frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 12 January 2023
DOI 10.3389/fenrg.2022.1071973

OPEN ACCESS

EDITED BY
Hanlin Zhang,
Qingdao University, China

REVIEWED BY
Linqiang Ge,
Columbus State University, United States
Yalong Wu,
University of Houston–Clear Lake, United States

*CORRESPONDENCE
Qingyu Yang,
yangqingyu@mail.xjtu.edu.cn

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 17 October 2022
ACCEPTED 21 December 2022
PUBLISHED 12 January 2023

CITATION
Li D, Yang Q, Ma L, Peng Z and Liao X (2023), Offense and defence against adversarial sample: A reinforcement learning method in energy trading market. *Front. Energy Res.* 10:1071973.
doi: 10.3389/fenrg.2022.1071973

Offense and defence against adversarial sample: A reinforcement learning method in energy trading market

Donghe Li[1], Qingyu Yang[1,2]*, Linyue Ma[3], Zhenhua Peng[1] and Xiao Liao[3]

[1]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China, [2]State Key Laboratory Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an, China, [3]State Grid Information and Telecommunication Group Co., LTD, Beijing, China

The energy trading market that can support free bidding among electricity users is currently the key method in smart grid demand response. Reinforcement learning is used to formulate optimal strategies for them to obtain optimal strategies. Non-etheless, the security problem raised by artificial intelligence technology has been paid more and more attention. For example, the neural network has been proved to be able to resist adversarial example attacks, thus affecting its training results. Considering that reinforcement learning is also widely used for training by neural networks, the security problem can not be ignored, especially in scenarios with high security requirements such as smart grids. To this end, we study the security issues in reinforcement learning-based bidding strategy method facing by the adversarial example. First of all, regarding to the electric vehicle double auction market, we formalize the bidding decision problem of EVs into a Markov Decision Process, so that reinforcement learning is used to solve this problem. Secondly, from the perspective of attackers, we have designed a local Fast Gradient Sign Method which affects the environment and the results of reinforcement learning by changing its own bidding. Then, from the perspective of the defender, we designed a reinforcement learning training network containing an attack identifier based on the deep neural network, so as to identify malicious injection attacks to resist against adversarial attacks. Finally, comprehensive simulations are conducted to verify our proposed method. The results shows that, our proposed attack method will reduce the auction profit by influencing reinforcement learning algorithm, and the protect method will be able to completely resist such attacks.

KEYWORDS

double auction, markov decision process, reinforcement learning, adversarial example, fast gradient sign method, adversarial example detection

1 Introduction

With the application of more and more Internet of Things equipment and information technology, the traditional purely physical power grid has gradually transformed into the Cyber Physic System-based (CPS) Smart Grid (SG) Zhang et al. (2016); Hong et al. (2019); Bandyszak et al. (2020); Zhao et al. (2021); An et al. (2022). Smart grid provides bi-direction information flow and power flow through advanced information technology, and realize effective interconnection of power generation, transmission, distribution and others Grigsby (2007); Haller et al. (2012). The most important function of smart grid is to plan and guide users to actively adjust their power load by taking advantage of the bi-direction transmission of information between the grid and users, so as to achieve the effect of peak load shifting, which is called Demand Response (DR) Croce et al. (2017); Albadi and El-Saadany (2007); Huang et al. (2019).

With the development of science, technology and society, almost all equipment depends on electric power transportation. People are increasingly dependent on electricity, which brings great pressure to the stable operation of the power grid. It is urgent to use demand response methods to alleviate the pressure. The mainstream demand response methods are divided into two categories, one is price-based DR and the other is incentive-based Hahn and Stavins (1991); Pyka (2002); Liu et al. (2005) DR. The price-based DR method guides users to adjust the load actively by setting the price, such as Time of Use Price (TOU), Real Time Pricing (RTP), and so on Ding et al. (2016); Cheng et al. (2018); Samadi et al. (2010). The incentive-based DR method realizes load migration by directly managing the user's load, such as Direct Load Control (DLC), Energy Trading Market, and so on Wu et al. (2015); Ruiz et al. (2009); Ng and Sheble (1998).

Due to the continuous increase of renewable energy Hosseini et al. (2021); Giaconi et al. (2018) and the popularity of flexible load and energy storage Miao et al. (2015); Liu et al. (2018) equipment such as electric vehicles, the energy trading market, which allows users to freely bid and transmit electric energy, has received extensive attention Kim et al. (2019); Esmat et al. (2021). Generally speaking, in typical energy trading market, the electricity users (or electric energy company) with surplus energy will act as sellers, the electricity uses with insufficient energy will act as buyers. Regarding to the winner decision mechanism in energy trading market, considering that the market has strong uncertainty, and the market needs to ensure the benefits of participants, fairness and other properties to attract more participants, the auction mechanism has better performance than the optimization algorithm, which is the mainstream of current research. In recent years, most scholars have devoted themselves to studying a more efficient auction mechanism from the perspective of auction platform. For example, two example of auction mechanism.

With further research, scholars found that determining optimal bidding strategy from the perspective of participants also affects the performance of the energy trading market. Reinforcement learning is a branch of machine learning, which focuses on interactive goal oriented learning Mohan and Laird (2014); Erhel and Jamet (2016). It can independently explore the environment and constantly optimize its own strategies driven by rewards. Deep reinforcement learning combines the independent exploration ability of reinforcement learning with the strong fitting ability of neural network, and has been widely studied Yu et al. (2022); Zhang et al. (2019). Deep reinforcement learning technology is widely used in the optimal decision-making of smart grid due to its strong perception and understanding ability and sequential decision-making ability Barto et al. (1989); Roijers et al. (2013). For example, two example of RL bidding.

In recent years, deep learning technology has made unprecedented development and has been widely used in many fields. However, its security problems have become increasingly prominent. Szegedy et al. (2013) found that the deep neural network is extremely vulnerable to the attack of adding disturbance to the confrontation sample image. This attack will cause the neural network to classify the image with high confidence, and the human can hardly distinguish the confrontation sample from the original image with their eyes. For instance, in **Figure 1**, the original panda image is judged as a panda by the depth learning image classification model with 57.7% confidence, but after adding small random noise, the model will misjudge the image as a gibbon with high confidence Goodfellow et al. (2014). The sample, which is carefully created or generated and leads to the wrong prediction of the deep learning model, is called Adversarial Example (AE) Szegedy et al. (2013). The training process of deep reinforcement learning also relies on neural networks, so theoretically, there is also a risk of being attacked by adversarial example. Moreover, the smart grid system, which requires high reliability, will have a great impact once the reinforcement learning algorithm is attacked by the adversarial example.

As introduced above, it is urgent to study the security problems of reinforcement learning algorithm applied in smart grid. In our paper, we mainly focus on the attack and defense of the bidding strategy algorithm based on reinforcement learning of double energy trading mechanism. At present, the research on counter attack has been carried out for several years, but the following problems still exist. 1) The application scenarios of reinforcement learning algorithms are mostly game environments. The research on adversarial attack is mainly carried out on images, and the effectiveness of scenes other than images is hardly explored. 2) It is worth exploring the adversarial attack and defense effects when the state observation is very limited information.

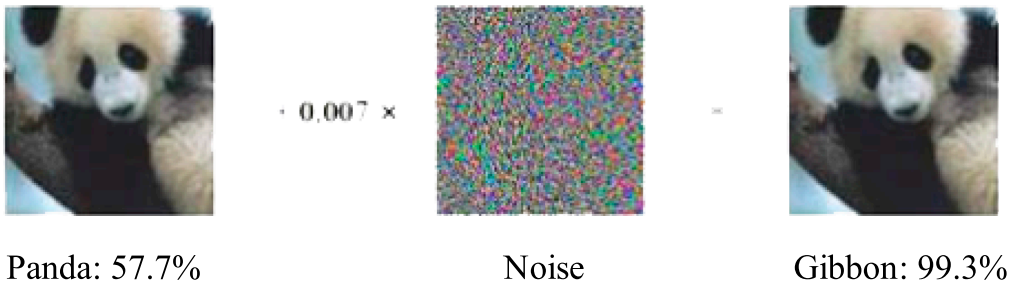

FIGURE 1
Examples of adversarial attack.

To this end, in this paper we will study the security issue aiming at the reinforcement learning-based bidding strategy method in Electric Vehicle double energy trading market. Specifically, we first conduct a double auction model/mechanism of EV double energy trading market. And we formalize the EVs' bidding strategy as a Markov Decision Process model so as to solve it by deep reinforcement learning. After that, we studied a method of generating Adversarial Example based on fast gradient sign method from the adversary's point of view, and explore the impact of AE on deep reinforcement learning algorithm. Then, we designed a deep reinforcement learning network that contains a deep neural network-based adversarial example discriminator to resist such attacks from the perspective of the defender. Finally, comprehensive simulations are conducted to verify our methods.

The remainder of this paper is organized as follows. In **Section 2**, we briefly review the research efforts related to energy trading market, reinforcement learning method and the adversarial example. In **Section 3**, we introduce the preliminaries of this paper. In **Section 4**, a local-fast gradient adversarial example generating method is proposed. In **Section 5**, the deep neural network-based adversarial example discriminator is proposed to protect the reinforcement learning method. In **Section 6**, the simulations are conducted. Finally, we conclude this paper in **Section 7**.

2 Related work

With the development of distributed energy and energy storage equipment, the electricity trading market between users has become an important research content in smart grid demand response. For example, Jin et al. (2013) studied the electric vehicle charging scheduling problem from the perspective of energy market, and proposed a mixed integer linear programming (MILP) model and a simple polynomial time heuristic algorithm to provide the best solution. Zeng et al. (2015) introduced a group sales mechanism for electric vehicle demand response management in the vehicle to grid (V2G) system and designed a group auction transaction mechanism to realize the bidding decision of electric vehicle users. The results show that this mechanism can reduce the system cost. Zhou et al. (2015) proposed an online auction mechanism to solve the demand response in smart grid, expressed the problem of maximizing social welfare as an online optimization problem in the form of natural integer linear programming, and obtained the optimal solution.

Reinforcement learning, as a powerful artificial intelligence tool in sequential decision-making problems, has been increasingly applied to the scheduling, decision-making and energy trading strategies in smart grid. For instance, Zhang et al. (2018) summarized the application research work of deep learning, reinforcement learning and deep reinforcement learning in smart grid. Wan et al. (2018) proposed a deep reinforcement learning real-time scheduling method considering the randomness of EV users' behavior and the uncertainty of real-time electricity price for a single EV user, designed a representation network to extract identification features from electricity prices and a deep Q network to approximate the optimal action value function to determine the optimal strategy.

As introduced above, the research and application of reinforcement learning in smart grid has been very extensive, so its security must be guaranteed. While as the application potential of deep reinforcement learning algorithm is gradually exploited, the adversarial attack and defense against deep reinforcement learning has gradually attracted the attention of scholars. For example, Huang et al. (2017) proved the effectiveness of adversarial attack against the neural network strategy in reinforcement learning. Lin et al. (2017) proposed two adversarial attack methods against the reinforcement learning neural network, and verified the effectiveness of the attack in a typical reinforcement learning environment. Qu et al. (2020) proposed a "minimalist attack" method for the deep reinforcement learning strategy network, and formulated countermeasures by defining three key settings and verified the effect of the attack. Although the above research is aimed at

reinforcement learning, in fact, the adversarial examples are aimed at environmental information mainly based on pictures. In the application of smart grid reinforcement learning, most of the environmental information is digital, so the research in this area needs to be carried out urgently.

3 Preliminaries

In this section, we will first introduce the Electric Vehicle double auction model, and then introduce the definition of deep reinforcement learning and adversarial attack.

3.1 Electric vehicle double auction energy Trading Market

System Model: In this paper, we consider a Electric Vehicle energy trading market which is shown in **Figure 2**. Specifically, the EVs which need to charge act as buyers, and the EVs with surplus energy and want to discharge to get some profits act as sellers. They are allowed to submit their charing/discharing request freely. The bidding information always including the valuation, demand/supply volume, arriving/departuring time. These bidding information would submitted to the auctioneer, which is acted by microgrid control center. The auctioneer will make a fair, effective determination within these information. In general, the auctioneer is always assumed as a trust platform, which means auctioneer will not steal or tamper the bidding information to threat the auction market. Note that, in our paper, the auction determination rule is considered as the typical double auction mechanism: McAfee mechanism. Due to the limit of the space, we will not introduce the work flow in detail.

In the above auction market, the bidding strategy of EVs is the key issue affecting their profits in the market. However, in such a game market, the information of competitors and environment is constantly changing, and it is impossible to obtain an optimal bidding strategy through traditional methods. And reinforcement learning can get an optimal strategy to adapt to different environments in the future by constantly exploring the environment. Therefore, at present, using reinforcement learning to find the optimal strategy is the mainstream to solve the bidding strategy problem.

Threat Model: Non-etheless, reinforcement learning is an artificial intelligence method, and neural networks are often used in the solution process. The neural network has been proved to be vulnerable to attacks against samples, that is, by adding a little noise to the samples, the training results of the neural network are affected. In our EV double auction market, the reinforcement learning bidding strategy will be attacked by this attack. So in our paper, we assume the adversary is one participant in the auction market. He/she modifies his/her own bidding information, thereby affecting the reading of the environment by reinforcement learning, and thus affecting the bidding strategy of other users. Specific attack methods will be given in **Section 4**.

3.2 Deep reinforcement learning

Deep reinforcement learning (DRL) combines the perceptual capability of deep learning (DL) with the decision-making capability of reinforcement learning (RL), where agent perceives information through a higher dimensional space and applies the obtained information to make decisions for complex scenarios. Deep reinforcement learning is widely used because it can achieve direct control from original input to output through end-to-end learning. Initially, due to the lack of training data and computational power, scholars mainly used deep neural networks to downscale high-latitude data, which were later used in traditional reinforcement learning algorithms Lange and Riedmiller (2010). Then Mnih of DeepMind proposed Deep Q-networks (DQN) Mnih et al. (2013), and people gradually started to study them in a deeper level while applying them to a wider range of fields. In recent years, research in deep reinforcement learning has focused on DQN, which combines convolutional neural networks with Q-learning and introduces an experience replay mechanism that allows algorithms to learn control policies by directly sensing high-dimensional inputs. As the most basic reinforcement learning algorithm, because of its good training speed and effect, it is widely used in various practical scenes.

4 Adversarial attack method against reinforcement learning -based trading market

4.1 Adversarial attack

Deep learning algorithms have been widely used in many fields, but the ensuing security issues deserve attention. Adversarial attack is an important risk. Since the input of deep neural network is a numerical vector, the attacker can maliciously design a targeted numerical vector (called adversarial sample) to make the deep neural network make a misjudgment. In the field of deep learning, we assume that x is the input and f represents a deep neural network, the production of adversarial samples can be represented as:

$$\min_{\delta} d(x, x+\delta) \tag{1}$$

subject to

$$f(x) \neq f(x+\delta) \tag{2}$$

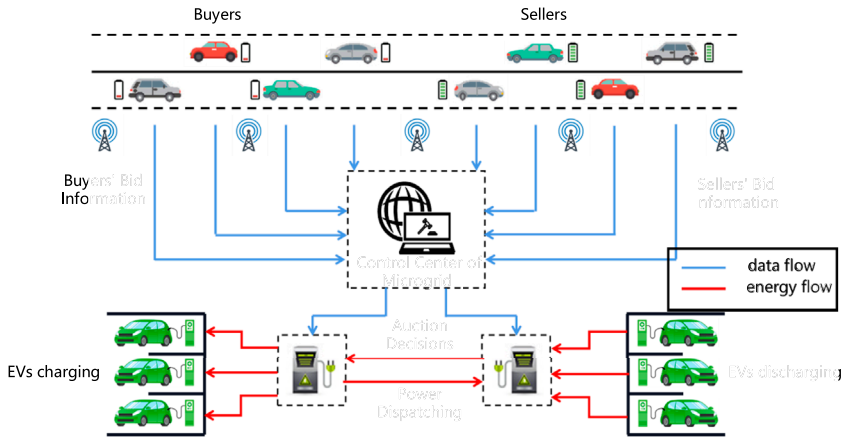

FIGURE 2
Ev double auction model.

where d represents the distance metric, which is calculated by l – norm.

The above equation also shows that the attacker tries to find the minimal perturbation δ that can make the deep neural network output wrong results.

Deep reinforcement learning (DRL) algorithms integrate deep neural networks based on the theory of reinforcement learning, which also leads to the risk of suffering from adversarial attacks. In value-based RL algorithms, adversarial samples can make the neural network misestimate the value of a specific action at a specific state and guide the agent to choose the wrong action. In policy-based RL algorithms, the attacker can make the agent unable to use the policy gradient to select the optimal policy through the adversarial sample.

4.2 Double auction and bidding strategy formalization

In the double auction scene model of electric vehicles, there are mainly the following three participants: auctioneer, buyer and seller. Among them, the microgrid control center serves as the auctioneer of the trading market, the electric vehicle with insufficient electric energy serves as the buyer, and the electric vehicle with surplus electric energy serves as the seller. In the electricity trading market, there are multiple buyers and sellers who can participate in the auction using their mobile devices or Internet of vehicles systems. The winning bidder trades the electric energy through the charging pile, avoiding the transmission loss of electric energy in the traditional power grid system.

According to the characteristics of the auction process, this paper discretizes the transaction process and adopts the integer set $T = \{1, 2, \cdots\}$ to represent the time series in the transaction process. B is the set of buyers, and the number of buyers is $|B|$. S is the set of sellers, and the number of sellers is $|S|$.

At time slot t, the actual power demand of the ith buyer is $d_{i,t}$, and its bidding information is denoted as a triplet:

$$\chi_{b,i,t} = \left\{i, p_{b,i,t}, q_{b,i,t}\right\}, \quad i \in B \tag{3}$$

where i represents the buyer ID, $p_{b,i,t}$ represents the valuation of one unit electricity submitted by ith buyer, and $q_{b,i,t}$ represents the submitted volume.

Similarly, at time slot t, the actual power supply of the jth seller is $u_{j,t}$, and its bidding information is:

$$\chi_{s,j,t} = \left\{j, p_{s,j,t}, q_{s,j,t}\right\}, \quad j \in S \tag{4}$$

where j represents the seller ID, $p_{s,j,t}$ represents the valuation of one unit electricity submitted by jth seller, and $q_{s,j,t}$ represents the submitted volume.

The actual power supply/demand and the submitted volume:

$$q_{b,i,t} \le d_{i,t}, t \in T, \quad i \in B \tag{5}$$

$$q_{s,j,t} \le u_{j,t}, t \in T, \quad j \in S \tag{6}$$

In the energy trading market, electric vehicle users with insufficient and excessive electric energy report bidding information according to their own wishes. The microgrid control center, as the auctioneer, organizes a double auction to determine the winning buyer and seller, and then determine the transaction price and volume of each buyer and seller. Subsequently, the auction results (including the winning buyer/seller) are released to all participants in the system to ensure the fairness and verifiability of the auction.

4.3 Rational

Research on adversarial attack theory in deep learning has made some progress. At present, deep learning plays an important role in the field of computer vision. Most of the adversarial attack methods for deep learning are based on the image-based system. The latest research on adversarial attack methods in deep reinforcement learning algorithms is also mainly oriented at game scenarios, and the observations of agents are also images. Note that the application of deep reinforcement learning algorithm is also likely to face the threat of adversarial samples in the scenario of electric vehicle energy trading.

Theoretically, the observation of agents in smart grid is mainly the digital data of electric energy, electricity price and so on. Similarly with the image data, these digital data is also the numerical vectors, and it has fewer input features. This makes it possible to produce adversarial samples for deep reinforcement learning algorithms in energy trading market theoretically. Once affected by the attacker's malicious interference, it may have a negative impact on the benefits of users in the power grid.

In the process of participating in smart grid energy trading, electric vehicle users need to continuously submit their bidding information to participate in double auction, and achieve their optimal benefits through multi-step decision-making. There is correlation between continuous decisions. Deep reinforcement learning algorithm can exactly give full play to its unique advantages in this process. Considering the current situation of actual power grid charging and discharging, it is appropriate to discretize the bidding price and trading volume of energy trading. Deep-Q-network (DQN) is a good choice in power trading scenarios. This paper considers the adversarial attack research on deep Q learning algorithm in power transaction of smart grid.

4.4 Adversarial attack method against reinforcement learning-based bidding strategy

In the double auction process, the attacker can affect the benefits of the other auctioneer by maliciously modifying his real demand/supply, making the average cost of the buyer group rise or making the average profit of the seller group decrease. Because each participant in the double auction has limited observations, and some state quantities cannot be explicitly modified, once changed, they are easy to be screened out and lose the attack effect. Therefore, this paper considers that attackers can change the state of agents in the system by submitting false bidding information. As a result, the deep-Q-network selects non-optimal bidding strategies to reduce the average reward.

To be specific, in the bilateral auction market, it is considered that there is an attacker in the buyer group. His purpose is to influence the state observation of other buyers by maliciously modifying his quantity demanded, so other buyers will make non-optimal bidding strategy. Ultimately, it affects the utility of the buyer group. Similarly, for the seller group, this paper also considers the existence of an attacker and studies the impact of the generated adversarial samples on the seller group's revenue. Adversarial attacks in electrical energy trading are shown in **Figure 3**.

At present, most of the adversarial sample production method for deep reinforcement learning borrows from the methods in deep learning. The Fast Gradient Sign Method (FGSM) make adversarial perturbations and add them to the observations, so as to attack the DRL agent. The core idea is to add perturbations along the direction where the deep neural network model gradient changes the most to induce the model output error results. Formally, adversarial samples generated by FGSM can be expressed as follows:

$$x' = x + \varepsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right) \tag{7}$$

where ε is the size of the disturbance, J represents the cross-entropy loss function, θ is the parameter of the neural network, x represents the model input, and y represents the sample label (here refers to the optimal action term). The cross-entropy loss function here measures the difference between the distribution of the label y and the distribution that puts all the weight on the optimal action.

Inspired by the original FGSM, to address the problem that the attacker can only modify some observations to avoid being detected by the system, in this paper, a local-FGSM is proposed to make adversarial samples by modifying some components of the agent state vector, which can be expressed as follows:

$$x' = x + \varepsilon \cdot sign\left(\nabla_x J(\theta, x, y)\right) \cdot \mu \tag{8}$$

where μ is a vector whose dimension is equal to the dimension of input x, the value of the dimension corresponding to the component to be modified by the agent state vector is 1, and the rest is 0.

The attack process is shown in **Algorithm 1**.

5 Adversarial sample recognition-based reinforcement learning-based energy Trading Mechanism

In this section, we propose a adversarial sample recognition-based reinforcement learning method for the above double auction.

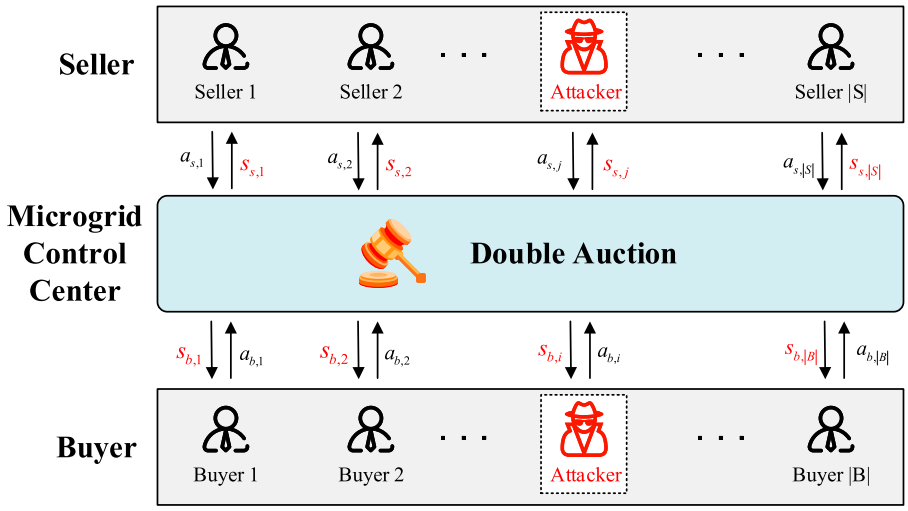

FIGURE 3
Adversarial attacks in electricity trading.

Input the number of buyers $|B|$ and the number of sellers $|S|$.
Select a well-trained buyer Q-network and seller Q-network.
for *epoch= 1 to* E_1 **do**
  Initialize the states of buyers and sellers.
  for *step= 1 to* v **do**
    Make bidding decisions according to the buyer Q-network and get reward r_{step}.
    Prepare the cross-entropy loss function J.
    Randomly select a buyer as the attacker and make adversarial sample $x' = x + \varepsilon \cdot sign\left(\nabla_x J\left(\theta, x, y\right)\right) \cdot \mu$
    Make bidding decisions according to the buyer Q-network and get reward r'_{step} again.
    Update the states of buyers and sellers.
  end
  Compare the buyers' average cumulative rewards with and without adversarial attacks.
  Initialize the states of buyers and sellers.
  for *step= 1 to* v **do**
    Make bidding decisions according to the buyer Q-network and get reward r_{step}.
    Prepare the cross-entropy loss function J.
    Randomly select a seller as the attacker and make adversarial sample $x' = x + \varepsilon \cdot sign\left(\nabla_x J\left(\theta, x, y\right)\right) \cdot \mu$
    Make bidding decisions according to the seller Q-network and get reward r'_{step} again.
    Update the states of buyers and sellers.
  end
  Compare the sellers' average cumulative rewards with and without adversarial attacks.
end

Algorithm 1. The process of local-FGSM adversarial attack.

5.1 Markov Decision Process model

We construct the EV electric energy trading double auction scenario as a Markov Decision Process (MDP) with discrete time steps, which can be expressed as a quadruple $\{S, A, P, R\}$.

S stands for the state space. S_B and S_S denote the state space sets of buyers and sellers, respectively. Electric vehicle users can be informed of the total demand and total supply during the current period. Assuming that each participant conducts v auctions in the trading market, the variable σ is introduced to indicate whether the participant currently participated in the last auction or not. In time slot t, the states of the *ith* buyer and the *jth* seller are denoted as:

$$s(b,i,t) = \{d_{i,t}, D_t, U_t, \sigma\} \tag{9}$$

$$s(s,j,t) = \{u_{j,t}, D_t, U_t, \sigma\} \tag{10}$$

A stands for action space. After acquiring observations at each time slot, buyers and sellers need to submit bidding information to participate in the bilateral auction, and the decision of bidding price and bidding volume will have an impact on their respective profits. In this system, the bidding information submitted by buyers and sellers is regarded as their respective actions. In time slot t, the actions of buyer EV users and seller EV users are denoted as

$$a(b,i,t) = \{p_{b,i,t}, q_{b,i,t}\} \tag{11}$$

$$a(s,j,t) = \{p_{s,j,t}, q_{s,j,t}\} \tag{12}$$

R is the reward function. The immediate reward of the buyer and seller of time slot t is denoted as $r(t)$. For the buyer, if he wins the bid in the bilateral auction, the cost is $p_{b,i,t} \cdot q_{b,i,t}$. The buyer's goal is to keep the cost as low as possible, but win the auction as much as possible. For the seller, if he succeeds in winning the bid in the bilateral auction, then his profit from selling electric energy is $p_{s,j,t} \cdot q_{s,j,t}$; otherwise, if he fails to win the bid, his profit $p_{s,j,t} \cdot q_{s,j,t}$ is 0. Then, the reward function of the buyer in time slot t is denoted by:

$$r(b,i,t) = \begin{cases} -p_{b,i,t} q_{b,i,t}, & i \in M_B \\ -2p_{b,i,t} q_{b,i,t}, & i \notin M_B \end{cases} \tag{13}$$

Setting the buyer's reward function as negative can make the optimal strategy for buyers and sellers using deep reinforcement learning algorithms formally consistent, and the goal is to

Input the number of buyers $|B|$ and the number of sellers $|S|$.
Randomly initialize the parameters of the deep Q-network θ.
Initialize the parameters of the target network θ'.
Initialize the capacity of replay buffer N.
Initialize the greedy coefficient ε.
for *epoch= 1 to* E_2 **do**
  Obtain the initial states of the buyers and sellers.
  for *t= 1 to T* **do**
    if a random number $x \leq \varepsilon$ **then**
      Randomly choose the action a_t from action space.
    else
      Select $a_t = \arg\max_a Q(s_t, a, \theta)$.
    end if
    Perform double auction and get the reward r and next state s_{t+1}.
    Store the transition $[s_t, a_t, r_t, s_{t+1}]$ in experience replay.
    Randomly choose mini-batch B from the replay buffer.
    Calculate $Q_{evel} = Q(S_t, A_t, \theta)$.
    Calculate $Q_{target} = r_t + \gamma \cdot Q\left(s_{t+1}, \arg_{a'} Q(s_{t+1}, a', \theta), \theta'\right)$.
    Perform the gradient descent on $\sum_{i=1}^{B} (Q_{target} - Q_{evel})^2$ with respect to the Q-network parameters.
    Update the greedy coefficient ε.
    Every C steps update $\theta' = \theta$.
  end
end

Algorithm 2. The training process of deep Q-learning algorithm in double auction.

maximize their own long-term benefits. The buyer's cost is the absolute value of the reward. For the buyer who fails to win the bid, it may need to spend more money to buy the much-needed power, so a large penalty coefficient is added to it to encourage the buyer to avoid the failure as much as possible.

The seller's reward function is denoted as:

$$r(s,j,t) = p_{s,j,t} \cdot q_{s,j,t} \tag{14}$$

P represents the state transition function. Function p_t is defined as a transition function. The state transition probability from state s_t to state s_{t+1} is expressed as:

$$p_t : s_t \times a_t \rightarrow s_{t+1} \tag{15}$$

5.2 Solution *via* reinforcement learning

In the electric energy trading market model designed in this paper, deep Q-learning algorithm is used to learn the optimal bidding strategy for buyers and sellers in microgrid bilateral auction respectively. In order to estimate the state action value function, this paper defines a multi-layer perceptron as a deep-Q-network for buyers and sellers respectively, taking the state as input and the state action value $Q(s,a) \approx Q(s,a,\theta)$ as output, where theta is the neural network parameter. Deep-Q-network is a fully connected neural network with two hidden layers.

In the process of training deep-Q-network, the state s_t, action a_t, reward r_t and next state s_{t+1} obtained from each interaction with the system environment can form an empirical tuple, denoted as $[s_t, a_t, r_t, s_{t+1}]$. For buyers and sellers, a experience replay is set to store the corresponding experience tuples respectively, and its capacity is N.

In addition, a target network with the same structure as the deep-Q-network is defined to solve the correlation and stability problems. Both the deep-Q-network and the target network initially have the same parameters. In the training process, the target network's parameter θ' is updated to the deep Q network's parameter θ every C steps. At each training session, a mini-batch sample of size B is sampled from the experience replay and used as input to the main network, and the output is selected to calculate the Q-value:

$$Q_{evel} = Q(S_t, A_t, \theta) \tag{16}$$

The target Q-value is:

$$Q_{target} = r_t + \gamma \cdot Q\left(s_{t+1}, \arg_{a'} Q(s_{t+1}, a', \theta), \theta'\right) \tag{17}$$

The γ is a discount factor, indicating the extent to which the future reward affects the current reward. The smaller the γ, the more the agent focuses on the current reward, and *vice versa*.

The loss function is calculated from the difference between the target Q-value and the estimated Q-value, and the parameters of the main network θ are updated by gradient descent. The loss function is:

$$L(t) = \sum_{i=1}^{B} \left(Q_{target} - Q_{evel}\right)^2 \tag{18}$$

The training process of deep Q-learning algorithm is shown in **Algorithm 2**.

5.3 Defense Strategy Architecture

At present, deep learning mainly achieves defense effect by modifying network structure, objective function or training process, but most defense methods cannot meet the practical application scenarios of DRL. From the perspective of data security and reliability, this paper considers the use of additional network to preprocess the data of the perturbed observation vector and screen out the adversarial samples to ensure the system security.

When EV users participate in the electric energy trading market, they obtain their current state according to the data published by the microgrid control center. Based on the deep Q learning algorithm proposed above, deep Q network is used to help EV users make optimal bidding decisions. In order to avoid adversarial samples that may appear in the process of electric energy trading, the state information of all EV users is screened out by an adversarial sample discriminator before bilateral auction. In this way, only real samples can be allowed to participate in the auction, and then bidding decisions can be made based on the deep-Q-network, and further transaction decisions and scheduling optimization can be made by the microgrid control center. **Figure 4** shows the architecture of the adversarial defense model.

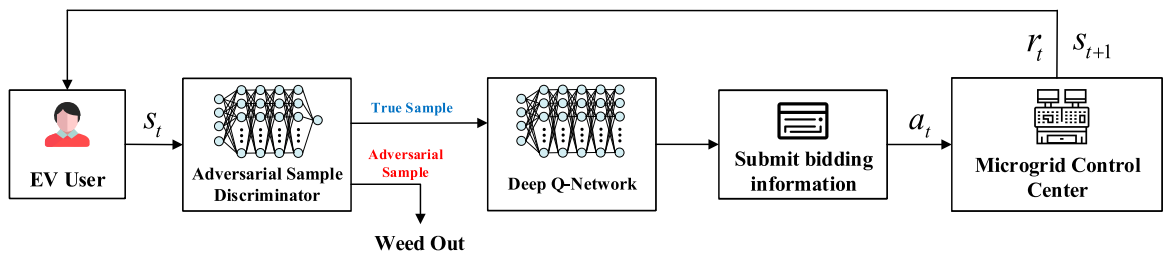

FIGURE 4
Defense strategy architecture.

The adversarial sample discriminator is designed by a fully connected deep neural network with four hidden layers. The specific network structure is shown in the following table. The input layer consists of four neurons corresponding to the four elements of the electric vehicle user's state. The output is 0 or 1, representing the input EV states as adversarial and real samples, respectively.

The adversarial sample discriminator is essentially a binary classifier, which is used to judge whether the input EV state sample is an adversarial sample, and its training process is a supervised learning process. Firstly, select a buyer deep-Q-network and train it well, so that users can make the optimal bidding decision according to it. Then the data set is collected and made. In each episode of electric energy trading, after the user state is initialized, 10 bilateral auctions are conducted successively, and the next state of the user will be obtained after each auction. The local-FGSM is used to make adversarial samples, and the real next time state and adversarial samples are stored with labels being made. After that, by using the collected adversarial samples and real samples, the training set and test set are divided to train the adversarial sample discriminator, and the weight is updated by using the back propagation to reduce the loss function value. Finally, the effectiveness of the adversarial sample discriminator is verified by the test set. Similarly, the adversarial sample discriminator of sellers' Q-network is trained.

The training process of the adversarial sample discriminator is shown in **Algorithm 3**.

Select a well-trained Q-network.
Randomly initialize the parameters of the adversarial sample discriminator θ.
for *epoch= 1 to* E_3 **do**
  Obtain the initial states of the buyers and sellers.
  for *t= 1 to T* **do**
    Make bidding decisions according to the Q-network.
    Perform double auction and get the reward r and next state s_{t+1}.
    Store real samples $[s_t, a_t, s_{t+1}]$ and make labels.
    Make the adversarial sample s'_{t+1} according to Equation (8).
    Store fake samples $[s_t, a_t, s'_{t+1}]$ and make labels.
  end
end
Divide the training set and test set.
for *epoch= 1 to* E_4 **do**
  Sample the mini-batch from the training data.
  Obtain the output of the adversarial sample discriminator.
  Calculate the loss function and update the adversarial sample discriminator.
end

Algorithm 3. The training process of the adversarial sample discriminator.

TABLE 1 Buyer's average cost per round.

Number of buyers	5	10	15	20
Buyers' average cost (DQN)	6.5649	6.4973	5.0431	4.8707
Buyers' average cost (random)	29.7003	31.5305	32.2234	32.5878

TABLE 2 Seller's average profit per round.

Number of sellers	5	10	15	20
Sellers' average profit (DQN)	11.0802	11.2976	11.4352	11.5323
Sellers' average profit (random)	8.1247	8.0215	7.9866	7.9864

6 Performance evaluation

In this section, we conduct several comprehensive evaluations to verify the performance of our proposed method. In the following, first the evaluation settings are given. Then the results of our proposed method is introduced. Finally, the comparison results are shown.

In this section, we conduct several comprehensive evaluations to verify the performance of our proposed method. In the following, first the evaluation settings are given. Then the results of our proposed method is introduced. Finally, the comparison results are shown.

6.1 Evaluation settings

6.1.1 Environment settings

Consider a microgrid in which energy trading is performed 10 times per round, that is, each round of bilateral auction is divided into 10 time slots, and 8,000 rounds of bilateral auction are conducted to train the deep Q-learning algorithm. In order

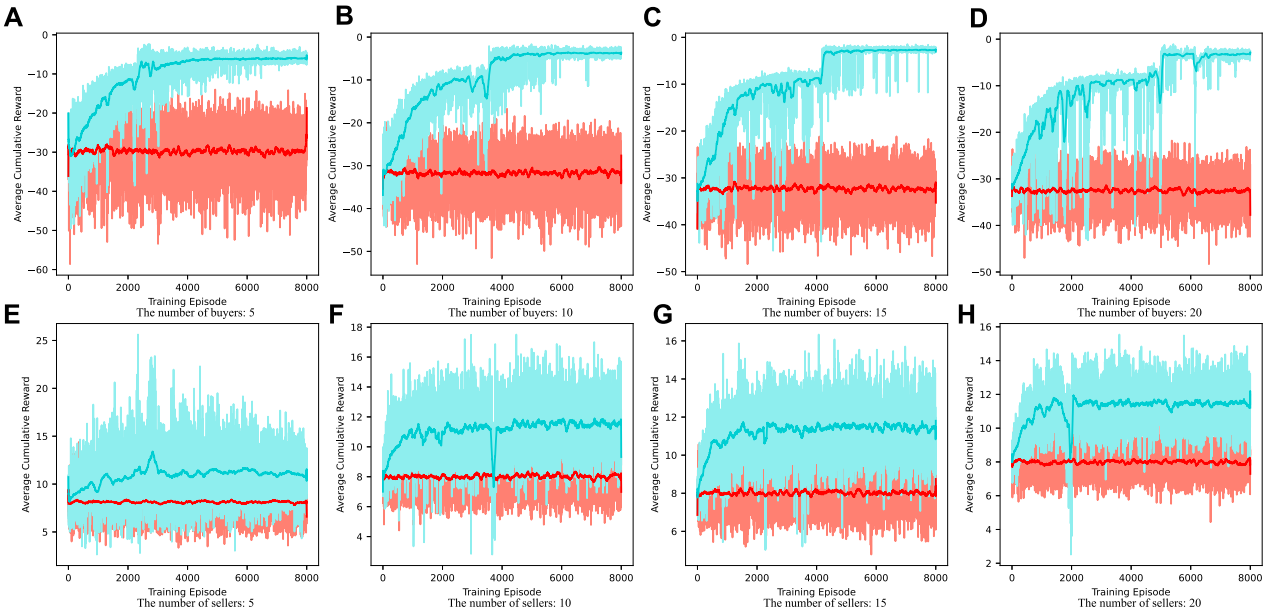

FIGURE 5
Convergence process of deep-Q-network with the numbers of buyers are **(A)** 5; **(B)** 10; **(C)** 15; **(D)** 20, and the numbers of sellers are **(E)** 5; **(F)** 10; **(G)** 15; **(H)** 20.

TABLE 3 Success rate of adversarial attack.

Magnitude of perturbation		0.1 (%)	0.2 (%)	0.3 (%)	0.4 (%)	0.5 (%)	0.6 (%)	0.7 (%)	0.8 (%)	0.9 (%)
Buyers' Q-network	5 buyer	43.3	48.1	47.3	47.7	46	47.7	47.8	46.7	45.6
	10 buyer	34.6	33.8	33.6	34.5	32.5	36.3	31.8	32.1	32.3
	15 buyer	47.1	49.6	49.6	48.5	50	48.5	48.7	46.5	48.2
	20 buyer	49.3	52.1	47.5	48.1	48.8	50.3	49	46.9	49
Sellers' Q-network	5 seller	23.5	36	44.4	48.2	50.1	52.5	51.9	49.5	47.8
	10 seller	39.5	48	56.7	59.7	57.1	62.2	60.3	60.4	62.3
	15 seller	39.9	56.4	61.1	63.5	67.5	66.8	66.2	66.2	66.7
	20 seller	45.4	59.4	61.7	66.7	65.8	66.4	68.8	66.7	68.8

to make the simulation fit the actual transaction as much as possible and avoid the dimension explosion problem, this paper discretizes the bid price and bid volume of the buyer and seller. The bid price is selected from $[0.6, 1.5]$ with a spacing of .1, a total of 10 bid price schemes, and the bid quantity is selected from $[0.5, 5]$ with a spacing of .5, a total of 10 bid volume schemes. In order to facilitate the simulation, the number of EVs of the buyer is assumed to be equal to the number of EVs of the seller in each training process, and the number of the two parties is considered to be 5, 10, 15 and 20 respectively. In each round of 10 auctions, the emerging demand or supply generated by each participant is a discrete number chosen from the set $(0.5, 1.5]$. Assuming that the unmet demand or supply from the previous step will be inherited to the next auction with an inheritance rate of .9, then

$$d_{i,t+1} = 0.9\left(d_{i,t} - w_{b,i,t}\right) + \varphi, \quad i \in B \tag{19}$$

$$u_{j,t+1} = 0.9\left(u_{j,t} - w_{s,j,t}\right) + \varphi, \quad j \in S \tag{20}$$

where, $w_{b,i,t}$ and $w_{s,j,t}$ respectively represent the transaction volume of the *ith* buyer and the *jth* seller in time slot t, and the value of φ satisfies the uniform distribution on $(0.5, 1.5]$.

6.1.2 Reinforcement learning settings

For the deep Q-learning algorithm, the learning rate is set to .001, the buyer discount rate is set to .99, the seller discount rate is set to .7, and the time interval for replacing the target

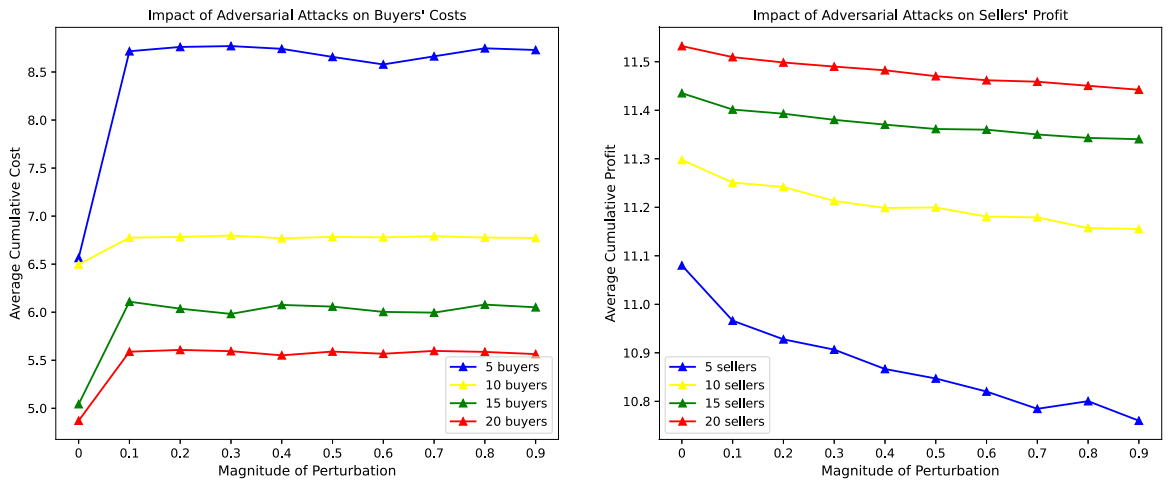

FIGURE 6
Effect of adversarial attacks.

network parameter θ' with the deep-Q-network parameter θ is set to 5. The size of the experience replay buffer is set to 3,000 for both buyer and seller, and the mini-batch size B sampled from it during training is set to 32. The input layer of the deep-Q-network is set to four neurons, the output layer is set to 100 neurons, and the number of neurons in the four hidden layers is 20, 512, 256 and 128, respectively. The greedy coefficient satisfies the following relation:

$$\varepsilon = \varepsilon_2 + (\varepsilon_1 - \varepsilon_2)\, e^{-\frac{t}{8000}} \tag{21}$$

where ε_1 and ε_2 are the values of .99 at the beginning of training and 0 at the end of training, respectively.

6.2 Effectiveness analysis of deep Q-Learning algorithms

In the case of different number of participants, deep Q-learning algorithm and random strategy are respectively used to compare the average cost per round of buyers and the average profit per round of sellers in the last 1,000 rounds. The results are shown in Table 1 and Table 2.

It can be seen from Table 1 and Table 2 that buyers and sellers can obtain more significant benefits when making bidding decisions based on deep-Q-network compared with random strategy. The average cost of buyers is the negative value of the cumulative reward in each round. It can be seen from Table 1 that under the deep Q-learning algorithm, with the increase of the number of buyers, the average cost of buyers participating in electric energy trading will also decrease. The average profit of sellers is the cumulative reward in each round. It can be seen from Table 2 that under the deep Q-learning algorithm, with the increase of the number of sellers, the average profit of sellers participating in electric energy trading will also rise. This shows the effectiveness of the algorithm and fully considers the willingness and interests of the participants. It can also encourage EV users to participate in the electric energy trading market and contribute to the peak regulation of the power grid. The convergence process of deep-Q-network training is shown in Figure 5.

6.3 Effectiveness evaluation of adversarial attacks

The effectiveness evaluation of adversarial attacks can be considered from two aspects. One is the success rate, and the other is the extent to which adversarial attacks affect participants' utilities. For the setting of user states in this paper, the attacker affects other users' state observations mainly by maliciously modifying its own demand/supply. Therefore, in local-FGSM for buyer-Q-network, the values in the first two dimensions of vector u are one and the rest are 0. The value of the first and third dimensions of the vector u of local-FGSM for the seller-Q-network is 1.

When attacking the buyer Q-network, a buyer is selected as the attacker in each auction, and its state is modified to affect the state observation of other buyers, then a non-optimal bidding strategy is selected to participate in the bilateral auction. Similarly, the seller Q-network is also attacked. If the buyer's average cumulative cost per turn increases or the seller's average cumulative profit per turn decreases, the attack is successful. The success rate against the attack is shown in Table 3.

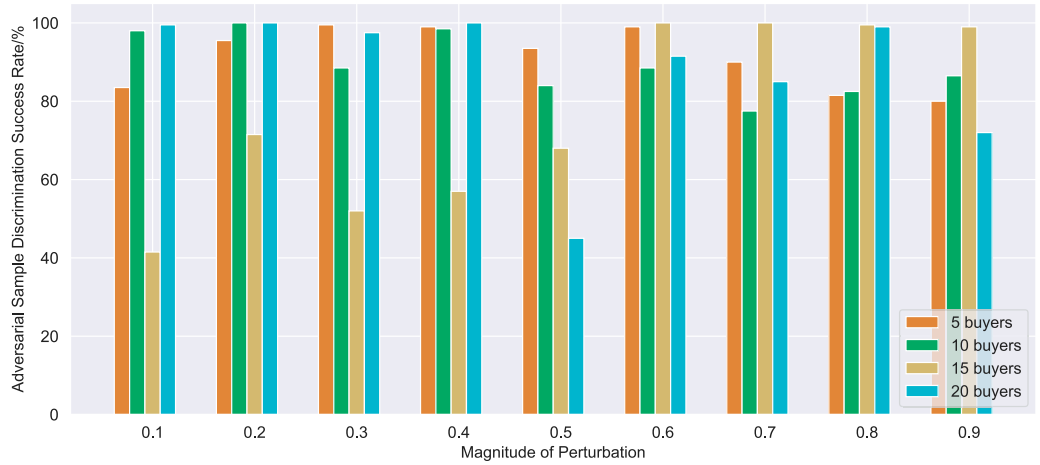

FIGURE 7
Effect of adversarial sample discriminator for buyer Q-Network.

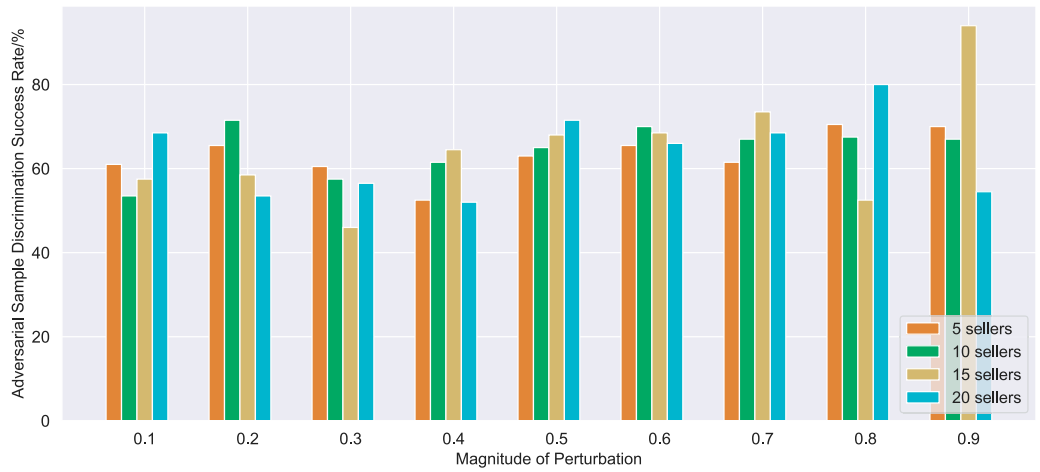

FIGURE 8
Effect of adversarial sample discriminator for seller Q-Network.

The impact of adversarial attacks on user benefits is shown in Figure 6. As can be seen from Figure 5, when adversarial samples are added, the average cumulative cost of buyers per round increases, especially when the number of buyers is small, the impact is greater. When adversarial samples are added, the average cumulative profit of sellers in each round decreases, and with the increase of disturbance size, the profit becomes lower and lower. When the number of sellers is small, the profit decreases more significantly.

6.4 Effectiveness evaluation of defense strategy

The adversarial sample discriminator is trained by supervised learning, and the well-trained buyer Q-network and seller Q-network are selected to collect 20,000 real samples and 4,000 adversarial samples in the process of adversarial attack for 2000 training times. The learning rate is set to .001. The final defense effect of the adversarial sample discriminator is shown in Figure 7 and Figure 8.

It can be seen from Figure 7 and Figure 8 that the adversarial defense method proposed in this paper can achieve defense effect in most cases. The buyer adversarial sample discriminator has a good screening effect on adversarial samples with different disturbance sizes, and can basically achieve a screening success rate of more than 80% in trading scenarios with different number of buyers. Compared with buyer adversarial sample discriminator, seller adversarial sample discriminator has a poor performance, but the success rate of adversarial sample screening generally reaches more than 60%, and it can also play a good adversarial defense effect in most cases.

7 Conclusion

In this paper, focusing the EV double auction market, we study the security issue of bidding strategy based on reinforcement learning raised by adversarial example. First, we construct a Markov Decision Process for EV energy trading, and use DQN to solve this problem. Second, we design a local-fast gradient sign method to try to counter attacks on DQN from the perspective of attackers. Third, from the perspective of defenders, we choose the method of adding additional network, and use the deep neural network to build the adversarial example discriminator to screen the adversarial example. Finally, the simulation results shows that adversarial example would have an impact on the deep reinforcement learning algorithm, and different disturbance sizes will have different degrees of negative impact on market profits. While after adding the discriminant network, it can almost completely resist such attacks.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DL: Conceptualization, Methodology, Investigation, Results Analysis, Writing—Original Draft; QY: Conceptualization, Supervision, Writing—Review and Editing; ZP: Survey of Methods, Simulation; XL: Results Analysis; LM: Data Processing, Writing—Review and Editing.

Funding

The work was supported in part by Key Research and Development Program of Shaanxi under Grants 2022GY-033, in part by the National Science Foundation of China under Grants 61973247 and 61673315, in part by China Postdoctoral Science Foundation 2021M692566, in part by the operation expenses for universities' basic scientific research of central authorities xzy012021027.

Conflict of interest

Authors LM and XL were employed by the company State Grid Information and Telecommunication Group Co., LTD, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Albadi, M. H., and El-Saadany, E. F. (2007). "Demand response in electricity markets: An overview," in 2007 IEEE power engineering society general meeting (IEEE), Tampa, FL, USA, 24-28 June 2007, 1–5. doi:10.1109/PES.2007.385728

An, D., Zhang, F., Yang, Q., and Zhang, C. (2022). Data integrity attack in dynamic state estimation of smart grid: Attack model and countermeasures. *IEEE Trans. Automation Sci. Eng.* 19, 1631–1644. doi:10.1109/TASE.2022.3149764

Bandyszak, T., Daun, M., Tenbergen, B., Kuhs, P., Wolf, S., and Weyer, T. (2020). Orthogonal uncertainty modeling in the engineering of cyber-physical systems. *IEEE Trans. Automation Sci. Eng.* 17, 1–16. doi:10.1109/TASE.2020.2980726

Barto, A. G., Sutton, R. S., and Watkins, C. (1989). *Learning and sequential decision making*. Amherst, MA: University of Massachusetts.

Cheng, J., Chu, F., and Zhou, M. (2018). An improved model for parallel machine scheduling under time-of-use electricity price. *IEEE Trans. Automation Sci. Eng.* 15, 896–899. doi:10.1109/TASE.2016.2631491

Croce, D., Giuliano, F., Tinnirello, I., Galatioto, A., Bonomolo, M., Beccali, M., et al. (2017). Overgrid: A fully distributed demand response architecture based on overlay networks. *IEEE Trans. Automation Sci. Eng.* 14, 471–481. doi:10.1109/TASE.2016.2621890

Ding, J.-Y., Song, S., Zhang, R., Chiong, R., and Wu, C. (2016). Parallel machine scheduling under time-of-use electricity prices: New models and optimization approaches. *IEEE Trans. Automation Sci. Eng.* 13, 1138–1154. doi:10.1109/TASE.2015.2495328

Erhel, S., and Jamet, E. (2016). The effects of goal-oriented instructions in digital game-based learning. *Interact. Learn. Environ.* 24, 1744–1757. doi:10.1080/10494820.2015.1041409

Esmat, A., de Vos, M., Ghiassi-Farrokhfal, Y., Palensky, P., and Epema, D. (2021). A novel decentralized platform for peer-to-peer energy trading market with blockchain technology. *Appl. Energy* 282, 116123. doi:10.1016/j.apenergy.2020.116123

Giaconi, G., Gündüz, D., and Poor, H. V. (2018). Smart meter privacy with renewable energy and an energy storage device. *IEEE Trans. Inf. Forensics Secur.* 13, 129–142. doi:10.1109/TIFS.2017.2744601

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. arXiv preprint arXiv:1412.6572.

Grigsby, L. L. (2007). *Electric power generation, transmission, and distribution*. Boca Raton: CRC Press.

Hahn, R. W., and Stavins, R. N. (1991). Incentive-based environmental regulation: A new era from an old idea. *Ecol. LQ* 18, 1.

Haller, M., Ludig, S., and Bauer, N. (2012). Bridging the scales: A conceptual model for coordinated expansion of renewable power generation, transmission and storage. *Renew. Sustain. Energy Rev.* 16, 2687–2695. doi:10.1016/j.rser.2012.01.080

Hong, Z., Wang, R., Ji, S., and Beyah, R. (2019). Attacker location evaluation-based fake source scheduling for source location privacy in cyber-physical systems. *IEEE Trans. Inf. Forensics Secur.* 14, 1337–1350. doi:10.1109/TIFS.2018.2876839

Hosseini, S. M., Carli, R., and Dotoli, M. (2021). Robust optimal energy management of a residential microgrid under uncertainties on demand and renewable power generation. *IEEE Trans. Automation Sci. Eng.* 18, 618–637. doi:10.1109/TASE.2020.2986269

Huang, S., Papernot, N., Goodfellow, I., Duan, Y., and Abbeel, P. (2017). *Adversarial attacks on neural network policies*. arXiv preprint arXiv:1702.02284.

Huang, W., Zhang, N., Kang, C., Li, M., and Huo, M. (2019). From demand response to integrated demand response: Review and prospect of research and application. *Prot. Control Mod. Power Syst.* 4, 12–13. doi:10.1186/s41601-019-0126-4

Jin, C., Tang, J., and Ghosh, P. (2013). Optimizing electric vehicle charging with energy storage in the electricity market. *IEEE Trans. Smart Grid* 4, 311–320. doi:10.1109/tsg.2012.2218834

Kim, H., Lee, J., Bahrami, S., and Wong, V. W. (2019). Direct energy trading of microgrids in distribution energy market. *IEEE Trans. Power Syst.* 35, 639–651. doi:10.1109/tpwrs.2019.2926305

Lange, S., and Riedmiller, M. (2010). "Deep auto-encoder neural networks in reinforcement learning," in The 2010 international joint conference on neural networks (IJCNN), Barcelona, Spain, 18-23 July 2010 (IEEE), 1–8. doi:10.1109/IJCNN.2010.5596468

Lin, Y.-C., Hong, Z.-W., Liao, Y.-H., Shih, M.-L., Liu, M.-Y., and Sun, M. (2017). *Tactics of adversarial attack on deep reinforcement learning agents*. arXiv preprint arXiv:1703.06748.

Liu, P., Zang, W., and Yu, M. (2005). Incentive-based modeling and inference of attacker intent, objectives, and strategies. *ACM Trans. Inf. Syst. Secur. (TISSEC)* 8, 78–118. doi:10.1145/1053283.1053288

Liu, Y., Duan, J., He, X., and Wang, Y. (2018). Experimental investigation on the heat transfer enhancement in a novel latent heat thermal storage equipment. *Appl. Therm. Eng.* 142, 361–370. doi:10.1016/j.applthermaleng.2018.07.009

Miao, L., Wen, J., Xie, H., Yue, C., and Lee, W.-J. (2015). Coordinated control strategy of wind turbine generator and energy storage equipment for frequency support. *IEEE Trans. Industry Appl.* 51, 2732–2742. doi:10.1109/tia.2015.2394435

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). *Playing atari with deep reinforcement learning*. arXiv preprint arXiv:1312.5602.

Mohan, S., and Laird, J. (2014). "Learning goal-oriented hierarchical tasks from situated interactive instruction," in Proceedings of the AAAI Conference on Artificial Intelligence. doi:10.1609/aaai.v28i1.8756

Ng, K.-H., and Sheble, G. B. (1998). Direct load control-a profit-based load management using linear programming. *IEEE Trans. Power Syst.* 13, 688–694. doi:10.1109/59.667401

Pyka, A. (2002). Innovation networks in economics: From the incentive-based to the knowledge-based approaches. *Eur. J. Innovation Manag.* 5, 152–163. doi:10.1108/14601060210436727

Qu, X., Sun, Z., Ong, Y.-S., Gupta, A., and Wei, P. (2020). Minimalistic attacks: How little it takes to fool deep reinforcement learning policies. *IEEE Trans. Cognitive Dev. Syst.* 13, 806–817. doi:10.1109/tcds.2020.2974509

Roijers, D. M., Vamplew, P., Whiteson, S., and Dazeley, R. (2013). A survey of multi-objective sequential decision-making. *J. Artif. Intell. Res.* 48, 67–113. doi:10.1613/jair.3987

Ruiz, N., Cobelo, I., and Oyarzabal, J. (2009). A direct load control model for virtual power plant management. *IEEE Trans. Power Syst.* 24, 959–966. doi:10.1109/tpwrs.2009.2016607

Samadi, P., Mohsenian-Rad, A.-H., Schober, R., Wong, V. W., and Jatskevich, J. (2010). "Optimal real-time pricing algorithm based on utility maximization for smart grid," in 2010 First IEEE International Conference on Smart Grid Communications, Gaithersburg, MD, USA, 04-06 October 2010 (IEEE), 415–420. doi:10.1109/SMARTGRID.2010.5622077

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., et al. (2013). *Intriguing properties of neural networks*. arXiv preprint arXiv:1312.6199.

Wan, Z., Li, H., He, H., and Prokhorov, D. (2018). Model-free real-time ev charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid* 10, 5246–5257. doi:10.1109/tsg.2018.2879572

Wu, Y., Tan, X., Qian, L., Tsang, D. H., Song, W.-Z., and Yu, L. (2015). Optimal pricing and energy scheduling for hybrid energy trading market in future smart grid. *Ieee Trans. industrial Inf.* 11, 1585–1596. doi:10.1109/tii.2015.2426052

Yu, B., Lu, J., Li, X., and Zhou, J. (2022). Salience-aware face presentation attack detection via deep reinforcement learning. *IEEE Trans. Inf. Forensics Secur.* 17, 413–427. doi:10.1109/TIFS.2021.3135748

Zeng, M., Leng, S., Maharjan, S., Gjessing, S., and He, J. (2015). An incentivized auction-based group-selling approach for demand response management in v2g systems. *IEEE Trans. Industrial Inf.* 11, 1554–1563. doi:10.1109/tii.2015.2482948

Zhang, D., Han, X., and Deng, C.Taiyuan University of Technology, and China Electric Power Research Institute (2018). Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* 4, 362–370. doi:10.17775/cseejpes.2018.00520

Zhang, K., Sprinkle, J., and Sanfelice, R. G. (2016). Computationally aware switching criteria for hybrid model predictive control of cyber-physical systems. *IEEE Trans. Automation Sci. Eng.* 13, 479–490. doi:10.1109/TASE.2016.2523341

Zhang, W., Song, K., Rong, X., and Li, Y. (2019). Coarse-to-fine uav target tracking with deep reinforcement learning. *IEEE Trans. Automation Sci. Eng.* 16, 1522–1530. doi:10.1109/TASE.2018.2877499

Zhao, S., Li, F., Li, H., Lu, R., Ren, S., Bao, H., et al. (2021). Smart and practical privacy-preserving data aggregation for fog-based smart grids. *IEEE Trans. Inf. Forensics Secur.* 16, 521–536. doi:10.1109/TIFS.2020.3014487

Zhou, R., Li, Z., and Wu, C. (2015). "An online procurement auction for power demand response in storage-assisted smart grids," in 2015 IEEE Conference on Computer Communications (INFOCOM), Hong Kong, China, 26 April 2015 - 01 May 2015 (IEEE), 2641–2649. doi:10.1109/INFOCOM.2015.7218655