Reciprocal Learning for Cognitive Medium Access

Xianfu Chen1, Zhifeng Zhao3, David Grace3, and Honggang Zhang2,4

1VTT Technical Research Centre of Finland, P.O. Box 1100, FI-90571 Oulu, Finland
2Department of ISEE, Zhejiang University, Zheda Road 38, Hangzhou 310027, China
3Department of Electronics, University of York, Heslington, York YO10 5DD, UK
4Université Européenne de Bretagne (UEB) and Supélec/ETR, CS 47601, 35576 Cesson-Sévigné Cedex, France

Email: xianfu.chen@vtt.fi, zhaozf@zju.edu.cn, dg@ohm.york.ac.uk, honggang.zhang@supelec.fr

Abstract—This paper considers designing efficient medium access strategies for secondary users (SUs) to select frequency channels to sense and access in cognitive radio networks. The interaction among the SUs is considered as a learning problem, in which every SU behaves as an intelligent agent. Each SU believes that its competitors alter their future medium access strategies in proportion to its own current strategy change. These beliefs adapt in accordance with limited information exchange. In this way, each SU can obtain the behavior feature of other users through conjecture, optimize the medium access strategy, and finally achieve the goal of reciprocity, based on which two learning algorithms are proposed. We show that the SUs’ stochastic behaviors and beliefs converge to a steady state under some conditions. Numerical results are provided to evaluate the performance of the two algorithms, and show that the achieved system performance gain outperforms some existing protocols.

I. INTRODUCTION

Recently, there has been significant research on cognitive radio networks (CRNs) which bridges the enormous gulf in time and space between the regulation and the potential spectrum efficiency [1], [2]. One of the critical challenges is how to realize the coexistence of primary users (PUs) and secondary users (SUs) accessing the same part of the spectrum. In such networks, the PUs have priority in accessing the spectrum, while the SUs transmit opportunistically when the PUs are not active. Therefore, to avoid colliding with the licensed network, the SUs must sense before transmission to determine whether there are primary activities over the selected channel. Moreover, due to the resource and hardware constraints, we assume in this work that each SU can only access one channel at a time slot.

The goal of this paper is to design a cognitive medium access protocol, based on which the SUs distributively decide which channels to use in different time slots, in order to fully and efficiently utilize the spectral opportunities. The decision process can be enhanced by taking into the account the statistical information about the licensed traffic. In the presence of multiple SUs in CRNs, the medium access protocol must also account for the possibility of competition among users over the same channel. In general, game theoretic approaches have been exploited to determine the communication resources of multiple interacting users. For instance, non-cooperative spectrum access in CRNs is studied in [3], where the authors demonstrate the existence of Nash equilibrium (NE) and derive the relevant equilibrium flow settings. In [4], the authors translate the problem of throughput maximization of a multi-input multi-output (MIMO) CRN into a noncooperative game, and further derive an optimal pricing policy that drives the game to an NE.

Game theory bases its solution on equilibrium. A user behaving within an equilibrium is often explained in terms of its beliefs about the strategies of its opponents. How users reach such beliefs through interactions is by learning. The distinction between learning and non-learning users is simply that the former change their beliefs, whereas the latter’s beliefs are static. In [5], the authors propose two stochastic learning algorithms for wireless users to dynamically and efficiently allocate discrete power levels, and show that pure and mixed equilibria exist under given certain conditions. A reinforcement learning (RL) algorithm with low implementation complexity is designed for the competitive and autonomous CRNs [6], where it is shown that the learning dynamics asymptotically converge to the NE.

This paper is concerned with developing distributed learning protocols for medium access in CRNs from not only a learning, but game-theoretic perspective. To encourage potential cooperation among the SUs, we adopt the conjectural variation model introduced by Bowley [7] to enable them to build beliefs about how other SUs react to their own strategy changes. The SUs’ belief functions reflect an awareness that there are strategic interaction mechanisms in which they need not to correctly perceive how the future strategies of their competitors depend on the past. Specifically, by implementing such a behavior model in CRNs, the SUs will no longer adopt myopic behaviors. Instead, they will form beliefs about how their channel selection strategy changes will influence the responses of other SUs and, based on these beliefs, they will try to maximize their own performance.

Particularly, the main contributions of this paper are as follows. To cultivate cooperation among the SUs in CRNs, we enable the SU to form independent conjectural beliefs about how the competing SUs’ strategies vary as a function of their own strategy changes. Based on the belief model, we design two simple distributed learning enhanced medium access algorithms, in which all SUs’ conjectural beliefs and strategies will be revised over time by observing the outcomes of past mutual interaction and exchanging limited information.
among the SUs. For both algorithms, we derive the sufficient conditions that guarantee their convergence. The rest of this paper is organized as follows. In Section II, we formulate the medium access problem in a CRN scenario. In Section III, we propose two simple distributed learning algorithms to solve the cognitive medium access problem. The numerical results included in Section IV verify the validity, and efficiency of the proposed algorithms. Finally, we present in Section V a conclusion of this paper.

II. NETWORK MODEL AND PROBLEM FORMULATION

Let \( \mathcal{N} = \{1, \ldots, N\} \) be the index set of SUs and \( \mathcal{M} = \{1, \ldots, M\} \) be the index set of orthogonal channels with equal bandwidth. For simplicity, the bandwidth is normalized to be 1. All users in the network are operated in a time-slotted fashion. In each time slot \( k \), the PU transmits over channel \( m \in \mathcal{M} \) with probability \( 1 - \mu_m \geq 0 \). In other words, let \( S_m(k) \) be a random variable, such that

\[
S_m(k) = \begin{cases} 
0, & \text{channel } m \text{ is occupied in time slot } k; \\
1, & \text{otherwise.}
\end{cases}
\]

\( S_m(k) \) is independent for each \( m \) and \( k \). In what follows, we suppose that \( \mu = [\mu_1, \ldots, \mu_M] \) is initially known to all SUs. For the case that \( \mu \) is unknown, the SUs can estimate \( \mu \) by using their historical channel sensing results [8], [9], which is beyond the scope of this paper.

At the beginning of time slot \( k \), each SU \( n \in \mathcal{N} \) selects one channel \( c_n(k) \in \mathcal{M} \) according to its medium access strategy \( \pi_n \) to sense. A strategy \( \pi_n^k \) is defined to be a probability vector \( \pi_n^k = (\pi_n^k(1), \ldots, \pi_n^k(M)) \), where \( \pi_n^k(m) \) means the probability with which SU \( n \) chooses channel \( m \) at time slot \( k \). We assume that channel sensing is perfect at all SUs. If the sensing result indicates that the channel \( c_n(k) \) is free, i.e., \( S_{c_n(k)}(k) = 1 \), the SUs selecting this channel compete to transmit. Specifically, we adopt in this paper the collision model with no avoidance mechanisms, under which a SU always transmits, and if two or more SUs transmit over the same channel then none of the transmissions are successful. At the end of same time slot, SU \( n \) receives an acknowledgement \( Z_{n,c_n(k)}(k) \) that equals 1 if the transmission goes through and equals 0 otherwise \(^1\).

The utility that each SU \( n \in \mathcal{N} \) obtains by accessing channel \( c_n(k) \) is the number of bits that it can transmit during time slot \( k \), namely,

\[
W_n(k) = S_{c_n(k)}(k)Z_{n,c_n(k)}(k).
\]

It is clear that \( W_n(k) \) is a random variable that depends on the PUs’ traffics and, more importantly for us, the medium access strategies deployed by all SUs. Therefore, the overarching goal in the rest of this paper is to design the strategies \( \{\pi_n^k | n \in \mathcal{N}\} \) that maximize every SU \( n \)'s expected utility, which is expressed as

\[
U_n(\pi_n^k, \pi_{-n}^k) = E \{S_{c_n(k)}(k)Z_{n,c_n(k)}(k)\} = \sum_{m \in \mathcal{M}} \mu_m \pi_n^k(m) \prod_{i \in \mathcal{N} \setminus \{n\}} (1 - \pi_i^k(m)) ,
\]

where \( \pi_n^k = (\pi_n^k(1), \ldots, \pi_n^k(M)) \). From Eq. (1), we can see that SU \( n \)'s expected utility at each time slot \( k \) depends not only on its own medium access strategy, but also the other SUs’ strategies.

III. DISTRIBUTED LEARNING ALGORITHMS WITH DYNAMIC CONJECTURES

In this section, to promote cooperation, we propose a simple and intuitive rule each SU has, which links the other SUs’ medium access strategies to that of its own. SUs exchange the strategy information in previous time slot and behave optimally with respect to their beliefs.

A. The Conjecture Model

Assuming all the SUs are reasonable players, then each SU thinks any change in its current medium access strategy will induce other SUs to make well-defined changes in the next time slot. Specifically, we need to express the expected contention measure \( \hat{b}_n^k(m) = \prod_{i \in \mathcal{N} \setminus \{n\}} (1 - \pi_i^k(m)) \) in Eq. (1) through belief function

\[
\hat{b}_n^k(m) = \hat{b}_n^{k-1}(m) - \delta_{n,m} (\pi_n^k(m) - \pi_n^{k-1}(m)),
\]

with belief factor \( \delta_{n,m} > 0 \), for all \( n \in \mathcal{N} \) and \( m \in \mathcal{M} \). Particularly, we set \( \pi_n^0(m) = \pi_n^1(m) \) when \( k = 1 \). An intuitive explanation for Eq. (2) is that, SU \( n \) believes a change of \( (\pi_n^k(m) - \pi_n^{k-1}(m)) \) in its medium access strategy at time slot \( k \) will induce a change of \( \delta_{n,m} (\pi_n^k(m) - \pi_n^{k-1}(m)) \) in the expected contention measure exactly corresponding to the strategies of other SUs. Although SU \( n \) may be aware that other SUs are subject to many influences on their strategies, when making its own decision, it is only concerned with other SUs’ reactions to itself. In another word, SU \( n \) does not take into account whether or not SU \( I \in \mathcal{N} \setminus \{n\} \) might react to changes in medium access strategy made by SU \( j \in \mathcal{N} \setminus \{n, I\} \). Substituting Eq. (2) into Eq. (1), we have

\[
U_n(\pi_n^k, \hat{b}_n^k) = \sum_{m \in \mathcal{M}} \mu_m \pi_n^k(m) \hat{b}_n^k(m),
\]

where \( \hat{b}_n^k = (\hat{b}_n^k(1), \ldots, \hat{b}_n^k(M)) \). Each SU \( n \) rationally chooses a strategy \( \pi_n^k \) to maximize its expected utility at each time slot \( k \), which it would achieve if other SUs react to its strategy change according to \( \delta_{n,m} (\pi_n^k(m) - \pi_n^{k-1}(m)) \). On one hand, the SUs may unilaterally think that others are not as sophisticated as they are. On the other hand, the SUs may be aware that other SUs are as sophisticated as they are, but “mistakenly” take the attitude that forming the beliefs that others are equally sophisticated will complicate their decision making too much. In either case, the SUs may regard Eq. (2) as a good approximation of what medium access strategies the others choose.
B. A Best Response Learning Algorithm

Along with the previous discussion, we develop a best response medium access strategy for each SU in this subsection.

1) The Best Response Strategies: We treat $b^k_n(m)$ and $\pi^k_n(m)$ as initial parameters and then find an optimal strategy for SU $n$ that consists of a sequence of single slot policy functions $\pi^k_n(m) = f_{n,m}(\pi^{k-1}_n(m), b^{k-1}_n(m))$, which gives best response behavior for SU $n$ at any time slot $k$ given its beliefs $\delta_n,m$.

Theorem 1. The best response medium access strategy for each SU $n \in \mathcal{N}$ is given by

$$\pi^k_n(m) = \begin{cases} \frac{\pi^{k-1}_n(m)}{2} + \frac{1}{2\delta_n,m} \left( b^{k-1}_n(m) + \frac{\lambda^k_n}{\mu_m} \right), & \text{if } \mu_m > 0; \\ 0, & \text{if } \mu_m = 0; \end{cases}$$

where $\lambda^k_n$ is a parameter that satisfies $\sum_{m \in \mathcal{M}} \pi^k_n(m) = 1$. Here, $[x]^k_b$ with $b > a$, denotes the Euclidean projection of $x$ onto the interval $[a,b]$; i.e., $[x]^k_b = a$ if $x < a$, $[x]^k_b = x$ if $a \leq x < b$, and $[x]^k_b = b$ if $x > b$.

Proof: The best response medium access strategy $\pi^k_n$ at time slot $k$ maximizes $U_n$, i.e.,

$$\max_{\pi^k_n} U_n \left( \pi^k_n, b^k_n \right) \quad \text{s.t. C1: } \sum_{m \in \mathcal{M}} \pi^k_n(m) = 1, \text{ and C2: } \pi^k_n(m) \geq 0.$$  

In our formulation, the optimization problem is convex with linear constraints C1-C2. So the Lagrangian function for SU $n$ can be written as,

$$J_n \left( \pi^k_n, b^k_n, \lambda^k_n, \gamma^k_n \right) = U_n \left( \pi^k_n, b^k_n \right) + \lambda^k_n \left( \sum_{m \in \mathcal{M}} \pi^k_n(m) - 1 \right) + \sum_{m \in \mathcal{M}} \gamma^k_{n,m} \pi^k_n(m),$$

where $\gamma^k_n = (\gamma^k_{n,1}, \ldots, \gamma^k_{n,k})$ and $\lambda^k_n$ and $\gamma^k_{n,m}$ are Lagrangian multipliers. The Karush-Kuhn-Tucker (K.K.T.) conditions [10] are

$$\frac{\partial J_n}{\partial \pi^k_n(m)} = \mu_m \left( -2\delta_n,m \pi^k_n(m) + b^{k-1}_n(m) \right) + \lambda^k_n + \gamma^k_{n,m} = 0,$$

$$\pi^k_n(m) \geq 0$$

$$\gamma^k_{n,m} \pi^k_n(m) = 0$$

$$\sum_{m \in \mathcal{M}} \pi^k_n(m) = 1.$$

It’s straightforward to get (4). This concludes the proof.

Remark 1: We can see from Theorem 1 that it’s not rational for each SU $n \in \mathcal{N}$ to follow the same medium access strategy $\pi^k_n$ obtained at the current time slot in the future. This is because $\pi^k_n$ is based on the conjectural beliefs about other competing SUs’ current strategies which, in general, are dynamic. Thus, a SU needs to recalculate another medium access strategy in the same way at the following time slot.

The detailed description of the proposed best response learning protocol for cognitive medium access is summarized in Algorithm 1.

Algorithm 1: A Best Response Learning Algorithm for Medium Access in CRNs

Initialization:

(a) $k = 1$, initialize the medium access strategies $\pi^k_n(m)$ and the parameters $\delta_n,m > 0$ in SU $n$’s belief functions, for $\forall n \in \mathcal{N}$ and $\forall m \in \mathcal{M}$.

End Initialization

Learning:

(b) Set $k \leftarrow k + 1$.

(c) For $\forall n \in \mathcal{N}$ and $\forall m \in \mathcal{M}$, do Eq. (4).

(d) Each SU $n$ decides to access channel $m$ at time slot $k$ with probability $\pi^k_n(m)$

End Learning

2) Global Convergence Analysis: With best response medium access strategies, the SUs should consider whether their beliefs have any negative effects. Our conjectural belief model expressed in Eq. (2) suggests that errors exist in the beliefs, and the SUs learn from the previous observations to try to improve the utilities at next time slot. For this reason, we shall assume that the dynamics of CRNs will appear reasonably consistent to the SUs if the values of beliefs stabilize as the time passes. We will show in Theorem 2 that the dynamic network is stable if each function $f_{n,m}$ ($\forall n \in \mathcal{N}$ and $\forall m \in \mathcal{M}$) is a contraction mapping.

Theorem 2. Suppose that belief factor $\delta_n,m \geq N - 1$, for $\forall n \in \mathcal{N}$ and $\forall m \in \mathcal{M}$, the dynamic CRN has a unique steady state; that is, regardless of any initial values chosen for $(\pi^1_n, \ldots, \pi^k_N)$, the best response medium access strategies $(\pi^k_n[n \in \mathcal{N}])$ converge to $(\pi^\infty_n, \ldots, \pi^\infty_N)$.

Proof: Without loss of generality, we assume that $\mu_m > 0$, for $m \in \mathcal{M}$. At the moment, Eq. (4) can be rewritten as

$$\pi^k_n(m) = \frac{1}{2} \pi^{k-1}_n(m) + \frac{1}{2\delta_n,m} \left[ \prod_{i \in \mathcal{N}\setminus\{n\}} \left( 1 - \pi^{k-1}_i(m) \right) + \frac{\lambda^k_n}{\mu_m} \right].$$

The sum of the absolute values of the partial derivatives of $\pi^k_n(m)$ with respect to $(\pi^{k-1}_1(m), \ldots, \pi^{k-1}_N(m))$ can be easily obtained, that is

$$Q = \frac{1}{2} + \sum_{i \in \mathcal{N}\setminus\{n\}} \left[ \prod_{j \in \mathcal{N}\setminus\{i,n\}} \left( 1 - \pi^{k-1}_j(m) \right) \right] \leq \frac{1}{2} + \sum_{i \in \mathcal{N}\setminus\{n\}} \frac{1}{2\delta_n,m} = \frac{1}{2} + \frac{N - 1}{2\delta_n,m}.$$
over the competing SUs’ strategies. A Gradient Ascent Learning Algorithm

During the stochastic learning procedure, there is limited information exchange among SUs. This results in an extra level of policy learning, the purpose of which is to learn a best response to the other SUs’ medium access strategies. How strategies of multiple agents interacting with one another evolve over time is an important aspect in multi-agent reinforcement learning [13], [14]. The techniques, for tackling reinforcement learning problems, then match well the topic we discuss in this paper. In what follows, we propose a gradient ascent learning algorithm with conjectural gradients, which uses the beliefs provided by Algorithm 2 converge to \( \{\pi_n^*\} \) obtained by Algorithm 2. We think that our belief model and the dynamics of the medium access strategies of multiple agents interacting with one another evolve over time is an important aspect in multi-agent reinforcement learning [13], [14]. The techniques, for tackling reinforcement learning problems, then match well the topic we discuss in this paper. In what follows, we propose a gradient ascent learning algorithm with conjectural gradients, which uses the beliefs of the SUs regarding others’ strategies. This results in an extra level of policy learning, the purpose of which is to learn a best response to the other SUs’ medium access strategies. As de¿ned in Eq. (5), that is, SU \( n \) ’s belief functions, for all \( m \in \mathcal{M} \), the initial strategies are set to be \( \pi_n^k(m) = \pi_n^{k-1}(m) + \eta_n \frac{\partial J_n(\pi_n, \xi_n, \lambda_n, \gamma_n)}{\partial \pi_n(m)} \bigg|_{\pi_n(m) = \pi_n^{k-1}(m)} \). (6)

Herein, the positive scalar \( \eta_n \) is the step size. So effectively, \( \frac{\partial J_n(\pi_n, \xi_n, \lambda_n, \gamma_n)}{\partial \pi_n(m)} \bigg|_{\pi_n(m) = \pi_n^{k-1}(m)} > 0 \) means the probability of accessing a good channel increases by a rate. On the contrary, the probability of choosing a bad channel decreases by a rate. Substituting Eq. (5) to Eq. (6), thus we get

\[
\pi_n^k(m) = \left[ \pi_n^{k-1}(m) + \eta_n \left( \mu_m + \beta \right) \right]_0^1,
\]

where \( \lambda_n^k \) is chosen such that \( \sum_{m \in \mathcal{M}} \pi_n^k(m) = 1 \). The steps concerning gradient ascent learning algorithm are summarized in Algorithm 2.

Algorithm 2: A Gradient Ascent Learning Algorithm for Medium Access in CRNs

Initialization:

(a) \( k = 1 \), initialize step size \( \eta_n \), the medium access strategies \( \pi_n^0(m) \), and the parameters \( \delta_{n,m} \), \( \gamma_n \), \( \lambda_n^k \), \( \mu_m \) in SU \( n \)’s belief functions, for all \( n \in \mathcal{N} \) and all \( m \in \mathcal{M} \).

End Initialization

Learning:

(b) Set \( k \leftarrow k + 1 \).

(c) For all \( n \in \mathcal{N} \) and all \( m \in \mathcal{M} \), do Eq. (7).

(d) Each SU \( n \) decides to access channel \( m \) at time slot \( k \) with probability \( \pi_n^k(m) \).

End Learning

Theorem 3. Suppose that for all \( n \in \mathcal{N} \) and all \( m \in \mathcal{M} \), the belief factor \( \delta_{n,m} \geq 1 \), and the step size \( \eta_n \) is small enough, the medium access strategies \( \{\pi_n^k\} \) obtained by Algorithm 2 converge to \( \{\pi_n^*, \ldots, \pi_n^*\} \).

Proof: If the step size \( \eta_n \) is small enough, Eq. (7) may be expressed by

\[
\pi_n^k(m) = (1 - \eta_n \mu_m) \pi_n^{k-1}(m) + \eta_n \mu_m \prod_{i \in \mathcal{N} \setminus \{n\}} (1 - \pi_i^{k-1}(m)) + \eta_n \lambda_n^k.
\]

The rest of the proof can be obtained similarly as Theorem 2, and is thus omitted.

Remark 3: Compare Theorem 3 with Theorem 2, we can find that, given the same target \( \{\pi_n^*, \ldots, \pi_n^*\} \) or belief factors \( \delta_{n,m} \) (for all \( n \in \mathcal{N} \) and all \( m \in \mathcal{M} \)), Algorithm 2 exhibits similar properties of Algorithm 1 in terms of global convergence, provided that the step size \( \eta_n \) (for all \( n \in \mathcal{N} \)) is sufficiently small. In other words, the stochastic behavior of these two learning algorithms are similar. However, the best response learning algorithm converges faster, while the gradient ascent learning algorithm with small step size evolves smoothly at the cost of sacrificing its convergence speed.

IV. NUMERICAL RESULTS

In this section, we present simulation experiments for evaluating the performance of the two algorithms developed earlier in this paper. First, we consider a relatively simple scenario where there are two SUs and two channels with idle probabilities 0.6 and 0.8. Denote the probability of SU 1 choosing channel 1 by \( \alpha \) and choosing channel 2 by \( 1 - \alpha \). In the same way, SU 2 chooses channel 1 with probability \( \beta \) and chooses channel 2 with probability \( 1 - \beta \). For all SUs, the initial strategies are set to be \( \alpha(1) = 0.5000 \) and \( \beta(1) = 0.5000 \), the belief factors \( \delta_{n,m} \) are uniformly distributed between 2 and 7, and the step size in Algorithm 2 is \( \eta_n = 0.0300 \). Fig. 1 compares the trajectory of the medium access strategy updates in both Algorithm 1 and Algorithm 2, under the assumption that each SU \( n \) can perfectly conjecture the contention measure \( \prod_{i \in \mathcal{N} \setminus \{n\}} (1 - \pi_i^k(m)) \). The curves show that Algorithm 1 converges in around 12 iterations and Algorithm 2 converges in more iterations, and Algorithm 2 achieves the same optimal strategies after about 40 iterations. The initial strategies \( \alpha(1) \) and \( \beta(1) \) do not affect the convergence of the two algorithms to the optimal strategies. To show this, we can set \( \alpha(1), \beta(1) \) to be \( (0.5000, 0.5000) \) and \( (0.7000, 0.8000) \). It is shown in Fig. 2 that, as we expect, the trajectory converges to the same optimal medium access strategies. Different SUs have different learning abilities, that is, they may have different belief factors \( \delta_{n,m} \) in the algorithms. If the SUs have the same beliefs of what the opponents react
Cognitive Medium Access Strategies

Algorithm 1, Algorithm 1, Algorithm 2, Algorithm 2,

Fig. 1. Strategy dynamics of Algorithm 1 and Algorithm 2.

Algorithm 1
Algorithm 2
Algorithm 1
Algorithm 2

(1)=0.5000, (1)=0.5000
(1)=0.7000, (1)=0.8000
(1)=0.5000, (1)=0.5000

Further, for a more general case, we consider that the CRN consists multiple SUs competing for 6 channels with idle probabilities characterized by Bernoulli distributions with evenly spaced parameters ranging from 0.4 to 0.9. The initial medium access strategy \( \pi_{1}^{(m)} \) is set to be 1/6 for all SUs, and the belief factors \( \delta_{m,n} \) are randomized in accordance with the number of SUs (as indicated in Theorem 2 and Theorem 3). The step size in Algorithm 2 is the same as in previous simulation scenario. As following, we numerically compare the overall network performance of the two proposed learning algorithms in terms of accumulated utilities and energy efficiency with two existing protocols, i.e. adaptive random medium access scheme and no learning scheme:

1. Adaptive random medium access scheme: When there is no communication among multiple SUs, they need to adapt to a collision-free configuration to ensure that the collisions are logarthmic [9]. If the SUs randomize in every time slot, there is a finite probability of collisions in every time slot.
2. No learning scheme: Without learning capability, each SU accesses channel \( m \) according to its strategy \( \pi_{n} \), which is given by [8]

\[
\pi_{n}(m) = \frac{\mu_{m}}{\sum_{i\in\mathcal{M}} \mu_{i}}.
\]

As shown in Fig. 4, the proposed algorithms and the adaptive random medium access scheme achieve better system performance than the no learning scheme. In addition, we can find from Fig. 4 that the accumulated utilities of the medium access solutions increase versus the number of SUs, but decrease when the number of SUs exceeds the number of channels. The reason is obvious: when \( N \leq M \), with more SUs, the utilization of channels is better exploited; yet, when \( N > M \), the collisions among the SUs cannot be avoided, thus cause the reduction in overall network performance.

Suppose that the SUs always transmit with same transmission power level, the energy efficiency (EE) of the CRN can be defined by

\[
\xi_{\text{EE}} = \frac{\text{Energy Consumed for Successful Transmissions}}{\text{Overall Energy Consumption}}.
\]

The curves in Fig. 5 show the energy efficiency of different medium access protocols. By utilizing learning algorithms proposed in this paper, severe collisions among the SUs can be alleviated, thus, achieving reduction in the number of retransmissions and the goal of energy saving.

Then, we evaluate the fairness of different medium access protocols using the quantitative fairness index [15],

\[
F = \frac{\varrho(U_{n})}{\varrho(U_{n}) + \sigma(U_{n})},
\]

where \( \varrho \) and \( \sigma \) are, respectively, the mean and the standard deviation of each SU \( n \)'s utility \( U_{n} \) over all the data flows. Fig. 6 evaluates the fairness characteristics of the two proposed algorithms in this paper, and the other two protocols. We can see that they are comparable in their fairness performance and the achieved fairness indexes are nearly the same.
V. Conclusion

We have studied distributed medium access problems in CRNs in this paper. In order to encourage cooperation among the SUs, we proposed two learning algorithms based on the conjectural variation theory to improve their performance. Each SU forms its own beliefs about the influence of strategy changes to other competing SUs, and thus learns the optimal medium access strategies from the interaction outcomes among SUs. Additionally, we prove the convergence of the dynamic networking environment under the two learning solutions. We have also presented simulations to demonstrate that the proposed two algorithms achieve significant performance, compared with existing protocols.

Acknowledgements

This work is performed in the framework of the VTT strategic project AWARENESS. The work is also partially supported by the National Basic Research Program of China (973 Program 2012CB316000).

References