Wireless Pers Commun (2013) 69:771–791
DOI 10.1007/s11277-012-0611-9

Reinforcement-Learning-Based Double Auction Design for Dynamic Spectrum Access in Cognitive Radio Networks

Yinglei Teng · F. Richard Yu · Ke Han · Yifei Wei · Yong Zhang

Published online: 7 April 2012

Abstract In cognitive radio networks, an important issue is to share the detected available spectrum among different secondary users to improve the network performance. Although some work has been done for dynamic spectrum access, the learning capability of cognitive radio networks is largely ignored in the previous work. In this paper, we propose a reinforcement-learning-based double auction algorithm aiming to improve the performance of dynamic spectrum access in cognitive radio networks. The dynamic spectrum access process is modeled as a double auction game. Based on the spectrum access history information, both primary users and secondary users can estimate the impact on their future rewards and then adapt their spectrum access or release strategies effectively to compete for channel opportunities. Simulation results show that the proposed reinforcement-learning-based double auction algorithm can significantly improve secondary users' performance in terms of packet loss, bidding efficiency and transmission rate or opportunity access.

Keywords Reinforcement learning · Cognitive radio networks · Dynamic spectrum access (DSA) · Double auction

Y. Teng (✉) · K. Han · Y. Wei · Y. Zhang
Beijing University of Posts and Telecommunications, Beijing, China
e-mail: lilytengtt@gmail.com

K. Han
e-mail: Hanke@bupt.edu.cn

Y. Wei
e-mail: weiyifei@bupt.edu.cn

Y. Zhang
e-mail: yongzhang@bupt.edu.cn

F. R. Yu
Defense R&D Canada, Ottawa, ON, Canada
e-mail: richard_yu@carleton.ca

1 Introduction

Recently, with a variety of existing and emerging wireless applications, the radio spectrum demand has been increased dramatically. The current wireless networks are characterized by the fixed spectrum assignment policy, where the wireless spectrum is assigned and restricted to the licensed users on a long-term basis. However, some licensed frequency bands lie idle at spatial and temporal dimensions under the current static spectrum policy, leading to under utilization of a significant amount of spectrum. Cognitive radio has emerged as a key enabling technology for dynamic spectrum access (DSA), which provides the capability to share the wireless spectrum with licensed users to improve spectrum efficiency and network performance by adaptively coordinating different users' access according to spectrum dynamics [1].

Since secondary users (SUs) (i.e., unlicensed users) are allowed to access the spectrum allocated to primary users (PUs) (i.e., licensed users) in the cognitive radio network, a fundamental requirement is to avoid the interference to PUs in their vicinity [2]. Therefore, SUs should be able to detect the presence of PUs through spectrum sensing [3,4] occasionally. Another important issue in cognitive radio networks is how to share the detected available spectrum among different SUs to improve the network performance [5,6].

Most existing work in the area of cognitive radio focuses on the technical aspects of spectrum sensing and dynamic spectrum sharing. Recently, market theories [7–13] bring a novel approach of spectrum sharing from economic aspects. In [8], the authors study a decentralized dynamic spectrum access scheme using the theory of multivariate global game, and Bayesian Nash equilibrium of the resulting global game is investigated. A non-cooperative game is formulated in [9] to obtain the spectrum allocation schemes for SUs. The authors also consider the case of bounded rationality, in which SUs gradually and iteratively adjust their strategies based on the observations on their previous strategies. Considering the selfishness of SUs that may deteriorate the efficiency of DSA seriously, the authors of [10] propose a collusion-resistant dynamic spectrum sharing scheme, and a belief-assisted dynamic pricing approach is studied in [11] to improve the robustness and spectrum efficiency. The authors of [12,13] analyze spectrum sharing and competition from the economic perspective, compare several applicable market theories, and propose a market-equilibrium-based approach to settle the trading price.

Although some work has been done in using market theories for dynamic spectrum access in cognitive radio networks, the *learning* capability of cognitive radio networks is largely ignored in the previous work. However, since the cognitive radio paradigm imposes human-like characteristics in wireless networks, the learning mechanism is one of the important characteristics in cognitive radio networks [1]. Particularly, cognitive radio networks can learn from the history of spectrum usage, which can be used for more efficient and effective dynamic spectrum access in cognitive radio networks. The learning mechanism has been studied extensively in other disciplines, such as artificial intelligence [14]. In deed, machine learning has been central to artificial intelligence from the beginning [15]. Recently, *reinforcement learning* (RL) [16] has become a topic of intensive research, which has been used successfully in solving the quality of service provisioning problem in wireless multimedia networks [17] and the interference management problem in OFDM networks [18], among others. However, little work has been done to consider the learning capability of cognitive radios for dynamic spectrum access.

In this paper, we propose a reinforcement-learning-based double auction algorithm (RL-DA) aiming to improve the performance of dynamic spectrum access in cognitive radio networks. Specifically, we model the dynamic spectrum access process as a double auction

game [19], in which potential buyers (SUs) submit their bids and potential sellers (PUs) simultaneously submit their ask prices to an auctioneer, and then the auctioneer chooses some price that clears the market. Moreover, during the repeated spectrum access interactions between SUs and PUs, cognitive users (users involved in the cognitive radio network, including PUs and SUs are called cognitive users in this literature) can partially observe the history information. Based on this history information, they can estimate the impact on their future rewards and then adapt their spectrum access strategies effectively to compete for channel opportunities or release channels. Simulation results show that not only the bidding price coming from Q-Learning mechanism improves the bidding efficiency but also the reserved price for PUs helps to combat collision occurs among SUs, which might distort the supply and demand relationship of spectrum resource. The proposed RL-DA algorithm can significantly improve cognitive users' performance in terms of packet loss, bidding efficiency and transmission rate or access opportunity.

The remainder of the paper is organized as follows. Section 2 presents the the system models and auction market formulation. In Sect. 3, we formulate the Q-Learning (QL) methodology and propose a RL-based double bidding algorithm. Section 4 is devoted to present the simulation results and discussion, comparing RL-DA with other bidding strategies of SUs and different reserved price of PUs. Finally, Sect. 5 concludes the paper.

2 System Models and Auction Market Formulation

In this section, we present the system models considered in this paper. Then, the dynamic spectrum access problem is modeled as a spectrum auction process.

2.1 Cognitive Radio Network

Consider a cognitive radio network with I PUs and J SUs, indicated by the set $\mathbf{P} = \{p_1, p_2, \ldots, p_I\}$ and $\mathbf{S} = \{s_1, s_2, \ldots, s_J\}$ respectively, as illustrated in Fig. 1. Typically, each PU is assigned with licensed bands by a centralized primary base station (PBS). SUs

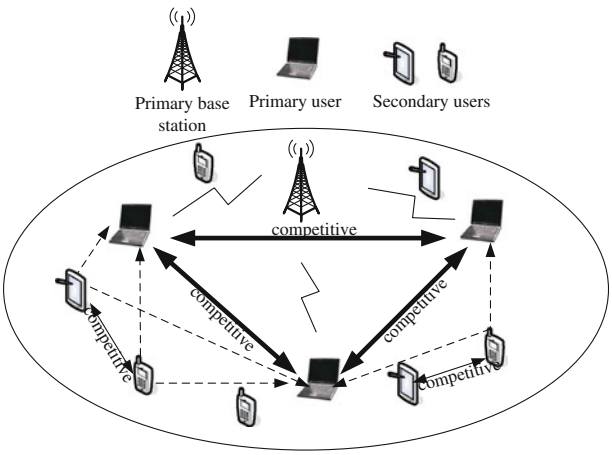

Fig. 1 A cognitive network with multiple primary users and multiple secondary users

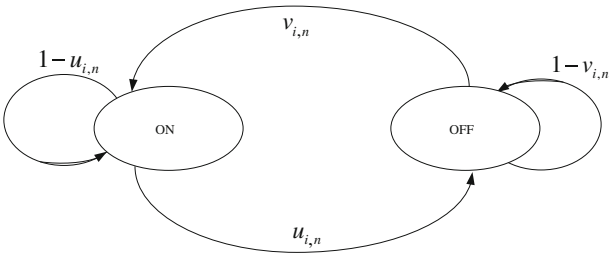

Fig. 2 Spectrum opportunities modeled as a two-state Markov chain

coexisting around seek for the spectrum access opportunities competitively and exclusively. Assume that they register in different network operators, for instance, PUs work as the spectrum owners and market auctioneers connected to a core network and SUs come from various unlicensed networks, e.g. self-organized networks (SONs).

Assume that both PUs and SUs access the channel in an OFDMA fashion. There are totally N_C subchannels in the system. PU $p_i, i \in \{1, 2, \ldots, I\}$ is assigned with N_i OFDM subchannels at time t indicated by a channel index vector $V_i^t = \{\delta_{i,n}^t\}$, where $\delta_{i,n}^t = \{0, 1\}$, $n \in \{1, 2, \ldots, N_C\}$. Herein, $\delta_{i,n}^t = 1$ means that the channel n is assigned to p_i currently. Meanwhile, each SU can simultaneously access multiple channels exclusively. If some anti-out-of-band emission measures are imposed, the OFDM subchannels are assumed to be perfect orthogonal without interference. Furthermore, we assume that wireless users move slowly, and thus, the experienced channel condition changes slowly. The cognitive users deploy constant transmission power and experience no interference during one resource allocation period.

2.2 Spectrum Opportunity

Due to the PU's action of joining or leaving the network occasionally, the available transmission opportunity in each channel changes over time and can be modeled as a two-state Markov chain [20,21], alternating between state ON (active) and state OFF (inactive). An ON-state represents the time period when the licensed channel is occupied by the PU, while an OFF-state is regarded as a potential opportunity for unlicensed radio networks. Suppose that the usage pattern of PU in each channel is independent and identically distributed (i.i.d.). Let $u_{i,n}$ be the transition probability of p_i from ON to OFF on the n channel and $v_{i,n}$ the reverse way. The spectrum opportunity in the cognitive network is illustrated in Fig. 2.

2.3 Spectrum Auction Model

Considering that PUs may not take up the owning spectrum all the time, it is possible for them to release the spectrum usage to SUs by marketing approaches. Specifically, PUs can offer the spectrum price and size that are to share with SUs to maximize their revenues. Similarly, the SUs can adapt their strategies to buy the spectrum opportunities competitively and provide sellers with a certain payoff in terms of performance and price [22]. Therefore, SUs pursue the access opportunities at a payoff while PUs snatch the revenue by leasing the vacant spectrum to the lessees, such that both of them will benefit from the win-win business. In the auction market, each participant (PU or SU) will attempt to maximize his profit in the long run by the bidding and asking process. Like in the real market, both PUs and

SUs have valuations of the goods (spectrum, transmission slot, etc.) to trade in an auction. In the double auction, not only the potential buyers (SUs) bid for the available spectrum, but also the potential sellers (PUs) try to rent out their band items and asking prices for monetary gain.

The following assumptions are placed throughout this paper:

- SUs work in a half-duplex mode at all time, thus either sense vacant channels at a particular time slots, or access one or more channels for data transmission exclusively. Meanwhile, PUs can receive payoff indicator information from all the appealing SUs and choose one as the favorite buyer for available channels.
- All the auction participants are risk neutral and they would like to maximize their own valuations according to their private action with reinforcement learning function.
- We assume a symmetric independent private value (SIPV) [16] cases where buyers and sellers only know their own valuation but no competitors' while getting the opposing parties' payoff information unilaterally and simultaneously.
- We consider only the case that channels are available for SUs to transmit once successful trading, focusing only on the cognitive spectrum access behavior, thus ignore the task of sensing the frequency spectrum, which has been well studied by many researchers [2,3].
- A common channel either dedicated or dynamic allocated, is assumed between PUs and SUs to carry the interactive information in our scheme and maintains synchronization between SUs so that SUs bid timely and tune to the correct channel to receive transmissions.

3 RL-Based Double Auction Algorithm for Dynamic Spectrum Access

In this section, we provide a RL-DA algorithm for SUs to access the network. Cognitive user will learn to improve its asking policy (reserved cost for the PU) and its bidding policy (bidding price for the SU) by participating in the auction. The optimal bidding policy for PU p_i is to generate an asking price that represents the cost for leasing the channels with respect to the marketing fluctuation, while the optimal bidding policy for SU s_j is to bring out a bid vector that illuminates its preference for using different channels. In this paper, we adopt a Q-Learning algorithm to acquire the reinforcement learning through an action-value function that gives the expected utility of taking a given action in a given state and following a fixed policy thereafter. We define two RL algorithms $\mathcal{L}_{Q1}$, $\mathcal{L}_{Q2}$ for PU and SU as functions taking the observation as input and having the auction policy as output, respectively.

3.1 Spectrum Auction with Reinforcement Learning

Generally, reinforcement learning systems are composed of a policy, a reward function and a value function, by which we can cope with the separate spectrum access problem. At each moment t, a secondary user s_j perceives the environment's state $s^t_{(s_j)}$ and performs an action $a^t_{(s_j)}$. One time step later, in part as a consequence of its action, the primary user receives a reward $r^{t+1}_{(p_i)}$, and s_j resides in a new state $s^{t+1}_{(s_j)}$. Primary user p_i will follow the same process. Herein and afterwards, we use the superscript as the time index, subscript as the user index or channel index, and a vacant bracket () indicates either p_i or s_j for simplicity.

Assume that each user in the cognitive radio network is able to make spectrum access decisions by itself. Q-Learning algorithm is used by PUs and SUs to dynamically access the

channels according to the history of states visited and the utility received due to the current choice of action.

To apply the Q-Learning algorithm, it is necessary to first define the *state*, *action*, *reward* and *learning policy* for users.

3.1.1 State

We define the occupation or availability of PU p_i on channels n at time t as the state $s^t_{p_i} = \{x^t_{p_i,n}\} \in \mathbf{S}$, where $x^t_{p_i,n} = z^t_{p_i,n} \cdot \delta^t_{p_i,n}$, $z^t_{p_i,n} \in \{0, 1\}$ represents the available transmission opportunity on channel n, i.e., $z^t_{s_j,n} = 1$ means that p_i has data to send on this channel currently and vice versa. Since the transition of $z^t_{p_i,n}$ to $z^{t+1}_{p_i,n}$ is determined by two stochastic events, i.e., spectrum occupation and no spectrum access, as described in the Markov chain of Sect. 2.2, the state transition from $s^t_{p_i}$ to $s^{t+1}_{p_i}$ is established, where $z^t_{p_i,n} = z^{t-1}_{p_i,n} \cdot (1 - u_{i,n}) + (1 - z^{t-1}_{p_i,n}) \cdot v_{i,n}$. Similarly, we define SU' current seized channels state as $s^t_{s_j} = \{y^t_{s_j,n}\} \in \mathbf{S}$, where $y^t_{s_j,n}$ means s_j taking transmission opportunity on the channel n at time t, and $y^t_{s_j,n} = 0$ otherwise. $\mathbf{S}$ is the finite set of possible state space, where $\mathbf{S} = \{S_k\}$, $S_k = \{s_n\}$, $s_n = \{0, 1\}$, $k = 1, \ldots 2^n$.

3.1.2 Action

Applying an action is to assign available channels from PUs to the current appealing SUs' access request in the network. For PUs, we define $a^t_{p_i} = \{\beta^t_{p_i,n}, \delta^t_{p_i,n} P^t_{p_i,n}\}$ as the action of choosing a bidder ($\beta^t_{p_i,n}$ as indicator) at an asking cost price $\delta^t_{p_i,n} P^t_{p_i,n}$. Meanwhile, for SUs, we define $a^t_{s_j} = \{\beta^t_{s_j,n}, P^t_{s_j,n}\}$ as the action of choosing a submitted channel ($\beta^t_{s_j,n}$ as indicator) at a bidding price $P^t_{s_j,n}$.

At each time step t, the cognitive user senses the current state $s^t_{()} \in \mathbf{S}$ of its environment and calculates an evaluation for each action $a^t_{p_i}$ and $a^t_{s_j}$ based on the Q-function. As a result, he receives an immediate reward $r^t_{()}$ and the environment's state changes from $s^t_{()}$ to a new state $s^{t+1}_{()}$ with certain transition probability.

3.1.3 Reward Function

The design of reward function $r^t_{()}(s^t_{()}, a^t_{()})$ is based on the thought that there is a reinforcement signal that directs the decision of action and brings benefit to system performance. In the cognitive radio scenario, what we expect is that on one side, PU chooses one of its most beneficial releasing task within its payoff surplus, i.e., to maximize PU's reward from this spectrum trade; On the other side, the SU chooses channels with the highest reward to access from the detected available channels. Hence, we define the reward in the above situations under current state $s^t_{()}$, performing the action $a^t_{()}$ for the PUs and SUs respectively as follows,

$$r^t_{p_i} = \sum_{n=1}^{m^t_{p_i}} \beta^t_{p_i,n} \delta^t_{p_i,n} P^t_{p_i,n} \tag{1}$$

$$r^t_{s_j} = \sum_{n=1}^{m^t_{s_j}} \beta^t_{s_j,n} P^t_{s_j,n} \tag{2}$$

$$\text{s.t.} \qquad \beta^t_{p_i,n} \in \{0, 1\}, \beta^t_{s_j,n} \in \{0, 1\} \quad \text{(C1)}$$

$$\sum_i \beta^t_{p_i,n} \le \sum_{i=1}^{I} N^i, \sum_j \beta^t_{s_j,n} \le \sum_{i=1}^{I} N^i \quad \text{(C2)}$$

Herein, $m^t_{p_i}$, $m^t_{s_j}$ are the channels assigned to p_i and s_j respectively. (C1)–(C2) are the constraints for action space of users. The reward for SUs $P^t_{s_j,n}$ is defined as the transmission capability in each subcarrier. Moreover, the SU pays for the spectrum renting by $P^t_{s_j,n}$. Assume that the payoff $P^t_{s_j,n}$ is transferred to the PU absolutely, i.e., ϵ=1 in (3). ϵ is the bargain factor. This means that the reward of PU equals the payoff of SU once spectrum access succeed.

$$P^t_{p_i,n} = \epsilon P^t_{s_j,n} \tag{3}$$

3.1.4 Learning Policy

A learning policy is used to map the history of states visited $s^t_{()}$, the probability of action chosen $a^t_{()}$, and the utility received $Q^{t+1}_{()}(s^t_{()}, a^t_{()})$, into current choice of action. Note that the learning policy being different between the PUs and SUs, the reward values for the two classes will have different orders and they learn optimally depending on their diverse situations and locations in the network. Thus we define two Q-functions $Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i})$ and $Q^{t+1}_{s_j}(s^t_{s_j}, a^t_{s_j})$ for PUs and SUs separately.

$$Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i}) = (1-\phi_{p_i})Q^t_{p_i}(s^t_{p_i}, a^t_{p_i}) + \phi_{p_i}[r_{p_i} + \varphi_{p_i} \max_{a\in A} Q^t_{p_i}(s^{t+1}_{p_i}, a^{t+1}_{p_i})] \tag{4}$$

$$Q^{t+1}_{s_j}(s^t_{s_j}, a^t_{s_j}) = (1-\phi_{s_j})Q^t_{s_j}(s^t_{s_j}, a^t_{s_j}) + \phi_{s_j}[r_{s_j} + \varphi_{s_j} \max_{a\in A} Q^t_{s_j}(s^{t+1}_{s_j}, a^{t+1}_{s_j})] \tag{5}$$

Q-Learning updates the action values and Q-functions by the following rules to approach the true values under the optimal policy $(a^{t+1}_{p_i})^*$ and $(a^{t+1}_{s_j})^*$, which are the expected sum of rewards discounted by $\varphi_{()}$ under $(a^{t+1}_{p_i})^*$ and $(a^{t+1}_{s_j})^*$, i.e.,

$$(a^{t+1}_{p_i})^* \leftarrow \text{argmax}(Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i})) \tag{6}$$

$$(a^{t+1}_{s_j})^* \leftarrow \text{argmax}(Q^{t+1}_{s_j}(s^t_{s_j}, a^t_{s_j})) \tag{7}$$

Therefore, cognitive entity comes up with a positive preference decision unilaterally. The learning rate $\phi_{()}$, $0 < \phi_{()} < 1$, is a convergent step-size parameter that determines the updating speed of the Q-function. When the learning rate parameter $\phi_{()}$ is close to 1, the reward changes rapidly in response to new experiences. The discount factor $\varphi_{()}$, $0 \le \varphi_{()} \le 1$, determines the present value of future rewards. When it is close to 1, future interaction plays a substantial role in defining the total utility values [23].

The Q-function updates its evaluation of the value for the action while taking into account (1) the immediate reinforcement value $r^{t+1}_{()}$ and (2) the estimated Q-value of the new state $Q^t_{()}(s^{t+1}_{()}, a^{t+1}_{()})$. The procedure that cognitive user follows to compete for the channel opportunities is illuminated as Table 1 in Sect. 4.3.

3.2 Convergence of the Q-learning Algorithm

The conditions for convergence of Q-learning with a time varying learning rate $\phi_{()}$ that uses the results derived from Robbins–Monro theory [24] are given here.

Table 1 Learning procedure

Algorithm1 Q-Learning Algorithm for p_i and s_j

1 Initialization: V_{p_i}, $Q^0_{p_i}$, $Q^0_{s_j}$

2 Learning: For every sensing period T, perform the following process:

At every edge node p_i and s_j

For $n = 1, \ldots, N_i$

Repeat

Initialize $Q^t_{()}(s^t_{()}, a^t_{()}), \forall s^t_{()}, \forall a^t_{()}, s^t_{()} \in S, a^t_{()} \in A$

At time t, Choose an action $a^t_{()}$, observe $r^t_{()}, s^{t+1}_{()}$

Repeat $s^{t+1}_{()} \in S, s^{t+1}_{()} \in A$

Update $Q^{t+1}_{()}(s^t_{()}, a^t_{()}) \leftarrow (1-\phi_{p_i})Q^t_{p_i}(s^t_{p_i}, a^t_{p_i}) + \phi_{()}[r_{()} + \varphi_{()}\max_{a\in A} Q^{t+1}_{()}(s^t_{()}, a^t_{()})]$

Until all the state and action space is terminal $(a^{t+1}_{()})^* = \text{argmax}(Q^{t+1}_{()}(s^t_{()}, a^t_{()}))$

Update $s^{t+1}_{()}, a^{t+1}_{()}$

Until convergence

End

Theorem 1 *The Q-learning algorithm given by* (4) *and* (5) *converges to the optimal* $(Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i}))^*$, $(Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i}))^*$ *values uniformly over* $s^t_{()}$ *and* $a^t_{()}$ *with probability of 1 if the following conditions are met* [25]*:*

(1) *The state and action spaces are finite.*
(2) $\sum_{t=0}^{+\infty} \phi_{()} = \infty$ *and* $\sum_{t=0}^{+\infty} \phi^2_{()} = \infty$.
(3) $Var\{r^t_{()}(s^t_{()}, a^t_{()})\}$ *is finite.*
(4) *If* $\phi_{()} = 1$*, all policies lead to a cost free terminal state with probability 1.*

Proof See Appendix. □

3.3 RL-based Double Auction Algorithm

We explore a double auction algorithm based on the above Q-learning iterative policy, the spectrum access policy (biding price) from the SU and the release policy (asking price) from the PU. The double auction can be analyzed as a game where potential buyers submit their bids and potential sellers simultaneously submit their asking prices to an auctioneer, and then the auctioneer chooses some price $P^t_{s_j,n}$ that clears the market.

In our proposed RL-DA algorithm, both the bidders and suppliers have some valuations of a good, that is their separate utility functions. Strategies made accord with their bids or asking prices for spectrum which depends on the private variation of buyers and sellers. Moreover, as for PUs, the asking price, also called reserved cost price, changes as the market fluctuates on the available subchannel. Meanwhile the payoff of SU depends on the private preference for the channel including the current transaction state and evaluation of the future behavior. Next, we explore the reserved cost prices for PUs and bidding prices for SUs deeply.

3.3.1 Reserved Cost Prices $C^t_{p_i,n}$ for the PU

It is intuitive that there is a cost price used to ensure the profitability of spectrum owners in the market competition. Also, it is necessary in the auction theorem to combat the collusive

behaviors of SUs. In the spectrum auction framework, the PU presents an adaptive reserved cost price $C^t_{p_i,n}$ by updating the Q-function during intervals, which means that the spectrum resource won't be sold lower than the reserved price. Therefore the design of reserved cost price ensures that it is lucrative for the sellers to get from this auction business. Next, we define the reserved cost $C^t_{p_i,n}$ in the context of the RL-DA model when using the channel n as the possible future reward variant with $Q^{t+1}_{p_i,n}(s^t_{p_i}, a^t_{p_i})$, $Q^{t+1}_{p_i,n}(s^t_{p_i}, a^t_{p_i}) \in Q^{t+1}_{p_i}(s^t_{p_i}, a^t_{p_i}), n \in \{1, 2, \ldots, Nc\}$.

$$C^t_{p_i,n} = Q^{t+1}_{p_i,n}(s^t_{p_i}, a^t_{p_i}) \tag{8}$$

Accordingly, PUs set the reserve cost price from the Q-Learning process obtained from (4). Different from fixed or without cost price, this QL-based reserved price involves a learning and updating process which takes consideration of the history visited and future expectation.

3.3.2 Bidding Prices $P^t_{s_j,n}$ from the SU

It is evident that SUs as buyers in the market have to bid at some price to achieve the spectrum transmission. How to produce such price is key to SUs and impacts the on-going of spectrum market. Herein, we use SU' preference $l^t_{s_j,n}$ over the channel n to express the optimal bid price. Assume that at each time slot t, SU s_j has preference $l^t_{s_j,n}$ over the channel n, which captures the benefit derived when using the channel. Thereby, the optimal bid that SU s_j can make is $P^t_{s_j,n} = l^t_{s_j,n}$, i.e., the optimal bid for SU s_j is to announce its true preference to the auctioneers.

Next, we define the preference $l^t_{s_j,n}$ as in (9). In the context of the RL-DA model, it can be interpreted as the accumulative total packets $B^t_{s_j,n}$ in the buffer plus the future reward $Q^{t+1}_{s_j,n}(s^t_{s_j,n}, a^t_{s_j,n})$ when s_j moves to the next state $s^{t+1}_{()}$. Herein, f is a parameter that regulates the tradeoff between the current packet and future market expectation.

$$l^t_{s_j,n} = f \cdot B^t_{s_j} + Q^{t+1}_{s_j,n}(s^t_{s_j}, a^t_{s_j}) \tag{9}$$

We follow the buffer state model of [21], assuming the number of packets arriving into the buffer during one time slot is a random variable independent of the time t and denoted as $A^t_{s_j,n}$. $A^t_{s_j,n}$ follows the Poisson distribution with the average arrival rate A packets per second. The buffer capacity is set to be X_{s_j}, therefore, the buffer state of s_j at time t can be calculated as

$$B^t_{s_j,n} = \min\{(B^{t-1}_{s_j,n} - R^{t-1}_{s_j,n})^+ + A^t_{s_j,n}, X_{s_j}\} \tag{10}$$

where $R^t_{s_j,n}$ is the immediate gain by transmitting the packets, and $(\bullet)^+ = \max\{0, \bullet\}$.

Consequently, SUs bid at the price with respect to the accumulated packet and learning process distributedly which relates the practical requirement and dynamic learning.

3.3.3 RL-DA Algorithm Implementation

Assume that the scheduling period of SU is T, and the real sensing time is much less than that. Also, we assume that the cost of PUs and payoffs of SUs remain unchanged over this period. In the supply falling short of demand case where PUs dominate in the market, we appoint PU the auctioneer in order to reduce the exchange information iteration. At each interval T, the auctioneer settles the spectrum access transaction once he receives the maximal bidding price from SU under the constraint of the received price. Note that the auctioneers' goal is

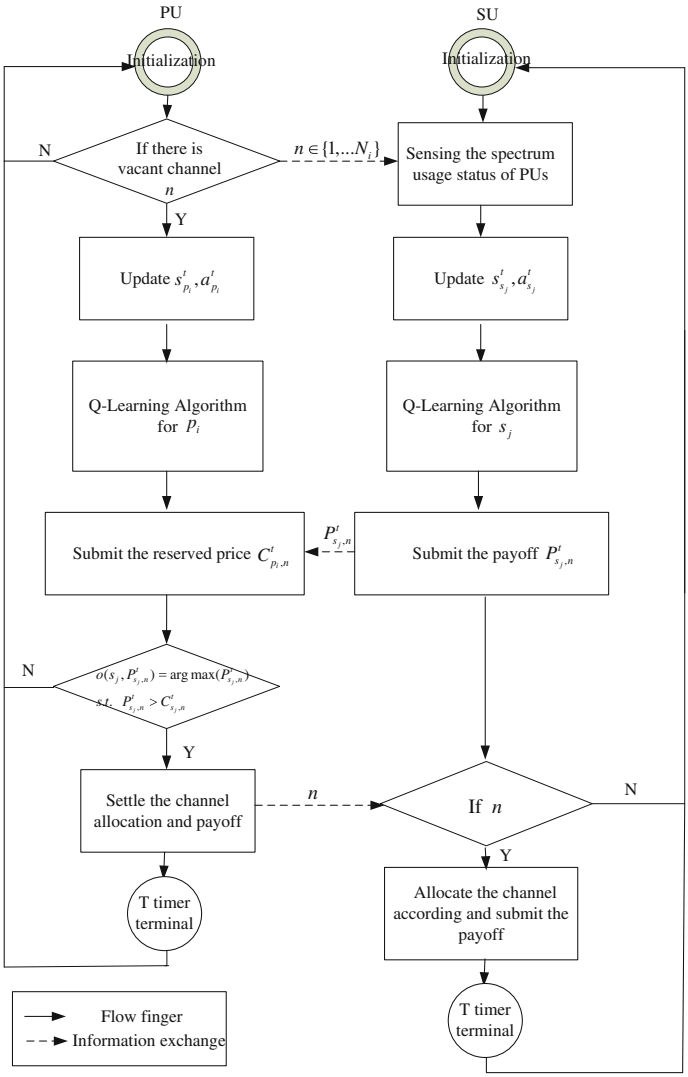

Fig. 3 Procedure and information exchange for PUs and SUs to play the auction game at time slot t

to maximize their own utility function, thereby the optimization problem of PUs is written as (11). The detailed algorithm is executed as illustrated in Fig. 3.

$$o(s_j, P^t_{s_j,n}) = \text{argmax}(r^t_{p_i,n}(s^t_{p_i,n}, a^t_{p_i,n})) \quad (11)$$
$$\text{s.t.} \quad P^t_{s_j,n} \geq C^t_{s_j,n} \quad \text{(C1)}$$

4 Simulation Results and Discussions

In this section, we show the performance of our proposed RL-DA framework via computer simulations. We justify the convergence of QL-based bidding prices and reserved price, test

Table 2 System parameters

Parameter	Value/assumption
Total bandwidth	1.25 (MHz)
power constraint for SU	1W
Noise power spectral density	−174 (dBm/Hz)
Bandwidth of subcarriers	10 (kHz)
Number of subcarriers	128
Modulation scheme	BPSK, QPSK, 16QAM, 64QAM
time period T (in s)	0.1
learning rate_default	As Eq. (15)
discount factor_default	0.5
Maximum Doppler shift	30 (H)z
Propagation model (in dB)	128.1 + 37.6 log10(R) (R in km)
Small-scale fading model	Six independent Rayleigh multipaths, exponential power delayprofile with decaying rate 2 and 10 μs delay speed

the effect of learning factor and discount factor. Also, we compare the performance of our proposed scheme with other bidding strategies for SUs and different reserved prices for PUs over time.

4.1 Parameter Setting

We simulate a cognitive radio environment with 5 PUs (i.e., $I = 5$) and 10 SUs (i.e., $J = 10$). Each mobile station's location is randomly generated and evenly distributed within the cell radius of 1,000 m. Besides, we assume that the SUs compete for the available spectrum opportunity to transmit delay-sensitive multimedia data. Set the uniform buffer size for all SUs X_{s_j} as 25 bit, and the average arrival rate of the Poisson distribution packets A is 10 bit/s. Other detailed values of the simulation parameters are shown in Table 2. All the performance evaluation is executed 20 times.

4.2 Convergence of Q-Learning

In Figs. 4 and 5, we test the convergence of Q-Learning algorithm for the bidding price of SUs and reserved cost of PUs. Generally, both of them converge within 20 iterations. Taking the default value (0.5, 0.5) for example, (all the legends in Figs. 4, 5 are labeled by the format of $(\phi_{()}, \varphi_{()})$, in the 10th iteration, it achieves about 96.4 % for bidding price in Fig. 4 and 93.1 % for reserved price in Fig. 5. Although 20 times iteration is acceptable, we can also appeal to different iterations considering the balance of the computation complexity and accuracy. In this paper, we set 20 times iteration targeting to reinforce the optimal effectivity of the proposed algorithm. We also examine the effect of discount factor $\varphi_{()}$ and learning rate $\phi_{()}$ for the convergence of the algorithm, the detailed analysis is given in the next subsection.

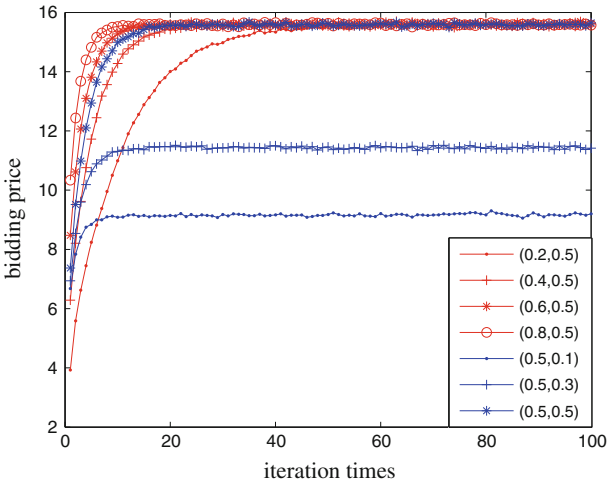

Fig. 4 Convergence of the bidding price for SUs

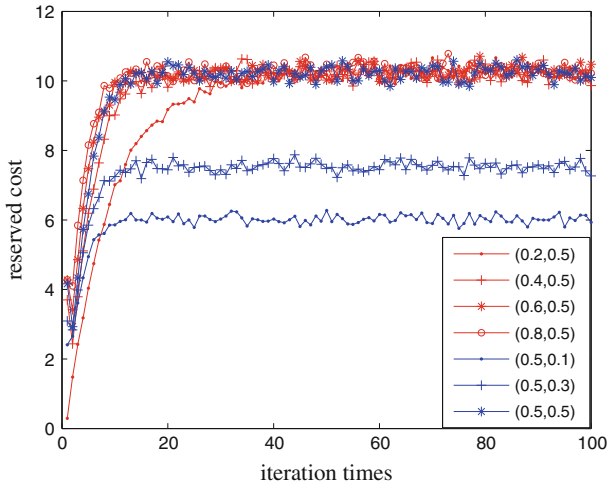

Fig. 5 Convergence of the reserved cost for PUs

4.3 Effect of Discount Factor and Learning Factor

Considering the restriction of discount factor, we compare the effect of $\varphi_()$ in the range of 0 and 1 for both the bidding price and reserved price in Fig. 6. We can see that the prices grow rapidly with $\varphi_()$ when it is over 0.8. This means that when the future rewards play a much more important role in the Q-function referring to (4–5), thus both the prices rise rapidly. However, the reserved price of PUs is higher than the bidding price of SUs when the discount factor is too low or too high. Therefore, there exists an effective range of [0.025, 0.735] for $\varphi_()$, which ensures that the auction suffices to trade between PUs and SUs. The reason of this is that PUs won't release the spectrum unless the bidding price is higher than the reserved

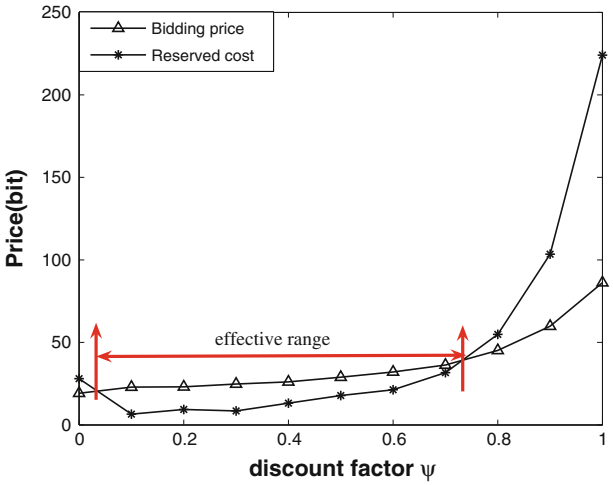

Fig. 6 Effect of the discount factor for the bidding price and reserved cost

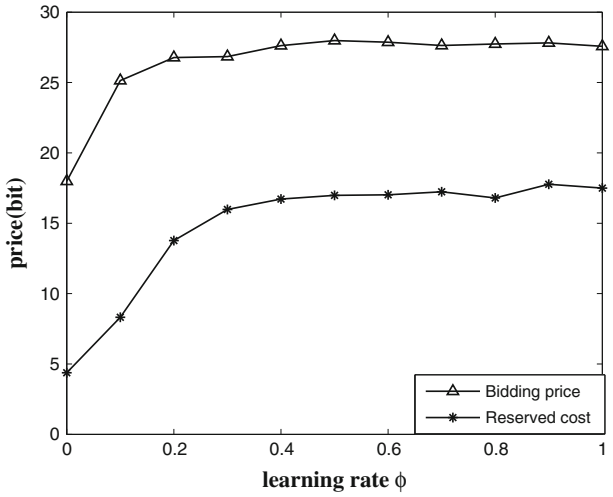

Fig. 7 Effect of the learning rate for the bidding price and reserved cost

cost price of PUs. Without loss of generality, we set the default value of $\varphi_{()}$ as 0.5 in the followed system simulation.

Also, Figs. 4 and 5 show effect of the discount factor from the comparison of the blue lines. We can see that the variance of $\varphi_{()}$ does not impact the convergence speed but only the converged value, which implicits that the component of current state or future reward in the final decision works on the future action correspondingly.

The learning rate $\phi_{()}$ determines the updating speed of the Q-function. Figure 7 depicts the effect of learning rate for values of the bidding price and reserved cost. When the learning rate parameter $\phi_{()}$ is over 0.3, both the bidding price and reserved price keep a roughly equal value, which means that the learning rate won't sway the value much. However, from the red lines illustrated in Figs. 4 and 5, we can see that with the variance of $\phi_{()}$, the converged value

keeps uniform, but gradient differs. This is to say, contrary to $\varphi_{()}$, the learning rate does not impact the converged value but only the convergence speed.

Meanwhile, in Fig. 7, the bidding price is over the reserved cost in the whole set of $\phi_{()} \in [0, 1)$, which means that all the value between 0 and 1 is fit for $\phi_{()}$ to perform the RL-DA algorithm but still required to satisfy requirement of the Theorem 1-(2).

4.4 Various Bidding Strategies of SUs

In this subsection, we highlight the metrics of the bidding prices in the double auction framework by comparing different bidding strategies for SUs. For comparison, we deploy the following three strategies when SUs are required to submit the biding vector,

(a) *Blind Bidding Strategy* $\pi_{s_j}^{blind}$: This strategy generates bidding vectors by considering the disturbance of transmission only based on the current buffer state; therefore it presents a uniform bidding prices on all the available channels as (12) without consideration of diversity on physical channels.

$$P_{s_j,n}^t = f \cdot B_{s_j}^t \tag{12}$$

(b) *Short-sighted Bidding Strategy* $\pi_{s_j}^{short}$: Different from $\pi_{s_j}^{blind}$, this strategy not only focuses on the internal state of the buffer state but the immediate transmission gains, (i.e., $R_{s_j,n}^t$ is the perceived rate on the available channels calculated according to Shannon Capability) and it presents as (13).

$$P_{s_j,n}^t = f \cdot B_{s_j}^t + R_{s_j,n}^t \tag{13}$$

(c) *QL Bidding Strategy* $\pi_{s_j}^{\mathcal{L}_Q}$: This strategy is produced using RL-DA algorithm proposed in Sect. 4. Under this consideration, SU deduces the learning based strategy by the historical observation related forecasting function as (14).

$$P_{s_j,n}^t \leftarrow (a_{s_j,n}^{t+1})^* \tag{14}$$

Note that Each SN announces a static price during the **T** period for once bidding on each channels according to the above three strategies simultaneously.

Figure 8 compares the aggregated packet loss with respect to the three scenarios. QL bidding strategy reduces the packet loss compared with $\pi_{s_j}^{blind}$ and $\pi_{s_j}^{short}$ strategies. This significant improvement lies in that cognitive users can accurately value the channel opportunities by modeling and caring for the experienced dynamics, i.e., channel availability. As time extends, the aggregated packet loss grows linearly for these three scenarios. However, $\pi_{s_j}^{\mathcal{L}_Q}$ strategy increases with the slowest speed and keeps lowest packet lost.

Figure 9 shows the average bidding efficiency with time, which is defined as rate per bidding price (bit/s/p). It can be seen that QL bidding strategy $\pi_{s_j}^{\mathcal{L}_Q}$ keeps the dominance uniformly. Averagely, the bidding efficiency of $\pi_{s_j}^{short}$ strategy is 2 % higher than $\pi_{s_j}^{blind}$, while $\pi_{s_j}^{\mathcal{L}_Q}$ is 4.4 % higher than $\pi_{s_j}^{blind}$. This is improved by higher packet delivered rate, since the learning based bidding model takes the future received reward of the agent into account. Notice that the twitter for the three scenarios is due to disturbance caused by the transmission opportunity, dynastic channel, varying anticipated reward, i.e., the disturbance of the network.

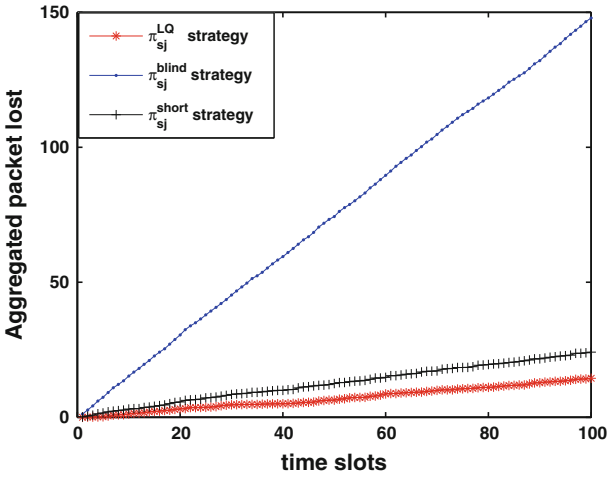

Fig. 8 Packet loss comparison of different bidding strategies

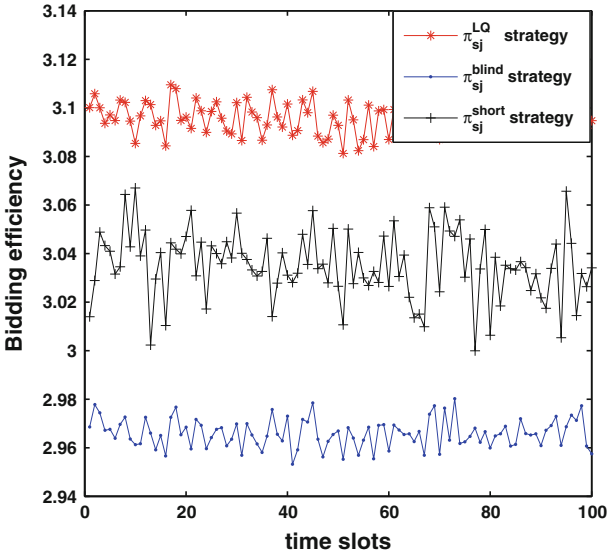

Fig. 9 Bidding efficiency of different bidding strategies

4.5 Reserved Prices for PUs (Without, With_fixed, With_QL)

We compare the operation of double auction framework with different reserved cost for PUs, i.e., *without* reserved cost, *with fixed* reserved cost or *with QL* reserved cost. The *with fixed* reserved price for the $\pi_{s_j}^{blind}$ and $\pi_{s_j}^{short}$ is equal to the average payoff of the previous bidding while the $\pi_{s_j}^{\mathcal{L}_Q}$ strategy acquires received price by the $\pi_{s_j}^{\mathcal{L}_{Q2}}$ as (10). Meanwhile, $C_{p_i,n}^t$ is set to be zero for the *without* case. Table 3 presents the user rate for successful bidders for

the two cases at 100 s instantaneous slot. We can observe from Table 3 horizontally for the comparison of different reserved prices strategies of PUs and vertically for that of different bidding strategies of SUs.

Horizontally for every three cases of the same SU, it gets a higher rate in *without* cases than *with fixed* or *with QL* cases, however, this higher rate may be caused by cheating or collusion behavior from multiple SUs, while with *with QL* reserved cost of PUs, the dynamic access is regulated dynamically, for *with QL* allows more access opportunities than *without* and *with fixed*.

Vertically for all SUs in every three cases, they get higher rate for the $\pi_{s_j}^{blind}$ and $\pi_{s_j}^{short}$ than $\pi_{s_j}^{\mathcal{L}_Q}$ if any, but less users are served in these two than $\pi_{s_j}^{\mathcal{L}_Q}$. This indicates that blind bidding from SUs causes resource amass on few users. On the contrary, more users get the transmission opportunity in $\pi_{s_j}^{\mathcal{L}_Q}$ than $\pi_{s_j}^{short}$ and $\pi_{s_j}^{blind}$ strategies, which makes fair by long run reserved price evolving with the learning rate ϕ_{s_j} and discounted by discount factor φ_{s_j}.

Figure 10 compares the average rate of reserved cost for three different strategies. Uniformly, strategies *without* reserved price achieve the highest average price; strategies *with fixed* reserved price get the lowest average price, while strategies *with QL* reserved price stands middle. This is because that *without* reserved cost strategies allow more SUs to access of the licensed band, thereby achieve the highest average rate for SUs; However, there is no protection for PUs' local interest which may be impaired by collusion or cheating behavior by SUs. In the *with QL* strategies, PUs can adjust the reserved cost according to the current status and future expectation so that the success rate of auction is bigger than the *with fixed* strategies.

On the other hand, it turns out that the average rate for $\pi_{s_j}^{\mathcal{L}_Q}$ outperforms $\pi_{s_j}^{blind}$ and $\pi_{s_j}^{short}$ in every cost price. The $\pi_{s_j}^{short}$ bidding price achieves the highest rate, the $\pi_{s_j}^{short}$ bidding price achieve the lower rate but $\pi_{s_j}^{blind}$ achieves the lowest rate. This result holds for all the three reserved prices cases.

4.6 Complexity Issue

After simulating the system performance of the proposed RL-DA algorithm, we further discuss the communication and computation complexity for implement. It is observed from Fig. 3 that there are at most two information exchange between each PU and SU. One for submission of $P_{s_j,n}^t$ to PU. As for the other, since in our proposed algorithm, the PU accept the submitted bidding price to clear the auction, there is no settled price feedback but only a backhaul channel assignment information. Thereby, the network communication iteration is maximally $(2 \times I \times J)$. In this paper, we appoint PU works as auctioneer in the market, so that the interaction process is greatly reduced.

As to the computation complexity, we measure it from the storage burden and Q-operation. To perform the QL algorithm, we need three tables to carry the matrix of state, reward, and Q-function. According to the data size, we set one integer for state, one float for reward and one float to store each iteration result of Q-function. Therefore, the total storage burden is $(20 \times [(N_c \times I) + (N_c \times J)])$, from which we can see that it grows linearly with the subcarrier and cognitive users. We can imagine that the main operation involved in our algorithm comes from the Q-operation. Each interval T has $(20 \times [(N_c \times (1 - z_{p_i,n}^t) \times \delta_{p_i,n}^t)(I + J)])$ Q-operation, from which we can note that the more spectrum opportunity costs the more operation computation.

Table 3 Rate for SU with or without reserved price (Bit/s)

	SU1			SU3			SU5			SU6			SU9		
	With fixed	With QL	Without	With fixed	With QL	Without	With fixed	With QL	Without	With fixed	With QL	Without	With fixed	With QL	Without
$\pi_{s_j}^{blind}$	0	0	0	235.23	249.03	645.08	0	0	0	0	0	0	0	0	0
$\pi_{s_j}^{short}$	0	0	0	186.21	188.10	219.26	0	0	0	70.41	71.77	87.39	200.21	210.11	262.78
$\pi_{s_j}^{\mathcal{L}_{gl}}$	154.57	154.57	154.57	172.53	172.53	177.53	134.75	140.21	143.14	40.23	51.65	62.78	103.01	116.79	199.41

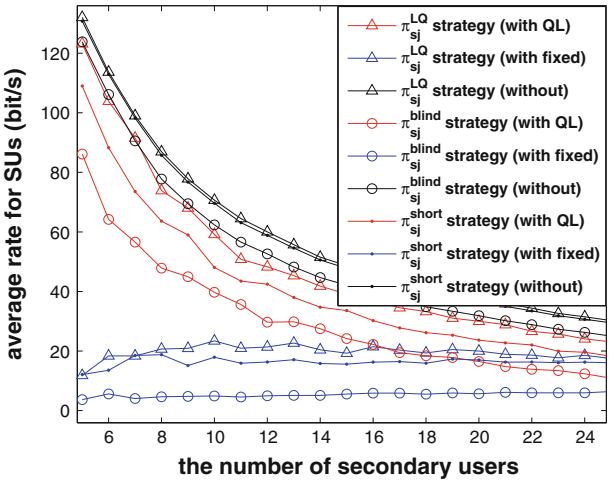

Fig. 10 Average rate comparison of reserved price for three different strategies

5 Conclusions and Future Work

In this paper, we have proposed a reinforcement-learning-based double auction algorithm for dynamic spectrum access in cognitive radio networks. In the proposed RL-DA algorithm, both SUs and PUs are allowed simultaneously and independently to make bid decisions on resource considering their current states, experienced environment and estimated future reward in the auction market, i.e., SUs are to generate a bidding prices according to the current transmission requirement and future expectation; meanwhile, PUs are to produce a reserved cost vector so as to combat the bidding collision and ensure the self-profit. Finally, we validate the practicality of RL-DA algorithm by convergence analysis and test of the adjusting variables. Also, simulation results confirm that the proposed RL-DA can improve not only the bidding efficiency but also average rate for RL-bidding strategies effectively.

We note that both the bidding and reserved prices are executed locally without impulsing information from outside environment, which might result in a trapped self-glorification condition. This is due to the intrinsical feature of Q-learning that it compares the expected utility of the available actions without requiring a model of the wide environment. Future work is in progress to improve the reinforcement learning method with observed information from a broad environment to reduce blind or inefficacy strategies.

Acknowledgments This work is supported by the National Natural Science Foundation of China under Grant No. 60971083, 61171097 and 61101107, the National International Science and Technology Cooperation Project under Granted NO.2010DFA11322, and the Chinese Universities Scientific Fund under Granted NO.2012RC0306.

Appendix

The four conditions of Theorem 1 can be proved as follows:

(1) In the Sect. 4.2, we defined $a_{()}^t$ and $s_{()}^t$, which are present owning channels and channels occupied or released with certain cost or bidding price, is finite. Hence (1) holds.

(2) In our RL-DA algorithm, we define

$$\phi() = \begin{cases} 1/t, \ t > 0. \\ 0, \quad \text{otherwise.} \end{cases} \tag{15}$$

therefore, it is easy to prove that (2) holds.

(3) According to the rewards defined in (1)–(2), we define $r_{s_j}^{\min} \leq r_{s_j} \leq r_{s_j}^{\max}$. Since $(r_{()}^t)^2$ is finite, $Var\{r_{()}^t\} = E((r_{()}^t)^2) - (E(r_{()}^t))^2$ is finite.

(4) When $\gamma = 1$, the infinite horizon model tends to the so called gain optimal policy wherein the objective is to maximize the long term average reward. In such a case, all policies lead to a terminal state with a probability 1 which is true for any finite horizon model and hence (4) holds.

References

1. Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications, 23*(2), 201–220.
2. Yucek, T., & Arslan, H. (2009). A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Communications Surveys & Tutorials, 11*(1), 116–130.
3. Haykin, S., Thomson, D. J., & Reed, J. H. (2009). Spectrum sensing for cognitive radio. *IEEE Proceedings, 97*(5), 849–877.
4. Yu, F. R., Huang, M., & Tang, H. (2010). Biologically inspired consensus-based spectrum sensing in mobile ad hoc networks with cognitive radios. *IEEE Networks, 24*(3), 26–30.
5. Yiping, X., Chandramouli, R., Stefan, M., & Sai Shankar, N. (2006). Dynamic spectrum access in open spectrum wireless networks. *IEEE Journal on Selected Areas in Communications, 24*(3), 626–637.
6. Zhu, J., & Liu, K. J. R. (2007). Cognitive radios for dynamic spectrum access—dynamic spectrum sharing: A game theoretical overview. *IEEE Communications Magazine, 45*(5), 1–5.
7. Wang, B., Wu, Y., Ji, Z., Liu, K. J. R., & Clancy, T. C. (2008). Game theoretical mechanism design methods. *IEEE Signal Processing Magazine, 25*(6), 74–84.
8. Krishnamurthy, V. (2009). Decentralized spectrum access amongst cognitive agents—an interacting multivariate global games approach. *IEEE Transactions on Signal Processing, 57*(10), 3999–4013.
9. Niyato, D., & Hossain, E. (2008). Competitive spectrum sharing in cognitive radio networks: A dynamic game approach. *IEEE Transactions on Wireless Communications, 7*(7), 2651–2660.
10. Zhu, J., & Liu, K. J. R. (2008). Multi-stage pricing game for collusion-resistant dynamic spectrum allocation. *IEEE Journal on Selected Areas in Communications, 26*(1), 182–191.
11. Zhu, J., & Liu, K. J. R. (2006). Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users. In *Proceedings of IEEE SECON'06*, pp. 119–127.
12. Niyato, D., & Hossain, E. (2008). Market-equilibrium, competitive, and cooperative pricing for spectrum sharing in cognitive radio networks: Analysis and comparison. *IEEE Transactions on Wireless Communications, 7*(11), 4273–4283.
13. Niyato, D., & Hossain, E. (2008). Spectrum trading in cognitive radio networks: A market-equilibrium-based approach. *IEEE Wireless Communications, 15*(6), 71–80.
14. Bobrow, D. G. (1994). *Artificial intelligence in perspective*. Cambridge, MA: MIT Press.
15. Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 71–80, 1–5.
16. Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
17. Yu, F. R., Wong, V. W. S., & Leung, V. C. M. (2008). A new QoS provisioning method for adaptive multimedia in wireless networks. *IEEE Transactions on Vehicular Technology, 57*(3), 1899–1909.
18. Bernardo, F., Agusti, R., Perez-Romero, J., & Sallent, O. (2011). Intercell interference management in OFDMA networks: A decentralized approach based on reinforcement learning. *IEEE Transactions on Systems, Man and Cybernetics, Part C, 41*(6), 1–9.
19. Gibbons, R. D. (1992). *Game theory for applied economists*. Princeton: Princeton University Presss.
20. Shankar, S., Chou, C. T., Challapali, K., & Mangold, S. (2005). Spectrum agile radio: Capacity and QoS implications of dynamic spectrum assignment. *Proceedings of IEEE Globecom'05*, pp. 2510–2516.
21. Fangwen, F., & Schaar, M. (2009). Learning to compete for resources in wireless stochastic games. *IEEE Transactions on Vehicular Technology, 58*(4), 1904–1919.

22. van der Niyato, D., Hossain, E., & Han, Z. (2009). Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach. *IEEE Transactions on Mobile Computing, 8*(8), 1009–1022.
23. Roy, N., Roy, A., & Das, S. K. (2005). A cooperative learning framework for Mobility-aware resource management in multi-inhabitant smart momes. In *Proceedings of IEEE MobiQuitous'05*, pp. 393–403.
24. Jakkola, T., & Singh, S.P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation, 6*(6), 1185–1201.
25. Venkatesh T., Kiran Y. V., Murthy C. S. R. (2009) Joint path and wavelength selection using Q-learning in Optical burst switching networks. In: *Proceedings of IEEE Globecom'09* (Vol. 15, no 6, pp. 1–5).

Author Biographies

Yinglei Teng received the B.S. degree from Shandong University, China, in 2005, and received the Ph.D. degree in electrical engineering from Beijing and University of Posts and Telecommunications (BUPT) in 2011. She is now working as an Assistant Professor in School of Electronic Engineering at BUPT. Her current research interests include multiuser MIMO, adaptive OFDM, cooperative diversity, cognitive radio network and cross-layer design technology.

F. Richard Yu received the Ph.D. degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2004, he was with Ericsson (in Lund, Sweden), where he worked on the research and development of dual mode UMTS/GPRS handsets. From 2005 to 2006, he was with a start-up in California, USA, where he worked on the research and development in the areas of advanced wireless communication technologies and new standards. He also held a position as a Postdoctoral Research Fellow with UBC in 2005 and 2006. He joined Carleton School of Information Technology and the Department of Systems and Computer Engineering (cross-appointment) at Carleton University, Ottawa, in 2007, where he is currently an Assistant Professor. He received the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and best paper awards at IEEE/IFIP TrustCom 2009 and Int'l Conference on Networking 2005. His research interests include cross-layer design, security and QoS provisioning in wireless networks. He serves on the editorial boards of several journals, including IEEE Communications Surveys & Tutorials, EURASIP Journal on Wireless Communications Networking, Ad Hoc & Sensor Wireless Networks, Wiley Journal on Security and Communication Networks, and International Journal of Wireless Communications and Networking.

Ke Han received the Ph.D. degree in electrical engineering from Beijing and University of Posts and Telecommunications (BUPT) in 2007. His current research interests include wireless communication system design and integrated circuit design of communication.

Yifei Wei received the B.Sc. and Ph.D. degrees in electrical engineering from Beijing University of Posts and Telecommunications (China), in 2004 and 2009, respectively. He is currently a faculty member in the School of Electronic Engineering, Beijing University of Posts and Telecommunications. He was involved in several projects funded by National High Technology Research and Development Program of China, and National Natural Science Foundation of China. He was invited to study at Carleton University (Canada) in electrical engineering for one year supported by China Scholarship Council. His research interests are in wireless mesh networks, heterogeneous converged networks, and cooperative relaying networks.

Yong Zhang is an Associate Professor in School of Electronic Engineering at the Beijing University of Posts and Telecommunications (P.R.C). He holds a Ph.D. degree in Beijing University of Posts and Telecommunications. His research interests include mobile communication, cognitive network, and self-organizing network.