Adaptive Service Function for System Reward Maximization Under Elastic Traffic Model

MohammadJavad NoroozOliaee, Bechir Hamdaoui, and Mohsen Guizani†
Oregon State University, Email: noroozom,hamdaoub@onid.orst.edu
† Qatar University, Email: mguizani@ieee.org

***Abstract*—We propose an adaptive service model that maximizes the amount of service that spectrum users (SUs) achieve from accessing DSA systems. The proposed model allows SUs to utilize available spectrum efficiently by enabling them to locate spectrum opportunities in a distributed manner, thereby maximizing the long-term rewards that SUs receive. In this model, SUs adapt their required level of service with time depending on the amount of service they received so far. This proposed model is suitable for and can be used by existing objective functions. It leads to the maximization of the amount of service that SUs receive through DSA. Our simulation results show that the proposed adaptive model is very scalable by performing well regardless of the number of users in the system, and allows users to achieve high service rewards by quickly locating spectrum opportunities in the system.**

***Index Terms*—Elastic traffic; distributed dynamic spectrum access; adaptive access techniques.**

I. Introduction

Dynamic spectrum access (DSA) has been a key solution to the spectrum shortage problem [1–3]. It allows spectrum users (SU) to dynamically and adaptively access the spectrum bands, leading to significant improvement of spectrum efficiency. DSA created significant research interests ranging from the development of sensing techniques [4, 5] and algorithms [6–8] to the design of new architectures [9, 10]. Learning-based techniques have also been key to promoting successful DSA, as they can easily be implemented in a decentralized manner without requiring users to have any prior knowledge of the dynamics and characteristics of the DSA environment. Learning-based techniques rely on learning algorithms (e.g., reinforcement learners [11]) to learn from users' interaction and experience to decide what to do in the future. More specifically, learning algorithms allow SUs to use their knowledge obtained from their interactions with the environment to take the appropriate actions that maximize the long-term amount of service that they receive from accessing the DSA system.

The challenge with learning techniques is that when SUs do not design their objectives carefully, learning algorithms can eventually lead to poor overall system performance. This is because the collective behavior of the SUs aiming to maximize poorly designed objectives is likely to result in low received system service, thereby worsening the amount of service each SU receives. Therefore, it is important to come up with the appropriate objective design so that when SUs go after their objectives, their behavior as a whole leads to the maximization of the amount of service that each SU receives from accessing the DSA system. In [12, 13], we proposed efficient objective functions that are suitable for DSA, in that they lead to the maximization of the total system service that SUs receive from using such systems.

In this work, we propose an adaptive service model that can be used by SUs to compute the rewards they receive from using the DSA system. This proposed service model complements the objective functions that we proposed in [12, 13], in that when used by these objective functions, it enhances the amount of service that each SU receives in the long run. Using simulations, we show that the proposed model, when used by the appropriate objective functions, promotes successful DSA. It enables SUs to achieve high rewards by allowing them to quickly locate and exploit spectrum opportunities. It is also very scalable in that it performs well regardless of the number of SUs in the system.

The rest of the paper is organized as follows. In Section II, we state the problem studied in this work. In Section III, we present the proposed service model and overview the objective functions used in this work. In Section IV, we derive the optimal performance behaviors. In Section V, we evaluate the performances of the proposed model and compare them with those of existing models. Finally, we conclude the paper in Section VI.

II. Problem Statement

Throughout this work, an *agent* is used to refer to any group of two or more SUs that want to communicate with each other. All members of each group must then switch to the same spectrum band prior to beginning their communication. We assume that spectrum is divided into m non-overlapping bands, and consider a time-slotted system where agents are assumed to arrive and leave at the beginning and at the end of time slots. At each time step, each agent using a band receives a service that is passed to it from that band. The amount of service that the band offers an agent can be measured in terms of, for example, amount of throughput, reliability of the communication, the signal to noise ratio, the packet success rates, etc. We assume that once the agent switches to a particular band, it can immediately quantify and measure the amount of service that it receives from using the band. The methods that agents use to quantify and measure the service received as a result of using any particular band are beyond the scope of this work. Let V_j be the total amount of service that spectrum band j offers.

In this work, our focus is on developing methods that allow agents to access and use the DSA system in a distributed

manner. Specifically, the assumption here is that agents ought to rely on and implement some learning algorithms (e.g., a reinforcement learner [11, 14]) to enable them to learn about and find good spectrum opportunities in the system. Agents, independently of one another, use their learners to select the best available spectrum bands, and do so either periodically (every time episode) or reactively (whenever their received rewards drop below a certain threshold). That is said, this work proposes techniques that complement the learning algorithms in that it enhances the amount of service that each agent receives in the long run when used by these algorithms. Although our proposed techniques are not developed for any specific learners, we choose to use throughout this work the ϵ-greedy Q-learner [11] with a discount rate of 0 and an ϵ value of 0.05 for the sake of evaluating our proposed model. For more details on the Q-learner, readers can refer to [11].

III. System Model

We present the proposed service model designed to fit elastic traffic in DSA systems, and for completeness, we also overview the objective functions used in this work.

A. Elastic Reward Model

In this paper, we consider the elastic traffic model, which is suitable for elastic applications such as file transfer and web browsing. Under this model, an agent's reward is the amount of service it receives from using the spectrum band when the agent's received level of service (LoS) is above a certain threshold. That is, the higher the amount of service, the greater the reward. But when the received LoS drops below a certain (typically low) threshold, the agent's reward becomes unacceptable very quickly. In other words, the reward can decrease exponentially with the received LoS when the received LoS is below the threshold. Formally, the reward of agent i, $r_i(t)$, at time t can be written as

$$r_i(t) = \begin{cases} S_i(t) & \text{if} \quad S_i(t) \geq Q_i(t) \\ Q_i(t) e^{-\beta \frac{Q_i(t) - S_i(t)}{S_i(t)}} & \text{otherwise} \end{cases} \tag{1}$$

where $Q_i(t)$ and $S_i(t)$ denote agent i's required and received LoS at time t, respectively, and β is a decaying factor.

From the system's perspective, the global or system reward, $G(t)$, can be defined as the sum of all agents' rewards. Formally, $G(t)$ at time step t can be expressed as

$$G(t) = \sum_{i=1}^{n} r_i(t) \tag{2}$$

where n denotes the total number of agents. As a special case, when all agents using band j are assumed to receive equal amount of service from their spectrum bands, and the required LoS of all agents is the same (i.e., $Q_i(t) = Q$ for all $t > 0$ and $i = 1, ..., n$), all agents in band j each receives $V_j/n_j(t)$ at time t where $n_j(t)$ is the number of agents using band j at that time. In this special scenario, the reward function can be expressed as

$$r_i(t) = \begin{cases} V_j/n_j(t) & \text{if} \quad n_j(t) \leq V_j/Q \\ Q e^{-\beta \frac{n_j(t)Q - V_j}{V_j}} & \text{otherwise} \end{cases} \tag{3}$$

and band j reward function can be expressed as

$$G_j(t) = \begin{cases} V_j & \text{if} \quad n_j(t) \leq V_j/Q \\ n_j(t) Q e^{-\beta \frac{n_j(t)Q - V_j}{V_j}} & \text{otherwise} \end{cases} \tag{4}$$

For illustration purposes, we show in Fig. 1 the elastic reward function $r_i(t)$ of agent i when using band j as a function of $n_j(t)$ for $\beta = 20$ and $V_j/Q = 4$.

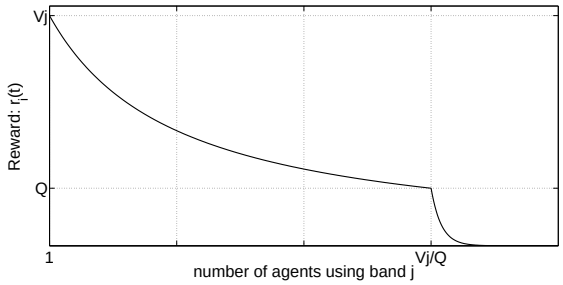


Fig. 1. Reward function $r_i(t)$: $\beta = 20$ and $V_j/Q = 4$ for all j.

B. Agent's Required Level of Service

We consider that each agent i has a total required LoS, Q_{total}, that should be received by a given target time period, T. We assume that the required LoS at time step t, $Q_i(t)$, changes adaptively based on the required LoS that has not been received yet and the target period T. Formally,

$$Q_i(t) = \frac{Q_{total} - \sum_{t'=1}^{t} S_i(t')}{T - t} \tag{5}$$

As mentioned earlier, at the end of each time step, by means of a reinforcement learner [11], each agent selects the "best" available band, and uses it during the next time step. Herein, we assume that the agents change their band and select the best spectrum band if and only if the agent has received less than the required LoS at that time step. The band selection method is as stated in [12]. In other words, an agent does not need to change its operating band unless it is not satisfied with the level of service it has received.

C. Objective Functions

Let $g_i(t)$ denote the objective function that agent i should go after in order to maximize the global received rewards. Here, the agents try to maximize their objectives by means of the Q-learner [11]. We now briefly describe the objective functions used in this work to evaluate the proposed model. More details on these functions can be found in [12, 13].

Agent reward function: each agent aims to maximize its own reward function; i.e., $g_i(t) = r_i(t)$ for each agent i.

Global reward function: each agent tries to maximize the global rewards received by all agents. That is, agent i's objective function is the same as the global reward $G(t)$; i.e., $g_i(t) = G(t)$ for each agent i.

Difference reward function: each agent aims to maximize its own contribution to the global reward, which is referred to as the difference function [15], and denoted by $D_i(t)$. That is, $g_i(t) = D_i(t)$ for each agent i, where

$$D_i(t) = G(t) - G_{-i}(t) \tag{6}$$

and $G_{-i}(t)$ is the global reward when agent i is absent from the system.

Team contribution reward function: each agent aims to maximize its team contribution to the global reward. This objective function is called team contribution function [13] and is denoted by $T_i(t)$. With this function choice, agent i's function, $g_i(t)$, is set to $T_i(t)$, where

$$T_i(t) = \begin{cases} \sum_{j=1}^{n} \delta_i(j) D_j(t) & \text{if} \quad S_i(t) \geq Q_i(t) \\ D_i(t) & \text{otherwise} \end{cases} \tag{7}$$

where $\delta_i(j)$ is equal to 1 if agent i and agent j are using the same band and 0 otherwise. Note that the focus of this work is not on the design of objective functions, but rather on the design of service models that are suitable for elastic traffic.

IV. Maximal Achievable System Rewards

We now analytically derive the maximum achievable global reward when the bands offer different LoS values. We define band j's capacity c_j as V_j/Q where V_j is again the total LoS offered by band j and Q is agents' required LoS. The following lemma will be used later for deriving the maximal achievable reward of our studied system.

Lemma 4.1: Consider two bands whose numbers of agents exceeding their capacities are the same. The system/global reward reduces less if a new agent joins the spectrum band with the least LoS value between the two bands.

Proof: Assume that there are two bands i and j offering V_i and V_j LoS values with band capacity c_i and c_j, respectively. Assume $V_i > V_j$ which implies $c_i > c_j$. Assume that the number of agents in bands i and j are $n_i = c_i + k$ and $n_j = c_j + k$, respectively for $k \geq 0$. Recall from Eq. (4) that when band j has $n_j \geq c_j$ agents, its reward is $G_j(n_j) = n_j Q e^{-\beta(\frac{n_j}{c_j}-1)}$. If a new agent joins this band, the reward becomes $G_j(n_j+1) = (n_j+1)Qe^{-\beta(\frac{n_j+1}{c_j}-1)}$. It can easily be shown that when $n_j = c_j + k \geq 1$, $G_j(n_j) > G_j(n_j+1)$; i.e., the reward when joining band j decreases by $\epsilon_j(k) \equiv G_j(c_j+k) - G_j(c_j+k+1)$. Now we can easily see that $\epsilon_i(k) > \epsilon_j(k)$. Hence, if the channel capacity is lower, the reward decreases less. ■

Let n and m denote the total number of agents and the total number of spectrum bands in the system.

Theorem 4.2: When there are n agents in the system and the bands capacities are $c_1, c_2, ..., c_m$, the global reward reaches its maximal only when the band with the lowest capacity, c_{min}, has $n - \sum_{l=1, c_l \neq c_{min}}^{m} c_l$ agents and any other band j has exactly c_j agents.

Proof: Without loss of generality, let's assume that $V_1 \leq V_2 \leq V_3 \leq ... \leq V_m$. Let $k = n - \sum_{l=1}^{m} c_l$ and let us refer to the agent distribution stated in the theorem as C. Note that C corresponds to when $n_j = c_j$, for $j = 2, 3, ..., m$ and $n_1 = c_1 + k$. We proceed with the proof by comparing C with any possible distribution C' among all possible distributions. Let $n_j = c_j + k_j \quad (k_j \geq -c_j)$ be the number of agents in band j in C', we know that

$$\sum_{j=1, k_j \geq 0}^{m} k_j \geq k. \tag{8}$$

since we are eliminating the negative values from the summation. Let $\epsilon_j(k_j)$ be the amount by which the global reward is reduced when an agent joins band j and the band j has already $n_j = c_j + k_j$ agents. From Lemma 4.1, it follows that $\epsilon_i(k') \leq \epsilon_j(k') > 0$ when $i < j$. And from [12], it follows that $0 < \epsilon_i(k') < \epsilon_i(k'+1)$ for $k \geq 0$. Note that for the distribution C, the global reward is reduced by $u = \sum_{i=0}^{k} \epsilon_1(i)$, and for C', it is reduced by $u' = \sum_{j=1, k_j \geq 0}^{m} \sum_{i=0}^{k_j} \epsilon_j(i)$. It remains to show that $u' - u > 0$ for any $C' \neq C$. $u' - u$ can be expressed as

$$u' - u = \sum_{\substack{j=1 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) - \sum_{i=0}^{k} \epsilon_1(i)$$

and as stated in Eq. (8), we know that the number of terms in part a is more than the number of terms in part b. We consider three different scenarios:

- $k_1 > k$: Here, we have

$$\begin{aligned} u' - u &= \sum_{\substack{j=1 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) - \sum_{i=0}^{k} \epsilon_1(i) \\ &= \sum_{i=k+1}^{k_1} \epsilon_1(i) + \sum_{\substack{j=2 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) \end{aligned}$$

  which is greater than zero since all the terms are positive.

- $k_1 = k$: In this scenario, we have

$$\begin{aligned} u' - u &= \sum_{\substack{j=1 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) - \sum_{i=0}^{k} \epsilon_1(i) \\ &= \sum_{\substack{j=2 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) \end{aligned}$$

  which is also greater than zero.

- $k_1 < k$: In this scenario, we have

$$u' - u = \sum_{\substack{j=1 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i) - \sum_{i=0}^{k} \epsilon_1(i)$$

$$= \underbrace{\sum_{\substack{j=2 \\ k_j \geq 0}}^{m} \sum_{i=0}^{k_j} \epsilon_j(i)}_{part\ a} - \underbrace{\sum_{i=k_1+1}^{k} \epsilon_1(i)}_{part\ b}$$

From Eq. (8), it follows that the number of terms in $part\ a$ is greater than that in $part\ b$. Thus, we can find a term from $part\ a$ for every term in $part\ b$ which is greater than that term since $\epsilon_i(k') > \epsilon_j(k')$ for $i > j$ as stated in Lemma 4.1 and $\epsilon_i(k'') < \epsilon_i(k')$ for $k'' > k'$ as proved in [12]. Moreover, the remaining terms in $part\ a$ are positive which implies that $u' - u > 0$.

In all scenarios, we showed that $u' - u > 0$. Therefore, the global reward for any distribution C' is smaller than that for the distribution C; i.e., C is the distribution that corresponds to the maximal achievable global reward. ■

Corollary 4.3: The system/global reward that a DSA system can achieve is at most

$$\sum_{j=2}^{m} V_j + (n - \sum_{j=2}^{m} c_j) \exp(-\beta(\frac{n - \sum_{j=2}^{m} c_j}{c_1} - 1)) \quad (9)$$

Proof: The stated achievable global reward can be calculated using Theorem 4.2. ■

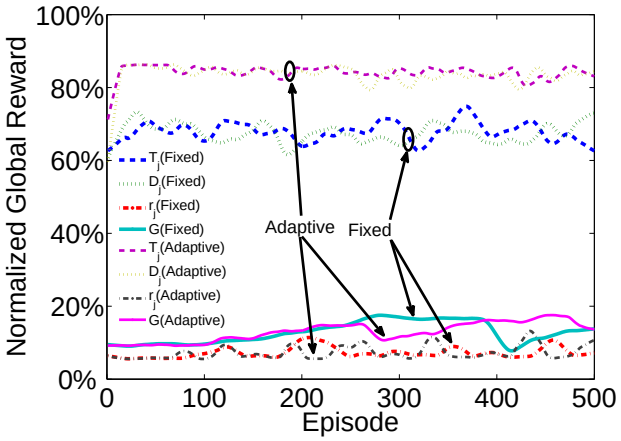
V. Performance Evaluation

In this section, we evaluate the performance of the proposed service model in terms of the achievable global reward, and compare it with that achieved under the model proposed in [12] for each of the aforementioned objective functions: agent reward ($g_i = r_j$), global reward ($g_i = G$), difference reward ($g_i = D_i$), and team contribution reward ($g_i = T_i$). In our model, the required LoS by agents adaptively changes through time whereas, in the model proposed in [12], the required LoS remains fixed at every time step. Thus, in this section, we refer to the our newly proposed model as *adaptive model* and the model used in [12] as *fixed model*.

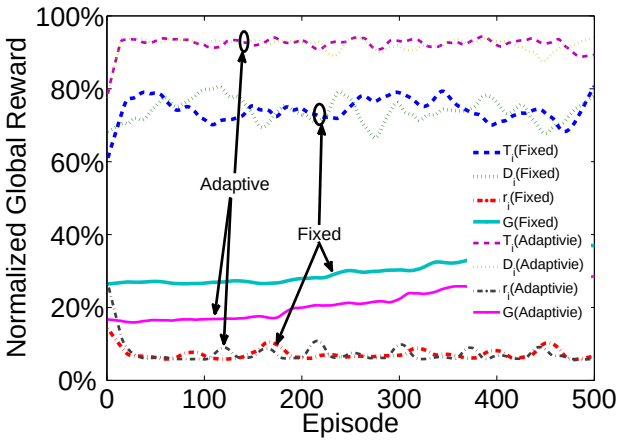
We define the variability of spectrum bands' LoS values, ψ, as the difference between the minimum, V_{min}, and the average, V_{avg}, of the total LoS values offered by the bands. We can write $\psi = \frac{V_{min} - V_{avg}}{V_{avg}}$. Unless stated otherwise, throughout this evaluation section, the decaying factor β is set to 2, the number of agents is set to 300, the number of bands is set to 10 and the average LoS, V_{avg}, is set to 20.

A. Global Reward Behavior

Fig. 2 shows global reward normalized with respect to optimal achievable rewards derived and stated in Corollary 4.3 under the studied service models and objective functions. In Fig. 2(a), the LoS offered by each band is the same for all bands (i.e., $\psi = 0\%$), whereas in Fig. 2(b), ψ is set to 80%. Note that the proposed adaptive model outperforms the fixed model under each of the studied objective functions. The adaptive model achieves about 90% of the optimal achievable reward while the fixed model achieves almost 70% of the optimal achievable reward.



(a) $\psi = 0\%$



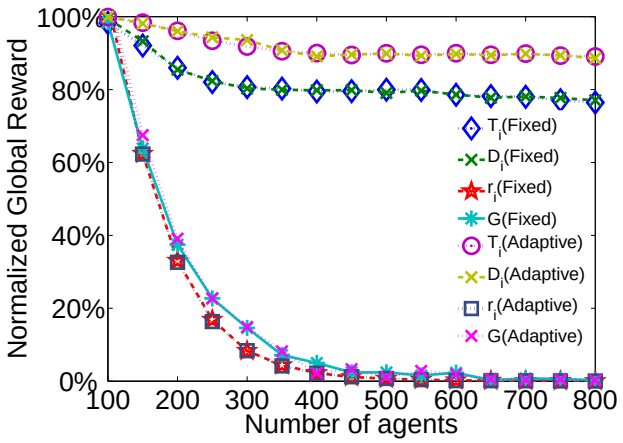
(b) $\psi = 80\%$

Fig. 2. Normalized global reward under various time steps. (a) $\psi = 0\%$ and (b) $\psi = 80\%$

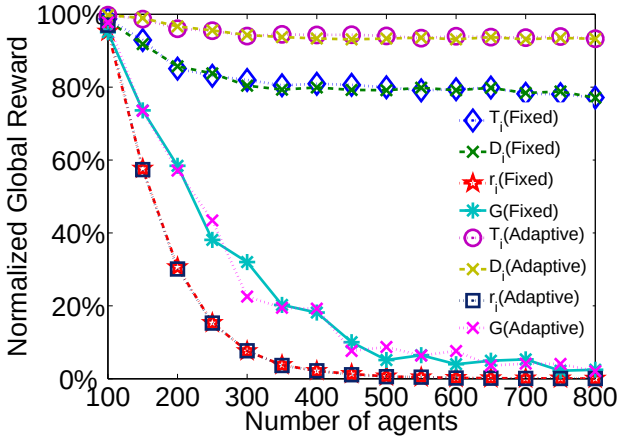
B. Scalability

In order to study the performance of the proposed model in terms of scalability, we compare in Fig. 3 the normalized global reward under each of the two service models while varying the number of agents, n, from 100 to 800 and fixing the number of bands m to 10. Figs. 3(a) and 3(b) show these results when $\psi = 0\%$ and $\psi = 80\%$, respectively.

The figures show that the proposed adaptive model outperforms the fixed model regardless of the number of agents in the system. These also show that both models are scalable when the difference and team contribution objective functions are used, but not scalable under the other two functions.



(a) $\psi = 0\%$



(b) $\psi = 80\%$

Fig. 3. Normalized global reward for various number of agents: (a) $\psi = 0\%$ and (b) $\psi = 80\%$

C. LoS Variability

Fig. 4 plots the normalized global reward for different values of ψ. Here, V_{avg} is kept equal to 20. The figure shows that the adaptive model outperforms the fixed one regardless of the value of ψ. While the fixed model achieves about 80% of the maximum possible achievable reward, the adaptive model achieves about 95% of that same maximum reward.

VI. Conclusion

This paper proposes an adaptive service model for promoting efficient DSA. The proposed model enables SUs to explore and exploit spectrum opportunities dynamically and efficiently, thereby maximizing the long-term rewards that SUs receive. In this model, SUs adapt their required LoS with time based on the LoS they received so far. This proposed service model complements existing objective functions in that it enhances the amount of service that each SU receives from accessing the DSA system. Our results show that the proposed model is very scalable and enables SUs to achieve high rewards by allowing them to quickly locate and exploit available spectrum opportunities.

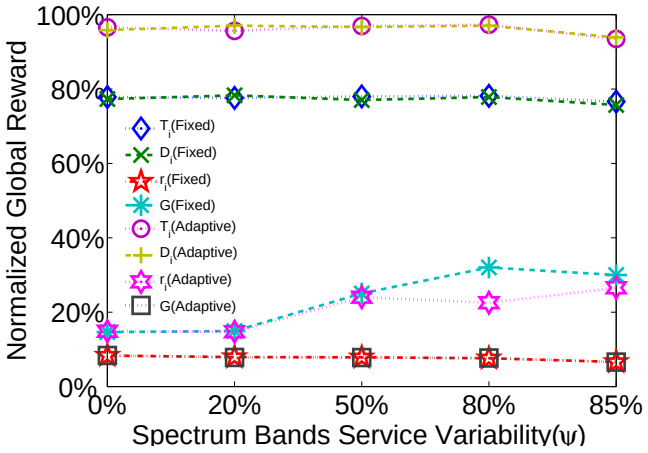


Fig. 4. Normalized global reward when varying LoS variabilities

VII. Acknowledgment

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.

References

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

[2] *FCC, Spectrum Policy Task Force (SPTF), Report of the Spectrum Efficiency WG, Report ET Docet no. 02-135, November*, 2002.

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

[4] A. Ghasmi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, 2008.

[5] R. Fan and H. Jiang, "Optimal multi-channel cooperative sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, March 2010.

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

[7] M. Ma and D. H. K. Tsang, "Joint design of spectrum sharing and routing with channel heterogeneity in cognitive radio networks," *Physical Communication*, vol. 2, no. 1-2, 2009.

[8] 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 Veh. Technology*, vol. 59, no. 1, 2010.

[9] G. Gur, S. Bayhan, and F. Alagoz, "Cognitive femtocell networks: an overlay architecture for localized dynamic spectrum access," *IEEE Wireless Communications*, vol. 17, no. 4, 2010.

[10] T. R. Newman, S. M. S. Hasan, D. Depoy, T. Bose, and J. H. Reed, "Designing and deploying a building-wide cognitive radio network testbed," *IEEE Communications Magazine*, vol. 48, no. 9, 2010.

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

[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*, May 2013.

[13] M. NoroozOliaee, B. Hamdaoui, and M. Guizani, "Maximizing secondary-user satisfaction in large-scale dsa systems through distributed team cooperation," *IEEE Trans. on Wireless Comm.*, Oct 2012.

[14] G. Tesauro, "Practical issues in temporal difference learning," *MLJ*, vol. 8, pp. 257–277, 1992.

[15] A. K. Agogino and K. Tumer, "Efficient evaluation functions for evolving coordination," *Evolutionary Computation*, vol. 16, no. 2, pp. 257–288, 2008.