OPPORTUNISTIC SPECTRUM ACCESS WITH TEMPORAL-SPATIAL REUSE IN COGNITIVE RADIO NETWORKS

Yi Zhang[⋆] *Wee Peng Tay*[⋆] *Kwok Hung Li*[⋆] *Moez Esseghir*[†] *Dominique Gaïti*[†]

[⋆] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[†] Charles Delaunay Institute, University of Technology of Troyes, France

ABSTRACT

We formulate and study a multi-user multi-armed bandit (MAB) problem that exploits the temporal-spatial reuse of primary user (PU) channels so that secondary users (SUs) who do not interfere with each other can make use of the same PU channel. We first propose a centralized channel allocation policy that has logarithmic regret, but requires a central processor to solve a NP-complete optimization problem at exponentially increasing time intervals. To avoid the high computation complexity at the central processor and the need for SU synchronization, we propose a heuristic distributed policy that incorporates channel access rank learning in a local procedure at each SU at the cost of a higher regret. We compare the performance of our proposed policies with other distributed policies recently proposed for opportunistic spectrum access. Simulations suggest that our proposed policies significantly outperform the benchmark algorithms when spectrum temporal-spatial reuse is allowed.

Index Terms— Cognitive radio, spectrum reuse, multi-armed bandit.

1. INTRODUCTION

In cognitive radio networks (CRNs), opportunistic spectrum access (OSA) alleviates the spectrum under-utilization problem. It has been extensively studied at the physical (PHY) and medium access control (MAC) layers, and various temporal [1, 2], spatial [3, 4], or spatial-temporal [5, 6] spectrum-sensing algorithms have been proposed to detect and utilize spectrum opportunities temporally and spatially with acceptable interference to PUs. To study the interactions among SUs in a distributed manner, game theory is widely used to design efficient distributed OSA mechanisms [7, 8]. However, most of these works do not exploit *spatial* spectrum reuse and assume that each SU interferes with every other SU in the CRN. To allow for spatial spectrum sharing amongst the SUs, graphical game algorithms have been proposed to investigate spatial reuse of the spectrum [9, 10]. However, these works assume that some information about PUs like the location or channel usage of PUs is known by all the SUs. Multi-armed bandit (MAB) techniques have been applied for OSA when PU channel information is unknown [7, 11–13]. All these methods assume that all SUs interfere with each other if they use the same channel, and spatial reuse of channels was not addressed. In our previous work [14], a three-step distributed channel allocation policy was proposed to exploit the temporal-spatial reuse of PU channels. However, it requires synchronization amongst the SUs.

If SUs are constrained to using different channels at the same time due to interference between them, as assumed in [11, 15], then the optimal allocation is to assign each SU a different channel with the best availability. However, spatially separated SUs can share the same channel without significant interference with each other. For example, consider the scenario depicted in Figure 1, where the expected network reward is given by the expected total number of interference-free channel uses by the SUs. An edge between two SUs indicates that they interfere with each other. Then, without interfering with each other, SU 1 and SU 4 can reuse the same PU channel with the highest idle probability. The scenario (iii) in Figure 1 achieves the highest expected network reward.

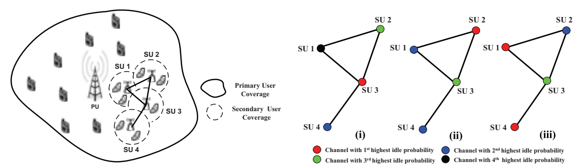

Fig. 1. Spatial spectrum reuse in a CRN with multiple SUs.

In this paper, we propose a centralized policy that uses a central processor to optimize the channel access ranks of the SUs at exponentially increasing time intervals, based on the idle probability estimates of an arbitrary SU. SUs then perform a local random ϵ-greedy channel learning algorithm. We call this the Centralized Channel Allocation (CCA) policy. To overcome the high computation complexity at the central processor in the CCA policy and to avoid the need for SU synchronization in the three-step distributed channel allocation policy we have previously developed in [14], we propose a Distributed Access Rank Learning (DARL) policy that embeds the channel access rank determination in the channel statistics learning process. We compare our policies with the random access policy [15], the time-division fair sharing (TDFS) policy [11] and the adaptive randomization policy [15]. Our simulation results suggest that our policies perform significantly better in terms of average regret than the benchmark policies.

The rest of this paper is organized as follows. In Section 2, we introduce our system model and problem formulation. We propose the Centralized Channel Allocation (CCA) policy in Section 3 and the Distributed Access Rank Learning (DARL) policy in Section 4. In Section 5, we present simulation results and we conclude in Section 6.

2. PROBLEM FORMULATION

Suppose that there are $M \geq 2$ secondary users and N orthogonal channels in a CRN. We model the SU network as a graph $G = (V, E)$, where V is the set of SUs, and E is a set of edges. Two SUs are connected by an edge if the mutual interference between them is above a predefined threshold. If two SUs are not connected via an edge, then we assume that they can utilize the same PU channel simultaneously.

Let $\mathcal{N}$ be the set of channels and $\mathcal{M}$ be the set of SUs. We divide time into equal intervals. In each time slot n, each channel $j \in \mathcal{N}$ is idle with probability $\mu_j \in (0,1)$, independent of all other channels. Without loss of generality, we assume that $\mu_1 > \mu_2 \geq \mu_3 \geq \ldots \geq \mu_N$ (SUs are not aware of this ordering). For each $j \in \mathcal{N}$, we use $S_j(n)$ to denote the channel state of a channel j in time slot n with $S_j(n) = 1$ if the channel j is idle and 0 otherwise.

Without knowing the channel idle probabilities, each SU needs to learn them through their sensing observations. In each time slot n, each SU can only sense and access one channel. Let $X_{i,j}(n) = 1$ if SU i chooses channel j and senses that it is idle, and $X_{i,j}(n) = 0$ otherwise. We assume that channel sensing is perfect for all SUs so that $X_{i,j}(n) = S_j(n)$ if channel j is chosen by SU i at time slot n. Let $Y_{i,j}(n)$ be the reward of a SU i from accessing a channel j in slot n after sensing it free. Let $\mathcal{M}_i$ be the set of neighboring SUs of the SU i in the graph G, not including SU i itself. $Y_{i,j}(n) = 1$ if channel j is idle and no other $k \in \mathcal{M}_i$ transmits over it in the same time slot n and 0 otherwise.

We are interested to design a policy ψ to learn the channel idle probabilities so as to maximize the total expected number of successful transmissions of all SUs by exploiting spatial channel reuse among SUs. The policy ψ is a rule that determines which channel $\psi_i(n) \in \mathcal{N}$ SU i chooses to sense in time slot n. The choice $\psi_i(n)$ can be made based on the sensing results of SU i at previous time slots $1, 2, \ldots, n-1$, and on the previous channel choices $\{\psi(i,l) : \text{ for } l < n\}$. If the channel $\psi_i(n)$ is idle, SU i will transmit over the channel. At the end of each time slot n, each SU is assumed to know whether it has transmitted successfully or not (e.g., through an acknowledgment from the SU receiver). Let $T_{i,j}(n)$ be the total number of time slots that the SU i has sensed the channel j in n time slots, and let $V_{i,j}(n) = \sum_{l \leq n} Y_{i,j}(l)$ be the total number of time slots that the channel j is successfully accessed by SU i up to time slot n.

The regret of the policy ψ until time slot n is defined as the difference between the total reward of a genie-aided rule and the expected reward of all SUs given by

$$R(n,\psi) = n \sum_{i=1}^{M} \mu_{\pi^*(i)} - \sum_{i=1}^{M} \sum_{j=1}^{N} \mu_j \mathbb{E}[V_{i,j}(n)], \qquad (1)$$

where $\pi^* : \{1, \ldots, M\} \mapsto \{1, \ldots, N\}$ is the optimal channel allocation if all channel idle probabilities are known, i.e., $\pi^*(i) = j$ if and only if $x_{ij} = 1$, where x_{ij}, for all $i \in \mathcal{M}$ and $j \in \mathcal{N}$ are the solutions to the following optimization problem:

$$\text{(P0)} \quad \max_{x_{ij}} \quad \sum_{i=1}^{M} \sum_{j=1}^{N} x_{ij} \mu_j \qquad (2)$$

$$\text{s.t.} \quad x_{ij} + \sum_{k \in \mathcal{M}_i} x_{kj} \leq 1, \ \forall i \in \mathcal{M}, j \in \mathcal{N}, \qquad (3)$$

$$\sum_{j=1}^{N} x_{ij} \leq 1, \ \forall i \in \mathcal{M}, \qquad (4)$$

$$x_{ij} \in \{0,1\}, \ \forall i \in \mathcal{M}, j \in \mathcal{N}. \qquad (5)$$

The above maximization problem is an integer linear program, which corresponds to a NP-complete decision problem (of which finding if there exists an independent set in the graph G for a given size is a special case [16]). In general, even if the channel idle probabilities are known a priori, it is analytically difficult for a genie to find the optimal channel allocations. To ensure that optimization is done within a reasonable amount of time, the genie can adopt an approximate method [17], which however leads to a linear regret as the number of time slots $n \to \infty$. For a distributed policy that does not know the channel idle probabilities a priori, the problem is even harder, and in general we cannot hope to learn the channel probabilities and an optimal channel allocation with sub-linear regret, unlike other MAB problems in which logarithmic regrets are common [11, 15].

For any channel allocation $\pi : \{1, \ldots, M\} \mapsto \{1, \ldots, N\}$, we say that $\pi(i)$ is the *channel access rank* of SU i because of the assumption that $\mu_1 > \mu_2 \geq \ldots \geq \mu_N$. Our main idea is to learn the optimal channel access rank of each SU and the idle probability of each channel in order to optimize the regret. In the following, we propose a centralized policy that has asymptotic logarithmic regret, but requires solving an analytically difficult optimization problem like (P0) at exponentially increasing time intervals. We also propose a heuristic distributed policy, which overcomes complexity of the centralized policy but has linear regret in general.

3. CENTRALIZED CHANNEL ALLOCATION (CCA) POLICY

In this section, we propose a centralized policy ψ^{CCA}, and show that it has asymptotic log regret. We assume that there is a central processor in the CRN capable of solving problem (P0) with the true channel idle probabilities μ_j, $j = 1, \ldots, N$, replaced by empirical estimates from an arbitrary SU. We call this optimization problem $\widehat{\text{(P0)}}$. However, since $\widehat{\text{(P0)}}$ corresponds to a NP-complete decision problem, we suppose that the central processor only performs this optimization at specific irregular time instances (see Figure 2) instead of at every time slot. For a time horizon n, let t_k, $k = 1, \ldots, \xi(n)$, be the $\xi(n)$ time instances at which the central processor solves $\widehat{\text{(P0)}}$ with updated empirical estimates of μ_j, $j = 1, \ldots, N$. We assume $t_1 < \infty$, i.e., there is at least one optimization time instance.

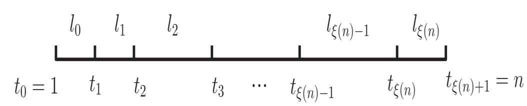

Fig. 2. Optimization time instances for ψ^{CCA}.

For $k = 0, \ldots, \xi(n)$, let $l_k = t_{k+1} - t_k - 1$, where $t_0 = 1$ and $t_{\xi(n)+1} = n$, be the number of time slots starting from the k-th optimization, and before the next optimization by the central processor. Let $\bar{X}_{i,j}(n) = \sum_{k=1}^{n} X_{i,j}(k)/T_{i,j}(n)$ be the empirical estimate of the idle probability of channel j by SU i. At each time instance t_k, $k = 1, \ldots, \xi(n)$, an arbitrarily chosen SU i sends $\{\bar{X}_{i,j}(t_k) : j \in \mathcal{N}\}$ to the central processor, which replaces μ_j with $\bar{X}_{i,j}(t_k)$ in problem (P0), and finds the optimal or near-optimal solution using the branch and bound algorithm [17]. Let $\{r_i(t_k) : i \in \mathcal{M}\}$ be the channel access ranks found by the central processor (i.e., $r_i(t_k) = j$ iff $x_{ij} = 1$ in the solution of $\widehat{\text{(P0)}}$. These ranks are then communicated to the SUs, which utilizes their assigned ranks in a local random ϵ-greedy channel learning algorithm: In each time slot of the channel learning period, each SU i chooses to sense a channel $j \in \mathcal{N}$ with probability ϵ, and with probability $1 - \epsilon$ chooses the $r_i(t_k)$-th best channel according to its empirical idle probability estimates $\{\bar{X}_{i,j}(t_k) : j \in \mathcal{N}\}$. This learning algorithm is an extension of the work in [18]. Let $\Delta_1 = \min_{j=1,\ldots,N-1} |\mu_j - \mu_{j+1}|$

and $\rho_i(n)$ be the channel chosen by SU i in time slot n. We call the above procedure the Centralized Channel Allocation Policy ψ^{CCA}, which is described in Algorithms 1 and 2.

Algorithm 1 Centralized channel allocation policy ψ^{CCA}

Input: SU interference network. At each time $t_k, k = 1, \ldots, \xi(n)$, empirical idle probability estimates $\{\bar{X}_{i,j}(t_k) : j \in \mathcal{N}\}$ from an arbitrarily chosen SU i.
Output: At each time $t_k, k = 1, \ldots, \xi(n)$, channel access ranks $r_i(t_k), i = 1, \ldots, M$
for $n = t_1, t_2, \ldots, t_{\xi(n)}$ **do**
    Central processor chooses an arbitrary SU i, which sends it $\{\bar{X}_{i,j}(t_k) : j \in \mathcal{N}\}$.
    Central processor solves the optimization problem ($\widehat{\text{P0}}$), and for each $i \in \mathcal{M}$, sets $r_i(t_k) = j$ if $x_{ij} = 1$.
    Central processor sends $r_i(t_k)$ to each SU $i, i \in \mathcal{M}$.
    Each SU performs the random ϵ-greedy channel learning algorithm in Algorithm 2.
end for

Algorithm 2 Random ϵ-greedy channel learning at each SU i

Input: $0 < \gamma < \min\{1, \Delta_1\}$, $\delta > \max\{2, 5\gamma^2\}$
for $n \geq 1$ **do**
    Set $\epsilon_n = \min\{1, \frac{\delta N}{\gamma^2 n}\}$.
    Let $t = \max\{t_k, k \geq 1 : t_k \leq n\}$.
    With probability $1 - \epsilon_n$, let $\rho_i(n)$ be a channel with the $r_i(t)$-th highest empirical idle probability estimate (with ties broken randomly), otherwise let $\rho_i(n)$ be chosen uniformly at random from the channel set $\mathcal{N}$.
    if channel $\rho_i(n)$ is sensed to be PU-free **then**
        SU i transmits over channel $\rho_i(n)$ and sets $X_{i,\rho_i(n)}(n) = 1$.
    else
        Set $X_{i,\rho_i(n)}(n) = 0$.
    end if
    Set

$$T_{i,\rho_i(n)}(n) = T_{i,\rho_i(n)}(n-1) + 1,$$
$$\bar{X}_{i,\rho_i(n)}(n) = \frac{\bar{X}_{i,\rho_i(n)}(n-1)T_{i,\rho_i(n)}(n-1) + X_{i,\rho_i(n)}(n)}{T_{i,\rho_i(n)}(n)}.$$

    For all $j \neq \rho_i(n)$, set $T_{i,j}(n) = T_{i,j}(n-1)$ and $\bar{X}_{i,j}(n) = \bar{X}_{i,j}(n-1)$.
end for

The following Theorem 1 shows that the regret using ψ^{CCA} is order-optimal for appropriately chosen optimization time instances. We omit the proof due to space constraints.

Theorem 1 *If $l_k > l_{k-1}$ for $1 < k < \xi(n)$ and $l_k \leq cl_{k-1}$ for all $k \geq 2$ and some $c > 0$, then $R(n, \psi^{CCA}) \in \Theta(\log n)$.*

4. DISTRIBUTED CHANNEL ACCESS RANKING AND LEARNING (DARL) POLICY

In this section, we propose a heuristic distributed policy to perform channel access ranking and learning. We know that finding the optimal channel access ranking is still NP-complete. However, as discussed in [14], the genie-aided channel allocation in (P0) becomes a graph coloring problem in which we wish to partition the graph into disjoint maximal independent sets $I_1, \ldots, I_{\chi(G)}$ (shown in non-increasing order), where $\chi(G)$ is known as the *chromatic number* of the graph G [19]. The SUs assigned to the same independent set are allocated the same channel, with a larger independent set being assigned a channel with a higher idle probability. Therefore, the regret (1) can then be equivalently written as

$$R(n, \psi) = n \sum_{j=1}^{\chi(G)} \mu_j |I_j| - \sum_{i=1}^{M} \sum_{j=1}^{N} \mu_j \mathbb{E}[V_{i,j}(n)],$$

where $|I_j|$ is the number of SUs in the maximal independent set I_j.

To reduce the overhead due to SU synchronization in the three-step distributed channel allocation policy [14], we propose a distributed policy DARL, denoted as ψ^{DARL}, which integrates the first two stages of the distributed policy in [14] into the channel statistics learning process (see Algorithm 3). At the start of DARL, the channel access ranks of SUs $r_i(1)$, $i \in \mathcal{M}$ are all set to be 1. In subsequent time slots $n > 1$, if there is no collision for SU i in the previous time slot, it continues to use the same channel access rank as $r_i(n-1)$. Otherwise, it generates a random number λ_i uniformly distributed in $[0, 1]$ and keeps on using the same channel access rank if λ_i has the largest value among all its neighbors who also have collisions in the previous time slot. If its random number λ_i is not the largest value, SU i is allocated a channel access rank uniformly and randomly from $\{1, \ldots, N\}$. SU i then performs the ϵ-greedy channel learning process. Since there is a higher likelihood for DARL to assign incorrect channel access ranks to the SUs, we expect DARL to have higher regret than CCA, as verified by simulations in Section 5.

The following Proposition 1 shows that the random access policy [15] denoted as ψ^{rand}, the time-division fair sharing (TDFS) policy [11] denoted as ψ^{TDFS}, the adaptive randomization policy [15] denoted as ψ^{adapt} and our proposed policy ψ^{DARL} all have $\Theta(n)$ regret under spatial spectrum reuse on an incomplete graph in general. The proof is omitted due to space constraints.

Proposition 1 *Under spatial spectrum reuse, if the graph G is incomplete and has a connected component of size at least two, then ψ^{rand}, ψ^{TDFS}, ψ^{adapt}, and ψ^{DARL} each has $\Theta(n)$ regret.*

5. SIMULATION RESULTS

In this section, we verify the performance of our proposed policies by simulations in large size random graphs that have $M = 100$ SUs and $N = 100$ orthogonal PU channels. The idle probabilities of the PU channels are $[0.9, 0.8, 0.7, 0.6, 0.5, 0.495, 0.490, \ldots, 0.025]$. For ψ^{CCA}, we let $l_0 = 2$ and $l_k = 2l_{k-1}$ for $k \geq 1$. We set $\delta = 5.1$ and $\gamma = 0.1$. We evaluate the performance of our proposed policies and that of ψ^{rand}, ψ^{TDFS} and ψ^{adapt} on the following:

(i) Erdös-Rényi (ER) graphs: 500 instances of Erdos-Renyi random graphs with M nodes and different probabilities of attachment [20]

(ii) Random connection (RC) graphs: 500 random graphs with M nodes and different number of edges. Edges are generated sequentially, and each edge is formed by choosing two distinct nodes uniformly at random and connecting them if they are not already connected.

We show the regrets in Figure 3 and Figure 4 when the graph is a randomly generated ER graph with attachment probability 0.05 and a RC graph with 200 edges. We compare the average regrets using

Algorithm 3 Distributed access rank learning (DARL) ψ^{DARL}

Input: $0 < \gamma < \min\{1, \Delta_1\}$, $\delta > \max\{2, 5\gamma^2\}$
Initialization: channel access rank $r_i(1) = 1$, for all $i \in \mathcal{M}$.
for $n \geq 1$ **do**
  Set $\epsilon_n = \min\{1, \frac{\delta N}{\gamma^2 n}\}$.
  if there was a collision in previous time slot $n-1$ **then**
    Broadcast $r_i(n-1)$ to $j \in \mathcal{M}_i$.
    Generate a random number λ_i uniformly distributed in $[0,1]$ and broadcasts λ_i to all $j \in \mathcal{M}_i$.
    Let $\bar{\mathcal{M}}_i$ be the set of SUs $j \in \mathcal{M}_i$ that also have collisions in time slot $n-1$.
    if $\lambda_i \geq \max_{j \in \bar{\mathcal{M}}_i} \lambda_j$ **then**
      Set $r_i(n) = r_i(n-1)$.
    else
      Set $r_i(n) = \min \left\{ \{1, \ldots, N\} \backslash \{ r_j(n-1) : j \in \bar{\mathcal{M}}_i \} \right\}$.
    end if
  else
    Set $r_i(n) = r_i(n-1)$.
  end if
  With probability $1-\epsilon_n$, let $\rho_i(n)$ be a channel with the $r_i(t)$-th highest empirical idle probability estimate (with ties broken randomly), otherwise let $\rho_i(n)$ be chosen uniformly at random from the channel set $\mathcal{N}$.
  if channel $\rho_i(n)$ is sensed to be PU-free **then**
    SU i transmits over channel $\rho_i(n)$ and sets $X_{i,\rho_i(n)}(n) = 1$.
  else
    Set $X_{i,\rho_i(n)}(n) = 0$.
  end if
  Set

$$T_{i,\rho_i(n)}(n) = T_{i,\rho_i(n)}(n-1) + 1,$$

$$\bar{X}_{i,\rho_i(n)}(n) = \frac{\bar{X}_{i,\rho_i(n)}(n-1) T_{i,\rho_i(n)}(n-1) + X_{i,\rho_i(n)}(n)}{T_{i,\rho_i(n)}(n)}.$$

  For all $j \neq \rho_i(n)$, set $T_{i,j}(n) = T_{i,j}(n-1)$ and $\bar{X}_{i,j}(n) = \bar{X}_{i,j}(n-1)$.
end for

500 trials for all the policies. We observe that ψ^{CCA} outperforms the policies ψ^{rand}, ψ^{TDFS} and ψ^{adapt}, where the regret of ψ^{CCA} is approximately a constant multiple of $\log n$ and regrets using other policies on both types of random graphs increase linearly over time. We also note that, without SU synchronization, ψ^{DARL} has worse regret than ψ^{CCA}, but still performs better than the other benchmark policies.

6. CONCLUSION

In this paper, we have investigated temporal-spatial channel reuse in cognitive radio networks using a multi-user MAB approach. We have proposed a centralized channel allocation policy for finding an optimal channel allocation and learning the statistics of the channels. To avoid the requirement of centralized processing and synchronization amongst the SUs, we proposed a heuristic distributed policy which let each SU determine their channel access ranks locally. Simulation results suggest that our proposed policies outperform current policies when spatial channel reuse is considered. Future work includes designing policies for mobile SUs where channel availabilities differ across the SU network.

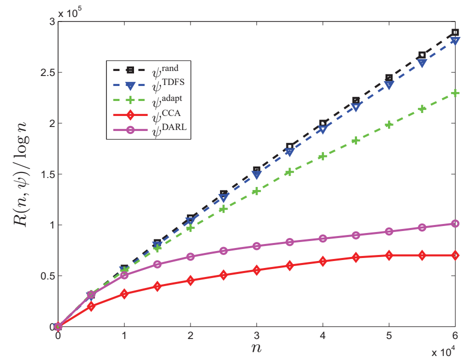

Fig. 3. Normalized regret $\frac{R(n,\psi)}{\log n}$ vs. time slot n on ER graphs.

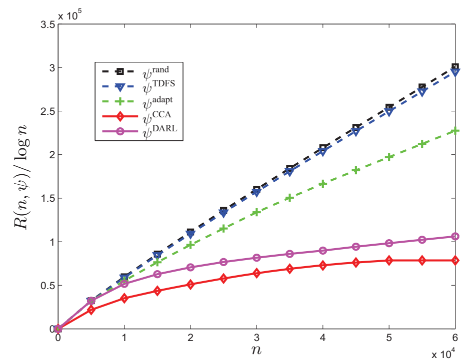

Fig. 4. Normalized regret $\frac{R(n,\psi)}{\log n}$ vs. time slot n on RC graphs.

7. ACKNOWLEDGEMENT

This work was supported in part by the Singapore Ministry of Education Academic Research Fund Tier 2 grants MOE2013-T2-2-006 and MOE2014-T2-1-028.

8. REFERENCES

[1] D. Duan, L. Yang, and J. C. Principe, "Cooperative diversity of spectrum sensing for cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3218–3227, 2010.

[2] G. Ozcan, M. cenk Gursoy, and S. Gezici, "Error rate analysis of cognitive radio transmissions with imperfect channel sensing," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1642–1655, 2014.

[3] S. M. Mishra, *Maximizing available spectrum for cognitive radios*, Ph.D. thesis, UC Berkeley, 2010.

[4] S. M. Yu and S.-L. Kim, "Optimal detection of spatial opportunity in wireless networks," *IEEE Commun. Lett.*, vol. 15, no. 4, pp. 395–397, 2011.

[5] Q. Wu, G. Ding, J. Wang, and Y. Yao, "Spatial-temporal opportunity detection for spectrum heterogeneous cognitive radio

networks: Two-dimensional sensing," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 516–526, 2013.

[6] Y. Zhang, W. P. Tay, K. H. Li, and D. Gaïti, "Distributed boundary estimation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 11, pp. 1–13, 2014.

[7] M. Maskery, V. Krishnamurthy, and Q. Zhao, "Decentralized dynamic spectrum access for cognitive radios: cooperative design of a non-cooperative game," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 459–469, 2009.

[8] 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, 2012.

[9] X. Chen and J. Huang, "Distributed spectrum access with spatial reuse," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 593–603, 2013.

[10] Y. Xu, Q. Wu, L. Shen, J. Wang, and A. Anpalagan, "Opportunistic spectrum access with spatial reuse: graphical game and uncoupled learning solutions," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4814–4826, 2013.

[11] 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, 2010.

[12] M. Zandi and M. Dong, "Learning-stage based decentralized adaptive access policy for dynamic spectrum access," in *IEEE ICASSP*, 2013.

[13] Y. Gai and B. Krishnamachari, "Distributed stochastic online learning policies for opportunistic spectrum access," *IEEE Trans. Signal Process.*, 2014.

[14] Y. Zhang, W. P. Tay, K. H. Li, M. Esseghir, and D. Gaïti, "Distributed opportunistic spectrum access with spatial reuse in cognitive radio networks," in *IEEE GlobalSIP*, 2014.

[15] A. Anandkumar, N. Michael, A. K. 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, 2011.

[16] R. M. Karp, "Reducibility among combinatorial problems," *Complexity of Computer Computations*, pp. 85–103, 1972.

[17] L. A. Wolsey, *Integer Programming*, Wiley-Interscience, 1998.

[18] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2-3, pp. 235–256, 2002.

[19] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*, PHI Learning Pvt. Ltd., 2004.

[20] P. Erdös and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.