IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 1, 2019, accepted August 15, 2019, date of publication September 2, 2019, date of current version September 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939000

Fuzzy Reinforcement Learning for Robust Spectrum Access in Dynamic Shared Networks

CHAOQIONG FAN[1], SHIJIAN BAO[2], YIWEN TAO[1], BIN LI[1], AND CHENGLIN ZHAO[1]

[1]School of Information and Communication Engineering (SICE), Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China
[2]China International Engineering Consulting Corporation (CIECC), Beijing 100048, China

Corresponding author: Bin Li (binli@bupt.edu.cn)

This work was supported in part by the Young Talents Invitation Program of the China Institute of Communications under Grant QT2017001, in part by the Natural Science Foundation of China under Grant U1805262, and in part by the BUPT Excellent Ph.D. Students Foundation under Grant CX2017209.

ABSTRACT The persistent increases of wireless terminals have brought about diverse shared networks, where robust and efficient spectrum reuse among heterogeneous users is of critical importance while still remains as a challenging task for practical application. In this paper, we study the problem of robust spectrum access (RSA) in a canonical wireless shared network (WSN) with fully considering the inherent dynamics of the wireless environment. The non-static features of WSNs result in uncertain channel state information (CSI) and complicated coupling interference, which can't be directly formulated as the well-accepted crisp game model, rendering most existing perfect CSI relied approaches inefficient or even unfeasible. To address this, by interpreting the estimated CSI with uncertainty as fuzzy number, a novel framework referred to as a non-cooperative fuzzy game (NC-FG) is adopted, whereby the user utility is mapped as a fuzzy value via the user-defined fuzzy utility function. Based on the derived property of the NC-FG that fuzzy Nash equilibrium (FNE) exists, a fuzzy-logic inspired reinforcement learning (FLRL) algorithm is proposed to achieve the FNE solutions of the constructed NC-FG to obtain the RSA in dynamic WSN, with which both the iterative learning and decision making procedures are implemented in a fuzzy-space, thus the sensitiveness of our scheme to environmental variations is alleviated. Finally, numerical simulations are provided to demonstrate the convergence, effectiveness, and superiority of our proposed FLRL algorithm in dynamic WSNs.

INDEX TERMS Dynamic wireless shared network, robust spectrum access, fuzzy space, non-cooperative fuzzy game, fuzzy-logic based reinforcement learning.

I. INTRODUCTION

As a crucial driven force of the next-generation communication system development [1], the explosive increases of data traffic which are resulted from the unprecedented growth of mobile devices and versatile applications, pose serious challenges to the limited spectrum resource. Consequently, the imminent spectrum shortage has produced a new impetus to seek practical solutions to improve the utilization efficiency of scarce spectrum resource in a shared manner. Therefore, the emerging 5G communication network [2], which is expected to possess the advantages of supporting a wide range of broadband accesses with higher data-rate as well as lower latency and providing ubiquitous connectivity for the advent of the Internet of Things (IoT) sector [3], continues to pursue advanced techniques of spectrum efficiency to fulfill the ever-increasing wireless service requirements.

In recent years, with sophisticated detection/estimation algorithms [4], [5] laying the foundation for more effective spectrum sharing schemes [6], [7], seamless spectrum shared access (SSA) through real-time channel perception and intelligent system adaption has become possible [8]. SSA can enhance the utilization efficiency of scarce spectrum resources by enabling spectrum reuse among diverse heterogeneous users and support the proliferating wireless service demands of dramatically growing mobile terminals, hence plays a paramount role in alleviating the spectrum scarcity and facilitating the deployment of 5G system [9].

The associate editor coordinating the review of this article and approving it for publication was Qilian Liang.

Triggered by SSA, wireless shared networks (WSNs) is likely to be the development trend of the next-generation network and has been attracted extensive attentions from both academic community and industrial practitioners [10]. The SSA technique based WSNs (such as small-cell networks and cognitive radio networks) which are usually installed in hot spot areas to provide wireless services for dense mobile users [11], maintain the favorable properties of low-power and high energy efficiency. Given more and more users are desired to be served with limited spectrum resource, the co-channel interference (CCI) [12], unless effectively eliminated, may become a major cause of users quality of service and system performance deterioration in dense WSNs. Therefore, efficient SSA schemes with the capacities of suppressing coupling interferences would become essential for WSNs.

SSA in diverse wireless communication networks has been extensively investigated [13]–[20]. With the mature cognitive radio (CR) concept, a clustering-based SSA scheme for multiuser orthogonal frequency division multiplexing (OFDM) CR network is studied in [13], and the authors first present an evolutionary game for joint spectrum sensing and access problem in [14], then considering the negative network externality as well as the sequential decision making structure of shared users, they propose a Bayesian social learning method for multi-channel sensing and access problem in [15]. For 5G heterogenous network, a comprehensive survey of advanced techniques for SSA is addressed in [16], and [17] specifically presents a distributed inter-cell interference coordination (ICIC) accelerated learning algorithm for SSA in LTE cellular systems, while [18] concentrates on synergistic SSA with a software defined network (SDN)-enabled approach. For emerging cloud/fog radio access networks, the authors develop a D2D-enabled distributed approach for spectral efficiency improvement in [19], and propose a deep reinforcement learning method for mode selection and resource management in [20]. Besides, some researches about SSA problem with varying environmental information have been conducted [21]–[23]. Based on a crisp game approach, [21] investigates the problem of SSA in time-varying environments with the help of the expectation and other-order moments of the channel capacity. The underlay spectrum sharing between the dynamic drone network and the traditional cellular network is studied in [22]. A user-centric paradigm for SSA under uncertain traffic model and dynamic network architecture is analyzed in [23].

Although SSA has been widely reported in the existing researches for its great promise in relieving the tension of spectrum resource requirements to support wireless service, its implementation for future application is not perfectly addressed. Since those approaches are generally carried out in the scenarios of static or quasi-static environments, i.e. the channel state information (CSI) and the mutual interference remain unchanged in scheduling slots. However, taking practical applications into account, the stochastic and dynamic features of the emerging 5G WSNs would pose some new formidable challenges to the SSA, and renders the performance of existing SSA schemes hard to be harvested.

To be specific, first, the varying network topology and the uncertain CSI make the channel quality evaluation tend to be inaccurate, which may result in a non-optimal channel selection. More importantly, the dynamic environmental conditions would severely deteriorate the stability and convergence of the previous learning algorithms [24]. Therefore, those existing methods relying on the ideal assumptions (i.e. definite network topology and perfect estimated CSI), though may lead to mathematical tractability, would be inadequate even infeasible in the realistic dynamic wireless environment. Furthermore, for the works concerning environmental variation, there still remains some deficiencies for practical applications, since the crisp game theory is not generally enough to formulate the SSA with dynamic and uncertain information. As such, a robust spectrum access (RSA) scheme for WSNs which possesses the capability of combating stochastic features of wireless environment should be further explored.

To address the requirement for a RSA scheme in the dynamic WSNs, in this paper, we fully consider the inherent variations and uncertainties of wireless environment. Rather than assuming the desirable/interference power gains are static and adopting the crisp game to describe the spectrum sharing competition, we model the uncertain CSI as fuzzy numbers and then introduce a novel non-cooperative fuzzy game (NC-FG) to formulate the RSA problem. On this basis, we propose a fuzzy-logic inspired reinforcement learning [25] (FLRL) algorithm to achieve the equilibrium solution of our formulated NC-FG, with which the decision is made in an *uncertainty immune* space, i.e. a fuzzy space, thereby the vulnerability of environmental variations can be remedied. Specific contributions of this paper are listed as follows.

(1) We formulate an optimization problem of RSA in a dynamic WSN with maximizing the system capacities. Given the stochastic features of the realistic wireless environment, the network topology and the CSI will become uncertain. By taking the heterogeneous interference structure into consideration, the objective is obtaining the robust and optimal spectrum access patten with reliable transmission assurance and imperfect knowledge constraint.
(2) We develop a NC-FG to characterize the formulated optimization problem with the dynamic and uncertain information restriction, and study its property to demonstrate the existence of fuzzy Nash equilibrium (FNE). Assisted by the appealing fuzzy logic, we introduce a fuzzy-logic space, in which the changing and uncertain CSI is interpreted as a fuzzy number and the fuzzy utility function (FUF) is defined as the fluctuated user data-rate.
(3) We propose a robust FLRL algorithm to achieve the equilibrium solution of the NC-FG to optimize the network performance, i.e. the overall system capacities. For the

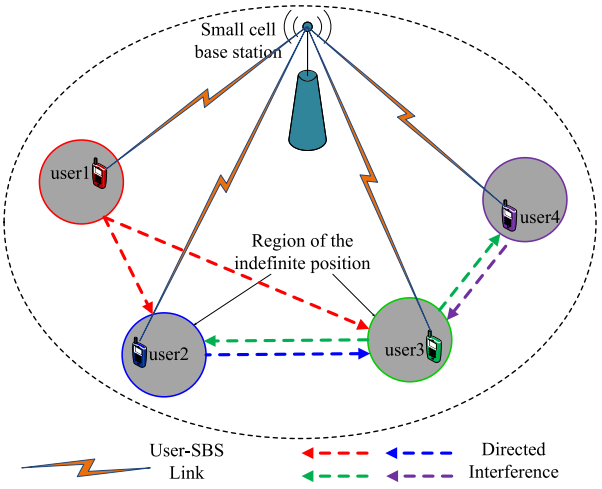


FIGURE 1. System model of the considered dynamic small-cell network.

designed NC-FG involving fuzzy component, we first apply fuzzy-logic computations to obtain a fuzzy preference relation (FPR), and then adopt it to calculate the priority vector of users. On this basis, the network manager can implement a robust decision making to achieve a FNE in the derived fuzzy space. Therefore, our scheme is innately resistant to the environmental uncertainties, and thereby can fulfill the optimal RSA.

The rest of this paper is structured as follows. We present the system model and the problem formulation in Section II. In Section III, premised on the preliminaries of fuzzy set theory, we formulate a NC-FG, and investigate its property. The FLRL algorithm for RSA in dynamic WSNs is proposed in Section IV, and its performances are demonstrated via numerical simulations in Section V. Finally, we make the conclusion of our work in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. DYNAMIC NETWORK MODEL

We study the RSA problem in a 5G small-cell network with fully considering the inherent dynamics and uncertainties of wireless environment. Assume that there are N shared users and M available channels in the cellular network, then for the purpose of revealing spectrum resource sharing and competition among users, without loss of generality we set $N > M$. For presentation, denote the set of shared users as $\mathcal{N}$, i.e. $\mathcal{N} = \{1, 2, \ldots, N\}$, and the set of available channel resource as $\mathcal{M}$, i.e. $\mathcal{M} = \{1, 2, \ldots, M\}$.

As shown in Fig. 1, a specific factor causing the non-negligible uncertainties in a small-cell network is the indeterminate positions of users [26], which may result from their stochastic movements. To comprehensively describe this indetermination and remedy the incomplete position information, for an arbitrary user $n \in \mathcal{N}$, we assume the unknown exact location $\boldsymbol{l}_n = \{x_n, y_n\}$ lies in a circular area with the imperfect estimated location $\hat{\boldsymbol{l}}_n = \{\hat{x}_n, \hat{y}_n\}$ as center and φ_n as radius, i.e.,

$$\boldsymbol{l}_n = \hat{\boldsymbol{l}}_n + \Delta\boldsymbol{l}_n, \tag{1}$$

$$\Delta\boldsymbol{l}_n \in \mathcal{L}_n = \{\Delta x_n^2 + \Delta y_n^2 \leqslant \varphi_n^2\}, \tag{2}$$

where $\Delta\boldsymbol{l}_n = \{\Delta x_n, \Delta y_n\}$ is the estimation error, and $\mathcal{L}_n$ is the uncertain region. Notably, due to the stochastic mobility of users, the uncertain region $\mathcal{L}_n$ may be irregular, but there is always a circle containing this irregular area. Thus, the formulated uncertain circular region $\mathcal{L}_n$ actually provides a conservative estimation for other irregular forms.

Clearly, the distance $d_{n,S}$ between user n and the small-cell base station (SBS) as well as the distance $d_{n,n'}$ between user n and user n' are indefinite, which are given by:

$$d_{n,S} = |\boldsymbol{l}_n - \boldsymbol{l}_S| = |(\hat{\boldsymbol{l}}_n + \Delta\boldsymbol{l}_n) - \boldsymbol{l}_S|, \tag{3}$$

$$d_{n,n'} = |\boldsymbol{l}_n - \boldsymbol{l}_{n'}| = |(\hat{\boldsymbol{l}}_n + \Delta\boldsymbol{l}_n) - (\hat{\boldsymbol{l}}_{n'} + \Delta\boldsymbol{l}_{n'})|, \tag{4}$$

respectively, where $|\bullet|$ represents the Euclidean norm, and $\boldsymbol{l}_S$ is the location of the SBS.

Here, the free space path-loss (PL) model with rayleigh fading [27] is adopted to describe the propagation of signal power gain, then the desirable (interference) power gain $h_{n,S}^m$ ($h_{n,n'}^m$) between node n and the SBS (another node n') on channel m can be given by:

$$h_{n,*}^m = \begin{cases} h_{n,S}^m = d_{n,S}^{-\alpha_m} \times \vartheta_m = |\boldsymbol{l}_n - \boldsymbol{l}_S|^{-\alpha_m} \times \vartheta_m, \\ h_{n,n'}^m = d_{n,n'}^{-\alpha_m} \times \vartheta_m = |\boldsymbol{l}_n - \boldsymbol{l}_{n'}|^{-\alpha_m} \times \vartheta_m, \end{cases} \tag{5}$$

where α_m is the PL exponent, and ϑ_m is the instantaneous random component of the PL on channel m ($m \in \mathcal{M}$).

It can be observed that the location uncertainty is finally transformed to the channel state. Therefore, the assumption of the position uncertainty can be understood from different perspectives. Concretely, on the one hand, it initially characterizes the unideal position estimation. On the other hand, it also can reflect the transmission/interference link uncertainty causing by the dynamic wireless environment.

Hence, we can redescribe the uncertainties of the dynamic WSNs in the aspect of CSI, which can be expressed as:

$$h_{n,*}^m = \{\hat{h}_{n,*}^m + \Delta h_{n,*}^m : |\Delta h_{n,*}^m| < \rho_{n,*}^m\}, \tag{6}$$

where $\hat{h}_{n,*}^m$ is the imperfect estimation of the power gain $h_{n,*}^m$ [28], and $\Delta h_{n,*}^m$ is the uncertain estimation error, which is limited by a boundary $\rho_{n,*}^m$ of the uncertainty.

B. DIRECTED INTERFERENCE MODEL

Considering the finite emission power of users, the transmission signal of one specific user only affects the users who are located in its interference region. Denote the transmission power of node n as P_n, and the interference range A_n of node n is defined as the range within which the received signal power from node n is higher than a threshold P_{th}, i.e.,

$$A_n = \max_{n' \in \mathcal{N}}\{d_{n,n'} : P_n h_{n,n'}^m \geqslant P_{\text{th}}\}. \tag{7}$$

Due to the diverse communication demands, the transmission powers of users are different, which leads to a heterogeneous interference range. That is to say, if node n' lies in

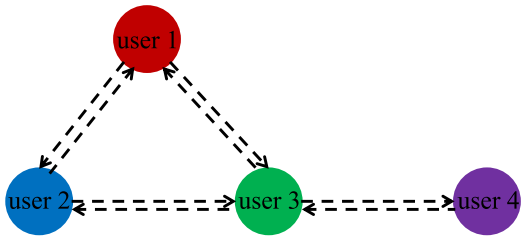


FIGURE 2. The indefinite interference graph.

the interference area of node n, whereas the converse may not hold. Therefore, we introduce a directed interference model, where two nodes n and n' are connected by a directed line from n to n', denoted by $n \rightarrow n'$, if the distance $d_{n,n'}$ is less than A_n, which means n' suffers from the interference of node n. In this case, the edge set of the directed interference graph is given by:

$$\mathcal{E} = \{(n \rightarrow n') : d_{n,n'} \leqslant A_n\}, \tag{8}$$

and the interference set of user n is denoted as:

$$\mathcal{I}_n = \{n' \in \mathcal{N} : (n' \rightarrow n) \in \mathcal{E}\}. \tag{9}$$

Note that the distance $d_{n,n'}$ would become changeable and uncertain, when we take the realistic dynamics and the incomplete information of users' position into consideration. As such, both the directed edge set $\mathcal{E}$ and the interference set $\mathcal{I}_n$ tend to be varying and indeterminate. For the network model shown in Fig. 1, the corresponding indefinite directed interference graph is drawn in Fig. 2, in which the dotted lines indicate the uncertain potential interference among users.

C. PROBLEM FORMULATION

Let s_n ($s_n \in \mathcal{S}_n = \mathcal{M}$) denote the channel selection of user n, $\mathcal{B}_m$ denote the set of users who select channel m for competition, i.e. $\mathcal{B}_m = \{n \in \mathcal{N} : s_n = m\}$, and the number of these users is represented as β_m, i.e. $\beta_m = \|\mathcal{B}_m\|$, where $\|\mathcal{X}\|$ indicates the number of elements in the set $\mathcal{X}$. Then with channel reuse in a WSN, the CCI I_n of user n is given by:

$$I_n = \sum\nolimits_{i \in \mathcal{I}_n \cap \mathcal{B}_{s_n}} P_i h_{n,i}^{s_n}. \tag{10}$$

To ensure the reliable communication quality, the suffered CCI of nodes should be below an interference threshold I_{th}, and a transmission of user n is successful if I_n is no more than the predefined I_{th}, i.e. $I_n \leqslant I_{\text{th}}$.

Then the instantaneous data-rate R_n of node n accessing channel m can be expressed as:

$$R_n = \theta_n B \log(1 + \frac{P_n h_{n,S}^m}{I_n + \sigma^2}), \tag{11}$$

where B represents the bandwidth of channel, σ^2 is the variance of the additive white Gauss noise (AWGN), and θ_n is a Bernoulli random variable indicating whether the channel competition of user n is successful or not, i.e.,

$$\theta_n = \begin{cases} 1, & I_n \leqslant I_{\text{th}}, \\ 0, & I_n > I_{\text{th}}. \end{cases} \qquad n \in \mathcal{N}$$

In addition, the total capacities C_m of channel m can be defined as the sum of the data-rate of users who choose channel m for transmission, i.e.,

$$C_m = \sum\nolimits_{n \in \mathcal{B}_m} R_n. \tag{12}$$

Based on the above descriptions, the aggregate network throughput, which can be calculated from the perspective of user data-rate or from the perspective of channel capacity, is given by:

$$U(s) = \sum\nolimits_{n \in \mathcal{N}} R_n = \sum\nolimits_{m \in \mathcal{M}} C_m, \tag{13}$$

where $s = \{s_1, s_2, \ldots, s_N\}$ is the channel selection pattern of all users.

To this point, the main purpose of this paper is to find a reliable and optimal spectrum access profile to maximize the aggregate network throughput, i.e.,

$$\mathcal{P}1 : s^* = \arg\max U(s). \tag{14}$$

The RSA problem in a dynamic WSN is formulated as eq. (14), which, intuitively, is an extremely challenging task for its intrinsic NP-hard nature, and more importantly for the concerned uncertain environmental information as well as non-static wireless network restrictions. On this basis, an effective learning scheme for RSA, one that can efficiently combat the uncertain wireless environment and steadily achieve the optimum solution is essential to be developed.

III. FUZZY GAME FOR RSA

Due to the definite information dependence, the conventional crisp-game only captures the channel competition problem with ideal environment assumptions (i.e. the CSI and the interference relationship remain unchanged), hence would lose effectiveness for the application in the considered realistic dynamic WSNs with uncertainties. Instead of adopting the crisp-game where all elements are definite, in this section, a game involving fuzzy factor-fuzzy game-is formulated to characterize the RSA problem with dynamic and uncertain information. Specifically, we first summarize some related definitions and notions of the fuzzy set theory [29]. Then, by projecting the uncertain CSI as fuzzy number, we establish a fuzzy space, in which the data-rate serves as a FUF of players. On this basis, a NC-FG [30] is developed, and its properties are investigated. Note that, thanks to the capabilities of formulating the spectrum access problem with uncertain information and combating the unpredictable stochastic features of WSNs, the fuzzy game based scheme is naturally more appropriate for the uncertain scenarios.

A. FUZZY SET THEORY

Definition 1 (Fuzzy Number): A real fuzzy number $\tilde{a}$ is precisely described as any fuzzy subset on the space of real numbers $\mathbb{R}$, whose membership function $\mu_{\tilde{a}}(x)$ satisfies the following conditions:

- $\mu_{\tilde{a}}(x)$ is a continuous mapping from $\mathbb{R}$ to the closed interval [0, 1].

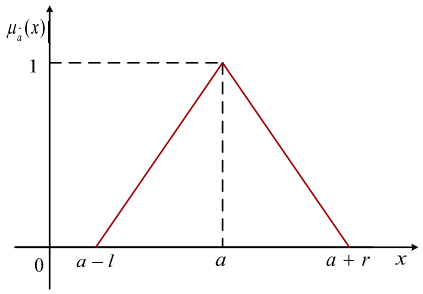


FIGURE 3. The membership function of TriFN $\tilde{a}$.

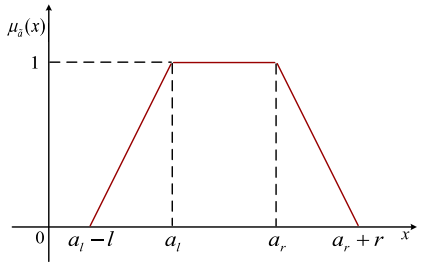


FIGURE 4. The membership function of TraFN $\tilde{a}$.

- $\mu_{\tilde{a}}(x)$ is constant on $[-\infty, a_l - l] \cup [a_r + r, +\infty]$ and $[a_l, a_r]$. Specifically, $\mu_{\tilde{a}}(x) = 0, \forall x \in [-\infty, a_l - l] \cup [a_r + r, +\infty]$ and $\mu_{\tilde{a}}(x) = 1, \forall x \in [a_l, a_r]$.
- $\mu_{\tilde{a}}(x)$ is strictly increasing and continuous over $[a_l - l, a_l]$, and strictly decreasing and continuous over $[a_r, a_r + r]$.

Here a_l, a_r, l and r are all real numbers, satisfying $a_l \leqslant a_r$, $l, r > 0$.

The membership function $\mu_{\tilde{a}}(x)$ presents a quantitative description of the fuzzy number $\tilde{a}$, which is a basic concept of fuzzy mathematics. A general membership function can be given by the following equation, i.e.,

$$\mu_{\tilde{a}}(x) = \begin{cases} \dfrac{x - a_l + l}{l}, & x \in [a_l - l, a_l], \\ \dfrac{a_r + r - x}{r}, & x \in [a_r, a_r + r], \\ 1, & x \in [a_l, a_r], \\ 0, & \text{else}. \end{cases} \tag{15}$$

Note that, when $a_l = a_r$ holds, eq. (15) presents a membership function of triangular fuzzy number (TriFN) i.e. $\tilde{a} = (a-l, a, a+r)$, which is illustrated in Fig. 3; and when $a_l < a_r$ holds, eq. (15) shows a membership function of trapezoid fuzzy number (TraFN) i.e. $\tilde{a} = (a_l - l, a_l, a_r, a_r + r)$, which is drawn in Fig. 4.

Here, we take a TriFN as an example, i.e. $a_l = a_r$, the operations of fuzzy numbers obey the following lemma.

Lemma 1: Let $\tilde{a}_1 = (a_1 - l_1, a_1, a_1 + r_1)$, $\tilde{a}_2 = (a_2 - l_2, a_2, a_2 + r_2)$ represent TriFNs, and v is a real number. It holds that:

- $\tilde{a}_1 + \tilde{a}_2 = (a_1 + a_2 - l_1 - l_2, a_1 + a_2, a_1 + a_2 + r_1 + r_2)$;
- $v\tilde{a}_1 = \Big(v(a_1 - l_1), va_1, v(a_1 + r_1)\Big)$;
- $\tilde{a}_2$ dominates $\tilde{a}_1$ (denoted by $\tilde{a}_2 \gtrapprox \tilde{a}_1$) if and only if $\max\{l_2 - l_1, 0\} \leqslant a_2 - a_1$ and $\max\{r_1 - r_2, 0\} \leqslant a_2 - a_1$.

It is noted that due to the ambiguous numeric values involving, ranking the fuzzy numbers just according to their magnitudes tends to be difficult. Therefore, multiple ranking approaches have been researched to fully explore the ambiguous characteristics of fuzzy numbers. Here, we adopt the method proposed in [31]. To facilitate the description, some relative conceptions and definitions are introduced as follows.

Definition 2 (Satisfaction Function): The SF between two fuzzy number $\tilde{a}$ and $\tilde{b}$ is defined as:

$$SF(\tilde{a} < \tilde{b}) = \frac{\int_{-\infty}^{+\infty} \int_{-\infty}^{y} \mu_{\tilde{a}}(x) \times \mu_{\tilde{b}}(y) dxdy}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mu_{\tilde{a}}(x) \times \mu_{\tilde{b}}(y) dxdy}, \tag{16a}$$

$$SF(\tilde{a} > \tilde{b}) = \frac{\int_{-\infty}^{+\infty} \int_{y}^{+\infty} \mu_{\tilde{a}}(x) \times \mu_{\tilde{b}}(y) dxdy}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mu_{\tilde{a}}(x) \times \mu_{\tilde{b}}(y) dxdy}, \tag{16b}$$

where $SF(\tilde{a} < \tilde{b})$ represents the possibility that $\tilde{a}$ is smaller than $\tilde{b}$. Similarly, $SF(\tilde{a} > \tilde{b})$ represents the possibility that $\tilde{a}$ is larger than $\tilde{b}$.

Definition 3 (Viewpoint): For a fuzzy number $\tilde{a}$, a fuzzy number $\tilde{b}$ which satisfies the following conditions is a viewpoint:

- $\sup(\tilde{a}) \subseteq \sup(\tilde{b})$, where $\sup(\tilde{a}) = \{x | \mu_{\tilde{a}}(x) \neq 0\}$;
- $\int_{-\infty}^{+\infty} \mu_{\tilde{b}}(x) \mathrm{d}x$ exists and it is not zero.

Without loss of generality, the defined viewpoint can be broadly divided into three categories: *optimistic*, *neutral* and *pessimistic*, with which the fuzzy numbers can be evaluated. The second condition is added to make sure that a viewpoint can be applicable to the SF.

Definition 4 (Evaluation Value): Based on the above defined SF and viewpoint, the evaluation value of the fuzzy number $\tilde{a}$ in the viewpoint $\tilde{b}$, $E_{\tilde{b}}(\tilde{a})$ can be given by:

$$E_{\tilde{b}}(\tilde{a}) = SF(\tilde{a} > \tilde{b}). \tag{17}$$

Definition 5 (Relative Index): The relative index of the fuzzy number $\tilde{a}$ in the viewpoint $\tilde{b}$, $T_{\tilde{b}}(\tilde{a})$, which shows how close $\tilde{a}$ is to the one having the best evaluation in viewpoint $\tilde{b}$, is defined as:

$$T_{\tilde{b}}(\tilde{a}) = \frac{E_{\tilde{b}}(\tilde{a})}{\max_{\tilde{a} \in \tilde{\mathcal{A}}} E_{\tilde{b}}(\tilde{a})}, \tag{18}$$

where $\tilde{\mathcal{A}}$ is the set of the sorted fuzzy numbers.

B. NON-COOPERATIVE FUZZY GAME

As shown by the previous explanations and subsequent simulations, a non-cooperative crisp game $\mathcal{G} \triangleq \Big(\mathcal{N}, \mathcal{S}, \mathcal{U}\Big)$, where $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of players (users), $\mathcal{S} = \otimes \mathcal{S}_n = \mathcal{M}^N$ is the set of strategy space of the game $\mathcal{G}$, and $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$ is the set of utility functions

for the players, is inadequate to characterize the spectrum sharing problem with dynamic and uncertain information for its variation-susceptible feature. Therefore, by extending the utility component $\mathcal{U}$ to a fuzzy set, we resort a NC-FG to reformulate the challenging problem of RSA in the dynamic WSN with uncertain environmental information.

Specifically, based on the above elaborated fuzzy set theory, we build a fuzzy space to map the uncertain link information, in which a fuzzy number $\tilde{h}^m_{n,*}$ is employed to describe the stochastic and uncertain CSI, i.e.,

$$h^m_{n,*} \to \tilde{h}^m_{n,*} = (\hat{h}^m_{n,*} - \Delta h^{m,\mathrm{l}}_{n,*}, \hat{h}^m_{n,*}, \hat{h}^m_{n,*} + \Delta h^{m,\mathrm{l}}_{n,*}), \tag{19}$$

where $\Delta h^{m,\mathrm{l}}_{n,*}$ and $\Delta h^{m,\mathrm{r}}_{n,*}$ are the left deviation and the right deviation of the fuzzy number $\tilde{h}^m_{n,*}$, indicating a random fluctuation of the uncertain channel gain.

As stated before, in our constructed NC-FG, the players set $\mathcal{N}$ and the strategy profiles $\mathcal{S}$ are definite, while the utility of each player which is related with the mapped fuzzy number $\tilde{h}^m_{n,*}$ is accordingly a fuzzy number. The definition of the FUF is presented as follows.

Definition 6 (Fuzzy Utility Function): The FUF of each player is defined as the uncertain instantaneous data-rate $\tilde{R}_n$ of user n, which is mapped by the fuzzy number $\tilde{h}^m_{n,S}$, i.e.,

$$\tilde{u}_n(s_n, s_{-n}, \tilde{h}^{s_n}_{n,S}) \triangleq \tilde{R}_n. \tag{20}$$

With the formulated FUF, the mathematical description of the NC-FG [32] is provided as follows.

Definition 7 (Non-Cooperative Fuzzy Game): The NC-FG is defined as:

$$\tilde{\mathcal{G}} \triangleq \left(\mathcal{N}, \mathcal{S}, \tilde{\mathcal{U}}(s, \tilde{\boldsymbol{h}})\right), \tag{21}$$

where $\mathcal{N}$ and $\mathcal{S}$ are identical with that in the crisp game $\mathcal{G}$, $\tilde{\mathcal{U}}(s, \tilde{\boldsymbol{h}}) = \{\tilde{u}_n(s, \tilde{h}^{s_n}_{n,S}) | n \in \mathcal{N}\}$ is the FUF set, and $\tilde{\boldsymbol{h}} = \{\tilde{h}^{s_n}_{n,S} | n \in \mathcal{N}\}$ is the vector of the uncertain CSI which is modeled by fuzzy number in the fuzzy space.

Based on the above analysis, the problem on eq. (14) can be reformulated as a NC-FG, in which the players attempt to attain a reliable and appropriate spectrum reuse pattern to maximize their fuzzy utility, i.e.,

$$\mathcal{P}2 : s^* = \arg\max_{s_n \in \mathcal{S}_n} \tilde{u}_n(s_n, s_{-n}, \tilde{h}^{s_n}_{n,S}), \ \forall n \in \mathcal{N}. \tag{22}$$

C. ANALYSIS OF FUZZY NASH EQUILIBRIUM

It is known to all that Nash Equilibrium (NE) is a stable solution to non-cooperative crisp games. Similarly, a stable solution to non-cooperative fuzzy games is called fuzzy Nash equilibrium (FNE) [33], which is defined as follows.

Definition 8 (Fuzzy Nash Equilibrium): A strategy pattern $s^* \in \mathcal{S}$ is called a FNE of the fuzzy game $\tilde{\mathcal{G}}$ if,

$$\tilde{u}_n(s^*_n, s^*_{-n}, \tilde{h}^{s^*_n}_{n,S}) \gtrapprox \tilde{u}_n(s_n, s^*_{-n}, \tilde{h}^{s_n}_{n,S}), \quad \forall n \in \mathcal{N}, \quad \forall s_n, s^*_n \in \mathcal{S}_n, \ s^*_{-n} \in \mathcal{S}_{-n}. \tag{23}$$

For our formulated fuzzy game, the following theorem is presented to validate the existence of a FNE.

Theorem 1: There exists a FNE solution for the formulated NC-FG in eq. (21).

Proof: In order to demonstrate the existence of the FNE, we first present the definition and the property of a fuzzy bi-matrix game.

Definition 9 (Fuzzy Bi-matrix Game): A fuzzy bi-matrix game $\tilde{\mathcal{G}}_B$ is defined as a bi-matrix game, which involves two players with fuzzy utilities [34], i.e.,

$$\tilde{\mathcal{G}}_B = \left(\mathrm{I}, \mathrm{II}, \mathcal{S}_\mathrm{I}, \mathcal{S}_\mathrm{II}, \tilde{\boldsymbol{u}}_\mathrm{I}, \tilde{\boldsymbol{u}}_\mathrm{II}\right), \tag{24}$$

where $\mathcal{S}_\mathrm{I}$ and $\mathcal{S}_\mathrm{II}$ are the strategies sets of Player I and Player II, respectively.

$$\tilde{\boldsymbol{u}}_n = \begin{bmatrix} \tilde{u}_{11} & \tilde{u}_{12} & \cdots & \tilde{u}_{1M} \\ \tilde{u}_{21} & \tilde{u}_{22} & \cdots & \tilde{u}_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{u}_{M1} & \tilde{u}_{M2} & \cdots & \tilde{u}_{MM} \end{bmatrix}_{M \times M} \quad n = \mathrm{I}, \mathrm{II}, \tag{25}$$

is the utilities matrix of the players. Each element $\tilde{u}_{m,m'}$ denotes the obtained fuzzy utility, when Player I adopts the strategy m while Player II adopts the strategy m'.

One key property of the fuzzy bi-matrix game is characterized by the following lemma.

Lemma 2: A fuzzy bi-matrix game has at least one FNE solution, if there exists a subset $\mathcal{N}_0 \subset \mathcal{N}$ such that the function $\sum_{n \in \mathcal{N}_0} \tilde{u}_n(s_n, s_{-n}, \tilde{h}^m_{n,S})$ is convex on $\mu_{\tilde{h}^m_{n,S}}(x)$ [35].

The attractive feature of fuzzy bi-matrix games provides a perspective to analyze the formulated NC-FG, i.e. exploiting mathematical induction (MI) to resolve it, with which the existence of FNE can be ensured.

The most fundamental situation in which the fuzzy game consists of two players is firstly investigated.

Theorem 2: There exists at least one FNE solution for the fuzzy game $\tilde{\mathcal{G}}_2$ with two players.

Proof: For the first constraint in **Lemma 2**, intuitively,

$$\tilde{\mathcal{G}}_2 = \left(1, 2, \mathcal{S}_1, \mathcal{S}_2, \tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{u}}_2\right), \tag{26}$$

is a fuzzy bi-matrix game.

For the second constraint, here, we choose $\mathcal{N}_0 = \{1\}$, then we have

$$\sum\nolimits_{n \in \mathcal{N}_0} \tilde{u}_n(s_n, s_{-n}, \tilde{h}^m_{n,S}) = \tilde{u}_1(s_1, s_2, \tilde{h}^{s_1}_{1,S}). \tag{27}$$

To analyze the concave-convex property of the function $\tilde{u}_1(s_1, s_2, \tilde{h}^{s_1}_{1,S})$, its second derivative $\tilde{u}''_1$ is given by:

$$\begin{aligned} \tilde{u}''_1 &= \frac{\mathrm{d}^2 \tilde{u}_1(s_1, s_2, \tilde{h}^{s_1}_{1,S})}{\mathrm{d}(\tilde{h}^{s_1}_{1,S})^2} \\ &= -\frac{\theta_1 B P_1^2}{\ln 2} \times \frac{1}{(I_1 + \sigma^2 + P_1 \tilde{h}^{s_1}_{1,S})^2} \leqslant 0. \end{aligned} \tag{28}$$

Therefore, the function $\sum_{n \in \mathcal{N}_0} \tilde{u}_n(s_n, s_{-n}, \tilde{h}^m_{n,S})$ is convex on $\mu_{\tilde{h}^m_{n,S}}(x)$.

Based on **Lemma 2** and the above elaborations, **Theorem 2** can be proved. □

Moving forward, we study a more general case and present the following theorem.

Theorem 3: Assuming that the fuzzy game $\tilde{\mathcal{G}}_N$ with N players has FNE solutions, then there exists at least one FNE solution for the fuzzy game $\tilde{\mathcal{G}}_{N+1}$ with $N+1$ players.

Proof: Based on the assumption that the FNE of the fuzzy game $\tilde{\mathcal{G}}_N$ exists, the players set $\mathcal{N}$ can be regarded as a whole unity. Denote the $N+1$th player as n_0, then the fuzzy game $\tilde{\mathcal{G}}_{N+1}$ can be expressed as:

$$\tilde{\mathcal{G}}_{N+1} = \left(n_0, \mathcal{N}, \mathcal{S}_{n_0}, \mathcal{S}_{\mathcal{N}}, \tilde{\boldsymbol{u}}_{n_0}, \tilde{\boldsymbol{u}}_{\mathcal{N}}\right). \tag{29}$$

Therefore, the fuzzy game $\tilde{\mathcal{G}}_{N+1}$ clearly can be treated as a fuzzy bi-matrix game, in which n_0 acts as Player I, and $\mathcal{N}$ acts as Player II.

After establishing the $\tilde{\mathcal{G}}_{N+1}$ as a fuzzy bi-matrix game, we analyze the concave-convex property $\tilde{u}_{n_0}$ and $\tilde{u}_{\mathcal{N}}$.

For the utility function $\tilde{u}_{n_0}(s_{n_0}, s_{\mathcal{N}}, \tilde{h}_{n_0,S}^{s_{n_0}})$ of Player I, let $\mathcal{N}_0 = \{n_0\}$, by performing the same steps in **Theorem 2**, the following equation can be obtained, i.e.,

$$\tilde{u}''_{n_0} = -\frac{\theta_{n_0} B P_{n_0}^2}{\ln 2} \times \frac{1}{(I_{n_0} + \sigma^2 + P_{n_0} \tilde{h}_{n_0,S}^{s_{n_0}})^2} \leqslant 0, \tag{30}$$

which demonstrates the convexity of $\tilde{u}_{n_0}$ on $\mu_{\tilde{h}_{n_0,S}^{s_{n_0}}}(x)$.

The utility function $\tilde{u}_{\mathcal{N}}$ of Player II is the sum utilities of all player n, $n \in \mathcal{N}$, i.e.,

$$\tilde{u}_{\mathcal{N}} = \sum\nolimits_{n \in \mathcal{N}} \tilde{u}_n(s_n, s_{-n}, \tilde{h}_{n,S}^{s_n}). \tag{31}$$

Obviously, $\tilde{u}_{\mathcal{N}}$ is a convex function on $\mu_{\tilde{\boldsymbol{h}}}(x)$, since $\tilde{u}_n(s_n, s_{-n}, \tilde{h}_{n,S}^{s_n})$ is convex on $\mu_{\tilde{h}_{n,S}^{s_n}}(x)$, $n \in \mathcal{N}$. Recall that $\tilde{\boldsymbol{h}}$ denotes the vector of the uncertain CSI.

On the basis of **Lemma 2** and the above analysis, **Theorem 3** can be proved. □

Finally, combining **Theorem 2** and **Theorem 3**, **Theorem 1** can be proved. □

With the above favorable property of our formulated NC-FG, we attempt to achieve the FNE solution to optimize the network performance with dynamic information constraint. It is noted that the procedure of finding the equilibrium pattern of a fuzzy game is far more different from that of a crisp game, which involves fuzzy number analysis and processing, rendering most existing learning methods no longer applicable. Therefore, a robust learning algorithm which can cope with the fuzzy parameters and achieve the optimal spectrum reuse in dynamic WSNs is required to be designed.

IV. FUZZY-LOGIC INSPIRED REINFORCEMENT LEARNING ALGORITHM

To achieve the FNE of our formulated NC-FG, in this section, we introduce a robust FLRL algorithm for the reliable transmission ensured RSA problem with uncertain information, and then demonstrate its convergence performance.

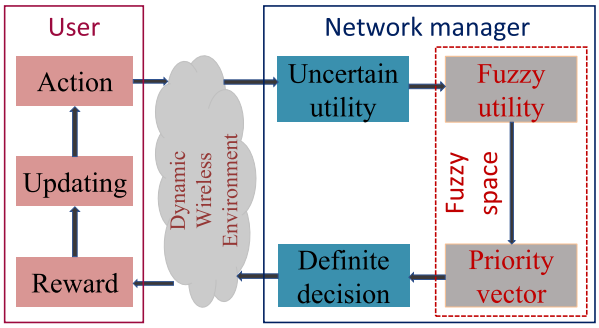


FIGURE 5. The framework of the FLRL algorithm.

A. DESCRIPTION OF THE FLRL ALGORITHM

In the considered future 5G WSNs, due to the random movement of nodes and the complex changing of environments, the link quality and the coupling interferences would become time-varying and uncertain, which can not be adopted directly as a metric to optimize the network performance. Consequently, existing crisp-game based learning methods, which rely on the definite utilities to make decisions and update strategies, would lose effectiveness to ensure convergence for lacking the ability to confront the encountered uncertainties.

To address the above issue, we propose a FLRL algorithm, whose key idea is to exploit the fuzzy space to handle the uncertain information. The algorithm framework is shown in Fig. 5, which involves interactions of users, network manager and dynamic wireless environment. Specifically,

(1) users take their channel selection actions in the dynamic wireless environment;
(2) the network manager acquires the uncertain utilities of the shared users and represents them as fuzzy utilities in the projected space. Assisted by the fuzzy logic, a priority vector is derived, with which the definite spectrum sharing decision can be obtained;
(3) users receive their reward based on the decision pattern, and adjust the channel selection probabilities accordingly.

To facilitate the elaboration on this algorithm, we denote the mixed strategy profile of $\tilde{\mathcal{G}}$ as P, i.e.,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & p_{nm} & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NM} \end{bmatrix}, \tag{32}$$

where p_{nm} is the probability of user n choosing channel m. Then the strategy probability vector of user n is denoted as:

$$p_n = (p_{n1}, \ldots, p_{nm}, \ldots, p_{nM} | m \in \mathcal{M}), \quad n \in \mathcal{N}, \tag{33}$$

and obviously we have $\sum_{m=1}^{M} p_{nm} = 1$.

At the beginning of each period k, every user n chooses a channel $s_n \in \mathcal{M}$ to access according to its current strategy probability vector $p_n(k)$. With the aid of fuzzy space mapping, the fluctuant utility with uncertain information is denoted as

the fuzzy utility $\tilde{u}_n(s_n, s_{-n}, \tilde{h}_{n,S}^{s_n})$, and then the robust and reliable spectrum sharing decision is made based on the derived priority vector $w_m(k) = (w_m^n(k)|n \in \mathcal{B}_m, m \in \mathcal{M})$. With the fuzzy-space implementation, the updating strategy of each user obeys the following rule:

$$p_{n,m}(k+1) =$$

$$\begin{cases} p_{n,m}(k) - \dfrac{p_{n,m}(k)\exp\left(\eta w_m^n(k)\right)}{\sum_{j\in\mathcal{B}_m}\exp\left(\eta w_m^j(k)\right)}, & m \neq s_n(k), \quad (34a) \\ p_{n,m}(k) + \dfrac{\left[1-p_{n,m}(k)\right]\exp\left(\eta w_m^n(k)\right)}{\sum_{j\in\mathcal{B}_m}\exp\left(\eta w_m^j(k)\right)}, & \text{else}, \quad (34b) \end{cases}$$

where $\eta > 0$ is the learning step parameter.

It is seen that, instead of the direct measurement, the updating rule of our designed FLRL scheme is executed based on the decision result $w_m^n(k)$ acquired in the fuzzy space. Obviously, the priority vector w_m is a principal parameter of fuzzy-space for decision making, which satisfies the normalizing condition $\sum_{n\in\mathcal{B}_m} w_m^n = 1$, $w_m^n \geqslant 0$. Quite a number of methods have been proposed to assess the priority vector of fuzzy numbers. Here, based on a fuzzy preference relation (FPR), we resort to a least deviation algorithm [36], with which a stable priority matrix $W = \{w_m|m \in \mathcal{M}\}$ of the users can be derived. The FPR matrix of users who attempt to access channel m is defined as $\boldsymbol{\Upsilon}_m = [\gamma_{i,j}]_{\beta_m\times\beta_m}$ with complementary matrix properties, i.e.,

$$\begin{cases} \gamma_{ij} + \gamma_{ji} = 1, \\ \gamma_{ij} \geqslant 0, \\ \gamma_{ii} = 0.5, \end{cases} \qquad \forall i, j \in \mathcal{B}_m, m \in \mathcal{M}, \quad (35)$$

where γ_{ij} denotes the preference degree of the network manager between ith user and jth user. For our specific situation, the γ_{ij} $(i, j \in \mathcal{B}_m)$ is given by:

$$\gamma_{ij} = \begin{cases} \min\left\{\delta H_{i,j} + (1-\delta)K_{i,j} + 0.5, 1\right\}, & H_{i,j} > 0, \\ 0.5, & H_{i,j} = 0, \\ 1 - \min\left\{\delta H_{j,i} + (1-\delta)K_{j,i} + 0.5, 1\right\}, & H_{i,j} < 0. \end{cases} \quad (36)$$

where

$$H_{i,j} \triangleq T_{\tilde{v}}\left[\tilde{u}_i(s_i, s_{-i}, \tilde{h}_{i,S}^{s_i})\right] - T_{\tilde{v}}\left[\tilde{u}_j(s_j, s_{-j}, \tilde{h}_{j,S}^{s_j})\right] \quad (37)$$

is the absolute difference, and

$$K_{i,j} \triangleq \frac{T_{\tilde{v}}\left[\tilde{u}_i(s_i, s_{-i}, \tilde{h}_{i,S}^{s_i})\right] - T_{\tilde{v}}\left[\tilde{u}_j(s_j, s_{-j}, \tilde{h}_{j,S}^{s_j})\right]}{T_{\tilde{v}}\left[\tilde{u}_j(s_j, s_{-j}, \tilde{h}_{j,S}^{s_j})\right]} = \frac{H_{i,j}}{T_{\tilde{v}}\left[\tilde{u}_j(s_j, s_{-j}, \tilde{h}_{j,S}^{s_j})\right]} \quad (38)$$

is the relative difference. The δ is used to fluctuate the weight of the absolute difference $H_{i,j}$ and the relative difference $K_{i,j}$. $\tilde{v}$ is the viewpoint of the network manager.

With the established FPR, we define the following function

$$G(\gamma_{i,j}, w_m^i, w_m^j) = 9^{2\gamma_{i,j}-1}\frac{w_m^j}{w_m^i}, \ m \in \mathcal{M} \quad (39)$$

which serves as a metric to evaluate the priority vector w_m. For $G(\gamma_{i,j}, w_m^i, w_m^j)$, we introduce the *evaluation difference* and the *evaluation ratio*, which are given by:

$$D_i = \sum_{j\in\mathcal{B}_m}\left[G(\gamma_{i,j}, w_m^i, w_m^j) - G(\gamma_{j,i}, w_m^j, w_m^i)\right], \ \forall i \in \mathcal{B}_m, \quad (40)$$

$$Q = \sqrt{\frac{\left[\sum_{j\in\mathcal{B}_m\setminus\lambda} G(\gamma_{\lambda,j}, w_m^\lambda, w_m^j)\right]}{\left[\sum_{j\in\mathcal{B}_m\setminus\lambda} G(\gamma_{j,\lambda}, w_m^j, w_m^\lambda)\right]}}, \quad (41)$$

where $\lambda = \arg\max_{j\in\mathcal{B}_m}\{|D_j|\}$.

Then the priority vector w_m is updated as:

$$w_m^n = \begin{cases} \dfrac{Q \times w_m^n}{\sum_{j\in\mathcal{B}_m\setminus\lambda} w_m^j + Q \times w_m^\lambda}, & n = \lambda, \\ \dfrac{w_m^n}{\sum_{j\in\mathcal{B}_m\setminus\lambda} w_m^j + Q \times w_m^\lambda}, & n \neq \lambda. \end{cases} \quad (42)$$

Based on the above elaborations, we show the pseudocode of a least deviation algorithm in **Algorithm 1**. With the **Algorithm 1** severing as a subfunction, the pseudocode of our proposed FLRL algorithm is presented in **Algorithm 2**. By introducing the fuzzy space to analyze the uncertain information and adopting the fuzzy-logic to quantify the mapped fuzzy utilities, the learning algorithm can intrinsically resist the non-static environments and thereby stably update to the FNE solution to optimize the network performance.

B. CONVERGENCE OF THE FLRL ALGORITHM

The convergence performance of the FLRL algorithm is discussed as follows. We concentrate on characterizing the long-term behavior of the matrix P and analyzing the dynamic mean of the FLRL algorithm. Based on the stochastic approximation theory [37] and the ordinary differential equation (ODE), the following theorem is presented.

Theorem 4: For our proposed FLRL algorithm for RSA in the dynamic WSNs, when the learning parameter η is sufficiently small, the FLRL algorithm asymptotically converges.

Proof: First, we rewrite the updating rule in eq. (34) as:

$$p_n(k+1) = p_n(k) + \Theta(k)\left(\mathbf{1}_{\{m=s_n\}} - p_n(k)\right), \quad (43)$$

where $\Theta(k) = \exp\left(\eta w_m^n(k)\right)/\sum_{j\in\mathcal{B}_m}\exp\left(\eta w_m^j(k)\right)$, and $\mathbf{1}_{\{m=s_n\}}$ is an unit vector with $m = s_n$ element being one, otherwise being zero.

We define the mapping from the mixed strategies $P(k)$ to the conditional expectation as:

$$\Phi(P) \triangleq \mathbb{E}\left[F\left(P(k), s(k), W(k)\right)|P(k)\right], \quad (44)$$

Algorithm 1 A Least Deviation Algorithm

Input: fuzzy utilities $\tilde{\mathcal{U}}_m = \{\tilde{u}_n | n \in \mathcal{B}_m(k)\}$, viewpoint $\tilde{v}$, threshold parameters ε.
Output: a stable priority vector w_m.

Initialize the priority vector $w_m = (\frac{1}{\beta_m}, \ldots, \frac{1}{\beta_m})$, $m \in \mathcal{M}$;
for n=1:$\beta_m(k)$ **do**
  calculate the relative index $T_{\tilde{v}}\left[\tilde{u}_n(s_n, s_{-n}, \tilde{h}^m_{n,S})\right]$ for the fuzzy utility $\tilde{u}_n(s_n, s_{-n}, \tilde{h}^m_{n,S})$ using eq. (18);
end
for i=1:$\beta_m(k)$ **do**
  for j=1:$\beta_m(k)$ **do**
    calculate the FPR $\boldsymbol{\Upsilon}_m$ for the relative index $T_{\tilde{v}}\left[\tilde{u}_i(s_i, s_{-i}, \tilde{h}^m_{i,S})\right]$, $i \in \mathcal{B}_m(k)$ using eq. (36);
  end
end
for n=1:$\beta_m(k)$ **do**
  calculate the *evaluation difference* D_n using eq. (40);
end
while $\max_{j\in\mathcal{B}_m}\{|D_j|\} > \varepsilon$ **do**
  $\lambda == \arg\max_{j\in\mathcal{B}_m}\{|D_j|\}$;
  calculate the *evaluation ratio* Q according to eq. (41);
  for n=1:$\beta_m(k)$ **do**
    update the priority vector $w^n_m(k+1)$ according to eq. (42);
  end
end

where the $F\Big(P(k), s(k), W(k)\Big) = P(k+1)$. Then the following lemma holds.

Lemma 3: With a sufficiently small step size η, the sequence $P(k)$, $\forall k \geqslant 0$ will weakly converge to the limiting point of the following ODE [38], i.e.,

$$\frac{\mathrm{d}P}{\mathrm{d}k} = \Phi(P), \quad P_0 = P(0), \tag{45}$$

where P_0 is the initial value of the ODE, which is equal to the initial channel selection probability matrix $P(0)$.

The dynamic mean in eq. (44) indicates that for a user n, if a channel s_n offers a better payoff, then the user will increase this channel access probability p_{n,s_n}, while decrease others access probability $p_{n,m}$, $m \neq s_n$ in future updating. Based on the stochastic approximation in eq. (45), the convergence of the FLRL algorithm can be obtained. Based on the above statements, **Theorem 4** is proved. □

V. SIMULATION RESULTS

In this section, numerical simulations are provided to demonstrate the performances of our proposed FLRL algorithm for RSA problem in dynamic and uncertain WSNs. In the following simulations, we configure the size of the square as $150 \times 150\text{m}^2$. The uncertain positions of users randomly

Algorithm 2 The FLRL Algorithm for RSA

Input: player nodes $\mathcal{N}$, strategy profile $\mathcal{S}$, positive constant C, learning parameter $\eta > 0$.
Output: a stable channel selection pattern s^*.

Set $k = 1$, and initialize the channel selection probability vector $p_n(k) = (\frac{1}{M}, \ldots, \frac{1}{M})$, $M = |\mathcal{M}|$, $n \in \mathcal{N}$;
while $\min\{\max p_n(k) | n \in \mathcal{N}\} < 0.99$ **do**
  for n=1:N **do**
    $s_n(k)$=randsrc(1,1,[$\mathcal{S}_n$;$p_n(k)$]);
  end
  for m=1:M **do**
    $\mathcal{B}_m(k) = \{n \in \mathcal{N} : s_n(k) == m\}$;
    $\beta_m(k)$=length$\Big(\mathcal{B}_m(k)\Big)$;
  end
  for n=1:N **do**
    calculate the fuzzy utility $\tilde{u}_n(s_n, s_{-n}, \tilde{h}^{s_n}_{n,S})$ according to eq. (20);
  end
  for m=1:M **do**
    for n=1:$\beta_m(k)$ **do**
      using **Algorithm 1** calculate the priority vector w^n_m of user n choosing channel m;
    end
  end
  for n=1:N **do**
    for m=1:M **do**
      if $m \neq s_n(k)$ **then**
        update selection probability $p_{n,m}(k+1)$ according to eq. (34a);
      else
        update selection probability $p_{n,m}(k+1)$ according to eq. (34b);
      end
    end
  end
  $k = k + 1$;
end

locate in their circle regions with a radius $\varphi = 5$m. Transmission power and interference range of users are $P_n \in [20, 30]$dBm and $A_n \in [30, 50]$m, respectively. The variance of AWGN and the PL exponent of channel link m are set as $\sigma^2_m = -80$dBm and $\alpha_m = 2.5$, $m \in \mathcal{M}$. Moreover, other constant parameters for carrying out the designed FLRL algorithm are set as $\delta = 0.5$, $\varepsilon = 0.8$ and $\eta = 0.2$. The viewpoint of the network manager is assumed to be *neutral*.

In the following studies, firstly we illustrate the simulation diagrams of the dynamic network model. Then we present the convergence performance of the FLRL algorithm. Finally, the system performances are demonstrated via different measurement standards. Note that all the results are obtained by independently simulating 50 network topologies, and 200 trials are implemented for each network topology.

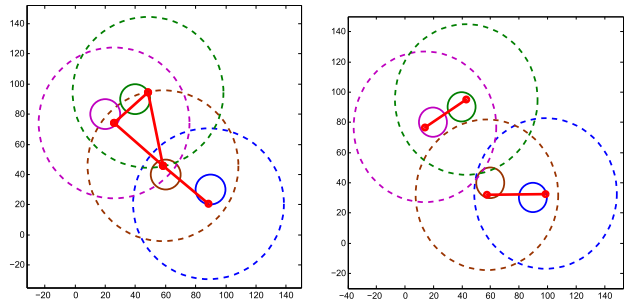


FIGURE 6. Simulation diagrams of the considered dynamic network model.

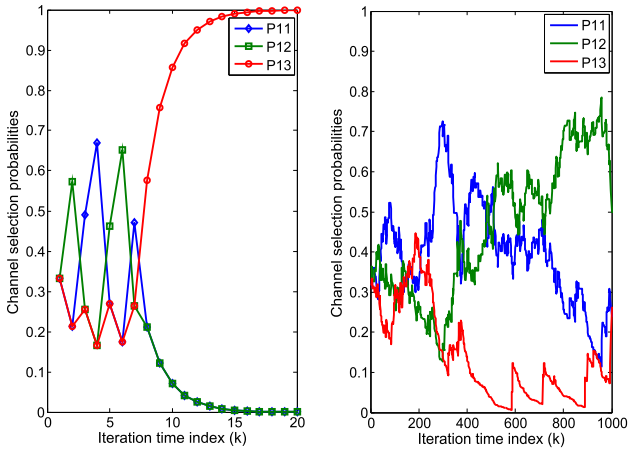


FIGURE 7. Evolution of the channel selection probabilities for an arbitrary user in dynamic WSN. (a) FLRL scheme; (b) conventional method.

A. DYNAMIC NETWORK SIMULATION

The simulation diagrams of our considered dynamic network model with multiple shared users are shown in Fig. 6. Here, for simplicity, we set $A_n = 40\text{m}$, $(n \in \mathcal{N})$. The solid circles represent the movement scope of each user, and the dots bounded in the circles are the random instantaneous position of users. The broken circles with the users' positions as centers represent the interference regions of users, and the red lines indicate the edges of the directed interference graph. It is intuitively seen from Fig. 6 that the generated directed interference graph would be different as the dynamics and uncertainty of wireless environment taking into account.

B. CONVERGENCE PERFORMANCE

To confirm the feasibility of our proposed FLRL algorithm for combating the dynamic environment, it is essential to demonstrate its convergence performance. To tackle this issue, first the evolution curves of the channel selection probability for an arbitrary user under uncertain environment with both our proposed algorithm and the conventional algorithm, are illustrated in Fig. 7(a) and Fig. 7(b), respectively. At the beginning, the user randomly selects the channels with equal probabilities. From Fig. 7(a), it is noted that with the

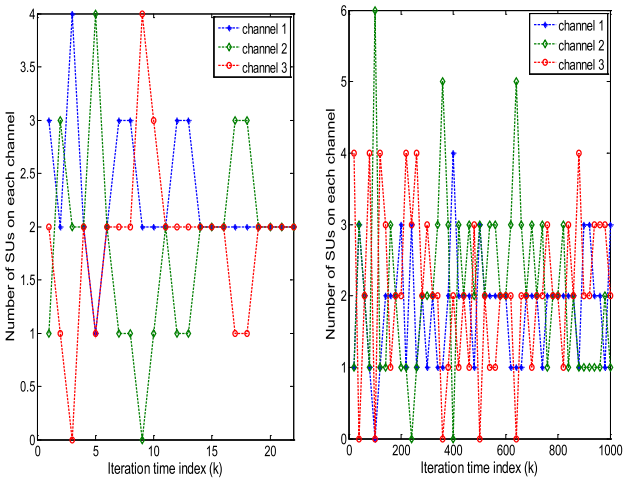


FIGURE 8. Evolution of the number of users selecting each channel in the considered dynamic WSN. (a) FLRL scheme; (b) conventional method.

algorithm iterating, the channel selection probability vector evolves from $\{1/3, 1/3, 1/3\}$ to $\{0, 0, 1\}$ in about 20 iterations, which demonstrates this user finally achieves a steady channel selection. On the contrary, as shown in Fig. 7(b), when the number of iteration is significantly large (approaching 1000), the selection probability fluctuates severely and no stable channel selection state can be obtained. Therefore, with the fuzzy space implementation, our approach can handle the uncertain information and guarantee convergence.

Besides, the evolution curves of the number of users selecting each channel in dynamic WSNs are shown in Fig. 8, which also include the proposed algorithm and the conventional method shown in Fig. 8(a) and Fig. 8(b), respectively. Since the users update their channel selection strategies continuously based on the selection probabilities, the number of users on each channel varies accordingly. It is revealed that when the proposed algorithm converges, all the users maintain their current channel selection strategies, and their number on each channel keeps unchanged, for example the result is $\beta_1 = \beta_2 = \beta_3 = 2$ in this specific case. Whereas the conventional algorithm can't reach a stable channel selection pattern, and the number of users on different channels is constantly changing, which further verifies its limitation for application in dynamic wireless environment.

For our designed FLRL algorithm, as the existence of convergence state has been certificated, we further assess the performance of convergence speed. Determined by the inherently stochastic nature of learning algorithms, the iterations needed to converge are random variables. Therefore, we compare the convergence speed from a statistical perspective. Specifically, the cumulative distribution functions (CDF) of the iterations needed for convergence of our proposed FLRL algorithm with uncertain environmental knowledge is shown in Fig. 9. It is noted that the iterations needed for convergence is positively related with the total user number, and the mean values of the required iteration number under different

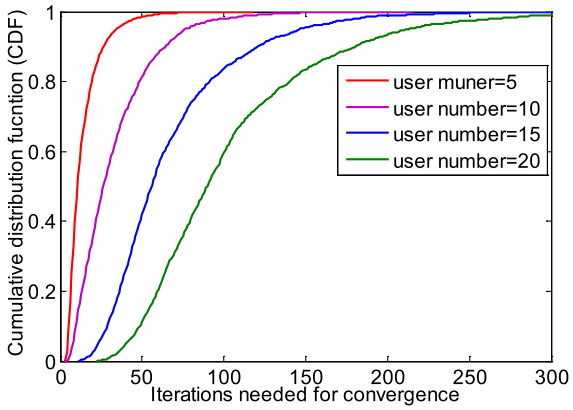


FIGURE 9. The convergence speed of the FLRL algorithm with different network scales.

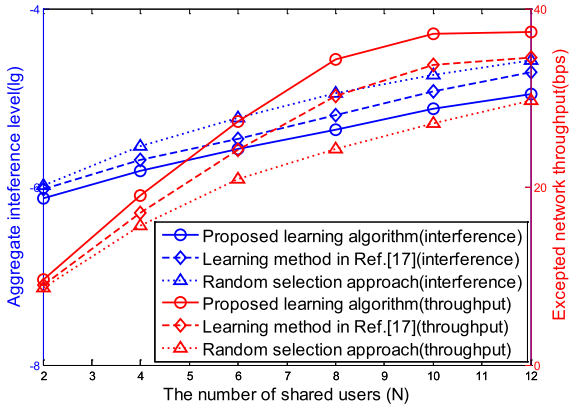


FIGURE 10. Comparisons of the system performance among the three spectrum access schemes.

network sizes ($N = 5, 10, 15, 20$) are about 15, 24, 53 and 91, respectively. Besides, we notice that, as the network size scale, the convergence rate (i.e. the required iteration number per users) is basically unchanged, which further demonstrates the stability and scalability of our designed mechanism in the dynamic WSN with uncertain information.

C. SYSTEM PERFORMANCE

In this subsection, we consider the system performance of our proposed FLRL algorithm for RSA in dynamic WSNs. Specifically, first we present the performance comparisons of network throughput and interference level with our proposed FLRL algorithm and the other two counterparts (the improved Q-learning method in [17] and the random access algorithm). Then we evaluate the influences of the membership function (triangle membership function and trapezoid membership function) to the convergence speed and the network performance of our proposed FLRL algorithm.

1) DIFFERENT ALLOCATION SCHEMES

First we show the performance gap of different spectrum access schemes, i.e. the proposed FLRL algorithm, the improved Q-learning method in [17] and the random selection approach. The number of available channels is $M = 8$. The comparison results of the expected network throughput and the aggregate interference level of the three spectrum access schemes with varying user number are shown in Fig. 10.

It is noted from the figure that, the performance (both the network throughput and the interference level) achieved by our proposed FLRL algorithm is superior to that of the improved Q-learning method and the random selection approach. Clearly, for the random selection approach, since no available learning scheme to refine the strategies of users, they just compete the channels stochastically, thus the performance of the random selection approach is the worst. More importantly, compared with the improved Q-learning method, our developed FLRL algorithm possesses the capability of efficiently combating the dynamic and stochastic wireless environment, thereby is more stable and robust under dynamic scenario. Hence can refine the strategies and obtain better performance. Owing to its significant advantage, our proposed learning scheme is promising for the considered realistic wireless communication network with dynamics and uncertainties.

2) DIFFERENT MEMBERSHIP FUNCTIONS

For our proposed FLRL algorithm, we will go further to research the influences of the membership function (triangle membership function and trapezoid membership function) on system performance, which contain two aspects: (1) the complexity of the FLRL algorithm, i.e. the iterations number needed for convergence; and (2) the system performance when achieving the optimal solution, i.e. the expected network throughput and the aggregate interference level.

The comparison result of the required iterations number for convergence of the developed FLRL algorithm applying the triangle membership function and the trapezoid membership function is shown in Fig. 11(a). The numbers of users and channels are set as $N = 5$ and $M = 3$, respectively. It is seen from the figure that the convergence of FLRL algorithm applying triangle membership function is more rapid than that applying trapezoid membership function. The expectation of the iteration number needed to converge of the former is about 15, while the latter approximately needs 21 iterations to reach the convergence.

The comparison result of the system performance of the FLRL algorithm applying the triangle membership function and the trapezoid membership function is shown in Fig. 11(b). The number of users varies from 2 to 12, and the channel number is $\|\mathcal{M}\| = 8$. It is seen that both the expected network throughput and the aggregate interference level of FLRL algorithm with the trapezoid membership function is superior to that with the triangle membership function.

Associating the above experimental results and analyzing the properties of different membership functions from a holistic perspective, we can draw the conclusion that compared with employing the triangle membership function, adopting

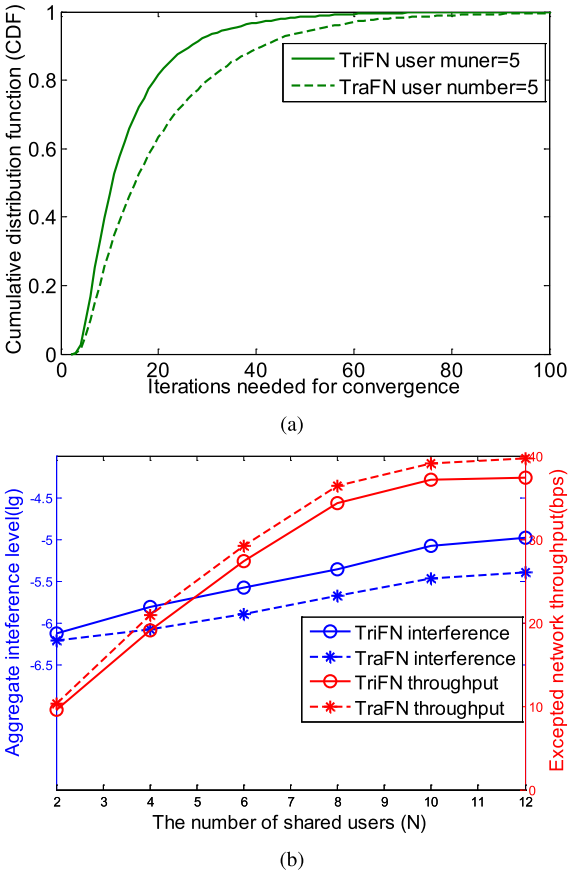


FIGURE 11. Comparison of the triangle membership function and the trapezoid membership function. (a) The iteration number for convergence; (b) The system performance.

the trapezoid membership function can get a superior network performance at the cost of a higher complexity, for the trapezoid membership function characterizing fuzzy number in a more comprehensive and detailed manner. This conclusion, which can be used to design the membership function of fuzzy number to balance the complexity and the performance of algorithms, is of significant importance to guide the future research.

VI. CONCLUSION

In this paper, we investigate the RSA problem regarding throughput maximization with fully considering the dynamic and uncertain information of the stochastic WSNs. By interpreting the time-varying CSI as a fuzzy number, we formulate the problem as a NC-FG, and thereby the random fluctuant data-rate can be represented as the fuzzy utility. On this basis, a robust FLRL paradigm is proposed to achieve the FNE solution under the dynamically uncertain wireless environment. Distinguishing from the most existing learning approaches relying on direct observations, our developed scheme executes the decision procedure according to the priority vectors in a mapped fuzzy space. Due to the fuzzy space interpolation, our scheme is inherently insensitive to uncertain CSI and changing utility, which effectively eradicates the decision fluctuation caused by environmental changes and attains a robust access in dynamic and uncertain wireless communication networks. The effectiveness and superiority of the FLRL algorithm are also demonstrated by comprehensive simulation results. As such, our proposed scheme will be of significant promise to the emerging diverse WSNs.

REFERENCES

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tut.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] L. Zhang, Y.-C. Liang, and M. Xiao, "Spectrum sharing for Internet of Things: A survey," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 132–139, Jun. 2019.

[4] B. Li, W. Guo, H. Zhang, C. Zhao, S. Li, and A. Nallanathan, "Spectrum detection and link quality assessment for heterogeneous shared access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1431–1445, Feb. 2019.

[5] B. Li, S. Li, A. Nallanathan, and C. Zhao, "Deep sensing for future spectrum and location awareness 5G communications," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1331–1344, Jul. 2015.

[6] 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.

[7] Y. Yang, B. Bai, and W. Chen, "Spectrum reuse ratio in 5G cellular networks: A matrix graph approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3541–3553, Dec. 2017.

[8] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 94–101, Apr. 2016.

[9] R. H. Tehrani, S. Vahid, D. Triantafyllopoulou, H. Lee, and K. Moessner, "Licensed spectrum sharing schemes for mobile operators: A survey and outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2591–2623, 4th Quart.,2016.

[10] C. Beckman and G. Smith, "Shared networks: Making wireless communication affordable," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 78–85, Apr. 2005.

[11] C. Liu and L. Wang, "Optimal cell load and throughput in green small cell networks with generalized cell association," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1058–1072, May 2016.

[12] R. Yin, G. Yu, A. Maaref, and G. Y. Li, "A framework for co-channel interference and collision probability tradeoff in LTE licensed-assisted access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6078–6090, Sep. 2016.

[13] J. Dai and S. Wang, "Clustering-based spectrum sharing strategy for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 228–237, Jan. 2017.

[14] C. Jiang, Y. Chen, Y. Gao, and K. J. R. Liu, "Joint spectrum sensing and access evolutionary game in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2470–2483, May 2013.

[15] C. Jiang, Y. Chen, and K. J. R. Liu, "Multi-channel sensing and access game: Bayesian social learning with negative network externality," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2176–2188, Apr. 2014.

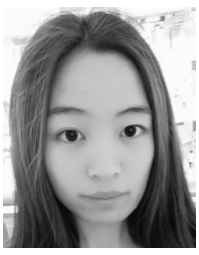
[16] L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 44–51, Oct. 2017.

[17] N. Morozs, T. Clarke, and D. Grace, "Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 817–825, Apr. 2016.

[18] A. M. Akhtar, X. Wang, and L. Hanzo, "Synergistic spectrum sharing in 5G HetNets: A harmonized SDN-enabled approach," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 40–47, Jan. 2016.

[19] Y. Sun, M. Peng, and H. V. Poor, "A distributed approach to improving spectral efficiency in uplink device-to-device-enabled cloud radio access networks," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6511–6526, Dec. 2018.

[20] Y. Sun, M. Peng, and S. Mao, "Deep reinforcement learning-based mode selection and resource management for green fog radio access networks," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1960–1971, Apr. 2019.
[21] 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.
[22] C. Zhang, Z. Wei, Z. Feng, and W. Zhang, "Spectrum sharing of drone networks," in *Handbook of Cognitive Radio*. Singapore: Springer, 2019, pp. 1279–1304.
[23] A. S. Shafigh, S. Glisic, E. Hossain, B. Lorenzo, and L. A. DaSilva, "User-centric distributed spectrum sharing in dynamic network architectures," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 15–28, Feb. 2019.
[24] C. Fan, B. Li, Y. Zhang, and C. Zhao, "Robust dynamic spectrum access in uncertain channels: A fuzzy payoffs game approach," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
[25] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[26] X. Ge, J. Ye, Y. Yang, and Q. Li, "User mobility evaluation for 5G small cell networks based on individual mobility model," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 528–541, Mar. 2016.
[27] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "QoE driven decentralized spectrum sharing in 5G networks: Potential game approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7797–7808, Sep. 2017.
[28] B. Li, J. Hou, X. Li, Y. Nan, A. Nallanathan, and C. Zhao, "Deep sensing for space-time doubly selective channels: When a primary user is mobile and the channel is flat rayleigh fading," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3362–3375, Jun. 2016.
[29] H.-J. Zimmermann, *Fuzzy Set Theoryąland and Its Applications*. Amsterdam, The Netherlands: Springer, 2011.
[30] M. Larbani, "Non cooperative fuzzy games in normal form: A survey," *Fuzzy Sets Syst.*, vol. 160, no. 22, pp. 3184–3210, Nov. 2009.
[31] H. Lee-Kwang and J.-H. Lee, "A method for ranking fuzzy numbers and its application to decision-making," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 677–685, Dec. 1999.
[32] D.-F. Li, "An effective methodology for solving matrix games with fuzzy payoffs," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 610–621, Apr. 2013.
[33] A. Chakeri and F. Sheikholeslam, "Fuzzy Nash Equilibriums in Crisp and Fuzzy Games," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 1, pp. 171–176, Feb. 2013.
[34] L. Cunlin and Z. Qiang, "Nash equilibrium strategy for fuzzy non-cooperative games," *Fuzzy Sets Syst.*, vol. 176, no. 1, pp. 46–55, Aug. 2011.
[35] F. Kacher and M. Larbani, "Existence of equilibrium solution for a non-cooperative game with fuzzy goals and parameters," *Fuzzy Sets Syst.*, vol. 159, no. 2, pp. 164–176, Jan. 2008.
[36] Z. Xu and Q. Da, "A least deviation method to obtain a priority vector of a fuzzy preference relation," *Eur. J. Oper. Res.*, vol. 164, no. 1, pp. 206–216, Jul. 2005.
[37] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, vol. 35. New York, NY, USA: Springer, 2003.
[38] P. S. Sastry, V. V. Phansalkar, and M. A. L. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 5, pp. 769–777, May 1994.



CHAOQIONG FAN received the bachelor's degree in communication engineering from the China University of Petroleum (UPC), Shandong, China, in 2015. She is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing, China. Her research interests include dynamic spectrum access for cognitive radios, resource management, game theory, and learning theory.

SHIJIAN BAO serves as an Engineer with China International Engineering Consulting Corporation (CIECC), Beijing, China. His research interest includes emerging technologies of 5G wireless communication.

YIWEN TAO received the bachelor's degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2016, where he is currently pursuing the Ph.D. degree. His research interests include statistical signal processing algorithms, sequential estimation and detection, target localization, and tracking.

BIN LI received the bachelor's degree in electrical information engineering from the Beijing University of Chemical Technology, in 2007, and the Ph.D. degree in information and communication engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2013. In 2013, he joined BUPT, where he is currently an Associate Professor with the School of Information and Communication Engineering. He has authored over 70 journal and conference papers. His current research interest includes statistical signal processing for wireless communications, such as millimeter-wave communications and cognitive radios. He received the 2011 ChinaCom Best Paper Award, the 2015 IEEE WCSP Best Paper Award, and the 2010 and 2011 BUPT Excellent Ph.D. Student Award Foundations.

CHENGLIN ZHAO received the bachelor's degree in radio technology from Tianjin University, in 1986, and the master's degree in circuits and systems and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1993 and 1997, respectively, where he currently serves as a Professor. His research interests include emerging technologies of short-range wireless communication, cognitive radios, and 60-GHz millimeter-wave communications.