A Self-organized Spectrum Assignment Strategy in Next Generation OFDMA Networks providing Secondary Spectrum Access

Francisco Bernardo, Ramon Agustí, Jordi Pérez-Romero and Oriol Sallent
Signal Theory and Communications Department
Universitat Politècnica de Catalunya (UPC)
08034 Barcelona, Spain
Email: [fbernardo, ramon, jorperez, sallent]@tsc.upc.edu

***Abstract*—This paper proposes a Self-organized Spectrum Assignment strategy in the context of next generation multicell Orthogonal Frequency Division Multiple Access networks. The proposed strategy is able to dynamically find spectrum assignments per cell depending on the spatial distribution of the users over the scenario, opening new spectrum access opportunities for secondary spectrum usage. Reinforcement Learning methodology has been employed to implement the strategy, which compared with other fixed and dynamic spectrum assignment strategies shows the best tradeoff between spectral efficiency and Quality-of-Service while releases spectrum in large geographical areas.**

***Index Terms*—Self-Organized Spectrum Assignment, OFDMA, Reinforcement Learning.**

I. Introduction

The detected spectrum scarcity and its underutilization in current networks [1] claim for a new paradigm of spectrum access that overcomes current regulatory and technological barriers and promotes the usage of the spectrum dynamically and opportunistically by accounting for the different temporal and spatial spectrum demands. To this end, different spectrum access management models have been identified being currently under consideration by the spectrum regulatory bodies. The *Private Commons* initiative [2] has arisen as a spectrum access model where primary spectrum owners (usually the operators) agree to open their spectrum and to generate spectrum opportunities for secondary usage at the same time that they may charge a fee for each commercial secondary spectrum access. Thus, within this initiative, primary operators aim at performing a Dynamic Spectrum Assignment (DSA) strategy to maximize spectral efficiency and maintain the Quality of Service (QoS) of their users while pieces of spectrum are released in large geographical areas to create spectrum access opportunities.

This work has been performed in the framework of the project E^3, which has received research funding from the Community's Seventh Framework programme. Also, the Spanish Research Council and FEDER funds under COGNOS grant (ref. TEC2007-60985) have supported this work. This paper reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein. The contributions of colleagues from E^3 consortium and the support of the Spanish Ministry of Science and Innovation via FPU grant AP20051165 are hereby acknowledged.

In this context and to cope with the DSA problem, primary operators face the challenge of deploying a flexible radio access interface that eases the spectrum assignment. In this sense, OFDMA (Orthogonal Frequency Division Multiple Access) is a multiple access technique that is in the main stream of current proposed systems (3G LTE, WiMax) and at the same time is suitable for cognitive radio usage [3].

Furthermore, primary operators should give to their network the proper self-organization mechanisms to automatically observe, analyze, learn and react to the diverse temporal and spatial spectrum demands in a typical cognition cycle [4]. Thus, operational costs can be reduced while adaptability and robustness is given to the network [5]. Therefore, self-organized spectrum management techniques that adapt the network to the environment in which it is immersed become crucial. In this way, Reinforcement Learning (RL) techniques show appealing cognitive capabilities since they try to solve a problem from the continuous interaction with an environment that returns a numerical representation of a goal achievement (*reward*) for each RL *action* [6].

In our previous work, we showed that RL is suitable for spectrum management tasks under static [7] and dynamic [8] variations of the network traffic demands. In this paper we present a self-organized spectrum management framework enabling secondary cognitive radio usage in a multicell OFDMA system within the Private Commons spectrum access model. The framework embraces the DSA problem at a network level by sharing the whole spectrum among different cells. Thus, cognitive network functionalities, like network observation, analysis decision and learning are introduced in order to provide the network with the ability to automatically decide the adequate spectrum assignment to the different base station transmitters. Moreover, the ability of RL to learn from interaction with the network to discover proper dynamic spectrum assignments of groups of contiguous OFDMA subcarriers or *chunks* to cells is exploited to propose a self-organized RL-based algorithm for the DSA problem.

The proposed RL-based algorithm is compared with the fixed frequency reuse schemes so far proposed for cellular OFDMA planning, showing that the proposed scheme (*i*)

improves spectral efficiency, (*ii*) maintains users' QoS satisfaction and, (*iii*) enables opportunistic spectrum access (by releasing some frequency bands in large geographical areas) in a way that secondary spectrum usage might be introduced without penalizing the primary licensed users.

Following introduction, section II describes the proposed RL-based model for DSA. Section III is devoted to present the RL-based algorithm where the RL REINFORCE [9] methodology adopted here is explained. Finally, section IV discusses results and section V concludes this work.

II. Dynamic Spectrum Assignment System Model

Fig. 1 depicts the proposed DSA framework for a multi-cell OFDMA network. It is composed of a centralized *DSA controller* providing the proper cell-by-cell chunk assignment based on a *RL-DSA* (Dynamic Spectrum Assignment) algorithm. It is located in an entity of the network able to control a set of cells (e.g., the aGW in the case of 3G LTE). Each cell has a *Short-Term Scheduler* (STS) in charge of scheduling in the short-term the users' transmissions into the available resources (i.e., chunks assigned by the DSA controller) in temporal frame-by-frame basis.

The network variable status is observed and analyzed by a *RL Trigger* entity to detect the instants when the current spectrum assignment is no longer valid and thus trigger the RL-DSA algorithm. Metrics such as the load per cell, the intercell interference patterns and the QoS indicators are combined to decide the execution of the RL algorithm.

The RL-DSA algorithm is executed and its intermediate actions (candidate spectrum assignments) are applied to a *Network Characterization Entity* (NCE) based on real measurements that mimics the behavior of the real network. Inputs to NCE are the deployment of base stations, the load per cell and the average pathloss of the users from serving and neighboring cells. These measurements can actually be obtained from real networks. As it is discussed in next section, RL-DSA converges to a solution based on the interaction with the NCE that returns the *reward* for a given action.

Finally, a *Decision Maker* analyses the status of the algorithm to decide when the RL-DSA algorithm converges and the final chunk-to-cell assignment. Also it implements the procedures to redeploy the new assignment in the real system.

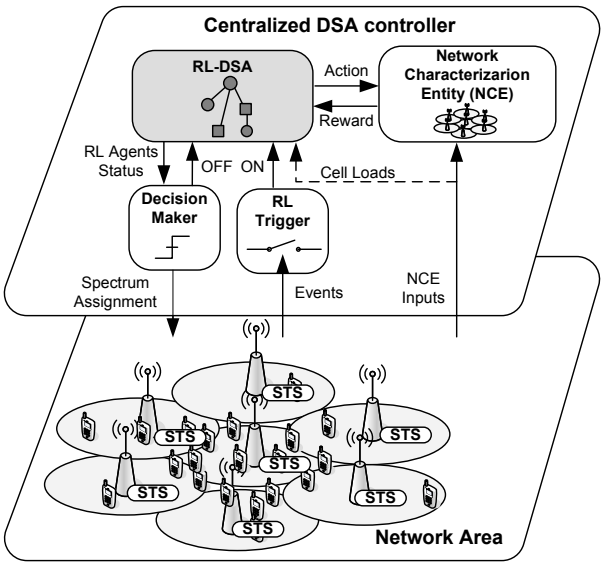

Fig. 1. Proposed DSA framework based on the RL-DSA algorithm.

III. RL-DSA Algorithm

Reinforcement Learning is learning the suitable set of actions to choose in order to maximize a numerical reward given that there is a continuous interaction with the outside world [6]. Then, we propose an algorithm that based on a set of RL agents (whose actions reflect possible spectrum assignments) learns the most suitable spectrum assignment that maximizes a given reward from the cellular network.

Consider N available chunks in a downlink OFDMA cellular system to distribute over K cells. Once it has been triggered, the purpose of the RL-DSA algorithm is to decide which chunks should be assigned to a cell and which not. To this end, the RL-DSA algorithm seeks chunk-to-cell assignments that maximize spectral efficiency per cell provided that the QoS per cell is assured while pieces of spectrum could be released for secondary radio usage. To cope with this problem, an RL system composed of KN RL agents based on the REINFORCE methods is proposed (Fig. 2(a)). This way, the network autonomously learns the best way of self-organize its spectrum in order to fulfill the operator's requirements.

A. Single RL Agent

Each RL agent is based on RL REINFORCE methods. Consider a single RL agent i represented in Fig. 2(b) that interacts with an environment in a succession of time steps. The total interaction with the environment of the agent is composed of a reward signal r and an input signal x_i biased by a weighting value w_i, that is $z_i{=}w_i x_i$ can be considered the effective input to the agent.

Then, the RL agent propagates the input z_i to the output y_i that is a binary number representing two possible actions. In fact, the RL agent is also called *Bernoulli-logistic unit* (BLU) because y_i is a Bernoulli random variable with parameter $p_i{=}f(z_i){=}1/\left(1{+}e^{-z_i}\right)$. That is, p_i represents the probability that the output y_i is 1, and contains the knowledge of the agent since it determines how often one of the two possible actions are chosen. Moreover, internal probabilities in the RL agents are always maintained within the interval $[p_{\text{explore}}, 1{-}p_{\text{explore}}]$ to assure a minimum exploratory probability p_{explore}. In this way, some actions that are not considered suitable *a priori* are selected even if internal probabilities tend to 0 or 1.

Notice that p_i depends on the input x_i and the weighting value w_i. Hence, the learning of the agent is condensed in the weighting value. Each time step t the agent learns from interaction with the environment by updating the weight as

$$\Delta w_i(t){=}\alpha(t)\left(r(t){-}\bar{r}(t{-}1)\right)\left(y_i(t{-}1){-}p_i(t{-}1)\right)x_i(t{-}1) \quad (1)$$

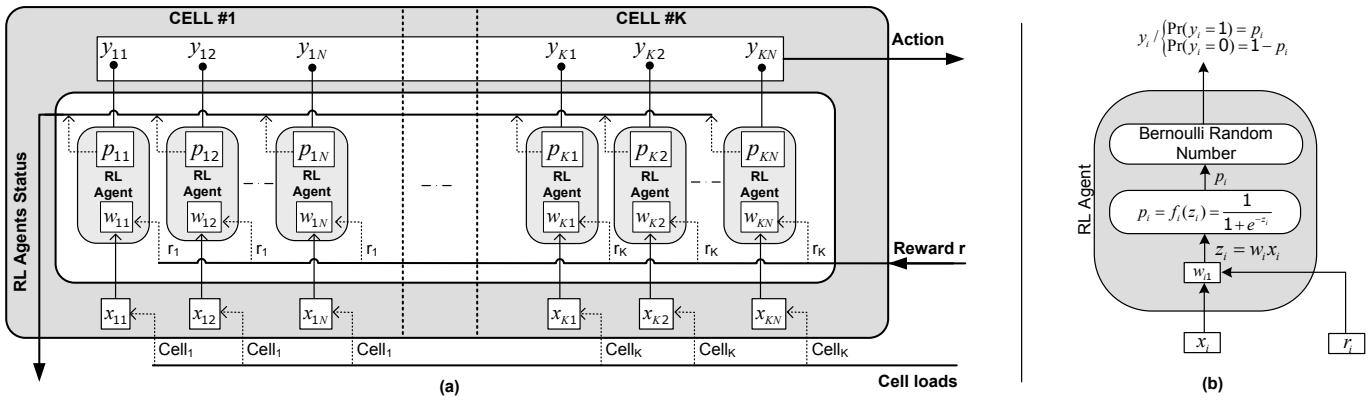

Fig. 2. (a) Proposed RL-DSA algorithm functional scheme. (b) Single RL REINFORCE agent

where $r(t)$ is the reward returned by the environment. $\bar{r}(t)$ is the reinforcement baseline or average reward calculated using a exponential moving average with parameter β. Parameter $\alpha(t)>0$ is called the learning rate and is decreased with the RL steps improves the convergence speed of the algorithm [6]. Thus, the learning rate is linearly decreased as $\alpha(t)=\alpha(t-1)-\Delta$ where Δ is a decreasing factor that should be small enough to assure a smooth transition between steps. Moreover, the maximum number of steps (MAX_STEPS) must be set so that negative values of α are avoided.

B. Proposed RL system

We propose a feed-forward network composed of KN REINFORCE agents to implement the RL-DSA algorithm (Fig. 2(a)) that interacts with the NCE. Notice that N RL agents are devoted to the k-th cell where each kn-th RL agent is dedicated to learn whether the n-th chunk should be assigned to that cell.

The RL-DSA algorithm works as follows. It is considered that in each interaction with the NCE the n-th chunk is assigned to the k-th cell if the output $y_{kn}(t)$ for the RL step t is 1. Thus, the action chosen by the RL system to interact with the NCE in each RL step is a binary assignment vector $A(t)=[y_{11}(t), y_{12}(t), ..., y_{KN}(t)]$ that contains the chunk-to-cell assignment given by the algorithm. Initially, full assignment is set, i.e., $y_{kn}(0)=1\ \forall n,k$.

For each assignment, the NCE returns a reward $r(t)$ that is used by the respective RL agents to update their weights and the average reward following 1. Initially average reward value is set to the first received reward. Then the agents capture the inputs for the next step $x_{kn}(t+1)$ that reflect the load status of a cell (i.e., the percentage of users in the k-th cell) so that the algorithm is able to adapt to homogeneous and heterogeneous spatial distributions of the load. Then, the next outputs $y_{kn}(t+1)$ are obtained with the procedure shown in Fig. 2(b).

Finally, the Decision Maker continuously checks whether the algorithm has converged and if so, decides the final chunk-to-cell assignment. As a convergence criterion the RL algorithm is executed until the variation of all internal probabilities between two consecutive steps is below a given threshold ε during S steps. After that, to take the final assignment to the real network, it is decided that a specific chunk is assigned to a cell if p_{kn} is greater than 0.5. Otherwise the chunk is not assigned.

C. Reward Signal Formulation

Theorem 1 in [9] assures that by updating the weights of each RL REINFORCE agent following 1 the average update vector $E\{\Delta\mathbf{W}|\mathbf{W}\}$ is proportional to $\nabla_{\mathbf{W}}E\{r|\mathbf{W}\}$, the gradient of the average reward r, where $\mathbf{W}=[w_{11}, w_{12}, \ldots, w_{KN}]$. Then, this property states that, the weighting vector $\mathbf{W}$ found when the algorithm converges (i.e.,$E\{\Delta\mathbf{W}|\mathbf{W}\}=0$) maximizes the average reward $E\{r|\mathbf{W}\}$. Thus, a reward signal that captures the final maximization objective and hence the performance of the cellular system in terms of spectral efficiency and QoS has to be defined to assure the success of the proposed strategy.

In this work, $P_k^{T_{th}}$ denotes the average user dissatisfaction per cell k and it reflects the percentage of time that the received throughput per user during a certain period is below a target throughput T_{th}. On the other hand, the spectral efficiency per cell is defined as

$$r_k(t)=\begin{cases}0, & \text{if } th_k(t) < T_{th}\\ \lambda\eta_k(t)+\mu(N-n_k(t)), & \text{otherwise}\end{cases} \quad (2)$$

where $r_k(t)$ and $th_k(t)$ are the reward and the average user throughput for cell k at step t respectively. $n_k(t)$ is the number of used chunks and N is the maximum number of chunks per cell. Finally, λ and μ are positive weight constants. In general λ should be greater than μ to mainly orient RL to choose solutions that enhance the spectral efficiency while the number of chunks is reduced. Notice that the maximum reward per cell is obtained when the users' average throughput is above target, the spectral efficiency is maximized and the minimum number of chunks is consumed in order to create spectrum usage opportunities for secondary spectrum markets.

IV. RESULTS

Results presented in this paper focus for the sake of brevity on the validation of the proposed algorithm by averaging the outcome of several trials with static users' distributions. Even with these static network conditions, where no load variations are considered within a trial, the RL-DSA algorithm remains dynamic since it adapts, if needed, the initial spectrum assignment (full assignment) to another one that improves the received reward. A system simulator has been developed to simulate the behavior of an OFDMA-based multicell network in downlink whose parameters are summarized in Table I. For RL-DSA simulations the same simulator has been employed for the NCE in Fig. 2(a) in order to predict the behavior of the real network.

The power devoted to every chunk is constant and does not change during simulation. The Signal to Interference plus Noise (SINR) ratio per each chunk is calculated as

$$\gamma_{m,n} = \frac{P_k G_{k,m} S_{k,m} F_{k,m,n}}{\sum_{j \in \Phi_n j \neq k} (P_j G_{j,m} S_{j,m} F_{j,m,n}) + \Upsilon_N}, \qquad (3)$$

where $\gamma_{m,n}$ represents the SINR in the n-th chunk for the m-th user, index i represents the serving cell and j any interfering cell. Φ_n is the set of cells using the n-th chunk. P_i, $G_{i,m}$, $S_{i,m}$ and $F_{i,m,n}$ denote respectively the transmitted chunk power, the distance dependant channel gain, the shadowing, and the fast frequency selective fading component that depends on the chunk n. Finally, Υ_N denotes the total thermal noise power including receiver noise figure. Therefore, the achievable rate $R_{m,n}$ that the m-th user can obtain in the n-th chunk in a frame is [11]

$$R_{m,n} = \frac{W}{N} \log_2 \left(1 - \frac{1.5\gamma_{m,n}}{\ln(5BER)}\right), \qquad (4)$$

where W/N the chunk bandwidth and BER stands for the target Bit Error Rate. The maximum theoretical spectral efficiency η_{max} is limited to 4 bits/s/Hz.

TABLE I
SIMULATION PARAMETERS

Number of cells	$K = 19$
Cell Radius	$R = 500$ meters
Antenna patterns	Omnidirectional
Frame time	2 ms
Number of chunks	$N = 12$
Chunk bandwidth	$B = 375$ KHz
Power per chunk	$P = 33$ dBm
Path loss in dB at d km	$128.1 + 37.6log10(d)$ [10]
Shadowing standard deviation	8 dB [10]
Shadowing decorrelation distance	5 m [10]
Small Scale Fading model	ITU Ped. A [10]
UE thermal noise	-174 dBm/Hz
UE noise factor	9 dB
RL parameters $[\alpha, \beta, \Delta]$	$\left[10, 0.01, 10^{-5}\right]$
Exploratory probability	$p_{exp} = 0.1\%$
Reward constant $[\lambda, \mu]$	$[100, 10]$
RL convergence criterion $[\varepsilon, S]$	$\left[10^{-4}, 5000\right]$
MAX_STEPS	100000

Users are distributed homogeneously within a cell and remain static during simulations. Their buffers are always full so that they demand as much capacity as possible. They are satisfied if the received throughput is above 128 kbps.

The scenario is composed of 19 cells and a maximum of 12 available chunks. Four spatial distributions of the users were simulated (Fig. 3). These patterns try to reflect a temporal evolution of the load in the scenario that progressively concentrates on a single cell. Thus, Fig. 3(a) depicts a homogeneous distribution and Fig. 3(d) a highly heterogeneous distribution. For each scenario, the load percentage per cell and the cell number is included in the figure.

Fig. 4 compares the results for the proposed RL-DSA algorithm with fixed frequency reuse factors FRF1 (full assignment) and FRF3 (1/3 of the bandwidth is assigned per cell), and a dynamic heuristic strategy named here DSA-heur [12]. Performance results are presented in terms of user's average dissatisfaction probability, spectral efficiency and the capacity of releasing spectrum. In order to quantify the capacity of each strategy for releasing spectrum for secondary usage the metric called Useful Released Surface (URS) presented in [13] is retained. The URS defines the surface where a given bandwidth can be used by secondary cognitive radio users respecting primary users' maximum interference level constrains. The same definition of the URS's parameters as in [13] is considered in this work.

Fig. 4(a) depicts the average spectral efficiency where RL-DSA obtains the best performance without compromising the satisfaction of the primary users (Fig. 4(b)) that is maintained or even improved respect the fixed spectrum assignment strategies. Then, RL-DSA obtains the best tradeoff between spectral efficiency and QoS of the primary users. For instance, in the case of the most heterogeneous load distribution scenario, RL-

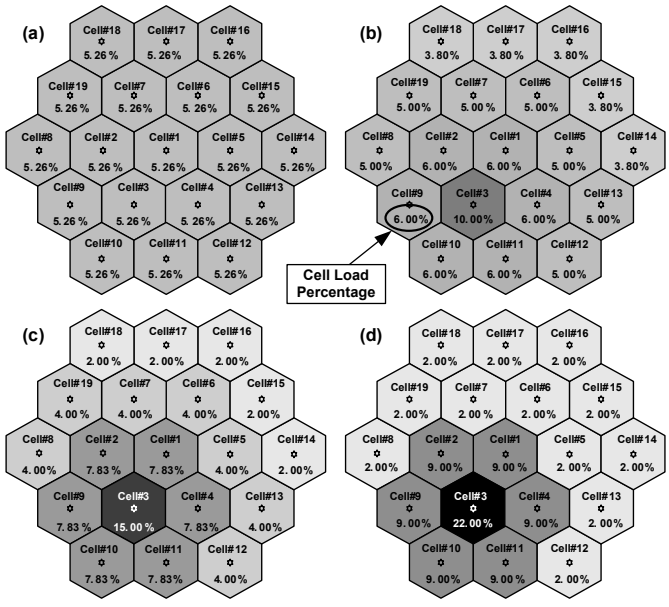

Fig. 3. Spatial distributions tested.

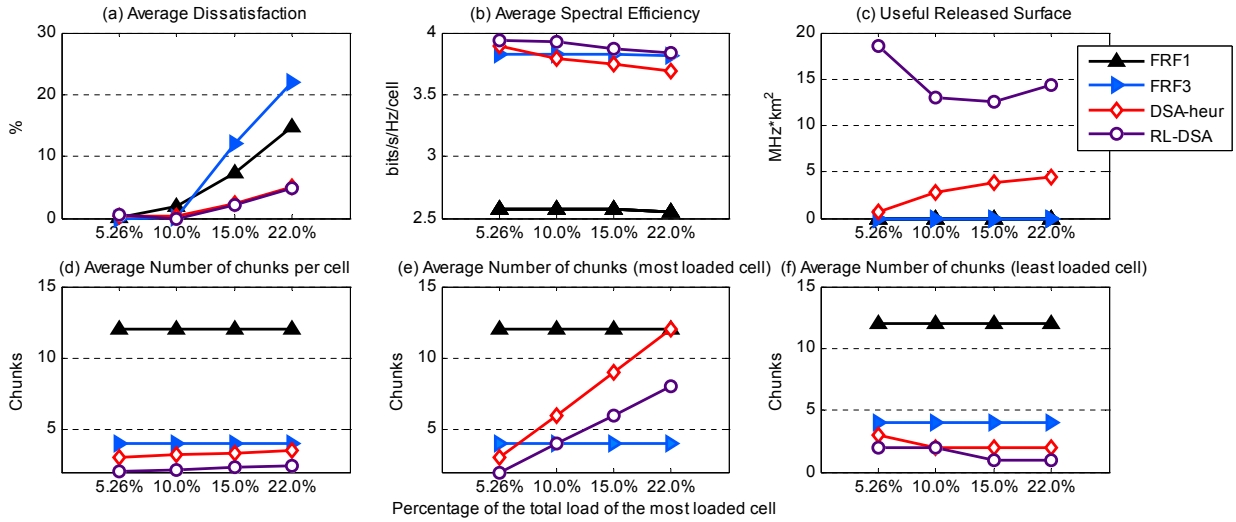

Fig. 4. Comparison results between RL-DSA algorithm, fixed spectrum assignment strategies (FRF1 and FRF3) and DSA-heur strategy. The abscissa represents the load of the most loaded cell in each one of the four studied scenarios.

DSA obtains the best spectral efficiency jointly with FRF3 but absolutely reduces the dissatisfaction of primary users in a 17%. Furthermore, RL-DSA demonstrates great capacity for releasing spectrum for secondary access (Fig. 4(c)). At least 10 MHz $*$ Km^2 of improvement in URS can be obtained respect the rest of studied strategies.

Therefore, the RL-DSA algorithm propitiates that the primary spectrum properly self-organizes by adapting to variable spatial needs of the spectrum. Fig. 4(d) shows the average number of chunks per cell. Notice that RL-DSA minimizes the number of chunks per cell but improves spectral efficiency and maintains dissatisfaction probability. Also, Fig. 4(e) and Fig. 4(f) show the average number of chunks for the most and least loaded cells respectively. It can be observed that RL-DSA wisely assigns the chunks per cell. For example, DSA-heur and RL-DSA increase(decrease) the number of assigned chunks to the most(least) loaded cell. However, RL-DSA organizes the chunks among cells in a way that the satisfaction of the primary users is maintained (Figure 4) while the number of chunks is reduced leading to improvements in the spectral efficiency and generation of spectrum gaps is geographical areas, as the Fig. 4(b)(c) reflect.

V. Conclusion

This paper has presented a self-organized Dynamic Spectrum Assignment (DSA) strategy in the context of next generation OFDMA-based networks by introducing a novel Reinforcement Learning-based algorithm (RL-DSA) whose foundations reside on an optimal RL methodology called REINFORCE. The primary operator's network, by using the RL-DSA strategy properly self-organizes its spectrum, demonstrating the best tradeoff between spectral efficiency, QoS fulfillment and generation of secondary spectrum access opportunities when compared with other fixed and dynamic spectrum assignment strategies under homogeneous and heterogeneous spatial. Concretely, improvements of 17% in satisfaction of the users can be obtained whereas at least 10 MHz $*$ Km^2 could be released for a primary operator for a secondary spectrum market, thus supposing new income streams for the operator.

References

[1] FCC, "Report of the spectrum efficiency working group," Spectrum Policy Task Force, Tech. Rep., 2002. [Online]. Available: http://www.fcc.gov/sptf/reports.html

[2] M. M. Buddhikot, "Understanding dynamic spectrum access: Models, taxonomy and challenges," in *IEEE New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2007, pp. 649–663.

[3] T. A. Weiss and F. K. Jondral, "Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency," *IEEE Commun. Mag.*, vol. 42, no. 3, pp. S8–14, 2004.

[4] B. Le, R. T.W., and B. C.W., "Cognitive radio realities," *Wireless Communications and Mobile Computing*, vol. 7, pp. 1037–1048, 2007.

[5] C. Prehofer and C. Bettstetter, "Self-organization in communication networks: principles and design paradigms," *IEEE Commun. Mag.*, vol. 43, no. 7, pp. 78–85, July 2005.

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, March 1998.

[7] F. Bernardo, R. Agustí, J. Perez-Romero, and O. Sallent, "A novel framework for dynamic spectrum assignment in multicell OFDMA networks based on reinforcement learning," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2009.

[8] ——, "Temporal and spatial spectrum assignment in next generation OFDMA networks through reinforcement learning," in *IEEE 69th Vehicular Technology Conference VTC2009-Spring*, 2009.

[9] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, no. 3, pp. 229–256, May 1992.

[10] 3GPP, "Physical layer aspects for evolved universal terrestrial radio access (UTRA)," 3GPP, Tech. Rep. TR 25.814 v7.1.0, September 2006, release 7.

[11] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Area. Comm.*, vol. 21, no. 2, pp. 171–178, 2003.

[12] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "Dynamic spectrum assignment in multicell OFDMA networks enabling a secondary spectrum usage," *Wireless Communications and Mobile Computing (WCMC)*, 2009.

[13] J. Nasreddine, J. Pérez-Romero, O. Sallent, and R. Agustí, "A primary spectrum management solution facilitating secondary usage exploitation," in *17th ICT mobile and wireless communications summit 2008.*, 2008.