Maximizing Secondary-User Satisfaction in Large-Scale DSA Systems Through Distributed Team Cooperation

MohammadJavad NoroozOliaee, Bechir Hamdaoui, and Mohsen Guizani

***Abstract*—We develop resource and service management techniques to support secondary users (SUs) with QoS requirements in large-scale distributed dynamic spectrum access (DSA) systems. The proposed techniques empower SUs to seek and exploit spectrum opportunities dynamically and effectively, thereby maximizing the SUs' long-term received service satisfaction levels. Our techniques are efficient in terms of optimality, scalability, distributivity, and fairness. First, they enable SUs to achieve high service satisfaction levels by quickly locating and accessing available spectrum opportunities. Second, they are scalable by performing well in systems with small as well as large numbers of SUs. Third, they can be implemented in a decentralized manner by relying on local information only. Finally, they ensure fairness among SUs by allowing them to receive equal amounts of service.**

I. INTRODUCTION

Dynamic spectrum access (DSA) has been recognized as a key solution for solving the recently observed spectrum shortage [1, 2]. In the DSA context, there are two types of users: primary users (PUs) and secondary users (SUs). While PUs have exclusive access rights to use their licensed spectrum bands at all time, SUs are allowed to use these bands only opportunistically. That is, prior to using any licensed band, SUs must first sense the band to make sure that it is vacancy. When a PU returns while SUs are using its band, SUs must also vacate immediately. Spectrum sensing and PU detection techniques are beyond the scope of this work; we assume that SUs use existing sensing [3–6] and signal classification [7, 8] techniques for detecting and coping with PU activities.

DSA has great potential for improving spectrum efficiency through distributed access and management of spectrum resources [9–15]. As a result, it has generated a lots of research interests in developing adaptive channel selection techniques [16–18]. Zhao et al. [16] propose a prediction model that captures the DSA environment's dynamics under periodic channel sensing. The authors use a simple, two-state Markovian model to mimic PUs' activities on each channel, and use this model to derive an optimal access policy that leads to the maximization of spectrum utilization. Similarly, Liu et al. [17] model PUs' activities as a discrete-time Markov chain, which is then used to develop channel decision policies for two SUs in a two-channel DSA system. Chen et al. [18] propose DSA access methods that integrate physical-layer's with MAC-layer's sensing and access policy. They also assume that PUs' activities follow a discrete-time ON/OFF Markov process.

This work was supported by NSF under NSF CAREER award CNS-0846044.
MohammadJavad NoroozOliaee and Bechir Hamdaoui are with the School of EECS at Oregon State University. Mohsen Guizani is with Qatar University. Emails: noroozom@eecs.orst.edu; hamdaoui@eecs.orst.edu; mguizani@ieee.org.

Most of the proposed models developed for deriving optimal spectrum selection make a Markovian process model assumption about PUs' activities, which may not be accurate. Unlike traditional communication environments, the DSA environment gives rise to some unique characteristics, making it too difficult to model its dynamics and behaviors. This fact has created research interests to develop new distributed techniques that promote effective DSA [19–23]. For instance, game-theoretic approaches have been the focus of many researchers who used game-theory to develop distributed dynamic access methods [19, 20]. The authors in [19] study a DSA system with multiple, non-cooperative SUs with restricted information exchange. In this work, ON/OFF PUs' activities are modeled as an i.i.d. Bernoulli process, and DSA is formulated as a multi-armed bandit problem with multiple, non-cooperating agents. In [20], the authors investigated distributed DSA networks with non-cooperative, selfish users by studying, through game-theoretic approaches, the impact of incomplete information on system performance. They show that the lack of information can degrade the performance substantially. Learning-based techniques are also of a particular interest to DSA because they can easily be implemented in a decentralized manner without requiring any prior knowledge of the DSA environment's dynamics. Instead, these learning algorithms allow SUs to use their knowledge acquired from past and present interactions with the environment to take the proper actions that lead to maximizing the long-term amount of service that the SUs receive from accessing the DSA system. In other words, SUs first define and choose their objectives, then rely on a learning algorithm as a means to maximize these objectives. However, when these objectives are not designed carefully, learning algorithms can lead to poor overall system performance. This is because the collective behavior of the SUs aiming to maximize poorly designed objective functions is likely to yield a low overall received system service, thereby worsening the amount of service each SU receives. It is, therefore, essential that SUs' objective functions be carefully designed so that when the SUs go after maximizing them, their behavior as a whole leads to the maximization of the amount of service that each SU receives from accessing the DSA system.

In this work, we propose efficient management techniques that allow SUs to maximize their received service satisfaction through efficient spectrum resource allocation. We consider a distributed DSA system with multiple, non-overlapping spectrum bands. We also assume that each SU implements a learning algorithm (e.g., a reinforcement learner [24]) so it can use to maximize its objective function, thus enabling it to locate and select the best available spectrum opportunities. We want to emphasize that the focus of this work is not on learning algorithms, but rather on the design of efficient techniques that can be used by any learning algorithm to promote effective resource utilization. We show that the proposed techniques: achieve high service satisfaction levels by allowing SUs to quickly locate and exploit available spectrum opportunities; are very scalable by performing well in systems with a small as well as a large number of SUs; can be implemented in a decentralized manner by relying on local information only; and ensure fairness among SUs by allowing them to receive approximately equal amounts of service.

The rest of the paper is organized as follows. Section II presents the model and describes our motivation. Section III presents our proposed resource and service management techniques. Section IV derives the optimal performance behaviors. Section V evaluates the performances of the proposed techniques. Finally, Section VI concludes the paper.

II. Problem Statement

When a group of SUs want to communicate, all members of the group must first select and switch to the same band prior to communicating. At each time step, each group using a band receives a service that is passed to it from that band. The amount of service that the band offers a group can be measured for e.g. in terms of amount of throughput, reliability of the communication, SNR, packet success rate, etc. We assume that once the group switches to a particular band, it can immediately quantify and measure the amount of service that it receives from using such a band. The methods that are used to measure the service received as a result of using a band are beyond the scope of this work. Throughout this paper, we let V_j be the total amount of service that spectrum band j offers, and we refer to communication groups as *agents*.

Although the proposed network management techniques can be used by all learning algorithms, we choose to use in this work the ϵ-greedy Q-learner [24] with a discount rate of 0 and an ϵ value of 0.05 for evaluation purposes. More details on the Q-learner can be found in [24]. We want again to reiterate that this work in not on learning, but rather on the development of management techniques for DSA that can be used by any learning algorithms.

A. Traffic Model

In this paper, we study the inelastic traffic model, in which an agent receives a *constant* service satisfaction level when the band it uses offers an amount of service that is greater than a certain required threshold, Q, and receives an almost zero service satisfaction level when the amount of service offered by the band is below the threshold. Under this inelastic traffic model, receiving an amount of service that is less than what is required (i.e., Q) is not acceptable (which explains why the service satisfaction level drops immediately to zero), while receiving an amount that is higher than what is required is not beneficial either (which explains why the service satisfaction level remains constant). This inelastic model suits well applications with QoS requirements, such as video and audio applications, where receiving a QoS level higher than what the application requires does not typically improve the quality, whereas if the received level is lower than the required one, the application experiences a significant degradation in its quality. Formally, the service satisfaction level, $s_j(t)$, any agent using band j receives at time step t can be written as:

$$s_j(t) = \begin{cases} 1 & \text{if} \quad n_j(t) \leq V_j/Q \\ e^{-\beta \frac{n_j(t)Q - V_j}{V_j}} & \text{otherwise} \end{cases} \tag{1}$$

where $n_j(t)$ is the number of agents using band j at episode t, and β is a decaying factor. Note that when $n_j(t)$ is greater than $c_j \equiv V_j/Q$, the service satisfaction level decreases exponentially. This means that none of the agents will be satisfied with the service they receive from band j if the band has more than c_j agents (c_j here is the maximum number of agents that the band can support while satisfying the agents' required service levels; i.e., band j's capacity). From the system's perspective, the global or system service satisfaction level can be regarded as the sum of all agents' service satisfaction levels. Formally, by letting m denote the number of available spectrum bands, the global service satisfaction level, $G(t)$, at time step t can be expressed as

$$G(t) = \sum_{j=1}^{m} n_j(t) s_j(t) \tag{2}$$

B. Motivation

The goal of this work is to develop efficient resource and service management techniques for large-scale, distributed DSA systems. Specifically, we aim to derive scalable and distributed objective functions for SUs that are aligned with system objective, so that when SUs (i.e., agents) aim to maximize them, they indeed lead to the maximization of their long-term received service satisfaction levels. By means of any learning algorithm, these functions will enable SUs to efficiently find and locate spectrum opportunities, thus increasing the long-term service satisfaction level that each SU can receive from accessing the DSA system. With this in mind, the question that arises now is which private objective function g_i should each agent i maximize so that its received service satisfaction level is maximized?

There are two intuitive objective function choices. One possible function choice is to have each agent i using band j maximize its inherent service satisfaction level s_j received from band j as defined in Eq. (1); i.e., $g_i = s_j$ for each agent i using band j. A second also intuitive choice is for each agent to maximize the global/total service satisfaction levels that all agents receive; i.e., $g_i = G$ for each agent i as defined in Eq. (2), hoping that maximizing the global received service satisfaction levels eventually leads to maximizing every agent's long-term average received service satisfaction level.

For illustration purposes, we measure and show in Fig. 1 the system/global service satisfaction levels received by all agents under each of these two private objective function choices. We consider a DSA system with $n = 1600$ agents and $m = 10$ spectrum bands. Now we make the following two key observations. First, note that when agents aim to maximize their own inherent received service satisfaction level (i.e., $g_i = s_j$ for each agent i using band j), the global/system service satisfaction level received by all agents presents an oscillating behavior: it ramps up quickly at first but then drops down rapidly too, and then starts to ramp up quickly and drop down rapidly again, and so on. With the inherent objective function, an agent's received service satisfaction level, by design, is sensitive to its

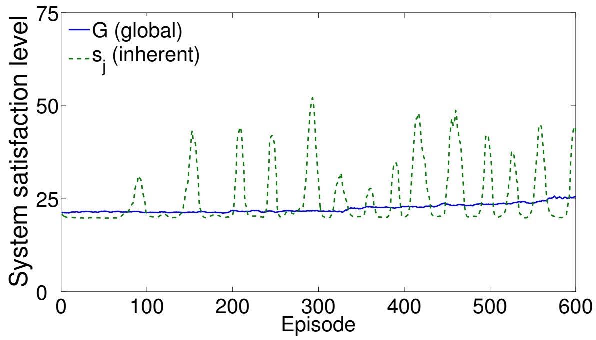


Fig. 1. System service satisfaction level under the two private objective functions: inherent choice ($g_i = s_j$) and global choice ($g_i = G$) for $m = 10$, $\beta = 2$, and $V_j/Q = 50$ for $j = 1, 2, \ldots, 10$.

own actions, which enables it to quickly determine the proper actions to select by limiting the impact of other agents' actions, thus learning about good spectrum opportunities fast enough. However, agents' inherent objectives are not aligned with one another, which explains the sudden drop in their received service satisfaction level right after learning about good opportunities.

Second, observe that, unlike the inherent function, the global function results in a steadier performance behavior where the system received service satisfaction increases continuously, but slowly. With this function choice, agents' objectives are aligned with one another by accounting for each other's actions, and thus are less sensitive to the actions of any particular agents. The alignedness feature of this function is the reason behind the observed monotonic increase in the overall system performance. However, the increase in the performance is relatively slow due to the function's insensitivity to one's actions, leading to slow learning rates.

Therefore, objective functions must be designed with two (conflicting) requirements in mind: (i) *alignedness*; when agents maximize their own private objectives, their collective behavior should indeed result in increasing each agent's long-term received service satisfaction level, and not in worsening it, and (ii) *sensitivity*; objective functions should be sensitive to agents' own actions so that proper action selections allow agents to learn about good opportunities fast enough.

C. Work Objective

Our goal is to derive efficient objective functions for large-scale DSA systems. Specifically, we aim to derive objective functions that $i)$ enable SUs to achieve high service satisfaction levels by allowing them to quickly locate and exploit available spectrum opportunities; $ii)$ are scalable by performing well in systems with a small as well as a large number of SUs/agents; $iii)$ are implementable in a decentralized manner by relying on local information only; and $iv)$ are fair by allowing agents to receive approximately equal amounts of service.

III. Resource and Service Management Techniques

The challenge in designing objective functions for DSA systems is to find the best balance between alignedness and sensitivity. Doing so will ensure that agents can learn to maximize their own objectives while also achieving good overall system performance. Throughout, we use g_i to denote the objective function of agent i that we aim to derive in this work.

In this section, we will first present the difference objective function, proposed in [25] and shown to perform well in various domains, such as multi-robot coordination [26], air traffic control [27], and opportunistic spectrum access [12, 28, 29]. This difference function will be used here as the basis for comparing the performance of our proposed function. Then, we present our proposed objective functions, whose performances, shown in Section V, are compared against those achievable under the two intuitive functions (s_j and G), against those achievable under the existing difference objective function, and against a theoretical upper bound that we also derive and state in Section IV.

A. Difference Objective Functions

Recall that, as illustrated in Section II-B, when agents set the global service satisfaction level, G, as their objectives (i.e., $g_i = G$ for each agent i), their collective behaviors did indeed result in increasing the total (system) service satisfaction levels, because agents' private objectives are aligned, in this case, with that of the system. However, because G depends on all agents, it is too difficult for agents (using G as their objective functions) to discern the effects of their own actions on their objectives, resulting then in low learnability rates. The authors in [25] address the above issue by proposing the difference objective functions, which provide a good balance between alignedness and sensitivity, leading to good system performance. The basic idea is that by removing the effects of all agents other than agent i from the function G, the resulting difference objective function will have higher learnability (or sensitivity) than G . These difference functions can formally be written as

$$D_i(t) \quad \equiv \quad G(t) - G_{-i}(t) \tag{3}$$

where $G_{-i}(t)$ is the system service satisfaction level at time step t when agent i is absent from the system. ($G(t)$ is given in Eq. (2).) Intuitively, since the second term evaluates the system satisfaction level without agent i, subtracting it from G provides an objective function that essentially measures agent i's contribution to the total received system service satisfaction level, making it more learnable. The difference function D_i can then be thought of as the *individual or agent contribution* to the system. Now by substituting Eq. (2) into Eq. (3) and after some algebraic manipulation, D_i for agent i selecting band j at time t can then be written as:

$$D_i(t) = n_j(t)s_j(n_j(t)) - (n_j(t)-1)s_j(n_j(t)-1) \tag{4}$$

B. Team Contribution Objective Functions

We now present our proposed functions. Our key idea is that instead of removing the impact of all agents other than agent i from the global service satisfaction level G (which led to the difference function design), we consider removing the impact of only those agents that may not be aligned with the agent itself. That is, in terms of contribution, we propose that an agent's objective function accounts for not only its contribution, but also for the contributions of all the agents that are aligned with it. More specifically, we propose that when the agents sharing a particular band/resource make, as a team, a positive contribution to the overall system performance, each agent in the team gets rewarded the team contribution; i.e., the sum of the contributions made by all agents in the team. But when the team contribution is negative (i.e., the resource is overcrowded, and hence none of the agents sharing it meet their required service), each agent in the team gets rewarded its own (negative) contribution only. The intuition is that when a group of agents (sharing a particular resource) succeed, they should celebrate their success as a team, but when they fail, each individual is only responsible for its own failure.

The proposed functions can then be thought of as the team or resource contribution to the entire system, and hence, they will be termed as *team (or resource) contribution objective functions*. Formally, when agent i chooses band j, its team contribution function can be written as

$$T_i(t) = \begin{cases} \sum_{k=1}^{n_j(t)} D_k(t) & \text{if} \quad n_j(t) \le V_j/Q \\ D_i(t) & \text{otherwise} \end{cases} \tag{5}$$

where again $n_j(t)$ is the number of agents using band j at episode t and $D_i(t)$ is the individual contribution function of agent i using band j, given in Eq. (4). Note that because D_i is the same for all agents sharing spectrum band j, Eq. (5) can be rewritten as

$$T_i(t) = \begin{cases} n_j(t) D_i(t) & \text{if} \quad n_j(t) \le V_j/Q \\ D_i(t) & \text{otherwise} \end{cases} \tag{6}$$

With the proposed team contribution function, SUs are capable of effectively distributing themselves across the bands in a way that benefits all of them by increasing the amounts of service they receive in the long-term. Thus, the proposed technique can be thought of as a resource allocation method that enables SUs to quickly locate best spectrum opportunities, and distribute themselves among the available bands effectively without cooperation.

C. Distributed Computation of Team Contribution Function T_i

Before proceeding with the performance evaluation of the proposed objective functions in terms of optimality, scalability, learnability, and fairness, we want to shed some light on their implementation aspects. Specifically, we want to discuss methods that agents can use to compute them in a distributed manner in spite of the large number of interacting agents, the restricted information sharing, and the limited communication and coordination capability among agents. Note that the design of computation methods for the proposed functions is beyond the scope of this work, and is in itself a different challenging problem. But here we only want to give some insights and reiterate on the distributed feature of these proposed functions.

Note that, by taking away agent i from the second term of the function D_i (as shown in Eq. (4)), the terms corresponding to all spectrum bands k except the band agent i is using cancel out, thus making the proposed functions implementable in a decentralized manner; i.e., each agent can implement them by relying on local information that can be observed locally by the agent itself. Let us now elaborate further on this. From the expression of $D_i(t)$ given in Eq. (4), note that $D_i(t)$ depends only on $n_j(t)$, the number of agents that happen to be contending with agent i for band j. Hence, in order to compute/estimate $D_i(t)$, one needs to estimate $n_j(t)$ given the information that agent i observes locally. Now an agent i using band j can easily/locally quantify the service, $a_i(t)$, it receives once it uses the DSA system, which can for e.g. be measured in terms of the amount of throughput the agent receives. Thus, assuming that all agents sharing a band will roughly receive the same amount of throughput, and that V_j is known to all agents, the number of agents, $n_j(t)$, using band j can be estimated to $V_j/a_i(t)$, which can then be used to estimate/compute $D_i(t)$. Hence, the function $D_i(t)$ can be computed by using information that an agent can observe/measure locally, and so can the function $T_i(t)$.

D. Performance Comparison: T_i versus D_i

We now want to compare the performance of T_i with that of D_i in terms of their ability to increase the overall achieved service satisfaction level. For this, we first provide an overview of the concept of "factoredness", which basically captures how aligned the agents' objectives are. Intuitively, the higher the degree of factoredness, the more likely it is that a change of state will have the same impact on the value of the objective function and on the achieved system satisfaction level. In other words, the more factored the objective function is, the more likely the system satisfaction level increases as agents maximize their objective functions, which eventually results in a higher long-term per-agent achieved service satisfaction level.

Let $z(t)$ characterize the joint move of all DSA agents in the system at time t. The global service satisfaction level, G, is then a function of $z(t)$, which can precisely be written as $G(z(t))$. The system state $z(t)$ basically captures the agent-channel assignment information and depends on the actions taken by the agents. For simplicity of notation, we often omit throughout the paper the dependency of these states on time t. With this, for systems with discrete states, the degree of *factoredness* for a given objective function g_i can formally be defined as [30]:

$$\mathcal{F}_{g_i} = \frac{\sum_z \sum_{z'} h[(g_i(z) - g_i(z'))\ (G(z) - G(z'))]}{\sum_z \sum_{z'} 1} \tag{7}$$

for all system states z and z' such that $z_{-i} = z'_{-i}$, where z_{-i} (or z'_{-i}) represents the system state that does not depend on the state of agent i (i.e., the parts of the system state controlled by all agents other than agent i), and $h[x]$ is the unit step function, equal to 1 if $x > 0$ and zero otherwise. A system is said to be fully factored when $\mathcal{F}_{g_i} = 1$.

Proposition 3.1: The degree of factoredness of the proposed team contribution objective function T_i is higher than that of the difference objective function D_i, i.e., $\mathcal{F}_{T_i} \ge \mathcal{F}_{D_i}$.

Proof: Note that the only term in $\mathcal{F}_{D_i}$ that is different from that in $\mathcal{F}_{T_i}$ is $g_i(z) - g_i(z')$; everything else is the same. Let us then compute and compare this term for T_i and D_i.
Let n_j and n'_j be the number of users in spectrum band j for system state z and z', respectively. Again, let us denote band j's capacity by c_j and define $c_j = V_j/Q$ (here, we assume $c_j = c$ for $j = 1, ..., m$). We consider the following four cases for n_j and n'_j that cover all possible cases:

- $n'_j > c > n_j$:
  In this case, from Eqs. (1) and (4), we can see that $D_i(z) - D_i(z')$ is positive. Similarly, from Eqs. (4) and (6), it follows that $T_i(z) - T_i(z') = n_j D_i(z) - D_i(z')$ is positive. Thus, the term $g_i(z) - g_i(z')$ is positive for both objective functions,

and hence, there is no difference between $\mathcal{F}_{D_i}$ and $\mathcal{F}_{T_i}$ since each depends on the sign of the term, and not on its value.

- $n_j > c > n'_j$:
Eqs. (1) and (4) imply that $D_i(z) - D_i(z')$ is negative, and similarly, from Eqs. (4) and (6), we easily see that the value of $T_i(z) - T_i(z') = D_i(z) - n'_j D_i(z')$ is also negative. Thus, the term $g_i(z) - g_i(z')$ is negative for both objective functions. Hence, $\mathcal{F}_{D_i} = \mathcal{F}_{T_i}$.
- $n'_j, n_j > c$:
Eq. (6) implies that $T_i(z) - T_i(z') = D_i(z) - D_i(z')$. Thus, the term $g_i(z) - g_i(z')$ has the same sign for both objective functions, and hence, there is no difference between $\mathcal{F}_{D_i}$ and $\mathcal{F}_{T_i}$ in this case either.
- $n'_j, n_j < c$:
From Eqs. (1) and (4), it follows that $D_i(z) - D_i(z')$ is zero, but from Eq. (6), it follows that the value of $T_i(z) - T_i(z')$ depends on n_j and n'_j and is not zero unless $n_j = n'_j$. This is the only case where the term $g_i(z) - g_i(z')$ in $\mathcal{F}_{D_i}$ is different from that in $\mathcal{F}_{T_i}$. So for these terms, the numerator in Eq. (7) is greater when $g_i = T_i$ than when $g_i = D_i$. This is because $D_i(z) - D_i(z')$ is equal to zero, and hence so is the step function value, whereas $T_i(z) - T_i(z')$ is not always equal to zero (when $n'_j, n_j < c$) and the step function value is equal to 1 for some values of n_j and n'_j. Thus, $\mathcal{F}_{T_i} \geq \mathcal{F}_{D_i}$. This completes the proof.

■

IV. Optimal Service Satisfaction

We now derive the optimal achievable service satisfaction level. This derivation will serve as a means of assessing how well the developed objective functions perform when compared with the optimal achievable performances.

Without loss of generality and for simplicity, assume that $V_j = V$ for $j = 1, 2, \cdots, m$. Let n denote the total number of agents in the system at any time. Let us also assume that $n > m\frac{V}{Q}$, since when $n \leq m\frac{V}{Q}$, the problem is trivial, and let $c = \frac{V}{Q}$, which denotes the capacity (in terms of the number of supported agents) of each spectrum band. Now, we start by proving the following lemma, which will later be used for proving our main result.

Lemma 4.1: The system/global service satisfaction level reduces less when a new agent joins a more crowded spectrum band than when it joins a less crowded band.

Proof: Recall that when a band j has $n' > c$ agents, the total service satisfaction level offered by the band is $G_j(n') = n' e^{-\beta(\frac{n'}{c}-1)}$. If a new agent joins this band, the new total service satisfaction level offered by the band becomes $G_j(n'+1) = (n'+1)e^{-\beta(\frac{n'+1}{c}-1)}$. First, it can easily be shown that when $n' > c \geq 1$, $G_j(n') > G_j(n'+1)$. Hence, the total service satisfaction level offered by a band j decreases by $\Delta_j(n') \equiv G_j(n') - G_j(n'+1)$ when a new agent joins the band. Now we can easily see that $\Delta_j(n')$ decreases when n' increases. Hence, the greater the number n' (i.e., the more crowded the band), the smaller the decrease in the total service satisfaction level when a new agent joins the band. ■

Theorem 4.2: When there are n agents in the system, the global service satisfaction level reaches its maximal only when $m-1$ bands (out of the total m bands) each has exactly c agents, and the m-th band has the remaining $n - c(m-1)$ agents.

Proof: Proof is in [31]. ■

Corollary 4.3: The system service satisfaction level is at most $(m-1)V/Q + (n-(m-1)V/Q)e^{-\beta(\frac{nQ}{V}-m)}$.

Proof: The proof follows from Theorem 4.2 by calculating the achievable global service satisfaction level for the derived optimal agent distribution. ■

Note that the optimal achievable system satisfaction level (stated in Corollary 4.3) is a theoretical upper bound on the sum of all agents' possible achievable service satisfaction levels. In the next section, we will evaluate the performances of the proposed functions and compare them with those of the difference functions as well as with this derived upper bound.

V. Performance Evaluation and Analysis

In this section, we evaluate the effectiveness of the proposed techniques by measuring and comparing their achievable satisfaction levels with those of the existing functions, inherent ($g_i = s_j$), global ($g_i = G$), and difference ($g_i = D_i$), as well with the optimal achievable performance, derived in Corollary 4.3.

A. Simulation Method and Setting

We consider a DSA system consisting of m non-overlapping spectrum bands and a large number of agents (SUs) all using the system opportunistically. We assume that each agent uses the Q-learning algorithm to implement the proposed objective function. Each agent does so independently from all other agents, and as long as it needs to access the DSA system. At each episode, each agent receives an amount of service (i.e., throughput) that is passed to it from the system. The learning algorithm utilizes this amount of service to compute and maximize its objective function so as to help the agent make the best spectrum decision/choice. All simulation scenarios are run (using MATLAB) until the measured achievable satisfaction level reaches its maximum peak. Each simulation point in all figures is averaged over all runs.

Unless stated otherwise, throughout this performance evaluation section, the decaying factor β is set to 2, the number of agents is set to 1600, the number of bands is set to 10, and the capacity V_j/Q is set to $c = 50$ for all j.

B. Static DSA without Primary Users

We begin by considering a static DSA system, in which all SUs enter and leave the system at the same time. We also ignore PU activities for the moment. Fig. 2 shows the system service satisfaction level normalized w.r.t. the optimal service satisfaction level (derived and stated in Corollary 4.3) achieved under each of the four functions: inherent, global, difference, and proposed. The figure shows that the proposed function, T_i, outperforms substantially the two intuitive functions, s_j and G, and outperforms the difference function, D_i, by about 25% in terms of the overall system service satisfaction levels. When compared to the optimal achievable performances, the proposed team function T_i is shown to achieve about 85 to 90% of the maximal achievable service satisfaction levels. Also, observe that our proposed function is very learnable as it enables agents to reach up their achievable service satisfaction levels quite quickly.

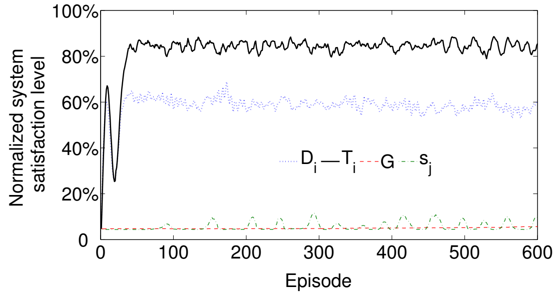
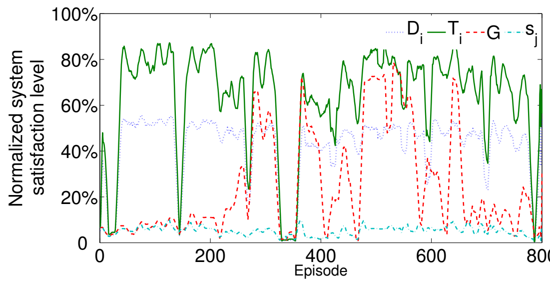
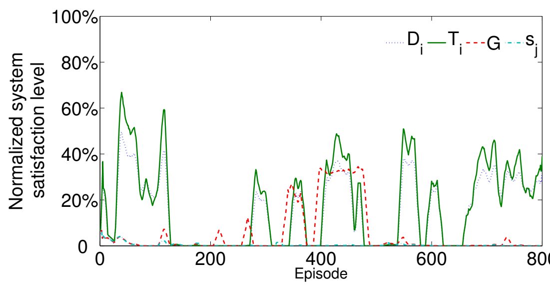


Fig. 2. Normalized system service satisfaction levels under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) at various time steps.



(a) $\eta = 10\%$



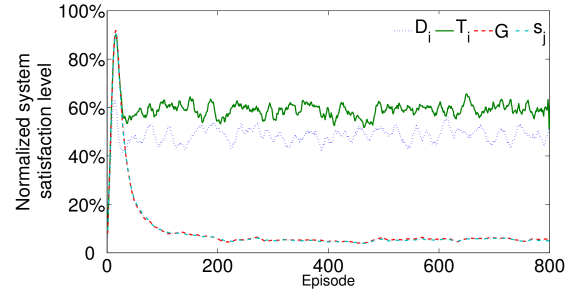
(b) $\eta = 50\%$

Fig. 3. Normalized system satisfaction level under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) at various time steps with PUs traffic load of $\eta = 10\%$ and 50%

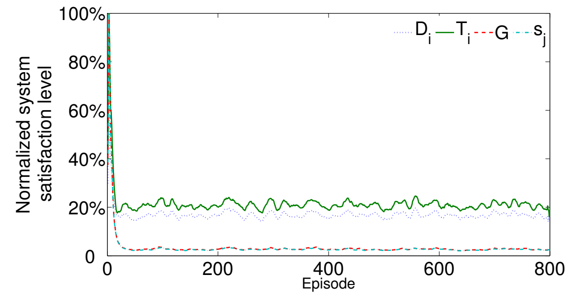
C. Static DSA with Primary Users

We again consider a static system where SUs enter and leave at the same time, but with the presence of PU activities. We model PU activities on each channel as a renewal process alternating between ON and OFF periods [32–34], which represent the time during which primary users are respectively present (ON) and absent (OFF). For each channel j, we assume that ON and OFF durations are exponentially distributed with means ν_j^{ON} and ν_j^{OFF}, respectively. We use $\eta_j \equiv \nu_j^{ON}/(\nu_j^{OFF} + \nu_j^{ON})$ to denote the *PU traffic load* on channel j.

Fig. 3 shows the service satisfaction levels under two different PU traffic loads. As it can be seen, even when considering PU activity, T_i still outperforms the other objective functions. However, the performance difference gap decreases as the PU traffic load increases. This is expected because the system satisfaction level, under any of the function, decreases as the PU load increases, since PU's presence makes the resources less available and hence, the overall system capacity decreases. Also, note that the achievable satisfaction levels, under any of the studied function, drop to zero whenever PUs come back, as it forces SUs to leave that channel, resulting then in an immediate decrease of the system service satisfaction levels.



(a) $\frac{\lambda}{\tau} = 1$



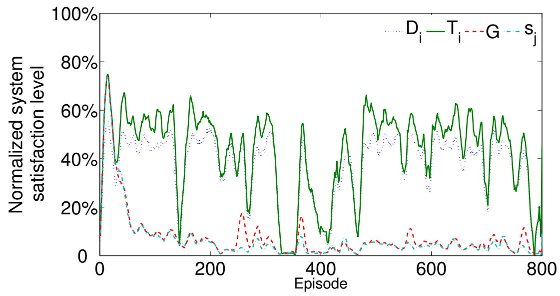
(b) $\frac{\lambda}{\tau} = 20$

Fig. 4. Normalized system satisfaction level under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) at various time steps under DSA agent traffic with $\frac{\lambda}{\tau} = 1$ and 20, and without PU activities (total number of agents $\kappa = 1600$).

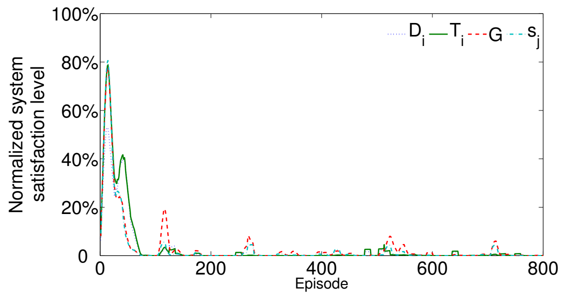
D. Dynamic DSA without Primary Users

Now, we consider a dynamic DSA system, in which SUs (i.e., the agents) can independently enter and leave the system at various different times. To model the dynamic behaviors of SUs, we assume that agents arrive according to a Poisson process with arrival rate λ. Each agent is characterized with an exponentially distributed duration of mean τ, during which the agent seeks and exploits available spectrum opportunities. We use $\kappa = \lambda\tau$ to designate the *DSA agent load*, which essentially represents the average number of agents that are using the system at any time. PU activities are ignored in this section and are considered in the next section.

In Fig. 4, we show the achieved performances under each of the four studied functions when considering dynamic behaviors of SUs: (Fig 4(a) for $\frac{\lambda}{\tau} = 1$; Fig 4(b) for $\frac{\lambda}{\tau} = 20$). Observe that the proposed objective function T_i outperforms all the other functions even when considering dynamic behaviors. Note that as the ratio $\frac{\lambda}{\tau}$ increases, the system satisfaction levels (under any of the function) decrease. This is because the higher the ratio $\frac{\lambda}{\tau}$, the lesser time (on average) SUs spend in the system (provided that κ is kept constant), and hence, the shorter the exploration time; i.e., SUs do not have enough time to explore better spectrum opportunities. This explains why the higher the $\frac{\lambda}{\tau}$, the smaller the system satisfaction level.

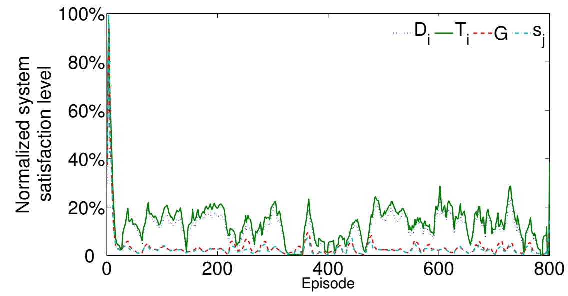


(a) $\eta = 10\%$

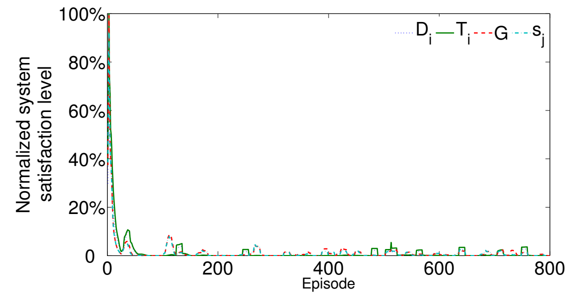


(b) $\eta = 50\%$

Fig. 5. Normalized system satisfaction level under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) at various time steps under DSA agent traffic of $\frac{\lambda}{\tau} = 1$ and with PU activities of $\eta = 10\%$ and 50%



(a) $\eta = 10\%$



(b) $\eta = 50\%$

Fig. 6. Normalized system satisfaction level under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) at various time steps under DSA agent traffic of $\frac{\lambda}{\tau} = 20$ and with PU activities of $\eta = 10\%$ and 50%

E. Dynamic DSA with Primary Users

We again consider a dynamic DSA system, but while also accounting for the activities of PUs. As in the previous scenario, we assume that agents arrive according to a Poisson process with arrival rate λ. Each agent is characterized with an exponentially distributed duration of mean τ. Figs. 5 and 6 show the system service satisfaction levels normalized w.r.t. the optimal service satisfaction level in a dynamic DSA system with PU activity for various combinations of the SU traffic ratio, $\frac{\lambda}{\tau}$, and the PU traffic load, η. In all cases, the proposed objective function T_i outperforms the other objective functions, but the performance gain depends on how loaded the system is. When the PU traffic load is relatively low as in Figs. 5(a) and 6(a) when $\eta = 10\%$, both the difference and the team contribution functions outperform the other two functions substantially. But as expected, when the PU load η increases, all functions achieve small service satisfaction levels, because the system is already loaded by PUs and hence there is no available spectrum for SUs to exploit. Likewise, as the ratio $\frac{\lambda}{\tau}$ grows (i.e., as the time each SU spends in the system decreases), the system satisfaction levels decrease, because when SUs spend lesser times in the system, it may not be enough for them to find good spectrum opportunities.

F. Scalability Performance

To also study scalability performance, we plot in Fig. 7 the normalized system service satisfaction level while varying number of agents, n, from 800 to 1600 while keeping m equal to 10. Since it takes some time for the technique to converge (to reach its maximum performance level), the performance values presented in this and the next subsections are measured after 600 episodes, which gave enough time for the performance to reach its best. Observe that T_i is highly scalable. Note that as the number of agents increases, T_i maintains high achievable system service satisfaction level, whereas the satisfaction level under s_j or G drops dramatically with the number of agents. When compared with the difference function D_i, our proposed team contribution function T_i still achieves satisfaction levels that are about 30% higher than those achievable under D_i.

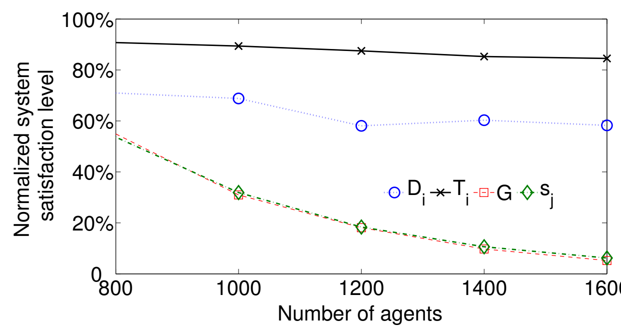


Fig. 7. Normalized system service satisfaction levels under inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) functions for various numbers of agents.

G. Fairness Performance

To also see how well the proposed functions do when it comes to fairness, we plot in Fig. 8 the coefficient of variations (CoV)[1] of the received system service satisfaction levels for various numbers of agents. Observe that the proposed function achieves

[1] CoV is the ratio of the standard deviation to the mean of the agents' received service satisfaction levels; we use this metric as a means of assessing the fairness, which reflects how close agents' received satisfaction levels are to one another.

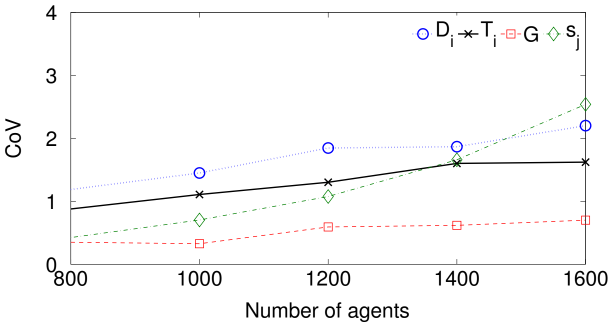


Fig. 8. Coefficient of variation (CoV) of satisfaction levels under inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed team ($g_i = T_i$) functions for various numbers of agents.

CoV values approximately similar to those achievable under any of the other three studied functions. These results show that not only does the proposed function achieve good performance in terms of optimality, scalability, and learnability, but also does so while ensuring a fairness quality as good as those achieved via the other approaches.

To summarize, we showed that the proposed team contribution objective functions achieve high satisfaction levels of agents' received service, are highly scalable as they perform well regardless of the number of agents, are highly learnable by enabling agents to reach up high values very quickly, are distributive as they require information sharing only among agents belonging to the same spectrum band, and are fair by allowing agents to receive similar amounts of service.

H. Discussion

There are two points that are worth mentioning and clarifying. Firstly, we want to reiterate that the reason for why our proposed objective functions are capable of achieving high service satisfaction levels is mainly because they lead to a distribution of agents across the available bands that is very close to the optimal agent distribution stated through Theorem 4.2, thus yielding near-optimal achievable performances and very scalable results regardless of the number of agents. That is, under the proposed techniques, $m-1$ bands will each have about c agents, whereas the rest of the agents will all go to the m^{th} band. In other words, unlike the other two functions which tend to jam all bands by distributing the agents uniformly across all bands, the proposed functions avoid band jamming by distributing the agents across the bands in a way that benefits all agents by increasing their long-term average achievable service. Now, one may think that this may be unfair to those agents that happen to be in the most crowded band (i.e., the m^{th} band), in that they will receive very low satisfaction levels when compared with those that happen to be in one of the other $m-1$ bands. Fortunately, this is not the case. Our experiments (not included in this paper due to limited number of figures) indicate that the most crowded band does not always contain the same set of agents. That is, agents belonging to this crowded band (which offers the least per-agent service) change over time, since agents move across bands at different time steps. The fact that the same agents do not get stuck in the most crowded channel is what ensures fairness among agents by allowing different agents to receive approximately equal amounts of service. This is justified via the fairness results shown in Fig. 8.

Secondly, in order for a group of SUs (e.g., a SU transmitter-receiver pair or an agent) to communicate on a given data channel, the SUs must first agree on the data channel before switching to it. SUs must rely on MAC protocols to do so. Most MAC designs for cognitive networks typically designate one channel, referred to as control channel, where all control messages needed for selecting data channels take place. Numerous MAC protocols have already been proposed in the literature to enable and coordinate multiple access in cognitive radio/dynamic spectrum access networks; [35, 36] present two surveys on MAC protocols. In this work, we assume that SUs use one the existing MAC protocols to negotiate data channels.

VI. Conclusion

In this paper, we proposed efficient resource and service management techniques that effectively support SUs in large-scale DSA systems. The proposed techniques allow SUs to exploit spectrum opportunities effectively, thereby maximizing the service satisfaction levels that SUs receive in the long term. We showed that the proposed techniques achieve high service satisfaction levels, are very scalable, are highly learnable by reaching up high values fast, are distributive by relying on local information only, and are fair.

References

[1] M. McHenry, "Reports on spectrum occupancy measurements, shared spectrum company," in *www.sharedspectrum.com/?section=nsf_summary*.

[2] M. McHenry and D. McCloskey, "New York city spectrum occupancy measurements," *Shared Spectrum Conf.*, Sept. 2004.

[3] H. Kim and K. G. Shin, "Efficient discovery of spectrum oppotunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, May 2008.

[4] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *Proceedings of IEEE DySPAN*, 2007.

[5] M. Gandetto and C. Regazzoni, "Spectrum sensing: a distributed approach for cognitive terminals," *IEEE Journal of Selected Areas in Communications*, April 2007.

[6] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, Februray 2008.

[7] T. J. O'Shea, T. C. Clancy, and H. J. Ebeid, "Practical signal detection and classification in gnu radio," in *Proceedings of the SDR 07 Technical Conference and Product Exposition*, 2007.

[8] J. Riihijarvi and P. Mohonen, "Exploiting spatial statistics of primary and secondary users towards improved cognitive radio networks," in *Proceedings of International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, 2008.

[9] H. Su and X. Zhang, "Cross-layer based oppotrunitic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE Journal on Selected Areas in Communications*, January 2008.

[10] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: a hardware-constrainted cognitive MAC for efficient spectrum management," *IEEE Journal on Selected Areas in Communications*, January 2008.

[11] B. Hamdaoui and K. G. Shin, "OS-MAC: An efficient MAC protocol for spectrum-agile wireless networks," *IEEE Transactions on Mobile Computing*, August 2008.

[12] M. NoroozOliaee, B. Hamdaoui, and K. Tumer, "Efficient objective functions for coordinated learning in large-scale distributed OSA systems," *IEEE Transactions on Mobile Computing (to appear)*.

[13] H. A. Bany Salameh, M. Krunz, and O. Younis, "Cooperative adaptive spectrum sharing in cognitive radio networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, 2010.

[14] M. Timmers, S. Pollin, A. Dejonghe, L. Van der Perre, and F. Catthoor, "A distributed multichannel MAC protocol for multihop cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, 2010.

[15] P. F. Marshall, "Dynamic spectrum access as a mechanism for transition to interference tolerant systems," in *Proceedings of IEEE DySPAN*, 2010.
[16] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *IEEE Transactions on Signal Processing*, February 2008.
[17] H. Liu, B. Krishnamachari, and Q. Zhao, "Cooperation and learning in multiuser opportunistic spectrum access," in *Proceedings of IEEE ICC*, 2008.
[18] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. on Inf. Theory*, May 2008.
[19] K. Liu and Q. Zhao, "Distributed learning in cognitive radio networks: multi-armed brandit with distributed multiple players," in *Submitted to IEEE Int. Conf. on Acousitcs, Speech, and Signal Processing*, 2010.
[20] A. Ozyagci and J. Zander, "Distributed dynamic spectrum access in multichannel random access networks with selfish users," in *Proceedings of International Wireless Communications and Networking Conference*, 2010.
[21] U. Berthold, M. Van Der Schaar, and F. K. Jondral, "Detection of spectral resources in cognitive radios using reinforcement learning," in *Proceedings of IEEE DySPAN*, 2008, pp. 1–5.
[22] J. Unnikrishnan and V. V. Veeravalli, "Algorithms for dynamic spectrum access with learning for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, August 2010.
[23] J. Unnikrishnan and V. V. Veeravalli, "Dynamic spectrum access with learning for cognitive radio," in *Proc. of Asilomar Conference on Signals Systems and Computers*, Oct. 2008.
[24] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 1998.
[25] A. Agogino and K. Tumer, "Multi agent reward analysis for learning in noisy domains," in *Proc. of the Fourth Int'l Joint Conf. on Autonomous Agents and Multi-Agent Systems*, Utrecht, Netherlands, July 2005.
[26] A. K. Agogino and K. Tumer, "Efficient evaluation functions for evolving coordination," *Evolutionary Computation*, vol. 16, no. 2, pp. 257–288, 2008.
[27] K. Tumer and A. Agogino, "Distributed agent-based air traffic flow management," in *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Honolulu, HI, May 2007, pp. 330–337.
[28] M. NoroozOliaee, B. Hamdaoui, and K. Tumer, "Achieving optimal elastic traffic rewards in dynamic multichannel access," in *Proc. of IEEE Conference on High Performance Computing and Simulation*, 2011.
[29] B. Hamdaoui, M. NoroozOliaee, K. Tumer, and A. Rayes, "Aligning spectrum-user objectives for maximum inelastic-traffic reward," in *Proc. of IEEE Int'l Conference on Computer Communication Networks*, 2011.
[30] A. K. Agogino and K. Tumer, "Analyzing and visualizing multiagent rewards in dynamic and stochastic environments," *Journal of Autonomous Agents and Multi Agent Systems*, vol. 17, no. 2, pp. 320–338, 2008.
[31] M. NoroozOliaee and B. Hamdaoui, "Distributed resource and service management for large-scale dynamic spectrum access systems through coordinated learning," in *Proceedings of International Wireless Communications and Mobile Computing Conference*, 2011.
[32] C. Jiang, Y. Chen, K. J. Ray Liu, and Y. Ren, "Renewal-theoretical dynamic spectrum access in cognitive radio networks with unknown primary behavior," *IEEE JSAC*, Submitted to.
[33] M. Levorato, L. Badia, , U. Mitra, and M. Zorzi, "An analysis of cognitive networks for unslotted time and reactive users," in *Proceedings of IEEE MASS*, 2010.
[34] D. Xue, E. Ekici, and X. Wang, "Opportunistic periodic mac protocol for cognitive radio networks," in *Proceedings of IEEE Globecom*, 2010.
[35] C. Cormiob and K. R. Chowdhurya, "A survey on MAC protocols for cognitive radio networks," *Ad Hoc Networks*, vol. 7, no. 7, 2009.
[36] A. De Domenico, E. Calvanese Strinati, and M.-G. Di Benedetto, "A survey on MAC strategies for cognitive radio networks," *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, vol. 14, no. 1, 2012.

PLACE PHOTO HERE

MohammadJavad NoroozOliaee received the BS degree in Computer Engineering from Sherif University of Technology, Iran, in 2009. He is currently working toward the PhD degree in Computer Science at Oregon State University. His research focus is on the design and development of distributed coordination techniques for wireless networking systems with dynamic spectrum access capabilities.

PLACE PHOTO HERE

Bechir Hamdaoui received the Diploma of Graduate Engineer (1997) from the National School of Engineers at Tunis, Tunisia. He also received M.S. degrees in both Electrical and Computer Engineering (2002) and Computer Sciences (2004), and the Ph.D. degree in Computer Engineering (2005) all from the University of Wisconsin at Madison. In September of 2005, he joined the RTCL Lab at the University of Michigan at Ann Arbor as a postdoctoral researcher. Since September of 2007, he has been with the School of EECS at Oregon State University as an assistant professor. His research focus is on cross-layer protocol design, system performance modeling and analysis, adaptive and learning technique development, and resource and service management for next-generation wireless networks and communications systems. He has won the NSF CAREER Award (2009), and is presently an Associate Editor for IEEE Transactions on Vehicular Technology (2009-present) and Wireless Communications and Computing Journal (2009-present). He also served as an Associate Editor for Journal of Computer Systems, Networks, and Communications (2007-2009). He served as the chair for the 2011 ACM MobiCom's Student Research Competition, and as the program chair/co-chair of the Pervasive Wireless Networking Workshop (PERCOM 2009), the WiMAX/WiBro Services and QoS Management Symposium (IWCMC 2009), the Broadband Wireless Access Symposium (IWCMC 2010), the Cooperative and Cognitive Networks Workshop (IWCMC 2011 and 2012), and the Internet of Things, Machine to Machine, and Smart Services Applications Workshop (CTS 2012). He also served on program committees of several IEEE conferences. He is a member of IEEE, IEEE Computer Society, IEEE Vehicular Society, and IEEE Communications Society.

PLACE PHOTO HERE

Mohsen Guizani S'85-M'89-SM'99-F'09) is currently a Professor and the Associate Vice President for Graduate Studies at Qatar University, Doha, Qatar. He was the Chair of the Computer Science Department at Western Michigan University from 2002 to 2006 and Chair of the Computer Science Department at 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, from Syracuse University, Syracuse, New York. His research interests include Computer Networks, Wireless Communications and Mobile Computing, and Optical Networking. He currently serves on the editorial boards of six 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 seven books and more than 270 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. Dr. Guizani served as the Chair of IEEE Communications Society Wireless Technical Committee (WTC) and Chair of TAOS Technical Committee. He was an IEEE Computer Society Distinguished Lecturer from 2003 to 2005. Dr. Guizani is an IEEE Fellow and a Senior member of ACM.