IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 15, 2018, accepted June 6, 2018, date of publication June 27, 2018, date of current version August 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2850879

A Multi-Domain Anti-Jamming Defense Scheme in Heterogeneous Wireless Networks

LULIANG JIA[1,2], (Student Member, IEEE), YUHUA XU[1,2], (Member, IEEE), YOUMING SUN[1], (Student Member, IEEE), SHUO FENG[1], (Student Member, IEEE), LONG YU[1], AND ALAGAN ANPALAGAN[3], (Senior Member, IEEE)

[1]College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China
[2]Science and Technology on Communication Networks Laboratory, Shijiazhuang 050002, China
[3]Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada

Corresponding author: Yuhua Xu (yuhuaenator@gmail.com)

This work was supported in part by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant BK20160034, in part by the National Science Foundation of China under Grant 61631020, Grant 61671473, Grant 61401508, and Grant 61401505, in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20130069 and Grant BK20151450, and in part by the Open Research Foundation of Science and Technology in the Communication Networks Laboratory.

ABSTRACT In this paper, we investigate the anti-jamming problem in heterogeneous wireless networks. Although there are many studies on the anti-jamming defense problem in both power domain and spectrum domain, these two important aspects were addressed separately. In this paper, to cope with the jamming attacks flexibly, we study the anti-jamming defense problem from a multi-domain perspective, which includes both power domain and spectrum domain, and a multi-domain anti-jamming scheme (MDAS) is proposed. To be more specific, a Stackelberg power game is formulated in the power domain to fight against the jamming attacks, and a multi-armed bandit-based channel selection with a channel switching cost and unknown channel availability state information is formulated in the spectrum domain. Besides, we analyze the performance of the formulated Stackelberg power game and derive the optimal power strategy and utility of a legitimate user. In addition, it is proved that the proposed anti-jamming scheme has a logarithmic regret. Finally, extensive simulations are conducted to validate the performance of the proposed MDAS.

INDEX TERMS Anti-jamming, power control, channel selection, Stackelberg game, multi-armed bandit (MAB).

I. INTRODUCTION

The security problem of wireless communication networks has gained growing attention in the past decade, and various attacks have been investigated, such as eavesdropping attack, data falsification attack, jamming attack, and so on [1]–[3]. Among these attacks, a jamming attack is a serious threat, which degrades the system performance severely, and therefore, we focus here on a jamming attack. To cope with a jamming attack, various countermeasures were proposed [4]–[6], such as Frequency Hopping Spread Spectrum (FHSS) and Uncoordinated Frequency Hoping (UFH). Unfortunately, these methods are inefficient and use wideband spectrum. In addition, they cannot be directly applied to the scenarios with a dynamic spectrum availability [7]. More importantly, the jammers with higher level intelligence pose new challenges to the existing anti-jamming approaches. Therefore, it is necessary to develop an efficient anti-jamming scheme in wireless networks with a dynamic spectrum availability. In this study, we study the anti-jamming defense problem in heterogeneous wireless networks with an unknown channel availability state information.

Some of the related studies are aimed to defend against the jamming attacks. The anti-jamming problem in power domain was investigated in [8]–[20], and the jamming defense problem in spectrum domain was studied in [7], [21]–[27]. The anti-jamming methods in power domain are traditional techniques, and the anti-jamming mechanisms in spectrum domain are promising schemes. However, in the existing works, the anti-jamming defense problem in power domain and spectrum domain were separately addressed. The anti-jamming transmission problem was investigated in [28], where the power and channel decision making were jointly considered, but the anti-jamming decision-making was made in all power channel combinations. In this study, to cope with

the jamming attacks flexibly, we investigate the anti-jamming problem from a new perspective. At the same time, we aim to develop a multi-domain anti-jamming scheme in heterogeneous wireless networks with an unknown channel availability state information. The multi-domain includes power domain and spectrum domain. Namely, to be more specific, we propose a Stackelberg power game to fight against a jamming attack in the power domain. In the case of a severe jamming attack, the channel switching is employed to combat the jamming attack effectively, and a multi-armed bandit (MAB) [29]–[32] channel selection problem of a certain channel switching cost is formulated.

Notably, the scenario considered in this study has the following challenges. Firstly, an unknown channel availability state information brings a great challenge, because it is necessary to explore channel availability and exploit an available channel. Therefore, we resort to the MAB model, which can strike a tradeoff between ''exploration" and ''exploitation" for the channel selection in unknown channel availability state information scenarios [31]. Moreover, a channel switching cost is another challenge, because it leads to the packet loss, delay, and protocol overhead reducing the performances. Therefore, a frequent channel switching should be avoided in the efficient anti-jamming scheme. As already mentioned, we investigate here the anti-jamming problem from a multi-domain perspective, and a multi-domain anti-jamming framework is formulated.

To the best of authors' knowledge, this is the first work on a multi-domain anti-jamming problem in heterogeneous wireless networks with an unknown channel availability state information. The main contributions of this paper are given as follows:

- An anti-jamming defense problem in heterogeneous wireless networks with an unknown channel availability state information is investigated in a multi-domain, and a multi-domain anti-jamming framework is proposed, wherein a Stackelberg power game is formulated in power domain, and a MAB-based channel selection problem with channel switching cost is formulated in spectrum domain.
- Both the Stackelberg power game and the MAB-based channel selection scheme are analyzed. For the formulated Stackelberg power game, the optimal transmission power and utility of a legitimate user are derived. For the MAB-based channel selection scheme, it is proved that it has the logarithmic order regret of the number of time slots.
- A multi-domain anti-jamming scheme (MDAS) is proposed to fight against jamming attacks effectively, and extensive simulations are conducted to validate the performance of the proposed MDAS.

The rest of this paper is organized as follows. The related work is presented in Section II. In Section III, the system model is given. In Section IV, the problem formulation and analysis are provided, and a multi-domain anti-jamming scheme is proposed. In Section V, extensive simulations are conducted. Lastly, the conclusions are presented in Section VI.

II. RELATED WORK

The anti-jamming defense problem is a hot research topic, and there are many studies aiming to cope with the jamming attacks. In [8], the power control problem was investigated, and a jamming game with a transmission cost was formulated. In [9], the joint power control and scheduling problem under jamming attacks were studied. In [10], a game theoretic framework was employed to model the interactions between a secondary user and a jammer. In [11], a generalized iterative power control water-filling algorithm was proposed, and the Nash equilibrium was analyzed. In [12], the strategic decision making for the anti-jamming problem was investigated in frequency-selective fading channels and AWGN channels. In [13], a jamming game was formulated to analyze and model the attacker-defender interactions, and a prospect theory was adopted to model the end-user behavior. Considering the sequential interactions between user and jammer, in [14]–[20], a Stackelberg game was adopted to analyze the anti-jamming defense problem. However, the above studies mainly focused on the anti-jamming problem in power domain, while little attention was paid to the spectrum domain.

Since the anti-jamming defense methods in spectrum domain are promising to fight against the jamming attacks, and some studies have been devoted to it. In [7], the anti-jamming channel access problem was analyzed, and two algorithms were introduced. The anti-jamming defense problem in the cognitive radio networks was studied in [21], and a flexible channel access method was given. In [22], Wang *et al.* studied the anti-jamming channel selection, and a stochastic game framework was formulated. A stochastic game that models the interactions between a secondary user and a malicious user was formulated, and a channel allocation problem was converted to a two-level auction in [23] and [24]. The jamming attack aimed to reduce the spectrum opportunities was investigated in [25], and a learning algorithm for an adversarial environment was proposed. In [26], we formulated a hierarchical learning framework to investigate the anti-jamming channel selection, and we proposed a hierarchical learning method to achieve a channel selection strategy. In [27], based on the deep reinforcement learning approach, we propose an anti-jamming defense scheme. However, all above-mentioned studies paid attention to the anti-jamming defense only in the spectrum domain.

Our work in this paper is different from above work, and we investigate the anti-jamming problem from a multi-domain perspective. The most related works are the ones presented in [33] and [34]. Compared with these two studies, here, the MAB-based channel selection is investigated, and the channel switching cost is considered. Therefore, the main improvements are: i) we consider a jamming attack, which is a potential threat to the wireless networks, while the works in [33] and [34] ignore it, and ii) we investigate the

anti-jamming from a multi-domain perspective, and a multi-domain anti-jamming scheme is proposed.

III. SYSTEM MODEL

We consider a wireless communication system consisting of two wireless devices, namely, a legitimate user (transmitter-receiver pair) and a malicious jammer. Denote a channel set as $\mathcal{M}$, and $|\mathcal{M}| = M$. The channel availability statistics vector can be denoted as $\theta = (\theta_1, \cdots, \theta_m, \cdots, \theta_M)$, $1 \leq m \leq M$. The channel availability vector of the heterogeneous spectrum opportunities can be expressed as $\tau = \{\tau_1, \tau_2, \cdots, \tau_m, \cdots, \tau_M\}$, and $\tau_m \in \{0, 1\}$ where value 0 denotes that channel is unavailable, and value 1 denotes that channel is available. We assume that legitimate user and malicious jammer are able to sense all the channels. Due to the hardware limitations, it is assumed that legitimate user selects only one channel for transmission at a time slot, and only one channel is jammed at a time slot by the malicious jammer [26], [35], [36]. For convenience, Table 1 lists the notations used in this study.

In order to realize a reliable transmission, it is necessary that legitimate user selects the optimal transmission power and channel to cope with a jamming attack effectively. Since the channel switching between different channels results in the packet loss, delay, and protocol overhead, and the channel switching cost is incurred in [33] and [34]. Therefore, to decrease this type of cost, it is necessary to avoid frequent channel switching. In this work, we propose a multi-domain anti-jamming scheme to fight against jamming attacks. The legitimate user transmits its traffic in the selected available channel with a certain power level. Similar to [26] and [35], it is assumed that channels undergo block fading, and the channel gain of the transmission link between legitimate transmitter and receiver can be denoted by [26], [35], [37]:

$$w_s = (d_s)^{-\alpha_1} \varepsilon_s, \tag{1}$$

where d_s denotes the distance between the legitimate transmitter and receiver, α_1 represents the path loss exponent, and ε_s is the instantaneous fading coefficient, e.g. the Lognormal fading. Similarly, the channel gain of the jamming link between the jammer and receiver can be expressed by:

$$w_j = \left(d_j\right)^{-\alpha_2} \varepsilon_j, \tag{2}$$

where α_2, d_j, and ε_j are the path loss exponent, distance and instantaneous fading coefficient between jammer and receiver, respectively. It is assumed that the legitimate user selects channel m at time slot t. Referring to [17]–[20], based on the Signal-to-Interference-plus-Noise Ratio (SINR), the utility of a legitimate user at time slot t can be defined by the SINR minus transmission cost, and can be given by:

$$u_{st}(P, J) = \frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} P}{BN_0 + \varepsilon_j \left(d_j\right)^{-\alpha_2} J} - C_s P, \tag{3}$$

where P and J respectively denote the transmission power of a legitimate user and a jammer, N_0 represents the noise power spectrum density, B is the channel bandwidth, τ_m denotes the availability of channel m, and C_s is the transmission cost per unit power of a legitimate user. Similarly, the utility of the jammer can be defined by the negative of the SINR minus transmission cost, and can be expressed by:

$$u_{jt}(P, J) = -\frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} P}{BN_0 + \varepsilon_j \left(d_j\right)^{-\alpha_2} J} - C_j J, \tag{4}$$

where C_j denotes the transmission cost per unit power of a jammer. The legitimate user aims to maximize its utility and realize reliable transmission. The jammer can learn the transmission power strategy of a legitimate user, and determines its optimal power strategy to prevent the communication of a legitimate user. Similar to [14]–[20], this problem can be modeled as a Stackelberg power game.

For a legitimate user that selects an available channel $\rho(t)$ to deliver a message at time slot t, the reward can be

TABLE 1. Summation of used notations.

Notations	Explanation
$\mathcal{M}$	set of channels
θ	channel availability statistics vector
τ	channel availability vector
w_s	channel gain of the transmission link
w_j	channel gain of the jamming link
α_1	path-loss exponent of the transmission link
α_2	path-loss exponent of the jamming link
d_s	distance between the transmitter and receiver
d_j	distance between the jammer and receiver
ε_s	fading coefficient of the transmission link
ε_j	fading coefficient of the jamming link
P	transmission power of the legitimate user
J	transmission power of the jammer
B	channel bandwidth
N_0	noise power spectrum density
C_s	power cost unit of the legitimate user
C_j	power cost unit of the jammer
$u_{st}(P, J)$	the utility of the legitimate user
$u_{jt}(P, J)$	the utility of the jammer
$\rho(t)$	selected channel of the legitimate user
r	reward of the legitimate user
δ	received SINR
T	a specified threshold
$A(t)$	action history
$H(t)$	reward history
$\chi_m(t)$	the number of channel m is selected
u_m	the expectation reward
$C(t)$	the accumulative reward
$S(t)$	the number of channel switching
c	the channel switching cost unit
$CW(t)$	the expected channel switching cost
$\tilde{C}(t)$	the long-term reward
g	the expected reward rate
$E[R(t)]$	the expected regret
$r_m(t)$	the total reward of the channel m
r_{nor}	the normalized reward

given by:

$$r\left(\rho\left(t\right),\tau_{\rho(t)}\right)=\begin{cases}B\log(1+\delta), & \delta\geq T,\\ 0, & otherwise,\end{cases} \tag{5}$$

where δ represents the received SINR, and T is a specified threshold, which is related to the service demands (i.e., data, image, and voice). If the received SINR δ reaches the specified transmission threshold T, then transmission can be decoded correctly; otherwise, the channel suffers severe jamming, and the reward is 0.

When the selected channel of the legitimate user is unavailable or suffers from severe jamming, the user has to select a new available channel in order to transmit its traffic. In our work, to effectively cope with the jamming attacks and avoid frequent channel switching, we suppose that legitimate user selects a new channel only when the current channel is unavailable or suffers from severe jamming; otherwise, the user stays in the current channel. An illustration of channel selection is shown in Fig. 1.

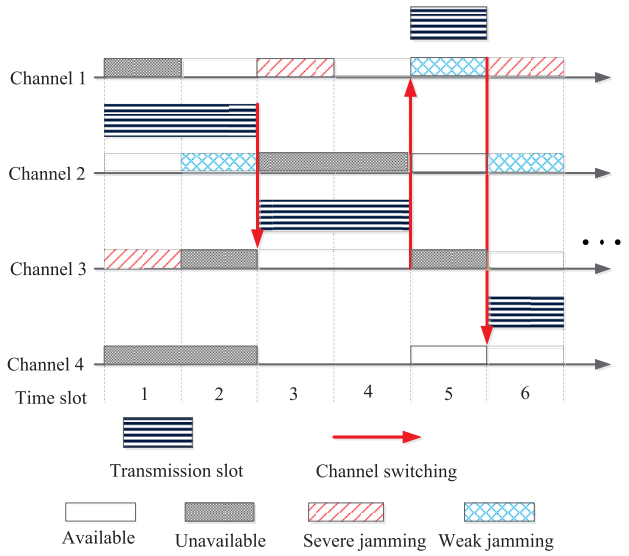


FIGURE 1. Illustration of channel switching.

IV. PROBLEM FORMULATION AND ANALYSIS

In this paper, we formulate a multi-domain anti-jamming framework to flexibly cope with the jamming attacks from a multi-domain perspective, and the structure is shown in Fig. 2. In the formulated multi-domain anti-jamming framework, the environment cognition is essential, and the context information includes channel availability state, location, access context, energy, and so on. The channel availability state shows whether the channel is available or not, the location shows where the legitimate user and jammer are, the access context is related to the service demands (i.e., data, image, and voice), and the energy is related to the transmission power. According to the context information, the multi-domain anti-jamming scheme can flexibly adopt various methods to fight against a jamming attack, such as avoidance (i.e., channel switching) and resistance (i.e., power control). In this paper, we focus on anti-jamming defense problem in both power domain and spectrum domain.

When the jamming attack is weak ($\delta \geq T$), we cope with a jamming attack in the power domain. On the other hand, when a jamming attack is severe ($\delta < T$), the channel switching is adopted to fight against a jamming attack, and an anti-jamming Stackelberg power game is formulated in a newly selected channel. The illustration of the proposed multi-domain anti-jamming scheme is shown in Fig. 3.

Remark 1: The most of the existing work [8]–[20] investigated the anti-jamming problem only in the power domain, and few of them were devoted to the channel selection. The works presented in [7] and [21]–[27] studied the anti-jamming mechanisms in the spectrum domain. However, these two important aspects were separately considered in the existing work. Therefore, in this study, we aim to formulate an anti-jamming scheme from a multi-domain perspective (power domain and spectrum domain) to cope with the jamming attacks effectively. To be specific, we first fight against a jamming attack in the power domain, and only if a jamming attack is severe, the channel switching is employed to combat the jamming attack.

A. ANTI-JAMMING STACKELBERG POWER GAME

As it is presented in Fig. 3, an anti-jamming Stackelberg power game is formulated in a selected available channel $\rho\left(t\right)\in\mathcal{M}$ at time slot t. Similar to [17]–[20], a legitimate user acts as a leader, and a jammer is assumed to be a follower. Therefore, a backward induction method is employed, and a jammer's optimization problem can be formulated as:

$$P1: \quad \max_{J\geq 0} u_{jt}\left(P,J\right), \quad \forall t, \tag{6}$$

$$s.t.\ J\leq J_{\max}. \tag{7}$$

Accordingly, the optimization problem of a legitimate user can be expressed as:

$$P2: \quad \max_{P\geq 0} u_{st}\left(P,J\right), \quad \forall t, \tag{8}$$

$$s.t.\ P\leq P_{\max}. \tag{9}$$

According to the problems $P1$ and $P2$, we can find the Stackelberg Equilibrium (SE) solution, which means neither the legitimate user nor the jammer has an incentive to deviate to improve the utility unilaterally.

For a selected available channel $\rho\left(t\right)$ at time slot t, a Stackelberg power game is formulated in order to cope with a jamming attack in the power domain. By employing a backward induction method [17]–[19], we first investigate the jammer sub-game. According to the problem $P1$, the jammer's optimal power can be expressed as:

$$J=\max_{0\leq J\leq J_{\max}} u_{jt}\left(P,J\right). \tag{10}$$

Motivated by [17]–[19], we derive the following Lemma 1.

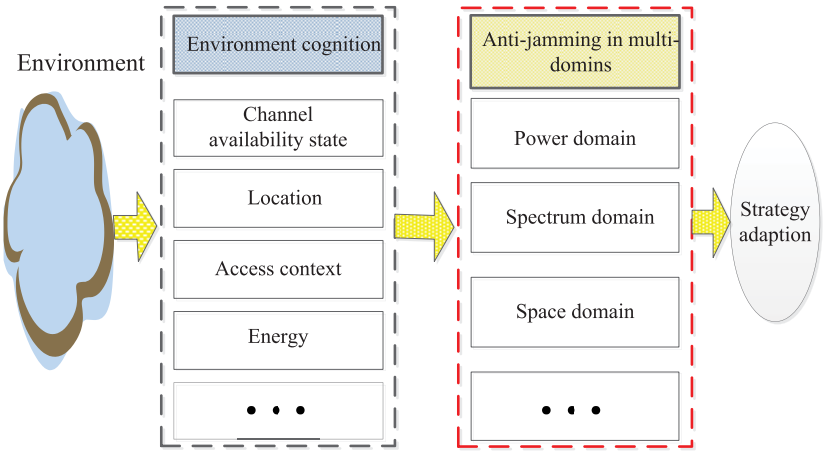


FIGURE 2. The structure of the multi-domain anti-jamming framework.

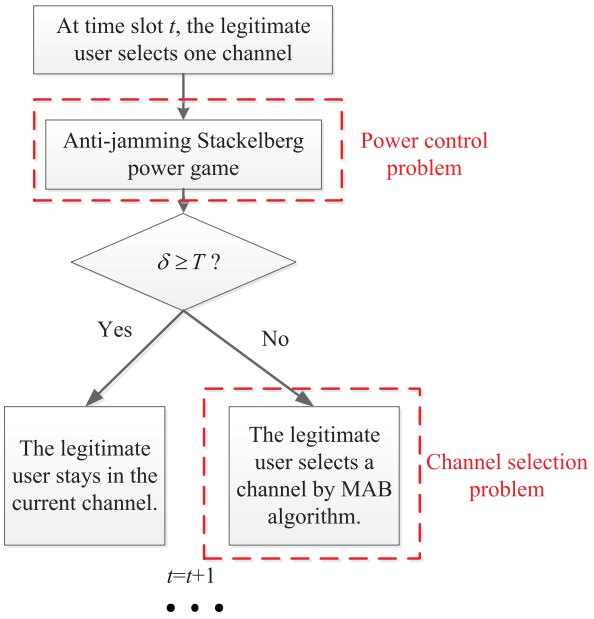


FIGURE 3. Illustration of the multi-domain anti-jamming scheme.

Lemma 1: The optimal transmission power of a legitimate user is defined by:

$$P^* = \begin{cases} \dfrac{\varepsilon_s (d_s)^{-\alpha_1} \left(C_j + \lambda\right)}{4(C_s+\mu)^2 \varepsilon_j \left(d_j\right)^{-\alpha_2}}, \\ \tau_m = 1, C_s \le \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{(2BN_0)} / - \mu, \\ \dfrac{\left(C_j + \lambda\right)\left(\delta_0^2\right)^2}{\varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} \left(d_j\right)^{-\alpha_2}}, \quad \tau_m = 1, \\ \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{(2BN_0)} / - \mu \le C_s \le \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0}, \\ 0, \quad \tau_m = 0, \ or \ \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0} < C_s. \end{cases} \tag{11}$$

Further, an optimal utility of a legitimate user is given by:

$$u_{st}^* (P, J) = \begin{cases} \dfrac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} \left(C_j + \lambda\right)}{2\varepsilon_j \left(d_j\right)^{-\alpha_2} (C_s + \mu)} \\ -\dfrac{\tau_m C_s \varepsilon_s (d_s)^{-\alpha_1} \left(C_j + \lambda\right)}{4\varepsilon_j \left(d_j\right)^{-\alpha_2} (C_s + \mu)^2}, \\ \tau_m = 1, \quad C_s \le \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{(2BN_0)} - \mu, \\ \dfrac{\left(\tau_m \varepsilon_s (d_s)^{-\alpha_1} - C_s \delta_0^2\right)\left(C_j + \lambda\right)\delta_0^2}{\tau_m \varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} \left(d_j\right)^{-\alpha_2}}, \\ \tau_m = 1, \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{(2BN_0)} - \mu \le C_s \le \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0}, \\ 0, \quad \tau_m = 0, \ or \ \dfrac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0} < C_s. \end{cases} \tag{12}$$

Proof: This proof follows the methods given in [17]–[19], where a backward induction method is adopted. The jammer's optimization problem is investigated at first. Then, the optimization problem of the legitimate user is solved.

Firstly, we investigate the optimization problem $P1$ of the jammer. Let P be a given transmission power strategy of a legitimate user. If it holds that $\tau_m = 1$, the jammer's utility function is concave of J, Since $\partial^2 u_{jt} / \partial J^2 = -\left(2\tau_m \varepsilon_s (d_s)^{-\alpha_1} \left(\varepsilon_j \left(d_j\right)^{-\alpha_2}\right)^2 P\right) / \left(BN_0 + \varepsilon_j \left(d_j\right)^{-\alpha_2} J\right)^3 < 0$. According to duality optimization theory [19], [38], the Lagrange function is given by:

$$L(P, J, \lambda) = -\frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} P}{BN_0 + \varepsilon_j \left(d_j\right)^{-\alpha_2} J} - C_j J - \lambda \left(J - J_{\max}\right). \tag{13}$$

where λ is a nonnegative dual variable. The Lagrange dual function can be expressed as:

$$d(\lambda) = \max_{J \ge 0} L(P, J, \lambda). \tag{14}$$

Accordingly, a dual optimization problem is $d^* = \min_{\lambda \geq 0} d(\lambda)$.

Based on the Karush-Kuhn-Tucker (KKT) conditions, we get the following:

$$\frac{\partial L(P, J, \lambda)}{\partial J} = \frac{\tau_m \varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} (d_j)^{-\alpha_2} P}{\left(BN_0 + \varepsilon_j (d_j)^{-\alpha_2} J\right)^2} - C_j - \lambda = 0. \tag{15}$$

According to (15), we derive:

$$J(P) = \frac{1}{\varepsilon_j (d_j)^{-\alpha_2}} \left(\sqrt{\frac{\tau_m \varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} (d_j)^{-\alpha_2} P}{C_j + \lambda}} - BN_0 \right)^+. \tag{16}$$

where $(\cdot)^+ \triangleq \max(\cdot, 0)$. If it holds that $\tau_m = 0$, the utility function of the jammer $u_{jt}(P, J) = -C_j J$ decreases with J, and in that case, its optimal transmission power is $J = 0$. Based on the analysis in [38], since the problem $P1$ is a convex optimization problem. According to the solution of the dual problem, we can obtain the optimal solution of the primal problem.

In the following, we investigate the optimization problem $P2$ of the legitimate user. According to the problem $P2$, the optimal power of the legitimate user is defined by:

$$P = \max_{0 \leq P \leq P_{\max}} u_{st}(P, J(P)). \tag{17}$$

By substituting (16) into the utility function of a legitimate user, we have

$$u_{st}(P, J(P)) = \begin{cases} \left(\frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1}}{BN_0} - C_s \right) P, & P \leq \Psi, \\ \sqrt{\frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} (C_j + \lambda) P}{\varepsilon_j (d_j)^{-\alpha_2}}} - C_s P, & P > \Psi \end{cases} \tag{18}$$

where $\Psi = (C_j + \lambda)(BN_0)^2 / \left(\tau_m \varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} (d_j)^{-\alpha_2} \right)$. If it holds that $P \leq \Psi$, the utility function of the legitimate user is linear with P. If it holds that $P > \Psi$, it is concave with respect to P, since $\partial^2 u_{st}(P, J(P)) / \partial^2 P = -(1/4P) \left(\sqrt{(\tau_m \varepsilon_s (d_s)^{-\alpha_1} (C_j + \lambda)) / \left(\varepsilon_j (d_j)^{-\alpha_2} P \right)} \right) < 0$.

Introducing a non-negative dual variable μ, we can obtain the following:

$$L(P, J(P), \mu) = \frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} P}{BN_0 + \varepsilon_j (d_j)^{-\alpha_2} J} - C_s P - \mu (P - P_{\max}). \tag{19}$$

Similar to the above analysis, we can obtain the optimal transmission power of a legitimate user, and it can be given by:

$$P_{opt} = \frac{\tau_m \varepsilon_s (d_s)^{-\alpha_1} (C_j + \lambda)}{4 (C_s + \mu)^2 \varepsilon_j (d_j)^{-\alpha_2}}. \tag{20}$$

Similar to the analyses in [17]–[19], if it holds that $C_s \leq \varepsilon_s (d_s)^{-\alpha_1} / (2BN_0) - \mu$, then, an optimal transmission power strategy of a legitimate user is $P^* = P_{opt}$. In that case, its optimal utility is expressed by $u_{st}^* = \left(\tau_m \varepsilon_s (d_s)^{-\alpha_1} (C_j + \lambda) \right) / \left(2\varepsilon_j (d_j)^{-\alpha_2} (C_s + \mu) \right) - \left(\tau_m C_s \varepsilon_s (d_s)^{-\alpha_1} (C_j + \lambda) \right) / \left(4\varepsilon_j (d_j)^{-\alpha_2} (C_s + \mu)^2 \right)$.

If it holds that $\varepsilon_s (d_s)^{-\alpha_1} / (2BN_0) - \mu \leq C_s \leq \varepsilon_s (d_s)^{-\alpha_1} / BN_0$, and an optimal transmission power strategy of a legitimate user is $P^* = \Psi$. Then, an optimal utility of a legitimate user is given by $u_{st}^* = \left(\Lambda (C_j + \lambda) BN_0 \right) / \left(\tau_m \varepsilon_s \varepsilon_j (d_s)^{-\alpha_1} (d_j)^{-\alpha_2} \right)$, where $\Lambda = \left(\tau_m \varepsilon_s (d_s)^{-\alpha_1} - C_s BN_0 \right)$.

If it holds that $\varepsilon_s (d_s)^{-\alpha_1} / BN_0 \leq C_s$, then, $P^* = 0$, and an optimal utility of a legitimate user is $u_{st}^* = 0$.

When $\tau_m = 0$, a utility function of a legitimate user is $u_{st}(P, J) = -C_s P$, and it decreases with P. An optimal transmission power strategy of a legitimate user is $P^* = 0$, and its optimal utility is $u_{st}^* = 0$. ■

In that context, if there does not exist a jammer in a selected available channel $\rho(t)$ at time slot t, a utility of a legitimate user is a linear function with respect to P and can be defined by $u_{st} = \left(\left(\varepsilon_s (d_s)^{-\alpha_1} \right) / BN_0 - C_s \right) P$. As it can be noticed, a utility function linearly decreases with P for $\varepsilon_s (d_s)^{-\alpha_1} / BN_0 \leq C_s$, where an optimal power strategy of a legitimate user is $P^* = 0$, and its optimal utility is $u_{st}^* = 0$. On the other hand, a utility function linearly increases with respect to P for $\varepsilon_s (d_s)^{-\alpha_1} / BN_0 > C_s$, where an optimal power strategy of a legitimate user is $P^* = P_{\max}$, and its optimal utility is $u_{st}^* = \left(\left(\varepsilon_s (d_s)^{-\alpha_1} \right) / BN_0 - C_s \right) P_{\max}$. Therefore, when a jammer does not jam a selected available channel $\rho(t)$, an optimal utility of a legitimate user can be expressed as:

$$u_{st}^* = \begin{cases} \left(\frac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0} - C_s \right) P_{\max}, & \frac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0} > C_s, \\ 0, & \frac{\varepsilon_s (d_s)^{-\alpha_1}}{BN_0} \leq C_s. \end{cases} \tag{21}$$

In this study, the received SINR can be defined as:

$$\delta = u_{st}^*. \tag{22}$$

Motivated by [19] and [39]–[41], the SE of the formulated Stackelberg power game can be defined as follows.

Definition 1: If neither the legitimate user nor the jammer has an incentive to deviate to improve its utility unilaterally, the strategy pair (P^*, J^*) is an SE. Thus, it can be expressed as:

$$u_{st}(P^*, J^*) \geq u_{st}(P, J^*), \tag{23}$$

$$u_{jt}(P^*, J^*) \geq u_{jt}(P^*, J). \tag{24}$$

Lemma 2: For a selected available channel m at time slot t, there exists a unique SE for the formulated anti-jamming Stackelberg power game.

Proof: We omit the proof for brevity, and readers can refer to [19]. ■

B. MAB-BASED CHANNEL SELECTION

By solving the above anti-jamming Stackelberg power game, we can obtain the received SINR δ, and if $\delta < T$ holds, it is necessary to switch the channel to cope with a jamming attack effectively. Otherwise, the legitimate user stays in the current channel. To be more specific, the legitimate user selects a channel $\rho(t)$ at the t-th time slot, and it will receive a reward $r\left(\rho(t), \tau_{\rho(t)}\right)$ at the end of the t-th time slot. If $\rho(t) \neq \rho(t-1)$, a channel switching occurs; otherwise, the legitimate user keeps on transmitting its traffic through the channel $\rho(t-1)$. The channel $\rho(t)$ is selected based on the action history $A(t) = \{\rho(1), \rho(2), \cdots, \rho(t-1)\}$ and reward history $H(t) = \{r(\rho(1)), r(\rho(2)), \cdots, r(\rho(t-1))\}$. Therefore, according to $A(t)$ and $H(t)$, a channel selection is made at each time slot.

Motivated by results presented in [33] and [34], we formulate the channel selection problem as a MAB problem where each channel can be regarded as an arm and each channel access is treated as a play of the formulated MAB problem. At time slot t, a player selects one of the arms to play, and it achieves a reward $r\left(\rho(t), \tau_{\rho(t)}\right)$.

Denote $\chi_m(t)$ as the number of time slots that channel m is selected in the first t time slots, and it can be expressed by:

$$\chi_m(t) = \sum_{i=1}^{t} f\{\rho(i) = m\}, \tag{25}$$

where $f\{\cdot\}$ represents the indicator function, which is defined by:

$$f\{\rho(i) = m\} = \begin{cases} 1, & \rho(i) = m, \\ 0, & \rho(i) \neq m. \end{cases} \tag{26}$$

The expectation reward is defined as $u_m = E\left[r\left(\rho(t), \tau_{\rho(t)}\right)\right]$, where $E[\cdot]$ denotes the operation of taking expectation. If $u_m, m \in \mathcal{M}$, is known, an optimal channel for a legitimate user can be defined by:

$$m^* = \arg\max_{m \in \mathcal{M}} u_m. \tag{27}$$

However, in practice, the information on channel availability and jammer is unknown. In this study, an accumulative reward during t time slots can be given by:

$$C(t) = \sum_{i=1}^{t} r\left(\rho(t), \tau_{\rho(t)}\right). \tag{28}$$

Then, the expected accumulative reward can be expressed by:

$$E[C(t)] = E\left[\sum_{i=1}^{t} r\left(\rho(t), \tau_{\rho(t)}\right)\right] = \sum_{m=1}^{M} u_m E[\chi_n(t)]. \tag{29}$$

As previously mentioned, channel switching incurs a switching cost. Similar to [33] and [34], a constant channel switching cost c is considered, and the number of channel switching during t time slots is defined by:

$$S(t) = \sum_{k=2}^{t} f(\rho(k) \neq \rho(k-1)), \tag{30}$$

where $f\{\cdot\}$ represents the indicator function. The expected channel switching cost during t time slots is then:

$$CW(t) = cE[S(t)]. \tag{31}$$

The channel switching cost can be treated as a performance loss, and the long-term reward can be defined by the accumulative reward minus channel switching cost. It can be expressed as:

$$\begin{aligned} \tilde{C}(t) &= E[C(t)] - CW(t) \\ &= \sum_{m=1}^{M} u_m E[\chi_n(t)] - cE[S(t)]. \end{aligned} \tag{32}$$

The expected reward rate can be expressed as:

$$g = \lim_{t \to \infty} \frac{\tilde{C}(t)}{t} = \lim_{t \to \infty} \frac{E[C(t)] - CW(t)}{t}. \tag{33}$$

Here, we aim at designing a channel selection policy that can maximize the long-term reward, which is defined as:

$$P3 : max \tilde{C}(t). \tag{34}$$

Remark 2: In [33]and [34], a MAB-based channel selection problem was studied. However, little attention was paid to the jamming attack, which is a serious threat to the wireless networks that deteriorates the performance of a legitimate user. In this paper, we investigate the MAB-based channel selection problem under a hostile jamming attack, and a multi-domain anti-jamming scheme is proposed.

In this subsection, a channel selection problem is formulated as a MAB problem. Besides, an online learning algorithm is proposed. In a MAB problem, the regret is an important metric, which determines the performance of a MAB algorithm and denotes the reward loss because the policy does not always play the best arm. As the channel switching incurs a certain cost. Motivated by [33], [34], and [42]–[44], the channel switching cost is considered in the regret function, and the regret after t time slots is given by:

$$\begin{aligned} E[R(t)] &= E\left[C^*(t)\right] - [E[C(t)] - CW(t)] \\ &= \sum_{m^* \neq m} (u_{m^*} - u_m) E[\chi_m(t)] + CW(t) \end{aligned} \tag{35}$$

where $u_{k^*} = \max u_k, k = 1, 2, \cdots, M$, $E\left[C^*(t)\right]$ is the upper bound of $E[C(t)]$, i.e., $E\left[C^*(t)\right] \geq E[C(t)]$. In a learning-based algorithm, the regret is minimized. In a MAB problem, minimizing the regret is equivalent to maximizing the long-term reward. However, it is challenging to find an optimal strategy, which can minimize the regret, even that is impossible. Therefore, in this study, we develop an online learning-based algorithm with a logarithmic-order regret, i.e., $E[R(t)] \sim O(\log n)$.

In that context, a channel selection problem without information on channel availability and jammer can be formulated as a MAB problem. Motivated by [33] and [34], based on UCB1 algorithm [45], an anti-jamming channel selection scheme is designed to achieve an effective channel selection

strategy, where a legitimate user selects an available channel and keeps on transmitting in the selected channel for successive t time slots until the selected channel is unavailable or starts suffering from severe jamming. Note that the number of time slots depends on channel availability and jammer.

In the proposed scheme, the reward is obtained by (5), and a normalized reward r_{nor} is employed. The legitimate user is related to four parameters, $\chi_m(t), S(t), \rho(t)$, and $r_m(t)$, where $\chi_m(t)$ represents the number of time slots when channel m is selected during t time slots, $S(t)$ is the number of channel switchings during t time slots, $\rho(t)$ is the selected available channel at time slot t, and $r_m(t)$ denotes the total reward of channel m during t time slots.

As mentioned, channel switching occurs when a selected channel is unavailable or suffers from severe jamming. An unsuccessful transmission event can be denoted by $\psi(t)$ at time slot t. If the transmission is unsuccessful, i.e., the selected channel is unavailable or suffers from severe jamming, then, $\psi(t) = 1$; otherwise, $\psi(t) = 0$. The average reward of channel m at time slot t can be estimated by:

$$\bar{r}_m(t) = \frac{r_m(t)}{\chi_m(t)}. \tag{36}$$

In the proposed scheme, when a channel is selected, a legitimate user will stay in the selected channel until the transmission fails. At first, the legitimate user selects each channel once. After that, the channel with the maximal index $\eta_m(t)$, $m \in \{1, 2, \cdots, M\}$ is selected. The index $\eta_m(t)$ is defined by:

$$\eta_m(t) = \bar{r}_m(t) + \sqrt{\frac{2\ln(S(t))}{\chi_m(t)}}, \tag{37}$$

where the first term means the average reward of the selected channel m, and the second term denotes the size of the one-sided confidence interval for the average reward. At time slot t, $\chi_m(t)$, $S(t)$, and $r_m(t)$ can be estimated as:

$$\chi_m(t) = \begin{cases} \chi_m(t-1) + 1, & if\ \psi(t-1) = 1 \\ & and\ \rho(t) = m, \\ \chi_m(t-1), & else. \end{cases} \tag{38}$$

$$S(t) = \begin{cases} S(t-1) + 1, & if\ \psi(t-1) = 1, \\ S(t-1), & else. \end{cases} \tag{39}$$

$$r_m(t) = \begin{cases} r_m(t-1) + r_{nor}, & if\ \psi(t-1) = 0 \\ & and\ \rho(t) = m, \\ r_m(t-1), & else. \end{cases} \tag{40}$$

Lemma 3: The proposed scheme has a logarithmic regret, i.e., $E[R(t)] \sim O(\log n)$, and whose upper bound is defined by:

$$E[R(t)] \le \sum_{k:\, u_k < u*} (\Delta_m + 2c)\left(\frac{8\ln t}{\Delta_m^2} + 1 + \frac{\pi^2}{3}\right). \tag{41}$$

Proof: Similar to Theorem 1 given in [45], we have

$$E[\chi_m(t)] \le \frac{8\ln(S(t))}{\Delta_m^2} + 1 + \frac{\pi^2}{3}, \tag{42}$$

where $\Delta_m = u^* - u_m$.

Algorithm 1 Multi-Domain Anti-Jamming Scheme (MDAS)

Initiate: $t = 0$, $S(0) = 0$, $\chi_m(0) = 0$, and $r_m(0) = 0$, $m \in \{1, 2, \cdots, M\}$. Select each channel once, and the legitimate user stays in the selected channel until the current channel is unavailable or suffers from severe jamming. Then, the next channel is selected in sequence, and the initialization will be finished after all channels have been tried once.

Loop for each time slot t,

Step 1: If a selected channel $\rho(t)$ is available, an anti-jamming Stackelberg power game is formulated, and an optimal utility of a legitimate user is derived by (12). If there does not exist a jammer in the selected available channel $\rho(t)$, then, an optimal utility of a legitimate user is derived by (21). If jamming is weak, the legitimate user stays in the current channel; otherwise, if the selected channel $\rho(t)$ is unavailable or suffers from severe jamming ($\psi(t) = 1$), go to step 2.

Step 2: A legitimate user selects the channel $\rho(t) = \arg\max_{0<m\le M} \left\{\bar{r}_m(t) + \sqrt{2\ln(S(t))/\chi_m(t)}\right\}$.

Step 3: Update $\chi_m(t)$, $S(t)$, and $r_m(t)$ using (38), (39), and (40), respectively.

End loop

Motivated by [34, Lemma 3.2], we can obtain the upper bound of the expected channel switching cost, and it can be expressed as:

$$\begin{aligned} CW(t) &= c\sum_{k=2}^{t} f(\rho(k) \ne \rho(k-1)) \\ &\le 2c \sum_{m:\, u_m < u*} \chi_m(t) \\ &\le 2c \sum_{m:\, u_m < u*} \left(\frac{8\ln(S(t))}{\Delta_m^2} + 1 + \frac{\pi^2}{3}\right) \\ &\le 2c \sum_{m:\, u_m < u*} \left(\frac{8\ln t}{\Delta_m^2} + 1 + \frac{\pi^2}{3}\right). \end{aligned} \tag{43}$$

Similar to [34] and [42], we can obtain the upper bound of the expected regret, and it can be given by:

$$\begin{aligned} E[R(t)] &\le \sum_{m:\, u_m < u*} (u_{m^*} - u_m) E[\chi_m(t)] + 2c \sum_{m:\, u_m < u*} \chi_m(t) \\ &\le \sum_{m:\, u_m < u*} (\Delta_m + 2c)\left(\frac{8\ln(S(t))}{\Delta_m^2} + 1 + \frac{\pi^2}{3}\right) \\ &\le \sum_{m:\, u_m < u*} (\Delta_m + 2c)\left(\frac{8\ln t}{\Delta_m^2} + 1 + \frac{\pi^2}{3}\right). \end{aligned} \tag{44}$$

■

The steps of the proposed multi-domain anti-jamming scheme (MDAS) are given in Algorithm 1.

V. SIMULATION RESULTS AND DISCUSSIONS

A. SIMULATION SETTINGS

In the simulations, a network with four channels is considered, and the channel availability statistics vector $\theta = (0.8, 0.7, 0.6, 0.5)$. Similar to [17]–[19] and [35], the parameters in this study are as follows: The path loss exponent is $\alpha_1 = \alpha_2 = 2$, the transmission cost per unit power is $c_s = c_j = 0.2$, the noise power spectral density is $N_0 = -135dB/Hz$, the bandwidth of all channels is $B = 1MHz$, the distance between legitimate transmitter and receiver is $d_s = 5km$, the distance between jammer and receiver is $d_j = 30km$, $P_{\max} = 20W$, and $J_{\max} = 50W$. In addition, the channel fading is modeled as a Lognormal fading model with a medium-scale loss [35], where the channel gain is expressed as e^K and K denotes a Gaussian variable with zero mean and variance η^2. In general, the Lognormal fading is expressed as a Decibel-spread form, and $\eta = 0.1\log(10)\eta_{dB}$ [35], [46]. In the simulations, the Decibel-spread is set to 12dB.

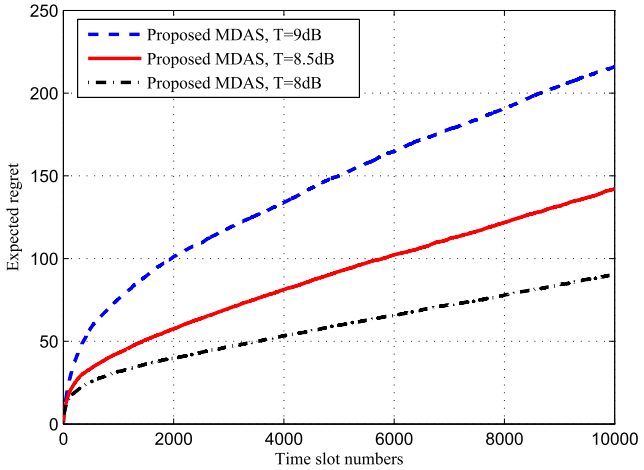


FIGURE 4. The expected regret for different transmission thresholds T ($c = 1$).

B. RESULTS

1) Expected Regret

The influence of the transmission threshold T on expected regret and expected switching cost is shown Fig. 4 and Fig. 5, respectively. The presented results are averaged results for 200 runs. Fig. 4 shows that expected regret of the proposed MDAS increases logarithmic with the number of time slots, which verifies the theoretical analysis as discussed in Lemma 3. In addition, the expected regret increases with the threshold T. As shown in Fig. 5, the expected switching cost also increases logarithmic with the number of time slots. Moreover, it is shown that higher threshold leads to larger switching cost, which is because higher threshold brings more frequent channel switching.

The influence of the switching cost unit c on expected regret and expected switching cost are shown in Fig. 6 and Fig. 7, respectively. In Fig. 6, larger switching cost unit c results in higher expected regret. The reason is that larger switching cost unit c causes more serious performance loss due to channel switching. Fig. 7 shows that switching cost increases with the switching cost unit c.

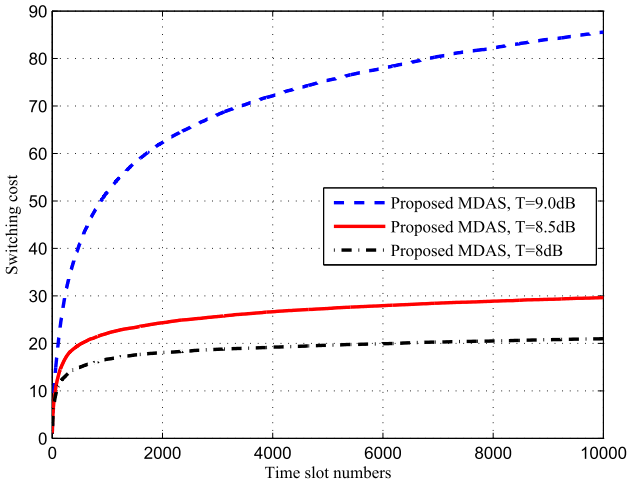


FIGURE 5. The expected switching cost for different transmission thresholds T ($c = 1$).

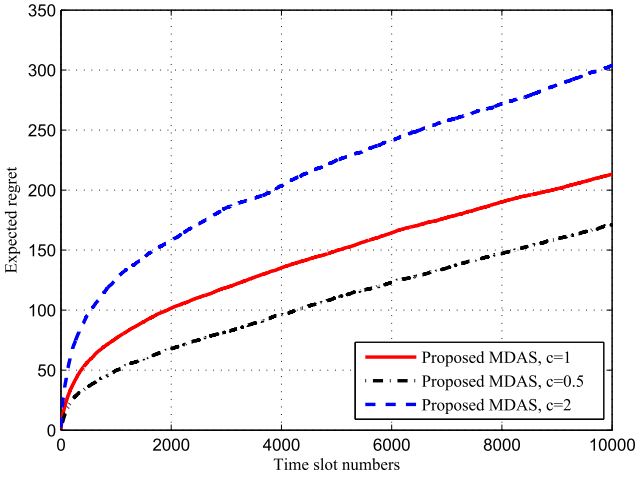


FIGURE 6. The expected regret for different switching cost units c ($T = 9$dB).

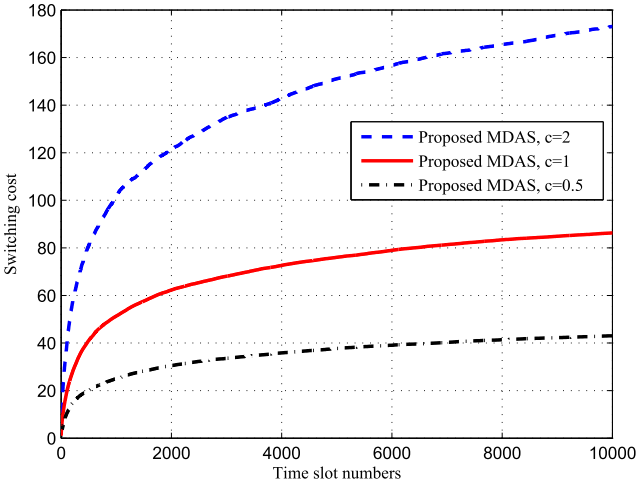


FIGURE 7. The expected switching cost for different switching cost units c ($T = 9$dB).

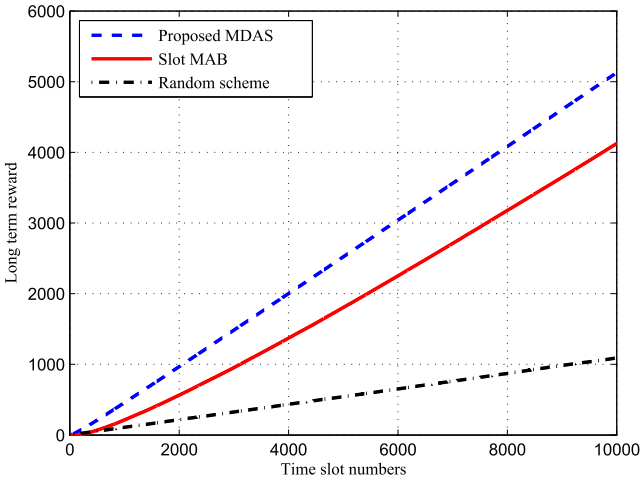


FIGURE 8. The long-term reward for different schemes ($T = 9$dB, $c = 1$).

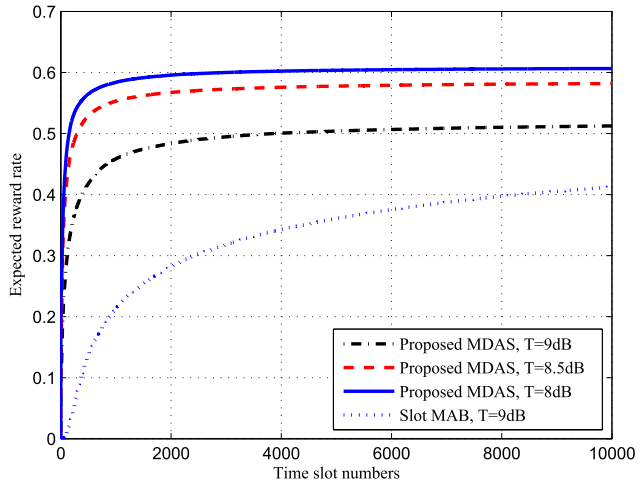


FIGURE 9. The expected reward rate for different transmission thresholds T ($c = 1$).

2) Long Term Reward

To evaluate the performance of the proposed MADS, the proposed MADS is compared with the following schemes:

- Slot MAB algorithm [45]: In the slot MAB algorithm, a channel is selected at each time slot. Thus, this algorithm is a traditional scheme, and it may increase the number of channel switching and result in high switching cost [34].
- Random scheme: In the random scheme, a channel is randomly selected, and it is an instinctive method, especially in an unknown environment.

As indicated in Fig. 8, the long-term reward of all schemes increases with the number of time slots. Moreover, the proposed MADS is superior to the traditional slot MAB algorithm and random scheme and yields the largest long-term reward. The reason is that the traditional slot MAB algorithm selects a channel at each time slot, which brings high switching cost due to frequent channel switching. In addition, the random scheme is an instinctive and inefficient approach.

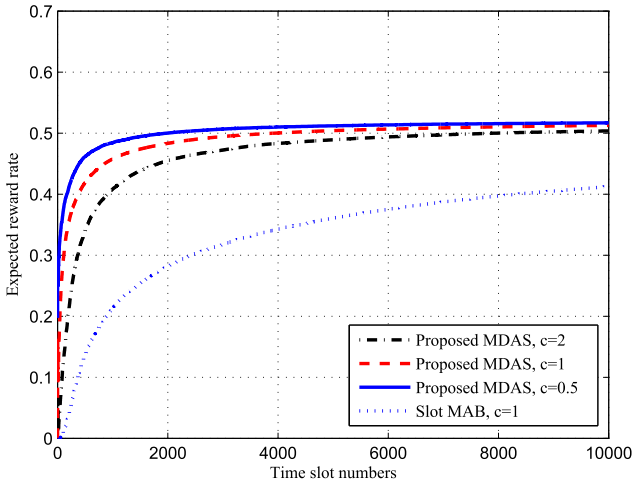


FIGURE 10. The expected reward rate for different switching cost units c ($T = 9$dB).

3) Expected Reward Rate

To better understand the performance of the proposed MDAS, we also show the expected reward rate g, which can be regarded as an instantaneous reward and can measure the performance of the proposed MDAS. Namely, the larger the expected reward rate is, the better MDAS performance will be. In Fig. 9, it can be seen that expected reward rate of the proposed MDAS outperforms the traditional slot MAB. Moreover, higher threshold leads to lower expected reward rate of the proposed MDAS. Fig. 10 shows that expected reward rate decreases with the switching cost unit c.

VI. CONCLUSION

In this paper, an anti-jamming defense problem in heterogeneous wireless networks without information on channel availability was investigated, and a multi-domain anti-jamming scheme (MDAS) was proposed. To be specific, a Stackelberg power game was formulated in the power domain, and a channel selection scheme based on multi-armed bandit (MAB) was formulated in the spectrum domain. Moreover, we analyzed both Stackelberg power game and MAB-based channel selection scheme. Finally, simulations were conducted to evaluate the performance of the proposed MDAS. Simulation results show that proposed MDAS outperformed the traditional slot MAB algorithm and the random scheme, and it had the largest long-term reward. In this study, a single-legitimate user scenario was investigated, and the scenarios with multiple legitimate users will be considered in our future work.

REFERENCES

[1] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.
[2] Y. E. Sagduyu, R. A. Berry, and A. Ephremides, "Jamming games in wireless networks with incomplete information," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 112–118, Aug. 2011.

[3] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, "Byzantine attack and defense in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1342–1363, 3rd Quart., 2015.

[4] H. Zhu, C. Fang, Y. Liu, C. Chen, M. Li, and X. S. Shen, "You can jam but you cannot hide: Defending against jamming attacks for geo-location database driven spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2723–2737, Oct. 2016.

[5] K. Pelechrinis, M. Iliofotou, and S. Krishnamurthy, "Denial of service attacks in wireless networks: The case of jammers," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 2, pp. 245–257, May 2011.

[6] L. Zhang, Z. Guan, and T. Melodia, "United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5733–5747, Aug. 2016.

[7] C. Chen, M. Song, C. Xin, and J. Backens, "A game-theoretical anti-jamming scheme for cognitive radio networks," *IEEE Netw.*, vol. 27, no. 3, pp. 22–27, May/Jun. 2013.

[8] E. Altman, K. Avrachenkov, and A. Garnaev, "A jamming game in wireless networks with transmission cost," in *Proc. NET-COOP*, Avignon, France, Jun. 2007, pp. 1–12.

[9] S. D'Oro, E. Ekici, and S. Palazzo, "Optimal power allocation and scheduling under jamming attacks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1310–1323, Jun. 2017.

[10] R. El-Bardan, S. Brahma, and P. K. Varshney, "Strategic power allocation with incomplete information in the presence of a jammer," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3467–3479, Aug. 2016.

[11] R. H. Gohary, Y. Huang, Z. Q. Luo, and J. S. Pang, "A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2660–2674, Jul. 2009.

[12] T. Song, W. E. Stark, T. Li, and J. K. Tugnait, "Optimal multiband transmission under hostile jamming," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 4013–4027, Sep. 2016.

[13] L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor, "User-centric view of jamming games in cognitive radio networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[14] L. Xiao, Y. Li, J. Liu, and Y. Zhao, "Power control with reinforcement learning in cooperative cognitive radio networks against jamming," *J. Supercomputing*, vol. 71, no. 9, pp. 3237–3257, 2015.

[15] F. Slimeni, V. L. Nir, B. Scheers, Z. Chtourou, and R. Attia, "Optimal power allocation over parallel Gaussian channels in cognitive radio and jammer games," *IET Commun.*, vol. 10, no. 8, pp. 980–986, May 2016.

[16] X. Tang, P. Ren, Y. Wang, Q. Du, and L. Sun, "Securing wireless transmission against reactive jamming: A stackelberg game framework," in *Proc. IEEE GLOBECOM*, Dec. 2015, pp. 1–6.

[17] D. Yang, G. Xue, J. Zhang, A. Richa, and X. Fang, "Coping with a smart jammer in wireless networks: A Stackelberg game approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4038–4047, Aug. 2013.

[18] L. Xiao, T. Chen, J. Liu, and H. Dai, "Anti-jamming transmission Stackelberg game with observation errors," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 949–952, Jun. 2015.

[19] L. Jia *et al.*, "Bayesian Stackelberg game for antijamming transmission with incomplete information," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 1991–1994, Oct. 2016.

[20] L. Jia, F. Yao, Y. Sun, Y. Xu, S. Feng, and A. Anpalagan, "A hierarchical learning solution for anti-jamming Stackelberg game with discrete power strategies," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 818–821, Dec. 2017, doi: 10.1109/LWC.2017.2747543.

[21] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Anti-jamming games in multi-channel cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 4–15, Jan. 2012.

[22] B. Wang, Y. Wu, K. J. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 877–889, Apr. 2011.

[23] M. A. Alavijeh, B. Maham, Z. Han, and S. Nader-Esfahani, "Efficient anti-jamming truthful spectrum auction among secondary users in cognitive radio networks," in *Proc. IEEE ICC*, Jun. 2013, pp. 2812–2816.

[24] M. A. Alavijeh, B. Maham, Z. Han, and W. Saad, "Truthful spectrum auction for efficient anti-jamming in cognitive radio networks," in *Proc. IEEE ISCC*, Jul. 2017, pp. 742–747.

[25] A. Garnaev and W. Trappe, "Bandwidth scanning when facing interference attacks aimed at reducing spectrum opportunities," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1916–1930, Aug. 2017.

[26] F. Yao, L. Jia, Y. Sun, Y. Xu, S. Feng, Y. Zhu, "A hierarchical learning approach to anti-jamming channel selection strategies," *Wireless Netw.*, pp. 1–13, Jul. 2017, doi: 10.1007/s11276-017-1551-9.

[27] X. Liu, Y. Xu, L. Jia, Q. Wu, and A. Anpalagan, "Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 998–1001, May 2018.

[28] T. Chen, J. Liu, L. Xiao, and L. Huang, "Anti-jamming transmissions with learning in heterogenous cognitive radio networks," in *Proc. IEEE WCNC*, Mar. 2015, pp. 293–298.

[29] Q. Wu, Z. Du, P. Yang, Y.-D. Yao, and J. Wang, "Traffic-aware online network selection in heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 381–397, Jan. 2016.

[30] S. Amuru, C. Tekin, M. van der Schaar, and R. M. Buehrer, "Jamming bandits—A novel learning method for optimal jamming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2792–2808, Apr. 2016.

[31] P. Zhou and T. Jiang, "Toward optimal adaptive wireless communications in unknown environments," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3655–3667, May 2016.

[32] M. Qiu, H. Su, G. Quan, and X. Qin, "Online learning anti-jamming cognitive radio network based on Markov model for green clouds," *IEEE Trans. Inf. Forensics Security*, to be published, doi: 10.1109/TIFS.2015.2417835.

[33] L. Chen, S. Iellamo, and M. Coupechoux, "Opportunistic spectrum access with channel switching cost for cognitive radio networks," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.

[34] Z. Qin, J. Wang, J. Chen, Y. Sun, Z. Du, and Y. Xu, "Opportunistic channel access with repetition time diversity and switching cost: A block multi-armed bandit approach," *Wireless Netw.*, vol. 24, no. 5, pp. 1683–1697, 2018, doi: 10.1007/s11276-016-1428-3.

[35] Q. Wu, Y. Xu, J. Wang, L. Shen, J. Zheng, and A. Anpalagan, "Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning," *IEEE Trans. Veh. Tech.*, vol. 62, no. 9, pp. 4524–4538, Nov. 2013.

[36] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 2, pp. 180–194, Apr. 2012.

[37] R. El-Bardan, S. Brahma, and P. K. Varshney, "Power control with jammer location uncertainty: A game theoretic perspective," in *Proc. CISS*, Mar. 2014, pp. 1–6.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] Z. Han *et al.*, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[40] Y. Xu, J. Wang, Q. Wu, J. Zheng, L. Shen, and A. Anpalagan, "Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5305–5318, Dec. 2017.

[41] Y. Sun, J. Wang, F. Sun, and J. Zhang, "Energy-aware joint user scheduling and power control for two-tier femtocell networks: A hierarchical game approach," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2016.2580560.

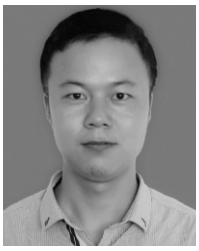
[42] Z. Du, Q. Wu, and P. Yang, "Learning with handoff cost constraint for network selection in heterogeneous wireless networks," *Wireless Commun. Mob. Comput.*, vol. 16, no. 4, pp. 441–458, 2016.

[43] R. Agrawal, M. V. Hegde, and D. Teneketzis, "Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost," *IEEE Trans. Autom. Control* , vol. 33, no. 10, pp. 899–906, Oct. 1988.

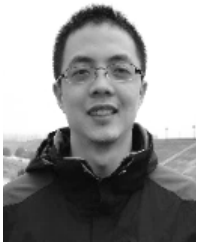
[44] J. Huang, X. Gan, and X. Feng, "Multi-armed bandit based opportunistic channel access: A consideration of switch cost," in *Proc. ICC*, Jun. 2013, pp. 1651–1655.

[45] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2, pp. 235–256, 2002.

[46] G. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA, USA: Kluwer, 2001.



LULIANG JIA received the B.S. degree in communications engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2011, and the M.S. degree in communications and information systems from the College of Communication Engineering, PLA University of Science and Technology, Nanjing, China, in 2014. He is currently pursuing the Ph.D. degree with the Army Engineering University of PLA. His current research interests include game theory, learning theory, and communication anti-jamming technology.



YUHUA XU received the B.S. degree in communications engineering and the Ph.D. degree in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, in 2006 and 2014, respectively. He is currently with the College of Communications Engineering, Army Engineering University of PLA, as an Associate Professor. He has published several papers in international conferences and reputed journals in his research area. His research interests focus on UAV communication networks, opportunistic spectrum access, learning theory, and distributed optimization techniques for wireless communications.

Prof. Xu received the Certificate of Appreciation as an Exemplary Reviewer for the IEEE Communications Letters in 2011 and 2012, respectively. He was selected to receive the IEEE Signal Processing Society's 2015 Young Author Best Paper Award and the Funds for Distinguished Young Scholars of Jiangsu Province in 2016. He served as an Associate Editor for *Transactions on Emerging Telecommunications Technologies* (Wiley) and *KSII Transactions on Internet and Information Systems*, and the Guest Editor of the Special Issue on *The Evolution and the Revolution of 5G Wireless Communication Systems* for *IET Communications*.

YOUMING SUN received the B.S. degree in electronic and information engineering from Yanshan University, Qinhuangdao, China, in 2010, and the M.S. degree from the National Digital Switching System Engineering & Technological Research Center, Zhengzhou, China, in 2013, where he is currently pursuing the Ph.D. degree in communications and information system. His research interests include resource allocation in small cell networks, cognitive radio networks, game theory, and statistical learning. He has acted as a Technical Program Committees Member for the IEEE International Conference on Wireless Communications and Signal Processing 2015. He currently serves as a regular reviewer for many technical journals, including the IEEE Journal on Selected Areas in Communications, the IEEE Systems Journal, the IEEE Access, the *Wireless Networks*, the *IET Communications*, and the *KSII Transactions on Internet and Information Systems*.

SHUO FENG (S'15) received the B.Sc. degree (Hons.) in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the M.Sc. degree in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2014. He is currently pursuing the Ph.D. degree with the Army Engineering University of PLA. His research interests include cognitive radio networks, machine learning, cognitive dynamic system, and information geometry. He served as the Technical Program Committee Member for the IEEE WCSP 2015 and WCCN 2015. He was a co-recipient of the Best Paper Award from the IEEE VTC 2014-Fall. He is an invited reviewer for several journals such as the IEEE Journal on Selected Areas in Communications and the IEEE Communications Letters.

LONG YU received the B.S. degree in mobile communications and the M.S. degree in communications and information systems from the Institute of Communications Engineering, Nanjing, China, in 2003 and 2006, respectively. He is currently pursuing the Ph.D. degree in communications and information systems with the Institute of Communications Engineering, PLA University of Science and Technology. His research interests focus on wireless security, communication anti-jamming techniques, and game theory.

ALAGAN ANPALAGAN (SM'04) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto. He is currently a Professor with the Department of Electrical and Computer Engineering with Ryerson University, where he directs a research group involved in radio resource management and radio access and networking areas within the WINCORE Lab. He has co-authored three edited books, *Design and Deployment of Small Cell Networks* (Cambridge University Press, 2014), *Routing in Opportunistic Networks* (Springer, 2013), and *Handbook on Green Information and Communication Systems* (Academic Press, 2012). He served as an Editor for the IEEE Communications Surveys and Tutorials from 2012 to 2014, the IEEE Communications Letters from 2010 to 2013, *Springer Wireless Personal Communications* from 2011 to 2013, and the *EURASIP Journal of Wireless Communications and Networking* from 2004 to 2009. Prof. Anpalagan is currently a registered Professional Engineer in the Province of Ontario, Canada, and a fellow of the Institution of Engineering and Technology. He was a recipient of the Dean's Teaching Award in 2011; the Faculty Scholastic, Research and Creativity Award in 2010, 2014, and 2017; and the Faculty Service Award in 2011 and 2013 from Ryerson University. He also received the Exemplary Editor Award from the IEEE ComSoc in 2013 and the Editor-in-Chief Top10 Choice Award in Transactions on Emerging Telecommunications Technology in 2012. He currently serves as the TPC Vice-Chair of the IEEE VTC Fall-2017, and served as the TPC Co-Chair of the IEEE GLOBECOM'15: SAC Green Communication and Computing, the IEEE WPMC'12 Wireless Networks, the IEEE PIMRC'11 Cognitive Radio and Spectrum Management, and the IEEE CCECE'04/08. He served as the ComSoc Toronto Chapter Chair from 2004 to 2005, the IEEE Toronto Section Chair from 2006 to 2007, the IEEE Canada Professional Activities Committee Chair from 2009 to 2011, and the IEEE Canada Central Area Chair from 2012 to 2014. He has been serving as the Vice Chair of the IEEE SIG on Green and Sustainable Networking and Computing with Cognition and Cooperation since 2015.