2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)
Allerton Park and Retreat Center
Monticello, IL, USA, September 24-27, 2019

A Distributed Stable Strategy Learning Algorithm for Multi-User Dynamic Spectrum Access

Tomer Gafni, Kobi Cohen

***Abstract*— We consider the problem of multi-user dynamic spectrum access (DSA) in cognitive radio networks. The shared bandwidth is divided into K orthogonal channels, and M (secondary) users aim at accessing the spectrum, where $K \geq M$. Each user is allowed to choose a single channel for transmission at each time slot. The state of each channel is modeled by a restless unknown Markovian process. By contrast to existing studies that analyzed a special case of this setting, in which each channel yields the same expected rate for all users, in this paper we consider the more general model, where each channel yields a different expected rate for each user. This general model adds a significant challenge of how to efficiently learn a channel allocation in a distributed manner so as to yield a global system wide objective. We adopt the stable matching utility as the system objective, which is known to yield strong performance in multichannel wireless networks, and develop a novel Distributed Stable Strategy Learning (DSSL) algorithm to achieve the objective. We prove theoretically that the DSSL algorithm converges to the stable matching allocation, and the regret, defined as the loss in total rate with respect to the stable matching solution, has a logarithmic order with time. Finally, we present numerical examples that support the theoretical results and demonstrate strong performance of the DSSL algorithm.**

I. Introduction

Consider the problem of multi-user dynamic spectrum access (DSA) in cognitive radio networks. The shared bandwidth is divided into K orthogonal channels, and M (secondary) users aim at accessing the spectrum, where $K \geq M$. We adopt the stable matching utility as the system objective, which is known to yield strong performance in multichannel wireless networks [1].

The stable matching problem for multi-user DSA was first introduced in [1] under the assumption that the expected rates are known, and a distributed opportunistic CSMA algorithm that solves the problem was proposed. The model with unknown expected rate matrix and rested setting (i.e., the states of the Markovian process do not change if not observed by the user) was studied in [2], [3]. A regret (with respect to the optimal allocation) of near-$O(\log t)$ was achieved. However, the algorithms require an intensive communication between users in order to apply the auction algorithm [4]. In [5], the authors reduces the amount of communication requirements, but without guarantees on the achievable regret. Recently, it was shown in [6] that achieving a sum-regret of near- $O(\log t)$ is possible without communication between users, but only for the case of i.i.d channels. Finally, in this paper we consider the general restless Markovian channel model, in which both used and unused channels change states (see details in Section II), which adds significant challenges in algorithm design and regret analysis.

There exist a number of studies that developed distributed learning algorithms for a special case of the restless Markovian channel model considered in this paper, where each channel yields the same expected rate for all users [7]–[9]. This special case significantly simplifies the channel allocation problem and the analysis (for instance, switching between assigned users does not affect the resulting regret in this special case). In this paper, we consider the general model where each channel yields a different expected rate for each user. This models the situation of a different channel fading across users and channels in spatial distributed networks, and adds a significant challenge of how to learn the desired channel allocation in a distributed manner to achieve a global system wide objective.

Another set of related work on multi-user channel allocation was studied from game theoretic and congestion control ([10]–[19] and references therein), and graph coloring ([20]–[23] and references therein) perspectives. Game theoretic aspects of the problem have been investigated from both non-cooperative (i.e., each user aims at maximizing an individual utility) [11], [12], [16], [18], [24], and cooperative (i.e., each user aims at maximizing a system-wide global utility) [10], [19], [25], [26] settings. Model-free learning strategies were developed in [27]–[29]. Graph coloring formulations have concerned with modeling the spectrum access problem as a graph coloring problem, in which users and channels are represented by vertices and colors, respectively. Thus, coloring vertices such that two adjacent vertices do not share the same color is equivalent to allocating channels such that interference between neighbors is being avoided (see [20]–[23] and references therein for related studies). Finally, all these studies did not consider the problem of achieving provable stable strategies in the learning context under unknown restless Markovian dynamics as considered in this paper.

In this paper, we adopt the stable matching utility as the system objective, which is known to yield strong performance in multichannel wireless networks [1]. We develop a novel Distributed Stable Strategy Learning (DSSL) algorithm to achieve the objective. The DSSL algorithm is very simple for distributed implementation via carrier sensing technology. We prove theoretically that the DSSL algorithm converges to the stable matching allocation, and the regret, defined as the

Tomer Gafni is with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 8410501 Israel. Email:gafnito@post.bgu.ac.il

Kobi Cohen is with the School of Electrical and Computer Engineering, with the Cyber Security Research Center, and with the Data Science Research Center, at Ben-Gurion University of the Negev, Israel. Email: yakovsec@bgu.ac.il

This work was supported by the BGU Cyber Security Research Center under grant 076/16, and by the U.S.-Israel Binational Science Foundation (BSF) under grant 2017723.

978-1-7281-3151-1/19/$31.00 ©2019 IEEE

loss in total rate with respect to the stable matching solution, has a logarithmic order with time. Furthermore, the regret has better scaling with the system parameters as compared to existing approaches. Finally, we present numerical examples that support the theoretical results and demonstrate strong performance of the DSSL algorithm.

II. System Model and Problem Formulation

We consider a cognitive radio network consisting of K orthogonal channels indexed by the set $\mathcal{K} = \{1, 2, ..., K\}$ and M secondary users indexed by the set $\mathcal{M} = \{1, 2, ..., M\}$, where $K \geq M$. The secondary users aim at accessing the spectrum to send their data. Each secondary user is allowed to choose a single channel for transmission at each time slot, and transmit if the channel is not occupied by a primary user (which is represented by a channel state with zero rate). The users are operated in a synchronous time-slotted fashion. Due to spatial geographic dispersion, each secondary user can potentially experience different achievable rates over the channels. When secondary user i transmits on channel k (when the channel is free) at time slot t, its data rate is given by $r_{i,k}(t)$. This information is concisely represented by an $M \times K$ rate matrix $V(t) = \{r_{ik}(t)\}$, $i = 1, ..., M, k = 1, ..., K$.

We consider the case where the rate process $r_{i,k}(t)$ is Markovian and has a well-defined steady state distribution. The transition probabilities associated with the Markov chain are unknown to the users. The process $r_{i,k}(t)$ evolves independently of the user actions (i.e., external process). Furthermore, the channel states might change whether or not being observed (i.e., restless setting). Specifically, the rate of user i on channel k, $r_{i,k}(t)$, is modeled as a discrete time, irreducible and aperiodic Markov chain on a finite-state space $\mathcal{X}^{i,k}$ and represented by a transition probability matrix $P^{i,k} \triangleq (p^{i,k}_{x,x'} : x, x' \in \mathcal{X}^{i,k})$. Let $\vec{\pi}_{i,k} \triangleq (\pi^x_{i,k}, x \in \mathcal{X}^{i,k})$ be the stationary distribution of the Markov chain $P^{i,k}$.

Let $X_{i,k}(t)$ be the actual achievable rate for secondary user i on channel k at time t. If two or more users choose to access the same channel at the same time slot, a collision occurs. In this case, $X_{i,k}(t) = 0$. Otherwise, if user i has accessed channel k without colliding with other users, then $X_{i,k}(t) = r_{i,k}(t)$. The users implement carrier sensing to observe the current channel state at each time slot as typically done in cognitive radio networks [7], [17]. The transmission scheme for the multi-user DSA model is detailed in Section III.

The expectations $\mu_{i,k}$ are given by:

$$\mu_{i,k} = \sum_{x \in \mathcal{X}^{i,k}} x \cdot \pi^x_{i,k},$$

and we define σ_i, for $i = 1, ..., M$, as a permutation of $\{1, \ldots, K\}$ such that

$$\mu_{i,\sigma_i(1)} > \mu_{i,\sigma_i(2)} > \ldots > \mu_{i,\sigma_i(K)}.$$

A. A Stable Channel Allocation

Let $a_i(t) \in \mathcal{K}$ be a selection rule, indicating which channel is chosen by user i at time t, which is a mapping from the observed history of the process (i.e., all past actions and observations up to time $t-1$) to $\{1, ..., K\}$. The expected aggregated data rate for all users up to time t is given by:

$$R(t) = \mathbb{E}[\sum_{n=1}^{t} \sum_{i=1}^{M} X_{i,a_i(n)}(n)]. \quad (1)$$

A policy ϕ_i is a time series vector of selection rules: $\phi_i = (a_i(t), t = 1, 2, ...)$ for user i.

Definition 1 ([1]): A bipartite matching between channels and users is a permutation $P : \mathcal{M} \to \mathcal{K}$. The optimal centralized allocation problem is to find a bipartite matching:

$$\mathbf{k}^{**} = \arg\max_{\mathbf{k} \in P} \sum_{i=1}^{M} \mu_{i,k(i)}.$$

Definition 2 ([1]): A matching $S : \mathcal{M} \to \mathcal{K}$ is stable if for every $i \in \mathcal{M}$ and $k \in \mathcal{K}$ satisfying $S(i) \neq k$, if $\mu_{i,S(i)} < \mu_{i,k}$ then there exists some user $i' \in \mathcal{M}$ such that $S(i') = k$ and $\mu_{i',k} > \mu_{i,k}$.

Achieving the optimal allocation in Definition 1 requires to implement a centralized solution, or a distributed solution with heavy complexity and slow convergence [30]. Therefore, we are interested in developing a distributed algorithm with low complexity that converges to the stable matching solution in Definition 2 which is known to yield strong performance and very fast convergence (when the expected rates are known) by using distributed opportunistic CSMA (see Section III-B and [1] for more details on opportunistic CSMA for stable channel allocation).

We assume that the entries in the matrix U are all different as in [1], which holds in wireless networks due to continuous-valued Shannon rates[1]. Thus, there is a unique stable matching solution under our assumptions, and the expected aggregated rate under the stable matching solution S is given by: $\sum_{m=1}^{M} \mu_{i,S(i)}$. The channel $S(i)$ (i.e., the channel that user i selects under the stable matching configuration) is referred to as the *stable channel selection* of user i.

B. The Objective

Since the expected rates $\mu_{i,k}$ are unknown in our setting, the users must learn this information online effectively so as to converge to the stable matching solution. A widely used performance measure of online learning algorithms is the regret, defined as the reward loss with respect to an algorithm with a side information on the model. In our setting, we define the regret for policy $\phi = (\phi_i, 1 \leq i \leq M)$ as the loss in the expected aggregated data rate with respect to the stable matching solution that uses the true expected rates:

$$r_\phi(t) \triangleq t \cdot \sum_{i=1}^{M} \mu_{i,S(i)} - \mathbb{E}_\phi[\sum_{n=1}^{t} \sum_{i=1}^{M} X_{i,\phi_i(n)}(n)]. \quad (2)$$

A policy ϕ that achieves a sublinear scaling rate of the regret with time (and consequently the time averaged regret tends to zero) approaches the required stable matching solution. The essence of the problem is thus to design an algorithm that learns the unknown expected rates efficiently to achieve the best sublinear scaling of the regret with time.

[1]Otherwise, we can add noise to the matrix.

III. The Distributed Stable Strategy Learning (DSSL) Algorithm

To achieve the objective, as detailed in Section II-B, we divide the time horizon into three phases, namely exploration, allocation, and exploitation phases. These three phases are performed repeatedly during the algorithm according to judiciously designed policy rules, as detailed later.

The purpose of the exploration phase is to allow each user to explore all the channels to identify its M best channels (i.e., the M channels that yield the highest expected rates for the user). The users use the sample means as estimators for the expected rates of the channels to achieve this goal. This phase results in a regret loss, since users access sub-optimal channels to explore them, and the stable allocation is not performed. However, this phase is essential to identify the M best channels and consequently minimize the regret scaling with time. The purpose of the exploitation phase is to use the currently learned information to execute the stable matching solution. The allocation phase is used to allow users to allocate the channels among users properly in a distributed manner using opportunistic carrier sensing [31].

Since the rate process $r_{i,k}(t)$ might evolve even when channel k is not selected by user i, learning the Markovian rate statistics requires using channels in a consecutive manner for a period of time [7], [8]. Moreover, frequent switching between channels would cause a loss due to the transient effect. The high-level structure of the DSSL algorithm works as follows. Each user i computes its required number of samples $N_{i,k}(t)$ for each channel k at the end of every exploitation phase t. If the number of samples is greater than $N_{i,k}(t)$ for all k, user i performs another exploitation phase. Otherwise, if the number of samples is smaller than $N_{i,k}(t)$ for one or more channel, user i carries out an exploration phase for those channels. When no exploration phase is needed, an allocation phase is being performed. At the end of the allocation phase, each user identifies its stable channel selection, and an exploitation phase is carried out. We now discuss the structure of the DSSL algorithm in details.

A. The Structure of the Exploration Phase:

Let $n_O^{i,k}(t)$ be the number of exploration phases in which channel k was selected by user i up to time t. Each exploration phase is divided into two sub epochs: a Random size Epoch (RE), and a Deterministic size Epoch (DE). Let $\gamma^{i,k}(n_O^{i,k}(t)-1)$ be the last channel state observed at the $(n_O^{i,k}(t)-1)^{th}$ exploration phase. RE starts at the beginning of the exploration phase until state $\gamma^{i,k}(n_O^{i,k}(t)-1)$ is observed. This epoch ensures that the generated sample path (after removing the samples observed in the RE epochs) is equivalent to a sample path which are generated by continuously sensing the Markovian channel without switching. This step guarantees a consistent estimation of the expected rates. Then, DE starts by sensing the channel for a deterministic period of time $4^{n_O^{i,k}(t)}$. The deterministic period of time grows geometrically with time to ensure a relatively small number of channel switching.

B. The Structure of the Allocation Phase:

The allocation phase applies opportunistic CSMA among users. In opportunistic CSMA, the backoff function maps from an index (i.e., expected rate) to a backoff time [31]. The backoff function is monotonically decreasing with the rates, so that the user with the highest rate on a certain channel waits the minimal time before transmission. All other users sense that the channel is occupied and do not transmit on that channel. To obtain the stable matching allocation, this procedure continues until all M users occupy M channels. For more details on opportunistic CSMA for stable matching see [1].

We now describe the structure of the allocation phase. Let $\mathcal{T}_k$ be the set of all users that attempt to transmit on channel k at a certain stage of the allocation phase. We initialize the phase by declaring each user to be *unassigned*. We divide the time horizon of the allocation phase into two sub-phases. In the first sub-phase, referred to as S_1, we perform opportunistic CSMA for stable matching as in [1], while replacing the expected rates by the sample means. Specifically, each unassigned user attempts to transmit on its best channel, out of those it has not yet attempted using opportunistic CSMA. On each channel k, the best user out of $\mathcal{T}_k$ in this sub-phase (S_1) is declared to be assigned. All the other users in $\mathcal{T}_k$ store the sample mean of the assigned user (by mapping from the sensed backoff time to the sample mean). This sub-phase continues until all M users are assigned to M channels. The second sub-phase, referred to as S_2, is used for gaining a side information required for efficient learning. Specifically, the opportunistic CSMA is executed again, but the assigned users of each channel do not transmit. All other users that attempted to transmit in S_1 transmit again on the same channel k. The sample mean of the best user in S_2 (i.e., the second best user in $\mathcal{T}_k$ for each channel k) is stored by the assigned user. This sub-phase continues until all M users in S_2 were observed, and the phase ends.

C. The Structure of the Exploitation Phase:

Let $n_I(t)$ be the number of exploitation phases up to time t. In the exploitation phase, each user transmits on the channel it was assigned according to the last allocation phase (during S_1) for a deterministic period of time $2\cdot 4^{n_I(t)-1}$ (for the n_I^{th} exploitation phase). There are no channel switching and no sample mean updating during the exploitation phase.

D. Parameter Setting for Efficient Learning:

As discussed earlier, exploring the channels increases the regret since the stable matching allocation is not being used. On the other hand, it is essential to reduce the estimation error and consequently reduce the regret scaling order with time. In this section, we establish the sufficient exploration rate of each channel for each user to achieve efficient learning of the stable matching allocation. For the ease of presentation we assume that $\sigma_i(k)=k$, $\forall k\in\mathcal{K},\forall i\in\mathcal{M}$. We next establish two parameters used in the learning strategy (a detailed explanation of the parameter setting can be found in the extended version of this paper [32]).

1) Identifying M best channels: We show in the extended version of this paper [32] that a user (say user i) who is interested in distinguishing with a sufficiently high accuracy between two channels k, l that yield expected rates $\mu_{i,k}, \mu_{i,l}$, respectively, must explore them at least $\frac{4L}{(\mu_{i,k}-\mu_{i,l})^2}\cdot\log(t)$ times (where L is a constant which depends on the systems parameters). Let $\Delta_{k,l}^{(i)} \triangleq \mu_{i,k}-\mu_{i,l}$, and let $\mathcal{M}_i$ be the set of the M best channels of user i. For each channel $k\in\mathcal{M}_i$ we define the deterministic row[2] threshold as

$$D_{i,k}^{(R)} \triangleq \frac{4L}{\min\{(\Delta_{k,k+1}^{(i)})^2,(\Delta_{k,k-1}^{(i)})^2\}}, \tag{3}$$

and for channel $k\notin\mathcal{M}_i$,

$$D_{i,k}^{(R)} \triangleq \frac{4L}{(\Delta_{k,M})^2}. \tag{4}$$

Since the expected rates are unknown, we develop the estimate $\widehat{D}_{i,k}^{(R)}(t)$ of $D_{i,k}^{(R)}$ in [32], which guarantees the desired convergence.

2) CSMA protocol identification: In accordance with the opportunistic CSMA protocol described above, each user i needs to distinguish between channels $k\in\mathcal{T}_k$ (these channels are in $\mathcal{M}_i$ as well), and the best channel in $\mathcal{T}_k$ (and the second best channel in $\mathcal{T}_k$ if k is the best channel in $\mathcal{T}_k$). Hence, we define the deterministic column threshold for channels $k\in\mathcal{T}_k$ by:

$$D_{i,k}^{(C)} \triangleq \frac{4L}{(\mu_{i,k}-\max_{j\neq i}\mu_{j,k})^2}, \tag{5}$$

and its estimate by $\widehat{D}_{i,k}^{(C)}(t)$ (see [32]). The adaptive threshold rate of user i for channels $k\in\mathcal{M}_i\cap\mathcal{T}_k$ is given by:

$$\widehat{D}_{i,k} = \max\{\widehat{D}_{i,k}^{(R)}(t),\widehat{D}_{i,k}^{(C)}(t)\}. \tag{6}$$

E. Choosing between phases types:

At the end of the exploitation phases, the users check the condition:

$$T_{i,k}^{(O)}(t) > \max\left\{\widehat{D}_{i,k},\frac{2}{I}\right\}\cdot\log(t), \tag{7}$$

where I can be viewed as the rate function of the estimators among all channels, required to guarantee the desired convergence rate (see [32]). If the condition holds for user i, the user enters another exploitation phase by transmitting on the same channel it has transmitted in the last exploitation phase. Otherwise, if the condition does not hold, the user enters an exploration phase by sensing channel k. At the end of the phase, the user signals the other users that it has finished the exploration phase. If such an interruption occurred, all the users check again condition (7). If it holds for all users, they start an allocation phase. At the end of the allocation phase, an exploitation phase starts. A detailed pseudocode of the DSSL algorithm is provided in [32].

In the extended version of this paper, we establish a finite-sample bound on the regret with time, which results in a logarithmic scaling of the regret [32].

[2]This definition is consistent with the definition of the $M\times K$ expected rate matrix by $U=\{\mu_{ik}\}$, $i=1,...,M, k=1,...,K$.

F. Numerical Examples:

In this section, we analyze the performance of DSSL numerically as compared to DSEE [8] and RCA [7]. The RCA and DSEE algorithms were proposed to solve the special case of our problem, i.e., when each channel yields the same expected rate for all users. For the simulation comparison, we extended the RCA and DSEE algorithms by replacing their parameters with the corresponding general matrix parameters. The RCA algorithm performs random regenerative cycles until catching predefined states in each phase, which results in oversampling the channels, and therefore is expected to increase the regret as compared to DSSL. The DSEE algorithm overcomes this issue by performing deterministic sequencing for both exploration and exploitation phases. However, the deterministic sequencing requires the algorithm to explore all channels using the maximal exploration rate among all channels, which is expected to increase the regret as compared to DSLL (that learns the desired exploration rate for each channel) as well. It can be seen in Fig. 1 that the DSSL algorithm outperforms both RCA and DSEE in this case.

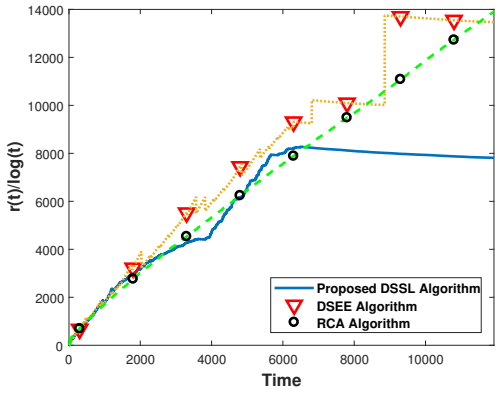

Fig. 1. The regret (normalized by $\log t$) under DSSL, extended DSEE, and extended RCA as a function of time. Parameter setting: 3 users, 6 channels, each with two states: 0, 1. Transition probabilities for all channels to transit from 0 to 1 and from 1 to 0: $[0.2, 0.3, 0.35, 0.38, 0.42, 0.46]$, Expected rates for channels at states 1 for users 1,2,3, respectively: $r_{1,k} = [7, 15, 3, 2.6, 2.2, 1.8]$, $r_{2,k} = [5, 7, 13, 2.5, 2.1, 1.7]$,$r_{3,k} = [11, 1.2, 19, 2.4, 2, 1.6]$, The expected rate for all channels at states 0 is $r = 1$ for all users.

IV. Conclusion

We considered the problem of multi-user DSA in cognitive radio networks. The state of each channel is modeled by a restless unknown Markovian process. By contrast to existing studies that analyzed a special case of this setting, in which each channel yields the same expected rate for all users, here each channel yields a different expected rate for each user. This general model adds a significant challenge of how to efficiently learn a channel allocation in a distributed manner so as to yield a global system wide objective. We developed a novel Distributed Stable Strategy Learning (DSSL) algorithm to achieve the objective, and proved theoretically convergence to the stable matching allocation with a logarithmic regret order. Numerical examples supported the theoretical findings and demonstrated strong performance of the DSSL algorithm.

REFERENCES

[1] A. Leshem, E. Zehavi, and Y. Yaffe, "Multichannel opportunistic carrier sensing for stable channel access control in cognitive radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 82–95, 2012.

[2] D. Kalathil, N. Nayyar, and R. Jain, "Decentralized learning for multiplayer multiarmed bandits," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2331–2345, 2014.

[3] N. Nayyar, D. Kalathil, and R. Jain, "On regret-optimal learning in decentralized multiplayer multiarmed bandits," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 597–606, 2016.

[4] D. P. Bertsekas, "The auction algorithm: A distributed relaxation method for the assignment problem," *Annals of operations research*, vol. 14, no. 1, pp. 105–123, 1988.

[5] O. Avner and S. Mannor, "Multi-user lax communications: a multi-armed bandit approach," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, IEEE, 2016.

[6] I. Bistritz and A. Leshem, "Distributed multi-player bandits-a game of thrones approach," in *Advances in Neural Information Processing Systems*, pp. 7222–7232, 2018.

[7] C. Tekin and M. Liu, "Online learning of rested and restless bandits," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5588–5611, 2012.

[8] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1902–1916, 2012.

[9] T. Gafni and K. Cohen, "Learning in restless multi-armed bandits using adaptive arm sequencing rules," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, pp. 1206–1210, Jun. 2018.

[10] Z. Han, Z. Ji, and K. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, 2005.

[11] I. Menache and N. Shimkin, "Rate-based equilibria in collision channels with fading," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1070–1077, 2008.

[12] U. O. Candogan, I. Menache, A. Ozdaglar, and P. A. Parrilo, "Competitive scheduling in wireless collision channels with correlated channel state," in *Game Theory for Networks, 2009. GameNets' 09. International Conference on*, pp. 621–630, 2009.

[13] I. Menache and A. Ozdaglar, "Network games: Theory, models, and dynamics," *Synthesis Lectures on Communication Networks*, vol. 4, no. 1, pp. 1–159, 2011.

[14] L. M. Law, J. Huang, and M. Liu, "Price of anarchy for congestion games in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3778–3787, 2012.

[15] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "Fasa: Accelerated S-ALOHA using access history for event-driven M2M communications," *IEEE/ACM Transactions on Networking (TON)*, vol. 21, no. 6, pp. 1904–1917, 2013.

[16] C. Singh, A. Kumar, and R. Sundaresan, "Combined base station association and power control in multichannel cellular networks," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 1065–1080, 2016.

[17] K. Cohen, A. Leshem, and E. Zehavi, "Game theoretic aspects of the multi-channel ALOHA protocol in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 2276–2288, 2013.

[18] K. Cohen and A. Leshem, "Distributed game-theoretic optimization and management of multichannel aloha networks," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1718–1731, 2016.

[19] K. Cohen, A. Nedić, and R. Srikant, "Distributed learning algorithms for spectrum sharing in spatial random access wireless networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 2854–2869, 2017.

[20] W. Wang and X. Liu, "List-coloring based channel allocation for open-spectrum wireless network," *In proc. of IEEE Vehic. Tech. Conf.*, 2005.

[21] J. Wang, Y. Huang, and H. Jiang, "Improved algorithm of spectrum allocation based on graph coloring model in cognitive radio," in *WRI International Conference on Communications and Mobile Computing*, vol. 3, pp. 353–357, 2009.

[22] A. Checco and D. J. Leith, "Fast, responsive decentralised graph colouring," *arXiv preprint arXiv:1405.6987*, 2014.

[23] A. Checco and D. Leith, "Learning-based constraint satisfaction with sensing restrictions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 811–820, Oct 2013.

[24] H. Cao and J. Cai, "Distributed opportunistic spectrum access in an unknown and dynamic environment: A stochastic learning approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4454–4465, 2018.

[25] A. Leshem and E. Zehavi, "Bargaining over the interference channel," in *IEEE International Symposium on Information Theory*, pp. 2225–2229, 2006.

[26] I. Bistritz and A. Leshem, "Approximate best-response dynamics in random interference games," *IEEE Transactions on Automatic Control*, vol. 63, no. 6, pp. 1549–1562, 2018.

[27] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for dynamic spectrum access in multichannel wireless networks," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, 2017.

[28] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 310–323, 2018.

[29] T. Sery and K. Cohen, "On analog gradient descent learning over multiple access fading channels," *arXiv preprint arXiv:1908.07463*, 2019.

[30] O. Naparstek and A. Leshem, "Fully distributed optimal channel assignment for open spectrum access," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 283–294, 2013.

[31] Q. Zhao and L. Tong, "Opportunistic carrier sensing for energy-efficient information retrieval in sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 2, pp. 231–241, 2005.

[32] T. Gafni and K. Cohen, "Distributed stable strategy learning for multi-user dynamic spectrum access," *Available at arXiv*, 2019.