Overcoming User Selfishness in DSA Systems Through Credit-Based Resource Allocation

Tamara AlShammari, Bechir Hamdaoui, Mohsen Guizani† and Ammar Rayes‡
Oregon State University | alshamta,hamdaoub@onid.orst.edu
† Qatar University | mguizani@ieee.org
‡ Cisco Systems | rayes@cisco.com

***Abstract*—We propose a credit-based resource allocation technique for wireless systems with dynamic spectrum access (DSA) capability, such as multichannel wireless sensor networks. The proposed technique is robust against selfish and malicious behaviors by achieving high performance independently of what users choose to pursue as their objectives. It also improves fairness by ensuring that different users are allocated equal amounts of spectrum service. Using simulations, we show that the proposed credit-based technique allows users to achieve high rewards/service even in the presence of misbehaved users that choose to pursue their greedy and selfish goals. We also show that it reduces the standard deviation of users' amounts of received service drastically, and hence, improves fairness among users substantially.**

I. INTRODUCTION

Dynamic spectrum access (DSA) is a new paradigm that allows spectrum users to seek and use spectrum bands opportunistically. DSA has great potential for improving the capacity of wireless mobile systems, such as wireless sensor networks, mobile networks, and cellular networks. In DSA systems, there are two types of users: licensed or primary users (PUs) and unlicensed or secondary users (SUs). DSA systems allow SUs to sense the licensed spectra, and to occupy and use any spectrum band when it is not used by its PUs. However, these SUs must be transparent to the PUs in that they have to leave the spectrum band as soon as they sense the presence of any PUs.

DSA has emerged as a potential solution for overcoming spectrum shortage problems [1]. As a result, many solutions have been proposed to address various aspects of DSA systems [2–4]. One of the key challenges arising from the complex and diverse nature of nowadays emerging wireless systems is the design of efficient DSA techniques that can be implemented in a distributed manner and can scale well with the number of users.

Learning-based techniques have been considered as a potential solution candidate for such a challenge due to their inherent distributed nature [5–7]. These techniques essentially propose distributed DSA methods that can perform without needing any control unit, thereby enabling SUs to distribute themselves among available bands/channels without any guidance or direction from any third party or entity. Learning-based techniques allow SUs to do so by using their knowledge to be acquired through their past and present environment interaction to decide what to do best in the future. For this, users often implement and go after some objective functions that they try to maximize by means of these learning algorithms.

It has been shown in literature that poorly designed objective functions can lead to poor system performances [8]. As a result, some research efforts have been put to develop efficient objective functions that are suitable for DSA systems [5, 9]. These previously proposed objective functions such as those proposed for elastic [5] and inelastic [9] traffic models are shown to have good performances in terms of optimality, scalability, and distributivity. However, they also have shortcomings: one, they are not robust against user misbehavior, in the sense that if some users choose (intentionally or unintentionally) to pursue selfish and greedy objectives, the overall system performance can degrade substantially. Two, they may be unfair to users, in that users that employ these proposed techniques may not receive equal amounts of service.

In this paper, we propose a credit-based spectrum resource allocation technique for DSA systems that, unlike the previous techniques, is robust against malicious and selfish behaviors and improves fairness among users. The robustness of our proposed technique against users' misbehavior (selfishness and maliciousness) lies in its ability to mask the impact that the users' pursued private objectives have on the overall system performance. Fairness improvements, on the other hand, are achieved by allocating service to users adaptively while accounting for the amount of service each user has received in the past. Our simulation results show that the proposed technique allows users to receive high service levels and ensures fairness among users even when the system contains misbehaved users.

The rest of this paper goes as follows. In Section II, we describe our system model. In Section III, we state our motivation and objective by illustrating the shortcomings of existing techniques. Section IV presents our proposed credit-based technique. In Section V, we evaluate and show the performance of our proposed technique. Section VI highlights and discusses some implementation and practical aspects of the proposed technique. Finally, we conclude the paper in Section VII.

II. System Model

We consider a wireless system with m non-overlapping spectrum bands (or channels). We assume that each band j offers an amount of service denoted by V_j; the service that the band offers could, for example, be throughput, reliability, data rate, etc. We also assume that there is an access point (or a monitoring agent) deployed in the system whose responsibility is to keep track of what and when users join the spectrum bands.

We consider the elastic traffic model in which a user's received reward corresponds to the amount of service it receives from using the spectrum when this received reward exceeds a certain threshold, Q. On the other hand, when the received amount of service is less than the threshold, the user's reward drops very quickly and becomes unacceptable. We assume that users do not leave their spectrum band (and try to find another band) unless their received level of service goes below their required level. In addition, we adopt the adaptive service model, proposed in [10], where the users' required level of service changes depending on what they have received so far. Mathematically, the reward, $r_i(t)$, of user i at time t can be written as [10]:

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

where $S_i(t)$ is user i's received level of service at time t, $Q(t)$ is the required level of service at time t, and β is the decaying factor. At last, we assume a time-slotted resource access and sharing scheme, where users are assumed to arrive at the beginning and leave at the end of time steps.

III. Fairness and Misbehavior

In learning-based DSA techniques, after a user determines its objective, it tries to maximize it using a learning algorithm. Two intuitive objective functions can be considered. The first one is the intrinsic reward function, r_i, given in Eq. (1), where here a user aims to maximize its own received reward; this function reflects the users' expected selfish behaviors when going after maximizing their own received rewards. The other one is the global/total reward function, G, which aims to maximize the total rewards received by all users. At time t, $G(t)$ can formally be written as

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

where $n(t)$ is the total number of users accessing the system at time t. The main drawback of using these two functions is that they lead to poor system performance. This is because in the intrinsic function case, users' objectives are not aligned with one another, and in the global function case, users' objectives are not sensitive enough to their own actions to lead to high rewards. Detailed and good explanations of such performance behaviors can be found in [5].

To address this performance issue, the difference objective function, D_i, has instead been used in DSA networks for supporting both elastic [5] and inelastic [9] traffic models, and is shown to achieve near-optimal performances by outperforming r_i and G substantially. The reason behind the high performance that D_i achieves lies in the fact that when the number of users in the system exceeds the channels' capacities, D_i leads to a near-optimal distribution of the users among the different available spectrum bands. As shown in [5], the optimal distribution occurs when $(m-1)$ channels/bands each has exactly a number of users equaling the channel's capacity, and the m^{th} band has all the other remaining users[1].

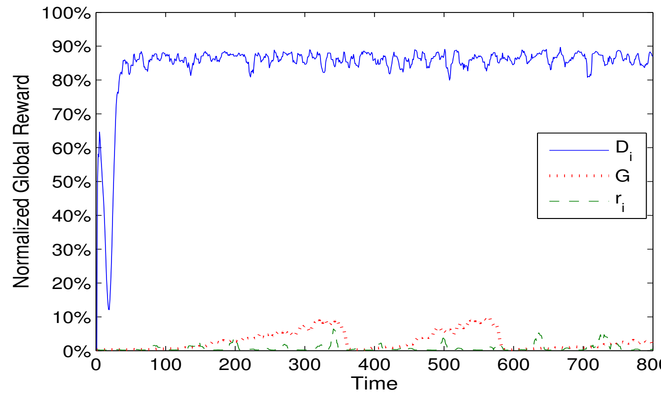

Fig. 1: Normalized global/total rewards achievable by all users under the functions D_i, r_i and G.

For illustration purposes, we consider in this section a DSA system with 10 bands and 500 users. Also, for simplicity and without loss of generality, we assume that all bands offer the same amount of service; i.e., V_j=V=20 for all j. In our figures, we normalize the global received reward with respect to an approximation of the maximal global achievable reward, given in [5]. We plot in Fig. 1 the normalized achievable global/total reward under D_i, r_i and G. As stated above, observe that r_i and G result in very poor performance, whereas D_i results in high performance. In addition to achieving high rewards, D_i is shown to scale well with the number of users, and can be implemented in a fully distributed manner in fully connected networks, as reported in [5].

Despite of its performance advantages, D_i has some shortcomings. First, it is unfair. This is because, under D_i, some users may end up staying in the most crowded channel more than others, thereby receiving smaller amounts of service. To illustrate, we show in Fig. 2 the standard deviation of users' received rewards under the D_i function for different number of users. As it can be seen from the figure, the standard deviations can be relatively high, implying that users may receive unequal amounts of service when D_i is used.

Second, the D_i function is not robust against misbehaved users. The issue is that even though D_i can increase the achievable performances, it can only do so when all users

[1]This is when all channels are assumed to offer the same service; i.e., $V_j = V$ for all j. Refer to [10] when V_js are not the same.

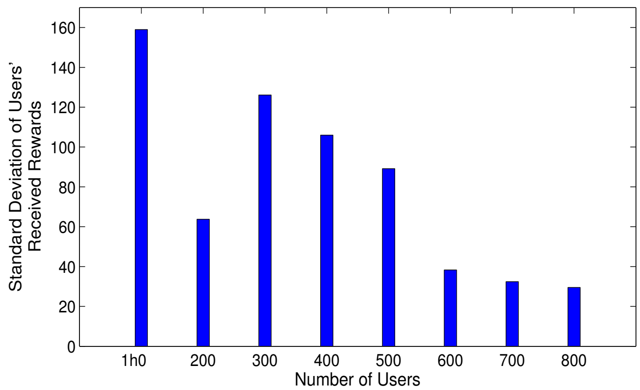

Fig. 2: Standard deviation of users' received rewards for different numbers of users under the D_i function.

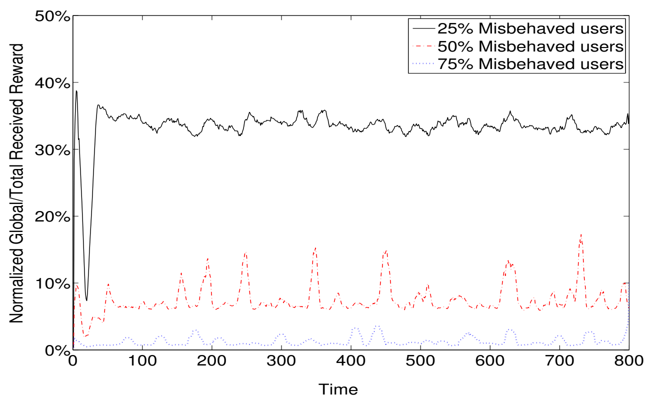

Fig. 3: Normalized global/total reward achievable by all users when D_i is used for various percentages of misbehaved users.

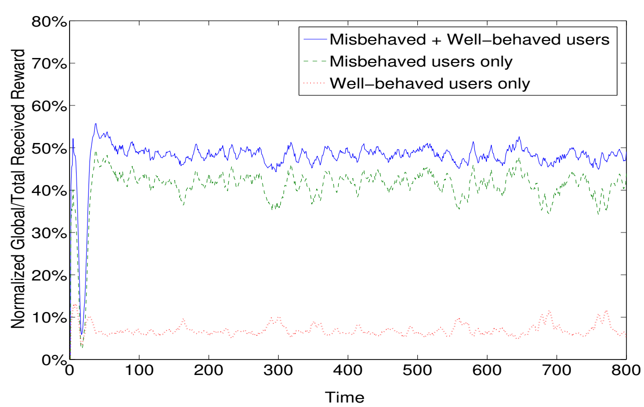

Fig. 4: Normalized global/total received reward when 20% of the users misbehave.

pursue it as their objectives. In other words, when some users choose (intentionally or unintentionally) not to pursue this function, the function can no longer lead to good performances. To illustrate this, we show in Fig. 3 the performance of D_i in the presence of misbehaved users. In the figure, "$x\%$ Misbehaved users" refers to the case when $x\%$ of the users choose to pursue their greedy objective function r_i, while the other $(100-x)\%$ users implement the D_i function. The figure shows that as the percentage (i.e., number) of misbehaved users increases, the overall system received reward decreases, and hence, so does the per-user average received reward. This overall performance degradation gets even worse when the percentage of misbehaved users becomes higher and higher.

What's even worse, not only do these misbehaved users lead to poor overall system performance, but also receive most of the available service, thereby leaving those well behaved users with no to little service. This is illustrated in Fig. 4. Note that the misbehaved users (which represent only 20% of all users) receive about 40% of the optimal/total amount of achievable service, whereas the well behaved ones (which represent 80% of all users) receive all together only about 10% of the total possible amount of service. Also, it is worth mentioning that because of the presence of these misbehaved users, the overall global/total system reward goes down from about 85% when all users behave well (as shown in Fig. 1) to about 50% only.

It is therefore important to devise efficient strategies and techniques that ensure fair allocation of resources among users while also maximizing the achievable system performance even in the presence of misbehaved users. Our proposed strategy for doing so consists of developing new resource allocation techniques that are immune from the users' objective function choices, in that even when users choose to deploy and pursue greedy goals and objectives, their collective selfish behavior does not lead to poor and unfair system performance. In addition, with our developed techniques, the amount of service a user receives depends not only on which channel the user selects, but also on how much service it has received so far so as to ensure fairness among users. To sum up, our proposed credit-based resource allocation technique possesses two characteristics:

- **Robustness against user selfishness.** It achieves high performance regardless of what objective functions users choose. In other words, it reduces the effect that objective functions have on the achievable performance.
- **Fairness among users.** It improves fairness among users by allocating service to users adaptively based on how much service each user has received in the past.

IV. Credit-based Resource Allocation

In this section, we present the proposed credit-based resource allocation technique designed to overcome selfish behaviors and ensure fairness among users. In this technique, each user is assigned a credit value with an initial value of one. This credit value determines the proportion of service that each user should receive; the greater the credit value, the higher the service to be received.

Under the proposed credit-based resource allocation technique, users' credit values get updated depending on the amount of service they receive when compared to the system fair-share. We define the system fair-share as the amount of service that each user should receive in order to ensure fair allocation of available service among all users. The maximum amount of system service is achieved when the users are distributed among the channels as follows [10]: each channel j contains exactly $b_j = V_j/Q$ users except the channel with the minimum capacity which should contain the remaining number of users. When $V_j = V$ for all channel j, this optimal distribution leads to a maximal value of global received reward that can be approximated to $\hat{G}_{max} = (m-1)V$ (derived in [10]). When the spectrum resources are allocated fairly among all users, each user i should then receive at time t a system fair-share, $R_i(t)$, that is equal to:

$$R_i(t) = \sum_{t'=t_i}^{t} \frac{(m-1)V}{n(t')} \tag{3}$$

where t_i is the time step at which user i joins the DSA system and $n(t)$ is again the total number of users present in the system at time t. When the number of users does not change over time (say $n(t) = n$ for all t) and all users enter and leave the system at the same time, $R_i(t) = \frac{(m-1)V}{n} t$ for every user i.

At the end of each time step, if the user receives less than the system fair-share, its credit value increases by one if it does not exceed a certain threshold, otherwise it is set to that threshold value. On the other hand, if the user receives more than the system fair-share, its credit value gets decreased by one when it is greater than the threshold, otherwise it is set to that threshold value. Consequently, users' credit values could be positive, zero, or negative. Whenever a user credit value reaches zero or a negative value, the user is no longer able to receive service until its credit value becomes positive again.

Indeed, when the user receives less than the system fair-share, our approach requires that its credit value does not exceed a certain threshold so as to prevent it from becoming very large. Otherwise, once the user receives its fair-share of the spectrum, it will take a relatively long time for its received service to be reduced so as to not exceed its fair-share. Likewise, in the case of receiving more than the system fair-share, a user credit value must not go below a certain threshold because otherwise its credit value would keep decreasing, and can reach a small value. When this happens, if this user, after sometime, wants to ramp up its share again, it will take it a long time before it can actually start receiving service.

Mathematically, a user i's credit value at time t, $Cr_i(t)$, is calculated as:

$$Cr_i(t) = \begin{cases} \max\{Cr_i(t-1)-1, Cr_i^{th}(t)\} & \text{if } R_i(t) < \sum_{t'=t_i}^{t} S_i(t') \\ \min\{Cr_i(t-1)+1, Cr_i^{th}(t)\} & \text{otherwise} \end{cases} \tag{4}$$

where the credit threshold bound is defined as:

$$Cr_i^{th}(t) = (R_i(t) - \sum_{t'=t_i}^{t} S_i(t'))/Q(t)$$

The numerator in the above equation represents either the missing service in case the user did not receive its whole fair-share or the extra service in case the user received more than its fair-share. In this equation, the amount of missing/extra service is represented as a multiple/fraction of user's required service at time t. This means that the user's credit value threshold represents how many Qs the user has to receive/miss in order to receive its fair-share.

For more clarification, let us consider the following example where we assume that at time t, user A is missing 4 units of service. If we further assume that user A's $Q(t)$ is equal to 2, then its credit threshold is 2. This means that user A needs $2Q$'s to compensate the missing service. Let us also assume, for illustration, that at the same time t, its credit value reaches that threshold, i.e. $Cr(t) = 2$. Since its credit value is positive, this user is able to receive service. At time $t+1$, this user receives an amount of service that is equal to Q plus the fair-share amount for this time step. Thus, its threshold at time $t+1$ gets updated, and becomes equal to 1. This implies that user A needs now just one Q to cover the missing service. Since its previous credit value is greater than its current threshold, its current credit becomes equal to the threshold, i.e. $Cr(t+1) = 1$. In the same manner, if this user at time $t+2$ receives $2Q$ plus the fair-share for one time step, its threshold value gets updated and becomes equal to -1. This means that this user is no longer missing any service, and has actually received one extra Q. The same credit updating process continues with reversing the condition for the credit value. That is, the user credit value must not go below the new threshold.

Using user i's credit at time $t-1$, the amount of service that user i receives from accessing band j at time t is:

$$S_i(t) = \begin{cases} \frac{Cr_i(t-1)}{\sum_{k \in B_j(t): Cr_k(t-1)>0} Cr_k(t-1)} V_j & \text{if } Cr_i(t-1) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $B_j(t)$ is the set of all users belonging to band j at time t.

It is worth mentioning that we here assume that our system is associated with an access point whose task is to monitor and keep track of users (their ID, their activities, their check-in and check-out times). One cannot rely on users to report their credit values as they might cheat in order to receive more service. Thus, we consider that both users and the access point are calculating and updating users' credit values. This way when a user lies about its credit value, the access point will know about the mismatch between the value it calculated and the value the user reported. As a result of this user behavior, the access point will block this user from accessing and using the DSA system.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed resource allocation technique in terms of the achievable global/total reward, the standard deviation of users' received rewards, and its robustness against user misbehavior. For simulation purposes, we use the ϵ-greedy Q-learning algorithm [11]. In this algorithm, at the end of each time step, the user selects the channel whose Q-value is the highest with probability 1-ϵ, and selects a random channel with probability ϵ. Whenever a user is tuned to a channel, it measures the service it receives[2], and then uses it to update the Q-value entry corresponding to that channel. Throughout this evaluation section, we assume that all users enter and leave the system at the same time; that is, the number of users is considered to be the same all time, and is equal to $n = 500$ unless stated otherwise. We also set m=10 and V_j=V=20 for all j.

A. Achievable Rewards

We show in Figs. 5 and 6 the normalized global/total reward achievable under each of the functions, r_i, G, and D_i, without and with the proposed credit-based resource allocation technique. Again, the results for the achievable rewards presented in this section are all normalized with respect to the maximal/total rewards that the system can achieve in the ideal scenario [5]. Fig. 5 shows that the proposed credit-based technique allows users to achieve high rewards/service even when they choose to pursue their greedy objective, r_i, or the global objective, G. The function D_i, on the other hand, already performs well in terms of the amount of achievable rewards, and adding the credit-based feature does only improve a little; this is illustrated in Fig. 6. But recall that, as mentioned in our early sections, the improvements of the proposed credit-based technique lies in its ability to combat selfishness/maliciousness by allowing users to achieve high amounts of service independently of what users pursue as objectives, and to ensure fairness among users also regardless of the objective function choice, as will be shown in next sections.

B. Robustness against objective function choice

We now show that the proposed credit-based technique reduces the impact of objective function choice on the system performance. Fig. 7 shows the system performance achievable in the presence of misbehaved users (users that choose not to use the D_i function as their objective) with and without the credit-based resource allocation technique. In the figure, "$x\%$ r_i, $(100-x)\%$ D_i" refers to the case when $x\%$ of the users choose to pursue their greedy objective function r_i, while the other $(100-x)\%$ users implement the D_i function. Consider, for e.g., the case when 25% of the users use the r_i function instead of the D_i function. Observe that when the proposed credit-based technique is not used, the normalized

[2]The methods used to calculate the received service are beyond the scope of this work

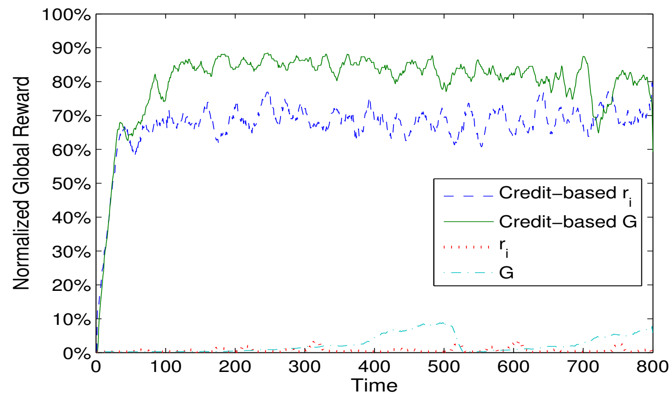

Fig. 5: Normalized global reward of r_i and G without and with the credit-based resource allocation technique.

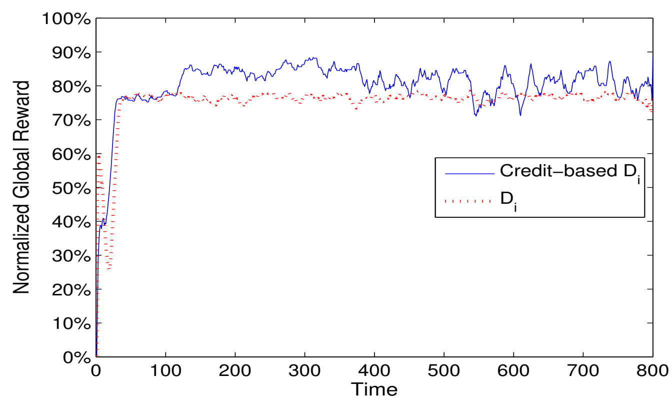

Fig. 6: Normalized global reward of D_i without and with the credit-based resource allocation technique.

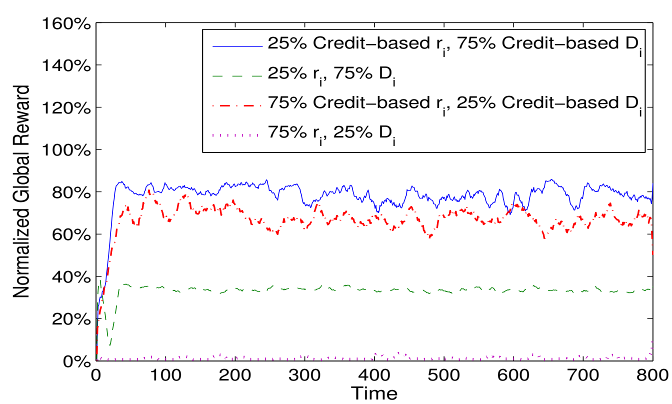

Fig. 7: Normalized global/total received reward without and with the credit-based resource allocation technique in the presence of misbehaved users.

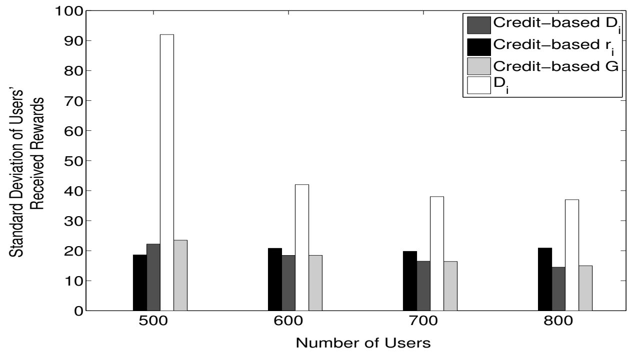

Fig. 8: Standard deviation of users' received rewards under different objective functions.

overall system performance is about 40% only, whereas when our technique is used, the system performance reaches about 80%. In this specific scenario, the adoption of our technique doubles the overall achievable performance. As you can see from the figure, this performance improvement can be even greater when the percentage of misbehaved users is higher. This is illustrated in the figure via the case corresponding to when 75% of the users use r_i.

C. *Fairness*

Fig. 8 shows the standard deviation of users' received rewards under the r_i, G, and D_i functions when using our proposed credit-based resource allocation technique for various numbers of users. The figure also shows the standard deviation of users' received rewards under D_i but without using the proposed technique. It can be seen from the figure that our proposed credit-based technique reduces the standard deviation drastically, and hence, it improves fairness among users substantially.

VI. Discussion

We have seen that the use of the proposed credit-based technique makes learning techniques robust against objective function choice and improves system fairness while still achieving high system rewards. Without this technique, the D_i function, on the other hand, does achieve good rewards as well, but only when all users use it as their objectives. In other words, if some (or all) users pursue other objectives, the overall system performance can degrade substantially. In addition, ensuring fairness can be very challenging due to the way users end up distributing themselves among the channels. One key advantage of the D_i function, however, lies in its fully distributed capability; it can be implemented and fully realized without needing any cooperation or centralized entity[2]. Ours can still be viewed as a distributed technique in the sense that users still choose and switch to their bands on their own will and without having any third entity tell them to do so. However, since our approach relies on and accounts for what users have received in the past to be able to decide what should be allocated in the future so that misbehavior is prevented and fairness is ensured, it requires the deployment of an access point to keep track of and monitor users' activities. This, however, is not unrealistic and can be done with minimum overhead.

[2]This depends, to a great extent, on the network topology and on other factors as well [5].

VII. Conclusion

This paper proposes a credit-based resource allocation technique that improves fairness and combats misbehavior in DSA systems. The proposed technique is robust against selfish behavior and achieves good performance independently of what users choose to pursue as their objectives. It also improves fairness by ensuring that users are allocated equal amounts of available spectrum service.

VIII. 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. The authors would also like to thank M. NoroozOliaee for his help with this work.

References

[1] W. Horne, P. Weed, and D. Schaefer, "Adaptive spectrum radio: a feasibility platform on the path to dynamic spectrum access," *Proc. of Int'l Symp. on Advanced Radio Technologies*, March 2003.

[2] Y. Lin, K. Liu, and H. Hsieh, "On using interference-aware spectrum sensing for dynamic spectrum access in cognitive radio networks," *IEEE Tran. on Mobile Computing*, vol. 12, pp. 461–474, March 2013.

[3] L. Cao and H. Zheng, "Balancing reliability and utilization in dynamic spectrum access," *IEEE/ACM Tran. on Networking*, vol. 20, June 2012.

[4] H. Li and Z. Han, "Catch me if you can: An abnormality detection approach for collaborative spectrum sensing in cognitive radio networks," *IEEE Tran. on Wireless Communications*, vol. 9, Nov. 2010.

[5] M. NoroozOliaee, B. Hamdaoui, and K. Tumer, "Efficient objective functions for coordinated learning in large-scale distributed OSA systems," *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, May 2013.

[6] 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.

[7] K. Liu and Q. Zhao, "Distributed learning in cognitive radio networks: multi-armed brandit with distributed mulitple players," *IEEE International Conference on Acousitcs, Speech, and Signal Processing*, 2010.

[8] K. Tumer and D. Wolpert, Eds., *Collectives and the Design of Complex Systems*, Springer, New York, 2004.

[9] B. Hamdaoui, M. NoroozOliaee, K. Tumer, and A. Rayes, "Coordinating secondary-user behaviors for inelastic traffic reward maximization in large-scale DSA networks," *IEEE Transactions on Network and Service Management*, December 2012.

[10] M. NoroozOliaee, B. Hamdaoui, and M. Guizani, "Adaptive service function for system reward maximization under elastic traffic model," *In Proc. of IEEE Global Communications Conf.*, December 2013.

[11] R.S. Sutton and A.G. Barto, *Reinforcement learning: An intoduction*, MIT, 1998.