Distributed Learning-based Cross-Layer Technique for Energy-Efficient Multi-carrier Dynamic Spectrum Access with Adaptive Power Allocation

Mahdi Ben Ghorbel, *Member, IEEE,* Bechir Hamdaoui, *Senior Member, IEEE,* Mohsen Guizani, *Fellow, IEEE,* and Bassem Khalfi, *Student Member, IEEE*

***Abstract*—This paper proposes energy and cross-layer aware resource allocation techniques that allow Dynamic Spectrum Access (DSA) users, by means of learning algorithms, to locate and exploit unused spectrum opportunities effectively. Specifically, we design private objective functions for DSA users with multiple channel access and adaptive power allocation capabilities. We also propose simple two-phase heuristics for allocating spectrum and power resources among users. The proposed heuristics split the spectrum and power allocation problem into two sub-problems, and solve each of them separately. The spectrum allocation problem is solved, during the first phase using learning. Two procedures to learn the channel selection are proposed and compared in terms of optimality, scalability, and robustness. The power allocation, on the other hand, is formulated as a real optimization problem and solved, during the second phase, by traditional optimization solvers. Simulation results show that energy and cross-layer awareness and multiple channel access capability improve the performance of the system in terms of the per-user average rewards received from accessing the dynamic spectrum access system. In addition, the two proposed methods for channel selection via learning represent a trade-off between optimality, scalability, and robustness.**

***Index Terms*—Cross-layer resource allocation, dynamic spectrum access, distributed resource sharing, private objective functions, cognitive radio networks.**

I. Introduction

Dynamic Spectrum Access (DSA) [1] has been one of the hot topics in wireless communications during the last decade due to its potential for improving spectrum utilization efficiency, and thus, addressing the spectrum shortage problem. DSA has been an important catalysis for numerous research works, ranging from protocol design [2] to performance optimization [3] and spectrum awareness techniques [4]. In particular, the resource allocation problem has been thoroughly addressed under different considerations. Nevertheless, a little attention was grasped to the context of large-scale DSA performances as it has more constrained requirements.

One of the important factors in the design of efficient wireless systems is the power consumption. Power and energy awareness have generated a continuous interest in the research community, as the importance of reducing energy consumption is becoming crucial not only in designing wireless systems, but also in any engineering systems due to other factors such as the environmental concerns (global warming, CO2 emissions, etc.) and the continuously increasing energy costs. Moreover, DSA is inherently a shared system for which the power needs to be efficiently used at each user to prevent mutual interference or interference to other users as DSA is mostly employed in the context of cognitive radio systems [5].

Recently, deriving optimal solutions and proposing low-complexity algorithms for DSA problems grasped a lot of research attention. In this perspective, the unprecedented growth of the number of users revealed that the conventional centralized approaches are no longer suitable to the recent developments in wireless communication systems. In fact, although among the merits of the centralized approaches is their ability to achieve optimal or near-optimal performances, these algorithms are not capable of ensuring the scalability since a large amount of overhead is mandatory. Thus, developing fully decentralized approaches is becoming more needed than ever due to the complexity of these emerging wireless systems. Though it can be very challenging to design them, decentralized approaches scale well, as they typically incur little to no communication and computational overhead while still performing relatively well.

In this context, an efficient distributed technique for spectrum access and allocation based on learning was proposed in [6]. The authors proposed a close-optimal, scalable, and highly learnable objective function that can be used for enabling efficient DSA. Although the proposed technique is shown to perform well in terms of the achieved throughput, it has some limits such as i) it considers ideal system with perfect channels (does not account for the effects of channel gains and noise), ii) its power consumption is not controlled which can result in an energy inefficient system, and iii) it limits each user to select a single band for their transmission which is not optimal.

In this work, we propose a joint dynamic multi-channel spectrum access with adaptive power allocation that extends the technique proposed in [6] to account for the power consumption and the cross-layer couplings. Specifically, we develop learning-based, distributed energy and cross-layer aware resource allocation techniques that allow DSA users, by means of learning algorithms, to locate and exploit unused spectrum

This work was made possible by NPRP grant # NPRP $5-319-2-121$ from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

This work is an extended version of the work accepted for presentation in IEEE International Conference on Computer Communications (INFOCOM Workshops), Toronto, Canada, April 2014.

Mahdi Ben Ghorbel and Mohsen Guizani are with Qatar University, Doha, Qatar (e-mail: mguizani@ieee.org).

Bechir Hamdaoui and Bassem Khalfi are with Oregon State University, Oregon, USA.

opportunities effectively. A key challenge of this work lies in how to propose an efficient algorithm that exploits the channel diversity to enhance the performance, but without suffering enormously from the added complexity of such an exploitation. To tackle this challenge, we proposed in [7] a two-phase heuristic approach that combines learning and optimization in a way that alleviates the computational complexity while still achieving good performances. The proposed heuristic splits the spectrum and the power allocation problem into two sub-optimal problems, and solves each of them separately. The spectrum allocation problem is solved using learning algorithms during the first phase, whereas the power allocation is formulated as an optimization problem and solved by traditional optimization solvers during the second phase. While the work in [7] generalizes the approach proposed in [6] to account for power consumption and takes into consideration variable channel gains, its complexity is still non-scalable for systems with a large number of bands, as it is an exponential function of the number of bands. In addition, it does not perform well in the case of dynamic availability of the channels for access due to primary users' activities. Thus, in this work we extended our previous work by proposing a suboptimal learning function that performs the search for the multichannel access on a channel by channel basis, thus resulting in 1) reducing the search cost to a linear function of the number of bands instead of being exponential, and 2) handling dynamic availability of channels effectively by allowing individual selection of channels to reduce interdependency of channels.

Our simulation results show that the proposed energy and cross-layer aware techniques coupled with the multiple channel access capability improve the DSA performances by increasing the per-user average reward that the users receive from accessing the DSA system. In addition, the two proposed methods for channel selection achieve a good DSA performance trade-off. While the (first) per-set channel selection method achieves better results in systems with a fixed and small number of bands, the (second) channel-by-channel selection method is more robust against changes in the cognitive radio environments.

The rest of this paper is organized as follows. We first overview in Section II the main research works done in the area of DSA. Section III introduces and states our proposed problem. In Section IV, we present our formulation of the DSA resource allocation problem, discuss the challenges of using learning in our multichannel DSA problem, and present the proposed suboptimal approach to be used to overcome these challenges. Then, in Section V, we present the new formulation of the problem in a cognitive radio framework with different interference scenarios and describe how to adapt the proposed approach to cope with the new context. In Section VI, we present simulation results and discuss the performance of the proposed algorithms under various system parameters. Finally, we conclude the paper in Section VII.

II. Literature Overview of Main Approaches for Successful DSA

Efficiently allocating the available resources is a key design problem lying at the core of DSA systems. For instance, as these systems are allowed to opportunistically access the spectrum, the need for effective techniques that allocate the resources for each user while meeting the DSA requirements becomes crucial. To this end, tremendous research efforts have been done to study the problem of power and spectrum allocation in DSA systems under different access paradigms. Broadly speaking, the proposed algorithms could be classified into two main categories: centralized and distributed approaches, depending on the decision-making process for each user, as whether the user allocates its resources based on its own observation and the information shared in the network or it relays its observation to a central unit that attributes to it the required resources [8].

Centralized resource allocation and scheduling techniques have been widely proposed as a potent means to efficiently allocate the power and the available spectrum in the network while preserving the energy and the spectral efficiency. The proposed algorithms range from joint spectrum and power allocation [9–12] to solely either power allocation [13] or spectrum allocation [14]. It is shown in this case that the centralized approaches can achieve optimal or near-optimal performances with the expense of a computational complexity. However, a large amount of overhead is required and may lead to a very high computational complexity with an intolerable delay. Hence, centralized approaches suffer from the problem of scalability.

To cope with this problem, distributed resource allocation has been seen as a promise to overcome the scalability problem of the centralized approaches [15, 16]. Different methods have been applied in this context. Game theoretical approaches have been seen as a good candidate [17, 18]. Pricing and bidding based approaches have also been the focus of some researchers [19]. While the proposed methods are shown to achieve good performance without the need of a central unit to perform the optimization and coordinate among users, significant communication overhead is still present in most of these methods.

Learning-based techniques have been considered as potential candidates for decentralizing the allocation of spectrum resources without the need of information exchange between users. Users rely mostly on observed behavior during past time slots to determine their best strategies for future resource allocation. Learning has been extensively used for resource allocation in wireless systems. For instance, [20] employs reinforcement learning for interference avoidance in heterogeneous networks through a cross-layer approach while [6] employs learning for optimal channel selection. Furthermore, [21] proposes some enhancements to the conventional learning algorithms to consider channel propagation and users' behavior while [22] combines learning with game theoretic approaches for a suitable technique for distributed DSA with noisy observation. A recent work by Xu *et al.* [23] proposes a multi-user learning algorithm to solve a dynamic spectrum allocation problem considering different moments of the targeted capacity instead of the expectation only.

In conclusion, learning approaches have shown great capabilities to be employed for DSA in wireless communication systems. Further work is still to be done to better exploit

available resources, such as multi-channel capabilities and adaptive power allocation, which have been shown as performance enablers through the additional diversity. While these capabilities could enhance performance, new challenges arise with the additional complexity.

III. System Model

We investigate the distributed resource allocation techniques for large-scale DSA networks. We consider a large-scale spectrum allocation problem with n users competing to access m spectrum bands ($n \gg m$), where each user selects and uses the spectrum bands among the m available bands to carry out its data communication. The interfering users, those that end up selecting the same band, are assumed to share the spectrum band using a Frequency Division Multiple Access (FDMA) scheme. We assume a fair band sharing so that each band is shared equally between the users who have selected it. We denote by $h_i^{(j)}(t)$ the instantaneous channel gain between the user i's transmitter and its receiver in the j^{th} channel band.

Unlike [6], where each user is only allowed to select and communicate over one channel band, we assume in this work that each user is allowed to select and use more than one channel band to communicate using a multicarrier scheme. In addition, this work also employs power control to reduce energy consumption, a factor that has not been taken into consideration in previous works. Moreover, we use channel gains to compute the received throughput which allow us to evaluate and analyze the performance under various channel conditions.

A. Throughput Expression

One contribution of this work is to study the problem taking into consideration the channel conditions for a better control of the energy consumption. For that, we express explicitly the received throughput in terms of the channel gains and the allocated power. Let B_j be the bandwidth of each channel sub-band. B_js are selected such that the channel gains are constant over each channel sub-band. The throughput of user i at instant t is expressed in terms of the allocated power per band $P_i^{(j)}(t)$ as

$$R_i(t) = \sum_{j=1}^{m} \frac{a_i^{(j)}(t)}{n_j(t)} B_j \, log_2 \left(1 + \frac{\left|h_{i,i}^{(j)}(t)\right|^2 P_i^{(j)}(t)}{N_0 B_j / n_j(t)} \right), \quad (1)$$

where $a_i^{(j)}(t)$ is the user-band occupation mapping index (i.e., $a_i^{(j)}(t) = 1$ if user i uses band j and $a_i^{(j)}(t) = 0$ otherwise), n_j is the number of users sharing band j at time slot t (i.e. $n_j(t) = \sum_{i=1}^{n} a_i^{(j)}(t)$). N_0 is the noise power density in dB/Hz.

B. Reward Functions

Throughout this paper, we aim to study different service models. Thus, the users' reward is expressed differently depending on the used service.

1) Elastic Traffic Model: In the elastic traffic model, the users utility increases as the received throughput increases given that it exceeds a minimum threshold. This traffic model is suitable for the cases of file download. Explicitly, the reward of user i at instant t, $r_i(t)$, can be expressed as

$$r_i^{ela}(t) = \begin{cases} R_i(t) & \text{if } R_i(t) \geq R_{th} \\ R_{th} e^{-\beta\left(\frac{R_{th}}{R_i(t)} - 1\right)} & \text{otherwise.} \end{cases} \quad (2)$$

2) Inelastic Traffic Model: In the inelastic traffic model, users need to receive a minimum amount of service. Below that threshold, the received service is not sufficient. But, any additional throughput received above the required threshold is not rewarded. Examples of applications of this type of reward include voice/video streaming and online gaming. Explicitly, the reward of user i at instant t, $r_i(t)$, can be expressed as

$$r_i^{inela}(t) = \begin{cases} R_{th} & \text{if } R_i(t) \geq R_{th} \\ R_{th} e^{-\beta\left(\frac{R_{th}}{R_i(t)} - 1\right)} & \text{otherwise.} \end{cases} \quad (3)$$

3) Energy Efficient Model: Due to the increasing energy costs, applications nowadays do not focus only on achieving high throughputs, but also on reducing power consumption. Thus, this reward model accounts for energy efficiency, and does so by combining the users' received throughput with the amount of energy consumed to achieve such throughput. This can be written as

$$r_i^{ee}(t) = \frac{R_i(t)}{P_0 + P_i(t)}, \quad (4)$$

where $P_i(t)$ is the power consumed during time episode t and P_0 is a fixed amount of power, capturing the non-radiated power consumed independently and regardless of the number of selected bands and of the power allocated for transmission.

C. Objective Function

Three types of objective functions are studied in this work to show the effectiveness of the proposed technique:

- **Intrinsic objective (selfish behavior).** The maximized objective for each user is equal to its own reward
$$g_i^{int}(t) = r_i(t). \quad (5)$$
- **Global objective (cooperative behavior).** The maximized objective for each user is equal to the sum of all users' rewards.
$$g_i^{glo}(t) = G(t) \triangleq \sum_{k=1}^{n} r_k(t). \quad (6)$$
- **Difference objective.** Inspired from [6], the basic idea of this objective lies in that removing the effects of all other users from the global objective gives us an objective function with higher learnability than the global objective function but while still ensuring objective alignment among users. Essentially, this function, referred as the *difference objective function*, measures user i's contribution to the total system received reward, making it more learnable without compromising its alignment

quality with other existing users in the system. Formally, the difference function can be written as

$$g_i^{diff}(t) = D_i(t) \triangleq \sum_{k=1}^{n} r_k(t) - \sum_{k=1, k\neq i}^{n} r_k^{-i}(t), \quad (7)$$

where $r_k^{-i}(t)$ denotes the reward of user k if user i's interference is removed.

It is worth stressing that the difference between the functions proposed in [6] and those proposed and studied in this work is three-fold: One, our proposed objective functions are cross-layer aware in a way that the reward a user receives depends on the consumed power level and the channel characteristics (this is provided via Eq. (1)). The second difference is that our proposed techniques are energy-aware in a way that the channel selection method (to be described later) accounts for the power consumption of the users via adaptive power control. Third, each user is allowed to access and use more than one spectrum band at the same time; that is, users have multi-channel access capabilities. Therefore, throughout the rest of the paper, we will refer to our proposed techniques as *energy and cross-layer aware objective functions* to distinguish them from those proposed in [6].

IV. Learning-based Multichannel DSA with Adaptive Power Allocation

The joint spectrum and power resource allocation problem can be formulated as

$$\begin{aligned} \max_{\{a_i^{(j)}, P_i^{(j)}\}} \quad & \sum_{i=1}^{n} r_i(t) \\ \text{s.t.} \quad & 0 \leq P_i^{(j)} \leq P_i^{(j),max} \quad i \in \{1...n\},\ j \in \{1...m\} \\ & \sum_{j=1}^{m} a_i^{(j)} P_i^{(j)} \leq P_i^{max} \quad i \in \{1...n\} \\ & 1 \leq \sum_{j=1}^{m} a_i^{(j)} \leq m_i^{max} \quad i \in \{1...n\} \end{aligned} \quad (8)$$

where P_i^{max} is user i's maximum transmit power, $P_i^{(j),max}$ is user i's maximum allowed power per sub-band j, and m_i^{max} is user i's maximum number of used sub-bands.

In this problem, the objective is to maximize the total obtained reward by all the users while respecting some constraints. The two first constraints serve to limit the energy consumption; the first is a budget power per user and the second is a maximum power per band which can serve to limit the generated interference, while the last constraint controls the multi-carrier complexity by limiting the number of sub-bands per user.

A. Challenge of the Learning Algorithm

Our proposed approach consists of extending the technique proposed in [6] to account for the power consumption when allocating the spectrum resources among the users. The approach consists of employing a learning algorithm, e.g. Q-learning [24], to evaluate the best possible channel/power allocation for each user. For that each user stores in a table a Q-value for each possible allocation called state. The Q-value gives an idea about the quality of the state using its reward history and future estimation. The Q-value of a selected state j by a user i at a time slot t is updated using the chosen objective function $g_i(t)$ as follows

$$Q_i^{(j)}(t) = (1-\alpha) Q_i^{(j)}(t-1) + \alpha g_i(t), \quad (9)$$

where α is a weighting factor chosen to control the importance of the effect of past information and present information in the Q-value. In single channel access the state is only a scalar representing the selected sub-band.

Accounting for the power resources when allocating the spectrum resources, though increases the spectrum efficiency as will be seen later, comes at a cost. The additional degree of freedom with the possibility of allocating different power levels on multiple channel sub-bands makes the resource allocation optimization problem a mixed integer programming (MIP) problem. The most intuitive approach consists in extending the state to accommodate the selected bands and allocated power per sub-band. For that, we consider L_i discrete possible power levels per sub-band for each user i. Then, instead of searching the selected band per user (m possibilities per user), the new unknown variable is a vector of $L_i \times m$ scalars (power allocated per sub-band).

The new state can take $\sum_{j=0}^{m_i^{max}} \binom{m}{j} L_i^j$ possibilities, where L_i^j is the number of non-zero possible power levels for user i. This quantity is upper-bounded by $\sum_{j=1}^{m} \binom{m}{j} L_i^j = (L_i + 1)^m$ since $m_i^{max} \leq m$.

The performance of the learning algorithm depends closely on the size of the searched space of the unknown variables in terms of computational complexity as well as optimality. In terms of computational complexity, searching the best action at each time slot requires to parse all possible actions to determine the best one. Thus as the search space's size increases, the computational complexity increases. In terms of optimality, learnability is affected by the decrease of the probability of finding the best action as the search state's increased. This results in needing more time slots to reach the optimal performance. On the other hand, decreasing the search space size results in lower degrees of freedom in allocating the available power among the used sub-bands which affects considerably the performance and makes the adaptive power approach non useful. Thus, we deduce that using discrete power levels is not a suitable solution for our problem.

B. Disjoint Channel and Power Allocation

To overcome the raised challenges of applying the learning algorithm in our problem, we propose, instead, a two-step algorithm that consists of using learning to determine only the channel-allocation mapping. Then, determining the power allocation by solving a pure real optimization problem.

1) Learning-Based Channel Selection: In this phase, learning will be employed to find only the set of channel bands to be used. Specifically, the best set of channel bands corresponds to

the available sub-bands with the maximal Q-table value since the users are allowed to select more than one band.

In this approach, the Q-table stores a Q-value for each possible set of bands that will be updated using Equation (9). Taking into consideration the constraint on the maximal number of the allowed bands per user, the total number of the possible sets is $\sum_{j=0}^{m_i^{max}} \begin{pmatrix} m \\ j \end{pmatrix}$. This quantity is upper-bounded by 2^m (since $m_i^{max} \leq m$ and $\sum_{j=0}^{m} \begin{pmatrix} m \\ j \end{pmatrix} = 2^m$). We note here a considerable decrease in the search space size which will result in a notable reduction of the computational complexity of the learning algorithm and an improved learnability on the price of a sub-optimality due to di-association of channels selection and power allocation.

2) Power Allocation Optimization: Having selected the channel sub-bands for each user after the first phase, the problem of determining the allocated power per sub-band for each user can be formulated as a constrained optimization problem. In this sub-problem, the sub-bands allocation indexes $\{a_i^{(j)}\}$ are known and the objective for each user is to determine the power to be allocated for each used sub-band $P_i^{(j)}$. The optimization problem for each user i is formulated as

$$\begin{aligned} \max_{\{P_i^{(j)}\}_{a_i^{(j)}=1}} \quad & r_i(t) \\ \text{S.t} \quad & \sum_{j=1}^{ml} a_i^{(j)} P_i^{(j)} \leq P_i^{max} \end{aligned} \tag{10}$$

In ordinary resource allocation problems where the reward $r_i(t)$ is exactly equal to the throughput $R_i(t)$, the solution to this problem can be found explicitly using a water-filling algorithm [25]. But when the reward is more complex, the problem becomes intractable analytically while being still convex if the reward function is concave in the power variables. In this case, a suitable numerical approach such as the interior point method could be used to determine the best power allocation over the selected bands. In the case of non-concave reward functions, more complex numerical solution approaches could be employed as in [26]. The studied reward functions (2), (3), and (4) are not necessarily concave which will make the optimization more challenging.

C. Independent Channels Selection Approach

Although, the proposed di-association between the channels and power allocation allowed to reduce the learning search space notably, it is still exponential as function of the number of bands which makes it unfeasible for systems with high number of bands. In addition, during the first phase, channel selection is done on a per group basis (i.e., the combination of bands with the highest Q-values). Thus, channels are highly dependent on each other which can cause a problem in case of unavailability of one of them or deterioration of its gain. Thus, in the following, we propose to lower the learning complexity by doing the bands selection for each band independently. In this approach, the Q-table stores a Q-value for each band. This method allows to reduce significantly the search space size of the Q-learner from exponential to linear as function of the number of bands ($m_i^{max} \times m$). In addition, it allows to remove the inter-band dependency which was forced by the group selection and which can harm the performance when there is a change of one or a part of the bands (dynamic scenario). In this method, a problem occurs as the reward function could not be computed for each channel independently. In fact, the reward corresponds to the use of the selected set of channels. In our work, we propose to compute the reward function per channel from the reward function of the selected set using two methods:

- **Equal reward:** In this method, we simply assume that all selected channels contributed equally to the obtained reward and assign to them the same reward as the obtained globally.
- **Proportional reward:** In this method, we compute the proportional contribution of each channel in the obtained reward. Even though this reward does not correspond to the real reward of that channel as it is dependent on the other bands selected with it, it can give an idea about the channel strength and occupancy and converges to the absolute reward of the channel with the learning as the channel will be selected with different combinations of other channels.

Note that in this approach, we modify only the channel selection phase while the power allocation procedure remains the same as in the previous approach.

Algorithm 1 summarizes the channels selection and power allocation procedure.

Algorithm 1 Learning-based channel and power allocation for large-scale DSA system.

Initialization: initialize Q-table as zero for all users and all channels;

for all time slot **do**

  for all DSA user **do**

1) **Channel Selection:** select the channels to be used either randomly with probability ϵ or by taking the ones with the highest value in the Q-table with probability $1 - \epsilon$;
2) **Power Allocation:** determine the optimal power allocation over the selected channels by solving the optimization problem (10);
3) **Reward Computation:** Measure the obtained throughput and compute the corresponding reward based on the considered traffic model using (2), (3), or (4);
4) **Q-table Update** update the Q-value of the used channels using (9);

  end for

end for

V. Spectrum and Power Allocation in Cognitive Radio Systems

In cognitive radios, unlicensed users (called cognitive users) are allowed to share the spectrum with the owners of the

spectrum or licensed users (called primary users) under certain constraints of interference that should be respected by the cognitive users. In the presence of primary users, the DSA problem is more challenging as additional constraints should be considered to limit the interference to the primary users. Next, we will consider different cognitive radio schemes with variable interference constraints and present the approach to adapt our proposed solution to solve the problem accordingly.

A. Opportunistic Cognitive Access

In this scheme, also called interweave scheme, cognitive users are urged to vacate the channels occupied by the primary users. Thus, the spectrum sensing is necessary to locate the occupied bands by the primary users to avoid colliding with them. Spectrum sensing methods are not within the scope of this paper and we assume that sensing is done perfectly by each user independently. As a result, in this scheme the set of the available bands for access to the cognitive users is variable over time. Each user will do a short spectrum sensing phase before starting its spectrum and power allocation. In order to account for this opportunistic access, we modify our algorithm by updating the list of the available channels for access at each time slot according to their occupation by the primary users. Thus, the bands selection phase will be limited over the unoccupied channels while the second phase of the power allocation remains similar to the initial algorithm.

B. Underlay Cognitive Access

In this scheme, the cognitive users are allowed to transmit in all the channels but urged to tune their transmission power in order to limit the interference caused to the primary users to a certain maximum threshold. Considering a peak power constraint of this form

$$\sum_{i=1}^{n} a_i^{(j)}(t)\left|h_{i,p}^{(j)}(t)\right|^2 P_i^{(j)}(t) \leq I_p^{(j)}(t),\ \forall j \in \{1...m\}, \quad (11)$$

where $\left|h_{i,p}^{(j)}(t)\right|^2$ is the gain of the interference channel from the secondary user j to the primary user and $I_p^{(j)}(t)$ is the peak interference threshold of the primary user at band j in time episode t. Adding this interference constraint to the DSA problem and using the band selection approach, we have

$$a_i^{(j)}(t)P_i^{(j)}(t) \leq \frac{I_p^{(j)}(t)}{\left|h_{i,p}^{(j)}(t)\right|^2},\ \forall j \in \{1...m\}, \forall i \in \{1...n\}. \quad (12)$$

Thus, with comparison to the original problem, the new constraint will represent the maximum power per band for each user to limit the interference to the primary user owning the channel. This additional constraint will not affect the first phase of the spectrum selection using learning but affects the second phase which consists of allocating the power optimally among the selected bands which should be tuned to take into consideration this new constraint.

C. Joint Opportunistic-Underlay Cognitive Access

We propose a new hybrid access mode for cognitive users which is a combination between the opportunistic and the underlay modes to allow more access opportunities to the cognitive users while still protecting the primary users. In this hybrid mode, the cognitive users perform initially the spectrum sensing. Then, they are allowed to transmit without any limit of interference in the unoccupied bands (like in the opportunistic scheme), while in the occupied bands, they can still transmit while respecting the interference thresholds fixed by the primary users. Explicitly, the new constraint is written as

$$\sum_{i=1}^{n} b_p^{(j)}(t)a_i^{(j)}(t)\left|h_{i,p}^{(j)}(t)\right|^2 P_i^{(j)}(t) \leq I_p^{(j)}(t),\ \forall j \in \{1...m\}, \quad (13)$$

where $b_p^{(j)}(t)$ is an index representing the primary user band occupancy (i.e., $b_p^{(j)}(t) = 1$ if band j is occupied by the primary user at time t and $b_p^{(j)}(t) = 0$ otherwise). We conclude that the problem in this mode can be formulated similarly to the underlay mode with the difference that the maximum power per band will only be forced on the occupied bands. The bands' selection phase remains similar to the original algorithms as all channels are accessible to the cognitive users for transmission.

VI. Performance Analysis

A. Complexity Analysis

As explained earlier, the computational complexity of the learning algorithm is proportional to the size of the search space which stores all possible states that can be allocated. In Table I, we summarize the complexity of each proposed algorithm by firstly presenting the learning table search space and then deducing the total complexity for each user. We denote by $C_{PA}(m)$ the complexity of the power allocation optimization over m channels to solve the problem (10). This complexity is not explicitly shown as it depends on the considered reward function.

B. Numerical Simulations

We consider an uplink cellular network where the cognitive users are generated randomly inside a circular cell of radius $d = 1$ Km. The base station (receiver for all users) is located in the center of the cell. The channel gains are generated according to a Rayleigh distribution [27] of mean power equal to the distance-based pathloss ($\frac{1}{d^\eta}$) with a pathloss exponent $\eta = 3$. The budget power per user is taken as $P_i^{max} = 0.1$ Watt. For the elastic and inelastic reward functions, the threshold of throughput is set to $R_{th} = 50\ kbps$ and the exponential decaying factor is set to $\beta = 2$ while for the energy efficiency reward, the minimum fixed consumed power is taken equal to P_i^{max}. A normalized channel width $B_j = 10$ KHz is used while the noise floor is set to $N_0 = -110$ dB/Hz. For the underlay cognitive radio scheme, the interference threshold is set to be equal to the noise floor. We denote by *per set selection* the method presented in subsection IV-B and by *per channel selection* the method presented in subsection IV-C. Unless explicitly noted, we consider the elastic reward function presented in Equation (2).

TABLE I
LEARNING SEARCH SIZE AND ALGORITHMS COMPLEXITY

Algorithm	Learning table size	Total Complexity
Joint Learning of Channels and Power Allocation	$\sum_{j=0}^{m_i^{max}} \binom{m}{j} L_i^j$	$O\big((L_i+1)^m\big)$
Disjoint Per-set Channels Learning and Power optimization	$\sum_{j=0}^{m_i^{max}} \binom{m}{j}$	$O\big(2^m + C_{PA}(m)\big)$
Disjoint Per Channel Learning and Power optimization	$m_i^{max} \times m$	$O\big(m^2 + C_{PA}(m)\big)$

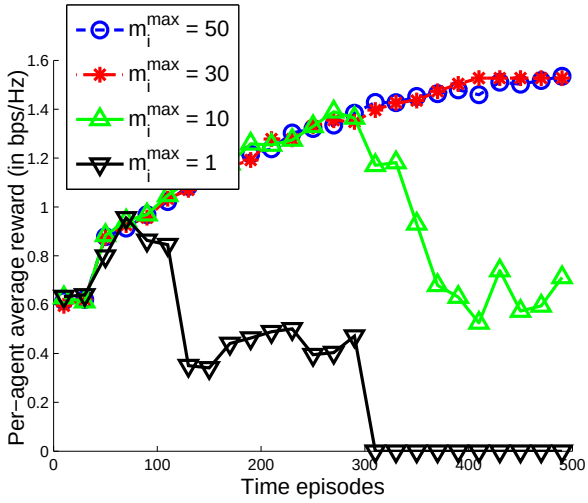

Fig. 1. Impact of multi-band spectrum access with adaptive power allocation capabilities on the achievable performance under the difference objective function with per set selection method (500 users and 50 bands).

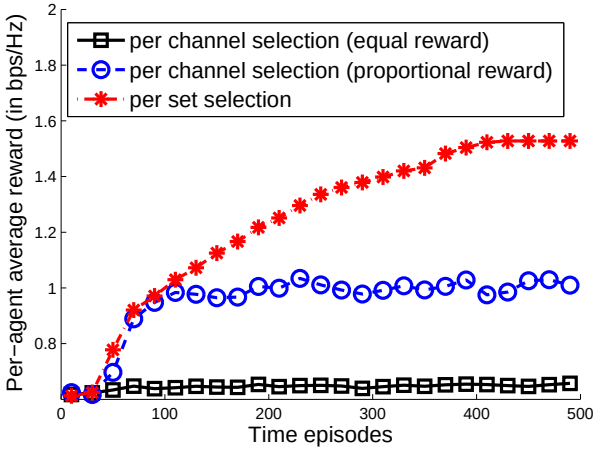

Fig. 2. Comparison between different methods for bands' selection with difference objective function (500 users, 50 bands, maximum 30 bands per user).

In Fig. 1, we show the performance (in terms of the per-user/agent average received rewards) of our proposed energy and cross-layer aware objective functions with multiple channel access and adaptive power allocation capabilities with different number of bands allowed per each user. We observe that the multiple channel allocation/access capability enhances the performance when compared with the single channel access by increasing the obtained average reward. This increased reward is a direct result of the benefit from the exploitation of channel diversity using this multiple channel access. In the proposed model, users are free to allocate more than one channel to maximize their reward. In addition, the proposed model allows the control of the energy consumption through the power constraint.

Fig. 2 shows the performance obtained under different methods for the channel selections in the learning phase when enabled with multiple access bands and adaptive power allocation capabilities. This figure shows the trade-off between complexity and optimality of the selection method. As shown earlier, the per set selection has an exponential complexity for the learning search but it achieves the best performance while the per channel selection allows to linearize the search price on the price of a performance degradation.

Fig. 3 shows the performance obtained under each of the three studied functions when enabled with multiple access bands and adaptive power allocation capabilities. First, This figure confirms the conclusions drawn in [6] on the efficiency of the difference objective function in achieving better performance than the intrinsic and global objective functions when enabled with our proposed cross-layer and energy aware features. Second, compared to the single channel scenario studied in [6], we note a relative deterioration of the performance obtained using the global and difference objective functions relatively to the intrinsic objective. This is explained by the fact that in this work, different actions affect the user's reward other than the users' selected channels which are the channel gains that vary across time slots and the power allocated over the selected sub-channels.

In Fig. 4, we plot the performance of the proposed energy and cross-layer aware techniques with multiple channel access and adaptive power allocation capabilities for two different network topologies by varying the cell radius. First, we observe that the difference objective functions outperform the other two regardless of the network topology. Also, note that this performance amelioration increases as the cell radius decreases. This is simply because the reward is asymptotically inversely proportional to the distance.

In Fig. 5, we show the performance (in terms of the per-

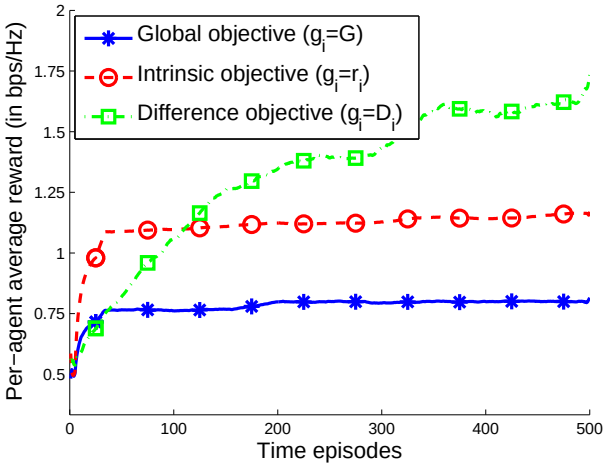

Fig. 3. Impact of multi-band spectrum access and adaptive power allocation capabilities on the achievable performance under the three studied objective functions: g_i, G, and D_i (100 users, 20 bands).

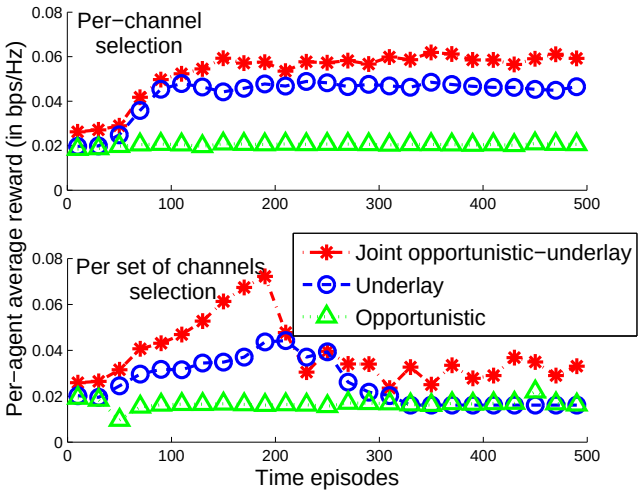

Fig. 5. Performance comparison in cognitive radio scenario (100 users, 10 bands).

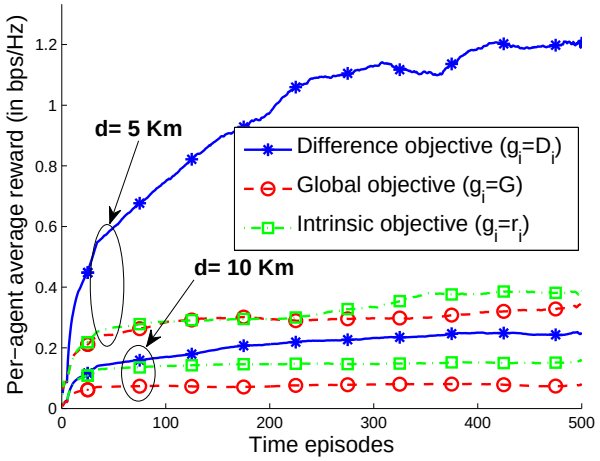

Fig. 4. Impact of network topology on the performance obtained under the three studied objective functions when enabled with multi-band spectrum access and adaptive power allocation capabilities (100 users, 20 bands).

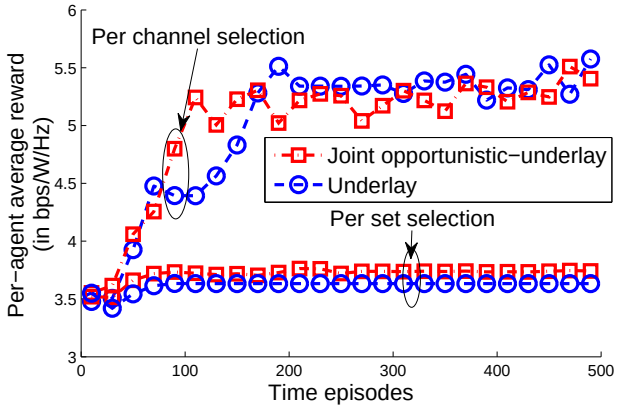

Fig. 6. Performance comparison for Energy efficiency reward model (100 users, 10 bands, maximum 5 bands per user).

user/agent average received rewards) of our proposed energy and cross-layer aware objective functions with multiple channel access and adaptive power allocation capabilities in a cognitive radio set-up where channel bands are occupied in average 50% of the time by the primary users. We plot the performance for different interference scenarios as shown in the figure's legend while using two methods for channel selections in the learning phase. This plot confirms the motivation for the proposed *per channel selection* method as the learning is not affected by the presence of primary users in the opportunistic and joint opportunistic-underlay scenarios oppositely to the case using the *per set selection* approach where the performance degrades with the primary users' presence, as the channels are forced to be selected in groups.

In Fig. 6, we present the performance result under the energy efficiency reward model presented in Equation (4) with 50% primary users' activity for two scenarios of the cognitive interference scheme. These results confirm the advantages of the new proposed method as it allows to achieve better performance with this reward as it is more sensitive to power consumption per band which reflects the real gain of the channel.

In Fig. 7, we present a comparison between the proposed Learning based technique to a conventional algorithm, namely the iterative water-filling algorithm known to achieve close-optimal solution in a linear complexity [28]. In this case, we consider the reward function as exactly the achieved rate and consider only the per set of channels method. This figure proves the efficiency of the learning approach as it allows to gradually outperform the water-filling algorithm and converge to a better solution thanks to the learning from previous experiences. In this figure, we also provide the error bar limits of the presented results based on the standard deviation to show the accuracy of the results. For instance, we observe a

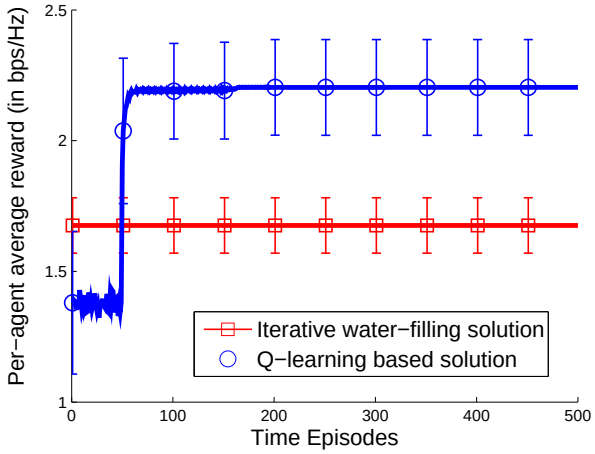

Fig. 7. Comparison of the proposed Learning based approach with an iterative water-filling based solution for a reward equal to the throughput (100 users, 10 bands).

stable error across our simulations which conserves the outperformance of our solution.

VII. Conclusion

This paper proposes learning-based, cross-layer and energy aware resource allocation techniques with multi-channel spectrum access and adaptive power allocation capabilities. It also proposes two heuristics for allocating spectrum and power resources among users. The proposed heuristics overcome the complexity issues by splitting this resource allocation problem into two sub-optimal problems, spectrum allocation problem and power allocation problem, and solves each of them separately. The spectrum allocation problem is solved using learning methods whereas, the power allocation one is formulated as an optimization problem. Our simulation results show that the proposed techniques perform well in terms of the per-user average achieved rewards because of their energy and cross-layer awareness and their multiple channel access capability.

References

[1] Q. Zhao and B. Sadler, "A Survey of Dynamic Spectrum Access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[2] B. Hamdaoui and K. Shin, "OS-MAC: An Efficient MAC Protocol for Spectrum-Agile Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 8, pp. 915–930, Aug 2008.

[3] Y. Ma, D. I. Kim, and Z. Wu, "Optimization of OFDMA-based cellular cognitive radio networks," *IEEE Transactions on Communications*, vol. 58, no. 8, pp. 2265–2276, August 2010.

[4] Z. Quan, S. Cui, and A. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, Feb 2008.

[5] J. Mitola and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug 1999.

[6] M. NoroozOliaee, B. Hamdaoui, and K. Tumer, "Efficient Objective Functions for Coordinated Learning in Large-Scale Distributed OSA Systems," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 931–944, May 2013.

[7] M. Ben Ghorbel, B. Hamdaoui, R. Hamdi, M. Guizani, and M. NoroozOliaee, "Distributed dynamic spectrum access with adaptive power allocation: Energy efficiency and cross-layer awareness," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS 2014)*, April 2014, pp. 694–699.

[8] S. Cicalo, V. Tralli, and A. Perez-Neira, "Centralized vs distributed resource allocation in multi-cell OFDMA systems," in *IEEE 73rd Vehicular Technology Conference (VTC Spring 2011)*, May 2011, pp. 1–6.

[9] Y. Liu, X. Tan, and A. A. Anghuwo, "Joint power and spectrum allocation algorithm in cognitive radio networks," *Journal of Systems Engineering and Electronics*, vol. 22, no. 4, pp. 691–701, Aug 2011.

[10] D. T. Ngo and T. Le-Ngoc, "Distributed resource allocation for cognitive radio networks with spectrum-sharing constraints," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3436–3449, Sept 2011.

[11] G. Bansal, M. Hossain, V. Bhargava, and T. Le-Ngoc, "Subcarrier and power allocation for OFDMA-based cognitive radio systems with joint overlay and underlay spectrum access mechanism," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1111–1122, March 2013.

[12] X. Zhou, G. Li, D. Li, D. Wang, and A. Soong, "Probabilistic resource allocation for opportunistic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2870–2879, September 2010.

[13] J. Zou, H. Xiong, D. Wang, and C. W. Chen, "Optimal power allocation for hybrid overlay/underlay spectrum sharing in multiband cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1827–1837, May 2013.

[14] G. Sun, X. Feng, X. Tian, X. Gan, Y. Xu, X. Wang, and M. Guizani, "Coalitional double auction for spatial spectrum allocation in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 6, pp. 3196–3206, June 2014.

[15] M. Guizani, B. Khalfi, M. Ben Ghorbel, and B. Hamdaoui, "Large-scale cognitive cellular systems: resource management overview," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 44–51, May 2015.

[16] R. Hamdi, M. Ben Ghorbel, B. Hamdaoui, M. Guizani, and B. Khalfi, "Implementation and analysis of reward functions under different traffic models for distributed dsa systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5147–5155, Sept 2015.

[17] M. R. Javan and A. R. Sharafat, "Distributed joint resource allocation in primary and cognitive wireless networks," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 1708–1719, May 2013.

[18] H. Li, Q. Guo, T. Tang, and Q. Li, "Distributed resource allocation for cognitive radio network with imperfect spectrum sensing," in *IEEE 78th Vehicular Technology Conference (VTC Fall 2013)*, Sept 2013, pp. 1–6.

[19] P. Yuan, Y.-C. Liang, and G. Bi, "Price-based distributed resource allocation for femtocell networks," in *IEEE International Conference on Communication Systems (ICCS 2012)*, Nov 2012, pp. 418–422.

[20] M. Bennis and D. Niyato, "A Q-learning based approach to interference avoidance in self-organized femtocell networks," in *IEEE GLOBECOM Workshops (GC Wkshps 2010)*, Dec 2010, pp. 706–710.

[21] L. Faganello, R. Kunst, C. Both, L. Granville, and J. Rochol, "Improving reinforcement learning algorithms for dynamic spectrum allocation in cognitive sensor networks," in *IEEE Wireless Communications and Networking Conference (WCNC 2013)*, April 2013, pp. 35–40.

[22] M. Derakhshani and T. Le-Ngoc, "Distributed learning-based spectrum allocation with noisy observations in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3715–3725, Oct 2014.

[23] Y. Xu, J. Wang, Q. Wu, J. Zheng, L. Shen, and A. Anpalagan, "Dynamic spectrum access with statistical qos provisioning: A distributed learning approach beyond expectation optimization," *CoRR*, vol. abs/1502.06672, 2015. [Online]. Available: http://arxiv.org/abs/1502.06672

[24] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, M. Cambridge, Ed. MIT Press, 1998.

[25] N. Papandreou and T. Antonakopoulos, "Bit and power allocation in constrained multicarrier systems: The single-user case," *EURASIP J. Adv. Signal Process*, vol. 2008, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1155/2008/643081

[26] M. Chiang, "Nonconvex optimization for communication networks," in *Advances in Applied Mathematics and Global Optimization*, ser. Advances in Mechanics and Mathematics, D. Y. Gao and H. D. Sherali, Eds. Springer US, 2009, vol. 17, pp. 137–196.

[27] B. Sklar and C. E. Services, "Rayleigh fading channels in mobile digital communication systems part I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, Jul. 1997.

[28] M. B. Ghorbel, H. Nam, and M. Alouini, "Generalized location-based resource allocation for OFDMA cognitive radio systems,"

in *Proceedings of the 2010 7th International Symposium on Wireless Communication Systems, ISWCS 2010, 19-22 September 2010, University of York, York, UK*, 2010, pp. 1011–1016. [Online]. Available: http://dx.doi.org/10.1109/ISWCS.2010.5624549

Mahdi Ben Ghorbel (S'10-M'14) is currently a postdoctoral fellow at Qatar University in Doha, Qatar since September 2013. He received the "Diplome d'Ingénieur" from Ecole Polytechnique de Tunisie (EPT) in Tunis, Tunisia in 2009 and the Ph.D in Electrical Engineering from King Abdullah University of Science and Technology (KAUST), Saudi Arabia in 2013. He received Excellency Fellowship for his studies at EPT and the top student award in his University in June 2009. He also received Provost Award and Discovery Fellowship to join KAUST in Sep 2009. His research interests include optimization of resource allocation and cooperative spectrum sensing performance for cognitive radio systems.

Bechir Hamdaoui (S'02-M'05-SM'12) is presently an Associate Professor in the School of EECS at Oregon State University. He received the Diploma of Graduate Engineer (1997) from the National School of Engineers at Tunis, Tunisia. He also received M.S. degrees in both ECE (2002) and CS (2004), and the Ph.D. degree in Computer Engineering (2005) all from the University of Wisconsin-Madison. His research focus is on distributed resource management and optimization, parallel computing, cooperative & cognitive networking, cloud computing, and Internet of Things. He has won the NSF CAREER Award (2009), and is presently an AE for IEEE Transactions on Wireless Communications (2013-present), and Wireless Communications and Computing Journal (2009-present). He also served as an AE for IEEE Transactions on Vehicular Technology (2009-2014) and for Journal of Computer Systems, Networks, and Communications (2007-2009). He served as the chair for the 2011 ACM MobiCom's SRC program, and as the program chair/co-chair of several IEEE symposia and workshops (including ICC 2014, IWCMC 2009-2015, CTS 2012, PERCOM 2009). He also served on technical program committees of many IEEE/ACM conferences, including INFOCOM, ICC, GLOBECOM, and PIMRC. He is a Senior Member of IEEE, IEEE Computer Society, IEEE Communications Society, and IEEE Vehicular Technology Society.

Mohsen Guizani (S'85-M'89-SM'99-F'09) is currently a Professor and Associate Vice President of Graduate Studies at Qatar University, Qatar. Previously, he served as the Chair of the Computer Science Department at Western Michigan University from 2002 to 2006 and Chair of the Computer Science Department at the University of West Florida from 1999 to 2002. He also served in academic positions at the University of Missouri-Kansas City, University of Colorado-Boulder, Syracuse University and Kuwait University. He received his B.S. (with distinction) and M.S. degrees in Electrical Engineering; M.S. and Ph.D. degrees in Computer Engineering in 1984, 1986, 1987, and 1990, respectively, all from Syracuse University, Syracuse, New York. His research interests include Wireless Communications and Mobile Computing, Computer Networks, Mobile Cloud Computing and Smart Grid. He currently serves on the editorial boards of many International technical Journals and the Founder and EIC of "Wireless Communications and Mobile Computing" Journal published by John Wiley (http://www.interscience.wiley.com/jpages/1530-8669/). He is the author of nine books and more than 400 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE Journals and Magazines. He also served as member, Chair, and General Chair of a number of conferences. He was selected as the Best Teaching Assistant for two consecutive years at Syracuse University, 1988 and 1989. He was the Chair of the IEEE Communications Society Wireless Technical Committee and Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005. Dr. Guizani is Fellow of IEEE, member of IEEE Communication Society, IEEE Computer Society, ASEE, and Senior Member of ACM.

Bassem Khalfi (S'14) is currently a Ph.D student at Oregon State University in Corvallis, OR, USA. He received the "Diplome d'Ingenieur" from Ecole Supérieure de Communications de Tunis (SUP'COM) in Ariana, Tunisia in 2012 and Masters of Science from Ecole Nationale d'Ingénieurs de Tunis (ENIT) in Tunis, Tunisia in 2014. His research interests include resource allocation in Dynamic Spectrum Access systems and performance analysis for cooperative spectrum sharing.