Rahman *et al. Hum. Cent. Comput. Inf. Sci.* (2016) 6:2
DOI 10.1186/s13673-016-0057-2

Human-centric Computing and Information Sciences

RESEARCH Open Access

An efficient transmission mode selection based on reinforcement learning for cooperative cognitive radio networks

Md. Arifur Rahman, Young-Doo Lee and Insoo Koo*

*Correspondence:
iskoo@ulsan.ac.kr
School of Electrical Engineering, University of Ulsan, 93 Daehak-ro, Nam-gu, Ulsan 44610, South Korea

Abstract

Cooperative communication systems use cooperative relays for transmitting their data packets to the destination. Cooperative communication schemes having all relays participating in transmission may cause unnecessary wastes of most valuable spectrum resources. So it is mandatory to effectively select a transmission mode for cooperative cognitive radio networks (CCRNs). In this paper, an efficient transmission mode scheme based on Q-learning algorithm is proposed. State, action, and reward are defined to achieve a good performance on time delay and energy efficiency in data transmission as well as the interference to primary users during secondary users transmission. The proposed scheme selects an optimal action on the networks environment to maximize the total revenue of the multilateral metric. The simulation result shows that the proposed scheme can efficiently support the determination for the transmission mode and outperforms conventional schemes for a single metric in CCRNs.

Keywords: Cooperative communication, Cognitive radio networks, Transmission mode, Relay selection, Q-learning

Introduction

Cognitive radio is a promising wireless technology to resolve the growing scarcity of the indispensable electromagnetic spectrum resources. In cognitive radio, secondary users (SUs) detect spectrum holes in their radio environment for data transmission, while primary users (PUs) use their own licensed spectrum band [1]. CR techniques for spectrum access (SA) are generally classified as underlay SA, overlay SA, and interweave SA [2]. Among it, the underlay SA has higher flexibility since it focuses on specific power control for satisfying a required interference to the PUs. Hence, the other can be understood as the special cases of the underlay SA. Basically, the underlay SA allows simultaneous transmission of SUs and PUs by sharing the spectrum as long as the interference to PUs during the secondary transmission is under a predefined threshold level called interference temperature [3]. Thus, to reduce the interference to PUs in a manner of the underlay SA is a critical issue.

As a method dealing with the interference to PUs, the cooperative communication (CC) can be applied into cognitive radio networks because it provides the effect of the transmission power reduction by employing relay nodes. Such a network in the paper is

termed as cooperative cognitive radio networks (CCRNs). The SUs in CCRNs can play a role of relay that carries the packets of neighbors. For the interference to PUs, the main problem in CCRNs is to devise a relay selection scheme that can meet quality of service requirements of a given application. Many of the literatures in CC have compared bit error rate (BER) or symbol error rate (SER) with closed-form solution according to the signal-to-noise ratio (SNR) in order to show the performance of relay selection schemes [4–7]. In, actual wireless communication networks, not only a single performance metric, such as (BER) or (SER), but a multilateral metric for quality of services (QoS) which may include time delay, energy efficiency, and actual interference should be considered. Thus, it is obvious that a relay selection scheme with consideration for a multilateral metric should be studied for CCRNs to cover the requirements of actual wireless communications. Furthermore, the consideration of a multilateral metric for QoS support in relay selection schemes must have even higher computational complexity due to contradictory difference among the requirements of QoS. It means the need of a relay selection scheme with low complexity for a multilateral metric.

Since source node in CC networks (CCNs) chooses its transmission mode according to the radio environment, actual relay selection can be said to be under transmission mode selection of the source node [8]. The transmission mode can be separated into the direct transmission, the mixture of the direct and the relay transmissions, and the only-relay transmission. The context of the transmission mode in CCNs becomes more complex in CCRNs due to the SA of the PUs. In this paper, therefore, an efficient transmission mode scheme based on the reinforcement learning for CCRNs is proposed to allow the SU source to effectively determine an optimal action for maximizing a given multilateral metric and to provide operations with low complexity through Q-learning that is a type of the reinforcement learning. As metrics for the QoS, it is considered the time delay, the energy efficiency, and the interference to PUs. In the proposed scheme, it is assumed that the channel state information (CSI) between the SU source and the PU receiver is not provided to the SUs. The channel gain between the SU source and the PU receiver is estimated by considering the well-known spectrum sensing threshold technique of CRNs, which is also a contribution point of this paper. The rest of this paper is organized as follows. Some of the "Related works" section with Q-learning are given as the detailed background and motivations of the paper. A "System model" section of CCRNs with single agent is presented for evaluate the performance. In order to demonstrate the performance of efficient transmission mode selection "Proposed scheme" section is briefly presented with state, action and relay selection, and reward. Simulation parameters of proposed scheme are discussed in "Simulation results" section. We compared our proposed scheme with other exist scheme which shows better performance in terms of the time delay, energy efficiency, and interference to PU receiver with low complexity. At last we concluded our works in "Conclusions" section.

Related works

To reduce the interference to PUs in CCRNs, a relay selection scheme plays an important role. Thus, several literatures are given in below to provide the background and recent researches of relay selection schemes. In [4] the authors propose a relay selection scheme based on instantaneous SNR either between the source and relay or the relay

and the destination. They evaluate the performance of their algorithm based on the perfect and the estimated CSI for both amplify-and forward (AF) and decode-and-forward (DF) protocols. In [5], the authors propose a joint relay selection (RS) with opportunistic source selection for an AF-based network in terms of outage probability and BER. Similarly in [6], the optimal RS scheme for maximizing the effective signal-interference-to-noise-ratio (SINR) is proposed, which significantly improves the system performance of full-duplex relaying. In [7] the authors propose a detect-and-forward (DetF) RS system and derive the average bit error probability by using the closed-form relay link SNR, and shows BER performance with SNR. Energy-efficient multi relay selection with power allocation strategy according to the total transmission power constraint is studied in [9] which achieve high energy efficiency performance in both low and high SNR region. However, all of the relay selection schemes from [4–7] only consider a single metric such as BER or SER. Since actual wireless systems aim to achieving a multilateral metric for QoS, the existing schemes could not support in practice.

Applying a given multilateral metric into CCRNs, it is obvious for relay selection schemes to be controlled by the transmission mode of the SU source. Practical issue in such a context is the complexity, which is rapidly increased due to in concinnity among the given metric. For this, Q-learning can be used as an efficient solution to lower the complexity with maximizing the total revenue of the given metric. The studies considering Q-learning for CRNs are as follows. Stochastic power adaptation with multi-agent Q-learning for cognitive wireless mesh networks is studied in [10] where the authors first extend single agent Q-learning to a multi-user context and propose a multi-agent Q-learning algorithm to achieve the optimal transmission strategies with the incomplete information of the environment. They propose a full learning-based algorithm for an intelligent SU which performs the Q-function updating based on the estimation over the other stochastic behaviors of the network environment. For dynamic channel selection (DCS) the authors in [11] apply Q-learning to select an available channel for data transmission, which denotes positive and negative values as a reward for successful and unsuccessful data transmission. Similarly in [12], for DCS the author denotes the reward as an amount of successful data transmission in CRNs. Cooperative channel sensing using Q-learning is studied in [13]. In [14], energy efficiency enhancement is studied which aims to reduce the total energy consumption in transmission. End-to-end delay, energy consumption, energy-efficient routing protocol are studied in [15], [16]. In [17], power consumption analysis of prominent time synchronization protocol for wireless sensor network is studied. Energy-efficient radio resource management for heterogeneous wireless networks and dynamic antenna shuffling scheme for MIMO wireless system are studied in [18, 19]. In cellular communication, sub-band selection method for cross- and co-tier interference in femtocell overlaid cellular network for evaluate the average cell throughput for macro and femtocell is studied in [20]. Spectrum sensing and data transmission in cognitive relay network with optimal power allocation strategy is studied in [21]. Performance evaluation of data aggregation in cluster based wireless sensor network, effective implementation of security based algorithmic approach in mobile ad-hoc network, and stochastic approach for dynamic power management in wireless sensor network are studied in [22–24]. However in cooperative communication an efficient relay selection scheme through self-learning is studied in [25] where the authors define

state, action, and reward to achieve a near optimum SER performance for relay selection. To get optimal SER performance, most of the literatures considered optimal power allocation in relay selection which requires a complicate optimization problem with higher complexity. Thus, Q learning algorithm can be a solution for such complexity and can provide a near optimal SER performance according to SNR with lower complexity.

System model

A CCRN considered in this paper is shown in Fig. 1, in which the SU denotes unlicensed user that opportunistically accesses the spectrum of the PUs under the required interference level, while the PU means licensed user of the spectrum. For simplicity, the system model consists of two PUs and N SUs.

It is assumed that the SU network is a time slot-based network. In the model, the SUs employ neighbors as relay in transmission to avoid the interference to the PUs as well as to maximize the total revenue of the multilateral metric comprised of the time delay, the energy efficiency, and the interference to the PUs. The SU relays carry the data packet of the neighboring SUs operating as the SU source to the SU destination. Since the relays are scattered throughout the network, it ought to be mandatory to select a best relay that maximize the given QoS requirements. It is assumed that the SU source is an agent of Q-learning in the networks and all the SUs have incomplete prior knowledge on the network environment, which means that the SUs just know the received signal strength of pilot signal from the PUs. Utilizing the pilot signal, the SUs estimate the channel gain and determine the interference level to the PUs.

In the case of the direct link, the ith SU source sends the packet to the kth SU destination during one time slot. In the mixture case of the direct and relay transmissions, at time slot T_1 the ith SU source forwards the data packet to the kth SU destination, and the jth SU relay receives the packet on the same time slot due to broadcast nature of communication and then finally forwards the received data packet to the kth SU destination on time slot T_2 in the manner of AF protocol.

In the case of the only-relay transmission, even though the protocol procedure is the same with the mixture case, it is different that the transmission target of the ith SU source is the jth SU relay, not the kth SU destination. Thus, the only-relay case also spends two time slots for data transmission. Figure 2 shows the summary of the whole case.

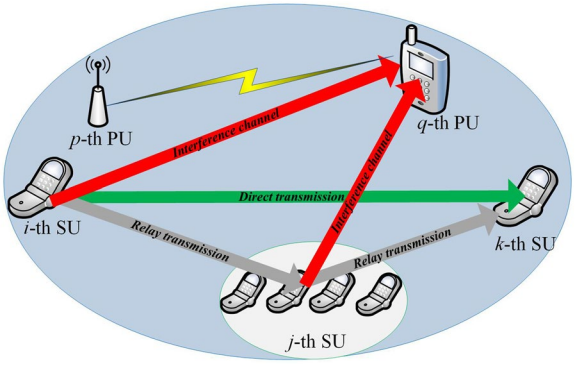

Fig. 1 System model

Since the SU source acts as an agent for Q-learning, it learns about state of the network environment at the beginning of packet transmission. The state is comprised of two elements: the direct transmission power between the SU source to the SU destination, and an estimated channel gain between the SU source and the PU receiver. Obtaining a state from the network environment, the SU source chooses one action in a given action set by referring to the Q-table. After executing the chosen action, the SU source assesses its revenue and updates the Q-table. Such a learning process is repeated until the Q-table is converged to a given level. And then finally the SU source performs actual operations with the learned Q-table in the network environment according to the state.

Proposed scheme

The proposed transmission mode selection scheme is based on Q-learning, thus in this section the proposed Q-learning scheme is presented by introducing the single agent Q-learning studied in [26]. The Q-learning model can be defined by the tuple of $\{S_i, A, R_{T,i}(S_i, A)\}$ which is shown in Fig. 3, where S_i, A, and $R_{T,i}(S_i, A)$ denotes the state of the ith SU, the action set, and the total reward function of the ith SU, respectively. The possible action for the agent on the environment can be expressed as $A = \{DT, DRT, RT, NA(no\ action)\}$. The goal of the agent is to choose an optimal action $a_i(t) \in A$ which maximizes the total reward function.

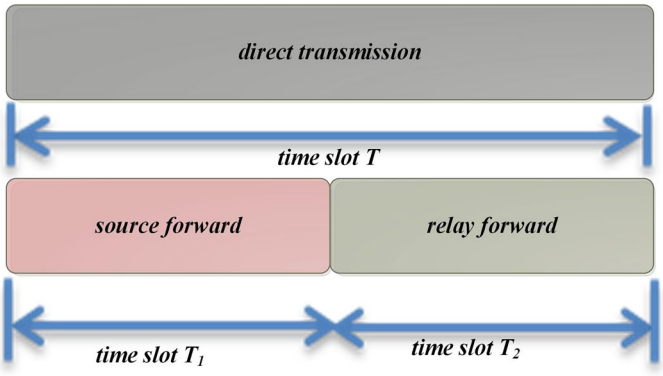

Fig. 2 Time frame structures for packet transmission in the considered CCRN

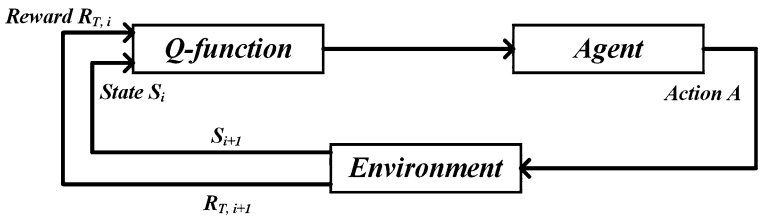

Fig. 3 Agent-environment interaction in Q-learning

The optimal action in Q-learning is determined by referring to Q-table which is constructed by the possible whole combination of the state and the action sets. After obtaining the value of the total reward function, the Q-table is updated as:

$$Q(S_i(t), a_i(t)) = (1-\alpha)Q(S_i(t), a_i(t)) + \alpha\left\{R_{T,i}(t) + \delta \max_a Q(S_i(t+1), a)\right\} \quad (1)$$

where α and δ is the learning rate and the discount factor, respectively. It was proven in [26, 27] that the Q-table converges to the optimal Q-value under certain conditions. In traditional Q-learning algorithms, there are two popular approaches to achieve the convergence to the optimal Q-value, which is associated with a trade-off between exploration and exploitation, namely, soft-max and ε-greedy [28]. The proposed Q-learning considers a random number ε at each step of the learning process to choose an action randomly and with $1-\varepsilon$ probability the action is selected according to the rule:

$$a_i(t) = arg \max_{a \in A} Q(S_i(t), a) \quad (2)$$

where $S_i(t)$ is the state of the ith SU source at time t and $a_i(t)$ means the selected action. The learning process is repeated until the given condition is satisfied as:

$$\left(\sum_{i=0}^{K_1}\sum_{j=0}^{K_2} Q_{t-1}(i,j)\right) - \left(\sum_{i=0}^{K_1}\sum_{j=0}^{K_2} Q_t(i,j)\right) \leq \eta \quad (3)$$

where η denotes a threshold for the convergence of Q-table.

State

Let γ_{DT} be the regulated SNR for direct transmission from the ith SU source to the kth SU destination, which can be expressed as follows:

$$\gamma_{DT} \leq \frac{P_{tx,ik}\left|h_{ik}\right|^2}{Z_{ik}} \quad (4)$$

where $P_{tx,ik}$ is the transmission power of the ith SU source to the kth SU destination, h_{ik} is the channel gain between the ith SU source to the kth SU destination, and Z_{ik} is an independent and identically distributed (IID) additive white Gaussian noise (AWGN). From the Eq. (4) the minimum decodable transmission power with given SNR is calculated in [29] as follows:

$$P_{tx,ik} = \gamma_{DT}\frac{Z_{ik}}{\left|h_{ik}\right|^2} \quad (5)$$

When a maximum allowable transmission power for SUs P_{tx}^{max} is given, the maximum transmission power has (K_1+1) feasible power intervals as follows:

$$P_n = \frac{n}{K_1}\left(P_{tx}^{max}\right), \quad n \in \{0, 1, \ldots, K_1\} \quad (6)$$

where $K_1 > 0$ is an integer. Thus, the actual transmission power of the ith SU source to the kth SU destination will be determined by:

$$\bar{P}_{tx,ik} = \begin{cases} P_n, & if\ P_{n-1} < P_{tx,ik} \leq P_n \\ P_{tx}^{max}, & otherwise \end{cases} \tag{7}$$

When a SU transmits during the transmission of the PU transmitter is active, the interference may occur on the PU receiver. Hence, the SINR level required by the PUs must be satisfied, which is expressed as:

$$SINR_{PU} = \frac{P_{tx,pq}\left|h_{pq}\right|^2}{I_{iq} + Z_{pq}} \tag{8}$$

where $P_{tx,pq}$ is the transmission power of the pth PU transmitter to the qth PU receiver, h_{pq} is the channel gain between the pth PU transmitter and the qth PU receiver, I_{iq} represents the interference to the qth PU receiver by the ith SU source, and Z_{pq} is an IID AWGN. The interference to the qth PU receiver by the transmission of the ith SU source can be expressed as:

$$I_{iq} = \frac{\bar{P}_{tx,ik}\left|h_{iq}\right|^2}{Z_{iq}} \tag{9}$$

where h_{iq} is the channel gain between the ith SU source to the qth PU receiver and Z_{iq} is an IID AWGN respectively. The interference to the qth PU receiver during transmission of the jth SU relay to the kth SU destination can be expressed as:

$$I_{jq} = \frac{\bar{P}_{tx,jk}\left|h_{jq}\right|^2}{Z_{jq}} \tag{10}$$

where h_{jq} is the channel gain between the jth SU relay to the qth PU receiver and Z_{jq} is an IID AWGN respectively.

Since it is assumed that the SUs can identify the pilot signals from the PUs, a reflector for $\left|h_{iq}\right|^2$ and $\left|h_{jq}\right|^2$ can be derived by considering the energy detector concept in CRNs. With a given reference threshold $\lambda_{th,i}$ and $\lambda_{th,j}$ of energy detector of the ith SU source as well as the jth SU relay, the reflector for the ith SU source to the qth PU receiver $\left|h_{iq}\right|^2$ and the jth SU relay to the qth PU receiver $\left|h_{jq}\right|^2$ can be defined as follows:

$$\Psi_i = \left|\left(\frac{\log \lambda_{th,i}}{\log \lambda_{iq}}\right)\right| \tag{11}$$

$$\Psi_j = \left|\left(\frac{\log \lambda_{th,j}}{\log \lambda_{jq}}\right)\right| \tag{12}$$

where λ_{iq} and λ_{jq} is the RSS value at the ith SU from the pilot signal of the qth PU receiver and at the jth SU relay from the qth PU receiver respectively. If the value of the given reference threshold and the value of the RSS are assumed as $\lambda_{th,i}, \lambda_{iq}, \lambda_{th,j}$, *and* $\lambda_{jq} < 1$,

then the estimated interference to the qth PU receiver by the ith SU source and the jth SU relay can be expressed as:

$$I_{iq} \approx \tilde{I}_{iq} = \bar{P}_{tx,ik} min(1, \Psi_i) \tag{13}$$

$$I_{jq} \approx \tilde{I}_{jq} = \bar{P}_{tx,jk} min\left(1, \Psi_j\right) \tag{14}$$

By applying the Eq. (13) into Shannon channel capacity form, the interference capacity level of the ith SU source is expressed by:

$$L_i = \frac{log_2\left(1 + \tilde{I}_{iq}\right)}{Y_0} \tag{15}$$

where Y_0 represents a normalizing factor. Since the maximum interference capacity level becomes 1 due to the normalizing factor Y_0, the interference capacity level can be expressed into $(K_2 + 1)$ intervals as follows:

$$l_n = \frac{n}{K_2} max(L_i) = \frac{n}{K_2}, n \in \{0, 1, \ldots, K_2\} \tag{16}$$

where $K_2 > 0$ is an integer. Thus, the actual interference capacity level of the ith SU source will be determined by:

$$\bar{L}_i = \begin{cases} l_n, & if\ l_{n-1} < L_i \leq l_n \\ 1, & otherwise \end{cases} \tag{17}$$

The state in Q-learning reflects the situation of the network environment, which is constructed by two elements in this paper: the direct transmission power from the ith SU source to the kth SU destination $\bar{P}_{tx,ik}$ and the interference capacity level of the ith SU source $\bar{L}_i$. Therefore, the state of the ith SU source for Q-learning is expressed as:

$$S_i = \left(\bar{P}_{tx,ik}, \bar{L}_i\right) \in \sum = P \times \Lambda, P = \left\{P_0, P_1, \ldots, P_{K_1}\right\}, \Lambda = \left\{l_0, l_1, \ldots, l_{K_2}\right\} \tag{18}$$

Action and relay selection

After obtaining the state, the ith SU source will take one action among the action set given as:

$$A = \{DT, DRT, RT, NA\} \tag{19}$$

where $DT, DRT, RT, and\ NA$ means the direct transmission, the mixture of the direct and the relay transmissions, the only-relay transmission, and no action, respectively. When, the destination may be located far away from the source and the direct transmission is impossible and when there are no available relays for cooperation, the ith SU source will keep silent. To consider these cases, no action (NA) is taken into account as an action set in this paper. In Q-learning, taking an action $a_i \in A$ of the ith SU source to a given state $S_i \in \sum$ is based on Q-table expressed as $\sum$ by A matrix, and the decision rule of the action is expressed as the Eq. (2). If DRT and RT actions are taken by the ith SU source, the relay selection is carried out as:

$$j^* = arg \max_{j \in \Omega} \{min(C_{ij}, C_{jk})\}$$
$$\text{subject to : } min(1, \Psi_i) > \Psi_j \tag{20}$$

where C_{ij} and C_{jk} is the channel capacity between the ith SU source and the jth SU relay, and the jth SU relay and the kth SU destination, respectively, Ω is the number of relays available for cooperation between the ith SU source and the kth SU destination, and Ψ_j is the reflector for $|h_{jq}|^2$ the channel gain between the jth SU relay to qth PU receiver which is calculated in the Eq. (12).

Reward

The SU source executes the selected action, which includes the relay selection step, and then the reward is evaluated as follows. Let T and τ_i be the maximum time delay and the time spent for transmission of the ith SU source, respectively. Then the time delay reward of the ith SU source in data transmission can be formulated as:

$$R_{D,i} = \frac{T - \tau_i}{T}$$
$$\text{subject to : } E(\tau_i)(1 + \sigma_{\tau_i}) \leq T \tag{21}$$

where $E(\tau_i)$ and σ_{τ_i} is the mean and the variance of τ_i, respectively.

The energy efficiency can be represented by the ratio of the transmission capacity to the total consumed energy for transmission. Thus, the energy efficiency of the ith SU source is evaluated in [30] as follows:

$$R_{EE,i} = Y_1 \frac{C_s}{P_{tx,i}} \tag{22}$$

where C_s and $P_{tx,i}$ denotes the transmission capacity of the system and the amount of the consumed energy of the ith SU source for the transmission, respectively, and Y_1 is the normalizing factor. The transmission capacity C_s can be calculated by Shannon channel capacity equation for both the direct and the relay transmissions. For the direct transmission, the transmission capacity can be expressed as:

$$C_s = log_2\left(1 + \frac{P_{tx,i}}{N_0}|h_{ik}|^2\right) \tag{23}$$

For the relay transmission, the transmission capacity can be expressed as:

$$C_i = min\left\{log_2\left(1 + \frac{P_{s,i}}{N_0}|h_{ij}|^2\right), log_2\left(1 + \frac{P_{j,k}}{N_0}|h_{jk}|^2\right)\right\} \tag{24}$$

where h_{ij} and h_{jk} is the channel gain between the ith SU source and the jth SU relay, and the jth SU relay and the kth SU destination, respectively. The interference by the transmission of the ith SU source is considered as penalty in the total reward and can be calculated as:

$$\Phi_{I,i} = Y_2 \, P_{tx,i} \min(1, \Psi_i) \tag{25}$$

where Y_2 is the normalizing factor. Consequently, the total reward of the ith SU source is calculated by:

$$R_{T,i} = R_{D,i} + R_{EE,i} - w_i \Phi_{I,i} \tag{26}$$

where w_i is a weighting parameter to meet the required interference level to PU receiver. Basically, w_i has a positive value, but a negative value only consider for the case of *NA*. Note that in Q-learning, the total reward has to be designed to have the most maximum value when the best action matched to a given state is selected.

Simulation results

In this section, the simulation result is presented in order to show the performance of the proposed transmission mode scheme based on Q-learning. Throughout the simulations, the main system parameters are used as follows. The frequency for large scale propagation model is 700 MHz, average path loss exponent $\varphi = 4$, discount rate $\delta = 0.5$, learning rate $\alpha = 0.5$, and $\in = 0.5$ for the $\in$-strategy, and weighting parameter $w_i = 2$. The whole simulation area is a 1 by 1 km square.

The considered CCRN consists of a pair of the PUs, one SU source, one SU destination, and twelve SU relays. It is assumed that the distance between the SU source and the SU destination is near to the maximum transmission range of SUs. For the simulation comparison, four different schemes, denoted as *Random RS*, *EECC*, *Shortest Path*, and *Knowlege-based*, are used and the proposed scheme is denoted as *Proposed*. *Random RS* is a random relay selection scheme where the SU source selects one of the surround relays randomly. *EECC* designates the energy efficient cooperative communication scheme in [9]. *Shortest Path* represents a delay-sensitive relay selection scheme where the SU source selects the relay making the shortest path to the destination. *Knowlege-based* is a relay selection scheme where the SU source has the whole information of the network environment.

Figure 4 shows the performance of the number of slots spent for data transmission according to average SNR of SU relays between the SU source and the SU destination,

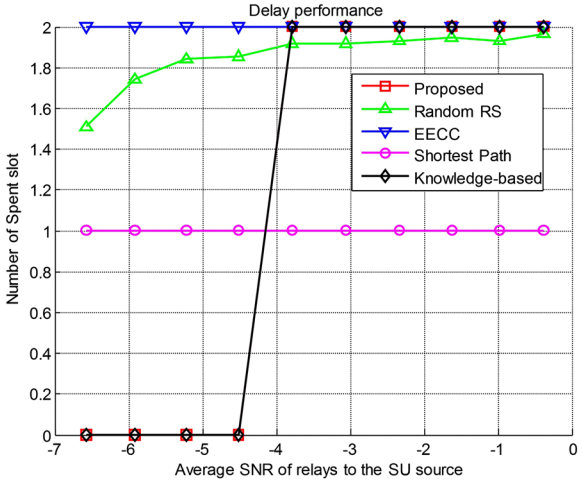

Fig. 4 Performance of time delay according to average SNR of SU relays

which means the performance of time delay. It is observed that the performance of *Proposed* is the same with that of *Knowledge-based*. They have zero time delay up to −4.5 dB in the given wireless environment, which represents there is no transmission due to the regulation of the interference to the PU receiver. Through this, it can be found that *Proposed* is well learned by Q-learning. In low region of the average SNR of SU relays, the probability that *Random RS* chooses the direct transmission is higher than the probability for the relay transmission. Hence, *Random RS* has lower time delay performance, compared to high region of the average SNR of SU relays. Since the purpose of *EECC* is to maximize the energy efficiency, *EECC* always employs a SU relay if there exists any SU relay available. Thus, *EECC* has two time slots for all the cases. *Shortest path* shows one time slot performance due to its high delay sensitivity.

Figure 5 shows the performance of the energy efficiency as average SNR of SU relays increases. EE of all the schemes increases with the average SNR of SU relays, except *Shortest Path*. *Proposed* and *Knowledge-based* has zero EE up to −4.5 dB due to the identical reason explained in Fig. 4. *Proposed* follows EE of *EECC* after −4.5 dB. The performance gap between *Proposed* and *Knowledge-based* after −4.0 dB comes from the fact that *Knowledge-based* seeks to maximize the transmission capacity. EE of *Random RS* increases according to the average SNR of SU relays. *EECC* has the best EE performance among the schemes on account of its operational goal. It is clear that *Shortest Path* has a very poor EE when the distance between the SU source and the SU destination is large, which is the reason why *Shortest Path* has EE value near to zero.

Figure 6 shows the interference performance as the average SNR of SU relays increases. The SU transmission in CCRNs may incur the interference to the PU receiver. Hence, it is mandatory to keep the interference level bellow required by the PUs. The required interference level depicted in the figure is the value when the SINR of the PUs is assumed as −5 dB. The trend of the interference performance for *Proposed* is identical to that of the *EECC*.

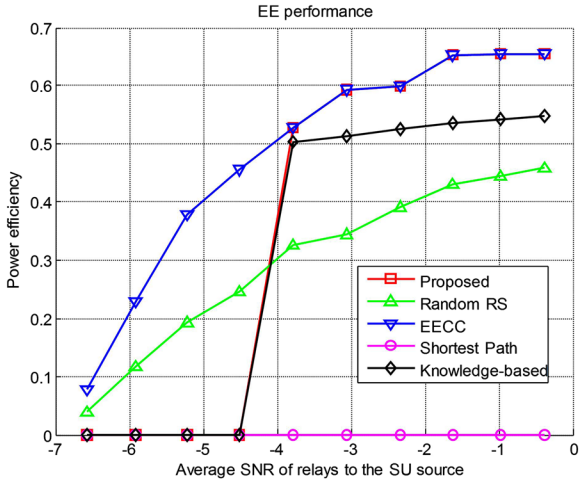

Fig. 5 Performance of energy efficiency according to average SNR of SU relays

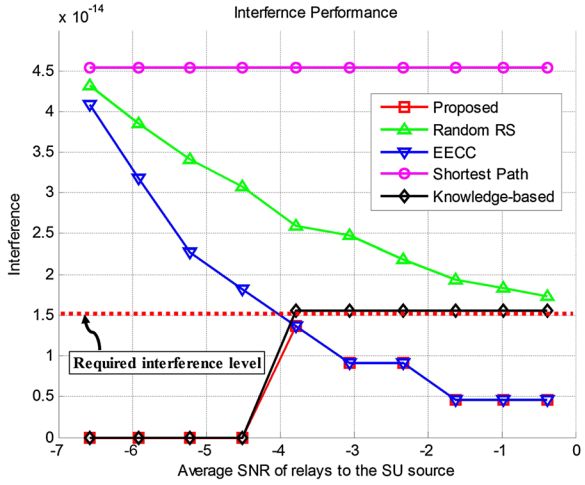

Fig. 6 Performance of interference to PU receiver according to the average SNR of SU relays

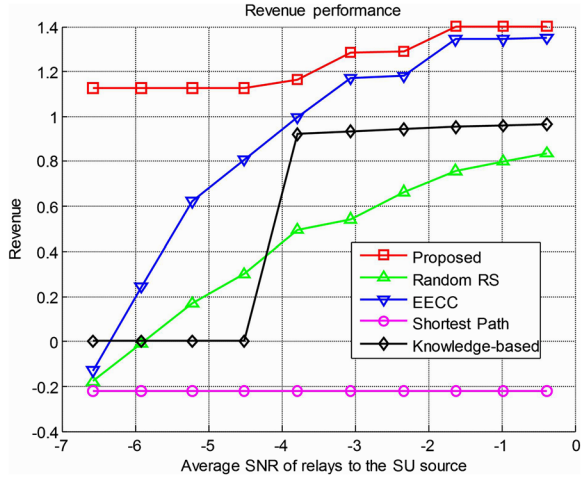

Fig. 7 Performance of the total revenue according to average SNR of SU relays

The reason for the performance gap between *Proposed* and *Knowledge-based* is also the same. It is observed that only *Proposed* and *Knowledge-based* keep the given interference level for low and high SNR region.

Figure 7 shows the performance of the total revenue of the time delay, the energy efficiency, and the interference, according to average SNR of SU relays between the SU source and the SU destination. It is found that *Proposed* outperforms the other schemes for all cases. Note that *Proposed* has a higher positive value up to −4.5 dB even though there is no transmission. The reason is that w_i in the Eq. (26) for the case of *NA* has a negative value, as mentioned above. *Shortest path* has negative revenues, because it is the delay sensitive transmission scheme and then the penalty from choosing the direct transmission is large.

Conclusions

The main purpose of CCRNs is to achieve not only maximizing QoS of SUs but also minimizing the interference to PUs. It, however, may require a complicated system approach because QoS requirements make up a multilateral metric and there exist contradictory system objectives. Unfortunately, the efficiency of its sophisticated implementation could not be ensured. For such context, Q-learning algorithm can be a good help for making the complicated system approach much simple because of virtue of learning algorithm.

In this paper, an efficient transmission mode scheme based on Q-learning for CCRNs has been proposed to effectively support the multilateral metric constructed by the time delay, the energy efficiency, and the interference to the PU receiver under the condition in which the network information is not provided. It has been demonstrated through the simulation result that the proposed scheme can have better performance than the representative schemes considering a single metric. As further works, the system throughput maximization based on generated load and the multi-agent Q-learning for SU relay needs to be studied in order to realize more efficient CCRNs.

Abbreviations

CCRNs: cooperative cognitive radio networks; PUs: primary users; SUs: secondary users; CR: cognitive radio; CRNs: cognitive radio networks; SA: spectrum access; CC: cooperative communication; QoS: quality of service; BER: bit error rate; SER: symbol error rate; SNR: signal to noise ratio; CCNs: cooperative communication networks; CSI: channel state information; AF: amplify and forward; DF: decode and forward; DCS: dynamic channel selection; RSS: receive signal strength; IID: independent and identically distributed; AWGN: additive white Gaussian noise; EE: energy efficiency.

Authors' contributions

IK provided the guideline to focus on issues, requiring solutions, and reviewed the overall manuscript. MAR and YDL conceived the study, drafting the article, revising it critically for intellectual content of the whole manuscript. They reviewed the technical contribution of the work and approved the final. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the KRF funded by the MEST (NRF-2013R1A2A2A05004535) and the Ministry of Education (2013R1A1A2063779).

Competing interests

The authors declare that they have no competing interests.

Received: 25 March 2015 Accepted: 6 February 2016
Published online: 05 April 2016

References

1. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23(2):201–220
2. Goldsmith A, Jafar SA, Maric´ I, Srinivasa S (2009) Breaking spectrum gridlock with cognitive radios: an information theoretic prospective. Proc IEEE 97(5):894–914
3. Gastpar M (2007) On capacity under receive and spatial spectrum-sharing constraints. IEEE Trans Inf Theory 53(2):471–487
4. El-Mahdy A, Tarek N (2013) Relay selection algorithm for wireless cooperative network. Conference on signal processing: algorithms, architectures, arrangements, and applications (SPA). IEEE, Poznan, pp 274–278
5. Ju M, Kim IM (2010) Joint Relay Selection and opportunistic source selection in bidirectional cooperative diversity networks. IEEE Trans Veh Technol 59(6):2885–2897
6. Cui H, Ma M, Song L (2014) Relay selection for two-way full duplex relay networks with amplify-and-forward protocol. IEEE Trans Wireless Commun 13(7):3768–3777
7. Yuan J, Li Y, Chu L (2010) Differential modulation and relay selection with detect-and-forward cooperative relaying. IEEE Trans Veh Technol 59(1):261–268
8. Wu D, Zhu G, Sun L, Zhao D (2012) Joint mode/route selection and power allocation in cellular networks with cooperative relay. In: IEEE international conference on communication, Ottawa, IEEE, pp 4144–4149. doi: 10.1109/ICC.2012.6363807

9. Xu Y, Bai Z, Wang B, Gong P, Kwak K (2014) Energy-efficient power allocation scheme for multi-relay cooperative communication. In: 16th international conference on advance communication technology (ICACT), GIRI, Korea (South)
10. Chen X, Zhao Z, Zhang H (2013) Stochastic power adaptation with multiagent reinforcement learning for cognitive wireless mesh network. IEEE Trans Mob Comput 12(11):2155–2166
11. Tang Y, Grace D, Clarke T, Wei J (2011) Multichannel non persistent CSMA MAC schemes with reinforcement learning for cognitive radio networks. In: 11th international symposium on communications and information technologies (ISCIT), IEEE, Hangzhou, pp 502–506. doi: 10.1109/ISCIT.2011.6092159
12. Li H (2009) Multi-agent Q-learning of channel selection in multiuser cognitive radio systems: a two by two case. In: international conference on systems, man and cybernetics (CSMC), USA, IEEE. pp 1893–1898. doi: 10.1109/ICSMC.2009.5346172
13. Lo BF, Akyildiz IF (2010) Reinforcement learning-based cooperative sensing in cognitive radio ad hoc networks. In: 21st international symposium on personal indoor and mobile radio communication (PIMRC). Turkey, IEEE, pp. 2244–2249. doi: 10.1109/PIMRC.2010.5671686
14. Zheng K, Li H (2010) Achieving energy efficiency via drowsy transmission in cognitive radio. In: 2010 global telecommunications conference (GLOBECOM). USA, IEEE. pp 1–6, doi: 10.1109/GLOCOM.2010.5683355
15. Peng J, Li J, Li S, Li J (2011) Multi-relay cooperative mechanism with Q-learning in cognitive radio multimedia sensor networks. In: international joint conference on IEEE TrustCom, China, IEEE. pp 1624–1629. doi: 10.1109/TrustCom.2011.225
16. Yoon M, Kim YK, Chang JW (2013) An energy-efficient routing protocol using message success rate in wireless sensor network. J Converg 4(1):15–22
17. Bae SK (2014) Power consumption analysis of prominent time synchronization protocols for wireless sensor networks. J Inf Processing Syst 10(2):300–313
18. Carvalho GH, Woungang I, Anpalagan A, Dhurandher SK (2012) Energy-efficient radio resource management scheme for heterogeneous wireless networks: a queueing theory perspective. J Converg 3(4):15–22
19. Chong JH, Ng CK, Noordin NK, Ali BM (2013) Dynamic transmit antenna shuffling scheme for mimo wireless communication systems. J Converg 4(1):7–14
20. Kwon YM, Choo H, Lee TJ, Chung MY, Kim M (2014) Femtocell subband selection method for managing cross- and co-tier interference in a femtocell overlaid cellular network. J Inf Processing Syst 10(3):384–394
21. Tishita TA, Akhter S, Islam MI, Amin MR (2014) Spectrum sensing and data transmission in a cognitive relay network considering spatial false alarms. J Inf Processing Syst 10(3):459–470
22. Sinha A, Lobiyal DK (2013) Performance evaluation of data aggregation for cluster-based wireless sensor network. Human-centric Comput Inf Sci 3(13). doi:10.1186/2192-1962-3-13
23. Singh R, Singh P, Duhan M (2014) An effective implementation of security based algorithmic approach in mobile Adhoc networks. Human -centric Comput Inf Sci 4(7). doi:10.1186/s13673-014-0007-9
24. Pughat A, Sharma V (2015) A review on stochastic approach for dynamic power management in wireless sensor network. Human-centric Comput Inf Sci 5(4). doi:10.1186/s13673-015-0021-6
25. Jung H, Kim K, Kim J, Shin OS, Shin Y (2012) A relay selection scheme using Q-learning algorithm in cooperative wireless communications. In: 18th Asia-Pacific conference on communications (APCC), Jeju Island, KICS. pp 7–11. doi: 10.1109/APCC.2012.6388091
26. Watkins CJCH, Dayan P (1992) Technical note Q-learning. J Mach Learn 8(3-4):279–292
27. F.S. Melo, Convergence of Q-learning: A simple proof, Institute of Systems and Robotics, Tech Rep
28. Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT Press, Cambridge
29. Sadek AK, Wei Y, Liu KJR (2006) When does cooperation have better performance in sensor network?. In: 3rd annual IEEE communication society on sensor and adhoc communications and networks (SECON), Reston, IEEE. pp 188–197. doi: 10.1109/SAHCN.2006.288423
30. Miao G, Himayat N, Li YG, Bormann D (2008) Energy efficient design in wireless OFDMA, Communications. In: IEEE international conference on communications, (ICC), Beijing, IEEE. pp 3307–3312. doi: 10.1109/ICC.2008.622