Hindawi
Security and Communication Networks
Volume 2021, Article ID 4773894, 15 pages
https://doi.org/10.1155/2021/4773894

Research Article

Defense Strategy Selection Model Based on Multistage Evolutionary Game Theory

Yanhua Liu, Hui Chen, Hao Zhang, and Ximeng Liu

College of Computer and Data Science, Fuzhou University, Fujian 350108, Fuzhou, China

Correspondence should be addressed to Yanhua Liu; lyhwa@fzu.edu.cn

Received 31 August 2021; Revised 18 October 2021; Accepted 25 October 2021; Published 11 November 2021

Academic Editor: Wenjuan Li

Evolutionary game theory is widely applied in network attack and defense. The existing network attack and defense analysis methods based on evolutionary games adopt the bounded rationality hypothesis. However, the existing research ignores that both sides of the game get more information about each other with the deepening of the network attack and defense game, which may cause the attacker to crack a certain type of defense strategy, resulting in an invalid defense strategy. The failure of the defense strategy reduces the accuracy and guidance value of existing methods. To solve the above problem, we propose a reward value learning mechanism (RLM). By analyzing previous game information, RLM automatically incentives or punishes the attack and defense reward values for the next stage, which reduces the probability of defense strategy failure. RLM is introduced into the dynamic network attack and defense process under incomplete information, and a multistage evolutionary game model with a learning mechanism is constructed. Based on the above model, we design the optimal defense strategy selection algorithm. Experimental results demonstrate that the evolutionary game model with RLM has better results in the value of reward and defense success rate than the evolutionary game model without RLM.

1. Introduction

The rapid development of IT infrastructures, such as cyber-physical systems and Internet of Things, has brought convenience to individuals and enterprises. But it also brings unprecedented security problems [1, 2]. Data management and communication layers in cyber-physical systems and the Internet of Things are vulnerable to cyberattacks such as DDoS attacks, APT, and vulnerability attacks, which seriously threaten network security [3]. According to the Crystal Market Research (CMR) report, to resist the increasingly severe attacks, the investment in the network security market is expected to increase from $ 58.13 billion to $ 173.57 billion from 2012 to 2022 [4]. Thus, it can be seen that the network attack and defense are increasingly severe, and network security defense has become an important problem to be solved in the field of network information [5]. Unfortunately, much research [6, 7] shows that improving information security technology alone cannot provide enough protection against persistent attacks. A new method is needed to guide the implementation of the defense strategy.

Network attack and defense have the characteristics of opposite objectives and noncooperation relationships, which are consistent with the characteristics of the game. The game theory [8] applied to the network attack and defense process, from the perspective of the defender exploring how to get the best defense strategy, has become an important research direction of network security defense [9–12]. Traditional game theory assumes that both sides of the game are in a complete information game scene and are required to be rational completely [13, 14]. Complete information requires the game players to know the information of the entire environment [15]. Complete rationality assumes that game players can choose their own best game strategy after obtaining the other's strategy and its revenue [16]. Evolutionary game theory [17, 18] starts from the condition of opaque information of players, takes the learning mechanism as the core, and influences the selection behavior of players through factors such as previous experience, learning, and imitating the behavior of others. Evolutionary game theory can better express the process of the mutual game between the attackers and defenders. It is widely used in the research work of the network attack and defense game [19–21].

However, there are still some problems and challenges in applying evolutionary games in network attack and defense. (1) Existing studies using evolutionary game theory have introduced some relevant parameters to express evolutionary game ideas under incomplete information and bounded rationality [22, 23]. The use of these parameters is feasible in certain application scenes and shows a certain application value. However, the introduced parameters are calculated manually and need to be quantified by experts. There is currently no automatic calculation method. (2) In the multistage network attack and defense game, network attacks develop over time using new methods and targets [24], and the defense strategy may partially or completely fail. The existing evolutionary game model cannot effectively feed the failure information back to the next game stage, which leads to the shortcomings of the best defense strategy selection algorithm in terms of timeliness, accuracy, and efficiency. (3) Network attack and defense are a dynamic, multistage process [25]. The confrontation between attack and defense is not limited to one round or a certain stage.

To address the above problems, we propose a reward value learning mechanism (RLM), a novel method for updating the reward value based on the game information in the previous stage. Inspired by machine learning, we use RLM to learn the attacker's strategy and the change of its reward to predict the return value in the next stage. Furthermore, RLM is introduced to the evolutionary game model to select the best defense strategy for each stage. The main research work and contributions are as follows:

(1) Aiming at the multistage network attack and defense game scene, a reward value learning mechanism is designed. By updating the reward value of each stage, the mechanism improves the active defense ability when the defense strategy fails.

(2) Construct a new multistage evolutionary game model with a reward value learning mechanism, and solve the Nash equilibrium of each stage of the attack and defense game by constructing the replication dynamic equations (RD) of each stage.

(3) Based on the proposed multistage evolutionary game model, an optimal defense strategy selection algorithm is proposed. Experiments show that the algorithm can effectively improve the effective strategy selection probability, defense revenue, and defense success rate in a single defense strategy failure scene and multiple defense strategy failure scene.

The remainder of this paper is organized as follows. Section 2 briefly reviews evolutionary game theory and the related literature in network attack and defense. Section 3 describes the evolutionary game model based on Q-learning replication dynamic equations (QRD). Section 4 designs RLM and gives a new evolutionary game model based on RLM-QRD. Section 5 proposes the optimal defense strategy selection algorithm. Section 6 presents the experimental results for evaluating our model and compares it to the model without RLM. Finally, Section 7 concludes this paper and discusses the future works.

2. Related Works

The existing network defense works based on evolutionary games mainly include network defense based on the static evolutionary game and network defense based on the dynamic evolutionary game. The following are two aspects.

2.1. Static Network Defense Evolutionary Game. The static game assumes that the information of both sides of the game remains unchanged, and it is a one-shot game [26].

Ruan et al. [27] established the attack and defense evolutionary game model using the lightweight broadcast authentication protocol to achieve security assurance and minimum resource cost. Abdalzaher et al. [28] used the scalability and low complexity of wireless sensor networks to propose a trust model that uses evolutionary game theory to make decisions to resist network attacks. Bouhaddi et al. [29] established a Bayesian game model to analyze the interaction between defenders and potential malicious nodes in the network. Aimed at the problems of free service users and users breaking system rules in peer-to-peer networks, Shareh et al. [30] established an evolutionary game model to resist network attacks from these two types of users. To explore and calculate more revenue of existing defense strategies, Jin et al. [31] combined Q-learning with replication dynamics equation to obtain Q-learning replication dynamics equation and proposed an evolutionary game model based on QRD. Aiming at the problem of the limited learning ability of players in the static network attack and defense, Liu et al. [32] established a network attack-defense evolutionary game model and designed an optimal defense strategy selection algorithm. Shi et al. [33] proposed an evolutionary game model based on honeypot technology to improve the security of the honeypot system.

2.2. Dynamic Network Defense Evolutionary Game. Compared with the static network attack and defense evolutionary game, the dynamic network attack and defense evolutionary game divides the game into multistage of attack and defense confrontation between the players, which is more in line with the network attack and defense.

To realize the optimal defense decision in network attack and defense, Huang et al. proposed two dynamic evolutionary game models. One of the game models is the Markov-based time game model. This model selects the best defense strategy by constructing a revenue discount factor and all possible network system states [34]. Another game model used the best-response dynamic learning mechanism to study the evolutionary law of network defense strategy selection [35].

To resist the invasion of the virus code, Hayel and Zhu [36] established an evolutionary Poisson game model by defining the number of players participating in the interaction to follow a Poisson process at a specific rate. Aiming at the security issues in radio networks, Fang et al. [37] established an evolutionary game model, using evolutionary stable strategy algorithms to defend dynamically against internal attacks. Mengibaev et al. [22] introduced parameters used to measure the dependence of game players on opponents into the evolutionary game model and applied them to the privacy protection of network users. Wang et al. [38] designed three evolutionary game models for different attack scenes in the dynamic network and introduced parameters that denote the degree of sensitivity of the players to the difference in revenue in these models.

Hu et al. proposed different dynamic evolutionary game models for the problems in network attack and defense. In [39], a multistage Bayesian attack and defense evolutionary game model is proposed for the difficulty of selecting the optimal defense strategy in a dynamic confrontation network. At the same time, the selection intensity factor was introduced to improve the replication dynamic equation and enhance the randomness of the evolution process. More recently, to improve the timeliness and predictability of network attack and defense game, Hu et al. [40] proposed a dynamic evolutionary game model based on Logit Quantal Response Dynamics (LQRD), which introduced parameters into the evolutionary game to describe the rationality of attack and defense sides.

3. Preliminary

This section first introduces Q-learning and then replication dynamic equations. Finally, the definition of the evolutionary game model is proposed.

3.1. Q-Learning. Q-learning [41] is a reinforcement learning method, which can be regarded as an asynchronous dynamic programming method. Q-learning is also an adaptive value iteration method, which is based on the state-action value $Q_t(s', a')$ and guides the estimation of the state-action value $Q_{t+1}(s, a)$ at time $t+1$. Among them, the state-action value $Q_t(s, a)$ is the expected revenue after action a is taken by state s at time t. State s' is the state the learner reaches after using action a in state s. The Q-learning formula is given as follows:

$$Q_{t+1}(s,a) \longleftarrow (1-\partial)Q_t(s,a) + \partial\left(r + \gamma \max_{a'} Q_t(s', a')\right). \tag{1}$$

∂ is the usual step size parameter, r is immediate reinforcement, and γ is the discount factor.

The principle of Q-learning is to move in a discrete, finite state and select one from a finite set of actions every time, forming a controlled Markov process. Continuously, it aims to improve its evaluation of the quality of specific actions in a specific state to find a strategy with the most profit, which is consistent with the game's goal.

3.2. Replication Dynamic Equation. The replication dynamic equation is a dynamic differential equation. It describes the frequency or probability of a certain strategy used in a specific group of people [42] and the degree to which the probability of the game's main body choosing a strategy during the game. Its basic principle is that the game players gradually adopt more strategies with a revenue better than the average revenue. In addition, the replication dynamic equation can ensure that the evolutionary stable strategy is the Nash equilibrium, thereby obtaining the strategy that benefits the most. The replication dynamic equations of attackers and defenders in network attack and defense are given as follows:

$$x_i'(t) = \frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i\left[Q_{\mathrm{AS}_i} - \overline{Q}_{\mathrm{AS}}\right]. \tag{2}$$

$$y_j'(t) = \frac{\mathrm{d}y_j}{\mathrm{d}t} = y_j\left[Q_{\mathrm{DS}_j} - \overline{Q}_{\mathrm{DS}}\right]. \tag{3}$$

In the previous formulas, $x_i'(t)$ represents the change rate of the probability of the attacker selecting the attack strategy AS_i over time, $y_j'(t)$ represents the change rate of the probability of the defender selecting the defense strategy DS_j over time, Q_{AS_i} represents the expected revenue of the attacker's selection of the attack strategy AS_i, Q_{DS_j} represents the expected revenue of the defender's selection of the defense strategy DS_j, $\overline{Q}_{\mathrm{AS}}$ represents the average revenue of the attack strategy set, and $\overline{Q}_{\mathrm{DS}}$ represents the average revenue of the defense strategy set.

3.3. Definition of Evolutionary Game Model Based on QRD. To extend the evolutionary game model to the dynamic network environment, this section introduces the stage definition into the evolutionary game model based on QRD. The model is defined below.

Definition 1. The evolutionary game model based on QRD is represented as 6 tuples, and its elements are defined as follows:

(1) $N = (N_A, N_D)$ is the space of both sides in the game, and N_A, N_D are the attacker and the defender.

(2) K is the number of stages in the multistage attack and defense game; $K = 1, 2, 3, \ldots, T$.

(3) $S = (\mathrm{AS}, \mathrm{DS})$ is the strategy space of both sides in the game, where $\mathrm{AS} = (\mathrm{AS}_1, \mathrm{AS}_2, \ldots, \mathrm{AS}_n)$ is the strategy set of the attacker, AS^K is the strategy set of the attacker in stage K, $\mathrm{DS} = (\mathrm{DS}_1, \mathrm{DS}_2, \ldots, \mathrm{DS}_m)$ is the strategy set of the defender, and DS^K is the strategy set of the defender in stage K. n and m represent the number of strategies of the attacker and the defender, respectively. Therefore, both n and m are integers and $n \geq 2, m \geq 2$.

(4) $\theta = (P_A, P_D)$ is the belief set of attack and defense game, $P_A = (x_1, x_2, \ldots, x_n)$ is the probability

distribution of the attacker's overall strategy set AS; that is, any $x_i \in P_A$ is the attacker's choice of strategy AS_i with probability x_i to implement network attack, and $P_D = (y_1, y_2, \ldots, y_m)$ is the probability distribution of the defender's overall strategy set DS; that is, any $y_j \in P_D$ is the defender's choice of strategy DS_j with probability y_i to implement defense strategy. In addition, the parameters satisfy the relationships $1 \le i \le n$, $1 \le i \le n$, $\sum_{i=1}^{n} x_i = 1$, $\sum_{j}^{m} y_j = 1$.

(5) $Q = (Q_A, Q_D)$ is the revenue function set of the attack-defense game, and Q_A and Q_D represent the revenue function of the attacker and the defender, respectively, that is, the revenue of the attacker and the defender from the game strategy combination $(\mathrm{AS}_i, \mathrm{DS}_j)$.

(6) τ is the exploration factor of both sides of the game, indicating the degree of their exploration of game information. The larger the τ is, the greater the degree of exploration is and the more the attack and defense sides explore the unknown game information and make better decisions. The smaller the τ is, the smaller the degree of exploration is. The attack and defense sides mainly make the best decision based on the current known game information.

4. Game Model Based on RLM-QRD

To solve the problem of invalidation of specific defense strategies in network attack and defense scene, we put forward RLM with incentive and punishment mechanisms. RLM uses parameter α to calculate the attack and defense reward value in the next stage.

4.1. Definition of Game Model. By combining RLM with the evolutionary game based on QRD [31], this paper designs an evolutionary game model based on RLM-QRD; that is, $\mathrm{RG} = (N, K, S, \theta, Q, \tau, \alpha)$. N, K, S, θ, Q, and τ have been defined, and the definition of α is given below.

Definition 2. α is the incentive and punishment factor of reward value, which means the reward value of the corresponding strategy combination should be stimulated or punished when RLM is triggered. The value of α affects the probability of multistage strategy selection.

In the first stage of the game, the incentive and punishment factor α formula of reward value is as follows:

$$\alpha = \frac{1}{2} \times \mathrm{RV}. \tag{4}$$

In stage K of the game, the formula of incentive and punishment factor α is as follows:

$$\alpha = \frac{\mathrm{AN}}{\mathrm{SN}} \times \mathrm{RV}. \tag{5}$$

The parameters in formulas (4) and (5) are defined as follows:

(1) RV is the reward variable, which represents the largest variable of the reward value in a single stage. Its value is determined by the influence of the other player's strategy on itself. In general, RV is equal to the minimum reward value of the strategy combination of the defender.

(2) SN is the maximum number of learning stages of RLM, which denotes the maximum number of the learning stages the defender can learn from the previous game. Therefore, the maximum value of SN is the maximum number of stages of game T.

(3) AN is the number of a specific attack strategy in the past SN stages. If the attacker implements strategy AS_i in the previous stage, then AN is equal to the number of AS_i in the previous SN stages. The formula is as follows:

$$\mathrm{AN} = \mathrm{num}(\mathrm{AS}_i). \tag{6}$$

4.2. Framework of Game Model. The proposed framework of the multistage evolutionary game model is based on RLM-QRD, as shown in Figure 1. It can be seen from Figure 1 that our model mainly includes an evolutionary game based on QRD and RLM.

The evolutionary game model based on QRD contains payoff quantification and QRD. Payoff quantification uses the information of the initial stage of the game to calculate the revenue of the attack and defense strategy in the initial stage. By QRD using the strategy revenue at the current stage, the optimal defense strategy is obtained by calculation.

RLM is responsible for connecting all stages of the game. According to the known game information, it automatically incentives or punishes the attack and defense revenue of the next stage. Based on the above methods, the model can give a better defense strategy in the next stage.

4.3. Payoff Quantification of Attack and Defense Strategy. Network attack and defense strategy and its cost-reward analysis are the basis of achieving the optimal network security defense, so reasonable attack and defense payoff quantification affect the selection of defense strategy directly, thus affecting the defense effect. Here, we give some related definitions.

Attack payoff matrix AM comprises attack revenue value a_{ij} generated by the attacker under attack and defense strategy combination $(\mathrm{AS}_i, \mathrm{DS}_j)$. According to the definition of the game model, the formula is as follows:

$$a_{\mathrm{ij}} = Q_A\left(\mathrm{AS}_i, \mathrm{DS}_j\right) = \mathrm{AR} - \mathrm{AC}. \tag{7}$$

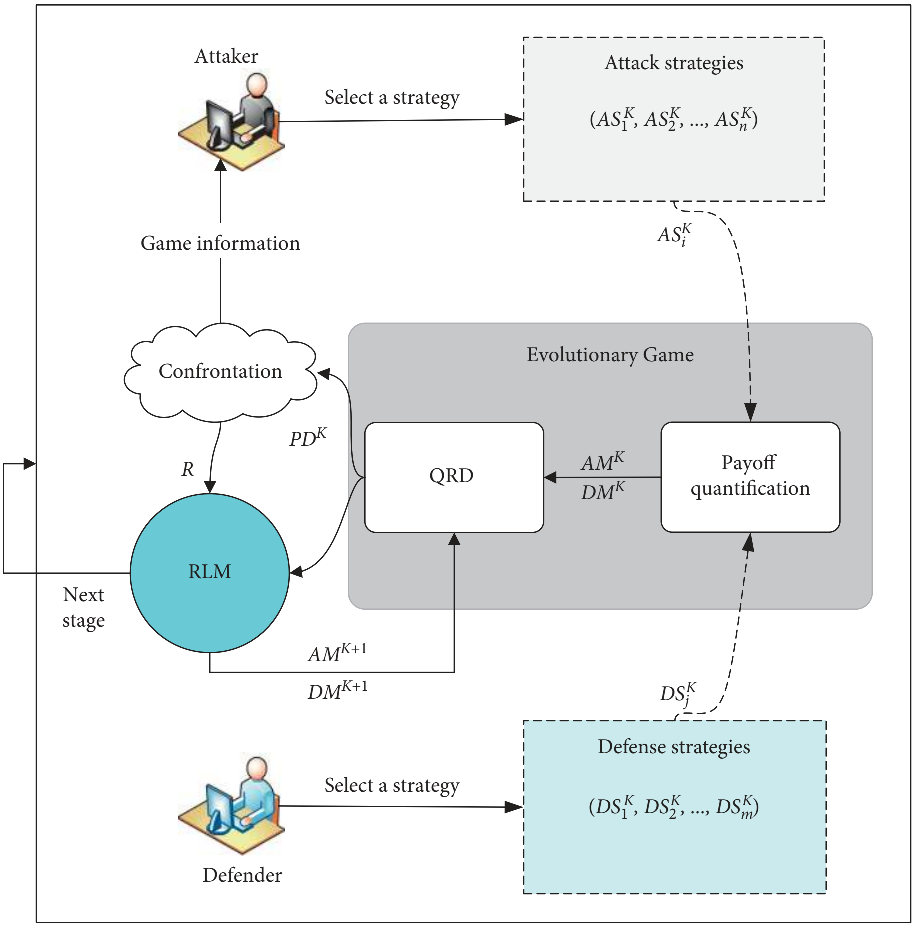


Figure 1: Framework of evolutionary game model based on RLM-QRD.

AR and AC represent attack revenue and attack cost, respectively.

The attack payoff matrix in stage K is as follows:

$$\mathrm{AM}^K = \begin{bmatrix} a_{11}^K & a_{12}^K & \cdots & a_{1m}^K \\ a_{21}^K & a_{22}^K & \cdots & a_{2m}^K \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^K & a_{n2}^K & \cdots & a_{\mathrm{nm}}^K \end{bmatrix}. \tag{8}$$

Defense payoff matrix DM comprises defense revenue value d_{ij} generated by the defender under attack and defense strategy combination $(\mathrm{AS}_i, \mathrm{DS}_j)$. According to the definition of the game model, the formula is as follows:

$$d_{\mathrm{ij}} = Q_D\left(\mathrm{AS}_i, \mathrm{DS}_j\right) = \mathrm{DR} - \mathrm{DC}. \tag{9}$$

DR and DC represent defense revenue and defense cost, respectively.

The defense payoff matrix in stage K is as follows:

$$\mathrm{DM}^K = \begin{bmatrix} d_{11}^K & d_{12}^K & \cdots & d_{1m}^K \\ d_{21}^K & d_{22}^K & \cdots & d_{2m}^K \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1}^K & d_{n2}^K & \cdots & d_{\mathrm{nm}}^K \end{bmatrix}. \tag{10}$$

4.4. Q-Learning Replication Dynamic Equation. According to the Nash equilibrium theorem [43], it can be seen that, in a game with limited players in a limited strategy set, a mixed strategy Nash equilibrium must exist, and strategy (x^*, y^*) is called a mixed strategy Nash equilibrium. When the game reaches Nash equilibrium, no player is worth changing his strategy unilaterally. In this case, if the attacker chooses strategy x^* and the defender chooses strategy y^*, the attack and defense benefits are expressed as $Q_A(x^*, y^*)$ and $Q_D(x^*, y^*)$, respectively, satisfying the following conditions:

$$
\begin{aligned}
&Q_A(x^*, y^*) \geq Q_A(x, y^*), \forall x \in P_A,\\
&Q_D(x^*, y^*) \geq Q_D(x^*, y), \forall y \in P_D.
\end{aligned} \tag{11}
$$

The following is the calculation of the expected revenue $\mathrm{Q}_{\mathrm{AS}_i}$ of attack strategy AS_i, the expected revenue $\mathrm{Q}_{\mathrm{DS}_j}$ of defense strategy DS_j, the average revenue of attack strategy set $\overline{\mathrm{Q}}_{\mathrm{AS}}$, and the average revenue of defense strategy set $\overline{\mathrm{Q}}_{\mathrm{DS}}$:

$$
\begin{aligned}
\mathrm{Q}_{\mathrm{AS}_i} &= \sum_{j=1}^{m} a_{\mathrm{ij}} y_j,\\
\mathrm{Q}_{\mathrm{DS}_j} &= \sum_{i=1}^{n} d_{\mathrm{ij}} x_i,\\
\overline{\mathrm{Q}}_{\mathrm{AS}} &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{\mathrm{ij}} x_i y_j,\\
\overline{\mathrm{Q}}_{\mathrm{DS}} &= \sum_{j=1}^{m}\sum_{i=1}^{n} d_{\mathrm{ij}} x_i y_j.
\end{aligned} \tag{12}
$$

The Boltzmann probability distribution is used to represent the attack and defense strategy, and the *Q*-learning algorithm is introduced into the replication dynamic equation to get the QRD equation. The probability of strategy selection for QRD is given as follows:

$$
x_i(k) = \frac{\exp\left(\tau \cdot \mathrm{Q}_{\mathrm{AS}_i}(k)\right)}{\sum_{l=1}^{n} \exp\left(\tau \cdot \mathrm{Q}_{\mathrm{AS}_i}(k)\right)}. \tag{13}
$$

$$
y_j(k) = \frac{\exp\left(\tau \cdot \mathrm{Q}_{\mathrm{DS}_j}(k)\right)}{\sum_{l=1}^{m} \exp\left(\tau \cdot \mathrm{Q}_{\mathrm{DS}_l}(k)\right)}. \tag{14}
$$

Here, $x_i(k)$ and $y_j(k)$ obey the Boltzmann probability distribution. $x_i(k)$ denotes the probability that the attacker selects the attack strategy AS_i in the k-th attack and defense confrontation at the same game stage. $y_j(k)$ denotes the probability that the defender selects the defense strategy DS_j in the k-th attack and defense confrontation at the same game stage. $\mathrm{Q}_{\mathrm{AS}_i}(k)$ denotes the expected revenue obtained by the attacker choosing the attack strategy AS_i in the k-th attack and defense confrontation at the same game stage. $\mathrm{Q}_{\mathrm{DS}_j}(k)$ denotes the expected revenue obtained by the defender choosing the defense strategy DS_j in the k-th attack and defense confrontation at the same stage. The *Q*-learning replicated dynamic equations formulas (15) and (16) are derived from the correlation formulas (2), (3), (15), and (14):

$$
x'(t) = \frac{\mathrm{d}x_i}{\mathrm{d}t} = \underbrace{x_i\left[\mathrm{Q}_{\mathrm{AS}_i} - \overline{\mathrm{Q}}_{\mathrm{AS}}\right]}_{\mathrm{RD}} + \underbrace{\frac{1}{\tau} x_i \sum_{k=1}^{n} x_k \ln(x_k/x_i)}_{\mathrm{ME}}. \tag{15}
$$

$$
y'(t) = \frac{\mathrm{d}y_j}{\mathrm{d}t} = \underbrace{y_j\left[\mathrm{Q}_{\mathrm{DS}_j} - \overline{\mathrm{Q}}_{\mathrm{DS}}\right]}_{\mathrm{RD}} + \underbrace{\frac{1}{\tau} y_j \sum_{l=1}^{m} y_l \ln\left(y_l/y_j\right)}_{\mathrm{ME}}. \tag{16}
$$

QRD consists of replication dynamic equation (RD) and mutation equation (ME). RD selects the most profitable strategy under current information. ME is to try different new strategies in unknown network attack and defense scenes, and constantly try and make error, and learn and adjust the strategies, which better reflects the diversity and uncertainty of network attack and defense.

From the definition of evolutionary equilibrium, when the strategies of players reach evolutionary equilibrium, there is

$$
x'(t) = 0 \text{ and } y'(t) = 0. \tag{17}
$$

Solution (x^*, y^*) of the above formula is an evolutionary stable equilibrium point. At this time, the τ value in the above equation needs to be large enough to make the selection probability of each strategy stable.

4.5. Reward Value Learning Mechanism. As shown in Algorithm 1, RLM calculates the incentive and punishment factor α according to reward variable RV and the proportion of the number *AN* of a certain type of attack strategy in the past SN stage. According to α and the defense result *R* of the last stage, RLM changes the reward value of the corresponding attack and defense strategy to change the attack and defense reward value of the next stage.

If the defense successfully resists an attack in the last stage $R = 1$, RLM increases the defense reward value of the specific strategy combination d_{ij}^{K}, and decreases the attack reward value of the specific strategy combination a_{ij}^{K}. If the defense failed in the last stage $R = 0$, RLM decreases the defense reward value d_{ij}^{K} of the specific strategy combination and increases the attack reward value a_{ij}^{K} of the specific strategy combination in the last stage.

5. Optimal Defense Strategy Selection Algorithm

In this paper, the Nash equilibrium solution of the multistage evolutionary game is regarded as a set of equilibrium solutions of a multistage evolutionary game. Each stage learns the known game information through the reward value learning mechanism to change the reward value of defense strategy in the current stage. According to the optimal defense strategy of each stage, the multistage optimal defense strategy set is constructed.

The multistage network attack and defense game tree based on our game model is shown in Figure 2. The black dots in Figure 2 represent the attacker at stage *K*, picked out with probability $P_A = (x_1, x_2, \ldots, x_n)$ and executed the attack strategy $\mathrm{AS} = (\mathrm{AS}_1, \mathrm{AS}_2, \ldots, \mathrm{AS}_n)$. The blue dots denote the defender at stage *K*, picked out with probability $P_D = (y_1, y_2, \ldots, y_m)$ and executed the defensive strategy $\mathrm{DS} = (\mathrm{DS}_1, \mathrm{DS}_2, \ldots, \mathrm{DS}_m)$.

As shown in Figure 2, the process of multistage network attack and defense games is as follows:

(1) In the initial stage of a network attack and defense game, the corresponding model parameters and

```
Input: revenue of strategy combination in the last stage $(\mathrm{AM}^{K01}, \mathrm{DM}^{K01})$. Current stage $K$ attack and defense strategy in the last stage $(\mathrm{AS}_i^{K01}, \mathrm{DS}_j^{K01})$. Defense results of the last stage $R$.
Output: revenue value of each strategy combination in the current stage $(\mathrm{AM}^K, \mathrm{DM}^K)$.
Initialize SN, AN, RV
if $K = 1$ then
    Calculate $\alpha$ from equation (4)
else if $K > 1$ then
    Calculate $\alpha$ from equation (5)
end if
if $R = 0$ //The result of the last stage of defense was failure then
    $a_{\mathrm{ij}}^K = a_{\mathrm{ij}}^{K-1} + \alpha$ and $d_{\mathrm{ij}}^K = d_{\mathrm{ij}}^{K-1} - \alpha$
else
    $a_{\mathrm{ij}}^K = a_{\mathrm{ij}}^{K-1} - \alpha$ and $d_{\mathrm{ij}}^K = d_{\mathrm{ij}}^{K-1} + \alpha$
end if
return $(\mathrm{AM}^K, \mathrm{DM}^K)$
```

ALGORITHM 1: Reward value learning mechanism.

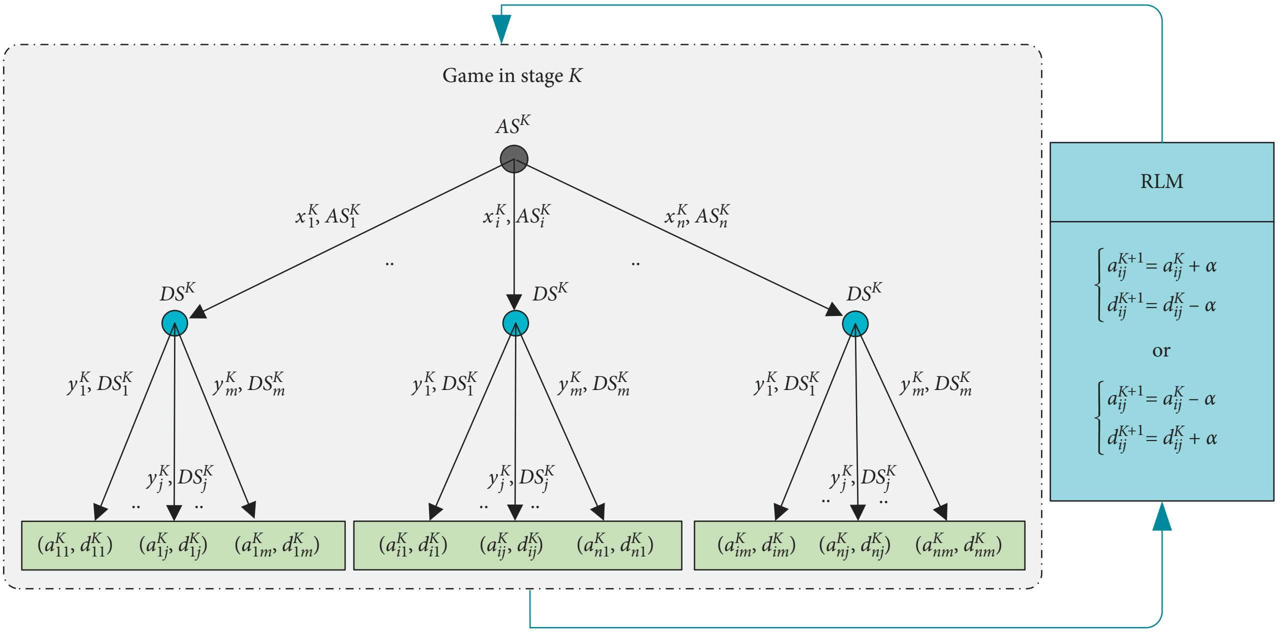


FIGURE 2: Multistage network attack and defense game tree.

reward values are calculated based on the current information of players, and the respective optimal decision in the current stage is reached by solving the Nash equilibrium and conserving information such as game results R of the current stage, the strategy combination $(\mathrm{AS}_i^K, \mathrm{DS}_j^K)$, and the attack and defense revenue $(\mathrm{AM}^K, \mathrm{DM}^K)$ for use in the calculation of the attack and defense revenue at the next stage.

(2) When the network attack and defense games enter stage $K\,(K > 1)$, as K grows larger and the attacker and defender gain more information gradually, the players of the game tend to be completely rational, and the revenue of the various combinations of strategies is highly likely to change. At this time, according to Algorithm 1, the incentive and punishment factor α of this stage is calculated, the attack and defense revenue $(\mathrm{AM}^K, \mathrm{DM}^K)$ of this stage is obtained from α and the information saved in the previous stage. The optimal defense strategy of the current stage is solved through the Nash equilibrium solution, and the defense result R, attack and defense strategy $(\mathrm{AS}_i^K, \mathrm{DS}_j^K)$, and attack and defense revenue $(\mathrm{AM}^K, \mathrm{DM}^K)$ of this stage are kept.

Next, we propose an optimal defense strategy selection algorithm based on the evolutionary game model based on RLM-QRD.

As shown in Algorithm 2, if K is the number of stages of the game, n and m represent the number of strategies of the attacker and the defender, respectively; generally, there is $K > 0$, $n > 0$, $m > 0$. The time cost of Algorithm 2 in each

```
     Input: evolutionary game model based on RLM-QRD.
     Output: probability set of optimal defense strategy in $K$-th stage $P_D^K$.
 (1) Initialize RG = $(N, K, S, \theta, Q, \tau, \alpha)$
 (2) for $i \longleftarrow 1$ to $n$ do
 (3)     for $j \longleftarrow 1$ to $m$ do
             Calculate $\mathrm{AM}^K, \mathrm{DM}^K$ from equations (8) and (10)
 (4)     end for
 (6) for$K \longleftarrow 1$ to $T$ do
 (7)     for$j \longleftarrow 1$ to $m$ do
 (8)         Construct $y'(t)$ from equation (16)
 (9) end for
(10) for $i \longleftarrow 1$ to $n$ do
(11)     Construct $x'(t)$ from equation (15)
(12)     end for
(13)     Calculate $\tau$ and $P_D^K$ from equation (17)
(14)     Calculate $(\mathrm{AM}^{K+1}, \mathrm{DM}^{K+1})$ from Algorithm 1
(15)     Output $P_D^K = (y_1^K, y_2^K, \ldots, y_M^K)$
(16) end for
```

Algorithm 2: The optimal defense strategy selection algorithm based on RLM-QRD evolutionary game.

stage is mainly concentrated on Step 8 and Step 11. In Step 8, the QRD of each defense strategy is constructed in turn, and the computational complexity of solving the equation in Step 8 is $\mathrm{O}(m^2 + \mathrm{nm})$. In Step 11, the QRD of each attack strategy is constructed in turn, and the computational complexity of solving the equation in Step 11 is $\mathrm{O}(n^2 + \mathrm{nm})$. The total time complexity of the algorithm is $\mathrm{O}(K(m+n)^2)$. The storage cost of the algorithm mainly focuses on the storage of the payoff matrix. The storage of the payoff matrix has high complexity, which contains the total number of nm memory cells. Therefore, the spatial complexity is $\mathrm{O}(\mathrm{nm})$.

Table 1 shows the comprehensive comparison between our model and other models in the literature. The following are some discussions:

(1) Payoff quantification: references [31, 36] have no extra parameters for the payoff quantification. References [22, 23] and [40] introduce specific parameters into the game model and set those parameters to quantify the attack and defense payoff in their respective application scenarios. We also introduce some parameters and calculate the parameters by the past game information, to better quantify the attack and defense payoff under the scenario of defense strategy failure.

(2) Equilibrium solution: the equilibrium solution represents the method used to solve the Nash equilibrium in the game model. References [23, 36] use RD, and reference [22] uses fermi function. All of these methods have had some success in their application. References [31, 40] improve on RD by proposing QRD and LQRD, respectively, and have been successful in optimal defense strategy selection. For the scenario of defense strategy failure, we propose RLM-QRD, which aims to realize automatic calculation of revenue, maximize defense revenue, and select the optimal defense strategy.

(3) Game type and Algorithm complexity: reference [31] is a static game, which has the advantage of low algorithm complexity. Although the complexity of the dynamic game algorithm is high, the dynamic game is more suitable for network attack and defense.

Based on the above discussion, our model is more suitable for failure scenarios of defense strategies in dynamic network attack and defense.

6. Experiment and Analysis

In this section, we verify the effectiveness of our model in the scenario of policy failure. Firstly, we give the attack strategy set and defense strategy set. We also introduce the strategy failure scenes. Secondly, we calculate the exploration factor τ. Thirdly, we analyze the defense strategy selection probability of our model under the single strategy failure scene and multistrategies failure scene. Fourthly, we compare the revenue of our model with those of the model without RLM. Finally, we compare the defense success rate of our model with those of the model without RLM.

6.1. Experimental Setup. In our experiment, the attack and defense behavior database of MIT [44] and China National Vulnerability Database of Information Security (CNNVD) [45] are used to analyze the attack and defense atomic strategy, as shown in Tables 2 and 3.

The attacker makes use of the vulnerability in the network information system to choose some atomic attack strategies. The defender selects several atomic defense strategies to defend against network attacks [46]. The attack and defense strategies in this experiment are composed of several atomic attack and defense strategies. For both sides of network attack and defense, set attack strategies $\mathrm{AS}_1 = \{a_1, a_2, a_5\}$ and $\mathrm{AS}_2 = \{a_3, a_4\}$ and defense strategies

Table 1: Comprehensive comparison among our model and other models.

Ref.	Payoff quantification	Game type	Equilibrium solution	Algorithm complexity	Application
[31]	No extra parameter	Static	QRD	$O((n+m)^2)$	Defense strategy selection
[36]	No extra parameter	Dynamic	RD	$O(K(m+n)^2)$	Virus protection
[23]	Discount factor δ	Dynamic	RD	$O(K(m+n)^2)$	Defense strategy selection
[22]	Dependent parameter μ	Dynamic	Fermi function	$O(K(m+n)^2)$	Privacy protection
[40]	Rational parameter λ	Dynamic	LQRD	$O(K(m+n)^2)$	Defense strategy selection
Ours	RLM	Dynamic	RLM-QRD	$O(K(m+n)^2)$	Defense strategy selection under strategy failure

Table 2: Attack strategy and sequence atom attack strategy.

Number	Name	Attack strategy	
		AS_1	AS_2
a_1	Install Web	√	
a_2	Remote attack	√	
a_3	Obtain user privileges		√
a_4	Buffer overflow attack		√
a_5	Homepage attack	√	

Table 3: Defense strategy and sequence atom defense strategy.

Number	Name	Defense strategy	
		DS_1	DS_2
d_1	Limit packers from ports	√	
d_2	Modify account password		√
d_3	Restart database server	√	
d_4	Limit packets from ports		√
d_5	Reinstall listener program	√	√
d_6	Correct homepage	√	

$DS_1 = \{d_1, d_3, d_5, d_6\}$ and $DS_2 = \{d_2, d_4, d_5\}$. By referencing [31, 34] and the definitions of attack and defense reward and cost in Section 4.3, the attack and defense revenue matrix of the first stage is given, as shown in Tables 4 and 5. The larger the number in the table, the greater the revenue from attack or defense.

In summary, we set the probability that attack strategy $AS = (AS_1, AS_2)$ and defense strategy $DS = (DS_1, DS_2)$ are chosen as $P_A = (x, 1-x)$ and $P_D = (y, 1-y)$, respectively, and set the reward variable RV = 10. Since there is a logarithm in the Q-learning replication dynamic equation, we assume that the value of the probability of strategy selection ranges $[0.01, 0.99]$. To better show the results of the experiment, the maximum number of learning stages SN = 500 and the maximum number of stages $T = 1000$ of RLM are set in this experiment.

This experiment verifies the validity of the evolutionary game model based on RLM-QRD in the scene of specific defense strategy failure from the perspective of single defense strategy failure and multiple defense strategies failure.

This experiment assumes that the attacker has acquired a specific defense strategy method at a certain stage and proposes and implements a new attack strategy and its selection probability, which results in the invalidation of the defense strategy. As shown in Table 6, status I and status II are single defense strategy failure scenes, and status III is multiple defense strategies failure scene.

6.2. The Calculation of Exploration Factor. It can be seen from equations (5) and (6) that when the exploration factor τ is small, both sides of the game do not fully grasp each other's relevant information under the condition of one stage. The ME in the Q-learning replication dynamic equation has a better impact, and the probability of attack strategies and defense strategies (AS_1, AS_2, DS_1, DS_2) selection is unstable.

Figure 3 shows the influence of exploration factor τ on the attack and defense strategy evolution. As shown in Figure 3, with the acquisition and analysis of each other's information in a single stage, the exploration factor τ gradually increases. The effect of the mutation equation decreases, the replication dynamic equation gradually begins to play a bigger role. The probability of attack and defense strategy selection gradually tends to be stable. QRD gradually degenerates into the replication dynamic equation. The simulation results show that the selection probability of attack and defense strategy remains constant when $\tau \geq 3$.

To sum up, the larger the exploration factor τ is, the more information the offensive and defensive sides can finally obtain in this game stage. Usually, both sides of the

Table 4: Attack revenue matrix of the first stage.

Attack strategy	Defense strategy	
	DS_1	DS_2
AS_1	100	70
AS_2	60	80

Table 5: Defense revenue matrix of the first stage.

Attack strategy	Defense strategy	
	DS_1	DS_2
AS_1	70	10
AS_2	50	60

Table 6: Strategy invalidation scene.

Status	Description
I	The attacker selects attack strategy AS_1 to make defense strategy DS_1 invalid
II	The attacker randomly selects attack strategy AS to make defense strategy DS_1 invalid
III	The attacker randomly selects attack strategy AS to make defense strategy DS invalid randomly

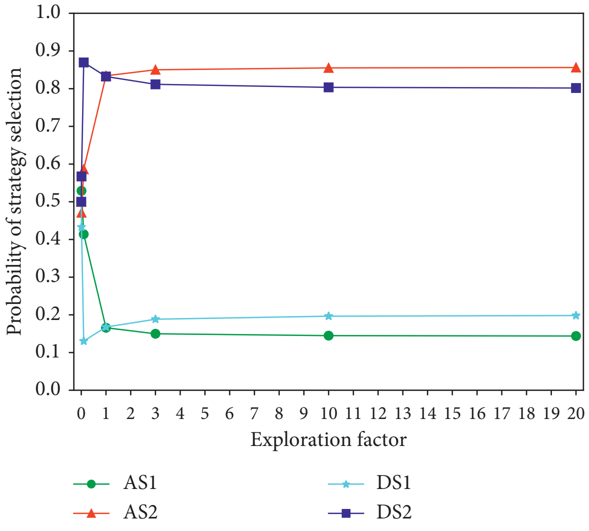


Figure 3: The influence of exploration factor τ on attack and defense strategy evolution.

game will change their strategies only when they know enough game information, and the game will enter a new stage. Therefore, this experiment sets the exploration factor $\tau = 100$.

6.3. *Probability of Defense Strategy Selection Experiment.*

To verify that the model can reduce the selection probability of failure strategy and increase the selection probability of effective strategy in single strategy failure and multistrategies failure scenes. This section studies the selection probability of optimal defense strategy when the status is I, II, and III, and the results are shown in Figures 4 and 5. Each point represents a stage. The red points and the orange points represent the failure of defense, which means the defender chooses the failure strategy. The green points and the blue points represent the success of the defense, which means the defender chooses the effective strategy. The x-axis and y-axis are the selection probabilities of game stage K and defense strategy DS_1, respectively.

6.4. *Single Strategy Failure Scene.*

Figure 4 shows the selection probability of defense strategy DS_1 in the number of stages $K = 30$ when the status is I and II. As shown in Figure 4, the selection probability of defense strategy DS_1 converges to around 0.01 after only about $K = 10$ stages in status I because defense strategy DS_1 has failed. The model implements defense strategy DS_2 with a high probability to resist the new attack strategy. The selection probability of defense strategy DS_1 converges after stage $K = 15$ in status II. Compared with the selection probability of defense strategy in status I, the selection probability of defense

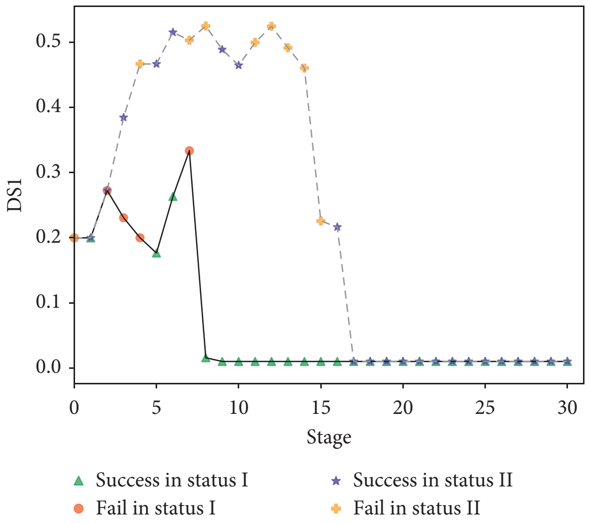


Figure 4: Selection probability of defense strategy under status numbers I and II.

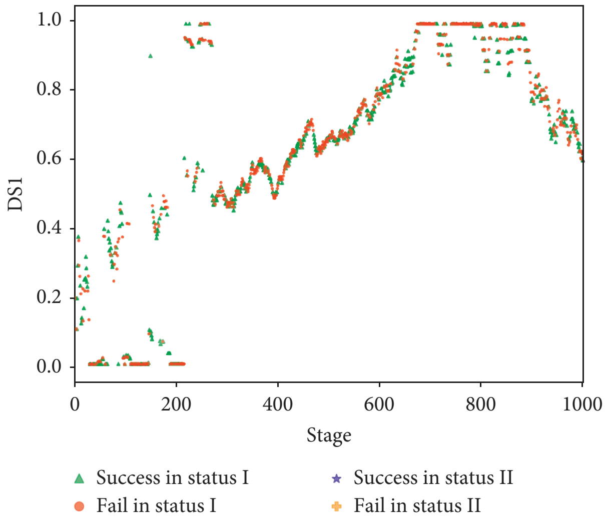


Figure 5: Selection probability of defense strategy under status number III.

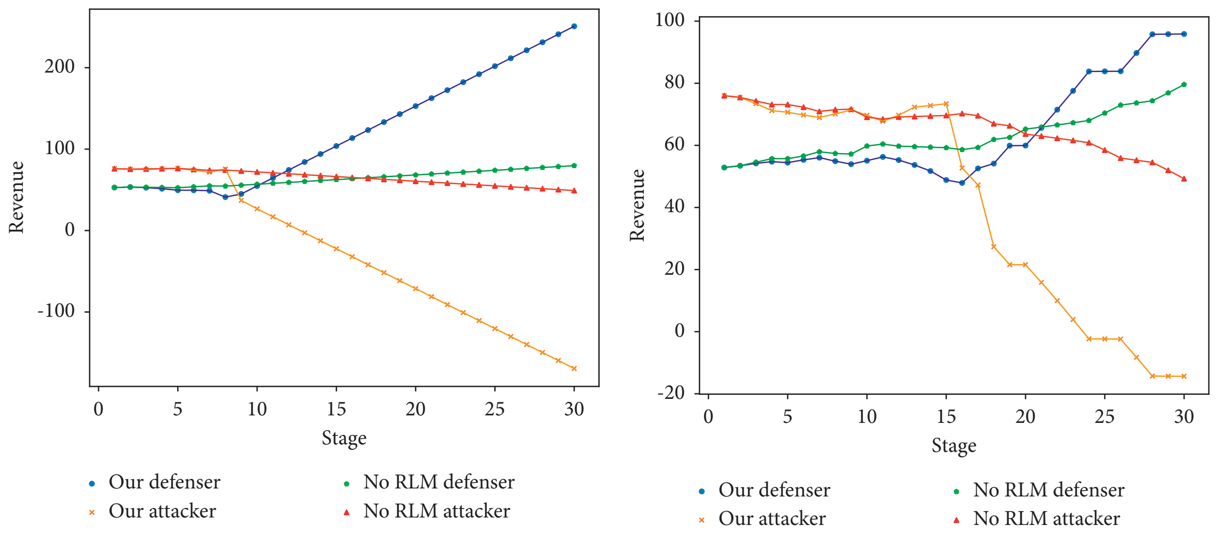
strategy in status II converges slowly because any strategy of the attacker can make defense strategy DS_1 invalid when the status is II.

As can be seen from Figure 4, the probability of failed defense strategy DS_1 shows a short upward trend before convergence and defense failure occur in the early stage in status I and status II because RLM is still learning attack strategies and their benefits at the beginning of the game stage. Due to the change of attack strategy, the strategy given by the defender is not optimal at this time. After several short stages of learning, RLM will converge the probability of failure strategy to 0.01. The defender has obtained the optimal defense strategy through several stages of learning. The game defense is successful.

6.5. Multistrategies Failure Scene. In status III, the attacker can choose any attack strategy to make any defense strategy invalid. In this scene, the defense strategy that failed may be DS_1 in stage K, and the defense strategy that failed may become DS_2 in stage $K+1$. So, the game becomes more complex.

Figure 5 shows the defense strategy selection probability in status III. As shown in Figure 5, the change of defense strategy selection probability is unstable, which is caused by the high complexity of the network state in this scene. In this scene, the defense side tends to give a more selection probability to the defense strategy DS_1.

In status III, the failed strategies of each stage are not necessarily the same, so we cannot judge the model from the



(a)

(b)

Figure 6: Comparison of attack and defense revenue under status I and II. (a) Status I. (b) Status II.

probability of defense strategy selection experiment. However, the model can be verified from attack and defense revenue experiment and defense success rate experiment in status III.

6.6. Attack and Defense Revenue Experiment. Figures 6(a), 6(b), and 7 show the comparative diagram of attack and defense revenue values based on our model and the evolutionary game model without RLM in statuses I, II, and III. The x-axis and y-axis are game stage and attack and defend revenue. The blue line and the yellow line represent the defense and the attack revenue of our model. The green line and the red line represent the defense and the attack revenue of the evolutionary game model without RLM.

6.7. Single Strategy Failure Scene. Figures 6(a) and 6(b) show the comparison of the attack and defense revenue values of the game model in the number of stages $K = 30$ when the status is I and II.

It can be seen from Figures 6(a) and 6(b) that the difference between the attack and defense revenues of the two models is not obvious at the beginning of the game stage, which indicates that RLM has not fully learned the attacker's information. At this time, the attack and defense revenues of the two models are unstable. Even because of the change of attack strategy, the revenues of the attacker may become more, and the revenues of the defender may become less.

After several stages, the defense revenues of the two models are positively correlated with the game stage, and the attack revenues are negatively correlated with the game stage. Compared with the game model without RLM, the model in Figure 6(a) has higher defense revenue and lower attack revenue after stage $K = 10$, and the model in Figure 6(b) has lower attack revenue after stage $K = 15$ and higher defense revenue after stage $K = 20$. Combined with the previous analysis, we can conclude that the revenue value in the game model with RLM changes drastically from the probability convergence of defense strategy. This situation also means that after the short-term learning stages of RLM, the defense revenue begins to rise, and the attack revenue begins to decline.

Compared with Figure 6(a), the change of attack and defense revenue in Figure 6(b) is relatively slow because any attacker's strategy may make defense strategy DS_1 invalid when the status is II. To sum up, this model can more effectively resist the attackers in the scenes of statuses I and II.

6.8. Multistrategies Failure Scene. Figure 7 shows the comparison of attack and defense revenue between our game model and the game model without RLM when the status is III. From the previous description, we can conclude that the scene is highly complex, and the attacker can adjust the strategy at each stage to invalidate the specific defense strategy. Therefore, with the continuous change of attack strategy, the attack and defense revenues fluctuate. Next, from the perspective of attack and defense revenues, we verify that our model has better defense ability than the model without RLM in status III.

As can be seen from Figure 7, the defense revenue of our model and the attack revenue of the game model without RLM rise with fluctuation, and the attack revenue of our model and the defense revenue of the game model without RLM decline with fluctuation. From stage $K = 400$, the defense revenue of our model is always higher than that of the game model without RLM. From stage $K = 300$, the attack revenue of this model is always lower than that of the game model without RLM.

To sum up, after learning multiple stages with RLM, the model can make the defense revenue greater and the attack revenue lower, to better resist network attacks.

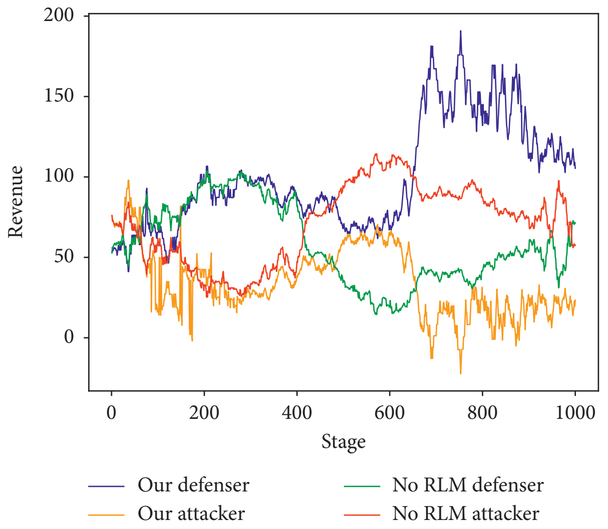


FIGURE 7: Comparison of attack and defense revenue under status III.

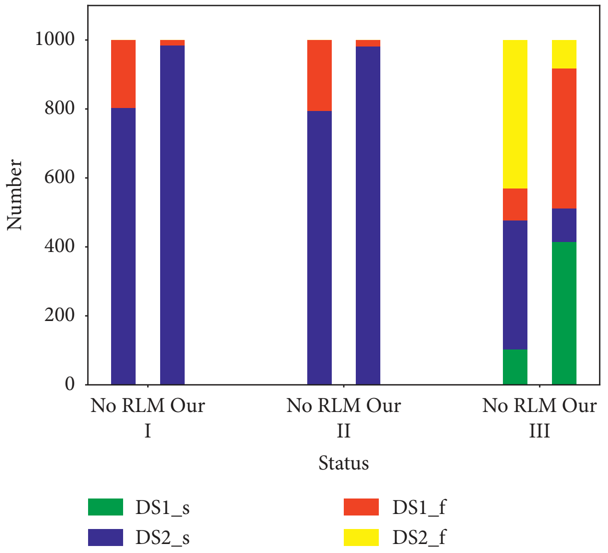


FIGURE 8: Comparison of defense success rate.

6.9. Defense Success Rate Experiment. Figure 8 shows the comparison of defense success rate between our model and the game model without RLM in stage $K = 1000$. The green column denotes the number of successful defense stages when the models choose defense strategy DS_1. The blue column denotes the number of successful defense stages when the models choose defense strategy DS_2. The red column denotes the number of failed defense stages when the models choose defense strategy DS_1. The yellow column denotes the number of failed defense stages when the models choose defense strategy DS_2.

It can be seen from Figure 8 that our model under statuses I, II, and III has a higher defense success rate than the game model without RLM. Compared with the game model without RLM, the defense success rate of this model is about 22.5% higher under state I. The defense success rate of our model increases by about 23.5% under status II. Our model improves by about 7.4% under status III. In conclusion, compared with the evolutionary game model without RLM, our model can better select the best defense strategy in the above three scenes.

7. Conclusion

Considering the existing problems in the application of the evolutionary game model in network attack and defense, this paper proposes a reward value learning mechanism. This mechanism overcomes the problem of quantifying incentives and punishments in the case of bounded rationality of attackers and defenders, which reduces manual involvement. An evolutionary game model with a multistage learning mechanism is constructed by combining the learning mechanism with a multistage game model. Furthermore, the optimal strategy selection algorithm of the game model is designed.

Our future work will study how to dynamically add new feasible defense strategies and reasonably expand the model when any defense strategy fails. In addition, we will also consider how to apply more intelligent methods, such as deep learning and machine learning, to the automatic calculation of reward and punishment factor α at every stage so that the model can better select the optimal defense strategy.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant nos. U1705262, 62072109, and U1804263) and the Natural Science Foundation of Fujian Province (Grant nos. 2021J01625, 2021J01616, and 2018J07005).

References

[1] M. A. Khan and K. Salah, “Iot security: review, blockchain solutions, and open challenges,” *Future Generation Computer Systems*, vol. 82, pp. 395–411, 2018.

[2] D. Kozhaya, J. Decouchant, V. Rahli, and P. Esteves-Verissimo, “Pistis: an event-triggered real-time byzantine-resilient protocol suite,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 9, pp. 2277–2290, 2021.

[3] W. Qian, H. Lai, Q. Zhu, and K.-C. Chang, “Overview of network security situation awareness based on big data,” in *Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications*, pp. 875–883, Springer, Cham, 05 March 2021.

[4] I. F. Kilincer, F. Ertam, and A. Sengur, “Machine learning methods for cyber security intrusion detection: datasets and comparative study,” *Computer Networks*, vol. 188, Article ID 107840, 2021.

[5] D. Woods, M. Abdallah, S. Bagchi, S. Sundaram, and T. Cason, "Network defense and behavioral biases: an experimental study," *Experimental Economics*, vol. 1–33, 2021.

[6] J. A. Paul and X. J. Wang, "Socially optimal it investment for cybersecurity," *Decision Support Systems*, vol. 122, Article ID 113069, 2019.

[7] V. Sivaraman, "A game-theoretic approach for enhancing data privacy in sdn-based smart grids," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10583–10595, 2021.

[8] K. Avrachenkov, L. Huang, R. Jason Marden, M. Coupechoux, and A. Giovanidis, "Game theory for networks," in *Proceedings of the 8th International EAI Conference, GameNets 2019*, vol. 277, Springer, Paris, France, 25 April 2019.

[9] J. Yingmo, K.-K. Raymond Choo, M. Li, L. Chen, and C. Guo, "Tradeoff gain and loss optimization against man-in-the-middle attacks based on game theoretic model," *Future Generation Computer Systems*, vol. 101, pp. 169–179, 2019.

[10] Z. Zhu, Y. Xu, and Z. Su, "A reputation-based cooperative content delivery with parking vehicles in vehicular ad-hoc networks," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1531–1547, 2021.

[11] X. Xu, G. Wang, J. Hu, and Y. Lu, "Study on stochastic differential game model in network attack and defense," *Security and Communication Networks*, vol. 2020, Article ID 3417039, 15 pages, 2020.

[12] W. Zhang and C. Peng, "Indefinite mean-field stochastic cooperative linear-quadratic dynamic difference game with its application to the network security model," *IEEE Transactions on Cybernetics*, pp. 1–14, 2021.

[13] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Computing Surveys*, vol. 45, no. 3, pp. 1–39, 2013.

[14] S. Du, X. Li, J. Du, and H. Zhu, "An attack-and-defence game for security assessment in vehicular ad hoc networks," *Peer-to-peer Networking and Applications*, vol. 7, no. 3, pp. 215–228, 2014.

[15] G. Cimini, C. Castellano, and A. Sánchez, "Dynamics to equilibrium in network games: individual behavior and global response," *PLoS one*, vol. 10, no. 3, Article ID e0120343, 2015.

[16] C. Liu, E. Zhu, Q. Zhang, and X. Wei, "Exploring the effects of computational costs in extensive games via modeling and simulation," *International Journal of Intelligent Systems*, vol. 36, no. 8, pp. 4065–4087, 2021.

[17] Su Yuan, Xi Zhang, L. Liu, S. Song, and B. Fang, "Understanding information interactions in diffusion: an evolutionary game-theoretic perspective," *Frontiers of Computer Science*, vol. 10, no. 3, pp. 518–531, 2016.

[18] C. Luo, C. Sun, and B. Liu, "Environment-based preference selection in spatial multigame with limited resource allocation and control," *Communications in Nonlinear Science and Numerical Simulation*, vol. 99, Article ID 105845, 2021.

[19] Z. Tian, X. Gao, S. Su, J. Qiu, X. Du, and M. Guizani, "Evaluating reputation management schemes of internet of vehicles based on evolutionary game theory," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 5971–5980, 2019.

[20] J. Moura and D. Hutchison, "Game theory for multi-access edge computing: survey, use cases, and future trends," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 260–288, 2018.

[21] H. Guo, X. Wang, H. Cheng, and M. Huang, "A routing defense mechanism using evolutionary game theory for delay tolerant networks," *Applied Soft Computing*, vol. 38, pp. 469–476, 2016.

[22] U. Mengibaev, X. Jia, and Y. Ma, "The impact of interactive dependence on privacy protection behavior based on evolutionary game," *Applied Mathematics and Computation*, vol. 379, Article ID 125231, 2020.

[23] Yu Yang, B. Che, Y. Zeng, Y. Cheng, and C. Li, "Maiad: a multistage asymmetric information attack and defense model based on evolutionary game theory," *Symmetry*, vol. 11, no. 2, p. 215, 2019.

[24] C. T. Do, N. H. Tran, C. Hong et al., "Game theory for cyber security and privacy," *ACM Computing Surveys*, vol. 50, no. 2, pp. 1–37, 2017.

[25] H. Zhang, L. V. Jiang, S. Huang, J. Wang, and Y. Zhang, "Attack-defense differential game model for network defense strategy selection," *IEEE Access*, vol. 7, pp. 50618–50629, 2018.

[26] M. Baykal-Guersoy, Z. Duan, H. Vincent Poor, and A. Garnaev, "Infrastructure security games," *European Journal of Operational Research*, vol. 239, no. 2, pp. 469–478, 2014.

[27] Na Ruan, L. Gao, H. Zhu, W. Jia, X. Li, and Qi Hu, "Toward optimal dos-resistant authentication in crowdsensing networks via evolutionary game," in *Proceedings of the 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*, pp. 364–373, IEEE, Nara, Japan, 27 June 2016.

[28] M. Abdalzaher, K. Seddik, M. Elsabrouty, O. Muta, H. Furukawa, and A. Abdel-Rahman, "Game theory meets wireless sensor networks security requirements and threats mitigation: a survey," *Sensors*, vol. 16, no. 7, p. 1003, 2016.

[29] M. Bouhaddi, M. S. Radjef, and K. Adi, "An efficient intrusion detection in resource-constrained mobile ad-hoc networks," *Computers & Security*, vol. 76, pp. 156–177, 2018.

[30] M. B. Shareh, H. Navidi, H. H. S. Javadi, and M. HosseinZadeh, "Preventing sybil attacks in p2p file sharing networks based on the evolutionary game model," *Information Sciences*, vol. 470, pp. 94–108, 2019.

[31] H. Jin, H. Zhang, C. Zhang, and H. Hu, "Research on active defense decision-making method based on qrd in complex network," *Netinfo Security*, vol. 20, no. 5, pp. 72–82, 2020, (in Chinese).

[32] X. Liu, H. Zhang, Y. Zhang, and L. Shao, "Optimal network defense strategy selection method based on evolutionary network game," *Security and Communication Networks*, vol. 2020, Article ID 5381495, 11 pages, 2020.

[33] L. Shi, X. Wang, and H. Hou, "Research on optimization of array honeypot defense strategies based on evolutionary game theory," *Mathematics*, vol. 9, no. 8, p. 805, 2021.

[34] J. Huang, H. Zhang, and J. Wang, "Markov evolutionary games for network defense strategy selection," *IEEE Access*, vol. 5, pp. 19505–19516, 2017.

[35] J.-ming Huang and J.-dong Wang, "Heng-wei Zhang and Na Wang. "Network defense strategy selection based on best-response dynamic evolutionary game model," in *Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 2611–2615, IEEE, Chongqing, China, 25-26 March 2017.

[36] Y. Hayel and Q. Zhu, "Epidemic protection over heterogeneous networks using evolutionary Poisson games," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1786–1800, 2017.

[37] F. He, Li Xu, J. Li, and K.-K. Raymond Choo, "An adaptive trust-stackelberg game model for security and energy

efficiency in dynamic cognitive radio networks," *Computer Communications*, vol. 105, pp. 124–132, 2017.

[38] Z. Wang, C. Li, X. Jin, H. Ding, G. Cui, and L. Yu, "Evolutionary dynamics of the interdependent security games on complex network," *Applied Mathematics and Computation*, vol. 399, Article ID 126051, 2021.

[39] H. Hu, Y. Liu, H. Zhang, and R. Pan, "Optimal network defense strategy selection based on incomplete information evolutionary game," *IEEE Access*, vol. 6, pp. 29806–29821, 2018.

[40] H. Hu, Y. Liu, C. Chen, H. Zhang, and Y. Liu, "Optimal decision making approach for cyber security defense using evolutionary game," *IEEE Transactions on Network and Service Management*, vol. 17, no. 3, pp. 1683–1700, 2020.

[41] Y. Guo, L. Wang, Z. Liu, and Y. Shen, "Reinforcement-learning-based dynamic defense strategy of multistage game against dynamic load altering attack," *International Journal of Electrical Power & Energy Systems*, vol. 131, Article ID 107113, 2021.

[42] Z. Yang, Y. Shi, and Y. Li, "Analysis of intellectual property cooperation behavior and its simulation under two types of scenarios using evolutionary game theory," *Computers & Industrial Engineering*, vol. 125, pp. 739–750, 2018.

[43] J. Li, G. Kendall, and R. John, "Computing nash equilibria and evolutionarily stable states of evolutionary games," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 3, pp. 460–469, 2015.

[44] L. A. Gordon, P. Martin Loeb, W. Lucyshyn, and R. Richardson, "2005 csi/fbi computer crime and security survey," *Computer Security Journal*, vol. 21, no. 3, p. 1, 2005.

[45] J. Peng, M. Guo, and J. Quan, "Software vulnerability and application security risk," *Information Resources Management Journal*, vol. 32, no. 1, pp. 48–57, 2019.

[46] Q. Tan, Y. Gao, J. Shi, X. Wang, B. fang, and Z. Tian, "Toward a comprehensive insight into the eclipse attacks of tor hidden services," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1584–1593, 2018.