Learning to Thrive in a Leasing Market: An Auctioning Framework for Distributed Dynamic Spectrum Leasing (D-DSL)

Sudharman K. Jayaweera
School of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM, USA
Email: jayaweera@ece.unm.edu

Mario Bkassiny
School of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM, USA
Email: bkassiny@ece.unm.edu

***Abstract*—Recently proposed dynamic spectrum leasing (DSL) for dynamic spectrum sharing (DSS) presumes that there is a reward for primary users for accepting secondary activity. The reward primary users received in previous DSL proposals, however, were considered to be proportional to the secondary interference level. In this paper, we propose a new way to encourage primary users to lease their spectrum: The secondary users indicate their willingness to spend a portion of their transmit powers to relay the primary signals. Thus, the primary reward is the power saving achieved due to asymmetric cooperation. We develop a repeated first-price-auction model to realize the proposed DSS scheme in a completely distributed implementation. A reinforcement learning algorithm is proposed for allowing all users to learn their bids in subsequent auctions. Our results show that the proposed Distributed-DSL framework leads to efficient spectrum utilization and, in particular, the proposed learning helps to improve secondary throughput.**

***Index Terms*—Cognitive radios, cooperative relaying, DSL, distributed dynamic spectrum leasing, dynamic spectrum access, dynamic spectrum sharing, game theory, MAC protocols.**

I. INTRODUCTION

Cognitive Radios (CR's) were envisioned in [1] as radio devices that are able to learn and adapt to their RF environment. These characteristics make the CR's a suitable platform for achieving dynamic spectrum sharing (DSS). The proposals for DSS can be identified under three major categories: Dynamic Exclusive Use, Open Sharing and Hierarchical access [2]. The Hierarchical access model, which arguably has gained the most attention in recent years, can be realized under three paradigms: the *underlay*, the *overlay* and the *interweave* [3]. Spectrum underlay allows secondary users (SU's) to transmit simultaneously with primary users (PU's) as long as the interference caused to the primary users is below a certain level that is determined by the Quality of Service (QoS) requirements of primary users. Spectrum interweaving, also known as the Opportunistic Spectrum Access (OSA), consists of exploiting the so-called spectral holes. The OSA relies on spectrum sensing by the SU's to correctly locate the spectrum holes and to minimizing the potential collision with primary users [4]. All these hierarchical DSS proposals are broadly termed as *dynamic spectrum access* (DSA).

In [5], [6] the authors introduced the concept of *dynamic spectrum leasing* (DSL) as a new paradigm for DSA in cognitive radio networks. They were motivated by the observation that the passive participation, or the non-participation, of PU's in the DSS process as assumed in DSA proposals is inefficient in terms of fully utilizing the spectrum. This is because, in DSA, the SU's are *completely* responsible for managing the spectrum sharing process while not compromising the primary Quality-of-Service (QoS). The PU's do not have a stake in the process, and thus act completely oblivious to the existence of the SU's as well as the on-going DSS coexistence. On the other hand, in the DSL framework proposed in [5], [6], the primary users are allowed to proactively manage the amount of secondary activity in their licensed spectrum band. The DSL presumes that there is a *reward* for primary users for accepting secondary activity whenever it is affordable without compromising their own Quality-of-Service (QoS).

Previous DSL proposals, however, have focused only on spectrum underlay architectures. Thus, the reward primary users received were considered to be a monetary reward proportional to the amount of secondary interference they are willing to tolerate. Furthermore, the DSL in [5]–[8] assumed a centralized architecture: The primary users essentially act as a single system, although secondary users are allowed to be autonomous. In this paper, we propose a completely new way to encourage primary users to lease their spectrum whenever affordable: Rather than a monetary reward, in the proposed framework the primary reward is in terms of savings on their communication resources, namely the power. This is achieved by proposing an *asymmetric cooperative communications* architecture between primary and secondary users [9]. In [10] asymmetric cooperation was introduced to cognitive radio networks such that SU's relay the primary signals in return for being able to simultaneously transmit their own signals. However, the method in [10] was completely centralized, has considerably high computational complexity and no real cognitive capability such as learning was involved.

In the proposed framework, the secondary users indicate to the primary users their willingness to spend a portion of their transmit powers to relay the primary signals to their destinations, in order to gain access to the primary spectrum. In this formulation, the primary reward is the power saving they achieve due to cooperative relaying by SU's. We formulate a completely distributed DSL framework in which each *primary link* (i.e. transmitter and a receiver pair) and each secondary user acts autonomously. This leads to a new *distributed dynamic spectrum leasing* (D-DSL) architecture that is most suitable for future heterogeneous wireless network scenarios. The transmitter-receiver pair (TX-RX pair) of each primary link works together by specifying how much bandwidth it is willing to lease at a given time. Each secondary user places a set of bids for each primary link indicating how much power it is willing to spend for relaying that link's signals. Receivers of each primary link picks the bid that will lead to the greatest expected power savings as the winning bid for that link (first-price auction) and starts cooperative communications. The advantages of the proposed first-price auction auctioning based D-DSL framework includes that it is applicable to both spectrum interweave and underlay architectures. The whole proposed framework, including asymmetric cooperative relaying, can be implemented with very little inter-system information exchange.

We believe that sophisticated and autonomous learning to be the defining feature of future cognitive radios. In our proposal, the SU's who do not win a favorable channel at the beginning of a dynamic spectrum auction will employ cognitive learning to win a bid for a channel in subsequent bidding times. Each winning secondary node (one per each primary channel) also uses learning to revise its bid in subsequent bidding times to improve its own power savings. They all make use of the knowledge obtained from past experience in participating in the spectrum leasing auction to learn to come up with better future bids. In this paper, we develop a simple, yet robust, Reinforcement Learning (RL) mechanism to achieve such distributed and autonomous learning from the past experience without any supervision. The reinforcement learning approach permits the cognitive agents to learn by interacting with their environment either in centralized or decentralized mode [11]. Recently, there has been a great interest in applying the RL and other machine learning concepts to CR's [12], [13]. In this paper, we show that without any centralized control both primary and secondary radios can autonomously learn to arrive at an equilibrium. We provide performance results that show how effective the auction-based D-DSL framework in utilizing the spectrum resources as well as the significant impact of reinforcement learning in both improving overall spectrum utilization and meeting individual secondary user performance requirements.

The rest of this paper is organized as follows: Section II defines the system model. Section III details the proposed first-price-auction D-DSL based on asymmetric cooperative relaying. In Section IV we show the simulation results, and finally, Section VI concludes the paper by summarizing our results.

II. SYSTEM MODEL

The centralized DSL (C-DSL) [7], [8] assumes that all primary and secondary users coexist in the whole spectrum band of interest. However, in almost all wireless systems the total spectrum is usually divided into a multiple number of (primary) channels. A channel allocation scheme either dynamically or statically allocates these channels to primary users, as needed. In the following we assume that there are L primary users/links on L distinct primary channels. Thus, we will use the terms primary user, primary channel and primary link interchangeably. To be general, let us assume that the allocated bandwidth of channel $i \in \mathcal{C}$ is W_i, where $\mathcal{C} = \{1, \cdots, L\}$. At this point it is perhaps worth mentioning that, these channels do not have to be necessarily frequency channels. They could also be TDMA channels, in which case the channel resource would be the time slot length T_i of channel/user i. Also the proposed D-DSL architecture can be adapted for an OFDMA-based primary system, in which each primary user is allocated a set of OFDMA sub-carriers as shown in Fig. 1. Then, the i-th primary user can be assumed to be allocated an L_i number of sub-carriers. In the following, to save space, we will always discuss things in the context of primary channels being distinct FDMA channels. However, the method and the proposed framework can easily be adapted to either TDMA or OFDMA contexts with necessary changes by redefining the spectrum resource as above. For simplicity, in this paper, we will also assume that each SU has the capability to transmit only on one channel at a time, and that each primary TX-RX pair (link) is allowed to be leased to one secondary user at a time.

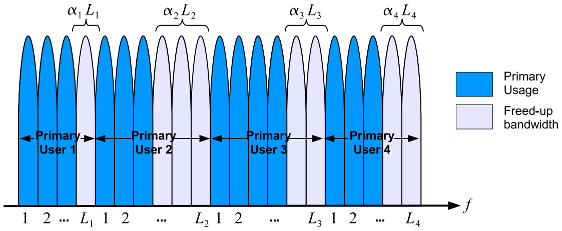


Fig. 1. Distributed dynamic spectrum leasing (D-DSL) in an OFDMA-based wireless network. Each user/link dynamically decides to lease an α_i fraction of its allocated sub-carriers.

The time horizon is assumed to be split into different time frames of duration T_f each by the primary system, and each time frame is divided into a number of equal-length time slots. We assume that the channel fading varies slowly within a time frame, and thus fading can be considered constant in this time duration. Suppose that the maximum transmit power of i-th primary user is $\bar{P}_i$. As required QoS, the RF interference and the observed channel fading (state) conditions change from one time frame to another, the i-th primary user may be able to achieve its required QoS by using only $(1 - \alpha_i)$ fraction of its allocated bandwidth W_i, for $\alpha_i \in [0, 1]$. This is the origin of the so-called spectrum holes and the resulting spectrum under-utilization. In existing proposals for dynamic spectrum sharing (DSS) based on DSA, the primary users do not pay any attention to this phenomenon, and the SU's are expected to sense the spectrum and detect these opportunities: Whichever the SU that successfully detects these seemingly random spectrum holes will get to access them, perhaps on contention-basis. Certainly there is no reason for the primary users to pay any attention to who accesses these white spaces, according to existing DSA proposals, because they do not have anything to gain. By default, in DSA proposals then the focus is on just utilizing the spectrum holes rather than efficient utilization of spectrum holes.

In contrast, according to our proposed D-DSL framework, if at the beginning of a time frame the i-th primary user determines that it can achieve its required QoS only by using $(1-\alpha_i)$ fraction of its bandwidth W_i (or sub-carriers L_i, or time slot T_i), then it consciously decides to free-up up to an α_i fraction of its bandwidth W_i for secondary users to lease. Note that, if there is frequency selective channel fading across the bandwidth W_i, then the i-th primary user will have the freedom to decide which parts of its allocated bandwidth to be freed-up. Although this may be an important aspect in practice, to avoid notational complexity in introducing the basic idea of D-DSL based on asymmetric cooperative communications and first-price auctioning, in this paper we will assume that each primary channel is frequency flat. Thus, to be concrete, we may assume that always the last $\alpha_i W_i$ portion of each channel will be freed-up. However, individual channels are completely independent in their choices of α_i's.

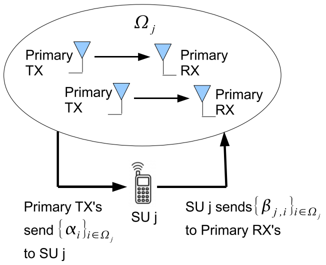


Fig. 2. Distributed dynamic spectrum leasing (D-DSL) based on first-price-auction

Suppose there are K_s number of secondary users, each with maximum transmit power $\bar{P}_j$ for $j \in \mathcal{K}_s$ where $\mathcal{K}_s = \{1, \cdots, K_s\}$. At the beginning of each time frame, each secondary user $j \in \mathcal{K}_s$ then receives all α_i's from $i \in \Omega_j$ where $\Omega_j \subseteq \mathcal{C}$ denotes the set of neighboring primary channels (i.e. the primary channels that can be sensed) of the j-th secondary user, as shown in Fig. 2. Note that the Ω_j sets are not necessarily disjoint. From the perspective of the j-th secondary user, which primary user a particular channel belongs to is irrelevant. All it cares about is what channels are available to transmit, and to what extent, and the quality of (i.e. CSI) those channels with respect to j-th SU's intended receiver. Assuming it has this CSI, the j-th secondary user then decides on the portion $\beta_{j,i}$ of its power $\bar{P}_j$, for $\beta_{j,i} \in [0,1]$ and $i \in \Omega_j$, that can be allocated to relay the primary signals.

All secondary users then place their computed $\beta_{j,i}$ values as bids to each of the primary links. Receivers of each primary link will be responsible for determining the winning bid for that channel, such that it maximizes some primary utility function. If there are any ties among secondary users for winning a particular channel, then the corresponding primary user will randomly pick one of them. Each primary user then informs its chosen winning secondary node of its bid being successful. In some cases, a particular secondary user's bids may be selected by more than one primary channel as winning bids. Then this secondary user, of course, decides to accept the invitation to cooperate with the primary channel that permits it to achieve the largest secondary rate. Once the bid selection is done, then the primary and winning secondary users start to transmit.

III. ASYMMETRIC COOPERATIVE COMMUNICATIONS BASED DISTRIBUTED DYNAMIC SPECTRUM LEASING

The assumed asymmetric cooperative system is depicted in Fig. 3: The secondary user $j \in \mathcal{K}_s$ *coherently* relays the signal of the primary user $i \in \mathcal{C}$ over a link with a fading coefficient $h_{j,i}$. For the sake of illustrating the method, we assume a genie-aided cooperation so that the secondary relay knows the primary message to be relayed instantaneously. Hence, the relayed signal is transmitted over the bandwidth $(1-\alpha_i)W_i$ that the primary user uses for its own transmission. Note that, this assumption can be easily dropped by adapting a practical cooperative protocol at the expense of more elaborate notation. We denote by h_i the fading coefficient between the primary TX i and the corresponding RX, and $h_j^{(i)}$ the fading coefficient between the secondary TX j and its corresponding secondary RX, when transmitting over channel i.

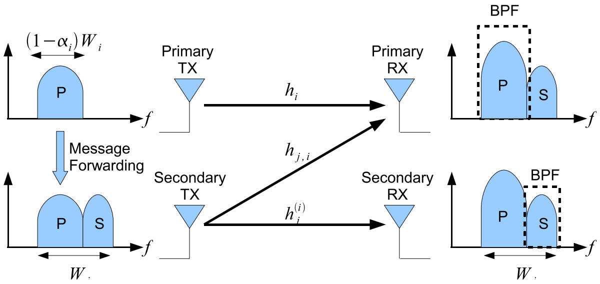


Fig. 3. Asymmetric cooperative communication is achieved by the primary users with the help of secondary relays.

A. Actions of Primary TX: Determination of How Much Spectrum to Lease

Let us denote by $j(i)$ the index of the secondary user who wins the bidding auction for the channel i. If the primary i does not reach a leasing agreement with any secondary user we will let $j(i) = 0$ (0 denoting a dummy SU). Suppose that the i-th primary user needs a minimum data rate of $R_i^{(\min)}$ on its link. The objective of primary users in the proposed distributed DSL framework is to minimize their own power expenditure by exploiting cooperative communications facilitated by secondary users. This objective is achieved by maximizing the following primary utility function:

$$u_i\left(\alpha_i, \beta_{j(i),i}\right) = \frac{\bar{P}_i - P_i(\beta_{j(i),i})}{\bar{P}_i} q\left(R_i(\alpha_i) - R_i^{(\min)}\right) \quad (1)$$

where $P_i(\beta_{j(i),i})$ is the primary i's transmit power with $P_i(\beta_{0,i}) = \bar{P}_i$ indicating that if primary i does not reach an agreement with any SU then it will be transmitting at its maximum transmit power, and

$$q(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

is the unit step-function.

While transmitting at its nominal power level of $\bar{P}_i$, the rate it can achieve by freeing up α_i fraction of its bandwidth for leasing is

$$R_i(\alpha_i) = (1-\alpha_i)W_i \log\left(1 + \Gamma_i(\alpha_i)\right) \quad (3)$$

where $\Gamma_i(\alpha_i) = \frac{|h_i|^2 \bar{P}_i}{(1-\alpha_i)W_i N_i}$ is the resulting SNR. Suppose that with the current realization of CSI h_i on the primary link, it can achieve a rate of $R_i^{(\min)}$ with only using a minimum of $(1-\alpha_i^{(\max)})$ fraction of its bandwidth (if it transmits at its nominal transmit power $\bar{P}_i$). Then, the primary i may free up $\alpha_i \in [0, \alpha_i^{(\max)}]$ fraction of its spectrum resource without degrading its QoS requirement. Note that, ideally it can set $\alpha_i = \alpha_i^{(\max)}$. However, by letting $\alpha_i \in [0, \alpha_i^{(\max)}]$ we allow the primary i to be more flexible in its choice (for example to leave out a safety margin).

B. Secondary Users Action: Determination of How Much Power To Be Spent for Cooperation

Each secondary user $j \in \mathcal{K}_s$ receives all α_i's from $i \in \Omega_j$ where $\Omega_j \subseteq \mathcal{C}$ is its neighboring primary TX set, as shown in Fig. 2, and computes a set of bids $\beta_{j,i}$'s for all $i \in \Omega_j$ so that $\beta_{j,i} \in [0, \beta_{j,i}^{(\max)}]$ where $\beta_{j,i}^{(\max)}$ is the maximum power it can allocate to relaying the i-th primary user signal while maintaining a minimum Bit Error Rate (BER) of ϵ with respect to its own receiver over the channel i (the portion $\alpha_i W_i$): i.e.

$$\beta_{j,i}^{(\max)} = \arg \max_{\beta_{j,i} \in [0,1]} \beta_{j,i} \qquad \text{s. t.} \quad P_e^{(j,i)}\left(\beta_{j,i}\right) < \epsilon \ .$$

where $P_e^{(j,i)}(.)$ is the BER of the j-th secondary link if transmitting on primary channel i. If a placed bid $\beta_{j,i}$ wins, then the secondary j will receive a utility of:

$$u_j(\beta_{j,i}, \alpha_i) = \alpha_i W_i \log\left(1 + \Gamma_j^{(i)}\right) q\left(\epsilon - P_e(\beta_{j,i})\right) \ , \quad (4)$$

where $\Gamma_j^{(i)} = \frac{\bar{P}_j(1-\beta_{j,i})|h_j^{(i)}|^2}{\alpha_i W_i N_i}$ is the SNR of the secondary user j on the leased channel of primary i. When employing BPSK transmission, $P_{e,j,i}(\beta_{j,i}) = Q\left(\sqrt{\gamma_{j,i}}\right)$ assuming $\alpha_i \neq 0$. Thus, we can numerically solve for $\beta_{j,i}^{(\max)}$ in (4). If $\alpha_i = 0$, we let $\beta_{j,i}^{(\max)} = 0$ for all $j \in \mathcal{K}_s$, meaning that SU's will not relay any primary user's signal who is not willing to lease any portion of its available bandwidth.

Each secondary user will aim to transmit at the lowest $\beta_{j,i}$ possible. However, this might reduce its chances in winning the bid because each primary user would prefer a secondary user that is willing to spent as much power as possible for relaying its signal. According to our proposal, each secondary user may pick its initial bid (or even the successive bids) using a particular distribution, or weighting, over $[0, \beta_{j,i}^{(\max)}]$ (not necessarily uniform). For example, if it has a large battery life remaining it can bid closer to $\beta_{j,i}^{(\max)}$ and vice versa. Thus, a possible method for picking up the initial bid could be as $\beta_{j,i}^{(\max)}\left(1 - e^{-aT_j^{(res)}}\right)$, where $T_j^{(res)}$ is the residual battery life of the j-th secondary user. We may also include any delay or priority conditions that the j-th user has into this bid selection formula. For example, if it has a high need to get access to the channel it can favor a bid closer to $\beta_{j,i}^{(\max)}$.

C. Actions of Primary RX: Selection of Winning Bids for Cooperative Communications

The i-th Primary RX selects the winning bid for i-th link as follows. It receives the bid from the j-th secondary at a received power level of $P_{i,j}^R = |h_{j,i}|^2 \bar{P}_j$. Then it may compute the received SNR it will get if the secondary j transmits at the bid power level of $\beta_{j,i}\bar{P}_j$ to be $\Gamma_{j,i} = \frac{\beta_{j,i} P_{i,j}^R}{(1-\alpha_i)W_i N_i}$.

The primary RX i will then use either the Maximum-Ratio Combining (MRC), Maximum-SNR Selection or Coherent Relay detection to compute the resulting overall SNR, if it combines the received signals from both paths: direct path from the i-th primary TX itself and the relayed path from secondary node j. To be specific, in the following we will assume coherent relay detection.

Denote by $R_i^f(P_i)$ the resulting final primary rate if the primary user i transmits at a power P_i:

$$R_i^f(P_i) = (1-\alpha_i)W_i \log\left(1 + \frac{P_i|h_i|^2 + \bar{P}_j \beta_{j,i}|h_{j,i}|^2}{(1-\alpha_i)N_i W_i}\right). \quad (5)$$

Let $P_i^{(\min)}(\beta_{j,i})$ be the minimum transmit power the i-th transmitter needs to transmit at to achieve $R_i^f(P_i) \geq R_i^{(\min)}$ if it accepts the j-th SU's bid for relaying:

$$P_i^{(\min)}(\beta_{j,i}) = \bar{P}_i \wedge \left[\frac{(1-\alpha_i)N_i W_i \bar{\gamma}_i(\alpha_i) - \bar{P}_j \beta_{j,i}|h_{j,i}|^2}{|h_i|^2}\right]^+, \quad (6)$$

where $\bar{\gamma}_i(\alpha_i) = 2^{\frac{R_i^{(\min)}}{(1-\alpha_i)W_i}} - 1$ and $x \wedge y \triangleq \min\{x, y\}$.

The i-th primary receiver then chooses the secondary user j that will lead to the smallest $P_i^{(\min)}$ such that $R_i^f(P_i) \geq R_i^{(\min)}$ as the winning bid for asymmetric cooperation on its channel:

$$\begin{aligned} j(i) \triangleq \ j^* \quad &= \arg\min\nolimits_{j \in \mathcal{K}_s} P_i^{(\min)}(\beta_{j,i}) & (7) \\ &= \arg\max\nolimits_{j \in \mathcal{K}_s} \beta_{j,i}\bar{P}_j|h_{j,i}|^2 \ . & (8) \end{aligned}$$

Although the above formulation seems to imply that cognitive devices need to be full-duplex, it can be interpreted and implemented easily in a time domain framework. In the absence of secondary cooperation, if we assume that a primary user has reduced its bandwidth to $(1-\alpha_i)W_i$ while transmitting continuously at a power P_i, then its transmission rate becomes $(1-\alpha_i)W_i \log\left(1+\frac{|h_i|^2 P_i}{(1-\alpha_i)N_i W_i}\right)$. This is equivalent to a primary user transmitting over a bandwidth of W_i for a $(1-\alpha_i)$ fraction of time and at a power of $\frac{P_i}{1-\alpha_i}$. Thus, transmitting continuously over a $(1-\alpha_i)W_i$ bandwidth and at a power P_i is equivalent to transmitting over the whole bandwidth W_i for a $(1-\alpha_i)$ fraction of the time and at a higher transmit power $\frac{P_i}{1-\alpha_i}$, which establishes the equivalence between time and frequency contraction.

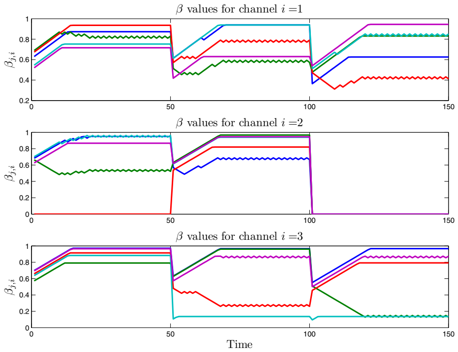


Fig. 4. Convergence of $\beta_{j,i}$ learning. $L=3$, $K_s=5$ and 3 time frames are shown.

IV. REPEATED FIRST-PRICE-AUCTION GAME MODEL FOR D-DSL AND REINFORCEMENT LEARNING

In the subsequent plays of the repeated game, if the channel conditions stay fixed, the secondary users can learn the others strategies and try to win the auction for spectrum leasing. At the beginning of each time slot, primary users take new bids. A simple reinforcement learning strategy to be used by the SU's are as follows:

$$\begin{aligned}
\text{Winning node:}\quad \beta_{j,i}^{(new)} &= \beta_{j,i}-\Delta\beta \quad \text{for} \quad \beta_{j,i}^{(new)}>0\\
\text{Loosing nodes:}\quad \beta_{j,i}^{(new)} &= \beta_{j,i}+\Delta\beta \quad \text{for} \quad \beta_{j,i}^{(new)}\leq \beta_{j,i}^{(\max)}
\end{aligned}$$

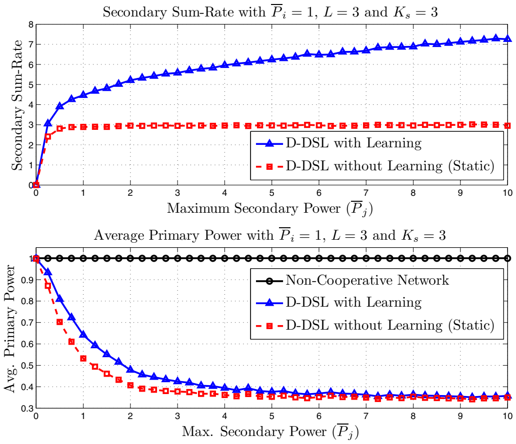


Fig. 5. Secondary Sum Rates and Average Primary Power v.s $\bar{P}_j$

where $\Delta\beta>0$ is some step size. Similarly, the primary users also learn and adapt their actions α_i at every time step. Primary users take distinct actions depending on whether a secondary user was selected for cooperation or not. The primary learning algorithm is as follows:

$$\begin{aligned}
\text{Coop.:}\quad \alpha_i^{(new)} &= \alpha_i-\Delta\alpha \quad \text{for} \quad \alpha_i^{(new)}>0\\
\text{No Coop.:}\quad \alpha_i^{(new)} &= \alpha_i+\Delta\alpha \quad \text{for} \quad \alpha_i^{(new)}\leq \alpha_i^{(\max)}
\end{aligned}$$

where $\Delta\alpha>0$ is some step size. However, when the primary users are adapting their α_i according to the secondary actions, the values of α_i might decrease and thus, the sum rate of secondary users might decrease as well. For that reason, we assume that primary users learn with a probability $\zeta\in[0,1]$, meaning that they adapt their actions in each time slot with a probability ζ. We note that secondary users should readjust their $\beta_{j,i}^{(\max)}$ values *within* a time frame whenever the primary user i changes α_i so that the secondary BER requirement stays satisfied.

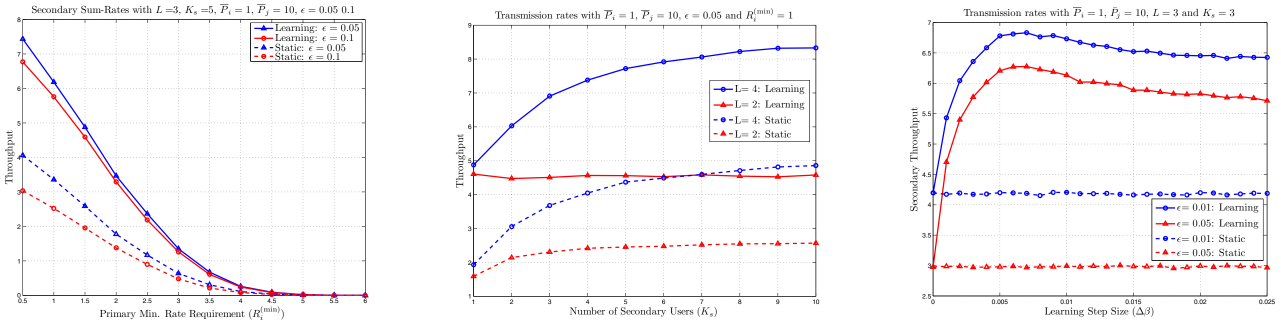
V. PERFORMANCE RESULTS

To illustrate the effectiveness of the proposed asymmetric cooperative communications based distributed DSL framework implemented as a first-price-auction, we consider a primary and a secondary systems having respectively $L=3$ and $K_s=5$ users, and assume all channels to be Rayleigh fading with $\mathbb{E}\left\{|h|^2\right\}=1$. We let $\bar{P}_i=1W$, $\bar{P}_j=10W$, $W_i=1Hz$, $N_i=0.1W/Hz$, $\epsilon=0.05$, $\zeta=0$, $\Delta\alpha=\Delta\beta=0.02$ and $R_i^{(\min)}=1bps$. First, we show in Fig. 4 the convergence of the secondary action variables $\beta_{j,i}$ as a function of time, over 3 time frames with 50 slots each. We observe that $\beta_{j,i}$ values are converging with time, and that there exists oscillations in the converging sequences when the users reach their lower bound for $\beta_{j,i}$. This is an expected behavior in all adaptive algorithms with non-zero step-sizes. Note that secondary users get faster access to primary channels if $\Delta\beta$ is larger, but this would increase the amplitude of oscillations in $\beta_{j,i}$ at equilibrium.

In Fig. 5, we plot the secondary sum-rate and the average primary power versus the maximum secondary power $\bar{P}_j$, in both learning and static scenarios. In the static case, the action variables are assumed to be fixed during the whole frame duration. Figure 5 shows that the primary users can reduce their power expenditure below $\bar{P}_i$ due to the asymmetric cooperation with secondary users.

Next, Fig. 6(a) shows the inverse dependence between the minimum primary rate requirement ($R_i^{(\min)}$) and the secondary sum-rate. In Fig. 6(b), we show that the sum-rate of secondary users increases with the number of secondary users (K_s) because of the increased diversity. Moreover, Fig. 6(b) visualizes the gain achieved by applying the learning procedure, when compared to the static case. Next, we analyze in Fig. 6(c) the impact of the learning step-size $\Delta\beta$ on the secondary throughput, and we observe that $\Delta\beta$ should be adjusted close to the point that ensures the fastest convergence and the highest performance.

Finally, we show in Fig. 7 the effect of the primary actions on the secondary throughput, and we see that the



(a) Secondary Sum Rates v.s $R_i^{(\min)}$.

(b) Secondary Sum Rates v.s K_s.

(c) Secondary Sum Rates v.s $\Delta\beta$.

Fig. 6. Secondary sum-rate performance with the proposed auctioning based D-DSL with asymmetric cooperative communications.

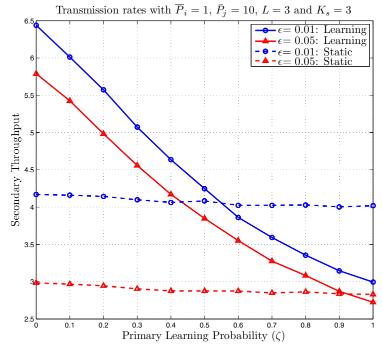


Fig. 7. Secondary Sum Rates v.s ζ

secondary performance degrades as the primary users try to learn more frequently. In fact, the primary learning procedure allows the primary users to decrease α_i, which reduces the available bandwidth for secondary transmission. Thus, it is more advantageous for secondary users to cooperate with a static non-adaptive primary system, which will will facilitate the adaptation of secondary users to their environment and prevent them from being exploited by the primary users.

VI. CONCLUSION

In this paper, we have proposed a D-DSL model that allows primary users to reduce their power expenditure by using the secondary users as relay nodes. In return for this asymmetric cooperative communication gains, primary users free-up a portion of their spectrum resources to secondary users. We developed a repeated first-price auction game in which the players learn by interacting with their environment so that they reach, distributively, an equilibrium point. We proposed a reinforcement learning algorithm for both primary and secondary users to learn and revise their actions and showed that the learning can significantly improve the overall performance.

ACKNOWLEDGEMENT

This research was supported in part by the National Science Foundation (NSF) under the grant CCF-0830545, and by the Space Vehicles Directorate of the Air Force Research Laboratory (AFRL), Kirtland AFB, Albuquerque, NM.

REFERENCES

[1] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology (KTH), Stockholm, Sweden, 2000.
[2] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Sig. Proc. Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.
[3] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894 –914, May 2009.
[4] Q. Xiao, Q. Gao, L. Xiao, S. Zhou, and J. Wang, "An optimal opportunistic spectrum access approach," in *IEEE International Conference on Communications Workshops. (ICC Workshops '09)*, Dresden, Germany, June 2009, pp. 1 –5.
[5] S. K. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 3300–3310, July 2009.
[6] S. K. Jayaweera, , G. Vazquez-Vilar, and C. Mosquera, "Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks," *IEEE Trans. Vehicular Technology*, vol. 59, no. 5, pp. 2328–2339, May 2010.
[7] S. K. Jayaweera and C. Mosquera, "A dynamic spectrum leasing (DSL) framework for spectrum sharing in cognitive radio networks," in *43rd Annual Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2009.
[8] K. Hakim, S. K. Jayaweera, G. El-Howayek, and C. Mosquera, "Efficient dynamic spectrum sharing in cognitive radio networks: Centralized dynamic spectrum leasing (C-DSL)," *IEEE Trans. Wireless Communications*, vol. 9, no. 9, pp. 2956–2967, Sep. 2010.
[9] O. Simeone, J. Gambini, Y. Bar-Ness, and U. Spagnolini, "Cooperation and cognitive radio," in *IEEE International Conference on Communications. (ICC '07)*, Glasgow, UK, June 2007, pp. 6511 –6515.
[10] M. Bkassiny and S. K. Jayaweera, "Optimal channel and power allocation for secondary users in cooperative cognitive radio networks," in *Special Session on Advanced Radio Access Techniques for Energy-Efficient Communications in 2nd International Conference on Mobile Lightweight Wireless Systems*, Barcelona, Spain, May 2010.
[11] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press, 1998.
[12] K.-L. Yau, P. Komisarczuk, and P. Teal, "Applications of reinforcement learning to cognitive radio networks," in *2010 IEEE International Conference on Communications Workshops (ICC)*, Capetown, South Africa, May 2010, pp. 1 –6.
[13] H. Li, "Multi-agent Q-learning of channel selection in multi-user cognitive radio systems: A two by two case," in *IEEE International Conference on Systems, Man and Cybernetics. (SMC 2009)*, San Antonio, TX, Oct. 2009, pp. 1893 –1898.