An Anti-Jamming Hierarchical Optimization Approach in Relay Communication System via Stackelberg Game

Zhibin Feng 1, Guochun Ren 1,*, Jin Chen 1, Chaohui Chen 2, Xiaoqin Yang 1, Yijie Luo 1 and Kun Xu 3

1 College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China
2 Haige Communications Group Incorporated Company, Guangzhou 510000, China
3 College of Information and Communication, National University of Defense Technology, Wuhan 430000, China

* Correspondence: guochunren@yeah.net

Received: 15 July 2019; Accepted: 10 August 2019; Published: 14 August 2019

Abstract: In this paper, we study joint relay selection and the power control optimization problem in an anti-jamming relay communication system. Considering the hierarchical competitive relationship between a user and jammer, we formulate the anti-jamming problem as a Stackelberg game. From the perspective of game, the user selects relay and power strategy firstly which acts as the leader, while the jammer chooses power strategy then that acts as follower. Moreover, we prove the existence of Stackelberg equilibrium. Based on the Q-learning algorithm and multi-armed bandit method, a hierarchical joint optimization algorithm is proposed. Simulation results show the user’s strategy selection probability and the jammer’s regret. We compare the user’s and jammer’s utility under the proposed algorithm with a random selection algorithm to verify the algorithm’s superiority. Moreover, the influence of feedback error and eavesdropping error on utility is analyzed.

Keywords: anti-jamming; relay selection; power control; Stackelberg game; Q-learning; multi-armed bandit

1. Introduction

The anti-jamming problem has always been a hot topic in wireless communication domain. As a major threat to reliable communication, the jammer can launch noise-like signals to legitimate receiver to decrease communication quality. In order to cope with malicious jamming attacks, a lot of anti-jamming works have been studied that focused on power control [1–10], frequency selection [11–20], and spatial domain [21–25], and many effective anti-jamming technologies were proposed. In this paper, we investigate the anti-jamming hierarchical optimization problem in relay communication system.

The power adjustment is considered as an effective method to respond to the jammer’s attacks directly. In [1], the authors made receivers more robust with adaptive arrays utilizing power inversion algorithm in jamming environment. In [2], based on the competitive relationship between user and jammer, a Stackelberg game was proposed to model the anti-jamming communication hierarchical optimization problem in the presence of smart jammer. Considering it is hard for the user and jammer to obtain the accurate power and channel state information, the authors analyzed the anti-jamming performance with observation error in [3–5]. In [6], the authors studied the communication problem between in WBANs via Stackelberg frame facing jamming attacks. In [7], the authors investigated cross-layer anti-jamming joint optimization problem based on Q-learning algorithm. In [8], the authors investigated discrete power strategy optimization problem with imperfect information.
the power allocation problem was studied in cognitive radio communication networks within different game frames. Taking the perspective of Stackelberg game, the authors combed the anti-jamming technologies under different scenarios in [10], moreover, an anti-jamming decision-making framework was also proposed based on the adversarial characteristics between user and jammer, incomplete information and so on.

In the domain of relay communication [26–39], the anti-jamming problem has also been widely investigated. In [27], the authors studied communication frequency selection optimization problem in unpredictable jamming environment. In [28], the authors formulated the relay selection and discrete power optimization problem under power constraint as a potential game, in which there existed more than one pure strategy Nash equilibrium (NE) and the optimal strategy profile was obtained. In [29], the cooperative relaying technology was investigated in vehicular networks. In [30], the authors studied the ongoing power control optimization problem via Stackelberg game and obtained the closed-form solutions of optimal power strategy. In [31], the authors decreased algorithm complexity and obtained the optimal joint strategy in a distributed wireless communication networks. In [32], the authors focused on game-theoretical approaches in cognitive communication and networking systems, in which a stochastic anti-jamming game was proposed to design the optimal adaptive defense strategies against cognitive malicious attackers. In [33], the authors designed an opportunistic wireless communication setting and modelled the asset selling problem as a game theoretic variant in the completely observable and the partially observable cases respectively.

With the increasing complexity of communication scenarios and jamming modes, in [34,35], the authors studied the optimal power allocation problem in friendly jamming and Gaussian noise jamming environments to increase secrecy sum rate in untrusted relay communication networks. In [36], the authors studied anti-jamming power optimization problem in multi-user relay communication system, a three-layer Stackelberg game was proposed and Stackelberg equilibrium (SE) was also obtained based on duality optimization theory. In [37], the authors studied the problem of resource assigning in a single carrier communication system in the presence of a jammer, a Bayesian jamming game framework was proposed and the Nash strategy was also compared to the Stackelberg strategy to verify its sensibility. In [38], the authors studied a new anti-jamming problem of unknown nodes in a peer-to-peer network attacked by a random jammer or an intelligent one. In [39], considering the idealized case and potential energy constraint, the authors investigated the jamming attacks optimization problem that the jammer could control the probability of jamming and transmission range to cause maximal damage to wireless network. The existing works [34–39] investigated the anti-jamming problems from the perspectives of adaptive defense strategies, potential energy constraint, jamming modes, ideal and non-ideal cases in different communication scenarios to perfect the study. Game theoretical approaches, such as stochastic game and Stackelberg game, are just methods which can help us analyze the anti-jamming problems effectively.

The main difference between this paper and the above anti-jamming works [29–39] is that, in the presence of an intelligent jammer, we investigate the relay selection problem under a power–energy constraint with imperfect information. Moreover, considering the existing power optimization methods in anti-jamming domain, there are still some problems needing to be solved. Firstly, it is hard for the user and jammer to get the math utilities and channel state information. Secondly, there exists error when a user measures the communication quality or a jammer measures the jamming effect. Thirdly, the optimizations of power strategy and relay selection need to be considered simultaneously.

In this paper, we investigate the joint relay selection and power control optimization problem in an anti-jamming relay communication scenario. Considering the competitive relationship between the user and jammer, we formulate the anti-jamming problem as a Stackelberg game, which is an effective tool to model the hierarchical optimization problem. In the game frame, the user acts as leader to select relay and power strategy firstly, and the jammer acts as follower then to choose power strategy. Moreover, the proof of existence of SE is given. Based on Q-learning algorithm and multi-armed bandit (MAB) method, a hierarchical joint optimization algorithm is proposed. Simulation results show the
user’s strategy selection probability and the jammer’s regret. We give utility comparison under the proposed algorithm and random selection algorithm with feedback error and eavesdropping error. In the last, we summarize the main contributions of this paper in the following:

- In relay communication scenario, considering the competitive relationship between the user and jammer, we formulate the anti-jamming joint optimization problem as a Stackelberg game, in which the user acts as leader and jammer acts as follower.
- We prove the existence of SE and propose a hierarchical joint optimization algorithm based on Q-learning and MAB method via Stackelberg game frame.
- Simulation results show the user’s strategy selection probability and jammer’s regret. Utility under the proposed algorithm is compared with random selection algorithm to verify the algorithm’s superiority. Moreover, the influence of feedback error and eavesdropping error on utility is analyzed.

The main differences from our previous work [36] are summarized as: (i) There exist multiple relays in the anti-jamming system which needs to realize the optimization of relay selection and power control simultaneously. (ii) The feedback error and eavesdropping error are introduced because it is hard for the the user and jammer to get accurate feedback information. (iii) The user and jammer do not need to know the power and channel fading information which is ignored in [36].

In the rest of this paper, we establish the anti-jamming model in the proposed relay communication scenario, and give the relative problem formulations in Section 2. In Section 3, we formulate the anti-jamming joint optimization problem as a Stackelberg game and prove the existence of SE. Moreover, a hierarchical joint optimization algorithm is proposed. Simulation results show the user’s strategy selection probability and the jammer’s regret in Section 4, besides, utility comparison under different algorithms with feedback error and eavesdropping error is also given. Finally, we draw a conclusion in Section 5.

2. System Model and Problem Formulation

2.1. System Model

We assume an anti-jamming relay communication scheme consisted of a base station (BS), a user, a jammer and a relay group, as shown in Figure 1. The user collects nearby information and sends it to the BS. Considering the serious channel fading of wireless communication, the user cannot transmit messages to BS directly, which communicates with BS through a relay group. In phase 1, the user transmits messages to the relay group that contains its collected information and instruction which relay is acquired to assist the user to forward messages to BS. In phase 2, the relay adopts apply and forward (AF) mode, and retransmits the messages received from the user to BS. The user and relay use the same channel. The jammer sends malicious noise-like signals to decrease communication quality. Moreover, both the user and jammer have ability to change power to achieve a better communication or jamming effect. Note that for simplicity, we consider relay’s power remains unchanged, because relay as the fixed station on the ground, the power constraint is not obvious compared to the user.

![Figure 1. Anti-jamming relay communication system.](image-url)
We consider the user and jammer update their respective strategies with different time scales. On the one hand, jammer updates its power strategy every epoch. On the other hand, in order to improve anti-jamming ability, the user updates its joint relay and power strategy quicker than jammer to adjust timely and obtain the optimal strategy. As shown in Figure 2, each epoch is divided as T time slots, and at the end of each time slot, BS feeds the signal-on-interference-plus-noise ratio (SINR) information of current time slot back to the user as the communication reward, which is also eavesdropped by jammer [40–42] to measure the jamming effect at the current epoch.

In the formulated model, we assume that the user has $M$ power levels and its power set is $\mathcal{P} = \{P_1, P_2, ..., P_M\}$. Jammer has $L$ power levels and its power set is $\mathcal{J} = \{J_1, J_2, ..., J_L\}$. The relay group contains $N$ relays and the relay set is $\mathcal{R} = \{R_1, R_2, ..., R_N\}$. The $n$th relay’s power is a fixed value and defined as $Q_n$, and all relays’ power set is $\mathcal{Q} = \{Q_1, Q_2, ..., Q_N\}$. The distance between the user, jammer, BS and the $n$th relay are $d_{u,r_n}, d_{j,r_n}, d_{B,r_n}$, respectively. The distance between BS and jammer is $d_{B,j}$. For convenience, some used notations are listed in Table 1.

### Table 1. Summation of used notation.

| Notation | Description |
|----------|-------------|
| $\mathcal{P}, \mathcal{J}$ | Power set of the user and jammer |
| $\mathcal{R}, \mathcal{Q}$ | Relay set and all relays’ power set |
| $P, J$ | Transmitting power of the user and jammer, i.e., $P \in \mathcal{P}, J \in \mathcal{J}$ |
| $a_{u,r_n}, \beta_{j,r_n}$ | The channel gain between the user, jammer and the $n$th relay |
| $a_{B,r_n}, \beta_{B,j}$ | The channel gain between the $n$th relay, jammer and BS |
| $N_{r_n}, N_B$ | The back noise power at the $n$th relay and BS |
| $G_n$ | The $n$th relay’s amplification factor |
| $X_{u}, X_{j}, X_{r_n}$ | The user’s, jammer’s and $n$th relay’s transmitted signal |
| $Y_{r_n}, Y_B$ | The $n$th relay’s and BS’s received signal |
| $C_u, C_j$ | The transmitting cost of the user and jammer |
| $\epsilon_1, \epsilon_2$ | Feedback error and eavesdropping error |
| $\gamma$ | The SINR received at BS |
| $\gamma', \gamma''$ | The feedback SINR received by the user and eavesdropped by jammer |

### 2.2. Problem Formulation

In the anti-jamming system, the communication process is divided into two phases. In phase 1, the user transmits messages and selects one relay to help forward messages to BS. When the jammer senses communication signals between the user and relay, it releases jamming signals immediately to destroy normal communication. Let $P$ and $J$ denote the user’s and jammer’s power, where $P \in \mathcal{P}$ and $J \in \mathcal{J}$. Inspired by [43], we define $a_{u,r_n} = d_{u,r_n}^{-\delta}$ and $\beta_{j,r_n} = d_{j,r_n}^{-\delta}$ as the channel gain between the
user, jammer and the nth relay, respectively, among which δ is the path-loss factor. The backward noise signal at the nth relay is \( m_n \). \( X_u \) and \( X_j \) denote the user’s and jammer’s transmitted signal, respectively. \( Y_{fr} \) denotes the nth relay’s received signal from the user and jammer, which is defined as follows:

\[
Y_{fr} = \sqrt{P_{ufr}} X_u + \sqrt{P_{jfr}} X_j + m_{fr}.
\]  

In phase 2, the relay adopts amplifies and forwards (AF) mode. Considering the nth relay is selected to help the user forward messages to BS, its amplification factor is expressed as follows:

\[
G_n = \frac{Q_n}{\sqrt{P_{ufr} + J_{jfr}^2 + N_n}}.
\]  

Let \( X_{fr} \) denote the nth relay’s retransmitted signal. We define the received signal at BS as \( Y_B \), and it is expressed as:

\[
Y_B = \sqrt{Q_n} a_{Bfr} X_{fr} + \sqrt{Q_n} \beta_{jfr} X_j + m_B
\]

\[
= G_n \sqrt{P_{ufr} a_{ufr} X_u} + G_n \sqrt{P_{jfr} \beta_{jfr} X_j} + G_n a_{Bfr} m_{fr} + \sqrt{P_{jfr} \beta_{jfr} X_j} + m_B,
\]  

where \( a_{Bfr} = d_{Bfr}^{-\delta} \) and \( \beta_{jfr} = d_{jfr}^{-\delta} \) denote the channel gain between the nth relay, jammer and BS, respectively. \( m_B \) is the backward noise signal at BS.

Thus the SINR received at BS can be defined as follows:

\[
\gamma = \frac{G_n^2 P_{ufr} a_{ufr}^2}{G_n^2 P_{fr} a_{fr}^2 + G_n^2 \beta_{jfr}^2 N_n + J_{jfr}^2 + N_B}.
\]  

For the user, it obtains feedback SINR \( \gamma' \) at the end of each time slot. Considering there may exists feedback information error, we introduce feedback error \( \varepsilon_1 = |\gamma - \gamma'|/\gamma \), which means the deviation degree of feedback SINR \( \gamma' \) with the actual SINR \( \gamma \). Inspired by [2–4], considering the transmitting cost, we give the user’s utility function based on the received SINR \( \gamma' \) in the following:

\[
U = \gamma' - C_u P.
\]  

For jammer, it eavesdrops the feedback SINR \( \gamma' \) from BS to the user at the end of each time slot and obtains eavesdropping result \( \gamma'' \). Similarly, considering that it is hard for a jammer to eavesdrop \( \gamma' \) accurately, we introduce eavesdropping error \( \varepsilon_2 = |\gamma - \gamma''|/\gamma \), which means the deviation degree of eavesdropped SINR \( \gamma'' \) with the actual SINR \( \gamma \). Similarly, considering the jamming cost, we give the jammer’s utility function based on the eavesdropped result \( \gamma'' \) in the following:

\[
V = -\gamma'' - C_j J.
\]  

3. The Joint Relay Selection and Power Control Optimization Method via Stackelberg Game

In this section, we formulate the anti-jamming joint optimization problem as a Stackelberg game firstly, which is an effective theoretical method to deal with the hierarchical confrontation relationship between the user and jammer. Then we prove the existence of Stackelberg equilibrium (SE). We propose a hierarchical joint optimization algorithm under the game frame in the last.

3.1. Stackelberg Game Model

In the anti-jamming relay communication, the user transmits messages firstly and selects a relay to help broadcast messages to base station, after having sensed the user’s signals, the jammer releases jamming signals immediately. Take a perspective of Stackelberg game, the user acts as
leader and jammer acts as follower in the proposed scenario. The proposed game can be denoted as $\mathcal{G} = \{\mathcal{P}, \mathcal{Q}, \mathcal{J}, \mathcal{U}, \mathcal{V}\}$, where $\mathcal{P}$ and $\mathcal{Q}$ denote the user’s power strategy space and relay selection strategy space respectively, $\mathcal{J}$ denotes jammer’s power strategy space, $\mathcal{U}$ and $\mathcal{V}$ denote the user’s and jammer’s utility, respectively.

Based on the utilities given in Section 2.2, both the user and jammer aim to maximize their utilities to get a better communication effect or jamming effect. For the jammer, given a user’s joint strategy $P \in \mathcal{P}, Q_n \in \mathcal{Q}$, it makes the optimal power strategy to destroy normal communication, and the optimization problem is expressed as:

$$\max_J V(P, Q_n, J).$$ (7)

For the user, it makes a joint relay selection and power control strategy to guarantee anti-jamming communication quality, and the user’s optimization problem can be expressed as follows:

$$\max_{P,Q_n} U(P, Q_n, J).$$ (8)

Based on the analysis above, the joint optimization problem can be solved by hierarchical decision-making method via Stackelberg game frame, which is shown as follows:

$$\begin{align*}
\max_{P,Q_n} U(P, Q_n, J) \\
\text{subject to: } P \in \mathcal{P}, Q_n \in \mathcal{Q} \\
The optimal solution: (J^*) \\
\max_J V(P, Q_n, J) \\
\text{subject to: } J \in \mathcal{J}
\end{align*}$$ (9)

3.2. Existence of Stackelberg Equilibrium

In the proposed scenario, we consider that the user adopts mixed strategy to fool the jammer due to the randomness of strategy, which can increase anti-jamming performance effectively. Let $q$ denotes the user’s mixed strategy, i.e., the probability distribution of the user’s optional relay selections and power strategies. Motivated by [4,8], we define the Stackelberg equilibrium (SE) and give the proof of existence of SE in the following.

**Definition 1.** If no player can improve the utility by deviating its optimal strategy unilaterally, the policy profile $(q^*, J^*)$ constitutes the SE, which satisfies the following conditions:

$$U(q^*, J^*) \geq U(q, J^*),$$ (10)

$$V(q^*, J^*) \geq V(q^*, J).$$ (11)

**Lemma 1.** There exists a user’s stationary strategy and a smart jammer’s stationary strategy, which constitute a SE [8].

**Proof.** Inspired by [44,45], every finite strategy game has a mixed strategy equilibrium [46], which means there exists a SE in the formulated game.

For the jammer, it aims to maximize the utility and makes strategy based on best-response:

$$J^* = \arg\max_J V(q, J).$$ (12)
Having known the jammer’s optimal strategy, the user’s optimal strategy can be obtained as follows:

$$q^* = \arg \max_q U(q, J^*(q)).$$ (13)

Based on the analysis above, the policy profile \((q^*, J^*)\) constitutes the SE.

### 3.3. Hierarchical Joint Optimization Algorithm

In this section, we propose a hierarchical joint optimization algorithm, which realizes the user’s strategy optimization based on Q-learning algorithm [47] and jammer’s strategy optimization based on multi armed bandit (MAB) method [48–50], respectively.

For the user, its mixed power and relay selection strategy at the \(t\)th time slot is expressed as

$$q(t) = \{q_{1,1}(t), q_{1,2}(t), ..., q_{1,N}(t), ..., q_{M,1}(t), q_{M,2}(t), ..., q_{M,N}(t)\},$$

where \(q_{m,n}(t)\) denotes the probability to select the \(n\)th relay and power \(P_m\) at the \(t\)th time slot. Then we update the user’s Q value as follows:

$$Q_{m,n}(t+1) = (1 - \kappa^t)Q_{m,n}(t) + \kappa^t r(t),$$ (14)

where \(\kappa^t\) denotes the user’s learning rate at the \(t\)th time slot, which satisfies \(\sum_{t=0}^{\infty} (\kappa^t)^2 < \infty\). \(r(t)\) denotes the user’s reward, i.e., its utility \(U\) at the \(t\)th time slot. Based on the analysis above, the user’s mixed strategy is updated as follows:

$$q_{m,n}(t+1) = \frac{\exp \left[ q_{m,n}(t)/\tau_0 \right]}{\sum_{1 \leq m \leq M, 1 \leq n \leq N} \exp \left[ q_{m,n}(t)/\tau_0 \right]} \text{ (15)}$$

where \(\tau_0\) controls the tradeoff of exploration-exploitation.

For the jammer, we formulate the finite strategy optimization as a MAB problem, and consider each power strategy as an arm to select, i.e., the jamming power strategy \(J_l\) is considered as the \(l\)th arm. \(J(k)\) denotes the jammer’s power strategy at the \(k\)th epoch. Thus the times of power \(J_l\) has been selected in the past \(K\) epochs is defined as:

$$A_l(K) = \sum_{k=1}^{K} \delta(J_l, J(k)), \text{ (16)}$$

where \(\delta(x, y)\) is the Kronecker delta function and it can be expressed as:

$$\delta(x, y) = \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases} \text{ (17)}$$

Thus, the jammer’s \(l\)th arm’s statistical average reward \(\mu_l\) in the past \(K\) epochs is defined as:

$$\mu_l = \frac{\sum_{k=1}^{K} V(k) \delta(J_l, J(k))}{A_l(K)} \text{ (18)}$$

\(B(K)\) denotes the jammer’s total utility in the past \(K\) epochs and is expressed as:

$$B(K) = \sum_{k=1}^{K} V(k). \text{ (19)}$$
Inspired by [48–50], we define jammer’s regret in the past $K$ epochs, which is an important index to represent the loss of utility so far because of the failure to select the optimal jamming power strategy, as shown in the following:

$$E[R(K)] = E[B^*(K)] - E[B(K)] = \sum_{l \neq l^*} (V_l - V_{l^*}) E[A_l(K)],$$

(20)

where $l^*$ denotes the estimated optimal arm of jammer based on the historical statistical average reward. $B^*(K)$ denotes jammer’s accumulative utility it could get if jammer had always chosen $l^*$ in the past $K$ epochs.

In the MAB problem, UCB1 [48–50] is an effective policy to realize the optimization of power strategy. Adopting the UCB1 method, jammer updates its strategy in the $(K + 1)$th epoch based on the following condition:

$$J(K + 1) = \arg \max_{l \in [1, 2, ..., L]} \left( \mu_l + \sqrt{\frac{2 \ln K}{A_l(K)}} \right).$$

(21)

According to the analysis above, based on the Q-learning algorithm and UCB1 method, both the user’s and jammer’s optimal strategy can be obtained through hierarchical joint optimization algorithm, which is shown in Algorithm 1.

**Algorithm 1: Hierarchical joint optimization algorithm (HJOA).**

**Initialization:**
- The number $T$ of time slots in one epoch, the number $k_{\text{max}}$ of all epochs.
- User selects power $P$ randomly, $t = 1, k = 0$.

**Outer Iteration:**
1. Jammer selects the optimal arm $J(k + 1)$ according to Equation (21).

**Inner Iteration:**
1. In the $t$th time slot, the user selects its joint relay and power strategy according to $q(t)$.
2. Obtain the user’s utility $U(t)$ through the feedback of SINR at the end of current time slot according to Equation (5).
3. Update the user’s Q values according to Equation (14).
4. Update the user’s mixed strategy $q(t)$ according to Equation (15).
5. Time slot $t = t + 1$.
6. Return to (1) until $t > T$.

**End Inner Iteration.**
2. Obtain jammer’s utility $V(k)$ through eavesdropping the feedback of SINR in the $k$th epoch according to Equation (6).
3. Update $A_l(k), \mu_l$ and $B(k)$ in turn.
4. Epoch $k = k + 1, t = 1$.
5. Return to 1 until $k > k_{\text{max}}$.

**End Outer Iteration.**

4. Simulation Results and Discussions

In this section, the simulation parameters and system location setting are given, and then we present some necessary simulation results and give brief discussions. In Section 4.1, we give the user’s power strategy probability and relay selection probability respectively. Moreover, the jammer’s regret is also given. In Section 4.2, we analyze the influence of feedback error and eavesdropping error, and we also compare the user’s and jammer’s utility under the proposed hierarchical joint optimization algorithm (HJOA) with random selection algorithm to verify the algorithm’s superiority.

In the simulation, we assume the user’s and jammer’s discrete power sets are $\mathcal{P} = \mathcal{J} = [0.5 \text{W}, 1 \text{W}, 1.5 \text{W}, 2 \text{W}, 2.5 \text{W}]$, and their transmission costs are $C_u = C_j = 0.1$. The relay set is
\( R = [R_1, R_2, R_3, R_4] \) and different relay has the same power \( Q = 2 \) W. The number of time slots in one epoch \( T = 100 \). Path-loss factor \( \gamma = 2 \).

As shown in Figure 3, there exist a BS, a user, a jammer and a relay group in the investigated scenario. The coordinates of BS and the user are (0 km, 10 km) and (10 km, 0 km). The jammer is located in (10 km, 10 km). The relay group consisted of four relays, which were located in (2.5 km, 2.5 km), (5 km, 2.5 km), (5 km, 5 km), (7.5 km, 5 km) respectively. The red arrow is the jamming signal and blue arrow is the communication signal.

![Figure 3](image)

Figure 3. Location of base station (BS), the user, jammer and relay group.

### 4.1. User’s Strategy Selection Probability and Jammer’s Regret

Figure 4 shows the user’s power selection probability in one epoch, and it converges to a stationary mixed strategy after about 70 time slots. All power strategies have possibilities to be selected which can increase the randomness of strategy and fool jammer effectively. It is easy to find that the user tends to choose power strategy \( P_3 \) and tends not to choose \( P_1 \) and \( P_2 \), that is because a higher transmission power can guarantee the SINR at BS. However, an exorbitant power causes the increase of power cost, so the possibility of \( P_3 \) is higher than \( P_4 \) and \( P_5 \), which realizes the tradeoff of anti-jamming effect and power cost at the same time.

![Figure 4](image)

Figure 4. User power selection probability.

Figure 5 shows the user’s relay selection probability in one epoch, and it converges to a stationary mixed strategy after about 50 time slots. We can find the selection probabilities of \( R_1, R_2, R_3, R_4 \) are
about 0.247, 0.355, 0.204, 0.194 respectively when they have achieved convergence. That is because though $R_3$ is closer to BS and the user, but the distance between $R_3$ and the jammer is also shorter. So in order to minimize the jamming influence of the jammer, the user tends to choose $R_2$ and $R_4$ even they are further from BS and the user compared to $R_3$ and $R_4$. Considering the influence of distance on channel fading, the user tends to select $R_2$ compared to $R_1$.

![Figure 5. Relay selection probability.](image)

Figure 5. Relay selection probability.

Figure 6 shows the jammer’s regret in 1000 epochs, which denotes the loss of payoff due to the fact that the optimal strategy is not always chosen during the decision-making process. We can find the jammer’s regret grows nearly logarithmically, i.e., it grows quickly at the start and grows slowly subsequently, which means the loss caused by the wrong choice of the optimal strategy is smaller and smaller. However, regret still grows because jammer selects other non-optimal strategies occasionally to avoid missing a future potential optimal strategy, which realizes the exploration and exploitation simultaneously.

![Figure 6. Regret of jammer.](image)

Figure 6. Regret of jammer.

### 4.2. Utility Comparison under Different Algorithms with Feedback Error and Eavesdropping Error

In one epoch, we average the user’s utility every 20 time slots and obtain the comparison of the user’s utility under the proposed HJOA algorithm and random selection algorithm, as shown in Figure 7. Moreover, we analyzed the influence of feedback error on utility. Under the HJOA
algorithm, we can find the user’s utility grows firstly and remains stable then, that is because the user has gradually obtained the optimal strategy with the iterations of algorithm. When the user can receive correct feedback, i.e., $\varepsilon_1 = 0$, we can find the user’s utility under the HJOA algorithm reaches the maximum value and is improved compared with random selection algorithm. The greater feedback error $\varepsilon_1$ is, the larger deviation between the user’s received feedback and actual value is, which causes the decrease of utility.

![Figure 7](image1.png)

**Figure 7.** User’s utility comparison under different algorithms with feedback error.

In Figure 8, we average the jammer’s utility every 20 epochs and obtain the comparison of the user’s utility under the proposed algorithm and random selection algorithm. We analyze the influence of eavesdropping error on utility. Under the proposed the HJOA algorithm, we can find that the jammer’s utility grows quickly firstly and grows slower then, that is because the jammer has gradually obtained the optimal strategy. When the jammer can eavesdrop the correct feedback, i.e., $\varepsilon_2 = 0$, its utility reaches the maximum value and is higher than utility under random selection algorithm. The greater feedback error $\varepsilon_2$ is, the lower the jammer’s utility is because the existence of $\varepsilon_2$ causes the jammer cannot eavesdrop the user’s feedback information accurately and influences its decision-making.

![Figure 8](image2.png)

**Figure 8.** Jammer’s utility comparison under different algorithms with eavesdropping error.

5. Conclusions

In this paper, we studied the joint relay selection and power control optimization problem in an anti-jamming relay communication system via Stackelberg game, in which a user acted as the leader and a jammer acted as follower. Based on the Q-learning algorithm and multi-armed bandit
method, a hierarchical joint optimization algorithm was proposed. Simulation results showed the user’s strategy selection probability and jammer’s regret. Moreover, we analyzed the influence of feedback error and eavesdropping error, and compared the user’s and jammer’s utilities under the proposed algorithm with random selection algorithm to verify the algorithm’s superiority. In the future, we will consider the dynamic change of the user’s and jammer’s position in the multi-user scenario to improve the anti-jamming performance.

Author Contributions: Z.F., G.R. and J.C. conceived of and designed the model. Z.F. performed the theoretical analysis and simulation. Z.F., G.R. and J.C. analyzed the simulation result and wrote the paper. C.C., X.Y., Y.L. and K.X. provided some valuable suggestions for this paper.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 61671473, No. 61771488 and No. 61801492, in part by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant No. BK20160034, in part by the National Science Foundation of China under Grant No. 61801492 and the National University of Defense Technology with contract No. ZK18-03-20.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Meng, D.; Feng, Z.; Lu, M. Anti-jamming with adaptive arrays utilizing power inversion algorithm. Tsinghua Sci. Technol. 2008, 13, 796–799. [CrossRef]
2. Yang, D.; Xue, G.; Zhang, J.; Richa, A.; Fang, X. Coping with a smart jammer in wireless networks: A Stackelberg game approach. IEEE Trans. Wirel. Commun. 2013, 12, 4038–4047. [CrossRef]
3. Xiao, L.; Chen, T.; Liu, J.; Dai, H.; Anpalagan, A. Anti-jamming transmission Stackelberg game with observation errors. IEEE Commun. Lett. 2015, 19, 949–952. [CrossRef]
4. Jia, L.; Yao, F.; Sun, Y.; Niu, Y.; Zhu, Y. Bayesian Stackelberg Game for Anti-Jamming Transmission with Incomplete Information. IEEE Commun. Lett. 2016, 20, 1991–1994. [CrossRef]
5. El-Bardan, R.; Brahma, S.; Varshney, P.K. Strategic power allocation with incomplete information in the presence of jammer. IEEE Trans. Commun. 2016, 64, 3467–3479. [CrossRef]
6. Chen, G.; Zhan, Y.; Chen, Y.; Xiao, L.; Wang, Y.; An, N. Reinforcement Learning Based Power Control for In-Body Sensors in WBANs Against Jamming. IEEE Access 2018, 6, 37403–37412. [CrossRef]
7. Han, C.; Niu, Y. Cross-Layer Anti-Jamming Scheme: A Hierarchical Learning Approach. IEEE Access 2018, 6, 34874–34883. [CrossRef]
8. Jia, L.; Yao, F.; Sun, Y.; Xu, Y.; Feng, S.; Anpalagan, A. A Hierarchical Learning Solution for Anti-Jamming Stackelberg Game with Discrete Power Strategies. IEEE Wirel. Commun. Lett. 2017, 6, 818–821. [CrossRef]
9. Slimeni, F.; Nir, V. I.; Scheers, B.; Chtourou, Z.; Attia, R. Optimal power allocation over parallel Gaussian channels in cognitive radio and jammer games. IET Commun. 2016, 10, 980–986. [CrossRef]
10. Jia, L.; Xu, Y.; Sun, Y.; Feng, S.; Anpalagan, A. Stackelberg game approaches for anti-jamming defence in wireless networks. IEEE Wirel. Commun. Mag. 2018, 25, 120–128. [CrossRef]
11. Gao, Y.; Xiao, Y.; Wu, M.; Xiao, M.; Shao, J. Game Theory-Based Anti-Jamming Strategies for Frequency Hopping Wireless Communications. IEEE Trans. Wirel. Commun. 2018, 17, 5314 – 5326. [CrossRef]
12. Xu, Y.; Anpalagan, A.; Wu, Q.; Shen, L.; Gao, Z.; Wang, J. Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions. IEEE Commun. Surv. Tutor. 2013, 15, 1689–1713. [CrossRef]
13. Zhang, L.; Wang, H.; Li, T. Anti-Jamming Message-Driven Frequency Hopping—Part I: System Design. IEEE Trans. Wirel. Commun. 2013, 12, 70–79. [CrossRef]
14. Xu, Y.; Wang, J.; Wu, Q.; Zheng, J.; Shen, L.; Anpalagan, A. Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization. IEEE Trans. Commun. 2017, 65, 5305–5318. [CrossRef]
15. Shi, R.; Du, Y. Analysis on frequency diversity and anti-jamming characteristic of TDCS signal. In Proceedings of the 2017 IEEE 17th International Conference on Communication Technology (ICCT), Chengdu, China, 27–30 October 2017; pp. 27–30.
16. Liu, X.; Xu, Y.; Jia, L.; Wu, Q.; Anpalagan, A. Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach. IEEE Commun. Lett. 2018, 22, 998–1001. [CrossRef]
17. Zhu, Y.; Guo, C.; Wang, X. Throughput analysis of dynamic spectrum anti-jamming multiple-access in HF communication systems. China Commun. 2018, 15, 85–94. [CrossRef]
18. Zhu, H.; Fang, C.; Liu, Y.; Chen C.; Li, M.; Shen, X. You Can Jam But You Cannot Hide: Defending against Jamming Attacks for Geo-Location Database Driven Spectrum Sharing. IEEE J. Sel. Areas Commun. 2016, 34, 2722–2737. [CrossRef]
19. Xu, Y.; Wang, J.; Wu, Q.; Anpalagan, A.; Yao, Y. Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution. IEEE Trans. Wirel. Commun. 2012, 11, 1380–1391. [CrossRef]
20. Salvatore, D.; Laura, G.; Giacomo, M.; Sergio, P.; Lin, C.; Fabio, M. Defeating Jamming with the Power of Silence: A Game-Theoretic Analysis. IEEE Trans. Wirel. Commun. 2015, 15, 2337 – 2352.
21. He, H.; Zhang, X. Application of improve subspace projection technique in anti-jam of GNSS space-time receiver. In Proceedings of the 2012 5th International Congress on Image and Signal Processing, Chongqing, China, 16–18 October 2012.
22. Du, Y.; Gao, Y.; Liu, J.; Xi, X. Frequency-Space Domain Anti-Jamming Algorithm Assisted with Probability Statistics. In Proceedings of the 2013 International Conference on Information Technology and Applications, Chengdu, China, 16–17 November 2013.
23. Liu, Y.; Zhang, S.; Shi, D.; Shen, Y.; Gao, Y. Anti-jamming space-time processor with digital beamformer for satellite navigation. In Proceedings of the 2015 7th Asia-Pacific Conference on Environmental Electromagnetics (CEEM), Hangzhou, China, 4–7 November 2015: pp. 31–35.
24. Wang, D.; Chen, Q. Reduced rank calculation of space-time anti-jamming for navigation receiver. In Proceedings of the 2013 International Conference on Mechatronics Sciences, Electric Engineering and Computer (MEC), Shenyang, China, 20–22 December 2013.
25. Xiao, A.; Zhong, B. Design for a dual-system satellite navigation anti-jamming receiver. In Proceedings of the 2015 Chinese Automation Congress (CAC), Wuhan, China, 27–29 November 2015.
26. Laneman, G. N.; Tse, D. N. C.; Wornell, G. W. Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior. IEEE Commun. Mag. 2004, 50, 3062–3080. [CrossRef]
27. Zhang, Z.; Wu, H.; Zhang, B.; Peng, J. Intelligent Anti-Jamming Relay Communication System Based on Reinforcement Learning. In Proceedings of the 2019 2nd International Conference on Communication Engineering and Technology (ICCET), Nagoya, Japan, 12–15 April 2019.
28. Zhong, W.; Chen, G.; Jin, S.; Wong, K. Relay Selection and Discrete Power Control for Cognitive Relay Networks via Potential Game. IEEE Trans. Signal Process. 2014, 62, 5411–5424. [CrossRef]
29. Gu, P.; Hua, C.; Khatoun, R.; Wu, Y.; Serhrrouchni, A. Cooperative Anti-Jamming Relaying for Control Channel Jamming in Vehicular Networks. IEEE Trans. Veh. Technol. 2018, 67, 7033–7046. [CrossRef]
30. Xiao, L.; Li, Y.; Liu, J.; Zhao, Y. Power control with reinforcement learning in cooperative cognitive radio networks against jamming. J. Supercomput. 2015, 71, 3237–3257. [CrossRef]
31. Liu, W.; Tan, D.; Xu, G. Low complexity power allocation and joint relay-jammer selection in cooperative jamming DF relay wireless secure networks. In Proceedings of the 2013 International Conference on Anti-Counterfeiting, Security and Identification (ASID), Shanghai, China, 25–27 October 2013.
32. Liu, K.J.R.; Wang, B. Cognitive Radio Networking and Security: A Game-Theoretic View; Cambridge University Press: Cambridge, UK, 2010.
33. Naveen, K.P.; Altman, E.; Kumar, A. Competitive Selection of Ephemeral Relays in Two-Way Untrusted Relaying Networks. IEEE J. Sel. Areas Commun. 2017, 35, 586–600. [CrossRef]
34. Kuhestani, A.; Mohammadi, A.; Yeoh, P.L. Optimal Power Allocation and Secrecy Sum Rate in Two-Way Untrusted Relaying Networks With an External Jammer. IEEE Trans. Commun. 2018, 66, 2671–2684. [CrossRef]
35. Mamaghani, M.T.; Kuhestani, A.; Wong, K. Secure Two-Way Transmission via Wireless-Powered Untrusted Relay and External Jammer. IEEE Trans. Veh. Technol. 2018, 67, 8451–8465. [CrossRef]
36. Feng, Z.; Ren, G.; Chen, J.; Zhang, X.; Luo, Y.; Wang, M.; Xu, Y. Power Control in Relay-assisted Anti-jamming Systems: A Bayesian Three-layer Stackelberg Game Approach. IEEE Access 2019, 7, 14623–14636. [CrossRef]
37. Garnaev, A.; Trappe, W.; Petropulu, A. Combating Jamming in Wireless Networks: A Bayesian Game with Jammer’s Channel Uncertainty. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019.
38. Garnaev, A.; Liu, Y.; Trappe, W. Anti-jamming Strategy Versus a Low-Power Jamming Attack When Intelligence of Adversary’s Attack Type is Unknown. *IEEE Trans. Signal Inf. Process. Netw.* 2016, 2, 49–56. [CrossRef]

39. Li, M.; Koutsopoulos, I.; Poovendran, R. Optimal Jamming Attacks and Network Defense Policies in Wireless Sensor Networks. In Proceedings of the 26th IEEE International Conference on Computer Communications, Barcelona, Spain, 6–12 May 2007.

40. Luo, Y.; Feng, Z.; Jiang, H.; Yang Y.; Huang, Y.; Yao, J. Game-theoretic Learning Approaches for Secure D2D Communications Against Full-duplex Active Eavesdropper. *IEEE Access* 2019, 7, 41324–41335. [CrossRef]

41. Tang, X.; Ren, P.; Wang, Y.; Hang, Z. Combating full-duplex active eavesdropper: A hierarchical game perspective. *IEEE Trans. Commun.* 2017, 65, 1379–1395. [CrossRef]

42. Qu, J.; Cai, Y.; Zheng, J. Power allocation for device-to-device communication underlaying cellular networks under a probabilistic eavesdropping scenario. *Ann. Telecommun.* 2016, 71, 389–398. [CrossRef]

43. Goldsmith, A. *Wireless Communications*; Cambridge University Press: Cambridge, UK, 2005.

44. Chen, X.; Zhang, H.; Chen, T.; Lasanen, M. Improving energy efficiency in green femtocell networks: A hierarchical reinforcement learning framework. In Proceedings of the 2013 IEEE International Conference on Communications (ICC), Budapest, Hungary, 9–13 June 2013; pp. 2241–2245.

45. Sun, Y.; Shao, H.; Liu, X.; Zhang, J.; Qiu, J.; Xu, Y. Traffic offloading in two-tier multi-mode small cell networks over unlicensed bands: A hierarchical learning framework. *TIIS* 2015, 9, 4291–4310.

46. Han, Z.; Niyato, D.; Saad, W.; Başar, T.; Hjørungnes, A. *Game Theory in Wireless and Communication Networks*; Cambridge University Press: Cambridge, UK, 2012.

47. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* 1992, 8, 279–292. [CrossRef]

48. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* 2002, 47, 235–256. [CrossRef]

49. Lai, T.; Robbins, H. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* 1985, 6, 4–22. [CrossRef]

50. Agrawal, R. Sample mean based index policies with O(log n) regret for the multi-armed bandit problem. *Adv. Appl. Probab.* 1995, 27, 1054–1078. [CrossRef]