Implementation and Analysis of Reward Functions under Different Traffic Models for Distributed DSA Systems

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

Abstract—**In this paper, we implement and analyze a resource allocation protocol for distributed dynamic spectrum allocation (DSA) systems. The DSA protocol is a learning-based protocol that allows secondary users (SU) to exploit the spectrum bands efficiently in a distributed manner without the need of information exchange. The implementation and test of the proposed protocol is done using ns3 assuming that the SUs selecting the same band share it in accordance with a carrier sense multiple access (CSMA) scheme. The evaluation of the proposed protocol is done under various traffic models. We show the importance of the objective function's choice; used as a utility to be maximized in the learning. We also show the impact of various practical aspects taken into consideration while implementing the protocol on the system's achieved performance.**

Index Terms—**dynamic spectrum allocation, distributed protocol, learning, traffic model, reward function, carrier sense multiple access.**

I. Introduction

The increase of utilization of wireless technology in the last two decades caused a serious shortage problem in the wireless spectrum. Moreover, studies done by the Federal Communications Commission (FCC) [1] show that some parts of the wireless spectrum are still under-utilized. Hence, DSA emerges as a potential solution for overcoming spectrum shortage. Many research attempts have recently tried to develop techniques and protocols for DSA systems that allow the exploitation of the wireless spectrum efficiently [2–6]. The focus has mostly been concentrated on the development of distributed approaches.

In order to allocate the channels distributively among the different users, the focus was initially on the game theoretical inspired solutions. [7–10]. For instance, an efficient DSA medium access control (MAC) protocol was developed in [7] based on game theory. In [8], the authors proposed game theoretical-based DSA techniques with respect to heterogeneous Quality-of-Service (QoS) requirements. In [10], the authors investigated a multi-channel network with random access. They compared the system performance between two scenarios: selfish systems and cooperative systems. They showed that incomplete information exchange significantly affects the system performance.

Recently, learning-based techniques have been proposed for distributed DSA. In [11, 12], it was shown that theses techniques can be easily implemented in a decentralized manner and achieve high performance without the need for excessive information exchange. The approach consists in allowing the users to locate and exploit the spectrum bands efficiently based on historic information about the channels' occupancy and quality using an adequate learning algorithm such as the Q-learning [13]. Q-learning is a distributed strategy that helps agents choosing their best actions, the channel to choose in our case, over time by storing and updating recursively an objective function corresponding to each possible action. This will allow to estimate future rewards of the actions to take based on the previous obtained rewards. The algorithm converges rapidly towards the best selection of channels by each user without needing lot of computation neither excessive information exchange.

The challenge consists of designing efficient objective functions that maximize the users' reward depending on the traffic model. In [11, 14], the authors proposed an efficient "difference" objective function for elastic traffic model that allows users to find and exploit spectrum bands efficiently. They have shown that this objective function maximizes spectrum users' received rewards and achieves near optimal performance with a high scalability. Also, they have shown its learnability as it converges very quickly towards the highest performance. In [12, 15–17], an inelastic traffic model is considered. The authors developed a "team-contribution" objective function. This function is based on cooperation between users to find and exploit the best spectrum band efficiently. Thus, the users need to work as a team to succeed together in obtaining a good level of the Quality of Service. The authors showed that the team-contribution objective function achieves close optimal performance.

Based on these findings, we design and implement a resource allocation protocol for distributed DSA systems [18]. The protocol supports different traffic models. The proposed protocol divides each time episode into three phases: 1) *Select phase,* in which each user will choose the best band based on Q-learning; 2) *Data Communication phase,* in which the transmission of data will be done; and finally 3) *Update*

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

This work is an extended version of the work accepted for presentation at the IEEE International Conference on Communications (ICC), Sydney, Australia, June 2014.

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

B. Hamdaoui and B. Khalfi are with Oregon State University, Corvallis, Oregon, USA.

phase, in which each user will update its Q-learning table. Moreover, we test the protocol with other reward models like the *Hybrid model* which combines elastic and inelastic behaviors. We evaluate the objective functions that will maximize the global achieved reward. In addition, we evaluate the system's performance assuming heterogeneous traffic model, where some users use the elastic reward and others use the inelastic reward, by comparing the performance obtained using different objective functions. The evaluation of the performance of the proposed protocol is done using ns3 [19]. We study the impact of different practical aspects resulting from the characteristics of DSA environments on the system's obtained performance. The implementation of the proposed protocol is very challenging since it takes into consideration various practical aspects such as traffic overhead due to control message exchanges, data collision, and the unequal share of the spectrum band due to the random access which will cause estimation errors of the received reward.

The main contributions of this paper with comparison to previous works are: (i) we show the validity of theoretic findings on efficiency of the proposed objective functions through real simulations using ns3; (ii) in achieving that, we take into consideration practical implementation limitations and propose convenient methods to avoid them; and then (iii) we extend the work to more generic reward models and propose suitable objective functions for them.

The paper is organized as follows. In Section II, we present the system model. In section III, the various traffic models and the various objective functions are described. Then, we present in Section IV the proposed distributed protocol and the implementation challenge due to the various practical aspects taken into consideration. In Section V, we evaluate the performance of the proposed protocol. Finally, we conclude the main results in Section VI.

II. System Model

We assume a cognitive radio network with n agents using m Data Channels (DCs). The total available spectrum is equally divided. At any time episode, each agent can use only one DC. The network is static, so the agents enter and leave the system at the same time. Also, we consider that the network is fully connected, so all agents interfere with one another. Each DC offers an amount of service V_j, which is the throughput in $Mbps$ in our case.

We consider that the agents are trying distributively to find the best DC using a Q-learning algorithm. The performance of the learning algorithm depends specifically on the objective function's choice to allow each agent to exploit the best DC by maximizing the long-term received reward. We assume that the agents selecting the same DC will share it in accordance to a carrier sense multiple access (CSMA) [20].

III. Learning Techniques in DSA Systems

Each agent implements a Q-learning algorithm to guide it in the DC selection. The learning algorithm is characterized by two key parameter functions: the reward function and the objective function. The reward function represents the level of satisfaction of the user depending on the adopted traffic model. It models the received service resulting from the access to a specific DC and the environment conditions. On the other hand, the objective function represents the utility that each agent uses to update its Q-learning table which will be exploited later to take a decision on which DC to select. In the following, we detail the description of the various reward and objective functions that we will be using throughout our work.

A. Reward Function

1) Elastic Reward: The elastic traffic model, as defined in [11], is used usually to model traffic hungry applications where the utility increases as the received service increases such as data transfer. In this traffic model, the received reward for each user is proportional to the amount of service offered by the selected band given that the service is above a minimum threshold service but when it is below that threshold, the reward drops exponentially as the level of service becomes insufficient. Using the CSMA scheme for the interfering users within the same bands, the total service of the band is split equally among them. Thus, the received service by each user is equal to the total band service divided by the number of users who selected that band. Thus, the elastic reward can be interpreted as if there is a maximum capacity of the band in terms of the number of users who could be served by the band and when the number of users selecting the band exceeds this threshold the received reward for all users decreases exponentially (see Fig. 1). Explicitly, the reward as a function of a user i at instant t, r_{ela} can be written as a function of the number of users selecting the band j, $n_j(t)$, and the band total service, V_j, as

$$r_{ela_i}(t) = \begin{cases} \frac{V_j}{n_j(t)} & \text{if } n_j \leq V_j/R_{th} \\ R_{th} \exp\left(-\beta \frac{n_j(t)R_{th} - V_j}{V_j}\right) & \text{otherwise,} \end{cases} \tag{1}$$

where β is an exponential decaying factor chosen to control the decay speed of the reward and R_{th} is the minimum service threshold.

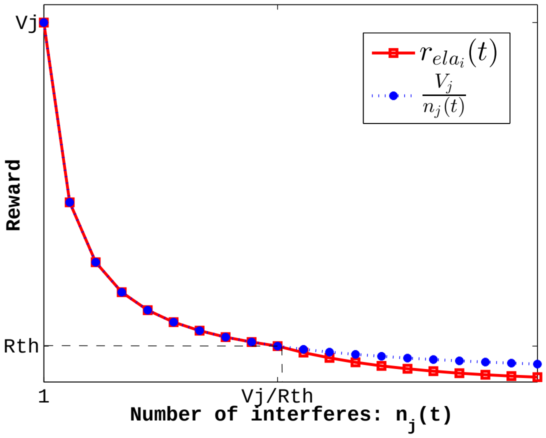

Fig. 1: Reward function for the elastic traffic model as a function of the number of interferers in the channel.

2) Inelastic Reward: The inelastic traffic model [12] is suitable for network applications that require a fixed amount of service such as voice calls. The received reward for each user is constant given that the received service exceeds a certain threshold but it drops exponentially if that threshold of service is not satisfied due to a high number of interferers exceeding the capacity of the band (see Fig. 2). Explicitly, the reward of a user i can be written in this case as follows

$$r_{inela_i}(t) = \begin{cases} R_{th} & \text{if } n_j \leq V_j/R_{th} \\ R_{th}\exp\left(-\beta\frac{n_j(t)R_{th}-V_j}{V_j}\right) & \text{otherwise,} \end{cases} \tag{2}$$

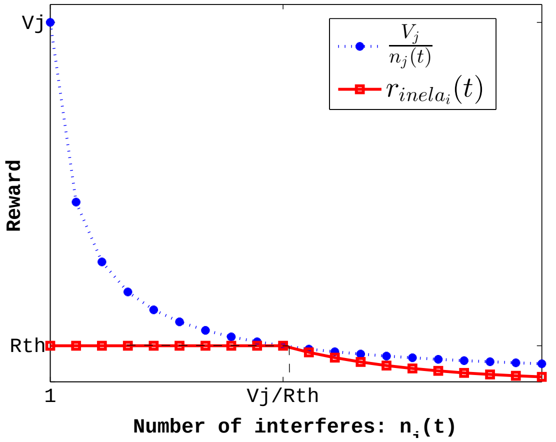

Fig. 2: Reward function for the inelastic traffic model as a function of the number of interferers in the channel.

3) Hybrid Reward: Recent applications, such as adaptive video streaming, can be modeled by a more generic reward where users' satisfaction increases proportionally with the received service (improved quality) starting from a minimum acceptable threshold until reaching the best rate where satisfaction is at its maximum.

An example of this reward function is presented in Fig. 3. We consider three threshold values for the level of service R_1, R_2, and R_3 such that $R_1 \geq R_2 \geq R_3$.

- When the level of service is less than R_3, the received reward drops exponentially to express a total non-satisfaction of the user.
- When the level of service is between R_3 and R_2, the user gets a fixed reward equals to R_2.
- When the level of service is between R_2 and R_1, the reward function is proportional to the level of service.
- When the level of service is higher than R_1, the user gets a fixed reward R_1.

Analytically, the reward function can be written as a function of the thresholds of the level of service as

$$r_{hyb_i}(t) = \begin{cases} R_1 & \text{if } n_j \leq V_j/R_1 \\ \frac{V_j}{n_j(t)} & \text{if } V_j/R_1 < n_j \leq V_j/R_2 \\ R_2 & \text{if } V_j/R_2 < n_j \leq V_j/R_3 \\ R_2\exp\left(-\beta\frac{n_j(t)R_3-V_j}{V_j}\right) & \text{otherwise,} \end{cases} \tag{3}$$

Intuitively, this reward could be used to describe systems with two characteristic levels of service R_2 and R_1, and a minimum acceptable level of service R_3. The behavior of this reward is elastic between R_2 and R_1 while it is inelastic above R_1 and between R_3 and R_2. This reward can be seen as a generalization of the elastic and inelastic models in a generic reward. The pure elastic behavior can be re-obtained by setting $R_2 = R_3$ and R_1 to V while the inelastic one is obtained by setting $R_1 = R_2$.

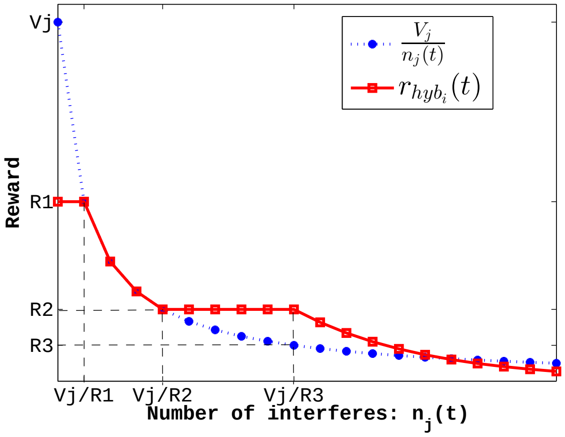

Fig. 3: Reward function for the hybrid traffic model as a function of the number of interferers in the channel.

B. Objective Functions

Since the distributed resource allocation for DSA will be based on a learning algorithm, a well-chosen objective function needs to be designed. Based on previous works [11, 12, 14–16], we present below the different objectives that we will be using in our implementation.

1) Intrinsic Objective Function: The agents using the intrinsic function $r_i(t)$ aim to maximize their own received reward. It had been shown in [11] that this objective function leads to an oscillating behavior and slow increase of the obtained performance due to non-coordination between users as each of them is trying to maximize its own reward without caring about the others' actions. This leads to mutual interference that creates these oscillations and prevent fast learning to improve performance of all users.

2) Global Objective Function: In the global objective function $G(t) = \sum_{i=1}^{n} r_i(t)$, the agents aim to maximize the total system performance with a goal to take care of other users' obtained reward. It had been shown in [11] that while the oscillating behavior is removed by considering global users' reward, the learning time using this objective remains very slow due to conflicts of interests between users.

3) Difference Objective Function: The difference objection function D_i for an agent i is defined to remove the effect of the other agents with the agent i from the global objective function. Hence, this function represents the effect of only the user itself on the global reward by taking out the effect of the other users. By doing so, the resulting objective function will have higher learnability and will achieve near optimal system's performance.

$$D_i(t) = G(t) - G_{-i}(t), \tag{4}$$

where $G_{-i}(t)$ represents the global objective function when the agent i is assumed absent. With the assumption of equal share of the received service between the interferers, [11] shows that D_i can be further simplified and written as follows

$$D_i(t) = n_j(t)r_i\big(n_j(t)\big) - \big(n_j(t)-1\big)r_i\big(n_j(t)-1\big) \tag{5}$$

Hence, the difference objective function can be implemented in a decentralized manner since each agent needs to know only the number of users with which the band is shared to compute the value of D_i at time episode t.

4) Team Objective Function: The team objective function $T_i(t)$ is defined as a contribution reward of the agents selecting the same band to force them to cooperate to enhance each other's reward. Thus, when they reach an acceptable service by selecting their DCs, they celebrate as a team and the affected reward is the sum of the team rewards, but when they do not receive enough service, each of them fails alone and receives its own reward only. This objective function will allow the SUs to find quickly the best DC with cooperation and achieves near optimal system's performance. The expression of the team contribution function is defined as

$$T_i(t) = \begin{cases} \sum_{k=1}^{n_j(t)} D_k(t) & \text{if } n_j \leq V_j/R_{th} \\ D_i(t) & \text{otherwise.} \end{cases} \tag{6}$$

In the assumption of an equal share of the received service between the interferers, T_i can be simplified [12] as follows

$$T_i(t) = \begin{cases} n_j(t)D_i(t) & \text{if } n_j \leq V_j/R_{th} \\ D_i(t) & \text{otherwise,} \end{cases} \tag{7}$$

which shows that in this case also the objective function can be implemented in a decentralized manner.

C. Learning Efficiency Criteria

On the objective of successful communication, the proposed protocol should satisfy the following four criteria:

- **Distributivity:** A centralized protocol requires a central unit to collect information from all agents and then compute optimal DCs allocation which results in a long delay and important communication overhead. Distributed protocols are preferred where each agent will compute its own objective based on the received service and use it for the DC choice for the next episode.
- **Optimality:** Achieving the highest performance in terms of Quality of Service is the target of any communication protocol. For our protocol, optimality is quantified as a function of the average received service of the users in the system.
- **Scalability:** Scalability is an important condition in recent wireless systems due to the continuous and dramatic increase of wireless systems' size. Thus, designed protocols should perform well in small size systems as well as in large scale systems.
- **Learnability:** Since the proposed protocol is based on learning, the most important criteria is the learnability which measures the rapidity of the system to converge to the optimal performance.

IV. Protocol Design and Implementation

We design a protocol that manages resource allocation for distributed DSA systems. The protocol is based on the learning approach described in Section II. We target a protocol that supports the different traffic models and can handle different objective functions to compare their behaviors using ns3 simulations.

A. Protocol Description

The proposed protocol divides time into episodes. Each time episode consists of three window durations: select window (SelWin), data communication window (DCWin), and update window (UpWin). Events occurring during each of these phases are briefly described as follows:

- **Select Phase:**
  Each SU stores a table Q (initialized to zero) containing m elements which correspond to the available DCs. The SUs use the epsilon-greedy strategy to pick their action at each time step. For instance, it selects a random DC with probability ϵ and chooses the best DC which corresponds to the index of the highest value in the table Q with probability $1-\epsilon$. The SU uses the selected DC until the end of the episode.
- **Data Communication Phase:**
  All SUs turned to the same DC use CSMA as the access method to share the DC for data communication. Each SU has a random access time on the shared data channel DC. It verifies the absence of other traffic (from other users) before transmitting its packets using a feedback from the receiver. If a carrier is sensed, the SU waits for the transmission in progress to finish before initiating its own transmission.
- **Update Phase:**
  At the end of the DCwin, each SU i updates its Q-table using the chosen objective function $g_i(t)$ as

  $$Q(j) = (1-\alpha)Q(j) + \alpha g_i(t), \tag{8}$$

  where j is the selected DC for user i in the time episode t and α is a weighting factor chosen to control importance of the effect of past information and present information in the Q-value. The challenge consists of evaluating the objective function which, as shown above, depends only on the number of interfering users. Using the hypothesis of "equal share", the number of interfering users can be estimated by dividing the total amount of service offered by the band over the amount of service received by the user (the measured received throughput) denoted by $\hat{r_i(t)}$.

The estimated number of interfering users is written as

$$\hat{n_j(t)} = \frac{V_j}{\hat{r_i(t)}}. \quad (9)$$

B. Evaluation of the Objective Function Challenge

In this implementation, the most challenging task is to evaluate the objective function. If we update the Q-learning algorithm with the intrinsic function r_i, each SU needs to measure its received throughput, and based on that it will evaluate the achieved reward. However, if we use the difference or team objective functions (D_i or T_i), each SU needs to estimate the number of interferers in the selected DC.

In theory, we assume an equal share among users of the total throughput offered by the DC. So, each SU computes the number of interferers on the selected DC as the total throughput offered by the band divided by its own received throughput. However, in practice a CSMA scheme results in a random access to the DC. So, SUs sharing the same DC may not receive equal instantaneous throughput, but only in average they will receive equal service. Hence, the difference objective function D_i can not be estimated correctly. Even if users exchange information about their received reward, the difference objective function D_i can not be estimated correctly due to the second term of the expression which is a virtual expression that computes the reward in the case of the absence of the actual user.

C. Throughput Estimation

In this section, we estimate the user's received throughput as a function of the number of interferers in the selected DC. We set a CSMA network where n users exchange data with a server via one DC. We assume a user datagram protocol (UDP) echo server protocol where clients receive from the server what they send to it. We simulate this data communication for many time episodes, we vary the number of users n and measure the received throughput as a function of n. We just fix the total throughout offered by the DC (Uplink and Downlink) to $V = 20\ Mbps$. Then, at the end of each episode, each user will estimate its received throughput by accessing the DC.

In Fig. 4, we plot first the instantaneous received throughput for a specific user for different simulations as a function of the number of interferers. We can see that each user receives different throughput on each time episode due to the random access to the DC. Thus, we verify that the received service is not exactly V/n. Then, we plot the average received throughput over time for each user accessing the DC. We deduce that users sharing the same DC will not receive the same throughput even in average. So, the assumption of the equal share of the band is not respected when using the CSMA scheme in practice. We also plot the average throughput over users as a function of the number of interferers n. It shows that users can only exploit about 80% of the total capacity of the DC due to the back-off algorithm and the overhead (i.e., control packets: Address Resolution Protocol (ARP) and Internet Control Message Protocol (ICMP) packets).

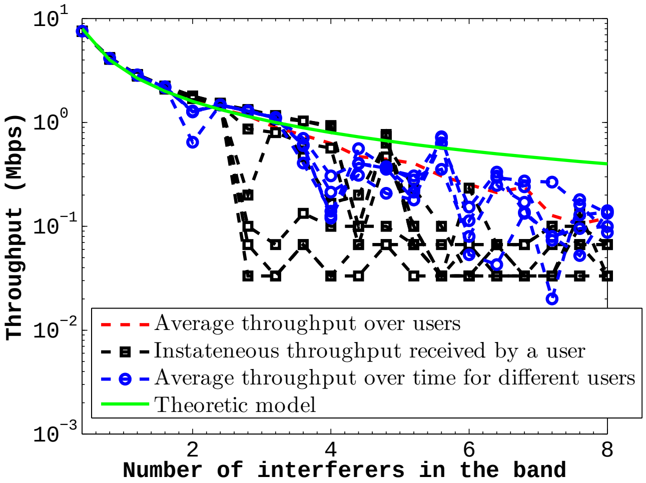

Fig. 4: CSMA throughput measurement.

V. Performance Evaluation

We evaluate the proposed protocol for different reward models. We consider the global system received reward $G(t) = \sum_{i=1}^{n} r_i(t)$ at each time episode t as the performance metric to evaluate the performance of the protocol. We measure the received throughput for each user at each time episode, then we compute the reward based on the considered traffic model.

A. Simulation Parameters

We consider an UDP echo client server application where $n = 100$ clients are sending packets to the server via $m = 5$ shared DCs. The server returns the received packets to the correspondent senders as an acknowledgment of reception. Each client generates random session, each of size Z bytes selected from a uniform distribution with mean $\overline{Z}$ and coefficient of variation δ_Z. Let $L = 1250\ Bytes$ be the length of each packet. Each DC has a capacity of $V = 20\ Mbps$. For the traffic reward models, we consider the threshold of acceptable throughput $R_{th} = 1.5$ Mbps and the exponential decay factor $\beta = 2$. Finally, for the learning algorithm, we use a learning rate $\alpha = 0.5$ and a randomness probability $\epsilon = 0.05$.

The parameters used in the simulations are shown in Table I.

TABLE I: Simulation Parameters.

Symbol	Description	Value
n	number of SUs	100
m	number of DCs	5
V_j	capacity of each DC	20 Mbps
R_{th}	service threshold	1.5 Mbps
β	reward decay factor	2
α	learning rate	0.5
ϵ	randomness probability	0.05

B. Simulation Results

We evaluate the proposed protocol under the different traffic models described in Section III. We compare the protocol's performance obtained using different objective functions using both ns3 simulations and theoretical results.

First, we run simulations considering the elastic traffic model. In Fig. 5, we plot the global achieved reward when using the D_i and r_i objective functions. We confirm that the difference objective function outperforms the intrinsic function for both simulations and theoretic results. The simulation achieved rewards are approximately equal to the theoretic results with the r_i objective function. However, when using the D_i objective in the Q-learning, the achieved reward becomes higher using simulations. This can be explained by the non-equal share condition caused by the random access to the channel via CSMA, some SUs will receive higher throughput than their interferers in the same DC, so they will receive higher rewards which will create distinction between users and push users in the same band to take different decisions in next time slots as they did not receive the same reward even if they accessed the same band which improves the learnability and allows to achieve higher performance.

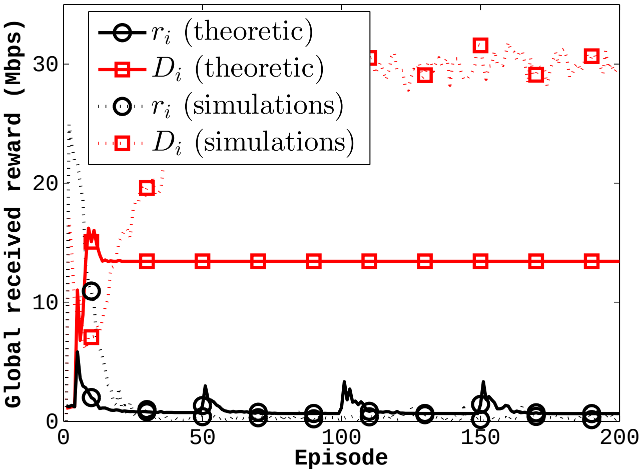

Fig. 5: Global achieved reward, assuming the elastic traffic model, as a function of the time episodes t under the two objective functions r_i and D_i: analytic and simulation results.

In Fig. 6, we consider also the elastic traffic model but we assume that DCs offer different service values to test the protocol in a more practical scenario of non homogeneous channels. We define $\psi = \frac{V_{avg}-V_{min}}{V_{avg}}$ as the variability of the DCs services. We set the capacity of the DCs respectively to 20 $Mbps$, 25 $Mbps$, 20 $Mbps$, 25 $Mbps$ and 16 $Mbps$, so the variability of the DCs services is equal to 32.5%. We show through this figure that the conclusions about the objective function choice are still maintained. The difference objective function D_i still achieves the best performance.

We stated that the evaluation of the objective function D_i in a distributed manner is challenging due to the estimation error of the other users' rewards. Thus, we compare the system's obtained performance using our fully distributed protocol to a semi-distributed protocol. In our fully distributed protocol, each user should estimate the number of interferers using its received service. In the semi-distributed protocol, we assume that there is a control channel where SUs exchange information about their selected bands. Thus, each SU can compute the D_i function with the exact number of interferers n_j. The simulation results given in Fig. 7 show that the obtained global achieved rewards are very close in the two scenarios which confirms that the estimation error's impact is minimal.

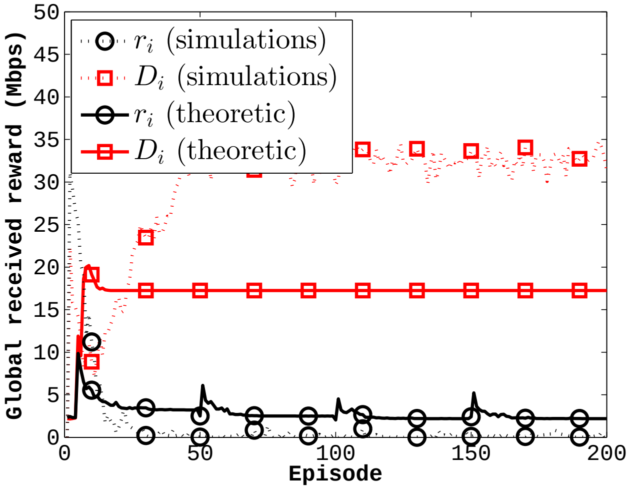

Fig. 6: Global achieved reward, assuming elastic traffic, as function of time episodes t under the D_i function: analytic and simulation results with variable channels service and $\psi = 32.5\%$.

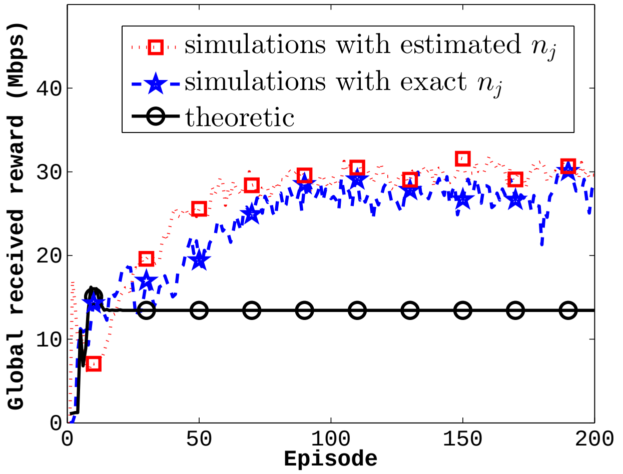

Fig. 7: Global achieved reward in an elastic traffic model as a function of time episodes t under the D_i objective function: analytic, simulation with estimated n_j, and simulation with exact n_j.

In Fig. 8, we study the impact of the service threshold R_{th} on the obtained performance. We plot the global achieved reward using the D_i and r_i objective functions with different values of R_{th}. Theoretically, decreasing R_{th} results in an increase of the bands' capacity to accommodate more users within the same band and thus obtaining higher rewards as shown with the analytic curves. But, with simulations, although rewards are still increasing when the satisfaction level

R_{th} decreases, the performance became closer for different values of R_{th}. This can be explained by the effect of the random share of the band instead of the equal share which allows some users to obtain a level of service higher than R_{th} although theoretically it should be less for all users.

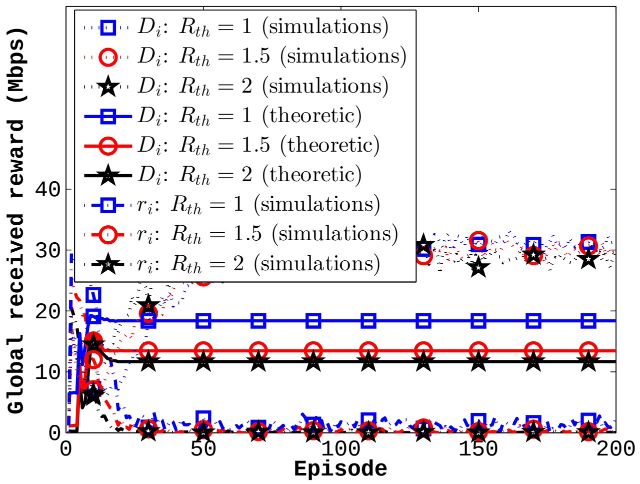

Fig. 8: Global achieved reward with elastic traffic model using the D_i and r_i objective functions for different values of R_{th}: analytic and simulation results.

In Fig. 9, we consider the inelastic traffic model. We plot the global achieved reward under the objective functions r_i, D_i, and T_i. The simulation results confirm the analytical results. Specifically, the team function T_i outperforms the difference objective function D_i with the inelastic traffic model. Similar to the results obtained with the elastic traffic model, we also observe that the simulation results obtained under the T_i function outperform the analytical results in terms of reward due to the non-equal share of the capacity offered by the CSMA channel which improves learnability in this case too.

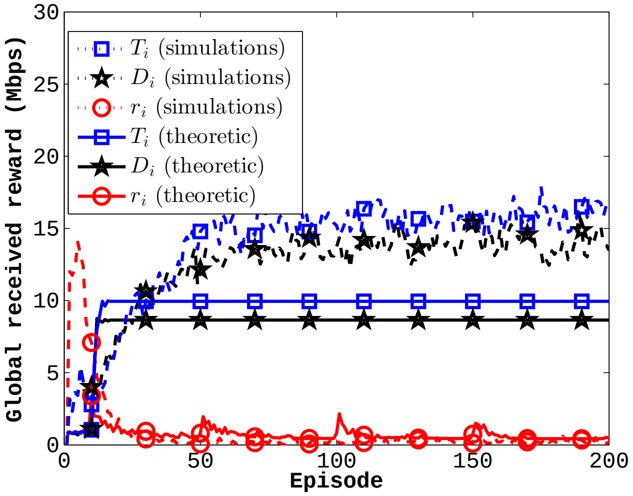

Fig. 9: Global achieved reward, assuming the inelastic traffic model, as a function of time episodes t under the objective functions r_i, D_i, and T_i: analytic and simulation results.

In Fig. 10, we consider the hybrid reward model. We show through this Figure that the protocol still works well when assuming a generic reward function. We obtain high global received reward with the two objective functions D_i and T_i for both simulation and analytical results but the intrinsic objective function r_i still leads to poor performance. To improve the system performance, we propose to update the Q-learning algorithm with a mixed objective function between D_i and T_i. For each SU, if the received throughput is between R_1 and R_2 or less than R_3, the Q-table is updated with D_i because the reward belongs to the elastic domain, else the T_i is used for the update since the reward belongs to the inelastic domain. The new defined mixed objective function M_i is expressed as follows

$$M_i(t) = \begin{cases} T_i(t) & \text{if } n_j(t) \leq V_j/R_1 \text{ or } V_j/R_2 < n_j(t) \leq V_j/R_3 \\ D_i(t) & \text{otherwise,} \end{cases} \tag{10}$$

When assuming the hybrid reward model, the mixed objective function inherits the advantages of the D_i objective function when the measured throughput is in the elastic part while it inherits the advantages of the T_i objective function when the measured throughput is in the inelastic part. We confirm in Fig. 10 that this proposed mixed objective function outperforms both the team contribution function T_i and the difference objective function D_i as it inherits the advantages of both of them by design.

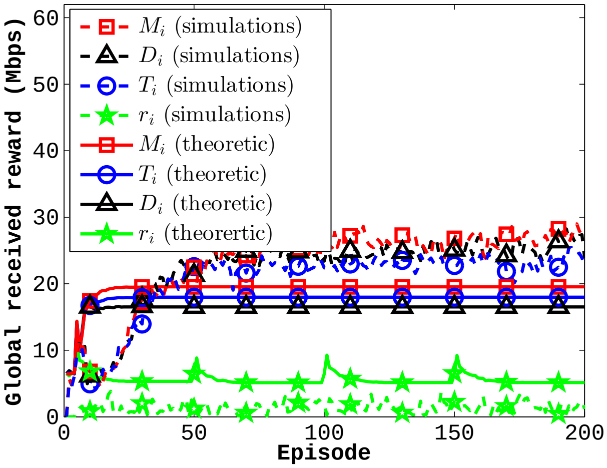

Fig. 10: Global achieved reward, assuming the hybrid reward model, as a function of time episodes t: analytic and simulation results.

We, now, consider a heterogeneous traffic model that we believe it is more practical in real life. In such a scenario, some SUs use the elastic traffic model and others use the inelastic traffic model. We will study the impact of traffic heterogeneity on the protocol's achievable performance.

We believe that it is impossible to evaluate the difference objective function D_i under the heterogeneous traffic model in a distributed manner since each SU needs to know the exact number of interferers using the elastic traffic model called $ne_j(t)$ and those using the inelastic traffic model called $ni_j(t)$.

Thus, for this scenario, we assume that there is a control channel where this information is shared among users.

Thus, for the SUs using the elastic traffic model, the difference objective function can be expressed as

$$D_{ela_i}(t) = ne_j(t)\ r_{ela_i}\big(n_j(t)\big) + ni_j(t)\ r_{inela_i}\big(n_j(t)\big) \\ -\big(ne_j(t)-1\big)\ r_{ela_i}\big(n_j(t)-1\big) - ni_j(t)\ r_{inela_i}\big(n_j(t)-1\big), \tag{11}$$

while for the SUs using the inelastic traffic model, the difference objective function can be expressed as

$$D_{inela_i}(t) = ne_j(t)\ r_{ela_i}\big(n_j(t)\big) + ni_j(t)\ r_{inela_i}\big(n_j(t)\big) \\ -ne_j(t)\ r_{ela_i}\big(n_j(t)-1\big) - \big(ni_j(t)-1\big)\ r_{inela_i}\big(n_j(t)-1\big) \tag{12}$$

where $r_{ela_i}(.)$ and $r_{inela_i}(.)$ are the elastic and inelastic traffic rewards defined in (1) and (2), respectively.

Fig. 11 shows the global achieved reward under the difference objective function D_i assuming three heterogeneous traffic models. The analytical and simulation results show that when the number of SUs under the elastic traffic model increases, the total system performance increases. That is explained by the reward capping in the inelastic traffic model even when the received service continue increasing.

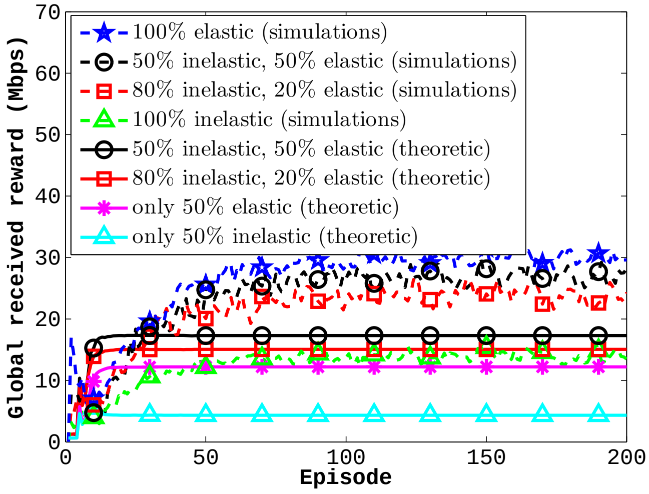

Fig. 11: Global achieved reward, assuming heterogeneous traffic model, as a function of time episodes t under the D_i objective function: analytic and simulation results.

In Fig. 12, we assume that 20% of the SUs are using an elastic reward and 80% of the SUs are using inelastic reward. We plot the analytical and simulation results obtained using different objective functions. First, we observe that the team contribution function T_i outperforms the difference objective function D_i because the percentage of SUs under the inelastic traffic model is higher than the percentage of SUs under the elastic traffic model. Then, we plot in red the global achieved reward assuming that the SUs under the elastic reward update their Q-table by the D_i function and those under the inelastic reward update their Q-table with the T_i function. Simulation and analytical results show that this objective outperforms both objective functions D_i and T_i since it is more adapted to this heterogeneous traffic.

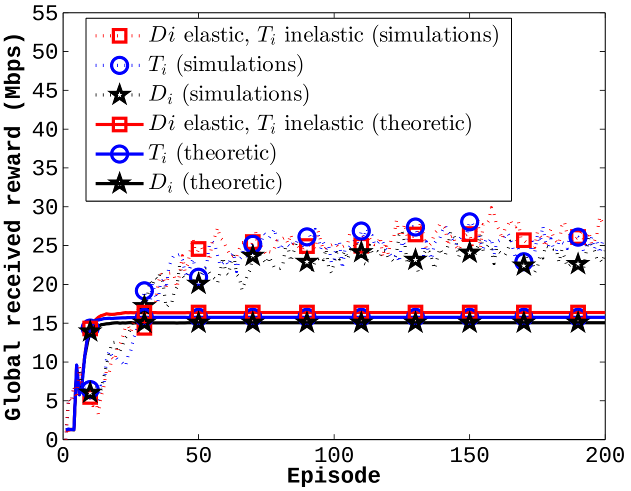

Fig. 12: Global achieved reward, assuming heterogeneous traffic model, as a function of time episodes t: 20% using elastic reward and 80% using inelastic reward.

VI. Conclusion

In this paper, we design a resource allocation protocol for distributed DSA systems. We evaluate the performance of the proposed protocol under various traffic models where each traffic model was represented by a reward function of the obtained service. The protocol design is evaluated using ns3 taking into consideration practical aspects. It was concluded that for each traffic model, we can design an efficient objective function that optimizes the performance under that reward model although the users' random access in shared bands reduces the targeted performance theoretically.

References

[1] F. C. Commission, "Spectrum policy task force," *Report ET Docket*, Nov 2002.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

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

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

[5] Q. Zhao, L. Tong, and A. Swami, "Decentralized cognitive MAC for dynamic spectrum access," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2005)*, Nov 2005, pp. 224–232.

[6] L.-F. Huang, S.-L. Zhou, D. Guo, and H.-C. Chao, "MHC-MAC: cognitive MAC with asynchronous-assembly line mode for improving spectrum utilization and network capacity," *Mathematical and Computer Modelling*, vol. 57, no. 11, pp. 2742–2749, 2013.

[7] C. Zou and C. Chigan, "On game theoretic DSA-driven MAC for cognitive radio networks," *Computer Communications*, vol. 32, no. 18, pp. 1944–1954, 2009.

[8] T. Jin, C. Chigan, and Z. Tian, "WLC05-4: Game-theoretic Distributed Spectrum Sharing for Wireless Cognitive Networks with Heterogeneous QoS," in *IEEE Global Telecommunications Conference (GLOBECOM '06)*, Nov 2006, pp. 1–6.

[9] Q. Ni, R. Zhu, Z. Wu, Y. Sun, L. Zhou, and B. Zhou, "Spectrum Allocation Based on Game Theory in Cognitive Radio Networks," *Journal of Networks*, vol. 8, no. 3, pp. 712–722, 2013.

[10] A. Ozyagci and J. Zander, "Distributed Dynamic Spectrum Access in Multichannel Random Access Networks with Selfish Users," in *IEEE Wireless Communications and Networking Conference (WCNC'10)*, April 2010, pp. 1–6.

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

[12] M. NoroozOliaee, B. Hamdaoui, and M. Guizani, "Maximizing Secondary-User Satisfaction in Large-Scale DSA Systems Through Distributed Team Cooperation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3588–3597, October 2012.

[13] C. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[14] M. NoroozOliaee, B. Hamdaoui, and K. Tumer, "Achieving optimal elastic traffic rewards in dynamic multichannel access," in *International Conference on High Performance Computing and Simulation (HPCS'11)*, July 2011, pp. 155–161.

[15] B. Hamdaoui, M. NoroozOliaee, K. Tumer, and A. Rayes, "Coordinating Secondary-User Behaviors for Inelastic Traffic Reward Maximization in Large-Scale OSA Networks," *IEEE Transactions on Network and Service Management*, vol. 9, no. 4, pp. 501–513, December 2012.

[16] ——, "Aligning Spectrum-User Objectives for Maximum Inelastic-Traffic Reward," in *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN'11)*, July 2011, pp. 1–6.

[17] M. NoroozOliaee and B. Hamdaoui, "Distributed resource and service management for large-scale dynamic spectrum access systems through coordinated learning," in *7th International Wireless Communications and Mobile Computing Conference (IWCMC'11)*, July 2011, pp. 522–527.

[18] R. Hamdi, M. Ben Ghorbel, B. Hamdaoui, and M. Guizani, "Design and implementation of distributed dynamic spectrum allocation protocol," in *IEEE International Conference on Communications Workshops (ICC'14)*, June 2014, pp. 274–278.

[19] ns developers, "ns-3 reference manual," *ns-3-dev*, August 2010.

[20] H. Kwon, H. Seo, S. Kim, and B. G. Lee, "Generalized CSMA/CA for OFDMA systems: protocol design, throughput analysis, and implementation issues," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4176–4187, August 2009.

Rami Hamdi is currently a Ph.D student at École Supérieure de Technologie (ÉTS), Montreal, Canada. He received the "Diplome d'Ingénieur" from Ecole Supérieure des Communications de Tunis (SUP'COM) in Ariana, Tunisia in 2012 and Masters of Science from Ecole Nationale d'Ingénieurs de Tunis (ENIT) in Tunis, Tunisia in 2014. He worked as research assistant at Qatar University (QU), Doha, Qatar, from May 2013 to May 2014. His research interests include wireless communication with focus on resource allocation in massive MIMO systems, learning for dynamic spectrum allocation protocols, and channel modeling and performance analysis of vehicle-to-vehicle communication systems.

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

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

Mohsen Guizani (S'85-M'89-SM'99-F'09) is currently a Professor and Associate Vice President of Graduate Studies at Qatar University, Qatar. Previously, he served as the Chair of the Computer Science Department at Western Michigan University from 2002 to 2006 and Chair of the Computer Science Department at the University of West Florida from 1999 to 2002. He also served in academic positions at the University of Missouri-Kansas City, University of Colorado-Boulder, Syracuse University and Kuwait University. He received his B.S. (with distinction) and M.S. degrees in Electrical Engineering; M.S. and Ph.D. degrees in Computer Engineering in 1984, 1986, 1987, and 1990, respectively, all from Syracuse University, Syracuse, New York. His research interests include Wireless Communications and Mobile Computing, Computer Networks, Mobile Cloud Computing and Smart Grid. He currently serves on the editorial boards of many International technical Journals and the Founder and EIC of "Wireless Communications and Mobile Computing" Journal published by John Wiley (http://www.interscience.wiley.com/jpages/1530-8669/). He is the author of nine books and more than 300 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE Journals and Magazines. He also served as member, Chair, and General Chair of a number of conferences. He was selected as the Best Teaching Assistant for two consecutive years at Syracuse University, 1988 and 1989. He was the Chair of the IEEE Communications Society Wireless Technical Committee and Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005. Dr. Guizani is Fellow of IEEE, member of IEEE Communication Society, IEEE Computer Society, ASEE, and Senior Member of ACM.

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