Distributed Algorithm for Learning to Coordinate in Infrastructure-Less Network

Rohit Kumar, Sumit J. Darak, Manjesh K. Hanawal, Ajay K. Sharma, and Rajiv K. Tripathi

***Abstract*— We consider the spectrum access in an unlicensed spectrum (i.e., no incumbent users) for infrastructure-less networks where the number of users are unknown and they cannot coordinate with others due to lack of a control channel or a central controller. Also, users do not have spectrum sensing capability due to size and power constraints in battery operated radios. Such a setup is being studied for Internet of Things applications to enable sensors to communicate sensed data without the need of dedicated spectrum and network infrastructure. Using multi-user multi-armed bandit-based learning framework, we propose a new distributed algorithm which achieves a lower regret (i.e., throughput loss) than existing algorithms while keeping the number of collisions low. Fewer collisions save power which would have been otherwise wasted due to re-transmissions. High confidence bounds on the regret and number of collisions along with simulation results validate the effectiveness of our algorithm.**

***Index Terms*— Infrastructure-less network, multi-armed bandit, distributed algorithm.**

I. Introduction

With the emergence of paradigms such as Internet of Things, the design of networks with thousands of sensors is itself a challenging problem due to the need of infrastructure to set-up such large-scale network and availability of dedicated spectrum. For such networks, infrastructure-less network are being explored where sensors/users can communicate the sensed information as and when needed over the unlicensed spectrum. This eliminates the need of a control channel overhead for the coordination among sensors and reduces the complexity of the sensor since they do not need sensing hardware to check the presence of other sensors and incumbent users [1]. The single user multi-armed bandit (MAB) algorithm offers an intuitive framework for identifying the top channels (Set of first N channels when arranged in the decreasing order of their quality, such as fading, throughput etc., [2], N is the number of users in the network) via exploration and exploitation trade-off. In the multi-user MAB model, users compete over the channels leading to collisions resulting in poor regret.[1] The collisions lead to retransmissions resulting in loss of battery power, spectrum and time. This becomes more challenging when N is unknown. Thus, how to minimize the regret without compromising on the number of collisions is a challenging problem and is the focus of this paper.

Various works such as [3]–[6] deal with coordination in infrastructure-less networks but they assume prior knowledge of N. The algorithms in [7], Multiuser $\epsilon-$greedy collision avoiding (MEGA) [8], musical chair (MC) [9], [10] are the only algorithms which do not need the prior knowledge of N. The algorithm in [7] allows users to sequentially hop among all channels without incurring collisions but its regret is high due to uniform channel selection. The MC outperforms MEGA and is the current state-of-the-art algorithm. The MC algorithm is extended to deal with non-ideal sensing detectors in [10]. The main drawback of MC is the large number of collisions leading to high regret and making it unfeasible for battery operated radios. Furthermore, random selection of channels makes the rendezvous, i.e., synchronization between the transmitter and receiver to lock on the same channel, challenging. Here, we propose distributed algorithm that aim to overcome these issues.

Manuscript received August 28, 2018; revised November 6, 2018; accepted December 3, 2018. Date of publication December 28, 2018; date of current version February 11, 2019. The associate editor coordinating the review of this paper and approving it for publication was Y. Chen. *(Corresponding author: Rohit Kumar.)*

R. Kumar and R. K. Tripathi are with the Department of ECE, National Institute of Technology at Delhi, New Delhi 110040, India (e-mail: rohitkumar@nitdelhi.ac.in; rajivtripathi@nitdelhi.ac.in).

S. J. Darak is with the Department of ECE, Indraprastha Institute of Information Technology at Delhi, New Delhi 110020, India (e-mail: sumit@iiitd.ac.in).

M. K. Hanawal is with the Department of IEOR, IIT Bombay, Mumbai 400076, India (e-mail: mhanawal@iitb.ac.in).

A. K. Sharma is with the Department of CSE, I. K. Gujral Punjab Technical University, Kapurthala 144603, India (e-mail: vc@ptu.ac.in).

Digital Object Identifier 10.1109/LCOMM.2018.2890226

[1]Regret is defined as difference between the best aggregate throughput achievable when all the users cooperate with prior knowledge of network parameters (channel statistic and N) and the throughput achieved without coordination and any prior knowledge of the network parameters [2]–[9].

II. Network Model

Consider an infrastructure-less network consisting of N users competing for K channels. We assume $K \geq N$ and time slotted communication where the horizon consists of T time slots, i.e., $t \in \{1, 2, .., T\}$. The throughput of the channel $k \in [K]$ is sampled independently from some distribution on [0,1] with mean μ_k. In each time slot, each user selects a channel and they aim to receive highest cumulative throughput over T slots. The same set-up has been considered in [2]–[10].

Collision occurs on selection of the same channel by more than one user otherwise data transmission is treated as successful. Our aim is to develop a distributed algorithm that keeps the regret, R, as small as possible where R is given as:

$$R = \sum_{t=1}^{T} \sum_{n \in N^*} \mu_n - \sum_{t=1}^{T} \sum_{n=1}^{N} \mu_{I_t^n} (1 - C_t^n) \tag{1}$$

where N^* denotes the set of channels with N highest mean throughputs, i.e., the set of top channels, I_t^n is the channel chosen by user n in time slot t. Also, $\mu_{I_t^n}$ and C_t^n denote the expected throughput and collision indicator on channel I_t^n at time t, respectively. In addition to lower regret, it is desirable that the algorithm should incur fewer collisions and guarantee fair channel allocation among users.

III. Proposed Algorithm

The proposed algorithm is implemented independently by each user and consists of three phases, namely Channel

Estimation (CE) phase, User Estimation (UE) phase and Orthogonalization (OR) phase that run sequentially.

A. CE Phase

This phase starts with the random selection of channels in each time slot till user successfully transmits on a channel and does not face collision. Next slot onwards, the user starts sequential hopping (SH) in which a channel with next higher index (up to modulo K) is selected in each slot. This phase runs for $T_{CE} = (T_{RH} + T_{SH})$ duration where expressions for T_{RH} and T_{SH} are given in Lemma 1 and 2, respectively. In each time slot, user checks whether there is collision or not. Based on the collision feedback, each user updates number of successful transmissions (S_k) and total throughput (V_k) received on each channel. At the end of CE phase, each user estimates the statistics $\hat{\mu}_k = \frac{V_k}{S_k}$ of all K channels. These estimates are then used to rank the channels and channel indices arranged in the decreasing order of $\hat{\mu}_k$ are stored in an array π. The duration T_{CE} guarantees that users are on distinct channels with high probability at the end of CE phase but they may not be necessarily in the top channels.

Lemma 1: For any $\delta_1 \in (0,1)$, all users will be orthogonalized on non-overlapping channels with probability $\geq 1 - \delta_1$ if they select channel uniformly randomly for T_{RH} (Eq. 2) time slots.

Proof: If P_C denotes the collision probability of an user when all the users are randomly hopping at any time t, and if none of the other users are on non-overlapping channel (worst-case) then probability that an user will find a non-overlapping channel within T_{RH} is given by: $\sum_{t=1}^{T_{RH}} P_C^{t-1}(1 - P_C)$. We want this probability to be at least $1 - \frac{\delta_1}{K}$ for each user. Thus

$$\sum_{t=1}^{T_{RH}} P_C^{t-1}(1 - P_C) \geq 1 - \frac{\delta_1}{K} \iff 1 - {P_C}^{T_{RH}} \geq 1 - \frac{\delta_1}{K}$$

$$T_{RH} \log P_C \leq \log\left(\frac{\delta_1}{K}\right) \iff T_{RH} = \frac{\log\left(\frac{\delta_1}{K}\right)}{\log P_C}.$$

We next give an uniform upper bound on P_C. Note that in any round some users may be selecting channels sequentially while others uniformly at random. Fix a round t and let $N_r \geq 1$ denotes the number of users selecting channels uniformly at random. Let P_{RH} and P_{SH} denotes the probability of no collision due to randomly hopping users and sequentially hopping users, respectively. We have

$$P_{NC} = 1 - P_C = (P_{RH} + P_{SH}) \geq \sum_{j=1}^{N_r} \frac{P_{RH}}{K}$$
$$\geq \frac{\left(1 - \frac{1}{K}\right)^{N_r - 1}}{K} \geq \frac{\left(1 - \frac{1}{K}\right)^{N-1}}{K} \geq \frac{\left(1 - \frac{1}{K}\right)^{K-1}}{K}$$

Substituting the upper bound on P_C thus obtained we get T_{RH} such that within T_{RH} number of time slots, all the users will orthogonalize with probability at least $1 - \delta_1$. We set

$$T_{RH} = \frac{\log\left(\frac{\delta_1}{K}\right)}{\log\left(1 - \frac{1}{K}\left(1 - \frac{1}{K}\right)^{K-1}\right)}. \tag{2}$$

Lemma 2: For any given $\delta_2 \in (0,1)$ and $\epsilon > 0$, if each SU follows collision free sequential hopping for T_{SH} (See Eq. 3)) time slots, then all the users will have ϵ-correct ($\forall \epsilon > 0$) ranking of channels with probability $\geq 1 - \delta_2$.

Proof: If for any user n it is true that $|\hat{\mu}_k - \mu_k| \leq \frac{\epsilon}{2}\ \forall k \in [K]$, then the user has an $\epsilon-$ correct ranking. We will upper bound the probability that no user has $\epsilon-$ correct ranking given the users have O_{min} observations of each channel. Assuming $\overline{X}$ denotes complement of an event X, we define three events:

J_n - User n has observed each channel at least $O_{\min}$ times.

A_n - User n has ϵ- correct ranking.

B_n - User n has at least O_{min} observations of each channel.

Now, we want to compute $Pr(\overline{A}_n|B_n) < \frac{\delta_2}{K}$ and we know,

$$Pr(\overline{A}_n|B_n)$$
$$\leq Pr\left(\exists k \in\ 1 \cdots K\quad s.t|\hat{\mu}_k - \mu_k| > \frac{\epsilon}{2}|\ B_n\right)$$
$$= \sum_{k=1}^{K} \sum_{j=O_{min}}^{\infty} Pr\left(|\hat{\mu}_k - \mu_k| > \frac{\epsilon}{2}|\ J_n = j\right)$$
$$\cdot Pr\left(J_n = j|\ B_n\right) \qquad \text{(By Union Bound)}$$
$$\leq \sum_{k=1}^{K} \sum_{j=O_{min}}^{\infty} 2 \cdot \exp\left(\frac{-j \cdot \epsilon^2}{2}\right) Pr\left(J_n = j|\ B_n\right)$$
$$\text{(By Hoeffding's Inequality)}$$
$$\leq \sum_{k=1}^{K} 2 \cdot \exp\left(\frac{-O_{min} \cdot \epsilon^2}{2}\right) \sum_{j=O_{min}}^{\infty} Pr\left(J_n = j|\ B_n\right)$$
$$\leq K \cdot 2 \cdot \exp\left(\frac{-O_{min} \cdot \epsilon^2}{2}\right)$$

We can apply Hoeffding's Inequality since each observation of the channel is independent of the number of times we observe that channel. In order for this to be $<\frac{\delta_2}{K}$, we set

$$K \cdot 2 \cdot \exp\left(\frac{-O_{min} \cdot \epsilon^2}{2}\right) < \frac{\delta_2}{K} \implies O_{min} > \ln\left(\frac{2 \cdot K^2}{\delta_2} \cdot\right) \frac{2}{\epsilon^2}$$

Since in each time slot the n^{th} user is getting one observation of a channel, thus the number of time slots required to obtain O_{min} observations of K channels, i.e., T_{SH} is given by:

$$T_{SH} = \frac{2 \cdot K}{\epsilon^2} \cdot \ln\left(\frac{2 \cdot K^2}{\delta_2}\right). \tag{3}$$

B. UE Phase

The UE phase, given in Subroutine 1, is used for estimation of number of active users in the network. In this phase, users can either be in SH (similar to CE phase) mode or collision sensing (CS) mode. The user can enter into the CS mode only once in the UE phase and only one user can be in the CS mode at a time (lines 4-5). After entering into CS mode (lines 8 – 13), user stops SH and jumps to channel with index 1 (lines 9 – 10). For subsequent $K - 1$ time slots, user transmits on this channel (line 10) and counts the number of SH users colliding with it (line 12). Based on the number of collisions during CS mode, it estimates the number of users, $\hat{N}$ (line 24). After $K - 1$ time slots, the user goes back to SH mode where it continues sequential hopping (lines 15 – 20) and other user enters into the CS mode.

Lemma 3: For any $\delta_3 \in (0,1)$, all the users will have correct estimate of N with probability $\geq 1 - \delta_3$ if they independently run UE phase for T_{UE} slots after the CE phase.

Proof: For correct estimation of N, each user must enter into CS mode and switch back to collision-free SH mode after estimating N. The proposed algorithm guarantees only one user becomes an initiator at a time by exploiting the fact that users are on different channels at the end of CE phase. This channel index is used to assign the entry and exit time slots of the CS mode for each user (lines 4 – 5 in Subroutine 1). The duration of the CS mode for a single user is $K-1$ time slots as it guarantees on collision with each SH user. Since the number of users in the network can be at most K, the duration of UE phase, $T_{UE} \geq K(K-1)$.

Subroutine 1 UE Phase

```
1: Input: ({μ̂_k}, I_{T_CE})
2: Output: N̂
3: Set C_n = 0
4: CS mode begins at t_str = T_CE + 1 + (I_{T_CE} − 1)(K − 1)
5: CS mode ends at t_end = t_str + K − 2
6: for t = T_CE + 1 ... T_CE + T_UE do
7:     if (t == T_CE + 1) then
8:         if (t ≥ t_str and t ≤ t_end) then
9:             Sensing mode, ζ_s = 1
10:            Transmit on channel I_t = 1
11:            if (collision) then
12:                Increment C_n by 1
13:            end if
14:        else
15:            if (ζ_s == 1) then
16:                I_t = 1
17:            else
18:                I_t = I_{t−1} + 1 modulo K
19:            end if
20:            SH mode, ζ_s = 0 and Transmit on channel I_t
21:        end if
22:    end if
23: end for
24: Estimate the number of users, N̂ = C_n + 1
```

C. OR Phase

The OR is the final phase which orthogonalizes users in the top channels. The OR phase takes π and $\hat{N}$ as inputs from the CE and UE phases, respectively and it is similar to CE phase except the user hops over top channels compared to K channels in the CE phase. Each user selects channel uniformly randomly among top channels till it successfully transmits on a channel without incurring collision. After that, user starts SH over top channels using the channel indexes in π. The OR phase continues till the end of horizon.

Lemma 4: For any $\delta_4 \in (0,1)$, all SUs will be orthogonalized in one of the top channels with probability $\geq 1-\delta_4$ if they independently run OR phase for T_{OR} (See Eq. 4) time slots.

Proof: In T_{OR} duration, the users will randomly hop on top $\hat{N}$ channels. Similar to T_{RH}, T_{OR} can easily be given as:

$$T_{OR} = \frac{\log\left(\frac{\delta_4}{\hat{N}}\right)}{\log\left(1-\frac{1}{\hat{N}}\left(1-\frac{1}{\hat{N}}\right)^{\hat{N}-1}\right)}. \tag{4}$$

By replacing $\hat{N}$ with K, we obtain upper bound on T_{OR}.

Theorem 1: For $\delta \in (0,1)$, the expected regret of the proposed algorithm over T rounds in a network of N users and K channels is upper bounded with probability $\geq 1-\delta$ is given as: $R_T \leq N[T_{RH} + T_{SH}\cdot(1-\frac{N}{K}) + T_{UE} + T_{OR}]$, where T_{RH}, T_{SH}, T_{UE} and T_{OR} are given in Lemmas 1, 2, 3 and 4, respectively.

Proof: Let Y denote the intersection of the below 4 events:

$A1$ – users are orthogonalized after T_{RH} time slots

$A2$ – users have ϵ correct ranking of channels after T_{SH} slots

$A3$ – users have correct estimate of N after T_{UE} time slots

$A4$ – users are orthogonalized in top channels in T_{OR} time slots

Using Lemmas 1-4, event Y holds with probability at least

$$\begin{aligned} Pr(Y) &= Pr(A1)Pr(A2|A1)Pr(A3|A2,A1) \\ &\quad \times Pr(A4|A3,A2,A1) \\ &\geq (1-\delta_1)(1-\delta_2)(1-\delta_3)(1-\delta_4) \end{aligned}$$

Setting $\delta_1 = \delta_2 = \delta_3 = \delta_4 = \frac{\delta}{4}$

$$Pr(Y) \geq (1-\delta/4)^4 \geq 1 - 4\frac{\delta}{4} \geq 1-\delta$$

For $t > T_{RH} + T_{SH} + T_{UE} + T_{OR}$, all the users are orthogonalized on the top channels, hence regret is zero with probability at least $1-\delta$. For any $t \leq T_{RH}+T_{SH}+T_{UE}+T_{OR}$, the regret due to each user can be upper bounded by t since regret per user per time slot is at most 1. Hence total regret is bounded by

$$R \leq N[T_{RH} + T_{SH} + T_{UE} + T_{OR}]$$

with probability at least $1-\delta$. Note that during the SH phase, each user spends N/K fraction of slots on the top N channels. Hence, above upper bound can be tightened as:

$$R \leq N[T_{RH} + T_{SH}\cdot(1-\frac{N}{K}) + T_{UE} + T_{OR}].$$

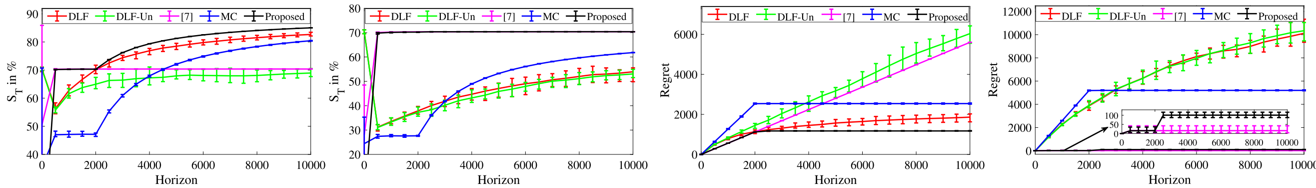
Thus, with an increase in the number of users, N, the regret of the proposed algorithm decreases whereas the regret of the state-of-the-art algorithms [4], [9] increases. We validate the same via simulation results.

Theorem 2: For any $\delta \in (0,1)$, the expected number of collisions incurred in the proposed algorithm over T rounds in a network of N users and K channels is upper bounded as $N\cdot(T_{RH} + T_{UE} + T_{OR})$ with probability $\geq 1-\delta$.

There will be no collision during T_{SH} and after OR phase due to orthogonalization of users in different channels. The proof of the Thm 2 is derived by noting that the collisions occur only in the OR phase and the UE phase. We omit it due to limited space constraints. The number of collisions are significantly lower than existing algorithms and we validate the same via simulation results presented next.

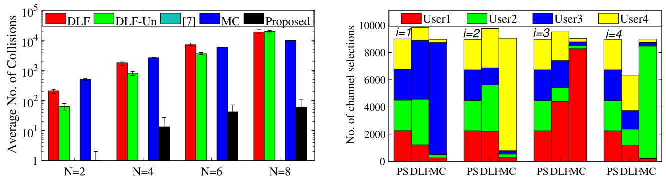
IV. Simulation Results and Analysis

We present the simulation results to compare the performance of the proposed algorithm with Distributed Learning Algorithm with Fairness (DLF) [4], its variant with unknown N (DLF-Un), MC algorithm [9] and algorithm in [7]. We consider $K = 8$ and $\mu_k = \{0.29, 0.36, 0.43, 0.50, 0.57, 0.64, 0.71, 0.78\}$ with the channel statistics of $\mu_{\lceil \frac{K}{2} \rceil} = 0.5$ and the minimum gap between the statistics of the k^{th} and $(k+1)^{th}$ channel, i.e. Δ, is 0.07. All the numerical results are averaged over 50 independent experiments with the horizon of 10000 time slots. The values of T_{CC} in the proposed algorithm and T_O in the MC is 2000.



(a) (b) (c) (d)

Fig. 1. The comparison of vacant spectrum utilization in % with $K = 8$ for (a) $N = 4$, and (b) $N = 8$. The comparison for average cumulative regret for various algorithms with $K = 8$ for (c) $N = 4$, and (d) $N = 8$. Average spectrum utilization should be higher while regret should be lower.



(a) (b)

Fig. 2. (a) Average number of collisions with $N = \{2, 4, 6, 8\}$. (b) Channel allocation fairness comparison for $N = 4$.

The vacant spectrum utilization, S_T in % for $K = 8$ and $N = \{4, 8\}$ is shown in Fig. 1a and 1b. From Fig. 1a, it is evident that the proposed algorithm offers higher number of collision-free transmission opportunities than DLF-Un, MC and [7] algorithms where N is unknown. This is due to proposed collision-free SH approach for N estimation as opposed to RH approach in MC and UCB based approach in DLF. Also, the performance of the proposed algorithm and [7] improves with the increase in N due to fewer number of collisions. This is interesting considering the fact that next generation networks will more likely to be ultra-dense. Due to uniform and collision-free channel selection approach, the performance of algorithm in [7] is identical to proposed algorithm for $N = K$. Note that in Fig. 1b, the plots for DLF and DLF-Un overlap as both are identical since $N = K$. As discussed in Section II, regret (i.e. loss of throughput) is another metric to compare the performance of the distributed algorithm. Fig. 1c and Fig. 1d compare the regret of various algorithms for $K = 8$ and $N = \{4, 8\}$, respectively. It is evident that the proposed algorithm significantly outperforms other algorithms in terms of regret as well. The constant regret after the OR phase confirms the orthogonalization of users in the top channels which is the objective of multi-player MAB framework.

The average number of collisions and hence, the number of retransmissions experienced by all users are compared in Fig. 2a. The y-axis is shown using the logarithmic scale for clarity of plots. It can be observed that the number of collisions are significantly lower in the proposed algorithm with probability of collision of 0.001 because of the use of the collision-free SH approach. Though algorithm in [7] offers almost zero collisions, it assumes either the presence of central controller or sensing hardware. Furthermore, the regret of [7] is significantly poor than the proposed algorithm for small N.

Next, we compare the channel allocations among users in a single experiment. Ideally, each user should get equal opportunity to transmit on the good quality channels in order to guarantee same average throughput for each user. As shown in Fig. 2b, proposed algorithm guarantees fairness in the channel allocation due to the SH over top channels after orthogonalization. In MC and DLF algorithms, such fairness is not guaranteed as some users get locked on one of the top channel after orthogonalization.

From complexity perspective, major tasks are estimation and sorting of channels which are carried out only once at the end of CE Phase. An estimation has computational complexity of at most $O(K)$ while sorting complexity is of order $O(K \log K)$. It is similar to MC [9] and slightly higher than [7]. In DLF [4], both tasks need to be performed in every time slot with additional operations such as square-root, log, multiplication and divisions for UCB. Hence proposed algorithm is computationally efficient and easy to implement.

V. Conclusions and Future Directions

In this work, a novel distributed algorithm for dynamic spectrum access in unlicensed spectrum for infrastructure-less networks is proposed. The theoretical and simulation results show that the proposed algorithm offers better performance in terms of regret (i.e. throughput loss) and number of collisions than state-of-the-art algorithms. Next, we will focus on the distributed algorithm for adversarial networks where channel statistics are non-stationary.

References

[1] M. Ozger *et al.*, "Clustering in multi-channel cognitive radio ad hoc and sensor networks," *IEEE Commun. Mag.*, vol. 56, no. 14, pp. 156–162, Apr. 2018.

[2] R. Kumar *et al.*, "Trekking based distributed algorithm for opportunistic spectrum access in infrastructure-less network," in *Proc. 16th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, Shanghai, China, May 2018, pp. 1–8.

[3] A. Anandkumar, *et al.*, "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.

[4] 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.

[5] M. Zandi *et al.*, " Distributed stochastic learning and adaptation to primary traffic for dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1675–1688, Mar. 2016.

[6] L. Besson and E. Kaufmann, "Multi-player bandits models revisited," in *Proc. 29th Int. Conf. Algorithmic Learn. Theory (ALT)*, Lanzarote, Spain, 2018, pp. 56–92.

[7] G. Zhang *et al.*, "Design and analysis of distributed hopping-based channel access in multi-channel cognitive radio systems with delay constraints," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 2026–2038, Nov. 2014.

[8] O. Avner and S. Mannor, "Concurrent bandits and cognitive radio networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, Nancy, France, Sep. 2014, pp. 66–81.

[9] J. Rosenski *et al.*, " Multi-player bandits—A musical chairs approach," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2016, pp. 155–163.

[10] R. Kumar *et al.*, "Channel selection for secondary users in decentralized network of unknown size," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2186–2189, Oct. 2017.