



ENERGY EFFICIENT CHANNEL SELECTION FRAMEWORK FOR COGNITIVE RADIO WIRELESS SENSOR NETWORKS

Joshua Abolarinwa, Nurul Mu'azzah Abdul Latiff, Sharifah Kamilah Syed Yusof and Norsheila Fisal

Faculty of Electrical Engineering, MIMOS-UTM Center of Excellence, Universiti Teknologi Malaysia, Johor, Malaysia

E-Mail: jaabolarinwa2@live.utm.my

ABSTRACT

Advancements in the field of cognitive radio technology have paved way for cognitive radio-based wireless sensor networks. Energy and spectrum efficiencies are two biggest challenges facing wireless sensor networks. This has impacted immensely on the network lifetime and performance. On the other hand, spectrum channel is scarce and limited. Hence, there is urgent need for energy efficient utilization of the scarce spectrum in cognitive radio wireless sensor networks. In this paper, we propose a flexible solution by reinforcement learning to address the problem of energy efficiency associated with channel selection in cognitive radio wireless sensor networks. A simple learning algorithm was developed to improve the secondary user throughput, channel availability in relation to the sensing time and energy efficiency. Comparing the results obtained from simulations with other non-learning channel selection methods-random channel assignment and dynamic channel assignment, the proposed learning algorithm produced up to 30% better performance in terms of throughput and energy efficiency. This signifies that, for better performance, intelligent learning is required in cognitive radio wireless sensor networks.

Keywords: intelligent, cognitive-radio, energy-efficiency, reinforcement-learning, throughput.

INTRODUCTION

The demand for radio spectrum has been on the increase due to continued evolution of different wireless applications. The existing wireless networks are characterized by static spectrum channel allocation, in which channels are assigned to a licensed user, otherwise known as primary user (PU) on a long-term basis. Whereas, some of the licensed spectrum remained spatial-time idle under the present static spectrum policy. This leads to inefficient utilization of large portion of the wireless spectrum.

Cognitive radio (CR) is a promising paradigm approach to spectrum utilization with increased quality of service (QoS) for wireless sensor networks. It is the key technology for dynamic spectrum access (DSA). It provides the capability to share the wireless spectrum with licensed users in order to improve spectrum efficiency and network performance by adaptively organizing channel access by different users according to the radio environment characteristics.

In CR, interference avoidance to the PU signal is of paramount importance to the by the other user trying to access the channel opportunistically. This other user is known as the secondary user (SU). It then becomes imperative for the SU to detect the presence or absence of the PU signal through channel sensing. The right of the PU to the channel must be protected by the SU while it maintains its own quality of service requirements. In order to address these questions posed to the SU cognitive radio wireless sensor network (CRWSN), we propose an intelligent reinforcement learning (RL) channel selection algorithm framework for CRWSN. In this RL channel selection approach, each CRWSN node learns and

dynamically decides when to sense, handoff or transmit data within a channel. For learning and decision accuracy, cooperation among inter-cluster CRWSN nodes is proposed. The problem is to learn a way of controlling the system so as to maximize the long term reward of energy and spectrum efficiencies. The learning problems differ in the details of how the data is collected and how performance is measured.

There exists some works in literature that proposed different channel access schemes other than RL for CRWSN. Their approach is either multi-radio or multi-channel based. In [1], the authors proposed two new channel selection strategies for the SU in order to improve channel utilization efficiency in cognitive radio network (CRN). The proposed solution tries its best to reduce collision and switching probabilities of the SU during data transmission. The authors in [2], presented recent developments and open research challenges in spectrum management based on CRN. While the authors focus on the development of CRN, the work did not offer solutions to the open issues raised. Reinforcement-learning-based double auction (RL-DA) algorithm for dynamic spectrum access in CRN was proposed by the authors in [3]. 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. However, with this approach, the PU does not have exclusive right to the channel as is expected of a typical cognitive radio network. In order to maintain channel's exclusive right of the PU, authors in [4] proposed distributed framework for testing and developing MAC protocols for cognitive radio sensor networks. The



framework is based on distributed algorithms and simulates nodes equipped with multi-channel radios. Even though the framework increases the delivery rate compared with conventional WSNs, it also introduced increase in latency and energy consumption, which is unsatisfactory to CRWSN.

Selection of appropriate channel for the SU in accordance with the application quality of service requirements and spectrum quality is the main focus of the work in [5]. The authors proposed automatic distributed spectrum decision (ADSD) using PU arrival probability to decide the quality of the available channel. The study in [6] the authors proposed an energy-aware SU selection algorithm for cognitive sensor networks. An optimization problem was solved to obtain the minimum required number of cognitive users, whereas satisfying the system requirements. The work in [7] proposed a pricing-based collision-resistant dynamic spectrum allocation to optimize overall spectrum efficiency. This approach was developed to avoid collision between PU and SU networks. As a multi-stage dynamic game approach, the problem of energy efficiency was not considered as the approach only focused on spectrum utilization.

Authors in [8] proposed a novel biologically inspired consensus-based cooperative spectrum sensing scheme for cognitive radio mobile ad hoc networks (CR-MANETs). Decision making in this scheme is without the use of common receiver for data aggregation. This is only applicable to distributed network. Centralized cluster-based CRWSN cannot adopt this scheme. To solve the problem of interference [9] proposed a systematic framework to produce conflict graphs based on physical interference characteristics. This framework optimizes conflict graph in order to produce spectrum allocations that match those derived from physical interference model. In [10] the authors considered CR system with one SU accessing multiple channels using periodic sensing and channel switching. Optimal spectrum sensing and access mechanism was proposed with focus on energy minimization in the presence of multi-constraints-sensing reliability, throughput and delay in the SU transmission. To minimize interference to PU and to maximize channel utilization [11] proposed a framework of constrained Markov decision processed, and optimal access policy based on linear program. Similar to previous work in [6], authors in [12] considered decentralized spectrum access using the theory of multivariate global game. Channel access is based on Bayesian estimate of the intention of the other SUs and based on expected throughput of the SU under consideration. Authors in [13] applied reinforcement learning to develop a new routing algorithm based on continuous link model. In the work, the idea of Q-learning was transferred into link-value. From the results obtained, it was shown that the reinforcement learning based algorithm gave improved performance in terms of link table and packet delivery ratio in ad hoc networks compared to AODV and DSR.

However, in this paper, we propose intelligent learning and decision making algorithm with RL backbone to address some of the problems associated with channel selection in CRWSN. The intelligence part of this approach ensures that the CR agent can learn from its previous experience and decide on an action that results in long-term expected energy efficient reward within the radio environment.

The rest of this paper is organized as follows; section II describes the system model and operation. Section III explains the fundamentals of RL and describes the proposed RL approach. Section IV discusses simulations and result analysis. Finally, we conclude with section V.

NETWORK MODEL DESIGN

Cognitive radio wireless sensor networks are sensor networks that employ radio cognition to adaptively and dynamically use the available communication channel. For the purpose of energy analysis, a cluster-based network topology is considered.

Cognitive radio wireless sensor network topology

For this cluster-based topology, there is a cluster head (CH) communicating with several other surrounding cluster members (MN) which are assumed non-mobile. This is shown in Figure-1. The CH is a full function device with cognitive radio capabilities within. It is the central controller for the MN within a cluster. The CH coordinates communication activities within a particular cluster, and aggregates data transmitted from MN to the sink.

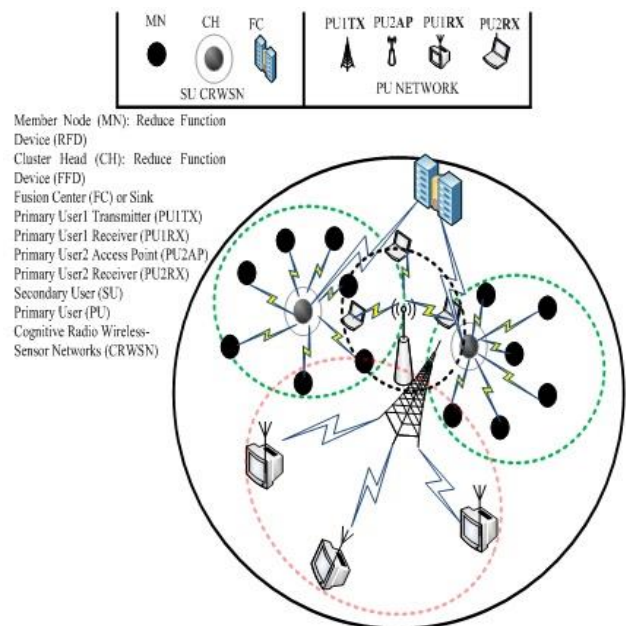


Figure-1. Cluster network topology.



Primary user channel behaviour

The PU channel behaviour and channel availability is modelled after continuous-time, probability process of channel utilization using two-state Markov decision process. As illustrated in Figure-2, PU activity in any given channel is not time-slotted because PU is the proprietary licensed user. PU switches between ON and OFF states, which corresponds to channel busy or available states. This behaviour represents the birth-death continuous time Markov chain (CTMC). Assuming, for each channel, the channel utilization by the PU is independent, identically distributed (i.i.d.) random variables exponentially distributed with constant ON and OFF mean time we define α and β as the channel transition probabilities of transiting from ON to OFF and OFF to ON respectively by the PU. The α and the β represents the birth and death rates respectively.

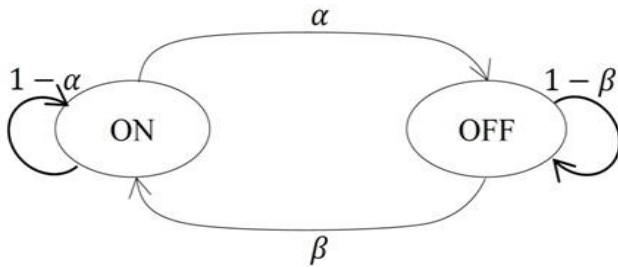


Figure-2. Two state Markov decision model of PU channel behavior.

Figure-3 shows the timing operation of the SU operation in a given channel. This is time-slotted operation that depends on the activity of the non-timed slotted PU. Each frame is divided into two time slots viz; (1) sensing slot and (2) transmission slot. Sensing operation is carried out periodically for duration t_1 . During the transmission slot, the CR transmit packet for duration t_2 . The probability that a given channel $s_k \in C$ is free (P_{OFF}) and the probability that the channel is occupied (P_{ON}) are determined as in (1) and (2) respectively;

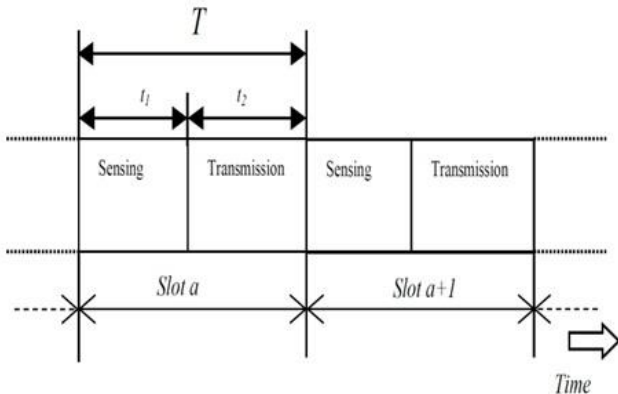


Figure-3. Slot durations of SU channel usage.

$$P_{OFF} = \frac{\alpha}{\alpha + \beta} \quad (1)$$

$$P_{ON} = \frac{\beta}{\alpha + \beta} \quad (2)$$

Probability of *channel idle* and probability of *channel busy* are depicted by (1) and (2) respectively. For C number of channels, the conditional probability that all the channels sensed busy by the SU in the next sensing slot given that they are currently busy is given as;

$$P(\beta | 1 - \alpha) = [(1 - \alpha)]^C \quad (3)$$

The conditional probability that at least one of the channels will be in idle state for the SU to use provided they are all in busy state previously is given as;

$$1 - P(\beta | 1 - \alpha) = 1 - [(1 - \alpha)]^C \quad (4)$$

Assuming energy detection spectrum sensing method, the probability of detection P_d and probability of false alarm P_f are given in (5) and (6) respectively.

$$P_d(t_1) = \left(\frac{\alpha}{\alpha + \beta} \right) Q \left(\frac{\lambda - 2t_1 B (\sigma_s^2 + \sigma_n^2)}{\sqrt{4t_1 B (\sigma_s^4 + \sigma_n^4)}} \right) \quad (5)$$

$$P_f(t_1) = \left(\frac{\alpha}{\alpha + \beta} \right) Q \left(\frac{\lambda - 2t_1 B \sigma_n^2}{\sqrt{4t_1 B \sigma_n^4}} \right) \quad (6)$$

Where, λ is the decision threshold value. The noise and PU signal variance are σ_n^2 and σ_s^2 respectively, B is the bandwidth, t_1 is the channel sensing time of the SU, Q is the Q -function.

REINFORCEMENT LEARNING AND MARKOV DECISION PROCESS

Reinforcement learning (RL) is an offshoot of dynamic programming (DP), and it is used to solve the Markov decision process (MDP) problems by having to construct the theoretical model. As a learning problem, it refers to learning to control a system so as to maximize some numerical value which represents a long-term objective function. Figure 4 shows a typical reinforcement learning operation. A controller receives the controlled system's state and a reward associated with the last state transition. It then calculates an action which is sent back to the system. In response, the system makes a transition to a new state and the cycle is repeated.

An MDP is defined as 4-tuple (S, A, R, P) where (S) is the set of states, (A) is the set of possible actions in a given state, (R) is the reward function for taken an action in a given state, and (P) is the transition probability function. The decision policy that maps the state set to the action set is $\pi: S \rightarrow A$. For finite state set $S =$



$s_1, s_2, \dots, s_k, \dots, C$ there are possible corresponding action set $A = a_1, a_2, \dots, a_k, \dots, A$. At a given decision instant k , the RL agent takes an action a_k from the set A while in a corresponding state s_k from the set S which leads to a new state s_{k+1} . The transition probability $P_{k \rightarrow k+1}$ provides a feedback reward function $r_k(s, a)$ for the RL agent. As mentioned earlier, this is an iterative process that intends to maximize the discounted reward or state value given as;

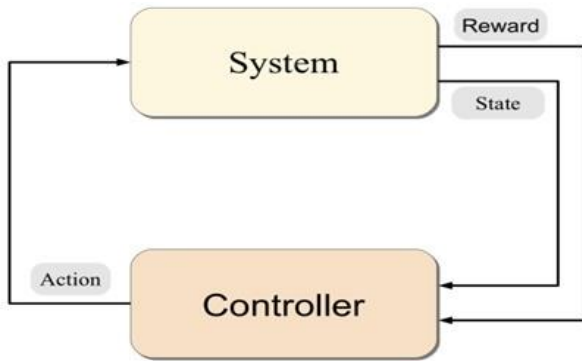


Figure-4. Reinforcement learning scenario.

$$V_k(\pi) = \max_{a \in A(k)} \left[r(k, a) + \lambda \sum_{k+1}^{|S|} P(k, a, k+1) V_{k+1} \right] \quad (7)$$

The expected immediate reward earned in state k when action a_k is selected is given as;

$$r(k, a) = \sum_{(k+1)=1}^{|S|} P(k, a, (k+1)) r(k, a, (k+1)) \quad (8)$$

For practical applications, the transition probability P and the reward function $R(s, \pi(s), a)$ are not known, which makes it difficult to evaluate the policy π . Q-learning variant of RL is one of the most effective and popular algorithms for learning from delayed reinforcement to determine an optimal policy in the absence the transition probability and reward function. In Q-learning, policies and the value function are represented by a two-dimensional lookup table of state-action pair. Formally, for each state s and action a , we define the Q value under policy π as;

$$Q(k, a) = \sum_{(k+1)=1}^{|S|} P(k, a, (k+1)) [r(k, a, (k+1)) + \lambda_d(7)] \quad (9)$$

The discount factor is λ_d . The iterative process of Q learning approximates the Q value function to optimal value. Hence, the updating rule for Q value is given as;

$$Q_k(a, s) = \begin{cases} Q_k(a, s) + \sigma \delta & \text{if } s_k = s, a_k = a \\ Q_k(a, s) & \text{otherwise} \end{cases} \quad (10)$$

Learning rate and temporal difference are σ and δ respectively.

The goal of the RL approach is to determine the

optimal state-action pair (policy- π) which maximizes the long-term anticipated reward. In this sense, throughput and energy efficiency maximizations are two important goals of our RL approach as applied to the CR aided sensor networks. It is undesirable to minimize energy consumption at the expense of unacceptable throughput. Mapping the general idea of state-action pair to CR aided sensor network, we formulate the learning approach as;

- Set of state S represents the set of communication channels $[S = s_1, s_2, \dots, s_k, \dots, C]$
- Set of actions A represents the three possible actions, sensing (a_s), transmit (a_t) and channel handoff (a_{hf}) $[A = a_s, a_t, a_{hf}]$.

The reward function R is the transmission capability of the SU in each channel.

State-action selection plan

We used softmax action selection strategy to determine the decision and access strategy to use. In this learning policy, the probability of selecting an action a_k in state s_k is given by;

$$P(s, a) = \pi(s, a) = \frac{\exp Q(s, a) / \tau}{\sum_1^{n_a} \exp Q(s, a) / \tau} \quad (11)$$

where $Q(s, a)$ is the state-action value function, n_a is the number of possible actions. In our case, $n_a = 3$, these are, a_s, a_t, a_{hf} . τ is the temperature parameter which controls the expected reward for the probability of a given action taken. When the value of τ is high, all the 3 actions are equally probable. On the other hand, for small τ , the action with maximum $Q(s, a)$ is selected. Hence, the state-action policy $\pi(s, a)$ can be rewritten as;

$$\pi(s, a) = \frac{\exp Q(s, a) / \tau}{\exp Q(s, a_s) + \exp Q(s, a_t) + \exp Q(s, a_{hf})} \quad (12)$$

Reinforcement learning algorithm

In our learning algorithm, at the end of each slot, the RL agent calculates the temporal difference, updates the Q-value and selects the next action in a predictive manner according to the learning policy $\pi(s, a)$. The RL agent decides which action to take at any decision time step with priority given to the action that yields the highest Q value. This is a form of greedy decision method in which the CR agent chooses an action randomly. This is done to maintain balance between exploitation of presumed optimal state-action pair and exploration of a new policy modification.

The learning algorithm is described below. The CR agent chooses one of the three actions possible based on the current value of $\pi(s, a)$. For a_s it updates the channel availability. For a_t it updates the long term reward for accessing a channel at a given time. The algorithm shows how the learning scheme dynamically adjusts the actual policy in (12). When the PU activity is absent in a



given channel, the CR agent stop sensing action, and access the channel to begin transmission within the transmission slot. As soon as the PU activities resume in the channel, the CR agent handoff the channel and begin sensing for another available channel. The delay due to sensing decreases the state-value of a channel, thus increases the probability to choose a different channel. The choice of another channel is dependent on which available channel has the least access delay and higher energy efficiency based on accumulated experience during past operations.

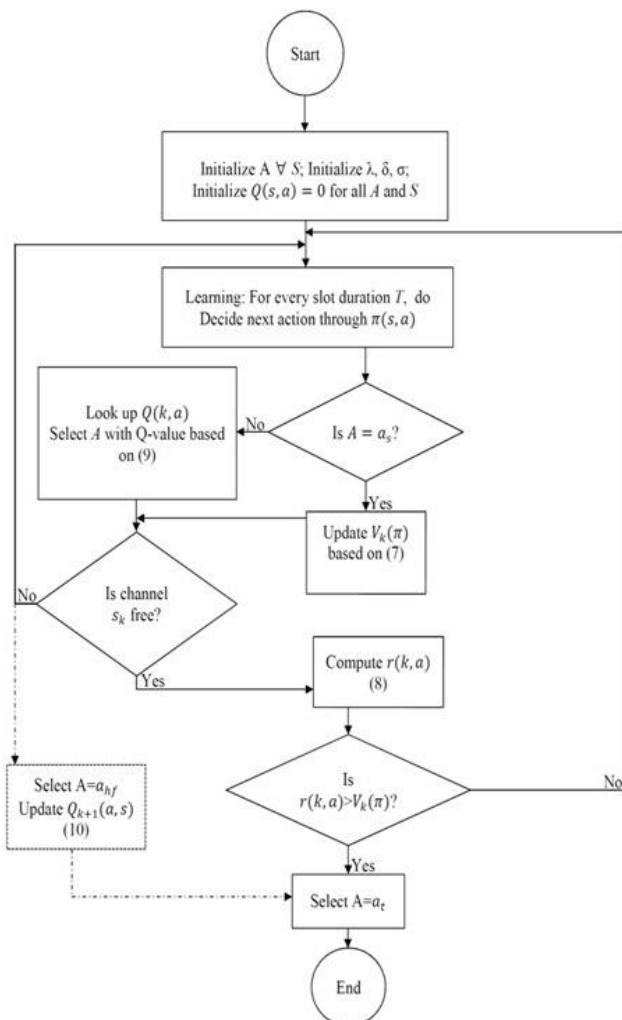


Figure-5. Flow diagram of the algorithm.

SIMULATION AND PERFORMANCE EVALUATION

In order to evaluate the performance of the proposed algorithm, we run simulations using CR-integrated MATLAB platform. The CRWSN scenario considered is set up using single hop communication consisting of 1 CH with several MN in each cluster. We simulate the CR environment using 2 PUs and 3 clusters of SU CRWSN. There are 3 CHs and 9 MN to form a SU

network composed of a total of 12 sensor nodes. We assume binary orthogonal frequency shift keying (FSK) modulation on a frequency non-selective Rayleigh fading channel. Based on the standard Friis transmission formula, the output power for SU and PU is held at -20dB and -39dB respectively and the path-loss factor is 0.125. Average distance between two SU nodes, and between SU node and PU, respectively is fixed at 8m and 120m. Other simulation parameters are shown in Table-1. Parameter values are set depending on the specific application and vary according to the operations pattern of PU on different spectrum bands. In addition, features of CR and attributes of traditional WSN were considered in making assumptions to realize a reliable CRWSN.

Performance comparison we made between the proposed scheme and the non-RL schemes such as-random channel assignment (RCA) scheme, and the dynamic channel assignment (DCA) scheme. In RCA, each cognitive radio CH chooses randomly the next action, that is, *transmit*, *sense* or *handoff*, at the end of each transmission slot. In RCA, no learning is involved. In the case of DCA, each CH of the CR network performs channel sensing during the sensing slot. When a PU is detected, the CH switches its cluster member nodes to the available channel in its neighbourhood. The DCA scheme constitutes a classical approach for spectrum management over cognitive radio sensor network.

Table-1. Simulation parameters.

Parameters	Description	Value
W	Channel Sampling frequency	1MHz
T	Slot duration	2.75s
λ_d	Discounting factor	0.5
λ	Detection threshold	3
α	Channel ON duration	0.08
β	Channel OFF duration	0.16
P_{ON}	Probability of busy channel	0.3
P_{OFF}	Probability of free channel	0.7
σ	Learning rate	0.6
σ_n^2	Noise variance	1
σ_s^2	PU signal variance	1
R	Data transmission rate	200Kbps

Figures-6, 7, 8 shows the performance comparison of the three schemes using different performance metrics. In Figure-6, the throughput and transmission duration of the SU network are measured. For a given available channel, the longer the transmission time, the higher the throughput. However, the random



approach suffers higher packet loss due to interference coming from random channel handoff. In the case of DCA, effect of PU interference is less, but experiences less optimal spectrum selection because it does not take into cognizance the quality of the channel unlike when channel selection is based on intelligent learning. In Figure-7, the average channel availability effect on the energy efficiency of transmission by the CRWSN is shown. As the channel availability increases, there is a corresponding increase in the energy efficiency in RCA, DCA and RL approaches. However, in the case of RCA, the energy efficiency experiences a downward trend at a certain peak. This downward trend is due to energy consumption as a result of continuous channel handoff and channel sensing. Energy efficiency in the case of RL approach is higher than in DCA approach. This is due to the impact of intelligent learning experience of the SU on the channel quality and status. The energy efficiency is calculated based on previous works in [14] and [15].

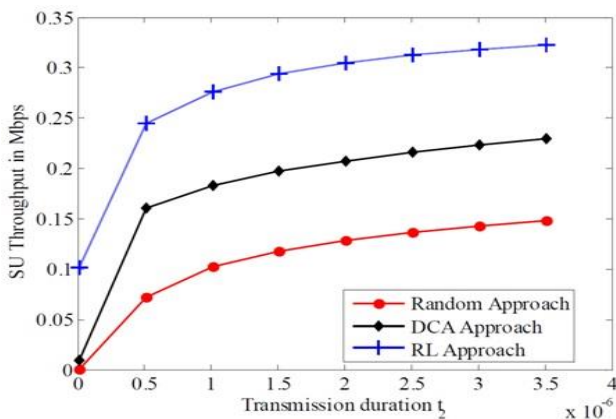


Figure-6. Effect of transmission time on throughput.

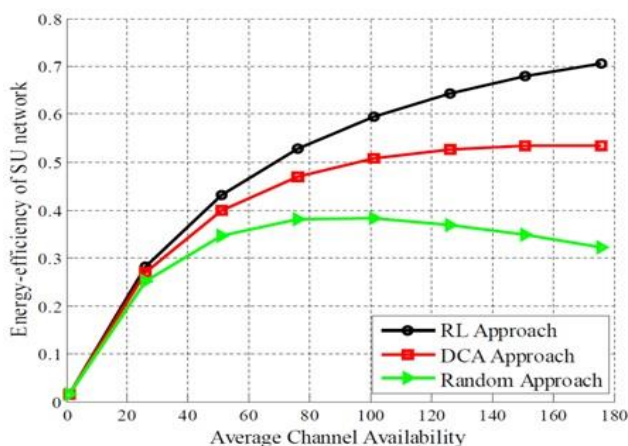


Figure-7. Channel availability effect on energy efficiency.

In the case of rate of free channel access variation with energy efficiency, it can be seen from Figure-8 that as

the rate of channel access increase, the energy efficiency increases. But, this performance metric as shown in Figure-8 emphasize that the RL learning approach performs better than DCA and RCA.

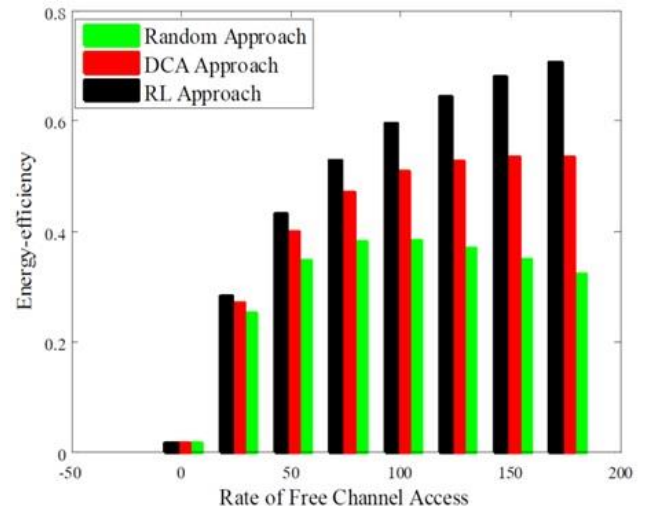


Figure-8. Channel free access rate effect on energy efficiency.

CONCLUSIONS

In this paper, we proposed a channel selection algorithm based on RL to optimize channel utilization experience of the SU CRWSN, RL based channel access demonstrates the best performance in terms of energy efficiency, throughput, and channel availability among other approaches in relation to sensing and transmission duration. We equally showed that the proposed RL based algorithm is able to converge to the optimal solution and adapting to the change in radio environment better than DCA and RCA. This is made possible by the intelligent learning capability incorporated in the algorithm. This work is extended to consider the SU activities without neglecting the PU channel behaviour. Simulation results revealed significant energy efficiency and throughput performance improvement up to 30% increment compared to non-learning channel access approaches.

ACKNOWLEDGEMENTS

The authors wish to thank the Ministry of Education (MOE) Malaysia and Research Management Center (RMC) of Universiti Teknologi Malaysia (UTM) for supporting this project under Research Grant no. Q.J130000.2523.08H80 and the authors also thank Federal University of Technology Minna, Nigeria for the study fellowship provided.



REFERENCES

- [1] X. Tan, H. Zhang, Q. Chen and J. Hu, 2014. Opportunistic channel selection based on time series prediction in cognitive radio networks. *Transactions on Emerging Telecommunications Technologies*, Vol. 25, pp. 1126-1136.
- [2] I. F. Akyildiz, L. Won-Yeol, M. C. Vuran, and S. Mohanty. 2008. A survey on spectrum management in cognitive radio networks. *IEEE Communications Magazine*. Vol. 46, pp. 40-48.
- [3] Y. Teng, F. R. Yu, K. Han, Y. Wei and Y. Zhang. 2013. Reinforcement-Learning-Based Double Auction Design for Dynamic Spectrum Access in Cognitive Radio Networks. *Wireless Personal Communications*. Vol. 69, pp. 771-791.
- [4] L. H. Correia, E. E. Oliveira, D. F. Macedo, P. M. Moura, A. Loureiro and J. S. Silva. 2012. A framework for cognitive radio wireless sensor networks. *IEEE Symposium on Computers and Communications*. pp. 611-616.
- [5] G. Yuming, S. Yi, L. Shan and E. Dutkiewicz. 2009. ADSD: An Automatic Distributed Spectrum Decision method in Cognitive Radio networks. *First International Conference on Future Information Networks*. pp. 253-258.
- [6] A. S. Zahmati, X. Fernando and A. Grami. 2014. Energy-aware secondary user selection in cognitive sensor networks. *IET Wireless Sensor Systems*. Vol. 4, pp. 86-96.
- [7] Z. Ji and K. R. Liu, 2008. Multi-stage pricing game for collusion-resistant dynamic spectrum allocation. *IEEE Journal on Selected Areas in Communications*. 26, pp. 182-191.
- [8] F. R. Yu, M. Huang, and H. Tang 2010. Biologically inspired consensus-based spectrum sensing in mobile ad hoc networks with cognitive radios," *IEEE Network*. 24, pp. 26-30.
- [9] L. Yang, L. Cao, and H. Zheng 2008. Physical interference driven dynamic spectrum management. *3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 1-12.
- [10] S. Wang, W. Yue, J. P. Coon, and A. Doufexi 2012. Energy-Efficient Spectrum Sensing and Access for Cognitive Radio Networks. *IEEE Transactions on Vehicular Technology*. 61, pp. 906-912.
- [11] Z. Qianchuan, S. Geirhofer, T. Lang, and B. M. Sadler. 2008. Opportunistic Spectrum Access via Periodic Channel Sensing. *IEEE Transactions on Signal Processing* Vol. 56, pp. 785-796.
- [12] V. Krishnamurthy 2009. Decentralized spectrum access amongst cognitive radios—An interacting multivariate global game-theoretic approach. *IEEE Transactions on Signal Processing*. Vol. 57, pp. 3999-4013.
- [13] Y. Fu and Q. Liu, 2013. Research of QoS Routing Algorithm in Ad Hoc Networks based on Reinforcement Learning. *Electronics and Electrical Engineering*. Vol. 19, pp. 65-68.
- [14] J. Abolarinwa, N. A. Latiff and S. S. Yusof, Energy Constrained Packet Size Optimization For Cluster-based Cognitive Radio-based Wireless Sensor Networks. *Australian Journal of Basic and Applied Sciences*. Vol. 7, pp. 138-150, 2013.
- [15] J. A. Abolarinwa, N. M. Abdul Latiff, S. K. Syed Yusof and N. Fisal. 2014. Learning-based algorithm for energy-efficient channel decision in cognitive radio-based wireless sensor networks. *International Conference on Frontiers of Communications, Networks and Applications*. pp. 1-6.