Multi-Armed Bandit Channel Access Scheme With Cognitive Radio Technology in Wireless Sensor Networks for the Internet of Things

JIANG ZHU[1], YONGHUI SONG[1], DINGDE JIANG[2], (Member, IEEE), AND HOUBING SONG[3], (Senior Member, IEEE)

[1]Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[2]Computer Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China
[3]Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV 25136, USA

Corresponding authors: J. Zhu (zhujiang@cqupt.edu.cn) and D. Jiang (jiangdd@mail.neu.edu.cn)

ABSTRACT The wireless sensor network (WSN) is one of the key enablers for the Internet of Things (IoT), where WSNs will play an important role in future internet by several application scenarios, such as healthcare, agriculture, environment monitoring, and smart metering. However, today's radio spectrum is very crowded for the rapid increasing popularities of various wireless applications. Hence, WSN utilizing the advantages of cognitive radio technology, namely, cognitive radio-based WSN (CR-WSN), is a promising solution for spectrum scarcity problem of IoT applications. A major challenge in CR-WSN is utilizing spectrum more efficiently. Therefore, a novel channel access scheme is proposed for the problem that how to access the multiple channels with the unknown environment information for cognitive users, so as to maximize system throughput. The problem is modeled as I.I.D. multi-armed bandit model with M cognitive users and N arms ($M < N$). In order to solve the competition and the fairness between cognitive users of WSNs, a fair channel-grouping scheme is proposed. The proposed scheme divides these channels into M groups according to the water-filling principle based on the learning algorithm UCB-K index, the number of channels not less than one in each group and then allocate channel group for each cognitive user by using distributed learning algorithm fairly. Finally, the experimental results demonstrate that the proposed scheme cannot only effectively solve the problem of collision between the cognitive users, improve the utilization rate of the idle spectrum, and at the same time reflect the fairness of selecting channels between cognitive users.

INDEX TERMS I.I.D. multi-armed bandit, fairness access scheme, channel-grouping, modified UCB-tuned.

I. INTRODUCTION

In future, WSNs are expected to be integrated into the Internet of Things [1], where reconfigurable, flexible, and intelligent sensors join the Internet dynamically, and use it to collaborate and accomplish their tasks for a wide range of applications various domains [2]–[16], big data applications, Internet of Things, E-commerce, multimedia medical device, virtual reality & augmented reality, and environment monitoring, etc. Multimedia applications of WSNs require substantially low power consumption, higher bandwidth and more available spectrum [17]–[23]. However, today's radio spectrum is very crowded for rapid increasing popularities of various

wireless applications. Therefore, cognitive radio technology stands as a key and spectrum-efficient communication approach for resource-constrained wireless sensor networks and future wireless network [24]–[26]. When cognitive users, namely, sensor nodes in wireless sensor networks access the spectrum competitively, in order to make effective usage of the spectrum and meet the throughput demand for multimedia applications, effective mechanisms are required to coordinate cognitive users' behaviors (transmission power control, spectrum access, et al.) [27]–[35]. At the same time, the wireless environment is gradually to be more complicated, which makes it more difficult for cognitive users to obtain the

complete network environment parameters information in the wireless network system. Also, the upcoming 5G networking architectures with a super high spectrum utilization and ultra-low power consumption, tends to support massive devices with limited bandwidth and minimum content delay, which would be a big challenge for the current crowded spectrum. Hence, distributed channel access under unknown environment information has been a hot research topic in cognitive radio based wireless sensor network.

Currently, many of the existing literatures (see [36]–[41]) have studied the problem of distributed spectrum access under unknown environment statistical information. Among these literatures, [36] and [37] have researched on the Upper Confidence Bound (UCB) index algorithm with single cognitive user based on multi-armed bandit model, while [38]–[41] focused on the multi-user situation, which transform the unknown system model into Non-Bayesian multi-armed bandit model. Considering the single cognitive user situation is too simple, the current trend is to study multi-armed bandit model with multi-user and multi-channel. Meanwhile, all cognitive users in the system do not know the channel environment statistical information. There will be collisions when more than one cognitive user accesses the same channel simultaneously. To address the problem, [42] proposes an adaptive random access scheme to reduce the collision among cognitive users. The priority access scheme and fair access scheme are also adopted in literature [43] and [44], which could effectively avoid the collision among cognitive users. Based on the UCB1 index algorithm, the literature [45] puts forward the d-UCB4 algorithm, which is suitable for the multi-user scenarios. Literature [46] proposes a distributed online learning access scheme called DSLA, which can be widely applied to distribution learning system and its channels have a variety of channel available probability.

Most of the current literatures, such as [47] and [48] are limited that one cognitive user can only select one channel each time. In this situation, when the selected channel is detected to be in a busy state, the data will not be transmitted and it has to wait for the next slot. However, there may be other channels that are not occupied presently, which will give rise to a waste of spectrum resources. For this problem, we propose a scheme which is based on channel grouping. In each time slot, cognitive users can sense all of the channels in the group one by one, until finding an idle channel or all of the channels are perceived once. Here, these channels are divided into several groups according to the water-filling principle based on the modified UCB-tuned index. On the other hand, many literatures don't take the fairness among the cognitive users into consideration. In this paper, we introduce the fairness access scheme based on the channel grouping method. Finally, the experimental results show that the proposed scheme can get a logarithmic regret with respect to time slot, and increase the selected number of the channels that has a small idle probability, so as to ensure the fairness among the cognitive users.

This paper is organized as follows. Section II introduces the system model. Then, section III introduces the proposed scheme of this paper, namely, the distributed learning algorithm, and the principle of channels grouping and the access principle. Section IV gives the scheme simulation and analysis. At last, we summarize the paper and point out the future research.

II. SYSTEM MODEL

As shown in Fig. 1, a CR-WSN coexisting with a licensed system is considered. The time is slotted and the N channels are independent of each other.
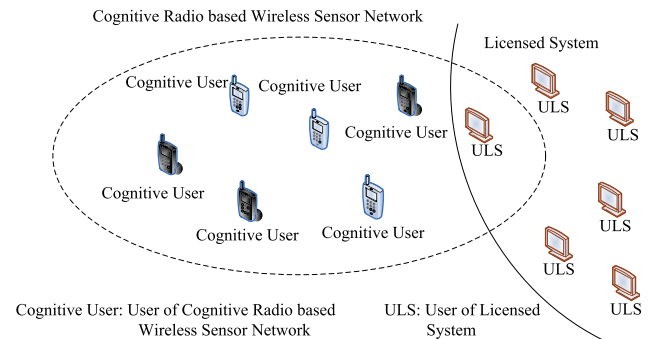

FIGURE 1. The structure of the system.

We denote time slot as n, and denote the reward (i.e. throughput) arbitrary of channel $i(1 \leq i \leq N)$ as random variable $X_i(n)$. The mean of random variable $X_i(n)$ is $\mu_i = E[X_i]$, where we normalize $X_i \in [0, 1]$. Different channels have different means μ_i, which is unknown for these cognitive users and there is no information exchange or communication between cognitive users.
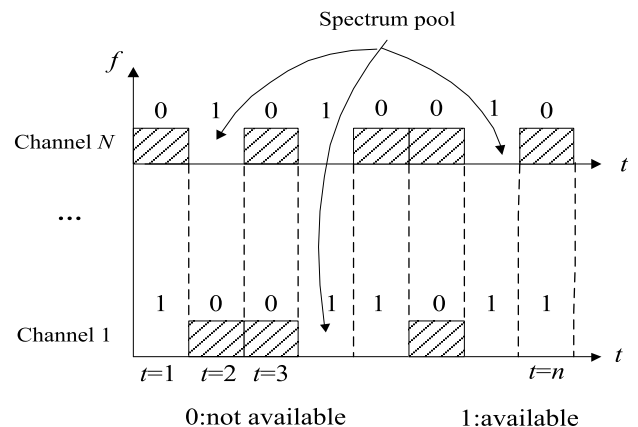

FIGURE 2. Channel model in the system.

In the distributed CR-WSN, each cognitive user can select one opportunity channel from the N channels. The channel set is $N \triangleq \{1, \ldots, N\}$. These channels are divided into time slots one by one. As shown in Fig 2, each channel has two states: $s_i(n) = 1$ and $s_i(n) = 0$. $s_i(n) = 1$ represents that in

each time slot, the channel is not occupied by the licensed user (i.e. the channel is available), otherwise, $s_i(n) = 0$ (i.e. the channel is not available). The set of the channels state space is $S = \{s_1(n), s_2(n), \ldots, s_i(n) \ldots, s_N(n)\}$. For cognitive user i, the closed-form normalized instantaneous reward at slot n is $X_i(n) = s_i(n) \times B, X_i \in [0, 1]$. Here let $B = 1$. These channels are independent and identically distributed with different Bernoulli parameters. The objective function is to maximize the total throughput or minimize the regret. So we formulate the problem as I.I.D. MAB model. In the model, cognitive users do not know the channel state information, and they have to estimate and predict channel availability by exploring and learning. The practical scheme is π. The scheme performance is evaluated by its regret value, which means that the difference between the obtained reward under an ideal environment and the obtained practical reward by taking some strategies. Then we can get the mathematical expression of the regret:

$$\Re^\pi(\Theta; n) = n \sum_{i \in O_M^*} \mu_i - E^\pi \left[\sum_{t=1}^{n} S_{\pi(t)}(t) \right] \quad (1)$$

Where, π is the adopted strategy, $\Theta = \{\mu_i, 1 \leq i \leq N\}$ is the channels availability set, o_M^* is the M channels set with M largest reward, E stands for the calculation of expectation value, $S_{\pi(t)}(t)$ stands for the total throughput by using the scheme, its mathematical expression is as shown below:

$$S_{\pi(t)}(t) = \sum_{m=1}^{M} \sum_{i=1}^{N} X_i(t) \times \Omega_{m,i}(t) \quad (2)$$

In formula (2), $\Omega_{m,i}(t)$ reflects the collision between cognitive users: in slot t, when cognitive user m is the only one to use channel i, then $\Omega_{m,i}(t) = 1$, otherwise $\Omega_{m,i}(t) = 0$. For facilitation, we let $V_{m,i}^\pi(n) = \sum_{t=1}^{n} \Omega_{m,i}(t)$, on account of $E[X_i(t)] = \mu_i$. Therefore, we have the expression of total regret:

$$\Re^\pi(\Theta; n) = n \sum_{i \in O_M^*} \mu_i - \sum_{m=1}^{M} \sum_{i=1}^{N} \mu_i \times E\left[V_{m,i}^\pi(t)\right] \quad (3)$$

III. THE PRINCIPLE OF THE PROPOSED SCHEME
A. DISTRIBUTED LEARNING ALGORITHM BASED ON MODIFIED UCB-TUNED INDEX
Due to the channel environment statistical information is completely unknown to cognitive users, we need to predict the channel information by learning algorithm. The proposed distributed learning algorithm in this paper is briefly called UCBT-K that modified based on the tuned upper confidence bound (UCBT) index, which introduces a variance factor in the index and it is generalized form of the modified UCBT. The UCBT-K can select arbitrarily channel with the K-th [49] largest index value. The process of UCBT-K on M cognitive users can be obtained as follows.

a) *Initialization*: Cognitive user m senses all the channelsone by one. At time slot t, the instantaneous throughput of channel i is $\hat{\mu}_{m,i}(t) = X_i(t)$. The sensed number of times of channel i is $T_i(t) = 1$.

b) *Channel Estimation*: Calculate the improved UCB-tuned index for each channel according to the mathematical formula (4):

$$\hat{\mu}_{m,i}(t-1) + \sqrt{\frac{\log(t)}{T_i(t-1)}}$$
$$\times \sqrt{\min\left(\frac{1}{4}, \delta_i^2(t-1) + \sqrt{2\log n / T_i(t-1)}\right)} \quad (4)$$

Where, $\delta_i^2(t)$ is the reward variance of the channel i after t time slots. Then, we construct a channel set o_K that contains the K largest index values.

c) *Channel Selection*: Select a channel k from o_K based on the following mathematical formula (5):

$$k = \arg\min_{i \in o_K} \hat{\mu}_{m,i}(t-1) + \sqrt{\frac{\log(t)}{T_i(t-1)}}$$
$$\times \sqrt{\min\left(\frac{1}{4}, \delta_i^2(t-1) + \sqrt{2\log(t) / T_i(t-1)}\right)} \quad (5)$$

d) *Update Information:* if the selected channel is idle, update the average reward $\hat{\mu}_{m,k}(t)$ of cognitive user m.

$$\hat{\mu}_{m,k}(t) = \frac{\hat{\mu}_{m,k}(t-1) \times T_k(t-1) + X_k(t)}{T_k(t-1) + 1} \quad (6)$$

And $T_i(t) = T_i(t-1) + 1$, otherwise, there are $\hat{\mu_k^m}(t) = \hat{\mu_k^m}(t-1)$ and $T_i(t) = T_i(t-1)$.

B. THE PRINCIPLE OF CHANNELS GROUPING
There are M distributed cognitive users and N channels in the system ($M < N$). The principle of channel grouping can be shown as the following Fig. 3.
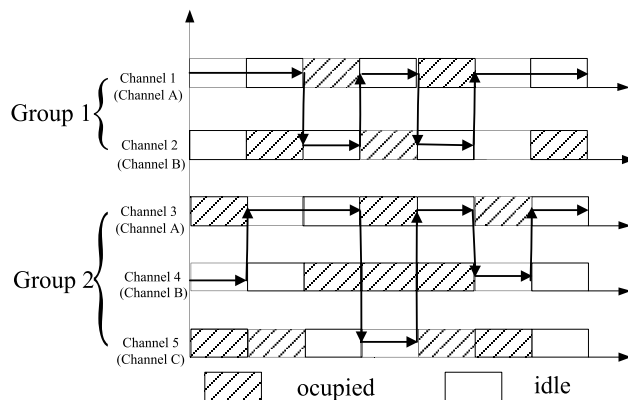

FIGURE 3. The channel sensing based on channel group.

Firstly, we let the channel with the largest index value in the set o_M as the channel A of the first channel-group ($G_1 = \{A\}$).

Call the channel with second largest index value as the channel A of the second group ($G_2 = \{A\}$), and the rest of the channels contained in o_M are grouped in the same way. Then, we will acquire M channel-group ($G_1, G_2, \ldots G_M$). Next, allocate the rest $N - M$ channels evenly to the M channel-group according to iterative water-filling principle: Put the channel with $(M + 1)$-th largest index value to the current channel-group with smallest sum of index values denoted as channel-B, the next channel added in this group is called channel-C. It is also follow the same principle from the channel with $(M + 2)$-th largest index value to channel with N-th largest index value, which forms a new channel-group set $o_M^* = \left\{ \begin{array}{l} G_1 = \{A, B \ldots\}, G_2 = \{A, B, \ldots\} \\ , \ldots \ldots, G_M = \{A, B, C \ldots\} \end{array} \right\}$. The set o_M^* still includes M channel-group, so the M cognitive users have a corresponding channel-group. For each group has at least one channel, therefore, we assume that the number of channel N is larger than cognitive users M. Assume that the time slot consists of three parts: the detection, the access process and the decision. We assume the first part and the third part is quite short and can be ignored. In this article we allow a cognitive user to sense more than one channel in a time slot.

C. THE ACCESS OF CHANNEL-GROUPS WITH FAIRNESS

Since there are multiple cognitive users in the system, it is necessary to find a reasonable channel access scheme to avoid collisions among cognitive users. In literature [42], the priority access scheme is adopted to avoid collisions. First, it assigns a rank of priority access for each cognitive user. Then each cognitive user chooses the corresponding channel according to its rank. For example, for cognitive user 1 its priority is the first, so it can choose the channel with largest index value every time. In term of the cognitive user 2, its priority is the second, it will always select the one with second largest index value. And the other cognitive users select in the same way. As a result, each cognitive user has a corresponding channel. Although this scheme can effectively avoid the collision among cognitive users, it does not reflect the fairness among cognitive users. Therefore, to solve the problem of unfairness, we propose an access scheme which is based on channel grouping, as shown in Fig. 4.
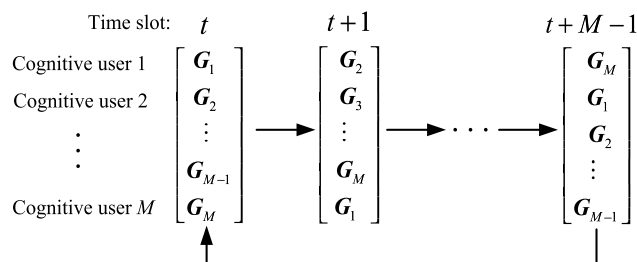
FIGURE 4. Channel-group access scheme with fairness.

As mentioned above, each cognitive user has a corresponding channel group. For arbitrary cognitive user m,

at time slot t, the selected channel-group G_j needs to meet the formula:

$$G_j = ((m + t) \bmod M) + 1 \tag{7}$$

The algorithm implementation of the proposed scheme can be shown in Fig. 5, which includes three parts: distributed learning, channel grouping and fair access scheme.

Lemma 1: The existence of regret value, assume that the parameter l is positive integer, n is the total number of time slot, $t, 1 \leq t \leq n$ is any arbitrary time slot. $K_m^*(t)$ is the optimal channel with the K-th largest index for the cognitive user m. $Q_i^m(n)$ is the number of times that cognitive user m chooses the non-optimal channel i (i.e. $i \neq K_m^*(t)$). $C_{t,n_i} = \sqrt{\frac{(L+1)\ln(t)}{n_i}}$, $\hat{\mu}_{i,n_i}$ is the average reward of the channel i. When $\mu_i < \mu_{K_m^*}$, there are:

$$\Pr\left\{\hat{\mu}_{j(t),n_{j(t)}} \leq \mu_{j(t)} - C_{t,n_{j(t)}}\right\} \leq t^{-4} \tag{8}$$

$$\Pr\left\{\hat{\mu}_{i,n_i} \geq \mu_i + C_{t,n_i}\right\} \leq t^{-4} \tag{9}$$

Proof: As $\mu_i < \mu_{K_m^*}$, there is at least one channel $j(t) \in o_{K_m^*}^*$. If the cognitive user chooses the channel i at slot t, then we can get:

$$\hat{\mu}_{j(t),Q_{j(t)}^m(t-1)} + C_{t-1,Q_{j(t)}^m(t-1)} \leq \hat{\mu}_{i,Q_i^m(t-1)} + C_{t-1,Q_{i(t)}^m(t-1)} \tag{10}$$

$Q_i^m(n)$ can be expressed as below:

$$Q_i^m(n) = 1 + \sum_{t=N+1}^{n} \| \{I_i(t)\} \leq l$$

$$+ \sum_{t=N+1}^{n} \| \{I_i(t), Q_i^m(t-1) \geq l\}$$

$$\leq l + \sum_{t=N+1}^{n} \left(\begin{array}{l} \| \{I_i(t), \mu_i < \mu_{K_m^*}, Q_i^m(t-1) \geq l\} \\ + \| \{I_i(t), \mu_i > \mu_{K_m^*}, Q_i^m(t-1) \geq l\} \end{array} \right) \tag{11}$$

Where, if the number of $Q_i^m(n)$ increases by one at time slot t, $I_i(t) = 1$, otherwise $I_i(t) = 0$. And when x is true, $\| (x) = 1$, otherwise, let $\| (x) = 0$.

According to the (10), there are:

$$\sum_{t=N+1}^{n} \| \{I_i(t), \mu_i < \mu_{K_m^*}, Q_i^m(t-1) \geq l\}$$

$$\leq \sum_{t=N+1}^{n} \| \left\{ \begin{array}{l} \hat{\mu}_{j(t),Q_{j(t)}^m(t-1)} + C_{t-1,Q_{j(t)}^m(t-1)} \\ \leq \hat{\mu}_{i,Q_i^m(t-1)} + C_{t-1,Q_i^m(t-1)}, Q_i^m(t-1) \geq l \end{array} \right\}$$

$$\leq \sum_{t=N+1}^{n} \| \left\{ \min_{0 < n_{j(t)} < t} \hat{\mu}_{j(t),n_{j(t)}} + C_{t-1,n_{j(t)}} \right.$$

$$\left. \leq \max_{l \leq n_i < t} \hat{\mu}_{i,n_i} + C_{t-1,n_i} \right\}$$
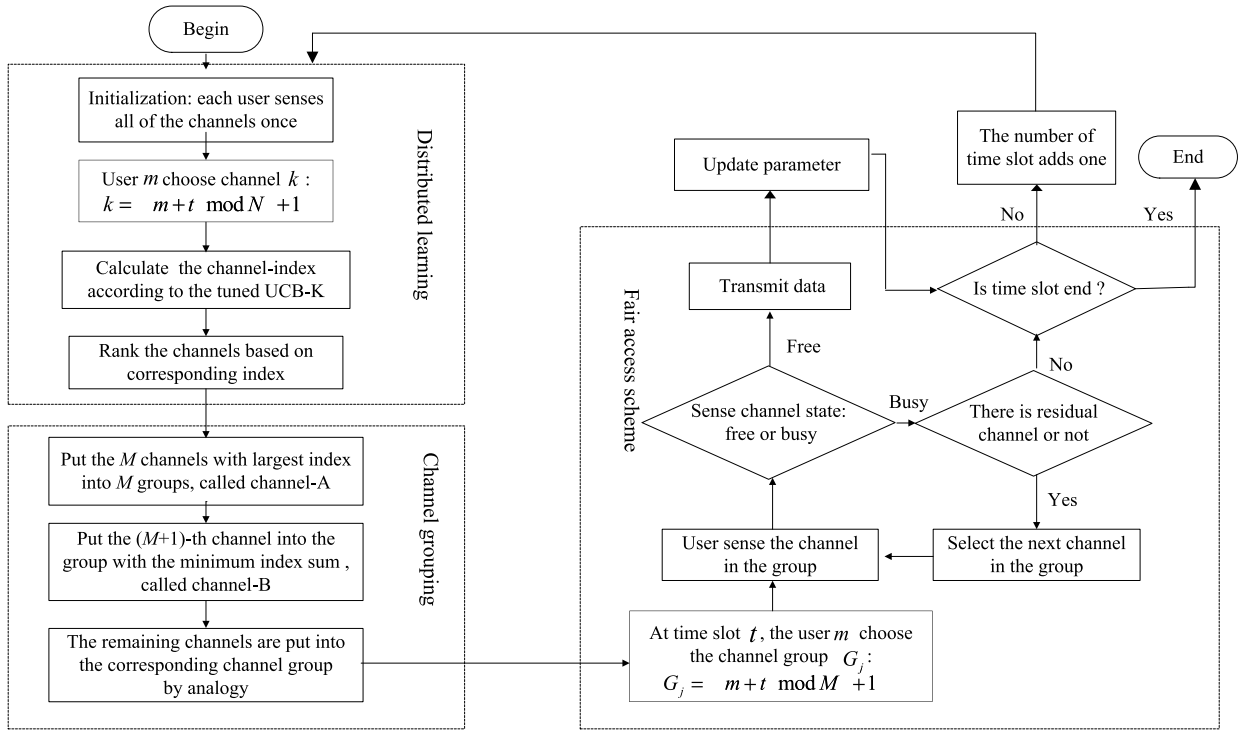
FIGURE 5. The flow chart of the proposed scheme.

$$\leq \sum_{t=1}^{\infty} \sum_{n_{j(t)}=1}^{t-1} \sum_{n_i=l}^{t-1} 1\{\hat{\mu}_{j(t),n_{j(t)}} + C_{t,n_{j(t)}} \leq \hat{\mu}_{i,n_i} + C_{t,n_i}\} \quad (12)$$

From the (12), for $\hat{\mu}_{j(t),n_{j(t)}} + C_{t,n_{j(t)}} \leq \hat{\mu}_{i,n_i} + C_{t,n_i}$, at least one of the following expressions is true:

$$\mu_{j(t)} < \mu_i + 2C_{t,n_i} \quad (13)$$

$$\hat{\mu}_{j(t),n_{j(t)}} \leq \mu_{j(t)} - C_{t,n_{j(t)}} \quad (14)$$

$$\hat{\mu}_{i,n_i} \geq \mu_i + C_{t,n_i} \quad (15)$$

For $l \geq \frac{8 \ln n}{\Delta_{K_m^*,i}^2}$, there are:

$$\mu_{j(t)} - \mu_i - 2C_{t,n_i}$$
$$\geq \mu_{K_m^*} - \mu_i - 2\sqrt{\frac{2\Delta_{K_m^*,i}^2 \ln t}{8 \ln n}}$$
$$\geq \mu_{K_m^*} - \mu_i - \Delta_{K_m^*,i} = 0 \quad (16)$$

So it is clearly false for the inequality (13). Similarly, we can get the inequality (14) and the inequality (15) hold.

According to the Chernoff-Hoeffding bound [46] for (14) and (15), we can get:

$$\Pr\{\hat{\mu}_{j(t),n_{j(t)}} \leq \mu_{j(t)} - C_{t,n_{j(t)}}\} \leq e^{-4\ln t} = t^{-4} \quad (17)$$

$$\Pr\{\hat{\mu}_{i,n_i} \geq \mu_i + C_{t,n_i}\} \leq e^{-4\ln t} = t^{-4} \quad (18)$$

■

Lemma 2: When $\mu_i > \mu_{K_m^*}$, similar to lemma 1, the following formula hold:

$$\Pr\{\hat{\mu}_{i,n_i} \leq \mu_i - C_{t,n_i}\} \leq t^{-4} \quad (19)$$

$$\Pr\{\hat{\mu}_{h(t),n_{h(t)}} \geq \mu_{h(t)} + C_{t,n_{h(t)}}\} \leq t^{-4} \quad (20)$$

Proof: In the case of $\mu_i > \mu_{K_m^*}$, we assume the channel i is selected at t slot by the cognitive user m. There are two possible situations:

(1) If $o_{K_m^*} = o_{K_m^*}^*$, there are:

$$\hat{\mu}_{i,Q_i^m(t-1)} - C_{t-1,Q_i^m(t-1)}$$
$$\leq \hat{\mu}_{K_m^*,Q_{K_m^*}^m(t-1)} - C_{t-1,Q_{K_m^*}^m(t-1)} \quad (21)$$

(2) If $o_{K_m^*} \neq o_{K_m^*}^*$, there is at least one channel $h(t) \notin o_{K_m^*}^*$ in the set $o_{K_m^*}$, we can get:

$$\hat{\mu}_{i,Q_i^m(t-1)} - C_{t-1,Q_i^m(t-1)}$$
$$\leq \hat{\mu}_{h(t),Q_{h(t)}^m(t-1)} - C_{t-1,Q_{h(t)}^m(t-1)} \quad (22)$$

For the above two situations, $\Phi_{K_m^*}$ is a set of channels with $K - th$ largest index. We let $o_{K_m^*-1}^* = o_{K_m^*}^* - \Phi_{K_m^*}$, the cognitive user m chooses channel i at t slot. There is a channel $h(t) \notin o_{K_m^*-1}^*$ that meets the following formula:

$$\hat{\mu}_{i,Q_i^m(t-1)} - C_{t-1,Q_i^m(t-1)} \leq \hat{\mu}_{K_m^*,Q_{K_m^*}^m(t-1)} - C_{t-1,Q_{K_m^*}^m(t-1)} \quad (23)$$

In the same way, we can get:

$$\hat{\mu}_{i,Q_i^m(t-1)} - C_{t-1,Q_i^m(t-1)} \leq \hat{\mu}_{h(t),Q_{h(t)}^m(t-1)} - C_{t-1,Q_{h(t)}^m(t-1)} \quad (24)$$

$$\mu_i < \mu_{h(t)} + 2C_{t,n_i} \quad (25)$$

$$\hat{\mu}_{i,n_i} \leq \mu_i - C_{t,n_i} \quad (26)$$

$$\hat{\mu}_{h(t),n_{h(t)}} \geq \mu_{h(t)} + C_{t,n_{h(t)}} \qquad (27)$$

From (25)-(27), there are at least one true inequality, for $l \geq \frac{8\ln n}{\Delta_{K,i}^2}$, there are:

$$\mu_i - \mu_{h(t)} - 2C_{t,n_i} \geq \mu_i - \mu_{K_m^*} - \Delta_{K_m^*,i} \geq 0 \qquad (28)$$

According to the inequality (28), the formula (25) is obviously wrong, which means the others are true.

According to the Chernoff-Hoeffding bound for the formula (26) and (27), we can get the following formula, respectively:

$$\Pr\left\{\hat{\mu}_{i,n_i} \leq \mu_i - C_{t,n_i}\right\} \leq e^{-4\ln t} = t^{-4} \qquad (29)$$
$$\Pr\left\{\hat{\mu}_{h(t),n_{h(t)}} \geq \mu_{h(t)} + C_{t,n_{h(t)}}\right\} \leq e^{-4\ln t} = t^{-4} \qquad (30)$$

∎

Theorem 1: The expected regret under the proposed scheme is bounded and grows logarithmically with time slot going on.

Proof: Because the selection times of the cognitive user m to the non-optimal channel i can be expressed as:

$$Q_i^m(n) = 1 + \sum_{t=N+1}^{n} \| \{I_i(t)\}$$

$$\leq l + \sum_{t=N+1}^{n} \| \{I_i(t), Q_i^m(t-1) \geq l\}$$

$$\leq l + \sum_{t=N+1}^{n} \left(\begin{array}{l} \| \{I_i(t), \mu_i < \mu_{K_m^*}, Q_i^m(t-1) \geq l\} \\ + \| \{I_i(t), \mu_i > \mu_{K_m^*}, Q_i^m(t-1) \geq l\} \end{array} \right)$$

$$(31)$$

Where, if the number of $Q_i^m(n)$ increases by one at time slot t, let $I_i(t) = 1$, otherwise $I_i(t) = 0$. And when x is true, $\|(x) = 1$, otherwise, let $\|(x) = 0$.

Here we need to discuss two cases: namely, $\mu_i < \mu_{K_m^*}$ and $\mu_i > \mu_{K_m^*}$. According to the above lemma 1 and lemma 2, we can get the expectation about $Q_i^m(n)$:

$$E\left[Q_i^m(n)\right]$$

$$\leq \left\lceil \frac{8\ln n}{\Delta_{\min,i}^2} \right\rceil$$

$$+ \sum_{t=1}^{\infty} \sum_{n_{j(t)}=1}^{t-1} \sum_{n_i=y}^{t-1} \left(\begin{array}{l} \Pr\left\{\hat{\mu}_{j(t),n_{j(t)}} \leq \mu_{j(t)} - C_{t,n_{j(t)}}\right\} \\ + \Pr\left\{\hat{\mu}_{i,n_i} \geq \mu_i + C_{t,n_i}\right\} \end{array} \right)$$

$$+ \sum_{t=1}^{\infty} \sum_{n_i=y}^{t-1} \sum_{n_{h(t)}=1}^{t-1} \left(\begin{array}{l} \Pr\left\{\hat{\mu}_{i,n_i} \leq \mu_i - C_{t,n_i}\right\} \\ + \Pr\left\{\hat{\mu}_{h(t),n_{h(t)}} \geq \mu_{h(t)} + C_{t,n_{h(t)}}\right\} \end{array} \right)$$

$$\leq \frac{8\ln n}{\Delta_{\min,i}^2} + 1 + 2\sum_{t=1}^{\infty} \sum_{n_{j(t)}=1}^{t-1} \sum_{n_i=1}^{t-1} 2t^{-4}$$

$$\leq \frac{8\ln n}{\Delta_{\min,i}^2} + 1 + \frac{2\pi^2}{3} \qquad (32)$$

For any cognitive user m, the regret can be expressed as follows:

$$\Re^\pi(\Theta, m, n) \leq \sum_{i=1}^{n} Q_i^m(n) \times \mu_{\max} + \sum_{h\neq m} \sum_{i\in o_m^*} Q_h^m(n)\mu_i \quad (33)$$

$$\Re^\pi(\Theta, n) = \sum_{m=1}^{M} \Re^\pi(\Theta, m, n)$$
$$\leq M\left(N + M \times (M-1)\right)$$
$$\times \left(\frac{8\ln n}{\Delta_{\min}} + 1 + \frac{2\pi^2}{3} \right) \times \mu_{\max} \qquad (34)$$

∎

IV. SIMULATION AND ANALYSIS
A. THE COMPLEXITY ANALYSIS OF REGRET

According to the conclusion of Theorem 1, when $n/\ln n \geq C$, (C is a constant which is related to cognitive users and channels, $C = 8(N+M)/\Delta_{\min}^2 + (1 + 2\pi^2/3)N + M$). We can get a bound about loose learning regret value when the total time slots n increases to a very large number.

$$R^\pi(\Theta, n) = M(N-M)\left(\frac{8\ln n}{\Delta_{\min}} + 1 + \frac{2\pi^2}{3} \right)\mu_{\max}$$
$$+ M^3\left(1 + \frac{2\pi^2}{3} \right)\mu_{\max} \qquad (35)$$

From the formula (35), the scale of the proposed scheme is $O(M(N-M)\ln n)$. By research on the existing literature, we can obtain a statistical table about regret complexity growth as shown in table 1:

TABLE 1. The complexity of regret.

ρ^{rand}	$O(M^2N)$
TDFS	$O(M(\max\{M^2, (N-M)M\}))$
DLP	$O(M(N-M))$
DLF	$O(M(N-M))$
Proposed	$O(M(N-M))$

It can be seen from table 1, the complexity of regret function produced by these regular schemes, such as ρ^{rand} and time division fair sharing (TDFS), is three times power about the number M of cognitive users and the number of channels N. However these schemes adapted in this paper obtain a lower complexity, which is two times power function. Therefore we will further study the performance between the comparison schemes (the DLP scheme, the DLF scheme) and the proposed scheme through experimental simulation.

B. SIMULATION RESULTS

In this experiment, we assume that these channels are mutually independent and identically distributed. The availability
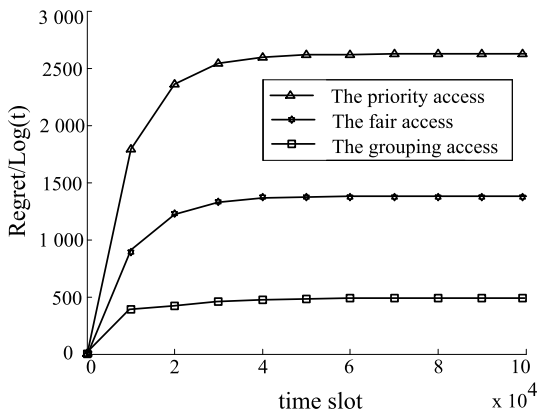
FIGURE 6. The total regret strategies ($n = 10^4$).

TABLE 2. The case of each channel selected with DLP.

CH i / User m	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$m=1$	9548	192	101	95	64
$m=2$	210	9440	202	105	43
$m=3$	115	220	9395	180	90

TABLE 3. The case of each channel selected with DLF.

CH i / User m	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$m=1$	3308	3282	3266	90	54
$m=2$	3315	3291	3265	85	44
$m=3$	3300	3310	3275	80	35

TABLE 4. The case of each channel selected with proposed scheme.

CH i / User m	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$m=1$	2776	2782	2729	899	814
$m=2$	2713	2740	2816	827	904
$m=3$	2738	2824	2787	825	826

FIGURE 7. The channel selection under DLP ($n = 10^4$).
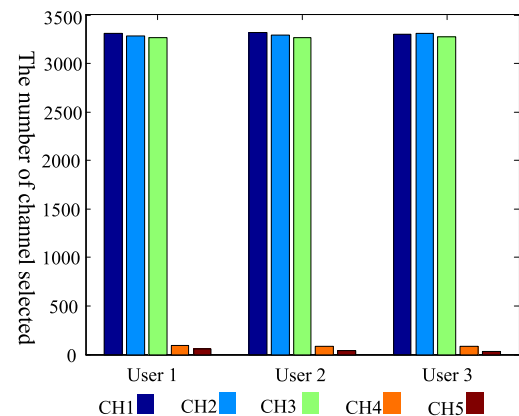
FIGURE 8. The channel selection under DLF ($n = 10^4$).

of channels obeys Bernoulli process with different parameters, while it is unknown to the M cognitive users. There are only two channel states: busy (i.e. $s_i(n) = 0$) or idle (i.e. $s_i(n) = 1$). When the channel is idle, the throughput is 1, otherwise it is 0. We run the simulation results for 100 times, and then take a statistical average. There are four cognitive users (i.e. $M = 4$) and nine channels (i.e. $N = 9$) with $\Theta = \{0.9, 0.8, \ldots, 0.2, 0.1\}$. The accumulate regret of the proposed scheme compared with the priority access scheme (DLP), the fair access scheme (DLF) and the regret under the three strategies is shown as Fig. 6, it can be obviously known that the regret is uniformly logarithmic with time-slot and the proposed scheme can yield lower regret.

In the case of $M = 3$ and $N = 5$ with $\Theta = \{0.9, \ldots, 0.6, 0.5\}$, the total time slot is $n = 10^4$. As shown in Fig. 6, when the number of slots increases to 4000 approximately, the simulation results will converge, namely, cognitive users can estimate the channel information after
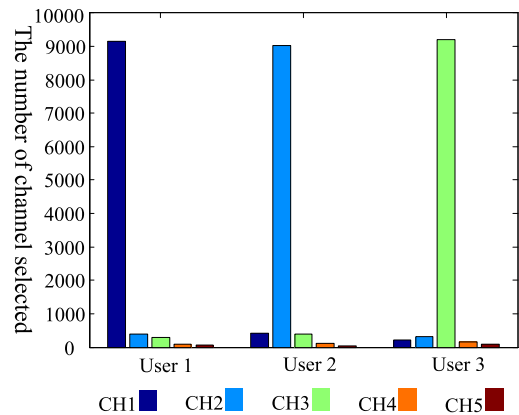
4000 slots approximately by using the proposed scheme. We verify the fairness under the three strategies. The results that the number of each channels selected by each cognitive user are shown in table 2-table 4, respectively, which clearly reflect that the proposed scheme can significantly increase the channel utilization rate with less idle probability.

At the same time, we provide the bar chart about the channel selected situation running on three strategies. The expression can be shown in Fig. 7, Fig. 8, Fig. 9, respectively.

Fig. 7 shows that there is a serious unfairness among various cognitive users. From Fig. 8, we can find the fairness between different cognitive users has been reflected, but the usage of the channels with smaller idle probability is low, such as channel $i = 4$ and $i = 5$. In Fig. 9, the proposed scheme not only reflects the fairness between various cognitive users, but also takes more advantages of channels with smaller idle probability. By fairness of channels selection, all the channels are selected more times, which can avoid the situation that only one or a small number of channels are selected while most of the others are unselected. The more channels are selected, the more opportunities for cognitive users to access channels, therefore, the efficiency of channel usage is improved.
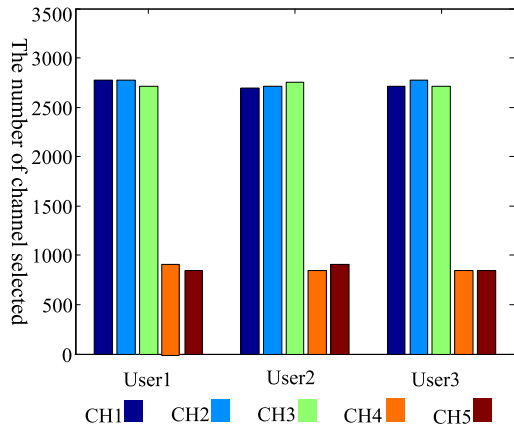
FIGURE 9. The channel selection under proposed scheme ($n = 10^4$).

V. CONCLUSION

In this paper, a multiple cognitive users and multiple channels system is considered in CR-WSNs. The channels statistical information is completely unknown for the cognitive users. To solve the problem of channel selection in CR-WSNs, a novel fair access scheme with channel grouping is proposed. Firstly, an online learning method called modified UCB-K based on the well-known UCB is used. Then these channels are divided into several groups according to the principle of channel grouping, which can improve the usage rate of the idle spectrum. Besides, we adopted the distributed learning with fairness to avoid collision between cognitive users and at the same time embody the fairness between cognitive users. Finally, the simulation results also show the superiority of the proposed scheme. With the development of the Internet of Things, the scale of networks is larger and larger. How to obtain and deal with the large amount of channel information for larger scale wireless sensor networks may be a promising research topic, and schemes based on wireless big data may be adopted.

REFERENCES

[1] L. Mainetti, L. Patrono, and A. Vilei, "Evolution of wireless sensor networks towards the Internet of Things: A survey," in *Proc. IEEE Int. Conf. Softw., Telecommun. Comput. Netw.*, Sep. 2011, pp. 1–6.

[2] T. Hua, M. Zheng, J. Tan, L. Zhu, and W. Miao, "Intelligent photovoltaic monitoring based on solar irradiance big data and wireless sensor networks," *Ad Hoc Netw.*, vol. 35, no. 5, pp. 127–136, 2015.

[3] B. He and Y. Li, "Big data reduction and optimization in sensor monitoring network," *J. Appl. Math.*, vol. 2014. pp. 1–8, Mar. 2014.

[4] Y. Lin, J. Yang, Z. Lv, W. Wei, and H. Song, "A self-assessment stereo capture model applicable to the Internet of Things," *Sensors*, vol. 15, no. 8, pp. 20925–20944, 2015.

[5] R. Roman, C. Alcaraz, J. Lopez, and N. Sklavos, "Key management systems for sensor networks in the context of the Internet of Things," *Comput. Elect. Eng.*, vol. 37, no. 2, pp. 147–159, 2011.

[6] J. Yang, S. He, Y. Lin, and Z. Lv, "Multimedia cloud transmission andstorage system based on Internet of Things," *Multimedia Tools Appl.*, pp. 1–16, Oct. 2015.

[7] E. Yaacoub, A. Kadri, and A. Abu-Dayya, "Cooperative wireless sensor networks for green Internet of Things," in *Proc. 8th ACM Symp. QoS Secur. Wireless Mobile Netw. (Q2SWinet)*, New York, NY, USA, 2012, pp. 79–80.

[8] G. Li and J. Shi, "Wireless sensor network technology and its application potentials for service innovation in supply chain management," *Int. J. Appl. Logistics*, vol. 1, no. 4, pp. 30–51, 2010.

[9] G. G. Messier and I. G. Finvers, "Traffic models for medical wireless sensor networks," *IEEE Commun. Lett.*, vol. 11, no. 1, pp. 13–15, Jan. 2007.

[10] Y. Sun *et al.*, "Design of medical wireless sensor network," *Adv. Mater. Res.*, vols. 314–316, pp. 2486–2490, Aug. 2011.

[11] D. Goldsmith, F. Liarokapis, G. Malone, and J. Kemp, "Augmented reality environmental monitoring using wireless sensor networks," in *Proc. IEEE 12th Int. Conf. Inf. Vis. (IV)*, London, U.K., Jul. 2008, pp. 539–544.

[12] T. Gulrez and M. Kavakli, "Precision position tracking in virtual reality environments using sensor networks," in *Proc. IEEE Int. Symp. Ind. Electron. (ISIE)*, Vigo, Spain, Jun. 2007, pp. 1997–2003.

[13] J. Hu, W. Wan, R. Wang, and X. Yu, "Virtual reality platform for smart city based on sensor network and OSG engine," in *Proc. IEEE Int. Conf. Audio, Lang. Image Process. (ICALIP)*, Shanghai, China, Jul. 2012, pp. 1167–1171.

[14] M. Othman and K. Shazali, "Wireless sensor network applications: A study in environment monitoring system," *Proc. Eng.*, vol. 41, pp. 1204–1210, Jul. 2012.

[15] J. Yang, J. Zhou, Z. Lv, W. Wei, and H. Song, "A real-time monitoring system of industry carbon monoxide based on wireless sensor networks," *Sensors*, vol. 15, no. 11, pp. 29535–29546, 2015.

[16] J. Hwang, C. Shin, and H. Yoe, "Study on an agricultural environment monitoring server system using wireless sensor networks," *Sensors*, vol. 10, no. 12, pp. 11189–11211, 2010.

[17] D. Jiang, Z. Xu, W. Wang, Y. Wang, and Y. Han, "A collaborative multi-hop routing algorithm for maximum achievable rate," *J. Netw. Comput. Appl.*, vol. 57, pp. 182–191, Nov. 2015.

[18] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, no. 4, pp. 921–959, 2007.

[19] D. Jiang, Z. Xu, P. Zhang, and T. Zhu, "A transform domain-based anomaly detection approach to network-wide traffic," *J. Netw. Comput. Appl.*, vol. 40, no. 2, pp. 292–306, Apr. 2014.

[20] H. Z. Eldin, M. A. Elhosseini, and H. A. AliAuthor, "Image compression algorithms in wireless multimedia sensor networks: A survey," *Ain Shams Eng. J.*, vol. 6, no. 2, pp. 481–490, 2015.

[21] D. Jiang, Z. Xu, Z. Chen, Y. Han, and H. Xu, "Joint time–frequency sparse estimation of large-scale network traffic," *Comput. Netw.*, vol. 55, no. 10, pp. 3533–3547, Oct. 2011.

[22] H. Shen, G. Bai, L. Zhao, and Z. Tang, "An adaptive opportunistic network coding mechanism in wireless multimedia sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 8, no. 12, pp. 1–13, 2012.

[23] P. G. V. Naranjo, M. Shojafar, H. Mostafaei, Z. Pooranian, and E. Baccarelli, "P-SEP: A prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks," *J. Supercomput.*, vol. 72, pp. 1–23, Jun. 2016.

[24] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[25] N. Cordeschi, D. Amendola, M. Shojafar, and E. Baccarelli, "Distributed and adaptive resource management in cloud-assisted cognitive radio vehicular networks with hard reliability guarantees," *Veh. Commun.*, vol. 2, no. 1, pp. 1–12, 2015.

[26] T. Haustein *et al.*, "Cognitive wireless communications—A paradigm shift in dealing with radio resources as a prerequisite for the wireless network of the future—An overview on the topic of cognitive wireless technologies," *Frequenz*, vol. 70, nos. 7–8, pp. 281–288, 2016.

[27] D. Jiang, Y. Wang, C. Yao, and Y. Han, "An effective dynamic spectrum access algorithm for multi-hop cognitive wireless networks," *Comput. Netw.*, vol. 19, vol. 84, no. 19, pp. 1–16, Jun. 2015.

[28] Z. Lin and M. van der Schaar, "Autonomic and distributed joint routing and power control for delay-sensitive applications in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 102–113, Jan. 2011.

[29] Y. Chen, G. Yu, Z. Zhang, H.-H. Chen, and P. Qiu, "On cognitive radio networks with opportunistic power control strategies in fading channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2752–2761, Jul. 2008.

[30] D. Jiang, Z. Xu, W. Li, and Z. Chen, "Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks," *J. Syst. Softw.*, vol. 104, pp. 152–165, Mar. 2015.

[31] S. M. Sanchez, R. D. Souza, E. M. G. Fernandez, and V. A. Reguera, "Rate and energy efficient power control in a cognitive radio ad hoc network," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 451–454, May 2013.

[32] D. Jiang, Z. Xu, J. Liu, and W. Zhao, "An optimization-based robust routing algorithm to energy-efficient networks for cloud computing," *Telecommun. Syst.*, vol. 63, no. 1, pp. 89–98, Mar. 2015.

[33] S. Gao, L. Qian, and D. R. Vaman, "Distributed energy efficient spectrum access in wireless cognitive radio sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Las Vegas, NV, USA, Mar. 2008, pp. 1442–1447.

[34] J. Yang, B. Chen, J. Zhou, and Z. Lv, "A low-power and portable biomedical device for respiratory monitoring with a stable power source," *Sensors*, vol. 15, no. 8, pp. 19618–19632, 2015.

[35] S. Maghsudi and S. Stanczak, "Hybrid centralized-distributed resource allocation for device-to-device communication underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2481–2495, Apr. 2016.

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

[37] T. A. Myrvoll and J. E. Håkegård, "Dynamic spectrum access in realistic environments using reinforcement learning," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, 2012, pp. 465–470.

[38] K. Liu and Q. Zhao, "Distributed learning in cognitive radio networks: Multi-armed bandit with distributed multiple players," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 3010–3013.

[39] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.

[40] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y.-D. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1380–1391, Apr. 2012.

[41] S. Vakili, K. Liu, and Q. Zhao, "Deterministic sequencing of exploration and exploitation for multi-armed bandit problems," *IEEE J. Sel. Topics Signal Process.*, vol. 59, no. 3, pp. 1902–1916, Oct. 2013.

[42] A. Anandkumar, N. Michael, A. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 731–745, Apr. 2011.

[43] Y. Gai and B. Krishnamachari, "Decentralized online learning algorithms for opportunistic spectrum access," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6.

[44] Y. Gai and B. Krishnamachari, "Distributed stochastic online learning policies for opportunistic spectrum access," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6184–6193, Dec. 2014.

[45] D. Kalathil, N. Nayyar, and R. Jain, "Decentralized learning for multiplayer multiarmed bandits," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2331–2345, Apr. 2014.
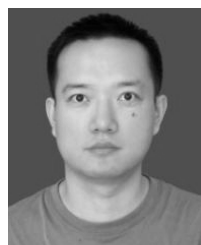
[46] M. Zandi, M. Dong, and A. Grami, "Decentralized spectrum learning and access adaptive to channel availability distribution in primary network," in *Proc. Signal Process. Adv. Wireless Commun. (SPAWC)*, 2013, pp. 130–134.

[47] M. Lelarge, A. Proutiere, and M. S. Talebi, "Spectrum bandit optimization," in *Proc. Inf. Theory Workshop (ITW)*, Sep. 2013, pp. 1–5.

[48] P. Tehrani, Q. Zhao, and L. Tong, "Multi-channel opportunistic spectrum access in unslotted primary systems with unknown models," in *Proc. Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, 2011, pp. 157–160.

[49] N. Torabi, K. Rostamzadeh, and V. C. M. Leung, "Rank-optimal channel selection strategy in cognitive networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 410–415.

[50] D. Pollard, *Convergence of Stochastic Processes*. Berlin, Germany: Springer, 1984.

YONGHUI SONG received the B.E. degree from the North China Institute of Aerospace Engineering, China, in 2014. He is currently pursuing the M.E. degree with the Chongqing University of Posts and Telecommunications. His research interests include wireless sensor networks and cognitive radio.

DINGDE JIANG (S'08–M'09) received the Ph.D. degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2009. He is currently an Professor with the School of Computer Science and Engineering, Northeastern University, Shenyang, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. His research focuses on network measurement, modeling and optimization, performance analysis, network management, network security in communication networks, particularly in software defined networks, information-centric networking, energy-efficient networks, and cognitive networks.

His research is supported by the National Science Foundation of China, the Program for New Century Excellent Talents in the University of Ministry of Education of China. He received the Best Paper Awards at the International Conferences. He has been serving as an Editor of the one international journals. He served as a Technical Program Committee Member for several international conferences.

HOUBING SONG (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, in 2012.

He joined the Department of Electrical and Computer Engineering, West Virginia University, Montgomery, WV, in 2012, where he is currently an Assistant Professor and the Founding Director of the Security and Optimization for Networked Globe Laboratory (SONG Lab), and West Virginia Center of Excellence for Cyber-Physical Systems sponsored by West Virginia Higher Education Policy Commission. In 2007, he was an Engineering Research Associate with the Texas A&M Transportation Institute. He is also the Editor of four books, including entitlted *Smart Cities: Foundations, Principles and Applications*, (Hoboken, NJ: Wiley, 2017), the *Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications*, (Chichester, UK: Wiley, 2017), the *Cyber-Physical Systems: Foundations, Principles and Applications*, (Waltham, MA: Elsevier, 2016), and the *Industrial Internet of Things: Cybermanufacturing Systems*, (Cham, Switzerland: Springer, 2016). He has authored over 100 articles. His research interests include cyber-physical systems, internet of things, cloud computing, big data analytics, connected vehicle, wireless communications and networking, and optical communications and networking.

Dr. Song is a member of ACM. He was the recipient of the Golden Bear Scholar Award, the Highest Faculty Research Award at the West Virginia University Institute of Technology, in 2016.

JIANG ZHU received the Ph.D. degree in communication and information from the University of Electronic Science and Technology, China, in 2009. He is currently an Associate Professor of the School of Telecommunications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. His current researches include cognitive radio, wireless sensor networks, and mobile communication security technology.

• • •