Channel Decision in Cognitive Radio Enabled Sensor Networks: A Reinforcement Learning Approach

Joshua Abolarinwa[#1], Nurul Mu'azzah Abdul Latiff[#2],
Sharifah Kamilah Syed Yusof[#3], Norsheila Fisal[#4]

Faculty of Electrical Engineering, MIMOS-UTM Center of Excellence,
Universiti Teknologi Malaysia, 81310 Johor, Malaysia
[1] jaabolarinwa2@live.utm.my
[2] nurulmuazzah@utm.my
[3] kamilah@fke.utm.my
[4] sheila@fke.utm.my

Abstract— Recent advancements in the field of cognitive radio technology have paved way for cognitive radio-based wireless sensor networks. This has been tipped to be the next generation sensor. Spectrum sensing and energy efficient channel access are two important operations in this network. In this paper, we propose the use of machine learning and decision making capability of reinforcement learning to address the problem of energy efficiency associated with channel access in cognitive radio aided sensor networks. A simple learning algorithm was developed to improve network parameters such as secondary user throughput, channel availability in relation to the sensing time. Comparing the results obtained from simulations with other channel access without intelligent learning such as random channel assignment and dynamic channel assignment, the learning algorithm produced better performance in terms of throughput, energy efficiency and other quality of service requirement of the network application.

Keyword- Channel, Decision, Energy, Reinforcement, Learning

I. INTRODUCTION

In recent time, 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 the static spectrum allocation, in which the wireless spectrum is assigned to a licensed user on a long-term basis. Whereas, some of the licensed spectra remained spatial-time idle under the present static spectrum policy. This leads to underutilization of large portion of the wireless communication spectrum.

Cognitive radio (CR) has emerged as a key enabling 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.

In CR, the fundamental requirement is to avoid interference to the signal of the licensed user, otherwise called the primary user (PU) by the other user trying to access the channel opportunistically. This other user is otherwise known as 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. In order to preserve the exclusive right of the PU, the SU is faced with the challenges of when to sense a channel, how long should the sensing operation take, when to switch from the channel and when to transmit using a given available channel.

In order to address these questions posed to the SU CR aided sensor network, we propose reinforcement learning (RL) channel access approach. In this RL channel decision and access protocol, each CR aided sensor node learn and adaptively decide when to sense, handoff or transmit data within a channel. For the purpose of learning and decision accuracy within the network, cooperation among inter-cluster CR enabled nodes are proposed. The problem is to learn a way of controlling the system so as to maximize the total reward. The learning problems differ in the details of how the data is collected and how performance is measured.

Few works abound in literature that proposed different channel access schemes other than RL in CR aided sensor networks. Their approach is either multi-radio or multi-channel based. Example of such is the configurable medium access control protocol. In [1], the authors presented recent developments and open research challenges in spectrum management based on CR networks. While the authors focus on the development of CR networks, the work did not address any specific solution to the issues raised. Reinforcement-learning-based double auction (RL-DA) algorithm for dynamic spectrum access in cognitive radio networks was proposed by the authors in [2]. 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 [3] 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 introduces an increase in latency and energy consumption, which is unsatisfactory to CR-aided sensor network.

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 [4]. The authors proposed automatic distributed spectrum decision (ADSD) using PU arrival probability to decide the quality of the available channel. In their work, [5] proposed a pricing-based collision-resistant dynamic spectrum allocation to optimize overall spectrum efficiency. This approach was developed to combat 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 [6] 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 CR-aided sensor network cannot adapt this scheme. Bothered by interference problem, [7] 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 [8] 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 [9] proposed a framework of constrained Markov decision processed, and optimal access policy based on linear program. Similar to previous work in [6], authors in [10] 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 [11] 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 have proposed the use of machine learning and decision making capability of reinforcement learning to address some of the problems associated with channel access in cognitive radio-aided sensor networks. This approach consists of intelligent learning algorithm developed to improve network parameters such as throughput in relation to the sensing time and interference within the network. 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 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.

II. System Model and Operation

Cognitive radio aided sensor networks are sensor networks that employ radio cognition to adaptively and dynamically use the available communication channel. In this section, we consider a cluster sensor network with star network topology.

A. Cognitive Radio Enabled Sensor Network Topology

In each cluster, there is one cluster head (CH) communicating with several other surrounding cluster members (MN) which are assumed non-mobile. This is illustrated in Figure 1. The CH carries with it the cognitive radio capabilities within its cluster. The CH acts as 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 node.

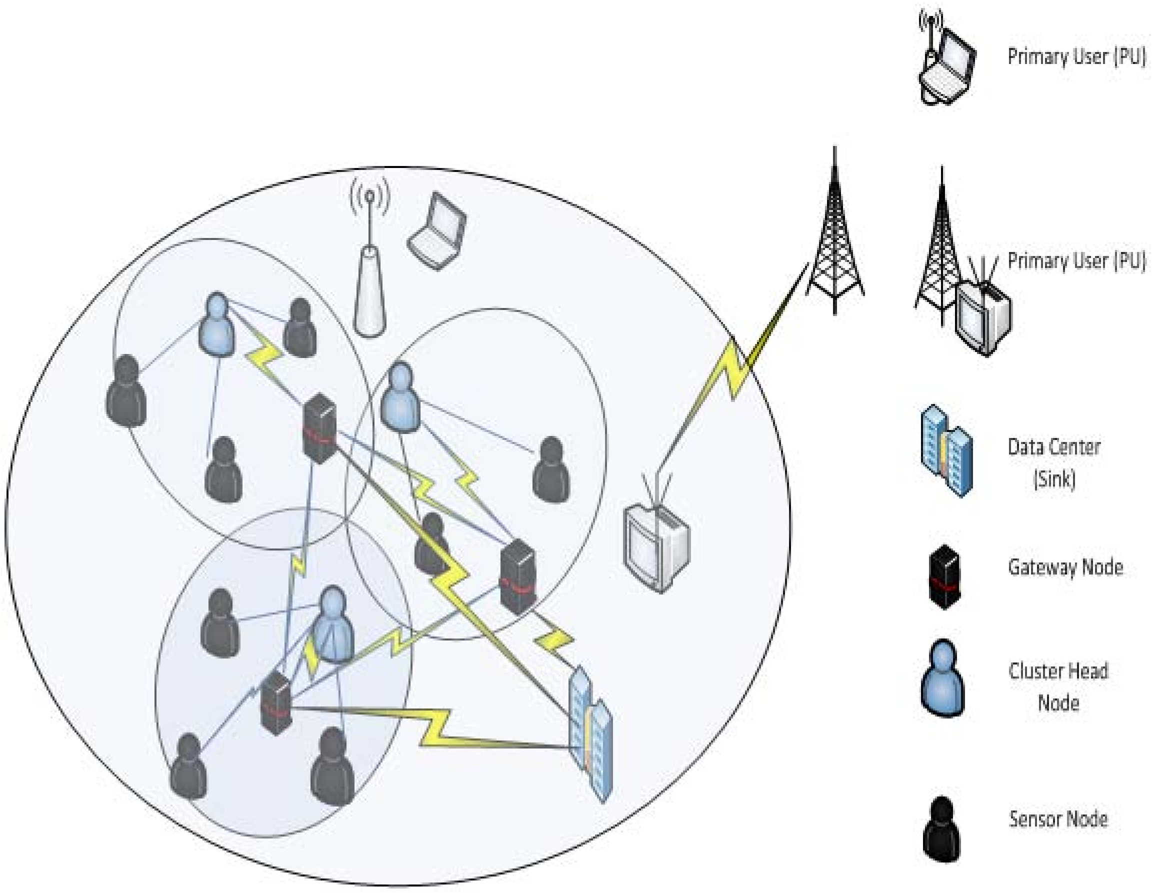

Fig. 1. Cluster Network Topology [12]

B. *Primary User Channel Activity and Channel Opportunity*

The PU channel behaviour and channel availability is modelled after continuous-time, stochastic process of channel utilization using two-state Markov decision process. This is shown 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). We assume that, 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. Let α and β be the channel transition probabilities of moving from ON to OFF and OFF to ON respectively by the PU. The α and the β represents the birth and death rates respectively.

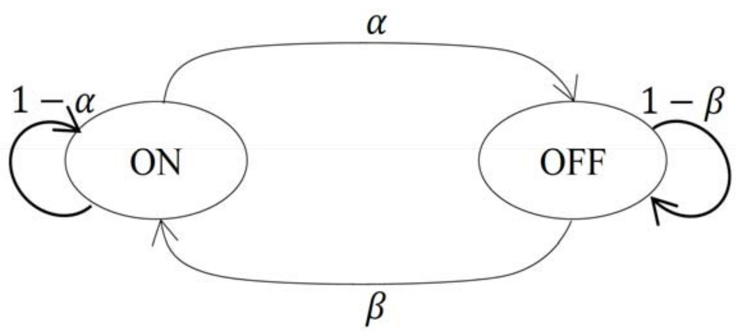

Fig. 2. Two state Markov decision model of PU channel behaviour [12]

Figure 3 shows the channel operation opportunity for the SU. 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 durationt_1. During the transmission slot, the CR transmit packet for durationt_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;

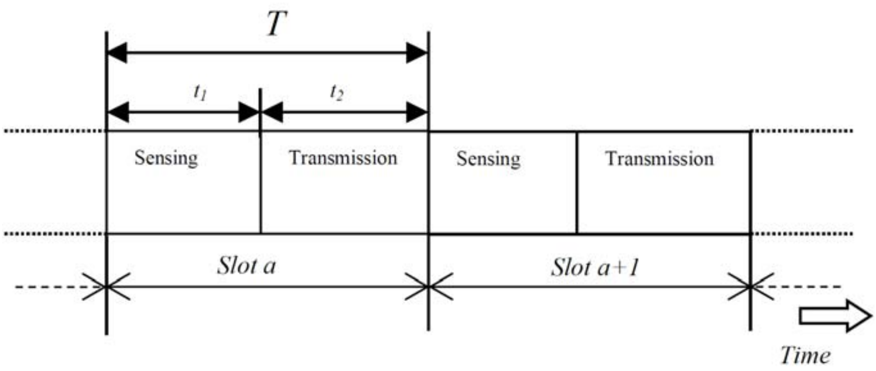

Fig. 3. Slot durations of SU channel usage [12]

$$P_{OFF} = \frac{\alpha}{\alpha+\beta} \tag{1}$$

$$P_{ON} = \frac{\beta}{\alpha+\beta} \tag{2}$$

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;

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

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;

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

We assume energy detection spectrum sensing. The probability of detection P_d and probability of false alarm P_f are given in (5) and (6) respectively.

$$P_d(t_1) = \left(\frac{\alpha}{\alpha+\beta}\right) Q\left(\frac{\lambda - 2t_1B(\sigma_s^2+\sigma_n^2)}{\sqrt{4t_1B(\sigma_s^4+\sigma_n^4)}}\right) \tag{5}$$

$$P_f(t_1) = \left(\frac{\alpha}{\alpha+\beta}\right) Q\left(\frac{\lambda - 2t_1B\sigma_n^2}{\sqrt{4t_1B\sigma_n^4}}\right) \tag{6}$$

λ is the decision threshold value. σ_n^2 and σ_s^2 are noise and PU signal variances, B is the bandwidth, t_1 is the channel sensing time of the SU, Q is the $Q-function$. Accurate probability of detection and high spectrum sensing efficiency in terms of near zero misdetection probability are two important aims of spectrum sensing operation.

III. REINFORCEMENT LEARNING AND MARKOV DECISION PROCESS

Reinforcement learning (RL) refers to both a learning problem and a subfield of machine learning. It is an offshoot of dynamic programming (DP), and it is used to solve the Markov decision process (MDP) with 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. A typical scenario where reinforcement learning operates is shown in Figure 4. 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. It is an iterative process.

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. $\pi: S \rightarrow A$ is the decision policy that maps the state set to the action set. For a finite state set of $S = s_1, s_2, \ldots. s_k \ldots.. C$ a corresponding action set $A = a_1, a_2, \ldots. a_k \ldots.. A$. 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;

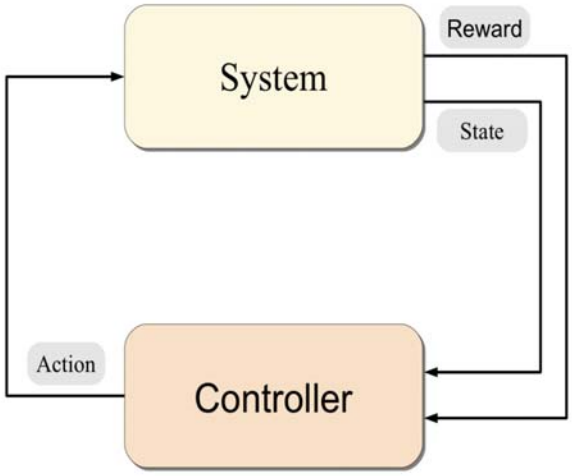

Fig. 4. Reinforcement Learning Scenario

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

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

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

In many practical situations, 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;

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

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

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

σ and δ are learning rate and temporal difference 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;

i. Set of *state S* represents the set of communication channels $[S = s_1, s_2, \dots s_k \dots C]$

ii. 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.

C. 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;

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

D. 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.

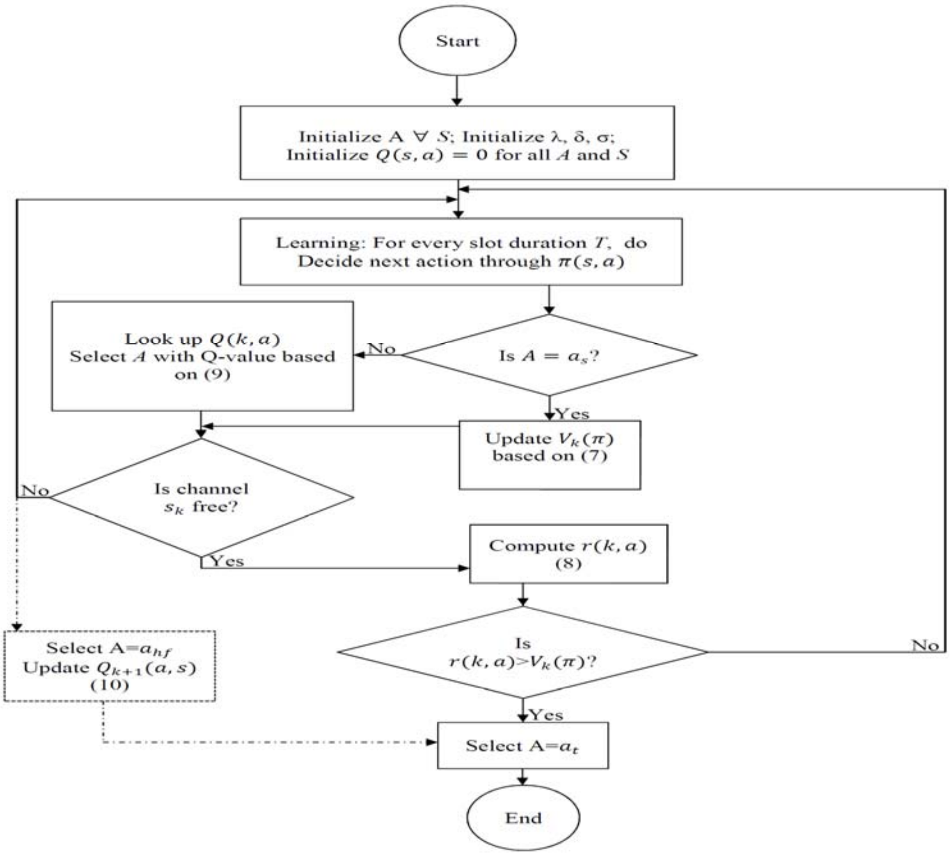

Fig. 5. Flow diagram of the Algorithm

IV. SIMULATION AND PERFORMANCE EVALUATION

In our attempt to evaluate the performance of the proposed algorithm, simulations were carried out using CR integrated MATLAB platform. The CR aided sensor network scenario considered is made up of clustered star topology, single hop communication network consisting of one cluster head and several member nodes in each cluster. We simulate the cognitive radio environment using 2 PUs and 3 clusters of SU network. 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 are 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 determined by numerical analysis are summarized in table I. 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 CR aided sensor network.

We compare the performance of the proposed scheme with 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 I. 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, 9, 10 and 11 shows the performance comparison of the three schemes using different performance metrics. Figure 6. shows the throughput and transmission duration of the SU network. When a channel is available for packet transmission, the longer the transmission time, the higher the throughput. The random approach suffers higher packet loss due to interference resulting from random handoff action. In the case of DCA approach, 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 in the case where channel selection is based on learning. Figure 7. shows the average channel availability effect on the energy efficiency of transmission by the CR sensor network. As the channel availability increases, there is a corresponding growth in the energy efficiency in RCA, DCA and RL.

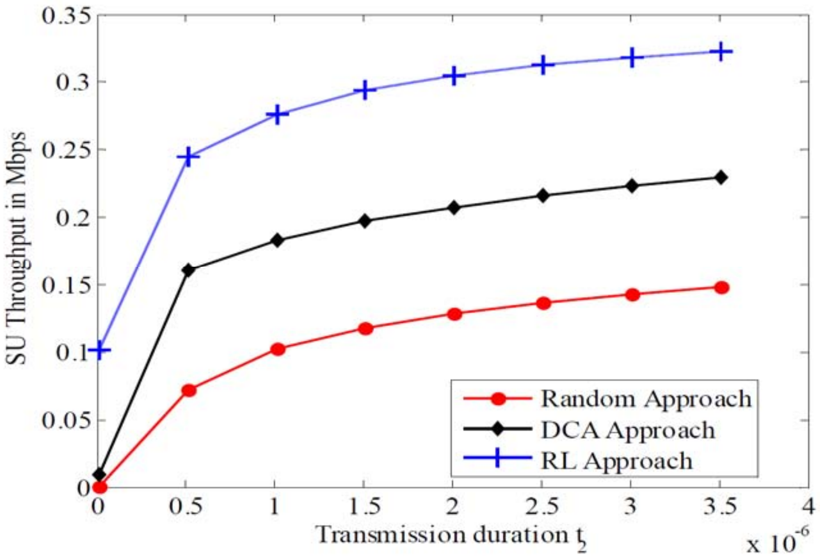

Fig. 6. Effect of Transmission time on throughput

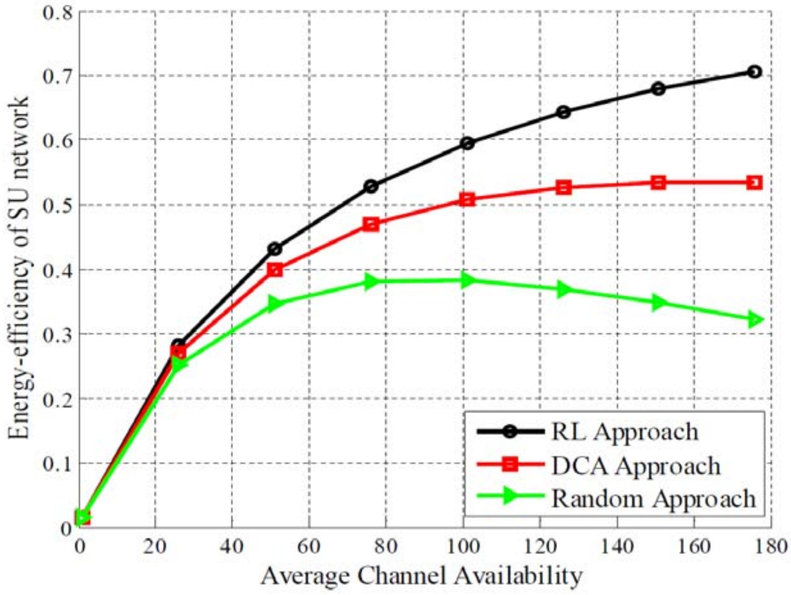

Fig. 7. Channel availability effect on energy efficiency

However, at a certain peak in the case of RCA, the energy efficiency experiences a downward trend. 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 faired higher than in DCA approach. This is due to the impact of the learning experience of the SU on the channel quality and status.

To further investigate the performance gains, Figure 8. shows the channel availability as a function of the sensing time. Figure 8 analyse the impact of sensing time on the possibility of a particular channel availability. With increase in sensing time of the SU for a given channel, this increase the chance of PU network activity within the channel. PU activity is not time-slotted, hence the shorter the sensing time, the higher the channel availability. RL approach performs better than the other DCA and RCA in that order because under RL approach, the SU decides which action to take in a given channel based on the state value and accumulated reward of received by the RL agent.

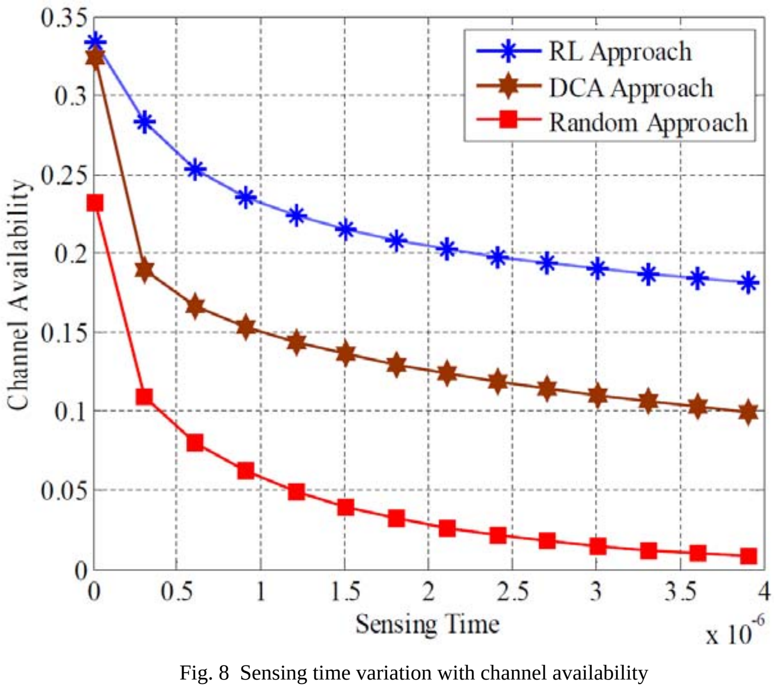

Fig. 8 Sensing time variation with channel availability

In Figure 9, we show the probability of channel handoff as a function of the sensing time. Both the RCA, DCA and the RL approach performs sense actions irrespective of the PU activity in the channel. In the DCA scheme, each CR cluster head alternate between *sense* and *transmit* action. In the proposed RL approach, each cluster head adjusts the sensing frequency based on the amount of PU activity detected on the current channel. Channel access decision is based on the reward function in (8) and the Q value in (9) and (10). For the three approaches, the average interference decreases with the reduced PU activity. Intuitively, the random approach incurs the highest probability to channel handoff due to PU activities, while DCA and RL approach provides the highest PU protection because each CR cluster head immediately handoff to a new channel each time a PU is detected in the current channel. In the RL-based approach, the sensing frequency is dynamically adjusted based on the amount of PU activity through the action selection function in (12). The difference with the DCA approach can be said to be due the fact that in our RL approach, each cluster head use a probabilistic policy $\pi(s, a)$ over the set of actions. Though randomization allows choosing suboptimal actions, Figure 9 shows that the impact is quite limited. Thus, our RL-based approach provides a similar impact on PU receivers than the DCA approach, but with a consistent performance gains for CR users.

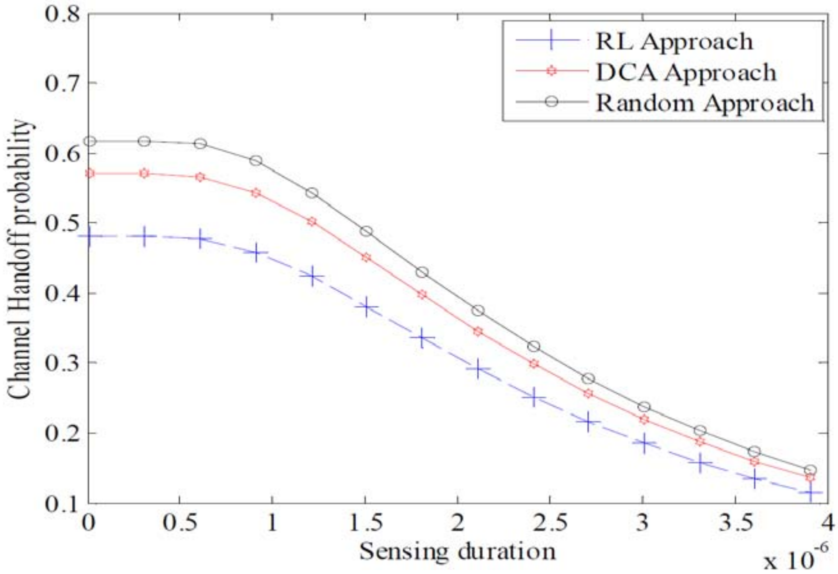

Fig. 9 Probability of handoff versus sensing duration

In case of rate of free channel access variation with energy efficiency, it can be seen from Figure 10 that as the rate of channel access increase, the energy efficiency increases accordingly. However, this performance metric as shown in Figure 10 reiterates that the RL learning approach performs better that DCA and RCA.

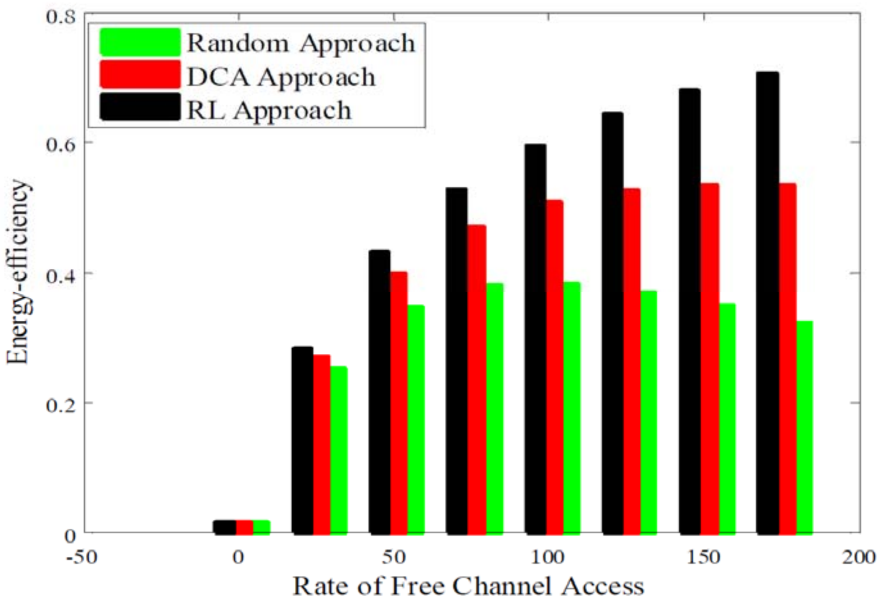

Fig. 10 Channel Free Access Rate effect on Energy Efficiency

PU activity is a major factor that determines how the CR user accesses a given channel. CR users operating in the underlay mode must vacate the channel as soon as the PU is detected. However, the PU usage is not time slotted. For the SU CR network to access the channel in energy efficient way, the rate of channel usage by the PU must be less or reduced in a given channel. To this end, Figure 11 shows that the lesser the rate of channel usage by the PU, the higher the energy efficiency of the channel access by the SU CR sensor network. At near 0 channel usage rate, the energy efficiency of channel access by the SU is about 1. This means, the efficiency is near 100%. Amazingly, this is the same for RCA, DCA and RL approach. However, with increase in rate of channel usage by PU, the energy efficiency of channel access drops accordingly. While the drop in the case of random approach is significantly noticeable, it is less noticeable in the DCA and the decline is least noticeable in our RL approach.

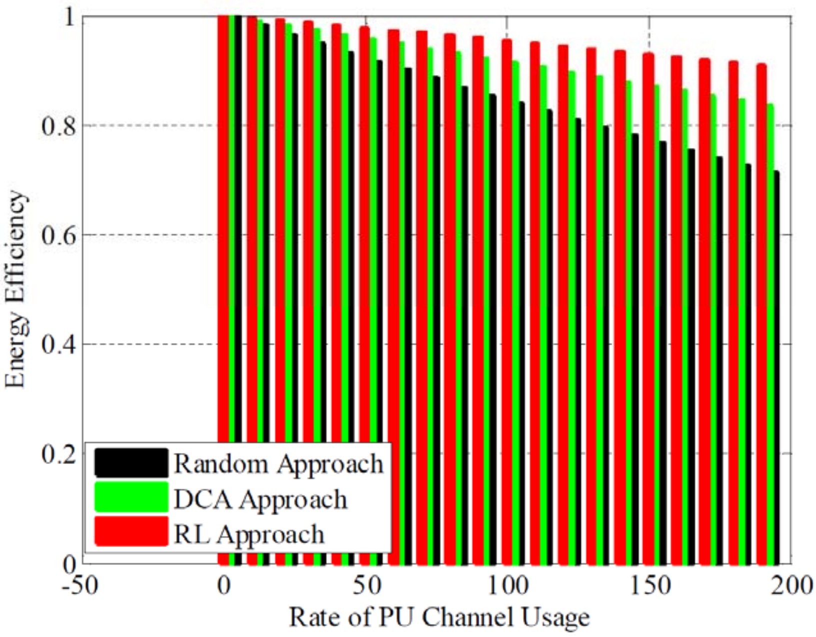

Fig. 11. Primary User Channel Usage effect on Energy Efficiency

V. CONCLUSION

In this paper, we proposed a channel access model based on reinforcement learning to optimize channel utilization experience of the secondary user cognitive radio enabled sensor network. With several learning methods available, RL-based channel access demonstrates the best performance in terms of energy efficiency, throughput, and channel availability in relation to sensing and transmission duration. We equally show that the proposed RL-based approach is capable of converging to the optimal solution and adapting to the change in radio environment better than DCA and RCA. Due to its learning capability, this work is extended to consider the activities of SU sensor network but not at the expense of PU model. Simulation results reveal significant energy efficiency and throughput performance improvement compared to non-learning channel access approaches.

ACKNOWLEDGMENT

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

REFERENCES

[1] Akyildiz I F, Won-Yeol L, Vuran M C, and Mohanty S 2008 "A survey on spectrum management in cognitive radio networks," IEEE Communications Magazine, vol. 46, pp. 40-48.
[2] Teng Y, Yu F R, Han K, Wei Y, and Zhang Y 2013 "Reinforcement-Learning-Based Double Auction Design for Dynamic Spectrum Access in Cognitive Radio Networks," Wireless Personal Communications, vol. 69, pp. 771-791.
[3] Correia L H A, Oliveira E E, Macedo D F, Moura P M, Loureiro A A F, and Silva J S 2012 "A framework for cognitive radio wireless sensor networks," IEEE Symposium on Computers and Communications, pp. 611-616.
[4] Yuming G, Yi S, Shan L, and Dutkiewicz E 2009 "ADSD: An Automatic Distributed Spectrum Decision method in Cognitive Radio networks," First International Conference on Future Information Networks, pp. 253-258.
[5] Ji Z and Liu K R 2008 "Multi-stage pricing game for collusion-resistant dynamic spectrum allocation," IEEE Journal on Selected Areas in Communications, vol. 26, pp. 182-191.
[6] Yu F R, Huang M, and Tang H 2010 "Biologically inspired consensus-based spectrum sensing in mobile ad hoc networks with cognitive radios," IEEE Network, vol. 24, pp. 26-30.
[7] Yang L, Cao L, and Zheng H 2008 "Physical interference driven dynamic spectrum management," 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 1-12.

[8] Wang S, Yue W, Coon J P, and Doufexi A 2012 "Energy-Efficient Spectrum Sensing and Access for Cognitive Radio Networks," IEEE Transactions on Vehicular Technology, vol. 61, pp. 906-912.
[9] Qianchuan Z, Geirhofer S, Lang T, and Sadler B M 2008 "Opportunistic Spectrum Access via Periodic Channel Sensing," IEEE Transactions on Signal Processing, vol. 56, pp. 785-796.
[10] Krishnamurthy V 2009 "Decentralized spectrum access amongst cognitive radios—An interacting multivariate global game-theoretic approach," IEEE Transactions on Signal Processing, vol. 57, pp. 3999-4013.
[11] Fu Y and Liu Q 2013 "Research of QoS Routing Algorithm in Ad Hoc Networks based on Reinforcement Learning," Electronics and Electrical Engineering, vol. 19, pp. 65-68.
[12] Abolarinwa J. A., Abdul Latiff N. M., Syed Yusof S. K., and Fisal N. 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 (ICFCNA 2014 - Malaysia), pp. 1-6.

AUTHOR PROFILE

J. Abolarinwa obtained Bachelor’s degree in Electrical and Computer Engineering and Master’s degree (M.Eng) in Electrical and Electronic Engineering. He is currently a Ph.D research student in the Faculty of Electrical Engineering, UTM. His research interests are in, Cognitive Radio Networks, Software Defined Radio, and Wireless Sensor Networks. He is a registered member of professional organizations such as IEEE and IET.

N. M. Abdul Latiff received her Bachelor of Engineering (B.Eng) degree in Electrical-Telecommunications in the 2001 from UTM, Malaysia. She obtained Master of Science (M.Sc) degree in Communications and Signal Processing, and Ph.D in Wireless Telecommunication Engineering from Newcastle University, UK in 2003 and 2008 respectively. Currently, she is a senior lecturer at the Faculty of Electrical Engineering, UTM and she is a member of the Telematic Research Group (TRG). Her research interest includes, Cognitive Radio, Wireless Sensor Networks, Mobile Ad Hoc Networks, Network Optimization, Bio-inspired Optimization Algorithm, Evolutionary Algorithm and Clustering Algorithm. She is a registered member of IEEE and IET.

S. K. Syed Yusof received BSc (cum laude) in Electrical Engineering from George Washington University USA in 1988 and obtained her MEE and PhD in 1994 and 2006 respectively from UTM. She is currently an Associate Professor with Faculty of Electrical Engineering, UTM, and collaborating with UTM-MIMOS Centre of Excellence. Her research interest includes wireless communication, Software define Radio, Network Coding and Cognitive radio. She is a member of Eta Kappa Nu (HKN), and Phi Beta Kappa society.

N. Fisal received her B.Sc. in Electronic Communication from the University of Salford, Manchester, U.K. in 1984. M.Sc. degree in Telecommunication Technology, and PhD degree in Data Communication from the University of Aston, Birmingham, U.K. in 1986 and 1993, respectively. Currently, she holds university professor position at the Faculty of Electrical Engineering, UTM and she is the director of UTM-MIMOS Centre of Excellence. Her research interests are in the areas of multimedia networking, wireless sensor network and cognitive radio.