Optimal Dwell Time for Frequency Hopping in a Stackelberg Game with a Smart Jammer

Long Yu(✉), Yonggang Zhu, and Yusheng Li

Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China
elong025@163.com

Abstract. Frequency hopping (FH) technique is usually used to anti-jamming communication. Frequency dwell time is an important parameter for FH communication. Short dwell time will reduce the communication efficiency due to frequency switching time, while long dwell time will increase the time to be jammed after the sensing of a smart jammer. The dwell time of the cognitive user and the sensing time of the jammer are interactive. We formulate the interactions between the user and the jammer as a Stackelberg game. The jammer first senses the user's operating frequency and then jams the user based on the sensing result. The user determines its dwell time according to the reward under the jamming. A tiered reinforcement learning algorithm is proposed to solve the game. The optimal dwell time of the user is given when the Stackelberg Equilibrium is achieved.

Keywords: Frequency hopping · Anti-jamming · Dwell time · Stackelberg game · Tiered reinforcement learning algorithm

1 Introduction

Wireless networks are suffering more and more security threats [1, 2]. Jamming attack is one of the vital threats where the jammer jams the communication process of the users by radiating high power signal. To cope with the jamming attack, various techniques have been proposed. Frequency hopping [3, 4] is one of the efficient anti-jamming techniques where the user's operating frequency hops from one to another with time slots. High-dimensional modulation [5–7], message driven methods [8, 9], M-ary orthogonal Walsh sequence keying modulation [10], families of sequences with good correlations [11], applied in the frequency hopping technique, have been researched in the previous works. However, these works have not considered the presence of the smart jammer with cognitive and reconfigurable abilities. In this paper, we focus on the anti-jamming strategy of FH system to cope with a smart jammer.

In [12, 13], a Stackelberg game was formulated, in which the players are a cognitive user and a smart jammer. Further, to solve the incomplete information problem in the game, an anti-jamming Bayesian Stackelberg game was proposed [14]. The optimal strategies based on duality optimization theory were derived. However, those works all based on the assumption that the jammer senses the user's power correctly. In practice

X. B. Zhai et al. (Eds.): MLICOM 2019, LNICST 294, pp. 410–418, 2019.
https://doi.org/10.1007/978-3-030-32388-2_35

scenario, the sensing results may be error, and the sensing performance is relevant with the sensing time, signal power, etc.

In this paper, we expect to get the optimal frequency dwell time of FH with consideration of the detection performance of a smart jammer. Frequency dwell time is an important parameter for FH communication. Short dwell time will reduce the communication efficiency due to the frequency switching time, while long dwell time will increase the time to be jammed after the sensing of the smart jammer. Sensing time is also a key parameter for the jammer. Short sensing time will decrease the sensing performance while long sensing time will shorten the jamming time. It is obvious that the dwell time of the user and the sensing time of the jammer are interactive. In this paper, we formulate the interaction as a Stackelberg game. The jammer first senses the user's operating frequency and then jams the user based on the sensing result. The user determines its dwell time according to the reward under the jamming. A tiered reinforcement learning algorithm is proposed to solve the game. The optimal dwell time of the user is obtained when the Stackelberg Equilibrium is achieved.

2 System Model

It consists of a user (a transmitter-receiver pair) and a jammer in the system. Both of the user and the jammer are equipped with a single radio and work with time slotted. The slot structure of the user and the jammer are shown in Fig. 1. The parameters for the user and the jammer are assumed to remain unchanged during a time slot.

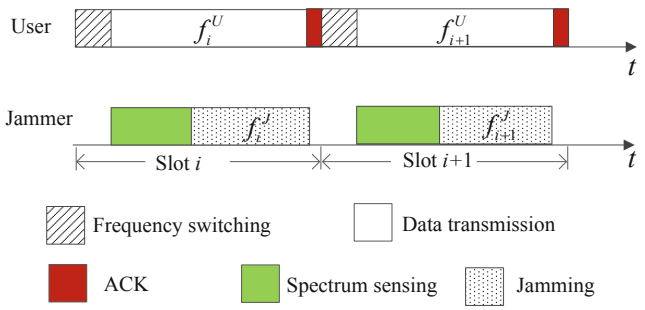

Fig. 1. Transmission structure of the user and the jammer.

The user hops from one frequency to another along with time slot to avoid jamming. The user's frequency f^U is selected from a frequency set $\mathcal{F}$ with $|\mathcal{F}| = M$. Denote Γ as the dwell time, which is the duration between two adjacent frequency switching points. Since transmission cannot be started immediately due to the settling time of radio frequency devices after tuning the frequency of the transceivers, a fixed frequency switching time t_c is considered in each time slot.

The jammer starts to transmit jamming signals after sensing at each time slot. This type of jammer is referred to as reactive jammer [2]. Since the user works with

frequency hopping mode, the jammer's sensing objective is to detect which frequency the user operates on. The sensing time of the jammer is denoted as τ. Based on the sensing result the reactive jammer obtains the user's operating frequency estimation f^J. Then, the reactive jammer will jam the frequency f^J.

3 Problem Formulation

After sensing, the jammer gets results. One result is that the jammer correctly detects the user's frequency, that is, $f^J = f^U$. The user will be jammed after the jammer sensing. The immediate payoff u_0 of the user in this case is defined as:

$$u_0 = \frac{1}{\Gamma}[\tau C_0 + (\Gamma - t_c - \tau)C_1], \tag{1}$$

where C_0 and C_1 represent the channel capability without and with jamming, respectively. The other result is that the jammer detects the user's frequency incorrectly, that is, $f^J \neq f^U$. The user will not be jammed during the slot in this case. The immediate payoff u_1 of the user in this case is defined as:

$$u_1 = \frac{1}{\Gamma}[(\Gamma - t_c)C_0]. \tag{2}$$

Based on the two results the user gets an expected payoff. Define $P_d(\tau)$ as the correct detection probability of the jammer with sensing time τ. The utility function of the user can be expressed as:

$$\begin{aligned} u(\Gamma, \tau) &= P_d(\tau)u_0 + (1 - P_d(\tau))u_1 \\ &= \frac{1}{\Gamma}((\Gamma - t_c)C_0 - P_d(\tau)(\Gamma - t_c - \tau)(C_0 - C_1)) \end{aligned} \tag{3}$$

$P_d(\tau)$ can be expressed as:

$$P_d(\tau) = \sum_{m=0}^{M-1} (-1)^m \binom{M-1}{m} \frac{1}{m+1} \exp\left(\frac{-m}{2(m+1)} \frac{gP\tau}{\sigma^2 M}\right), \tag{4}$$

where P is the user's transmission power, g is the channel gain between the user and the jammer, and σ^2 is the noise power.

The user expects to maximize the utility. The optimization problem for the user can be expressed as:

$$\max_{\Gamma} u(\Gamma, \tau). \tag{5}$$

Opposite to the user, the jammer expects to decrease the user's payoff. The instant return of the jammer in the jth time slot is expressed as:

$$v(j) = \frac{1}{T}[(\beta(t_d + \tau) + (1 - \beta)(T - t_c))]C_0, \tag{6}$$

where β is the indication function. $\beta = 1$ represents that the jamming is successful, and $\beta = 0$ represents that the jamming is failed. The utility of the jammer is defined as:

$$v_J(\Gamma, \tau) = I - u(\Gamma, \tau), \tag{7}$$

where I is a constant value to grantee v positive. From the perspective of the user, the optimization problem can be expressed as:

$$\max_{\tau} \ v_J(\Gamma, \tau). \tag{8}$$

It can be seen that the optimization problems of the user and the jammer are mutually influential. Hence, the problem can be formulated as a game. Since the scenario is that jammer adjusts its own strategy after sensing the user, a Stackelberg game can be used. The user is set as leader and the jammer is set as follower. Mathematically, the Stackelberg game is expressed as $\mathcal{G} = \{\mathcal{N}, \mathcal{T}, \mathcal{S}, u, v\}$, where $\mathcal{N}$ denotes the player set including the user and the jammer, $\mathcal{T}$ and $\mathcal{S}$ represent the strategy space of the user and the jammer, respectively. In the game, the user selects the dwell time Γ from a discrete strategy space $\mathcal{T}$ in each time slot, where $\mathcal{T} \triangleq \{T_1, T_2, \ldots, T_W\}$, and T_i is the ith optional action in the space $\mathcal{T}$. The jammer selects its sensing time τ from its strategy space $\mathcal{S}$, where $\mathcal{S} \triangleq \{S_1, S_2, \ldots, S_L\}$, and S_i is the ith optional action in the space $\mathcal{S}$.

4 Tiered Reinforcement Learning Algorithm

In the game, since there is no information interaction between the user and the jammer, the two parties can only choose to optimize their own strategies based on the observation on the other's strategy. Because the two strategies are mutually influential, it is very suitable to solve the game using a tiered reinforcement learning algorithm.

The algorithm is performed with two layers, the upper layer and the lower layer. The upper layer subject is the user, and the lower layer subject is the jammer. First, the user selects an action. The jammer learns the optimal response policy under this action. Then, the user calculates the reward under the policy selected by the jammer and updates its own action accordingly. Again, the jammer learns and loops until both the user and the jammer learn the optimal response policy.

The user and jammer updates their policies with different time scales. The frame structure of the tiered reinforcement learning algorithm is shown in Fig. 2. The user updates the policy each epoch, and the jammer updates its policy in each time slot. $R(k)$ represents the number of time slots in the kth epoch. Since the epoch duration is greater

than the slot duration, the user has plenty of time to coordinate the receiver and transmitter when the dwell time is changed.

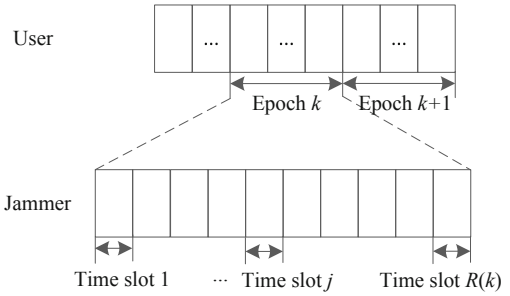

Fig. 2. The frame structure of the tiered reinforcement learning algorithm

In the tiered reinforcement learning procedure, both the user and the jammer expect an optimal long-term return, called an average cumulative reward. In the lower layer, the average cumulative reward vector of the jammer $\mathbf{Q}^J = \left(Q_1^J, Q_2^J, \ldots, Q_L^J\right)$, where Q_i^J is the average cumulative reward with the action S_i. When the jammer selects the action S_i, its average cumulative reward is updated as follows:

$$Q_i^J(j+1) = Q_i^J(j) + \frac{1}{\eta_i(j)}\left(v(j) - Q_i^J(j)\right), \tag{9}$$

where $\eta_i(j)$ is the times that the action S_i is selected within j time slot, and $v(j)$ is the instant return of the jammer in the jth time slot.

The jammer selects its action according to the following rules:

$$\tau(j) = \begin{cases} \text{selected from } \mathcal{S} \text{ randomly with uniform distribution}, & \text{with probability } \delta_J(j), \\ \arg\max_i Q_i^J(j), & \text{with probability } 1 - \delta_J(j). \end{cases} \tag{10}$$

where $\tau(j)$ is the action selected by the jammer during the jth time slot, and $\delta_J(j)$ is the temperature coefficient of the jammer during the jth slot. $\delta_J(j)$ is used to control the tradeoff between exploration and exploitation in the learning process.

In the upper layer, the average cumulative reward vector of the user is defined as $\mathbf{Q}^u$, where $\mathbf{Q}^u = \left(Q_1^u, Q_2^u, \ldots, Q_W^u\right)$. Q_i^u is the average cumulative reward with the action T_i. When the user selects the action T_i, its average cumulative reward is updated as:

$$Q_i^u(k+1) = Q_i^u(k) + \frac{1}{\kappa_i(k)}\left(\hat{u}_i(k) - Q_i^u(k)\right), \tag{11}$$

where $\kappa_i(k)$ is the times that the action T_i is selected within k epochs, and $\hat{u}_i(k) = \sum_{j=1}^{R(k)} u(j)/R(k)$, represents the average reward of the user in the kth epoch with the strategy T_i.

The user's action is updated according to:

$$\Gamma(k) = \begin{cases} \text{selected from } \mathcal{T} \text{ randomly with uniform distribution}, & \text{with probability } \delta_u(k), \\ \arg\max_i Q_i^u(k), & \text{with probability } 1-\delta_u(k). \end{cases} \tag{12}$$

where $\Gamma(k)$ is the action selected by the user during the kth epoch, and $\delta_u(k)$ is the temperature coefficient of the user during the kth epoch.

The specific flow of the algorithm is shown in Algorithm 1.

Step 1: Initilization. Set k=1, j=1. Initialize the average cumulative reward vector $\mathbf{Q}^u$ and $\mathbf{Q}^J$ 。

Step 2: In the kth epoch, the user selects an action from the strategy set according to Equation (12).

Step 3: The learning process of the jammer.

(1) In time slot j, the jammer selects its action from the strategy space according to Equation (10).

(2) The jammer measures its instant return using Equation (6).

(3) The jammer updates $\mathbf{Q}^J$ according to Equation (7).

(4) Update $j = j+1$, go to (1), until $j = R(k)$.

Step 4: The user measures its average reward $\hat{u}(k)$ in the learning process of the jammer.

Step 5: The user updates $\mathbf{Q}^u$ according to (11).

Step 6: Update $k = k+1$. Set j=1. Go to Step 2, and until the stopping criterion holds.

---

5 Numerical Results

In this section, simulation results are presented. The strategy space of the user is set as $\left\{\frac{1}{1500}, \frac{1}{500}, \frac{1}{200}, \frac{1}{10}\right\}$, while the strategy space of the jammer is set as $\{m/1000,\ m = 1, 2, \ldots, 9\}$. Each epoch contains 1000 time slots. The channel gain between the user and the jammer is 10^{-3}. The frequency switching time t_c is set as 50 μs. The temperature coefficient of the user and the jammer are given as $0.5/1.01^k$ and $0.5/1.001^j$, respectively.

The convergence behavior of the user is given in Fig. 3. After learning, the selection probability of the user converges to a stationary mixed strategy. The convergence behavior of the jammer is given in Fig. 4 where the user's dwell time is 1/1500 s. It is seen that the selection probability of the jammer also converges to a stationary mixed strategy.

Figure 5 shows the optimal average cumulative reward of the user under different transmission power. It can be found that as the number of epoch increases, the average cumulative reward of the user at different transmission power converges to a steady value. When the user power is 0.01 W, the average cumulative reward of the user hardly changes as the number of epoch increases. This is because when the user's power is very low, the detection probability of the jammer is almost zero. At this time, the average utility of the user is almost independent of the jammer's sensing time, but only related to the user's dwell time. Since the user's frequency switching time is almost negligible compared to the user's dwell time, no matter which action is selected by the user, the utility is almost the same. When the user power is 10 W, the steady-state value of the user's average cumulative reward is reduced compared to that when the user power is 0.1 W and 1 W. This is because as the power of the user increases, the detection probability of the jammer increases, and the optimal sensing time decreases. This increases the jamming duration, which makes the user's average accumulative reward reduced. Therefore, in the presence of a smart jammer, an increase in user's power does not necessarily increase the user's utility, but may reduce it.

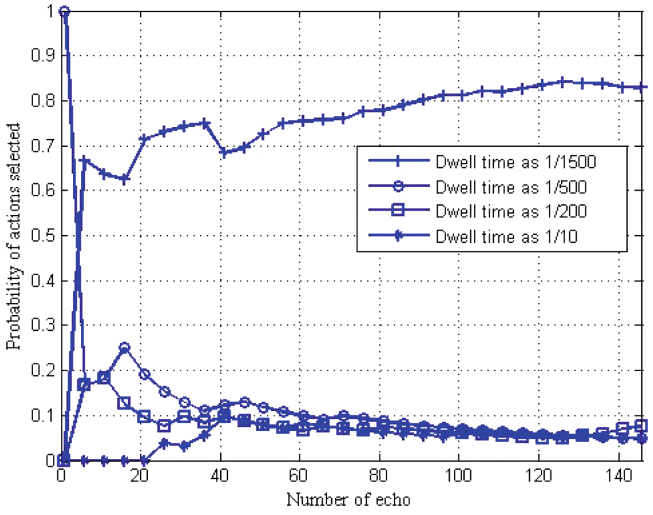

Fig. 3. The convergence behavior of the user.

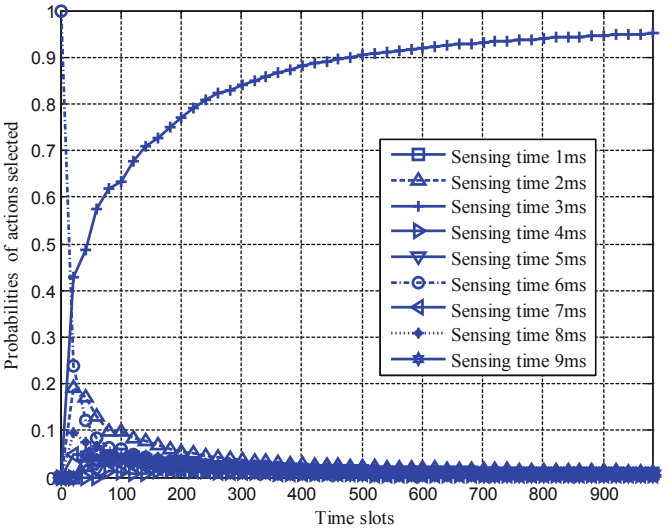

Fig. 4. The convergence behavior of the jammer.

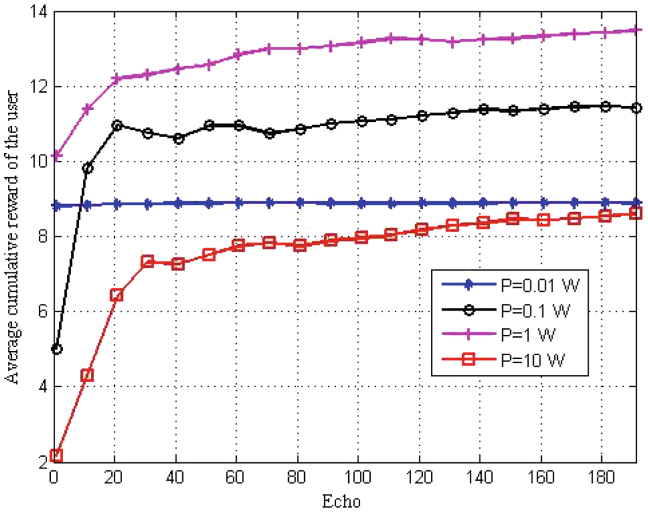

Fig. 5. Convergence process of the user's average cumulative reward under different transmission power.

6 Conclusions

In this paper, the interaction between the dwell time of the user and the sensing time of the jammer is formulated as a Stackelberg game. The jammer first senses the user's operating frequency and then jams the user based on the sensing result. The user

determines its dwell time according to the reward under the jamming. A tiered reinforcement learning algorithm is proposed to solve the game. The optimal dwell time of the user is given when the Stackelberg Equilibrium is achieved.

References

1. Sharma, R.K., Rawat, D.B.: Advances on security threats and countermeasures for cognitive radio networks: a survey. IEEE Commun. Surv. Tutorials **17**(2), 1023–1043 (2015)
2. Zou, Y., Zhu, J., Wang, X., Hanzo, L.: A survey on wireless security: technical challenges, recent advances, and future trends. Proc. IEEE **104**(9), 1727–1765 (2016)
3. Hanawal, M.K., Abdel-Rahman, M.J., Krunz, M.: Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems. IEEE Trans. Mob. Comput. **15**(9), 2247–2259 (2016)
4. Yu, L., Xu, Y., Wu, Q., et al.: Self-organizing hit avoidance in distributed frequency hopping multiple access networks. IEEE Access **5**, 26614–26622 (2017)
5. Simon, M., Polydoros, A.: Coherent detection of frequency-hopped quadrature modulations in the presence of jamming—Part I: QPSK and QASK modulations. IEEE Trans. Commun. **29**(11), 1644–1660 (1981)
6. Choi, K., Cheun, K.: Maximum throughput of FHSS multiple-access networks using MFSK modulation. IEEE Trans. Commun. **52**(3), 426–434 (2004)
7. Peng, K.-C., Huang, C.-H., Li, C.-J., Horng, T.-S.: High-performance frequency-hopping transmitters using two-point delta-sigma modulation. IEEE Trans. Microw. Theory Techn. **52**(11), 2529–2535 (2004)
8. Ling, Q., Li, T.: Message-driven frequency hopping: design and analysis. IEEE Trans. Wirel. Commun. **8**(4), 1773–1782 (2009)
9. Zhang, L., Wang, H., Li, T.: Anti-jamming message-driven frequency hopping—Part I: system design. IEEE Trans. Wirel. Commun. **12**(1), 70–79 (2013)
10. Cho, J., Kim, Y., Cheun, K.: A novel frequency-hopping spread spectrum multiple-access network using M-ary orthogonal Walsh sequence keying. IEEE Trans. Commun. **51**(11), 1885–1896 (2003)
11. Bao, J., Ji, L.: Frequency hopping sequences with optimal partial Hamming correlation. IEEE Trans. Inf. Theory **62**(6), 3768–3783 (2016)
12. Yang, D., et al.: Coping with a smart jammer in wireless networks: a Stackelberg game approach. IEEE Trans. Wirele. Commun. **12**(8), 4038–4047 (2013)
13. Xiao, L., et al.: Anti-jamming transmission stackelberg game with observation errors. IEEE Commun. Lett. **19**(6), 949–952 (2015)
14. Jia, L., et al.: Bayesian Stackelberg game for anti-jamming with incomplete information. IEEE Commun. Lett. **20**(10), 1991–1994 (2016)