IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 25, 2020, accepted July 12, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009702

A Distributed Collaborative Game-Theoretic Approach in Cognitive Satellite Communication Networks

JILI WANG [1], BANGNING ZHANG [1], (Member, IEEE), LULIANG JIA [2], (Member, IEEE), BING ZHAO [1], (Member, IEEE), AND DAOXING GUO [1]

[1]College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China
[2]School of Space Information, Space Engineering University, Beijing 100000, China

Corresponding author: Daoxing Guo (dxgguo@sina.cn)

ABSTRACT Due to the increasing spectrum scarcity in satellite communications and terrestrial communications, cognitive satellite communications have received widespread attention. In this paper, we take the inherent feature that the terrestrial cognitive users suffer from inter-beam interference. In particular, for the first time, we consider the coupling characteristics of channel access and power optimization in opportunistic spectrum access based on distributed frameworks. We first formulate a joint channel access and power optimization game, which is proven to be an exact potential game and accordingly has at least one pure Nash Equilibrium (NE) point. The sufficient conditions for interference-free between cognitive users and for maximization of system utility are given, respectively. Also, the lower bound of the system utility is deduced theoretically. For discrete power control strategies, we then propose a joint-strategy iteration algorithm (JID) to converge to the general NE in two-dimensional discrete strategy space. Especially, to solve the challenge of finding the optimal NE solution in two-dimensional strategy space, we propose a novel joint-strategy iteration algorithm based on exploration and exploitation (JIDEE). Simultaneously, these two algorithms are extended to the case where the power control strategies are continuous. Finally, simulations are conducted to confirm the effectiveness of the formulated game and the two proposed algorithms.

INDEX TERMS Cognitive satellite communication, spectrum sharing, exact potential game, exploration and exploitation.

I. INTRODUCTION

With the increasing demand for broadcast, multimedia, and interactive services, satellite communications and terrestrial communications face the challenges of insufficient spectrum resources [1], [2]. Spectrum sharing between them has become a promising option. As a spectrum sharing technology, cognitive radio has been extensively investigated in many fields, such as device-to-device communication [3], vehicular communication [4], unmanned aerial vehicle (UAV) communication [5], and wireless sensor network [6]. However, cognitive satellite communications have many inherent characteristics, such as considerable propagation delay, wide beam coverage, cumulative uplink interference, sensitivity to interference, and so on [7]. There are still many challenges to be be solved in cognitive satellite communications.

The associate editor coordinating the review of this manuscript and approving it for publication was Shree Krishna Sharma.

A. BACKGROUND AND MOTIVATIONS

In the spectrum sharing between satellite communications and terrestrial communications, some of the schemes work in the overlay mode [8], [9]. When operating in this mode, terrestrial users are often cognitive users, satellite users are often primary users, and both use the same frequencies. The interference caused by the terrestrial cognitive users to the primary users can be compensated by assisting the primary users in transmitting [10]. The relay nodes (cognitive users) usually receive the signals of the primary satellite network in the first phase and then transmit to the receivers of the primary satellite users in the second phase by amplify-and-forward (AF) relay or decode-and-forward (DF) relay mode. Therefore, it has lower power efficiency and a relatively large delay [11]. On the other hand, most spectrum sharing schemes between satellite and terrestrial communication systems operate in the underlay mode [12]–[19]. However, due to the large antenna gain, satellite receivers are very sensitive to interference [20], [21]. In this mode, the protection distance between terrestrial cognitive users and primary satellite users is large, which reduces the sharing efficiency and causes inconvenience to terrestrial cognitive users.

With the emergence of spectrum database technology [22] and the development of various spectrum sensing technologies in the cognitive satellite communication systems [23]–[25], it is possible to share the spectrum between satellite communications and terrestrial communications in the interweave mode. When working in this mode, cognitive users will temporarily access the idle spectrum unused by the primary user, and no power constraint is required to avoid exceeding the interference threshold [26]. Therefore, it has higher power efficiency and no requirement for protection distance, which is especially suitable for cognitive satellite communications. Consequently, it has practical significance and some advantages to investigate the spectrum sharing between satellite communication systems and terrestrial communication systems in the interweave mode. Despite a plethora of research in cognitive satellite communication in the overlay mode and underlay mode, very few works have focused on the interweave spectrum sharing between terrestrial communications and satellite communications.

B. CONTRIBUTIONS

In this paper, we investigate the interweave spectrum sharing between satellite communications and terrestrial communications, and perform distributed algorithms to optimize the system throughput. The main contributions are summarized as follows:

- The channel access and power optimization problem for cognitive satellite communication networks is formulated as joint channel access and power optimization game, taking the inherent feature that the terrestrial cognitive users suffer from inter-beam interference into account. It is noteworthy that the coupling characteristic of channel selection and power control in opportunistic spectrum access based on distributed framework is considered for the first time. The joint channel access and power optimization game is proven to be an exact potential game.
- Due to the differences in available channel quality, sufficient conditions for interference-free between terrestrial cognitive users and for maximizing system utility are given, respectively. Also, the lower bound of the system utility is deduced theoretically.
- For discrete power control strategies, the JID algorithm is proposed to converge to the general NE of the joint channel access and power optimization game. Especially, to solve the challenge of finding the optimal NE solution in the two-dimensional discrete strategy space, the novel JIDEE algorithm is proposed. Both of these two algorithms are extended to the case where the power control strategies are continuous.
- It is shown that the joint channel access and power optimization scheme achieves better system throughput performance than only considering channel access, which verifies the effectiveness of the formulated game.

It is noted that our earlier work [27] studies the spectrum sharing in cognitive satellite communication in the interweave mode, in which only channel selection is optimized (terrestrial cognitive users are non-cooperative and always transmit with the maximum power), thus significantly reducing the system throughput. Besides, the algorithm proposed in [27] converges slowly and does not always converge to the optimal NE. Motivated by all these observations, this work extends pure channel selection to joint channel selection and power control, which is also not investigated in dynamic spectrum access based on distributed frameworks [26], [28]–[35]. Besides, most of the distributed algorithms are mainly designed and applied to converge to the NE in one-dimensional strategy space [26]–[36], we solve the challenge of developing an algorithm with fast convergence speed and can converge to the optimal NE in two-dimensional strategy space.

C. RELATED WORKS

In recent years, many papers studied the distributed channel selection problem using game theory. In [28], [29], the authors examined the problem of distributed channel selection in opportunistic spectrum (OSA) networks. In [30], [31], the authors studied the problem of channel selection for interference mitigation in canonical networks. In [32], [33], the authors studied the problem of anti-jamming dynamic spectrum access in wireless networks. In [34], the author studied the spectrum access problem with channel bonding in Mesh Networks. These works only focus on the problem of channel selection without considering the coupling characteristic of channel access and power control in the spectrum access competition. In order to further improve system performance, it is necessary to consider the issues of channel access and power optimization together in the spectrum access competition.

The resource optimization problem in cognitive radio communications has been studied by game theory in some literature. Most of them only concentrated on power control or channel selection. In [37], [38], the authors focused on the power control strategy in cognitive radio networks. In [39], [40], the authors focused on the channel selection strategy in cognitive radio networks. In [16], [41], the authors investigated the power control and channel allocation problem using game theory. However, most of the works about the joint channel selection and power control in cognitive radio networks work in the underlay mode, and few works have studied the channel selection and power control in the interweave mode.

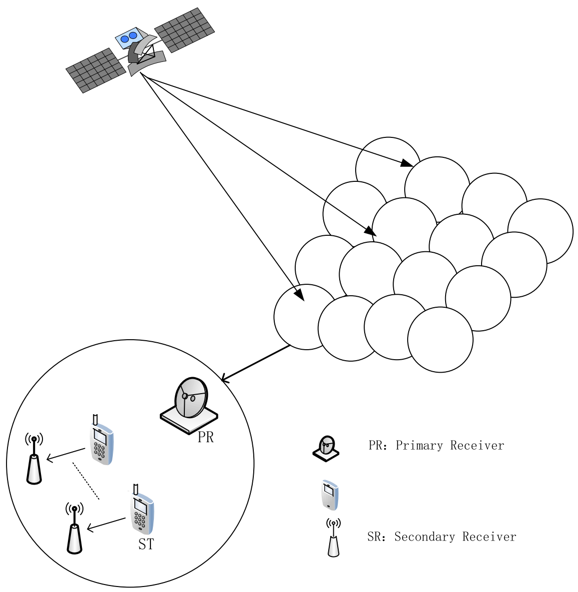


FIGURE 1. The system model of the Multibeam-based cognitive satellite communication system.

In the competition of spectrum access based on game theory, most studies assume that cognitive users only interfere with each other and are not interfered by the external environment [28], [30], [35]. Hence, the quality of all available channels is the same. However, cognitive users may be subject to different degrees of external interference, resulting in differences in available channel quality. In the research scenario of this paper, cognitive users are also interfered by the multi-beam satellite communication system, resulting in different channel quality.

The rest of this article is organized as follows. In Section II, we present our system model and problem formulation for joint channel access and power optimization problem. In Section III, we present the joint channel access and power optimization game model and investigate the properties of its NE. In Section IV, two distributed learning algorithms are proposed to achieve the general NE and optimal NE of the proposed game, respectively. Moreover, these two algorithms are extended to the case where the power strategies are continuous. Simulation results and analysis are presented in Section V. Concluding remarks are provided in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

We consider a multi-beam KU-band satellite communication system as shown in Fig.1, each beam covering a specific area. Frequency reuse is adopted between beams to improve spectrum efficiency. In downlink satellite communications, satellite receivers at any beam will suffer co-channel interference from other beams using the same frequency.

For any coverage area, the idle downlink channel of satellite communication is opportunistically accessed by the terrestrial cognitive users. Each cognitive user refers to a cognitive transmitter-receiver pair [35]. Through spectrum sensing [24], [25] or querying the local spectrum database [7], the terrestrial cognitive users can obtain information about whether each channel is occupied or idle. The available idle channels and the terrestrial cognitive users are denoted by $\mathcal{C} = \{1 \cdots C\}$ and $\mathcal{N} = \{1 \cdots N\}$, respectively. In addition, to facilitate the collaboration between cognitive users, channel state information and user transmission strategies need to be exchanged between cognitive users.

B. PROBLEM FORMULATION

By spectrum sensing or by querying the spectrum database, the terrestrial cognitive user n in beam $b \in \mathcal{B}$ selects an idle channel $s_n \in \mathcal{S}_n$ and power $p_n \in \mathcal{P}_n$ for transmission. The co-channel interference from satellite downlink signals experienced by cognitive user n can be expressed as

$$I_b^{s_n} = \sum_{j \in \mathcal{B}_n} p_j^{s_n} h_{j,b} G_{\max} \left(\frac{\lambda}{4\pi d} \right)^2, \tag{1}$$

where $\mathcal{B}_n$ is the set of satellite beams causing co-channel interference to terrestrial cognitive user n, $p_j^{s_n}$ is transmitted power allocated to channel s_n by beam $j \in \mathcal{B}_n$, $G_{\max}$ is the maximum transmission antenna gain of satellite, and $h_{j,b}$ $(0 < h_{j,b} < 1)$ represents the interference coefficient of beam $j \in \mathcal{B}_n$ to beam $b \in \mathcal{B}$. In addition, free space loss is expressed as $(\lambda/4\pi d)^2$, where d is the distance between satellite and the cognitive users, and λ is the working wavelength.

In addition to the co-channel interference from the downlink signals of satellite, cognitive user n also suffers the co-channel interference from other cognitive users occupying the same channel s_n. The received Signal to Interference plus Noise Ratio (SINR) of the terrestrial cognitive user n is given by

$$\eta_n(s_n, s_{-n}, p_n, p_{-n}) = \frac{p_n g_{nn}}{\sum\limits_{i \in \mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}, \tag{2}$$

where $\mathcal{F}_n(s_n)$ is the set of terrestrial cognitive users also choosing the same channel s_n as user n, σ denotes the background noise, p_n is the transmission power of cognitive user n, g_{nn} is the link gain from the transmitter to the receiver for cognitive user n, and $\sum_{i \in \mathcal{F}_n(s_n)} p_i g_{in}$ indicates the co-channel interference from other cognitive users to cognitive user n.

The achievable rate of cognitive user n is given by

$$R_n(a_n, a_{-n}, p_n, p_{-n}) = B \log(1 + \eta_n(a_n, a_{-n}, p_n, p_{-n})), \tag{3}$$

where B denotes the channel bandwidth. From the system optimization point of view, the total achievable throughput of all the terrestrial cognitive users can be expressed as

$$\begin{aligned} P1: \quad & \max_{s_n \in \mathcal{S}_n, p_n \in P_n} \sum_{n=1}^{N} R_n(s_n, s_{-n}, p_n, p_{-n}), \\ & s.t.\ 0 < p_n < p_n^{\max}, \quad \forall n \in \mathcal{N} \end{aligned} \tag{4}$$

where $p_n^{\max}$ indicates the maximum power constraint of cognitive user n.

However, solving the problem P1 is challenging for the following reasons: i) the strategy set of channel selection is discrete, and the strategy set of power control may be continuous or discrete. Even if we only consider the channel selection problem, it is a combinatorial optimization problem, which cannot be solved by standard optimization techniques with low complexity [30], and ii) there may be no centralized controller, so centralized optimization is not feasible. Based on the above analysis, we consider adopting game theory and distributed learning algorithms to solve this problem.

III. JOINT GAME MODEL FOR CHANNEL ACCESS AND POWER OPTIMIZATION

A. GAME MODEL

Denote the joint channel access and power optimization game as $\mathcal{G} = \{\mathcal{N}, \{\mathcal{A}_n\}_{n\in\mathcal{N}}, \{u_n\}_{n\in\mathcal{N}}\}$, where $\mathcal{N} = \{1, \cdots N\}$ is the cognitive user set, $\mathcal{A}_n$ is the strategy set of available actions for cognitive user n, $\mathcal{A}_n = \mathcal{S}_n \otimes \mathcal{P}_n$, where S_n is the strategy set of available channels for cognitive user n, $\mathcal{P}_n = \{p_n | 0 < p_n \le p_n^{\max}\}$ represents the power constraint of cognitive player n, and $\otimes$ denotes the Cartesian product. The utility function of player n is defined as

$$u_n(a_n, a_{-n}) = -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}}, \tag{5}$$

where $a_n \in \mathcal{A}_n$ and $a_n = s_n \otimes p_n$, a_n is the joint strategy of player n, $a_{-n} = a_1 \otimes \cdots a_{n-1} \otimes a_{n+1} \otimes \cdots \otimes a_N$ is the action profile of all the cognitive users excluding n, s_n and p_n indicate the channel selection and power control strategy of cognitive user n, respectively.

In equation (5), the first term indicates the negative reciprocal of SINR for cognitive user n, $\sum_{i\in\mathcal{F}_n(s_n)} (p_n g_{ni}/p_i g_{ii})$ is the penalty term, which represents the weighted interference of cognitive user n to other cognitive users, the weight factor is $p_i g_{ii} (i \in F_n(s_n))$.

B. ANALYSIS OF NASH EQUILIBRIUM

Definition 1 (Nash Equilibrium): The joint channel access and power control profile $a^* = (a_1^*, a_2^*, \cdots a_N^*)$ is a pure strategy NE if and only if no player can improve its utility by deviating unilaterally, i.e.,

$$u_n(a_n^*, a_{-n}^*) \ge u_n(a_n', a_{-n}^*), \qquad \forall n \in \mathcal{N}, \forall a_n' \in \mathcal{A}_n / \{a_n^*\}. \tag{6}$$

Theorem 1: The joint channel access and power optimization game $\mathcal{G}$ is an exact potential game, which has at least one pure-strategy NE point.

Proof: The following network potential function $\varphi(a_n, a_{-n}) = -\sum_{n=1}^{N} 1/r_n$ is constructed for the joint channel access and power optimization game, which is the sum of the negative reciprocal of SINR over all the users.

According to (2), the network potential function can be expressed as

$$\begin{aligned}
\varphi(a_n, a_{-n}) &= -\sum_{n=1}^{N} \frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} \\
&= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{j=1, j\ne n}^{N} \frac{\sum\limits_{i\in\mathcal{F}_j(s_j)} p_i g_{ij} + I_b^{s_j} + \sigma}{p_j g_{jj}} \\
&= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{j=1, j\ne n}^{N} \frac{p_n g_{nj}(s_n, s_j)}{p_j g_{jj}} \\
&\quad - \sum_{j=1, j\ne n}^{N} \frac{\sum\limits_{i\in\mathcal{F}_j(s_j), i\ne n} p_i g_{ij} + I_b^{s_j} + \sigma}{p_j g_{jj}},
\end{aligned} \tag{7}$$

where,

$$g_{nj}(s_n, s_j) = \begin{cases} 0, & s_n \ne s_j \\ g_{nj}, & s_n = s_j \end{cases} \tag{8}$$

Therefore, $\sum_{j=1, j\ne n}^{N} (p_n g_{nj}(s_n, s_j)/p_j g_{jj})$ can also be written as

$$\sum_{j=1, j\ne n}^{N} \frac{p_n g_{nj}(s_n, s_j)}{p_j g_{jj}} = \sum_{j\in\mathcal{F}_n(s_n)} \frac{p_n g_{nj}}{p_j g_{jj}} = \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}}. \tag{9}$$

Based on (7) and (9), we can get

$$\begin{aligned}
\varphi(a_n, a_{-n}) &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}} \\
&\quad - \sum_{j=1, j\ne n}^{N} \frac{\sum\limits_{i\in\mathcal{F}_j(s_j), i\ne n} p_i g_{ij} + I_b^{s_j} + \sigma}{p_j g_{jj}}.
\end{aligned} \tag{10}$$

According to (5) and (10), we have

$$\varphi(a_n, a_{-n}) = u_n(a_n, a_{-n}) - v(a_{-n}), \tag{11}$$

where $v(a_{-n}) = \sum_{j=1, j\ne n}^{N} \frac{\sum_{i\in\mathcal{F}_j(s_j), i\ne n} p_i g_{ij} + I_b^{s_j} + \sigma}{p_j g_{jj}}$ is independent of a_n. Thus, we also have

$$\varphi(a_n', a_{-n}) = u_n(a_n', a_{-n}) - v(a_{-n}). \tag{12}$$

Based on (11) and (12), we can get

$$u_n(a'_n, a_{-n}) - u_n(a_n, a_{-n}) = \varphi_n(a'_n, a_{-n}) - \varphi_n(a_n, a_{-n}). \tag{13}$$

According to the definition and properties of exact potential game given in [28], it can be seen that $\mathcal{G}$ is an exact potential game, which has at least one pure NE point. ∎

Theorem 2: When achieving the NE, if one user monopolizes a channel, the user on that channel will transmit at its maximum power.

Proof: The channel monopolized by cognitive user n is denoted by s_n^*. Terestrial cognitive user n will not interfere with other users, nor will it be interfered by other users, we readily get $\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} = 0$ and $\sum_{i\in\mathcal{F}_n(s_n^*)} (p_n g_{ni}/p_i g_{ii}) = 0$.

Assume that $0 < p_n < p_n^{\max}$, the joint strategy of the terrestrial cognitive user n is denoted by $a_n^* = s_n^* \otimes p_n^*$, and thus we can get

$$\begin{aligned} u_n(a_n^*, a_{-n}) &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n^* g_{ni}}{p_i g_{ii}} \\ &= -\frac{I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} \\ &< -\frac{I_b^{s_n^*} + \sigma}{p_n^{\max} g_{nn}}. \end{aligned} \tag{14}$$

Since channel s_n^* is only occupied by user n, it is easy to get $\sum_{i\in\mathcal{F}_n(s_n^*)} (p_n^{\max} g_{ni}/p_i g_{ii}) = 0$, which consequently leads to

$$\begin{aligned} -\frac{I_b^{s_n^*} + \sigma}{p_n^{\max} g_{nn}} &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} + \sigma}{p_n^{\max} g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n^*)} \frac{p_n^{max} g_{ni}}{p_i g_{ii}} \\ &= u_n(a_n, a_{-n}), \end{aligned} \tag{15}$$

where $a_n = s_n^* \otimes p_n^{\max}$. Combing (14) and (15), we have $u_n(a_n^*, a_{-n}) < u_n(a_n, a_{-n})$. To pursue its own payoff, terrestrial cognitive user n will deviate definitively the current strategy and to transmit at its maximum power, which contradicts with the assumption. This concludes the proof. ■

Theorem 3: In equally loaded or underloaded scenarios ($N \le C$), if $I_b^1 = I_b^2 = \cdots = I_b^C$, each channel is not occupied by multiple terrestrial cognitive users when achieving the NE.

Proof: We proof it by contradiction. Assuming that at the NE point, channel s_n^* is co-occupied by cognitive user n and other terrestrial cognitive users, there must be mutual interference between them. It is easy to get $\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} > 0$ and $\sum_{i\in\mathcal{F}_n(s_n^*)} (p_n g_{ni}/p_i g_{ii}) > 0$. The joint channel access and power control strategy is denoted by $a_n^* = s_n^* \otimes p_n^*$, thus we can get

$$\begin{aligned} u_n(a_n^*, a_{-n}) &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n^*)} \frac{p_n g_{ni}}{p_i g_{ii}} \\ &< -\frac{I_b^{s_n^*} + \sigma}{p_n^* g_{nn}}. \end{aligned} \tag{16}$$

Based on the assumption $I_b^1 = I_b^2 = \cdots = I_b^C$, it follows that

$$-\frac{I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} = -\frac{I_b^{s_n} + \sigma}{p_n^* g_{nn}} \le \frac{I_b^{s_n} + \sigma}{p_n^{\max} g_{nn}} \tag{17}$$

Since $N \le C$, according to the assumptions, at least one channel is not occupied by any cognitive user. Any one of the unoccupied channels is denoted by s_n. If the channel selection of cognitive user n is changed from s_n^* to s_n, and the strategies of other cognitive users remain unchanged, there must be no mutual interference between them. Thus, it is easy concluded that $\sum_{i\in\mathcal{F}_n(s_n)} p_i g_{in} = 0$ and $\sum_{i\in\mathcal{F}_n(s_n)} (p_n^{\max} g_{ni}/p_i g_{ii}) = 0$. Combining (16), (17), we have

$$\begin{aligned} u_n(a_n^*, a_{-n}) &< \frac{I_b^{s_n} + \sigma}{p_n^{\max} g_{nn}} \\ &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n^{\max} g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n^{max} g_{ni}}{p_i g_{ii}} \\ &= u_n(a_n, a_{-n}), \end{aligned} \tag{18}$$

where, $a_n = s_n \otimes p_n^{\max}$. Obviously, to pursue higher payoff, the cognitive user n will deviate from the selection s_n^*, s_n^* is not the channel selection of terrestrial cognitive user n at the NE point, which contradicts the former assumption. This concludes the proof. ■

Theorem 4: $\forall n \in \mathcal{N}$, $\forall s_n^* \in \mathcal{S}_n$ and $\forall p_n^* \in \mathcal{P}_n$, if $\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} \ge \max[I_b^1 \cdots I_b^C]$ holds, in equally loaded or underloaded scenarios, cognitive users do not interfere with each other when achieving the NE, i.e., each channel is occupied by at most one cognitive user.

Proof: We assume that the channel s_n^* is occupied by cognitive n and other cognitive users when achieving the NE, we have $\sum_{i\in\mathcal{F}_n(s_n^*)} (p_n g_{ni}/p_i g_{ii}) > 0$. The joint channel access and power control strategy is denoted by $a_n^* = s_n^* \otimes p_n^*$, we can get

$$\begin{aligned} u_n(a_n^*, a_{-n}) &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n^*)} \frac{p_n^* g_{ni}}{p_i g_{ii}} \\ &\le -\frac{\max[I_b^1 \cdots I_b^C] + \sigma}{p_n^* g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n^*)} \frac{p_n^* g_{ni}}{p_i g_{ii}} \\ &< -\frac{\max[I_b^1 \cdots I_b^C] + \sigma}{p_n^* g_{nn}}. \end{aligned} \tag{19}$$

Since $N \le C$, according to the assumption that the channel s_n^* is co-occupied by cognitive n and other cognitive users, there must exist one or more channels unoccupied by any other cognitive users. Without loss of generality, one of the unoccupied channel is denoted by s_n. If the terrestrial cognitive user n unilaterally changes its channel selection from s_n^* to s_n, it is easy to get $\sum_{i\in\mathcal{F}_n(s_n)} p_i g_{in} = 0$ and $\sum_{i\in\mathcal{F}_n(s_n)} (p_n^{max} g_{ni}/p_i g_{ii}) = 0$. Accordingly, it follows that

$$\begin{aligned} u_n(a_n^*, a_{-n}) &< -\frac{\max[I_b^1 \cdots I_b^C] + \sigma}{p_n^* g_{nn}} \\ &\le -\frac{I_b^{s_n} + \sigma}{p_n^* g_{nn}} \\ &\le -\frac{I_b^{s_n} + \sigma}{p_n^{\max} g_{nn}} \\ &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n^{max} g_{nn}} \\ &\quad - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n^{max} g_{ni}}{p_i g_{ii}} \\ &= u_n(a_n, a_{-n}), \end{aligned} \tag{20}$$

where $a_n = s_n \otimes p_n^{max}$. Obviously, to pursue its own higher payoff, terrestrial cognitive user n will choose channel s_n instead of s_n^*, which contradicts with the former assumption. This concludes the proof. ■

Theorem 4 can be intuitively understood as: If the inter-beam interference suffered on any channel is less than or equal to the total interference ($\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*}$) on the occupied channel, in equally loaded or underloaded scenarios, the cognitive user will select the unoccupied channel. This process continues until all channels are occupied by at most one user.

Theorem 5: When achieving NE, if there exists one or more channels not occupied by any users, the inter-beam interference on these channels must be greater than or equal to that on the occupied channels.

Proof: When achieving NE, without loss of generality, any one channel unoccupied and occupied by the terrestrial cognitive user n are denoted by $s_n \in S_n$ and $s_n^* \in S_n$, respectively. Thus, we can get

$$\begin{aligned} u_n(a_n^*, a_{-n}) &= -\frac{\sum\limits_{i\in\mathcal{F}(s_n^*)} p_i g_{in}(s_n^*, s_i) + I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} \\ &\quad - \sum_{i\in\mathcal{F}_n(s_n^*)} \frac{p_n g_{ni}}{p_i g_{ii}} \\ &\le -\frac{I_b^{s_n^*} + \sigma}{p_n^* g_{nn}} \\ &\le -\frac{I_b^{s_n^*} + \sigma}{p_n^{max} g_{nn}}. \end{aligned} \tag{21}$$

It is assumed that $I_b^{s_n^*} > I_b^{s_n}$, we have

$$-\frac{I_b^{s_n^*} + \sigma}{p_n^{max} g_{nn}} < -\frac{I_b^{s_n} + \sigma}{p_n^{max} g_{nn}}. \tag{22}$$

When only the channel selection of cognitive user n is changed from s_n^* to s_n, channel s_n is occupied only by cognitive user n, we have $\sum_{i\in\mathcal{F}_n(s_n)} (p_n^{max} g_{ni}/p_i g_{ii}) = 0$ and $\sum_{i\in\mathcal{F}_n(s_n)} p_i g_{in} = 0$. Combining (21) and (22), we have

$$\begin{aligned} u_n(a_n^*, a_{-n}) &< -\frac{I_b^{s_n} + \sigma}{p_n^{max} g_{nn}} \\ &= -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n^{max} g_{nn}} \\ &\quad - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}} \\ &= u_n(a_n, a_{-n}), \end{aligned} \tag{23}$$

where $a_n = s_n \otimes p_n^{max}$. Therefore, to pursue higher payoff, the terrestrial cognitive user n will definitively deviate its current channel selection from s_n^* to s_n, s_n^* is not the channel selection of terrestrial cognitive user n at the NE point, which contradicts with the former assumption. This concludes the proof. ■

Theorem 6: When the NE is achieved, the system utility of the joint channel access and power optimization game is lower bounded by $\sum_{n\in\mathcal{N}} u_n(a_n^*, a_{-n}) \ge -2J_O/C - \left((I + C\sigma)/C\right) \sum_{n\in\mathcal{N}} (1/p_n g_{nn})$, where $J_O = \sum_{n\in\mathcal{N}} \sum_{i\in\mathcal{N}, i\neq n} \left(p_i g_{in}/p_n g_{nn}\right)$ means the aggregated weighted interference experienced by all the cognitive users when occupying the same channel, and $I = \sum_{i\in\mathcal{C}} I_b^i$ is the sum of the inter-beam interference to all the available channels.

Proof: Based on the definition of NE, we have the following inequality:

$$u_n(a_n^*, a_{-n}) \ge u_n(a_n, a_{-n}). \tag{24}$$

where $\left(a_1^*, a_2^* \cdots a_N^*\right)$ refers to any pure-strategy NE point.

Based on (5) and (24), we have

$$u_n(a_n^*, a_{-n}) \ge -\frac{\sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}} \tag{25}$$

Summing the two sides of (25) upon all the available C channels to obtain

$$\begin{aligned} u_n(a_n^*, a_{-n}) \ge &-\frac{\sum\limits_{s_n\in\mathcal{C}} \sum\limits_{i\in\mathcal{F}_n(s_n)} p_i g_{in} + \sum\limits_{i\in\mathcal{C}} I_b^i + C\sigma}{C p_n g_{nn}} \\ &- \sum_{s_n\in\mathcal{C}} \sum_{i\in\mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{C p_i g_{ii}}. \end{aligned} \tag{26}$$

Note that $\sum_{s_n\in\mathcal{C}} \sum_{i\in\mathcal{F}_n(s_n)} p_i g_{in}$ refers to the aggregated interference suffered by cognitive user n when all the cognitive users occupy the same channel. Thus, we have:

$$\sum_{s_n\in\mathcal{C}} \sum_{i\in\mathcal{F}_n(s_n)} p_i g_{in} = \sum_{i\in\mathcal{N}, i\neq n} p_i g_{in}. \tag{27}$$

Also, $\sum_{s_n\in\mathcal{C}} \sum_{i\in\mathcal{F}_n(s_n)} p_n g_{nj}/p_i g_{ii}$ can be interpreted as the aggregated weighted interference experienced by all the cognitive users excluding cognitive user n, which corresponds to the following virtual scenario. That is, the terrestrial cognitive user n selects all the channels at the same time while all the other maintain their channel selection unchanged, the weighting factor is $1/p_i g_{ii}$. Therefore, we have the following result

$$\sum_{s_n \in \mathcal{C}} \sum_{i \in \mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}} = \sum_{i \in \mathcal{N}, i \neq n} \frac{p_n g_{ni}}{p_i g_{ii}}. \tag{28}$$

Combining (26)-(28), we can derive the following result

$$u_n(a_n^*, a_{-n}) \geq -\frac{\sum\limits_{i \in \mathcal{N}, i \neq n} p_i g_{in} + \sum\limits_{i \in \mathcal{C}} I_b^i + C\sigma}{C p_n g_{nn}} - \sum_{i \in \mathcal{N}, i \neq n} \frac{p_n g_{ni}}{C p_i g_{ii}}. \tag{29}$$

Therefore, the system utility of the formulated game can be expressed as

$$\sum_{n \in \mathcal{N}} u_n(a_n^*, a_{-n}) \geq -\sum_{n \in \mathcal{N}} \frac{\sum\limits_{i \in \mathcal{N}, i \neq n} p_i g_{in} + \sum\limits_{i \in \mathcal{C}} I_b^i + C\sigma}{C p_n g_{nn}} - \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} \frac{p_n g_{ni}}{C p_i g_{ii}}. \tag{30}$$

It is noted that $\sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} p_i g_{in}$ is the aggregated interference suffered by all the terrestrial cognitive users when selecting the same channel. $\sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} (p_i g_{in}/p_n g_{nn})$ and $\sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} (p_n g_{ni}/p_i g_{ii})$ both refer to the the aggregated weighted interference experienced by all the terrestrial cognitive users when they select the same channel. Thus, we have

$$\sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} \frac{p_i g_{in}}{p_n g_{nn}} = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} \frac{p_n g_{ni}}{p_i g_{ii}}. \tag{31}$$

Combining (30) and (31), we can derive the following results

$$\begin{aligned} \sum_{n \in \mathcal{N}} u_n(a_n^*, a_{-n}) &\geq -\frac{2J_O}{C} - \sum_{n \in \mathcal{N}} \frac{I + C\sigma}{C p_n g_{nn}} \\ &= -\frac{2J_O}{C} - \frac{(I + C\sigma)}{C} \sum_{n \in \mathcal{N}} \frac{1}{p_n g_{nn}}, \end{aligned} \tag{32}$$

where $J_O = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{N}, i \neq n} (p_i g_{in}/p_n g_{nn})$ means the aggregated weighted interference suffered by all the terrestrial cognitive users when occupying the same channel, and $I = \sum_{i \in \mathcal{C}} I_b^i$ is the sum of the inter-beam interference to all the available channels. The proof is completed. ■

Theorem 7: Suppose that at the NE point, each channel is not occupied by multiple cognitive users, then the system utility will reach the maximum if the user with the smaller product of power constraint and link gain occupy a better channel. That is, $\forall m, n \in \mathcal{N}$, if $p_m^{max} g_{mm} < p_n^{max} g_{nn}$, $I_b^{s_m} < I_b^{s_n}$ holds, $\sum_{n \in \mathcal{N}} u_n(a_n, a_{-n})$ will achieve the maximum.

Proof: At the NE point, when each channel is not occupied by multiple cognitive users, we have $\sum_{i \in \mathcal{F}_n(s_n)} (p_n g_{ni}/p_i g_{ii}) = 0$ and $\sum_{i \in \mathcal{F}_n(s_n)} p_i g_{in} = 0$. The system utility can be expressed as

$$\begin{aligned} \varphi 1 &= -\sum_{n=1}^{N} \frac{\sum\limits_{i \in \mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n g_{nn}} - \sum_{i \in \mathcal{F}_n(s_n)} \frac{p_n g_{ni}}{p_i g_{ii}} \\ &= -\sum_{n=1}^{N} \frac{\sigma + I_b^{s_n}}{p_n g_{nn}}. \end{aligned} \tag{33}$$

Theorem 2 has proven that if each channel is not occupied by multiple cognitive users, at the NE point, $p_n = p_n^{max}$ holds. Supposing that there exists one or more terrestrial cognitive users with the smaller product of power constraint and link gain occupy worse channels, i.e., $\exists\, m, n \in \mathcal{N}$, if $p_m^{max} g_{mm} < p_n^{max} g_{nn}$, $I_b^{s_m} > I_b^{s_n}$ holds. Thus, we have

$$\begin{aligned} \varphi 1 &= -\sum_{n=1}^{N} \frac{\sigma + I_b^{s_n}}{p_n g_{nn}} \\ &= -\sum_{n=1}^{N} \frac{\sigma + I_b^{s_n}}{p_n^{max} g_{nn}} \\ &= -\frac{\sigma + I_b^{s_n}}{p_n^{max} g_{nn}} - \frac{\sigma + I_b^{s_m}}{p_m^{max} g_{mm}} - \sum_{i=1, i \neq m, i \neq n}^{N} \frac{\sigma + I_b^{s_i}}{p_i^{max} g_{ii}}. \end{aligned} \tag{34}$$

When cognitive user m and n exchange their channel selection, and other users maintain their choices unchanged, the system utility can be expressed as follows

$$\varphi_2 = -\frac{\sigma + I_b^{s_m}}{p_n^{max} g_{nn}} - \frac{\sigma + I_b^{s_n}}{p_m^{max} g_{mm}} - \sum_{i=1, i \neq m, i \neq n}^{N} \frac{\sigma + I_b^{s_i}}{p_i^{max} g_{ii}}. \tag{35}$$

Combining (34) and (35), we can get

$$\begin{aligned} \varphi_2 - \varphi_1 &= \frac{I_b^{s_n} - I_b^{s_m}}{p_n^{max} g_{nn}} + \frac{I_b^{s_m} - I_b^{s_n}}{p_m^{max} g_{mm}} \\ &= \left(I_b^{s_m} - I_b^{s_n}\right) \left(\frac{1}{p_m^{max} g_{mm}} - \frac{1}{p_n^{max} g_{nn}}\right). \end{aligned} \tag{36}$$

According to the assumption that $p_m^{max} g_{mm} < p_n^{max} g_{nn}$ and $I_b^{s_m} > I_b^{s_n}$, we can get

$$\varphi_2 - \varphi_1 > 0. \tag{37}$$

Therefore, for any two cognitive users, if the cognitive user with the smaller product of power constraint and link gain occupies the worse channel, by exchanging their selections while the strategies of other cognitive users maintain unchanged, the system utility will increase. If this continues, the system utility will continue to increase until the user with the smaller product of power constraint and link gain occupies the better channel between any two cognitive users. At this time, the system utility reaches the maximum. The proof is completed. From theorem 7, we can see that fairness is well reflected in our game model. In other words, if each channel is not occupied by multiple cognitive users, cognitive users with a smaller $p_n^{max} g_{nn}$ will occupy a better channel when achieving the optimal NE. ■

IV. DISTRIBUTED ALGORITHMS

As the joint channel access and power control problem now formulated as an exact potential game, it is another essential work to achieve the NE. Many distributed algorithms can be applied to achieve the NE of the potential game, including best response [28], stochastic-learning automata [35], no-regret learning [30], Q-learning algorithm [42] and trial and error learning [27]. However, it is still not an easy work to achieve the NE solution in the channel access and power optimization game due to the following reasons: i) The strategies of channel access are discrete, while the power control strategies may be continuous. Accordingly, the number of strategies in the joint two-dimensional strategy space may be infinite. These conventional algorithms cannot be directly applied in cases where the number of the two-dimensional strategies is infinite, and it is more challenging to find the optimal NE solution. ii) Even though the power control strategies are also discrete, the number of joint strategies in the two-dimensional strategy space is large (the number of joint strategy in the two-dimensional strategy space in each iteration is $C * K$, C is the number of available channels, K is the number of power control strategies). It requires considerable computation complexity and storage space. Besides, for the combination optimization problem, standard optimization techniques cannot be applied directly to obtain the globally optimal solution [30]. If the exhaustive search approach is applied, a central controller and a considerable amount of computation are needed, which is impractical for implementation. (the number of combination strategies in the two-dimensional strategy space is N^{C*K}).

For discrete power control strategies, we propose the JID algorithm to achieve the general NE of the joint channel access and power optimization game. Besides, inspired by the process of exploration and exploitation in reinforcement learning algorithms [43], [44], the JIDEE algorithm is designed to achieve the optimal NE. Simultaneously, these two algorithms are extended to the case where the power control strategies are continuous. In the implementation of these algorithms, some information exchange ((e.g., user transmission strategies, channel state information) is needed, which can be regarded as a reflection of cooperation between terrestrial cognitive users.

A. JOINT-STRATEGY ITERATION ALGORITHM FOR DISCRETE POWER CONTROL STRATEGIES

As discussed above, if the strategies of channel access and power control are both discrete, the number of the joint two-dimensional strategies may be huge, which leads to considerable computation complexity and large storage requirements. To solve this problem, a decomposed joint-strategy iteration algorithm suitable for the discrete power control strategies is proposed, which can converge to the general joint-strategy NE point. The detail of the joint-strategy iteration algorithm suitable for the discrete power control strategies (JID) is listed in algorithm 1.

Algorithm 1 Joint-Strategy Iteration Algorithm for Discrete Power Control Strategies (JID)

Initialization: Set $t = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $s_n \in \mathcal{S}$ and selects power $p_n = p_n^{\max}/K$.
For $t = 2,\ 3, \cdots L$
  $\mathbf{a}^t = \mathbf{a}^{t-1}$ ($\mathbf{s}^t = \mathbf{s}^{t-1}$, $\mathbf{p}^t = \mathbf{p}^{t-1}$)
  For $n = 1 : N$,
    Fix the power at the value updated in the last iteration, terrestrial cognitive user n chooses a best response according to $s_n^t = \underset{s_n \in \mathcal{S},\, \mathbf{p}=\mathbf{p}^t}{\arg\max} \left(u_n\left(s_n,\ s_{-n}^t\right)\right)$.
    Fix the channel selection updated in the last iteration, terrestrial cognitive user n chooses a best response according to $p_n^t = \underset{p_n \in \mathcal{P}_n,\, \mathbf{s}=\mathbf{s}^t}{\arg\max} \left(u_n\left(p_n,\ p_{-n}^t\right)\right)$.
  End For
End For: until $\mathbf{a}^t = \mathbf{a}^{t-1}$ or reach the maximum number of iterations L.

Each terrestrial cognitive user independently makes decisions about its channel access and power control strategy. Therefore, no central controller is needed, which greatly reduces the computation complexity.

B. JOINT-STRATEGY ITERATION ALGORITHM FOR CONTINUOUS POWER CONTROL STRATEGIES

The proposed JID is designed for the case that the power control strategies are discrete. In this subsection, a joint-strategy iteration algorithm for continuous power control strategies (JIC) is designed, which is easy to implement and can also converge to the joint-strategy NE. The optimal power control strategy can be obtained by the partial derivative of utility function to power.

$$\frac{\partial u_n(a_n, a_{-n})}{\partial p_n} = \frac{\sum\limits_{i \in \mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{p_n^2 g_{nn}} - \sum_{i \in \mathcal{F}_n(s_n)} \frac{g_{ni}}{p_i g_{ii}} = 0. \tag{38}$$

Then, the optimal power strategy can be expressed as

$$p_n^* = \sqrt{\frac{\sum\limits_{i \in \mathcal{F}_n(s_n)} p_i g_{in} + I_b^{s_n} + \sigma}{g_{nn} \sum\limits_{i \in \mathcal{F}_n(s_n)} \frac{g_{ni}}{p_i g_{ii}}}}. \tag{39}$$

The detail of the JIC is listed in algorithm 2.

It is noted that although the JIC is updated by a step size δ to make it easier to find the global optimal solution, it does not necessarily guarantee to convergence to the optimal NE point. In addition, It is worth noting that although a larger step size δ will accelerate the convergence speed of the JIC, it may also make the value of $p_n^{i-1} + \delta(\min\left(p_n^*,\ p_n^{\max}\right) - p_n^{i-1})$ negative, so an appropriate value of δ needs to be carefully set.

Algorithm 2 Joint-Strategy Iteration Algorithm for Continuous Power Control Strategies (JIC)

Initialization: Set $i = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $s_n \in \mathcal{S}$ and sets the initial power $p_n = p_n^{\max}/2$.
For $i = 2,\ 3, \cdots L$
  $\mathbf{a}^i = \mathbf{a}^{i-1}$ ($\mathbf{s}^i = \mathbf{s}^{i-1}$, $\mathbf{p}^i = \mathbf{p}^{i-1}$)
  For $n = 1 : N$**,**
    Fix the power at the value updated in the last iteration, terrestrial cognitive user n chooses a best response according to $s_n^i = \underset{s_n \in \mathcal{S},\, \mathbf{p}=\mathbf{p}^i}{\arg\max} \left(u_n\left(s_n,\ s_{-n}^i\right)\right)$.
    Fix the channel selection updated in the last iteration, terrestrial cognitive user n updates its power according to

$$p_n^* = \underset{p_n \in \mathcal{P}_n, \mathbf{s}=\mathbf{s}^i}{\arg\max} \left(u_n(p_n,\ p_{-n}^i)\right). \tag{40}$$

$$p_n^i = p_n^{i-1} + \delta(\min\left(p_n^*,\ p_n^{\max}\right) - p_n^{i-1}). \tag{41}$$

  End For
End For: until $\mathbf{a}^i = \mathbf{a}^{i-1}$ or reach the maximum number of iterations L.

Theorem 8: The proposed JID and JIC converge to the general NE point.

Proof: According to the definition of exact potential game, the equation $u_n(a'_n, a_{-n}) - u_n(a_n, a_{-n}) = \varphi_n(a'_n, a_{-n}) - \varphi_n(a_n, a_{-n})$ always holds. That is, the change in the utility function caused by any unilateral deviation is consistent with the change in the exact potential function. In the JID, although the channel access and power control are decomposed into two steps, the best response algorithm for each step guarantees the improvement of the individual's utility when other strategies unchanged, which achieves a feasible improvement path in system utility. Besides, the feasible strategy profiles are finite; the improvement path must be limited and will terminate in a pure NE point. In the JIC, through the partial derivative of the utility function to the transmit power, and makes it to zero, the value of utility function will not decrease in each iteration. Thus, it will also eventually converge to the general NE. ■

C. JOINT-STRATEGY ITERATION ALGORITHM FOR DISCRETE POWER CONTROL STRATEGIES BASED ON EXPLORATION AND EXPLOITATION

Both of the above algorithms converge to a general NE in the joint two-dimensional strategy space, but it is challenging to find the optimal NE solution. Most existing algorithms converge to a general NE, such as best/better response [28], no regret learning [30], stochastically learning automata [35], and reinforcement learning [42]. Although the trial and error algorithm (TE) can statistically converge to the optimal NE [27], it needs lots of iterations and will deviate from the optimal NE with a certain probability. Moreover, finding the optimal NE in a two-dimensional strategy space is even more difficult. Inspired by the thought of exploration and exploitation in reinforcement learning algorithms [43], [44], a joint-strategy iteration algorithm for discrete power control strategies based on exploration and exploitation (JIDEE) is proposed, which will eventually converge to the optimal NE after a sufficient number of iterations. The detail of JIDEE is listed in algorithm 3.

Algorithm 3 Joint-Strategy Iteration Algorithm for Discrete Power Control Strategies Based on Exploration and Exploitation (JIDEE)

Initialization: Set $i = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $s_n^i \in S$ and selects power $p_n = p_n^{\max}/K$.
For $i = 2,\ 3, \cdots L$
  Initializaion: Set $j = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $\bar{s}_n^j \in \mathcal{S}$.
  For $j = 2,\ 3, \cdots Q$
    Fix the power at the value updated in the last iteration, each cognitive user $n \in \mathcal{N}$ sequentially updates its channel selection according to the best response, i.e.,

$$\bar{s}_n^j = \underset{\bar{s}_n \in \mathcal{S},\, \mathbf{p}=\mathbf{p}^{i-1}}{\arg\max} \left(u_n\left(\bar{s}_n,\ \bar{s}_{-n}^{j-1}\right)\right)$$

.
  End For: until $\bar{\mathbf{s}}^j = \bar{\mathbf{s}}^{j-1}$.
  Channel selection vector is set to $\mathbf{s}^i = \bar{\mathbf{s}}^j$, the power vector is set to $\mathbf{p}^i = \mathbf{p}^{i-1}$, and calculate the system utility SU(i) according to (7).
  If i>2
    If $SU(i) > \overline{SU}(i-1)$, the channel selection vector $\mathbf{s}(i) = \bar{\mathbf{s}}(j)$
    Else the channel selection vector $\mathbf{s}(i) = \mathbf{s}(i-1)$
    End If
  Else
  End If
  Fix the channel selection at the value updated in the last iteration, each terrestrial cognitive user n sequentially chooses a best response according to

$$p_n^i = \underset{p_n \in \mathcal{P}_n,\, \mathbf{s}=\mathbf{s}^{i-1}}{\arg\max} \left(u_n\left(p_n,\ p_{-n}^{i-1}\right)\right)$$

.
  Channel selection vector is set to $s1^i$, the power control vector is set to p^i, and compute the system utility $\overline{SU}(i)$ according to (7).
  If i==2, $SU(1) = SU(2)$
  Else
  End If
End For: until $\mathbf{a}^i = \mathbf{a}^{i-1}$ or reach the maximum number of iterations L.

The implementation of JIDEE can be summarized as follows: Based on power optimization in the last iteration, it randomly selects channels for exploration in step 3, and achieve the NE from step 4 to 6. Step 7, 8 and 10 are used to calculate the system utility, compare, and choose the better channel combination. Step 9 is used for power optimization. Continue to explore, find a better channel choice, and optimize the power, which cycles until it achieves the optimal NE.

Theorem 9: The proposed JIDEE will converge to the optimal NE point when the power control strategy set is sufficiently enough.

Proof: The JIDEE includes three stages: 1) channel selection process through best response, 2) power control process through best response, and 3) exploration and exploitation stage. The Theorem 8 has proved that stage 1 and 2 can make sure it converges to the general NE, the system utility may be locally optimal or globally optimal. Through the process of exploration and exploitation, when iterating a sufficient number of times, the system utility will eventually jump out of the local optimum and reach the global optimum.

It should be noted that in order to achieve the optimal NE, the power control strategies must be sufficiently large. In this way, in the power control stage, when the best response method is used to optimize the power, it is beneficial for potential function to produce more subtle changes. Therefore, it is favorable for the potential function to jump out of the local optimum in the channel exploration and exploitation stage. After enough times of exploration and exploitation, the potential function reaches the maximum, which corresponds to the optimal Nash equilibrium solution. The proof is completed. ■

D. JOINT-STRATEGY ITERATION ALGORITHM FOR CONTINUOUS POWER CONTROL STRATEGIES BASED ON EXPLORATION AND EXPLOITATION

The JIDEE converges to the optimal NE when the power control strategies are discrete. In this subsection, we study the joint-strategy iteration algorithm for continuous power control strategies based on exploration and exploitation (JICEE). As with the JIDEE, the JICEE also has a process of exploration and exploitation, with the main difference being the power optimization stage. Since the power control strategies are continuous, the power can be optimized by Eq. (39)-(41). The detail of the JICEE is listed in algorithm 4.

Theorem 10: The proposed JICEE will converge to the optimal NE point.

Proof: The JICEE includes three stages: 1) channel selection process through the best response method, 2) power control process, 3) exploration and exploitation stage. The Theorem 8 has proved that stage 1 and 2 can make sure it converges to the general NE, the system utility may be locally optimal or globally optimal.

It should be noted that in the power control stage, the power changes slightly, it is conducive for the potential function to produce subtle changes, which is similar to the case where the number of power control strategies in the JIDEE is sufficiently large. Therefore, after enough times of exploration and exploitation, the potential function will eventually achieve the maximum value, which corresponds to the optimal NE solution. The proof is completed. ■

Algorithm 4 Joint-Strategy Iteration Algorithm for Continuous Power Control Strategies Based on Exploration and Exploitation (JICEE)

Initialization: Set $i = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $s_n^i \in \mathcal{S}$ and sets the initial power $p_n^i = p_n^{\max}/2$.
For $i = 2, 3, \cdots L$
  Initializaion: Set $j = 1$, each cognitive user $n \in \mathcal{N}$ randomly chooses a channel $\bar{s}_n^j \in \mathcal{S}$.
  For $j = 2, 3, \cdots Q$
    Fix the power at the value updated in the last iteration, each cognitive user $n \in \mathcal{N}$ sequentially updates its channel selection according to the best response, i.e.,

$$\bar{s}_n^j = \underset{\bar{s}_n \in S,\, \mathbf{p}=\mathbf{p}^{i-1}}{\arg\max} \left(u_n \left(\bar{s}_n,\ s2_{-n}^{j-1} \right) \right)$$

.

  End For until $\bar{\mathbf{s}}^j = \bar{\mathbf{s}}^{j-1}$.
  The channel selection vector is set to $\mathbf{s}^i = \bar{\mathbf{s}}^j$, the power vector is $\mathbf{p}^i$, and calculate the system utility SU(i) according to (7).
  If i>2
    If $SU(i) > \overline{SU}(i-1)$, the channel selection vector $\mathbf{s}(i) = \bar{\mathbf{s}}(j)$
    Else the channel selection vector $\mathbf{s}(i) = \mathbf{s}(i-1)$
    End If
  Else
  End If
  Fix the channel selection at the value updated in the last iteration, each terrestrial cognitive user n sequentially chooses a best response according to

$$p_n^* = \underset{p_n \in \mathcal{P}_n, \mathbf{s}=\mathbf{s}^{i-1}}{\arg\max} \left(u_n(p_n,\ p_{-n}^{i-1}) \right)$$

$$p_n^i = p_n^{i-1} + \delta(\min\left(p_n^*,\ p_n^{\max}\right) - p_n^{i-1})$$

  Channel selection vector is set to $s1^i$, the power control vector is set to p^i, and compute the system utility $\overline{SU}(i)$ according to Eq. (7).
  If i==2, $SU(1) = SU(2)$;
  Else
  End If
End For: until $\mathbf{a}^i = \mathbf{a}^{i-1}$ or reach the maximum number of iterations L.

V. SIMULATION RESULTS AND BEHAVIOR

We consider a multi-beam GEO satellite communication system that has 10 beams. For each beam, there are multiple downlink channels, the bandwidth and power of each channel

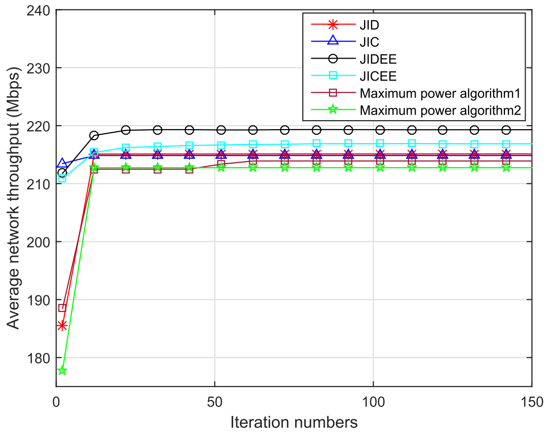


FIGURE 2. **The convergence of the proposed algorithms ($N = 10$, $C = 4$, $\delta = 0.5$, $K = 6$).**

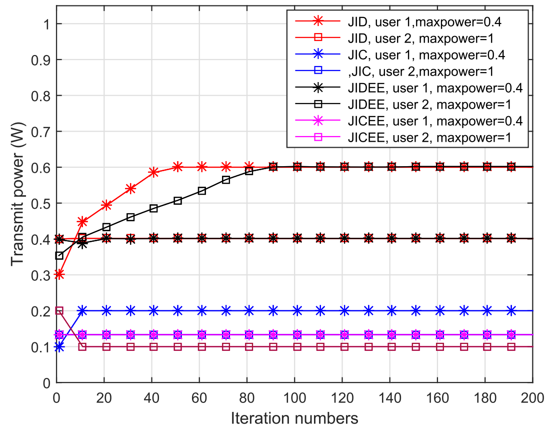


FIGURE 4. **Evolution of the power control for any two selected cognitive users in an area 150m × 150m ($N = 10$, $C = 4$, $\delta = 0.5$, $K = 6$).**

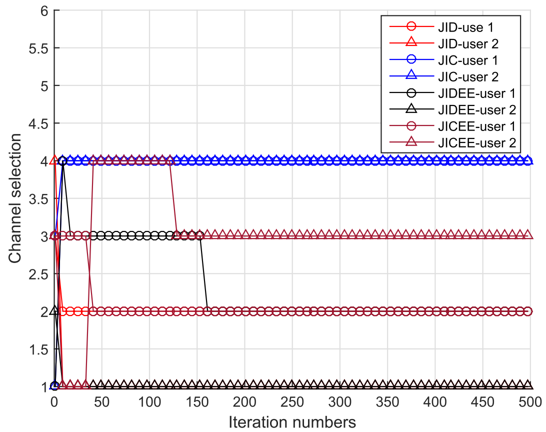


FIGURE 3. **Evolution of the channel selection for any two selected cognitive users cognitive users ($N = 10$, $C = 4$, $\delta = 0.5$, $K = 6$).**

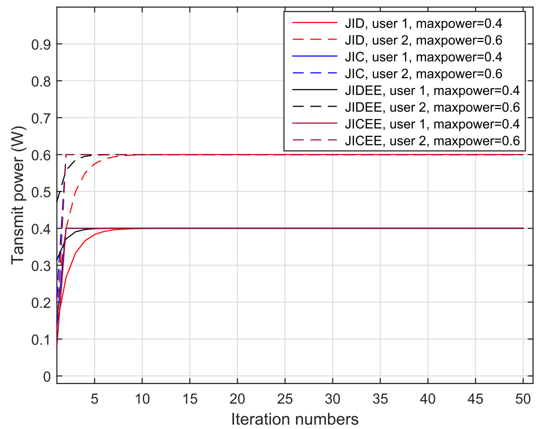


FIGURE 5. **Evolution of the power control for any two selected cognitive users in an area 150m × 150m ($N = 4$, $C = 4$, $\delta = 0.5$, $K = 6$).**

are set to $B = 3MHz$ and $p = 1W$, respectively. According to TTU-R.672, the maximum antenna gain of the satellite is set to $G_{max} = 50dBi$ [45]. In the converge of beam b, there are four available idle channels, and ten terrestrial cognitive users are randomly distributed in a 200*200m square area. Unless otherwise specified, the maximum transmission power for each cognitive user is set to $p_{max} = 0.8W$. The distance between the transmitter and receiver of each terrestrial cognitive user is set to $d = 20m$ [35]. In all simulations, the transmission loss between terrestrial cognitive users is calculated by ideal attenuation formulation $L = 32.5 + 20\lg D\,(km) + 20\lg F\,(MHz)$, where F is set to 14000 [27]. It is assumed that the cognitive users are interfered by three co-channel beams, the interference coefficient to beam b is set to $h = [0.3, 0.2, 0.1]$, respectively [46]. Also,we assume that the four available channels are independently occupied in the three co-channel beams with a probability of 1/2.

A. CONVERGENCE BEHAVIOR

The convergence performance of the proposed algorithms is shown in Fig.2-Fig.7. To verify the effectiveness of the formulated game model and the feasibility of the proposed algorithms, two other schemes based on best response are considered for comparison: (i) Maximum power algorithm 1, (ii) Maximum power algorithm 2. Both approaches do not consider the coupling characteristic between channel selection and power control, so the cognitive users alway transmit at their maximum power. The difference is that the Maximum power algorithm 2 does not consider the cooperation between users and directly optimizes the throughput. In contrast, the Maximum power algorithm 1 considers the cooperation between users, maximize the utility function (5) presented in this paper. The step size δ for the JIC and the JICEE is set to 0.5, and the number of power control strategies for the JID and the JIDEE is set to 6.

Fig.2 shows the convergence behaviors of the proposed algorithms. The results are obtained through the expectations of 200 independent trails. It can be seen that the JID and the JIC converge to a pure strategy in about 10 iterations, while the JIDEE and the JICEE generally need more than 100 iterations to converge to a pure strategy. The reason is that there is a process of exploration and exploitation in the implementation of the JIDEE and the JICEE. Besides, it can be seen that the proposed algorithms achieve better network throughput than the other two compared ones. The reason is that in the implementation of the two compared algorithms,

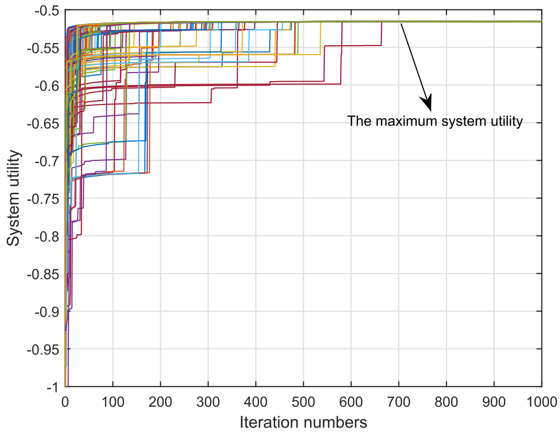


FIGURE 6. **The convergence behavior of the JIDEE with random initial channel selection ($N = 10$, $C = 4$, $K = 600$).**

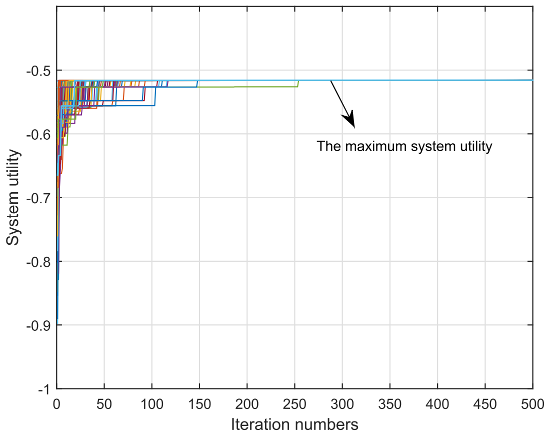


FIGURE 7. **The convergence behavior of the JICEE with random initial channel selection ($N = 10$, $C = 4$, $\delta = 0.5$).**

the cognitive users always transmit at their maximum power, and will cause more significant interference to other cognitive users. Not only that, but the Maximum power algorithm 2 also does not consider the cooperation between cognitive users, so its throughput performance is the worst. It should be pointed out that the JIDEE and the JICEE converge to different optimal NE point because they have different strategy spaces.

Fig.3 shows the evolution of channel selection for any two selected cognitive users. It can be seen that the terrestrial cognitive users remain their current channel selection strategies unchanged within 200 iterations. It can also be seen that the JIDEE and the JICEE has a longer convergence time than the JID and the JIC.

Fig.4 and Fig.5 show the evolution of power control for any two selected cognitive users. The present results show that, for the proposed algorithms, the cognitive users remain their power control strategies unchanged after several iterations, and the JIDEE and the JICEE has a longer convergence time than the JID and the JIC. Besides, it is seen from Fig.6 that when $N \leq C$, the cognitive users transmit at their maximum power at the NE point. The reason can be explained by Theorem 2 and Theorem 4. That is, in equally loaded

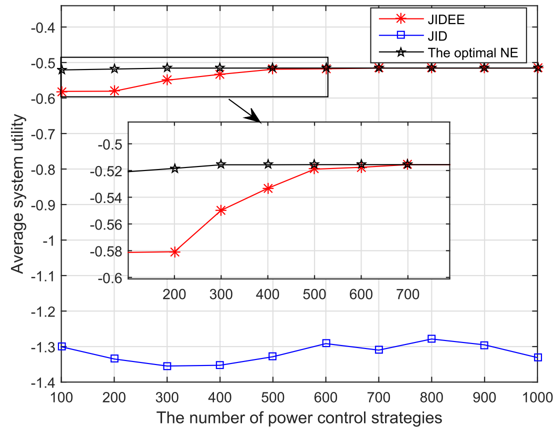


FIGURE 8. **The network throughput for the JID and the JIDEE with different number of power control strategies. ($N = 10$, $C = 4$).**

or underloaded scenarios, when $\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} \geq \max[I_b^1 \cdots I_b^{s_n}]$, each channel is occupied by at most one cognitive user, and each cognitive user will transmit at its maximum power when achieving the NE point.

The convergence behaviors of the JIDEE and the JICEE with random initial channel selection are shown in Fig.6 and Fig.7, respectively. The presented results are obtained by simulating 100 independent convergence curves. It can be seen from the two figures that through continuous exploration and exploitation, the system utility keeps increasing, and finally reaches the maximum value, which corresponds to the optimal NE point.

B. IMPACT OF DIFFERENT PARAMETERS

At different parameter values, we examine the network throughput, convergence speed and system utility for the proposed algorithms.

Fig.8 shows the network throughput performance for the JID and the JIDEE with different number of power control strategies. The results are obtained through the expectations of 200 independent trials. As can be seen from the figure, for the JID, with the increase in the number of power control strategies, the average system utility does not increase. But for the JIDEE, with the increase in the number of power control strategies, the average system utility increases. When K is large enough ($K \geq 600$), the JIDEE converges to the optimal NE point, which can be explained by Theorem 9.

Fig.9 shows the network throughput performance for the JIC with different step size δ. For each step size δ, we simulate 2 independent convergence curves. The figure shows that different δ may converge to different NE point, even for the same δ, different initial conditions may converge to different results. It can also be seen that when converging to the same NE point, the larger δ, the faster the convergence speed. It should be noted that the parameter δ should not be set too large, so as not to make the value of power negative according to Eq. (41).

Fig.10 shows the system utility for the JICEE with different step size δ. For each step size δ, we draw any one random

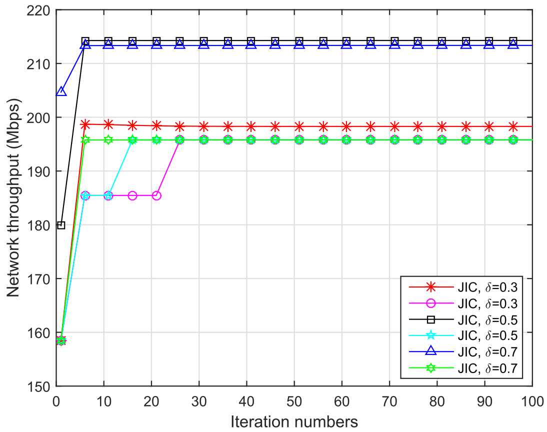


FIGURE 9. **The network throughput for the JIC with different step size δ ($N = 10$, $C = 4$).**

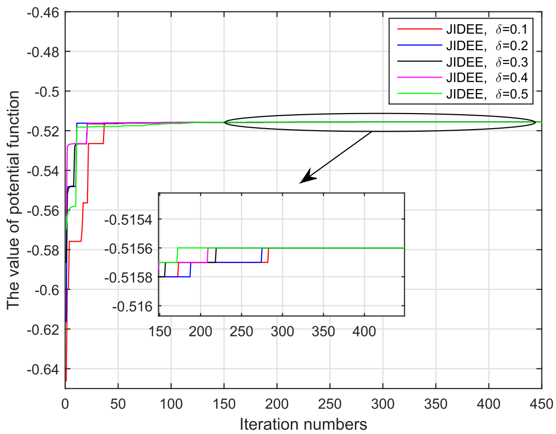


FIGURE 10. **The system utility for the JICEE with different step size δ ($N = 10$, $C = 4$).**

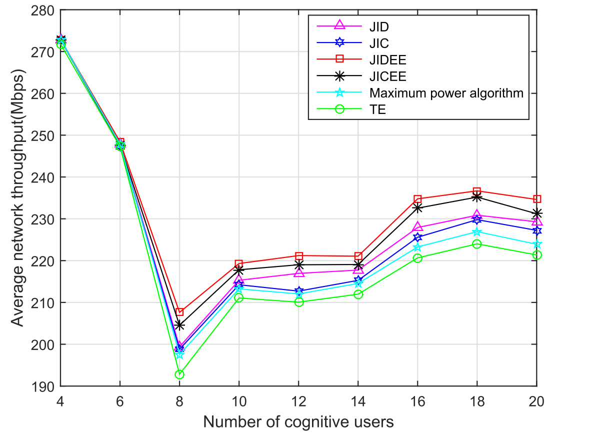


FIGURE 11. **The comparison of the network throughput when varying the number of cognitive users in an square area $200m * 200m$ ($C = 4$, $\delta = 0.5$, $K = 6$).**

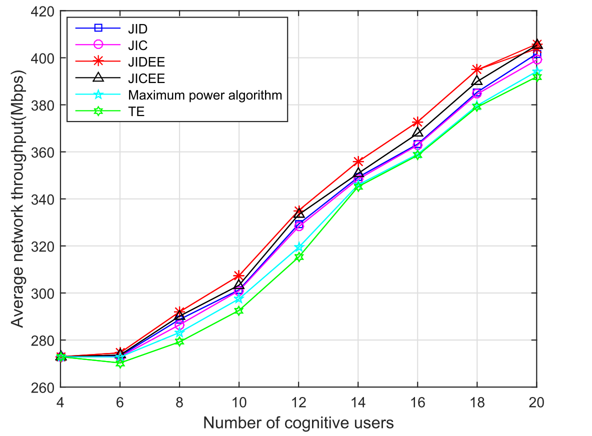


FIGURE 12. **The comparison of the network throughput when varying the number of cognitive users in an square area $400m * 400m$ ($C = 4$, $\delta = 0.5$, $K = 6$).**

convergence curve. As can be seen from the figure that whatever δ is, the JICEE algorithm will eventually converge to the optimal NE point. Also, it can be seen that as δ increases, the convergence speed becomes faster. As with the JIC, the step size δ for the JICEE cannot be set too large to make the value of power negative.

C. THROUGHPUT PERFORMANCE

In this subsection, we evaluate the throughput performance of the proposed algorithms. For comparison, two other approaches are considered: the Maximum power algorithm and the trial and error (TE) algorithm. 1) Maximum power algorithm: each terrestrial cognitive user always transmits with its maximum power, maximizes the utility function (5) presented in this paper, and selects the channel with the best response algorithm. 2) TE algorithm proposed in [27]: each user transmits with the maximum power, maximize its throughput, and selects the channel with the TE algorithm. The step size δ for the JIC and the JICEE is set to 0.5, and the number of power control strategies for the JID and the JIDEE is set to 6. In addition, as in [27], the learning parameter ϵ for TE algorithm is set to 0.005.

Fig.11 and Fig.12 show how network throughput varies with the number of cognitive users in different size regions. As can be seen from the two figures, compared with the Maximum power algorithm and the TE algorithm, our proposed algorithms achieve better network throughput performance. The reason is that both the Maximum power algorithm and the TE algorithm do not consider the coupling characteristic between the channel selection and power control. Besides, each cognitive user optimizes its own throughput without considering the cooperation between the cognitive users in the implementation of the TE algorithm. Therefore, the throughput performance of the TE algorithm is the worst.

As can be seen from Fig.11 and Fig.12, the system throughput does not always increase with the increase in the number of cognitive users. The reason is that as the number of cognition increases, the interference between them will become more serious, especially in a smaller area ($200 * 200$).

Fig.13 shows how network throughput varies with the number of available channels. As can be seen from the figure, the network throughput increases with the number of available channels, and our proposed algorithms achieve better performance in network throughput than the Maximum

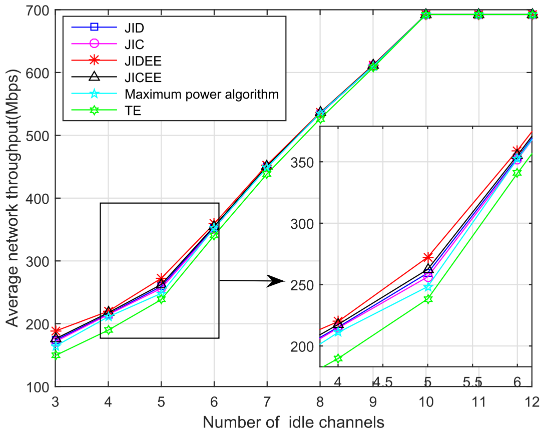


FIGURE 13. The comparison of the network throughput when varying the number of available channels ($N = 10$, $C = 4$, $\delta = 0.5$, $K = 6$, the square area is $200m * 200m$).

power algorithm and the TE algorithm. It should be noted that with the increase of available channels, the throughput performance of each algorithm gradually approaches. As described in Theorem 7, when the number of available channels is equal to or greater than the number of terrestrial cognitive users and the condition $\sum_{i\in\mathcal{F}_n(s_n^*)} p_i g_{in} + I_b^{s_n^*} \geq \max[I_b^1 \cdots I_b^C]$ is satisfied, each cognitive user occupies a channel alone. At this time, the network throughput reaches the maximum.

VI. CONCLUSION

In this paper, We studied the spectrum sharing in the cognitive satellite communication system, using a game-theoretic solution. The terrestrial cognitive users suffer from co-channel interference from other cognitive users and inter-beam interference. By taking the coupling characteristic of channel access and power control into account, we first formulated the channel access and power optimization problem as a joint channel access and power optimization game. It was proven to be an exact potential game and accordingly has at least one pure Nash Equilibrium (NE) point. The sufficient conditions for interference-free between cognitive users and the sufficient conditions for maximizing the system utility were given, and the performance bounds of the NE were theoretically analyzed. Then, for the case of discrete power control strategies, two distributed algorithms were proposed to converge to the general and optimal NE, respectively. Besides, these two algorithms were extended to the case where the power control strategies are continuous. Simulation results confirm the effectiveness of the formulated game and the feasibility of the proposed algorithms.

REFERENCES

[1] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "Cognitive radio techniques for satellite communication systems," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Las Vegas, NV, USA, Sep. 2013, pp. 1–5.

[2] B. Li, Z. Fei, Z. Chu, F. Zhou, K.-K. Wong, and P. Xiao, "Robust chance-constrained secure transmission for cognitive satellite–terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4208–4219, May 2018.

[3] A. Sultana, L. Zhao, and X. Fernando, "Energy-efficient power allocation in underlay and overlay cognitive device-to-device communications," *IET Commun.*, vol. 13, no. 2, pp. 162–170, Jan. 2019.

[4] J. Eze, S. Zhang, E. Liu, and E. Eze, "Cognitive radio-enabled Internet of vehicles: A cooperative spectrum sensing and allocation for vehicular communication," *IET Netw.*, vol. 7, no. 4, pp. 190–199, Jul. 2018.

[5] K. Kumar, A. Prakash, and R. Tripathi, "A spectrum handoff scheme for optimal network selection in NEMO based cognitive radio vehicular networks," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–16, Jan. 2017.

[6] C. Wu, Y. Wang, and Z. Yin, "Energy-efficiency opportunistic spectrum allocation in cognitive wireless sensor network," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, p. 13, Jan. 2018.

[7] M. Hoyhtya, A. Mammela, X. Chen, A. Hulkkonen, J. Janhunen, J.-C. Dunat, and J. Gardey, "Database-assisted spectrum sharing in satellite communications: A survey," *IEEE Access*, vol. 5, pp. 25322–25341, Nov. 2017.

[8] M. A. Vazquez, L. Blanco, and A. I. Perez-Neira, "Spectrum sharing backhaul satellite-terrestrial systems via analog beamforming," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 270–281, May 2018.

[9] X. Zhang, B. Zhang, K. An, Z. Chen, S. Xie, H. Wang, L. Wang, and D. Guo, "Outage performance of NOMA-based cognitive hybrid satellite-terrestrial overlay networks by amplify-and-forward protocols," *IEEE Access*, vol. 7, pp. 85372–85381, Jun. 2019.

[10] P. K. Sharma, P. K. Upadhyay, D. B. da Costa, P. S. Bithas, and A. G. Kanatas, "Performance analysis of overlay spectrum sharing in hybrid satellite-terrestrial systems with secondary network selection," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6586–6601, Oct. 2017.

[11] Z. Li, F. Xiao, S. Wang, T. Pei, and J. Li, "Achievable rate maximization for cognitive hybrid satellite-terrestrial networks with AF-relays," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 2, pp. 304–313, Feb. 2018.

[12] Y. Yan, W. Yang, D. Guo, S. Li, H. Niu, and B. Zhang, "Robust secure beamforming and power splitting for millimeter-wave cognitive satellite–terrestrial networks with SWIPT," *IEEE Syst. J.*, pp. 1–12, 2020.

[13] K. An, M. Lin, W.-P. Zhu, Y. Huang, and G. Zheng, "Outage performance of cognitive hybrid satellite–terrestrial networks with interference constraint," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9397–9404, Nov. 2016.

[14] K. An, M. Lin, J. Ouyang, and W.-P. Zhu, "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025–3037, Nov. 2016.

[15] D. Nguyen, M. T. Nguyen, and L. B. Le, "Cognitive radio based resource allocation for sum rate maximization in dual satellite systems," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.

[16] E. Lagunas, S. K. Sharma, S. Maleki, S. Chatzinotas, and B. Ottersten, "Resource allocation for cognitive satellite communications with incumbent terrestrial networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 1, no. 3, pp. 305–317, Sep. 2015.

[17] E. Lagunas, S. Maleki, S. Chatzinotas, M. Soltanalian, A. I. Perez-Neira, and B. Oftersten, "Power and rate allocation in cognitive satellite uplink networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[18] S. Shi, G. Li, K. An, Z. Li, and G. Zheng, "Optimal power control for real-time applications in cognitive satellite terrestrial networks," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1815–1818, Apr. 2017.

[19] M. Höyhtyä, "Sharing FSS satellite C band with secondary small cells and D2D communications," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1606–1611.

[20] L. Kuang, X. Chen, C. Jiang, H. Zhang, and S. Wu, "Radio resource management in future terrestrial-satellite communication networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 81–87, Oct. 2017.

[21] C. Zhang, L. Kuang, and C. Jiang, "Spectrum sharing between geostationary and terrestrial communication systems," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Sydney, NSW, Australia, Jun. 2017, pp. 1–5.

[22] M. Höyhtyä, J. Ylitalo, X. Chen, and A. Mämmelä, "Use of databases for dynamic spectrum management in cognitive satellite systems," in *Cooperative and Cognitive Satellite Systems*. 2015, ch. 14, pp. 453–480.

[23] S. Nallagonda, A. Chandra, S. D. Roy, and S. Kundu, "Analytical performance of soft data fusion-aided spectrum sensing in hybrid terrestrial-satellite networks," *Int. J. Satell. Commun. Netw.*, vol. 35, no. 5, pp. 461–480, Sep. 2017.

[24] M. Jia, X. Liu, X. Gu, and Q. Guo, "Joint cooperative spectrum sensing and channel selection optimization for satellite communication systems based on cognitive radio," *Int. J. Satell. Commun. Netw.*, vol. 35, no. 2, pp. 139–150, Mar. 2017.

[25] C. Zhang, C. Jiang, J. Jin, S. Wu, L. Kuang, and S. Guo, "Spectrum sensing and recognition in satellite systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2502–2516, Mar. 2019.

[26] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao, and J. Wang, "Decision-theoretic distributed channel selection for opportunistic spectrum access: Strategies, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1689–1713, Jan. 2013.

[27] J. Wang, D. Guo, B. Zhang, L. Jia, and X. Tong, "Spectrum access and power control for cognitive satellite communications: A game-theoretical learning approach," *IEEE Access*, vol. 7, pp. 164216–164228, 2019.

[28] Y. Xu, C. Wang, J. Chen, J. Wang, Y. Xu, Q. Wu, and A. Anpalagan, "Load-aware dynamic spectrum access for small-cell networks: A graphical game approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8794–8800, Oct. 2016.

[29] W. Zhang, Y. Sun, L. Deng, C. K. Yeo, and L. Yang, "Dynamic spectrum allocation for heterogeneous cognitive radio networks with multiple channels," *IEEE Syst. J.*, vol. 13, no. 1, pp. 53–64, Mar. 2019.

[30] J. Zheng, Y. Cai, Y. Xu, and A. Anpalagan, "Distributed channel selection for interference mitigation in dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4757–4762, Nov. 2014.

[31] J. Zheng, Y. Cai, N. Lu, Y. Xu, and X. Shen, "Stochastic game-theoretic spectrum access in distributed and dynamic environment," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4807–4820, Oct. 2015.

[32] L. Jia, Y. Xu, Y. Sun, S. Feng, and A. Anpalagan, "Stackelberg game approaches for anti-jamming defence in wireless networks," *IEEE Wireless Commun.*, vol. 25, no. 6, pp. 120–128, Dec. 2018.

[33] L. Jia, Y. Xu, Y. Sun, S. Feng, L. Yu, and A. Anpalagan, "A game-theoretic learning approach for anti-jamming dynamic spectrum access in dense wireless networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1646–1656, Feb. 2019.

[34] C. Pan, Y. Cheng, Z. Yang, and Y. Zhang, "Dynamic opportunistic spectrum access with channel bonding in mesh networks: A game-theoretic approach," in *Proc. Int. Conf. Mach. Learn. Intell. Commun.* Springer, 2018, pp. 381–390.

[35] Y. Xu, Y. Xu, and A. Anpalagan, "Database-assisted spectrum access in dynamic networks: A distributed learning solution," *IEEE Access*, vol. 3, pp. 1071–1078, 2015.

[36] L. Kuang, C. Jiang, Y. Qian, and J. Lu, *Terrestrial-Satellite Communication Networks*. 2018.

[37] Z.-Q. Wang, X.-Y. Wan, and Z.-F. Fan, "Fair power control algorithm in cognitive radio networks based on stackelberg game," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E100.A, no. 8, pp. 1738–1741, Aug. 2017.

[38] Z. Chen, D. Guo, G. Ding, X. Tong, H. Wang, and X. Zhang, "Optimized power control scheme for global throughput of cognitive satellite-terrestrial networks based on non-cooperative game," *IEEE Access*, vol. 7, pp. 81652–81663, 2019.

[39] Y. Lu and A. Duel-Hallen, "A sensing contribution-based two-layer game for channel selection and spectrum access in cognitive radio ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3631–3640, Jun. 2018.

[40] A. V. Kordali and P. G. Cottis, "On spectrum trading for 5G cognitive spectrum sharing networks with hybrid access: A game-theoretic approach," *Wireless Pers. Commun.*, vol. 97, no. 4, pp. 6089–6109, Aug. 2017.

[41] H. Dai, Y. Huang, R. Zhao, J. Wang, and L. Yang, "Resource optimization for device-to-device and small cell uplink communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1187–1201, Feb. 2018.

[42] C. Fan, B. Li, C. Zhao, W. Guo, and Y.-C. Liang, "Learning-based spectrum sharing and spatial reuse in mm-wave ultradense networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4954–4968, Jun. 2018.

[43] Z. Wang, Y. Chen, F. Liu, Y. Xia, and X. Zhang, "Power system security under false data injection attacks with exploitation and exploration based on reinforcement learning," *IEEE Access*, vol. 6, pp. 48785–48796, 2018.

[44] C. Li, L. Cao, X. Liu, X. Chen, Z. Xu, and Y. Zhang, "A study of qualitative knowledge-based exploration for continuous deep reinforcement learning," *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 11, pp. 2721–2724, Nov. 2017.

[45] K. Kang, J. M. Park, H. W. Kim, T. C. Hong, B. J. Ku, and D.-I. Chang, "Analysis of interference and availability between satellite and ground components in an integrated mobile-satellite service system," *Int. J. Satell. Commun. Netw.*, vol. 33, no. 4, pp. 351–366, Jul. 2015.

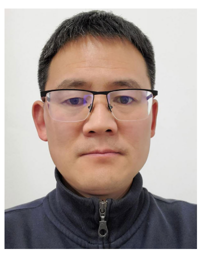
[46] H. Wang, A. Liu, X. Pan, and J. Li, "Optimization of power allocation for a multibeam satellite communication system with interbeam interference," *J. Appl. Math.*, vol. 2014, pp. 1–8, Nov. 2014.

JILI WANG received the B.S. degree from Tianjin University, Tianjin, China, in 2006, and the M.S. degree from Xinjiang University, Urumqi, China, in 2010. He is currently pursuing the Ph.D. degree with the Army Engineering University of PLA, Nanjing, China. His research interests include satellite communication, cognitive radio, game theory, and distributed optimization techniques for wireless communications.

BANGNING ZHANG (Member, IEEE) received the B.S. and M.S. degrees from the Institute of Communications Engineering (ICE), Nanjing, China, in 1984 and 1987, respectively. He is currently a Full Professor and the Head of the College of Communications Engineering. He has authored and coauthored more than 80 conference and journal articles and has been granted over 20 patents in his research areas. He has served as a reviewer for several journals in communication field. His current research interests include communication anti-jamming technologies, microwave technologies, satellite communications systems, cooperative communications, and physical layer security.

LULIANG JIA (Member, IEEE) received the M.S. and Ph.D. degrees in communications and information systems from the College of Communications Engineering, PLA University of Science and Technology, in 2014 and 2018, respectively. He is currently an Assistant Professor with the School of Space Information, Space Engineering University, Beijing, China. His research interests include game theory, learning theory, and communication anti-jamming technology.

BING ZHAO (Member, IEEE) received the B.S. and M.S. degrees from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2007 and 2009, respectively. She is currently a Full Lecturer with the Army Engineering University of PLA. She has been granted over ten patents in her research areas. Her current research interests include satellite communication systems and transmission technologies.



DAOXING GUO received the B.S., M.S., and Ph.D. degrees from the Institute of Communications Engineering (ICE), Nanjing, China, in 1995, 1999, and 2002, respectively. He is currently a Full Professor and a Ph.D. Supervisor with the PLA University of Science and Technology. He has authored and coauthored more than 40 conference and journal articles and has been granted over 20 patents in his research areas. He has served as a reviewer for several journals in communication field. His current research interests include satellite communications systems and transmission technologies, communication anti-jamming technologies, communication anti-interception technologies, including physical layer security, and so on.