substitute model*.

10) Attack success rate (ASR)

It represents the proportion of evasion samples that are misclassified as benign by *the utility detector*. Note, the attacker sends only the evasion samples that passes the substitute model to the utility, but the ASR metric is computed over all the evasion samples.

11) Transferability rate (TR)

It quantifies the proportion of evasion samples that successfully bypass both the substitute model and utility detector compared to the total number of evasion samples that only manage to bypass the substitute model. This metric, which is derived from the previous two metrics, provides an indication of the probability that a given sample, which evades the substitute model, will also successfully evade the utility detector. It is computed using the following equation:

$$TR(\%) = \frac{ASR}{EVR} \times 100 \tag{12}$$

B. EXPERIMENT #1

In this experiment, we focused on training DRL-based DDQN global electricity theft detectors using the first subset of electricity consumption readings. Our approach involved creating separate training and testing datasets by processing the electricity consumption readings, as explained in Section IV-C. The training dataset was then utilized to train three global DRL-based DDQN detectors, representing the first stage of defense stage. These detectors employed different neural network architectures, including FFNN, CNN, and GRU. The selection of these DL-based architectures was based on their demonstrated superior performance compared to shallow architectures and their widespread use in the literature. Subsequently, we evaluated the performance of the three DRL-based DDQN detectors using the testing dataset. Table 4 provides a comprehensive comparison of the detectors in terms of key metrics such as $ACC, Precision, Recall, FA, HD,$ and $F1$. From the table, it is evident that the FFNN detector exhibited the lowest performance, which can be attributed to its simpler architecture compared to CNN and GRU. The CNN detector showcased high performance due to its convolutional layers, enabling the extraction of important features from the electricity consumption readings for more accurate detection. Similarly, the GRU detector demonstrated good performance by effectively capturing temporal correlations within the electricity consumption readings in each input sample.

C. EXPERIMENT #2

In this experiment, we conducted training for DRL-based DDQN and FGSM-based attack models to generate adversarial evasion attack samples, as explained in detail in Section V-A. During this experiment, we considered a practical black box attack scenario, where the attackers lack any knowledge about the detector used by the utility. Furthermore, there could be variations in the neural network architectures

TABLE 5: Comparison between the performance of DRL-based DDQN and FGSM-based attack models against the detector.

Model architecture		DRL-based DDQN attack model				FGSM-based attack model			
DRL-based substitute model	DRL-based detector model	EVR (%)	ASR (%)	TR (%)	ACC (%)	EVR (%)	ASR (%)	TR (%)	ACC (%)
FFNN	FFNN	99.982	99.965	99.982	41.210	78.722	69.451	88.223	47.066
	CNN		99.704	99.721	46.563		48.479	61.582	65.585
	GRU		92.929	92.945	46.055		68.240	86.685	49.205
CNN	FFNN	99.982	99.965	99.982	41.210	94.214	60.747	64.477	58.971
	CNN		99.817	99.834	46.507		78.173	82.974	54.914
	GRU		95.364	95.381	44.838		69.983	74.280	55.440
GRU	FFNN	97.921	97.904	99.982	41.210	72.911	95.484	81.584	50.485
	CNN		97.886	99.964	46.442		55.563	76.206	58.358
	GRU		96.599	98.650	44.073		69.582	95.434	44.898

TABLE 6: Comparison between the defense performance of the hardened detector against DRL-based DDQN and FGSM-based evasion samples.

Model architecture		Adversarially trained hardened detector against DRL-based DDQN evasion samples				Adversarially trained hardened detector against FGSM-based evasion samples			
DRL-based substitute model	DRL-based detector model	ASR (%)	ACC (%)	ACC$_{adv}$(%)	ACC$_{all}$(%)	ASR (%)	ACC (%)	ACC$_{adv}$(%)	ACC$_{all}$(%)
FFNN	FFNN	2.197	85.048	92.310	88.853	8.330	84.679	84.814	84.686
	CNN	2.354	93.103	96.644	94.892	0.671	92.764	94.547	93.593
	GRU	9.207	85.865	95.724	90.991	1.679	86.784	92.086	89.300
CNN	FFNN	1.807	85.257	92.703	89.166	14.401	84.030	79.976	81.998
	CNN	1.885	93.263	97.004	95.175	1.681	89.113	89.969	89.387
	GRU	6.521	85.808	95.896	91.011	2.942	88.363	91.323	89.787
GRU	FFNN	2.640	85.509	93.565	89.846	14.120	84.550	82.556	83.575
	CNN	2.667	93.199	96.712	95.130	3.475	88.977	96.503	92.664
	GRU	4.027	85.718	95.960	91.122	0.868	87.183	92.608	89.844

employed by the attack model and the electricity theft detector, as well as differences in the training datasets used. Additionally, the attacker did not have direct access to the utility's detector. Therefore, the attacker attempted to create a substitute model to evaluate whether the generated evasion samples could successfully evade detection. The attacker's assumption was that if the evasion samples were capable of deceiving the substitute model, they would likely also be able to bypass the actual electricity theft detector implemented by the utility. The parameter configurations of the substitute and attack models are provided in the TABLE 2, and TABLE 3.

Once the attack models were trained, we utilized the test malicious samples from the second subset, as outlined in Section IV-B, to generate the evasion attack samples. These generated samples were subsequently employed to target the detectors developed in Experiment #1. The outcomes of this experiment, evaluated based on EVR, ASR, TR, and ACC, are presented in TABLE 5. These outcomes demonstrate the superior performance of the proposed DRL-based DDQN attack model compared to the FGSM-based attack model in terms of EVR, ASR, and ACC. The DRL-based DDQN attack model achieves outstanding success, with EVR ranging from 97.921% to 99.982%, ASR ranging from 92.929% to 99.965%, and ACC ranging from 41.210% to 46.563%. This indicates a significant reduction of 41.611% to 46.795% within 95% confidence interval (CI) of (44.66 ± 1.14)% compared to the corresponding values in TABLE 4. In the same context, the FGSM-based attack model exhibits EVR values between 72.911% and 94.214%, and ASR values ranging from 48.479% to 95.484%, and ACC ranging from 44.898% to 65.585%. This indicates a significant reduction of 26.768% to 42.768% within 95%CI of (35.022 ± 3.27)% compared to the corresponding values in TABLE 4. These outcomes highlight the severity of the evasion attacks and the effectiveness of our proposed DRL-DDQN based attack model in computing effective evasion samples that can deceive and bypass the electricity theft detector, particularly in the challenging context of a black box attack scenario and variations in neural network architectures employed by the attack model and the electricity theft detector.

D. EXPERIMENT #3

The previous experiment has revealed the detrimental impact of evasion attacks on the performance of electricity theft detectors. Our goal in this experiment, which represents the second stage of defense, is to propose a robust and hardened detector capable of maintaining a consistent detection performance against evasion attacks. To accomplish this, we

TABLE 7: The defense performance of the adversarially trained hardened detector using DRL-based evasion samples against FGSM-based evasion samples.

Model architecture			Adversarially trained hardened detector using DRL evasion samples	Adversarially trained hardened detector using DRL and FGSM-based evasion samples			
attack model	DRL-based substitute model	DRL-based detector model	ASR (%)	ASR (%)	ACC (%)	ACC$_{adv}$(%)	ACC$_{all}$(%)
FGSM	FFNN	FFNN	91.568	7.222	85.287	86.917	86.240
		CNN	62.008	0.604	93.833	96.328	95.161
		GRU	84.682	10.715	84.774	84.801	84.751

TABLE 8: The defense performance of the adversarially trained hardened detector using FGSM-based evasion samples against DRL-based evasion samples.

Model architecture			Adversarially trained hardened detector using FGSM evasion samples	Adversarially trained hardened detector using DRL and FGSM-based evasion samples			
attack model	DRL-based substitute model	DRL-based detector model	ASR (%)	ASR (%)	ACC (%)	ACC$_{adv}$(%)	ACC$_{all}$(%)
DRL	FFNN	FFNN	97.750	3.832	86.005	93.947	90.167
		CNN	97.462	2.197	93.077	95.526	94.308
		GRU	97.750	2.579	88.123	95.032	91.785

employ an adversarial training process for the electricity theft detector obtained in the experiment #1, leveraging the captured evasion samples from the first defense stage. The procedures of this experiment are visually illustrated through annotated steps 11 and 12 in FIGURE 4 and detailed in Algorithm 3. The results of this experiment, evaluated based on ASR, ACC, ACC_{adv}, and ACC_{all} are presented in TABLE 6.

These results validate the effectiveness of our proposed defense mechanism, achieved through the process of adversarial training. This training enables us to enhance the resilience and robustness of the electricity theft detection model against evasion attacks. Specifically, when considering DRL-based evasion samples and comparing to the corresponding values in TABLE 5, which represent the detector's performance without adversarial training, we observe a substantial decrease in ASR, ranging from 1.80% to 9.20%, indicating a significant reduction of 83.72% to 98.16% within $95\%CI$ of (94.09 ± 3.05)%. Additionally, there is a significant increase in ACC, ranging from 85.05% to 93.26%, representing an improvement of 39.81% to 46.75% within $95\%CI$ of (34.85 ± 1.6)%. Furthermore, notable enhancements are observed in ACC_{adv} and ACC_{all}, with values ranging from 92.31% to 97% and 88.85% to 95.18%, respectively. Similarly, for FGSM-based evasion samples, comparing the results to the corresponding values in TABLE 5, we find a decrease in ASR in the range of 0.67% to 14.4%. This showcases a reduction trend of 46.34% to 76.49% within $95\%CI$ of (63.05 ± 7.58)%. Moreover, there is a significant increase in ACC, ranging from 84.03% to 92.76%, indicating an improvement of 25.06% to 42.29% within $95\%CI$ of (33.50 ± 3.31)%. Additionally, significant improvements are observed in ACC_{adv} and ACC_{all}, with values ranging from 79.98% to 96.5% and 81.99% to 93.59%, respectively. The basis for this improvements in the detector's effectiveness lies in the adversarial training, which drives the RL agent to investigate and updates its policy to accommodate unexpected shifts in consumption behaviors or attack methods. This empowers the agent to become proficient at making optimal decisions and broadens its capability to handle diverse scenarios and actions, enabling informed choices even when confronting adversarial perturbations or attacks.

E. EXPERIMENT #4

Following the promising results of the previous experiment, the focus of this study shifts to investigating whether the DRL evasion samples-based defense model can defend against FGSM-based adversarial evasion samples, and vice versa. Our objective is to examine the model's ability to defend against evasion attacks originating from different attack methods. In the first phase of the experiment, we subject the adversarially trained hardened detector, which was initially trained exclusively on DRL-based evasion samples, to attacks using FGSM-based evasion samples. The outcomes, as presented in TABLE 7, reveal the detector's vulnerability, with an ASR ranging from 62.008% to 91.568%. To strengthen the defense mechanism, we proceed to apply adversarial training to the detector using both DRL-based and FGSM-based evasion samples. This comprehensive training approach yields remarkable performance, significantly reducing the ASR to a range of 0.604% to 10.715%. Additionally, notable improvements are observed in key evaluation metrics: ACC, ACC_{adv}, and ACC_{all}, with values ranging from 84.774% to 93.883%, 84.801% to 96.328%, and 84.751% to 95.161%, respectively.

Furthermore, in the second phase of the experiment, we examine the impact of using an adversarially trained hardened detector, initially trained exclusively on FGSM-based evasion samples, to defend against DRL-based evasion samples. The outcomes, presented in TABLE 8, also reveal the detector's vulnerability to the evasion attack, with an ASR ranging from 97.462% to 97.750%. However, when the hardened detector is adversarially trained on both DRL-based and FGSM-based evasion samples, a reduction in ASR is achieved, ranging from 2.197% to 3.832%. Moreover, significant improvements are observed in ACC, ACC_{adv}, and ACC_{all}, with values ranging from 86.005% to 93.007%, 93.947% to 95.526%, and 90.167% to 94.308%, respectively.

VII. CONCLUSION

This paper investigates the vulnerability of RL-based electricity theft detectors to adversarial evasion attacks. We propose a DRL-based DDQN attack model to generate adversarial evasion samples, leveraging the benefits of RL for determining the optimal values through exploration and exploitation mechanisms. By perturbing malicious samples, evasion samples are computed to evade the detectors and classify them as benign. The evasion attack is conducted in a black-box scenario, which is practical and challenging because attackers do not have any knowledge abut the defense model. Our experiments demonstrate the effectiveness of the proposed attacks compared to FGSM-based attacks. The results indicate that our attack model can significantly degrade the detector's performance, achieving an ASR ranging from 92.92% to 99.96%. Additionally, there is a notable decrease in ACC ranging from 41.21% to 46.56%, representing a significant reduction of 39.81% to 46.75%.

To counter evasion attacks, we train a defense model that utilizes adversarial training of a DRL-based detector to obtain a hardened detector. The experimental results showcase the robustness of the defense model against evasion attacks, reducing the ASR by 1.80% to 9.20%, which corresponds to a significant reduction of 83.72% to 98.15%. Moreover, there is a substantial increase in ACC ranging from 85.04% to 93.26%, resulting in an improvement of 39.81% to 46.75%. Finally, we evaluate the ability of the hardened defense model, which is adversarially trained on evasion samples, to defend against evasion attack samples from different attack method. The results suggest that the hardened defense model should be retrained on additional evasion samples originating from different evasion attack methods.

In summary, we'd like to emphasize the significance and benefits of developing a DRL-based defense model to counter fraudulent electricity theft attacks. These attacks, initiated by fraudulent customers manipulating consumption readings in smart power grids, bear practical implications for grid security. This approach yields multifaceted advantages for smart power grids, such as bolstering their security and reliability by detecting and preventing fraudulent actions. By doing so, it safeguards the grid's functionality, diminishing the risk of cascading failures that might disrupt services for legitimate users. Minimizing losses due to fraudulent activities allows for improved resource allocation towards maintenance and upgrades. Mitigating electricity theft enhances customer trust by ensuring fair billing and promoting positive customer relationships. The insights garnered from the RL-based detector offer valuable information on consumption patterns and vulnerabilities, guiding informed decisions for grid management, load monitoring and forecasting, and security enhancements. A dependable and secure power grid environment further stimulates innovation in smart grid technologies. This empowers utility providers to confidently invest in advanced solutions like renewable energy integration, smart metering, and demand response systems, ultimately enhancing operational efficiency. The reduction in losses attributed to electricity theft also contributes to increased revenue for utility companies.

ACKNOWLEDGEMENT

This work was supported by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Project RSPD2023R636.

REFERENCES

[1] M. M. Badr, M. Mahmoud, M. Abdulaal, A. J. Aljohani, F. Alsolami, and A. Balamsh, "A novel evasion attack against global electricity theft detectors and a countermeasure," IEEE Internet of Things Journal, 2023.

[2] M. M. Badr, M. Mahmoud, Y. Fang, M. Abdulaal, A. J. Aljohani, W. Alasmary, and M. I. Ibrahem, "Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids," IEEE Internet of Things Journal, 2023.

[3] M. M. Badr, M. Mahmoud, M. Abdulaal, A. J. Aljohani, F. Alsolami, and A. Balamsh, "A novel evasion attack against global electricity theft detectors and a countermeasure," IEEE Internet of Things Journal, 2023.

[4] M. I. Ibrahem, M. Nabil, M. M. Fouda, M. M. Mahmoud, W. Alasmary, and F. Alsolami, "Efficient privacy-preserving electricity theft detection with dynamic billing and load monitoring for AMI networks," IEEE Internet of things journal, vol. 8, no. 2, pp. 1243–1258, 2020.

[5] "Security and privacy preservation for smart grid ami using machine learning and cryptography," 2022.

[6] M. M. Badr, M. Mahmoud, Y. Fang, M. Abdulaal, A. J. Aljohani, W. Alasmary, and M. I. Ibrahem, "Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids," IEEE Internet of Things Journal, 2023.

[7] A. Takiddin, M. Ismail, M. Nabil, M. M. Mahmoud, and E. Serpedin, "Detecting electricity theft cyber-attacks in AMI networks using deep vector embeddings," IEEE Systems Journal, vol. 15, no. 3, pp. 4 189 297–4 192 978, 2020.

[8] M. M. Badr, M. I. Ibrahem, M. Mahmoud, M. M. Fouda, F. Alsolami, and W. Alasmary, "Detection of false-reading attacks in smart grid net-metering system," IEEE Internet of Things Journal, vol. 9297, no. 2, pp. 1386–1401, 2021.

[9] D. Gu, Y. Gao, K. Chen, J. Shi, Y. Li, and Y. Cao, "Electricity theft detection in AMI with low false positive rate based on deep learning and evolutionary algorithm," IEEE Transactions on Power Systems, vol. 37, no. 6, pp. 4568–4578, 2022.

[10] M. Nabil, M. Ismail, M. M. Mahmoud, W. Alasmary, and E. Serpedin, "PPETD: Privacy-preserving electricity theft detection scheme with load monitoring and billing for AMI networks," IEEE Access, vol. 7, pp. 92 976 334–92 976 348, 2019297.

[11] M. I. Ibrahem, S. Abdelfattah, M. Mahmoud, and W. Alasmary, "Detecting electricity theft cyber-attacks in CAT AMI system using machine learning," in proc. of IEEE International Symposium on Networks, Computers and Communications (ISNCC), 2021, pp. 1–6.

[12] M. M. Badr, M. I. Ibrahem, M. Baza, M. Mahmoud, and W. Alasmary, "Detecting electricity fraud in the net-metering system using deep learning," in proc. of IEEE International Symposium on Networks, Computers and Communications (ISNCC), 2021, pp. 1–6.

[13] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in AMI using customers' consumption patterns," IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216–226, 2015.

[14] V. Ford, A. Siraj, and W. Eberle, "Smart grid energy fraud detection using artificial neural networks," in proc. of IEEE symposium on computational intelligence applications in smart grid (CIASG), 2014, pp. 1–6.

[15] M. I. Ibrahem, M. M. Badr, M. M. Fouda, M. Mahmoud, W. Alasmary, and Z. M. Fadlullah, "PMBFE: Efficient and privacy-preserving monitoring and billing using functional encryption for AMI networks," in proc. of International Symposium on Networks, Computers and Communications (ISNCC), 2020, pp. 1–7.

[16] M. M. Badr, M. I. Ibrahem, M. Mahmoud, W. Alasmary, M. M. Fouda, K. H. Almotairi, and Z. M. Fadlullah, "Privacy-preserving federated-learning-based net-energy forecasting," in proc. of IEEE SoutheastCon, 2022, pp. 133–139.

[17] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of non-technical losses using smart meter data and supervised learning," IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 2661–2670, 2018.

[18] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, and A. Bretas, "Identifying nontechnical power loss via spatial and temporal deep learning," in proc. of 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016, pp. 272–279.

[19] M. Nabil, M. Ismail, M. Mahmoud, M. Shahin, K. Qaraqe, and E. Serpedin, "Deep recurrent electricity theft detection in ami networks with random tuning of hyper-parameters," in 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 740–745.

[20] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, "Hybrid deep neural networks for detection of non-technical losses in electricity smart meters," IEEE Transactions on Power Systems, vol. 35, no. 2, pp. 1254–1263, 2019.

[21] A. Kumari and S. Tanwar, "A reinforcement-learning-based secure demand response scheme for smart grid system," IEEE Internet of Things Journal, vol. 9, no. 3, pp. 2180–2191, 2021.

[22] M. Lapan, Deep reinforcement learning hands-on: Apply modern RL methods, with deep Q-networks, value iteration, policy gradients, TRPO, AlphaGo Zero and more. Packt Publishing Ltd, 2018.

[23] S. Levine, A. Kumar, G. Tucker, and J. Fu, "Offline reinforcement learning: Tutorial, review, and perspectives on open problems," arXiv preprint arXiv:2005.01643, 2020.

[24] M. Moradi, Y. Weng, and Y.-C. Lai, "Defending smart electrical power grids against cyberattacks with deep q-learning," P R X Energy, vol. 1, p. 033005, 2022.

[25] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," IEEE transactions on neural networks and learning systems, vol. 30, no. 9, pp. 2684–2695, 2019.

[26] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[27] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.

[28] T. T. Nguyen and V. J. Reddi, "Deep reinforcement learning for cyber security," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–17, 2021.

[29] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 3201–3211, 2020.

[30] A. T. El-Toukhy, M. M. Badr, M. M. E. A. Mahmoud, G. Srivastava, M. M. Fouda, and M. Alsabaan, "Electricity theft detection using deep reinforcement learning in smart power grids," IEEE Access, vol. 11, pp. 59 558–59 574, 2023.

[31] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1606–1615, 2017.

[32] A. Seyyedabbasi, R. Aliyev, F. Kiani, M. U. Gulle, H. Basyildiz, and M. A. Shah, "Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems," Knowledge-Based Systems, vol. 223, p. 107044, 2021.

[33] S. Y. Diaba, M. Shafie-Khah, and M. Elmusrati, "Cyber security in power systems using meta-heuristic and deep learning algorithms," IEEE Access, vol. 11, pp. 18 660–18 672, 2023.

[34] S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, "Game-theoretic models of electricity theft detection in smart utility networks: Providing new capabilities with advanced metering infrastructure," IEEE Control Systems Magazine, vol. 35, no. 1, pp. 66–81, 2015.

[35] C.-H. Lin, S.-J. Chen, C.-L. Kuo, and J.-L. Chen, "Non-cooperative game model applied to an advanced metering infrastructure for non-technical loss screening in micro-distribution systems," IEEE Transactions on Smart Grid, vol. 5, no. 5, pp. 2468–2469, 2014.

[36] T.-S. Zhan, S.-J. Chen, C.-C. Kao, C.-L. Kuo, J.-L. Chen, and C.-H. Lin, "Non-technical loss and power blackout detection under advanced metering infrastructure using a cooperative game based inference mechanism," IET Generation, Transmission & Distribution, vol. 10, no. 4, pp. 873–882, 2016.

[37] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," Tsinghua Science and Technology, vol. 19, no. 2, pp. 105–120, 2014.

[38] S. K. Singh, R. Bose, and A. Joshi, "Pca based electricity theft detection in advanced metering infrastructure," in proc. of the 7th IEEE International Conference on Power Systems (ICPS). IEEE, 2017, pp. 441–445.

[39] Y. Peng, Y. Yang, Y. Xu, Y. Xue, R. Song, J. Kang, and H. Zhao, "Electricity theft detection in AMI based on clustering and local outlier factor," IEEE Access, vol. 9, pp. 107 250–107 259, 2021.

[40] A. Takiddin, M. Ismail, and E. Serpedin, "Robust data-driven detection of electricity theft adversarial evasion attacks in smart grids," IEEE Transactions on Smart Grid, vol. 14, no. 1, pp. 663–676, 2022.

[41] T. S. Murthy, N. Gopalan, and V. Ramachandran, "A naive bayes classifier for detecting unusual customer consumption profiles in power distribution systems-apspdcl," in 2019 Third International Conference on Inventive Systems and Control (ICISC). IEEE, 2019, pp. 673–678.

[42] V. Badrinath Krishna, R. K. Iyer, and W. H. Sanders, "Arima-based modeling and validation of consumption readings in power grids," in Critical Information Infrastructures Security: 10th International Conference, CRITIS 2015, Berlin, Germany, October 5-7, 2015, Revised Selected Papers 10. Springer, 2016, pp. 199–210.

[43] C. She, C. Sun, Z. Gu, Y. Li, C. Yang, H. V. Poor, and B. Vucetic, "A tutorial on ultrareliable and low-latency communications in 6G: Integrating domain knowledge into deep learning," Proceedings of the IEEE, vol. 109, no. 3, pp. 204–246, 2021.

[44] D. Wu, C. Wang, Y. Wu, Q.-C. Wang, and D.-S. Huang, "Attention deep model with multi-scale deep supervision for person re-identification," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 5, no. 1, pp. 70–78, 2021.

[45] M. J. Abdulaal, M. I. Ibrahem, M. M. Mahmoud, J. Khalid, A. J. Aljohani, A. H. Milyani, and A. M. Abusorrah, "Real-time detection of false readings in smart grid ami using deep and ensemble learning," IEEE Access, vol. 10, pp. 47 541–47 556, 2022.

[46] T. Talaei Khoei and N. Kaabouch, "A comparative analysis of supervised and unsupervised models for detecting attacks on the intrusion detection systems," Information, vol. 14, no. 2, p. 103, 2023.

[47] T. T. Khoei and N. Kaabouch, "Densely connected neural networks for detecting denial of service attacks on smart grid network," in proc. of 13th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2022, pp. 0207–0211.

[48] "Irish social science data archive," https://www.ucd.ie/issda/data/ commissionforenergyregulationcer/, last accessed: Sep. 2020.

[49] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity theft detection in power grids with deep learning and random forests," Journal of Electrical and Computer Engineering, vol. 2019, pp. 1–12, 2019.

[50] M. Hasan, R. N. Toma, A.-A. Nahid, M. Islam, J.-M. Kim et al., "Electricity theft detection in smart grid systems: A CNN-LSTM based approach," Energies, vol. 12, no. 17, p. 3310, 2019.

[51] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," arXiv preprint arXiv:1312.6199, 2013.

[52] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv preprint arXiv:1412.6572, 2014.

[53] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "Deepfool: a simple and accurate method to fool deep neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2574–2582.

[54] A. Rozsa, E. M. Rudd, and T. E. Boult, "Adversarial diversity and hard positive generation," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 25–32.

[55] C. J. Watkins and P. Dayan, "Q-learning," Machine learning, vol. 8, no. 3, pp. 279–292, 1992.

[56] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, "Application of deep reinforcement learning to intrusion detection for supervised problems," Expert Systems with Applications, vol. 141, p. 112963, 2020.

[57] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in Proceedings of the 2017 ACM on Asia conference on computer and communications security, 2017, pp. 506–519.

[58] A. Thangaraju and C. Merkel, "Exploring adversarial attacks and defenses in deep learning," in 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE, 2022, pp. 1–6.

[59] J. Li, "Analyse of influence of adversarial samples on neural network attacks with different complexities," in 2022 2nd International Signal Processing, Communications and Engineering Management Conference (ISPCEM). IEEE, 2022, pp. 329–333.

[60] A. Takiddin, M. Ismail, and E. Serpedin, "Robust detection of electricity theft against evasion attacks in smart grids," in ICC 2021-IEEE International Conference on Communications. IEEE, 2021, pp. 1–6.

[61] J. Tian, B. Wang, J. Li, and Z. Wang, "Adversarial attacks and defense for cnn based power quality recognition in smart grid," IEEE Transactions on Network Science and Engineering, vol. 9, no. 2, pp. 807–819, 2021.

AHMED T. EL-TOUKHY received the B.Sc. and M.Sc. degrees in electrical engineering from Al-Azhar University, Cairo, Egypt, in 2011, and 2017, respectively. He is currently pursuing his Ph.D. degree at the Department of Electrical & Computer Engineering, Tennessee Tech. University, TN, USA. He is also holding the position of a Lecturer Assistant at the Faculty of Engineering, Al-Azhar University, Cairo, Egypt. His research interests include machine learning, reinforcement learning, cryptography, network security, 5G networks, and cognitive radio.

MOHAMED M. E. A. MAHMOUD received PhD degree from the University of Waterloo in April 2011. Currently, Dr. Mahmoud is an associate professor in Department Electrical and Computer Engineering, Tennessee Tech University, USA. The research interests of Dr. Mahmoud include security and privacy preserving schemes for smart grid, e-health, and intelligent transportation systems. Dr. Mahmoud has received NSERC-PDF award. He won the Best Paper Award from IEEE International Conference on Communications (ICC'09), Dresden, Germany, 2009. Dr. Mahmoud is the author for more than 100 papers published in IEEE conferences and journals. He serves as an Associate Editor in IEEE Internet of Things Journal and Springer journal of peer-to-peer networking and applications. He served as a technical program committee member for several IEEE conferences.

ATEF H. BONDOK received the B.Sc. degree in electrical engineering from Al-Azhar University, Cairo, Egypt, in 2011, and the M.Sc. degree in electronics and communications engineering from Egypt-Japan University of Science and Technology (EJUST), Alexandria, Egypt, in 2018. He is currently pursuing his Ph.D. degree at the Department of Electrical & Computer Engineering, Tennessee Tech. University, TN, USA. He is also holding the position of a Lecturer Assistant at the Faculty of Engineering, Al-Azhar University, Cairo, Egypt. His research interests include machine learning, cryptography and network security, and privacy preserving schemes for smart grid communication and AMI networks.

MOSTAFA M. FOUDA (Senior Member, IEEE) received the B.S. degree (as the valedictorian) and the M.S. degree in Electrical Engineering from Benha University, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in Information Sciences from Tohoku University, Japan, in 2011. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Idaho State University, ID, USA. He also holds the position of a Full Professor at Benha University. He was an Assistant Professor at Tohoku University and a Postdoctoral Research Associate at Tennessee Technological University, TN, USA. He has (co)authored more than 160 technical publications. His current research focuses on cybersecurity, communication networks, signal processing, wireless mobile communications, smart healthcare, smart grids, AI, and IoT. He has guest-edited a number of special issues covering various emerging topics in communications, networking, and health analytics. He is currently serving on the editorial board of IEEE Transactions on Vehicular Technology (TVT) and IEEE Access. He has received several research grants including NSF Japan-U.S. Network Opportunity 3 (JUNO3). He is a Senior Member of IEEE.

MAAZEN ALSABAAN received the B.S. degree in electrical engineering from King Saud University (KSU), Saudi Arabia, in 2004, and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Canada, in 2007 and 2013, respectively. He is currently an Associate Professor with the Department of Computer Engineering, KSU. From 2015 to 2018, he was the Chairperson of the Department. He serves as a consultant for different agencies and has been awarded many grants from KSU and King Abdulaziz City for Science and Technology (KACST). His current research interests include wireless communications and networking, surveillance systems, vehicular networks, green communications, intelligent transportation systems, and cybersecurity

• • •