Distributed Resource and Service Management for Large-Scale Dynamic Spectrum Access Systems Through Coordinated Learning

M. NoroozOliaee and B. Hamdaoui

Oregon State University, Corvallis, OR 97331
noroozom,hamdaoub@onid.orst.edu

***Abstract*— We develop resource and service management techniques to support spectrum users (SUs) with quality of service 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 long-term service satisfaction levels that SUs receive from accessing and using the DSA system. 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.**

Keywords: Distributed resource allocation; management techniques; dynamic spectrum access; cognitive networks.

I. Introduction

Dynamic spectrum access (DSA) has been recognized as a key networking solution for solving the recently observed shortage problem in spectrum supply [1–3]. It improves spectrum efficiency by allowing dynamic access and management of spectrum resources by spectrum users (SUs) themselves with no to little involvement of centralized regulatory bodies. As a result of DSA's apparent potential, there has been a significant research interest in the development of learning techniques to promote effective DSA [4–7]. Learning-based techniques are of a particular interest to DSA because they can easily be implemented in a decentralized manner without requiring any prior knowledge of the dynamics and characteristics of the DSA environment. Instead, these techniques rely on learning algorithms (e.g., reinforcement learners [8]) to learn from past and present interaction experience to decide what to do best in the future. More specifically, learning algorithms allow SUs to use their knowledge acquired from these 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.

The challenge with learning techniques is that when SUs do not choose and coordinate 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 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 an efficient use of the spectrum resources, thus in turn leading to the maximization of the amount of service that each SU receives from accessing the DSA system in the long term.

In this work, we propose efficient management techniques that improve the spectrum resource utilization by maximizing the total amount of service that a DSA system offers its SUs. We consider a time-slotted DSA system with multiple, non-overlapping spectrum bands, where SUs are assumed to arrive and leave at the beginning and at the end of time slots. We also consider that each SU implements and uses a learning algorithm (e.g., a reinforcement learner [8]) to allow it to maximize its own objective function, enabling it then to locate and select the best available spectrum opportunities. The proposed resource management techniques ensure that the collective behavior of SUs aiming to maximize their own objectives indeed leads to a good overall system performance, resulting in maximizing the amount of service that each SU receives in the long run.

Using simulations, we show that our proposed techniques are optimal, scalable, distributive, and fair. First, they enable SUs to achieve high service satisfaction levels by allowing them to quickly locate and exploit available spectrum opportunities. Second, they are very scalable as they perform well in systems with a small as well as a large number of SUs. Third, they can be implemented in a decentralized manner by relying on local information sharing only. Finally, they ensure fairness among SUs by allowing them to receive approximately equal amounts of service.

The rest of the paper is organized as follows. In Section II, we present the model and describe the motivation of this work. In Section III, we present our proposed resource and service management techniques. In Section IV, we derive the optimal performance behaviors. We evaluate the performances of the proposed techniques, and compare them with those achievable under existing approaches in Section V. Finally, we conclude the paper in Section VI.

II. Problem Statement

When the members of a group of two or more SUs want to communicate with each other, all members of the group must first select and switch to the same spectrum band to be able to carry out a communication among them; in the remainder of the paper, we will refer to these groups as *agents*. 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 such a 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. Throughout, let V_j be the total amount of service that spectrum band j offers.

Although the proposed resource and service management techniques can be used by all learning algorithms, we choose to use throughout this work the ϵ-greedy Q-learner [8] with a discount rate of 0 and an ϵ value of 0.05 for the purpose of evaluating these proposed techniques. For more details on the Q-learner, readers are referred to [8]. We want to mention that this work in not on learning, but rather on developing techniques 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 less than what is required (i.e., Q) is not acceptable, while receiving an amount higher than what is required is not beneficial either, which explains why the service satisfaction level remains constant. 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 the number of agents using band j 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 amount of service they receive from band j if the band has more agents than c_j (c_j here represents band j's capacity; i.e., the maximum number of agents that the band can support while satisfying their required service levels).

For illustration purposes, we show in Fig. 1 the service satisfaction level $s_j(t)$ each agent receives from using band j as a function of the number of agents $n_j(t)$ using band j for $\beta = 20$ and $V_j/Q = 4$.

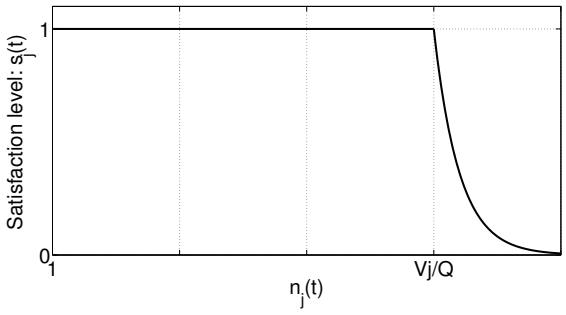

Fig. 1. Service satisfaction level: $\beta = 20$ and $V_j/Q = 4$ for all $j = 1, 2, \ldots, m$.

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 objective function g_i should each agent i maximize so that its received service satisfaction level is maximized?

Intuitively, one can think of two function choices. One possible objective 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. 2 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

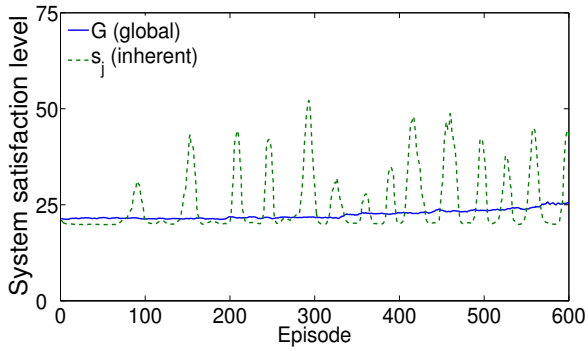

Fig. 2. System service satisfaction level under the two private objective functions: inherent choice ($g_i = s_j$) and global choice ($g_i = G$) for $m = 10$, $\beta = 2$, and $V_j/Q = 50$ for $j = 1, 2, \ldots, 10$.

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, which explains as follows. With the inherent objective function, an agent's received service satisfaction level, by design, is sensitive to its 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 choice (i.e., when each agent i sets its objective function g_i to the global service satisfaction level function G) results in a steadier performance behavior where the system received service satisfaction level 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.

III. Distributed Resource and Service Management Technique

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; i.e., their collective behavior will not worsen each other's received service satisfaction level. Throughout, g_i denotes the objective function of agent i that we aim to derive in this work.

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 [9] 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, yet without compromising its alignedness quality. These difference functions have been shown to perform well in various domains, such as multi-robot coordination [10] and air traffic control [11], and 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. 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 without compromising its alignedness level. The difference function D_i can be thought of as the *individual or agent contribution* to the system.

Now by substituting Eq. (2) into Eq. (3), 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 objective 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; i.e., those which share with it the resource. 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 all agents' contributions. 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 levels), 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

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

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

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

It is important to note that, by taking away agent i from the second term of the function D_i (Eq. (3)), the terms corresponding to all spectrum bands k, except the band j that agent i is using, cancel out. This explains why D_i, as shown in Eq. (4), depends on band j only. Therefore, the proposed function T_i is simpler to compute than the global function G. More specifically and importantly, it is fully decentralized as agents implementing/using it as their objectives need to gather and share information only with the agents that belong to the same band. This is one important property among few others (to be described later) that this proposed function has.

IV. Optimal Service Satisfaction

In this section, we theoretically derive the maximum/optimal achievable service satisfaction level. This derivation will serve as a means of assessing how well the developed objection functions perform when compared not only with existing objective functions, but also with the optimal achievable performances.

Without loss of generality and for simplicity, let us 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. In what follows, we assume that $n > m\frac{V}{Q}$ (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, its service satisfaction level is $G_j(n') = n' e^{-\beta(\frac{n'}{c}-1)}$. If a new agent joins this band, the new service satisfaction level becomes $G_j(n'+1) = (n'+1)e^{-\beta(\frac{n'+1}{c}-1)}$. First, it can easily be shown that when $n' > c \geq 1$, $G_j(n') > G_j(n'+1)$; i.e., the service satisfaction level when joining band j decreases by $\Delta_j(n') \equiv G_j(n') - G_j(n'+1)$. Now we can easily see that $\Delta_j(n')$ increases when n' increases. Hence, the greater the number n' (i.e., the more crowded the band), the smaller the decrease in the service satisfaction level. ■

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: Let $k = n - mc$, and let us refer to the agent distribution stated in the theorem as C. Note that C corresponds to when $m-1$ bands each has exactly c agents and the other m-th band has the remaining $c+k$ agents (since $n - c(m-1) = c + k$). We proceed with the proof by comparing C with any possible distribution C' among all possible distributions. Let $c + k_1$ be the number of agents in the most crowded band in C', $c + k_2$ be the number of agents in the second most crowded band in C', and so forth. We just need to deal with the bands that each contains more than c agents. If there are p bands each containing more than c agents, then we know that $\sum_{i=1}^{p} k_i \geq k$.

For each band having $c + k'$ agents, let ϵ_i be the amount by which the global service satisfaction level is reduced when agent i joins the band for $i = 1, 2, \cdots, k'$. From Lemma 4.1, it follows that $\epsilon_i > \epsilon_{i+1} > 0$, for all $i = 1, 2, \cdots, k' - 1$.

Note that for the distribution C, the global service satisfaction level is reduced by $t = \sum_{i=1}^{k} \epsilon_i$, and for C', it is reduced by $t' = \sum_{i=1}^{k_1} \epsilon_i + \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i$. It remains to show that $t' - t > 0$ for any $C' \neq C$. We consider three different scenarios:

- $k_1 > k$: Here, we have

$$\begin{aligned} t' - t &= \sum_{i=1}^{k_1} \epsilon_i + \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i - \sum_{i=1}^{k} \epsilon_i \\ &= \sum_{i=k}^{k_1} \epsilon_i + \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i \end{aligned}$$

  which is greater than zero.
- $k_1 = k$: In this scenario, we have

$$\begin{aligned} t' - t &= \sum_{i=1}^{k_1} \epsilon_i + \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i - \sum_{i=1}^{k} \epsilon_i \\ &= \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i \end{aligned}$$

which is also greater than zero.

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

$$\begin{aligned} t' - t &= \sum_{i=1}^{k_1} \epsilon_i + \sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i - \sum_{i=1}^{k} \epsilon_i \\ &= \underbrace{\sum_{i=1}^{k_2} \epsilon_i + \cdots + \sum_{i=1}^{k_p} \epsilon_i}_{part\ a} - \underbrace{\sum_{i=k_1}^{k} \epsilon_i}_{part\ b} \end{aligned}$$

Since $k_1 + k_2 + \cdots + k_p \geq k$, the number of ϵ_i terms in $part\ a$ is greater than the number of terms in $part\ b$. From Lemma 4.1, we know that the largest term in $part\ b$ is ϵ_{k_1}, which is smaller than the smallest term ϵ_{k_2} in $part\ a$. Hence, $part\ a$ is greater than $part\ b$, and thus $t' - t$ is greater than zero.

In all scenarios, we showed that $t' - t > 0$. Therefore, the global service satisfaction level 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 service satisfaction level. ■

Corollary 4.3: The system service satisfaction level that a DSA system can achieve 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 service satisfaction level (that we derived and 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 objective functions, and compare them against this upper bound.

V. Performance Evaluation

In this section, we evaluate the effectiveness of the proposed objective functions in terms of their achievable system service satisfaction levels, and comparing them with those achievable under each of the functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed ($g_i = T_i$). Unless stated otherwise, throughout this evaluation, 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 $c_j = V_j/Q$ is set to 50 for all j.

A. Service Satisfaction Behaviors

Fig. 3 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 clearly shows that the proposed function T_i outperforms substantially the two intuitive functions, s_j and G, and outperforms the difference function by about 25% in terms of the overall system 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.

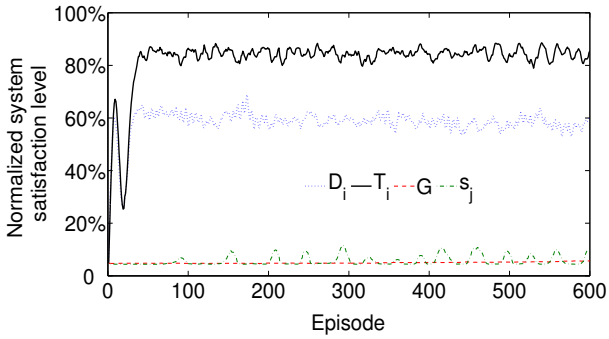

Fig. 3. Normalized system service satisfaction levels under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed ($g_i = T_i$) at various time steps.

B. Scalability Performance

In order to study the performance of the proposed functions in terms of scalability, we plot in Fig. 4 the normalized system service satisfaction level under each of the four studied objective functions when varying the number of agents, n, from 800 to 1600 while keeping the number of bands m equal to 10. In this and the next subsections, the system service satisfaction levels shown in the figures are all measured at episode 600 (basically, when the maximum level is attained). Observe that the proposed function T_i is highly scalable. Note

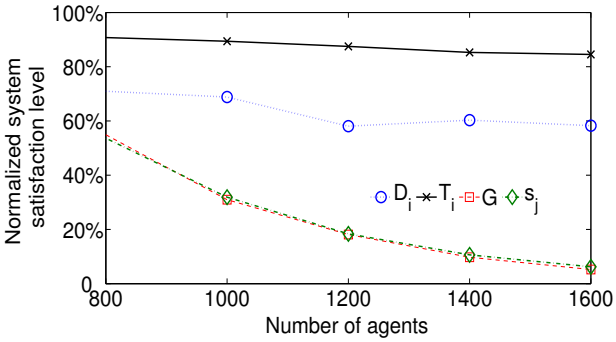

Fig. 4. Normalized system service satisfaction levels under inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed ($g_i = T_i$) functions for various numbers of agents.

that as the number of agents increases, T_i maintains high system service satisfaction levels, 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 function T_i still achieves satisfaction levels that are about 30% higher than those achievable under D_i.

We now plot in Fig. 5 the normalized system service satisfaction level achieved under each of the four functions, but for various values of the capacity c. The figure clearly shows that the proposed function T_i outperforms the other three function choices, even when varying the capacity of the spectrum bands.

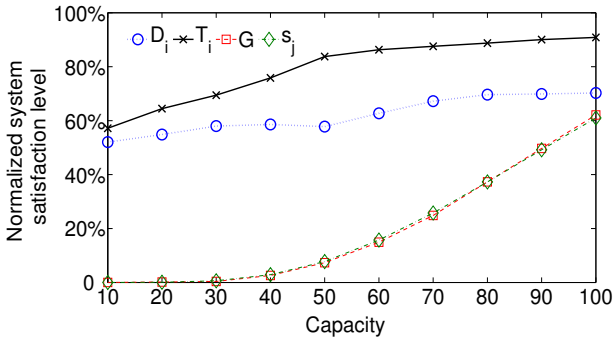

Fig. 5. Normalized system service satisfaction levels under the four studied functions: inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed ($g_i = T_i$) for various capacities.

C. Fairness Performance

To also see how well the proposed functions do when it comes to fairness, we plot in Fig. 6 the coefficient of variations (CoV)[1] of the received system service satisfaction levels for various numbers of agents. Observe that the proposed function achieves CoV values approximately similar to those achievable under any of the other three studied functions. These results show that not only 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.

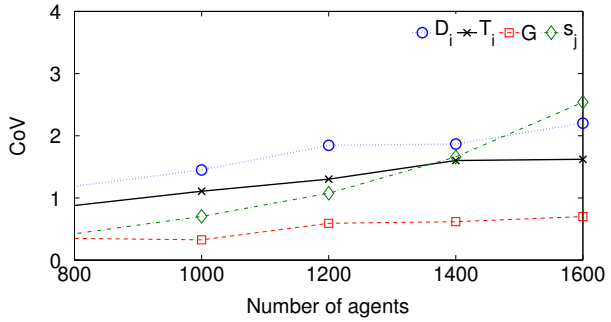

Fig. 6. Coefficient of variation (CoV) of satisfaction levels under inherent ($g_i = s_j$), global ($g_i = G$), difference ($g_i = D_i$), and proposed ($g_i = T_i$) functions for various numbers of agents.

VI. Conclusion

This paper proposed efficient resource and service management techniques to effectively support SUs in large-scale DSA systems. We showed that the proposed techniques achieve high service satisfaction levels, are very scalable by performing well in small- as well as large-scale systems, are highly learnable by reaching up high values fast, are distributive by requiring information sharing only among agents belonging to the same band, and ensure fairness among SUs by allowing them to receive equal amounts of service.

[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.

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] 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.

[5] 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.

[6] H. Liu, B. Krishnamachari, and Q. Zhao, "Cooperation and learning in multiuser opportunistic spectrum access," in *Proceedings of IEEE ICC*, 2008.

[7] 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.

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

[9] 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.

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

[11] 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.