UNIVERSITY OF MAIDUGURI
FACULTY OF ENGINEERING SEMINAR SERIES
VOLUME 8, MARCH 2017

A Reinforcement Learning Based Clustering and Cooperative Channel Sensing Algorithm for Cognitive Radio Wireless Sensor Network

I. Mustapha[1*], M. U. M. Bakura[1] , D. Mustapha[2], Borhanuddin M. Ali[3]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, University of Maiduguri.
[2]Department of Computer Engineering, Faculty of Engineering, University of Maiduguri
[3]Department of Computer and Communication Engineering, Universiti Putra Malaysia
*mustib@unimaid.edu.ng; +2347069266555

Abstract
Energy efficiency and spectrum efficiency are two main challenges in the realization of Cognitive Radio-Wireless Sensor Network (CR-WSN). Clustering is a well-known technique that could be used to achieve energy efficient communication and to enhance dynamic channel access in cognitive radio through cooperative sensing. While the energy efficiency issue has been well investigated in conventional wireless sensor networks, the latter has not been extensively explored. In this paper, a Reinforcement Learning (RL) based clustering algorithm is proposed to address energy and Primary Users (PUs) detection challenges in CR-WSN. The scheme minimizes network energy consumption, improves channel utilization and enhances PUs detection performance through RL based spectrum-aware clustering and cooperative channel sensing schemes. In the schemes, a cluster member node learns energy and cooperative sensing costs for neighboring clusters through exploration to select optimal cluster, while Cluster head learns channels dynamic behaviors in terms of channel availability, channel sensing energy cost and channel impairment to achieve optimal sensing sequence and optimal set of channels. Simulation results show convergence, learning and adaptability of the RL based algorithms to dynamic environment toward achieving the optimal solutions. Performance comparison of the proposed algorithm with other scheme indicates viability of the RL based scheme over the other bench mark schemes in terms of energy efficiency and PU detection performance.

Keywords: clustering; reinforcement learning; cooperative sensing; wireless sensor network; cognitive radio

1.0 Introduction

Tremendous growth in microelectronics technology and wireless communication applications have led to the widespread use of Wireless Sensor Networks (WSNs) in a wide variety of applications areas. In addition, WSN was envisioned to be the main enabling component in driving revolutionary technologies such as Internet of Things (IoT), Machine-to-Machine (M2M) and Web of Things (WoT) into reality (Ahmad et al., 2015). Wireless sensor node is a self-organizing entity equipped with communication, sensing and computing module that enables it to monitor certain events such as temperature, humidity, images, motion and seismic related signals in a geographical area and report the information to a sink node for further processing.

In general, wireless sensor nodes and many other wireless devices based on Wi-Fi, Zigbee and Bluetooth standards operate in unlicensed spectrum bands such as the Industrial Scientific and Medical (ISM) band which lack tight regulations. This leads to severe congestion in the useable unlicensed spectrum bands and causes harmful interference between the various wireless devices. On the other hand, licensed spectrum bands which are assigned to licensed users known as Primary Users (PUs) tend to become underutilized due to their fixed spectrum band allocation, as reported in the Federal Communications Commission (FCC) report (FCC, December, 2003). This necessitates the need for a paradigm shift from the conventional inefficient spectrum allocation policy to a dynamic and more flexible spectrum access management.

Cognitive Radio (CR) is an enabling technology to address the emerging demand for higher spectra efficiency through opportunistic access of vacant portions of the spectrum bands. The technology allows opportunists users known as Secondary Users (SUs) to operate on the vacant spectrum bands assigned to the licensed users known as PU (Mitola, 2000). The need to properly harness the potentials benefits of cognitive radio technology in WSN to improve spectrum utilization and support many applications that involve monitoring of sensitive and critical activities in an environment led to emergence of Cognitive Radio Wireless Sensor Network (CR-WSN) (Akan et al., 2009). A CR-WSN is a dispersed network of cognitive radio sensor nodes equipped with cognitive radio capability that dynamically utilize unused available spectrum bands to communicate sensed readings. This emerging technology is expected to be the most promising technology that has the potentials to address spectrum access challenges in conventional WSNs. However, practical realization of this breakthrough poses many challenges due to the resource constraints of the sensor nodes. Generally, CR sensor nodes inherent resource constraint of conventional wireless sensor nodes consequently are characterized by limited energy, constraint storage and processing resources. They are normally powered by battery and usually deploy in inaccessible terrain which make it difficult or impossible to replace and/or replenish the batteries (Anastasi et al., 2009). The additional task of opportunistic access to unused licensed channels through spectrum sensing incurs significant energy cost which drains more energy from the battery of the sensor nodes and hence shorten the life time of the network.

Therefore, the main challenges in CR-WSNs are energy efficient communication to extend the lifetime of the network and PU protection from unlawful interference. This paper presents a Reinforcement Learning based Clustering and Cooperative Channel Sensing (RL-CCCS) algorithm that enhances spectrum hole detection and minimizes network energy consumption in CR-WSNs.

2.0 Methodology

Spectrum sensing is the key fundamental function of CR for PU detections and licensed spectrum bands exploration. Autonomous spectrum sensing approach which requires individual SUs in a network to locally sense the spectrum bands and determine availability or otherwise of spectrum holes is usually susceptible to propagation impairments such as receiver uncertainty, multi-path fading, interference and shadowing which offset sensing performance as shown in Figure 1. Cooperative spectrum sensing has been proposed in (Singh *et al.*, 2012), (Lo and Akyildiz, 2013) as a viable option to address these issues. The key function of this approach is the exploration of multi-user sensing diversity to improve spectrum sensing performance (Mustapha *et al.*, 2015b). Although the approach achieves significant success in improving sensing performance, it also incurs heavy communications overhead especially in large-scale networks such as CR-WSN. To mitigate these problems and minimize network energy consumptions, multiple SUs can be logically grouped to form a cluster and a dedicated SU can be assigned to coordinate spectrum sensing.

Network clustering involves partitioning the network into logical groups of nodes that form clusters, each cluster comprises of a Clusterhead (CH) and none Clusterhead nodes are referred to as Member Nodes (MNs). The CH may serve as a central point to all nodes in the cluster, and it performs various tasks such as data aggregation and spectrum sensing coordination. In addition, it also provides inter-cluster communications by communicating with neighboring CHs and a Base Station (BS). The MN detects events and communicates its data to the associated CH through intra-cluster communications. Clustering of a network has several benefits and it has been widely explored in conventional wireless sensor networks. However, its application in CR-WSN to enhance PU protection has not been fully explored (Ibrahim Mustapha, 2014). Therefore, conventional clustering algorithms for WSNs or mobile ad hoc networks may not be suitable for CR-WSN due to the dynamic nature of the channels. The network is assumed to be static, consisting of N non-mobile homogenous fully functional cognitive radio sensor nodes uniformly distributed in a two-dimensional square area $\mathrm{N_A}$ of $\mathrm{L} \times \mathrm{L}$ square meters as shown in Figure 2. All member nodes lie within the radio range of their respective Clusterheads($d < R_{max}$) and communicate directly with the Clusterheads in a single-hop manner.

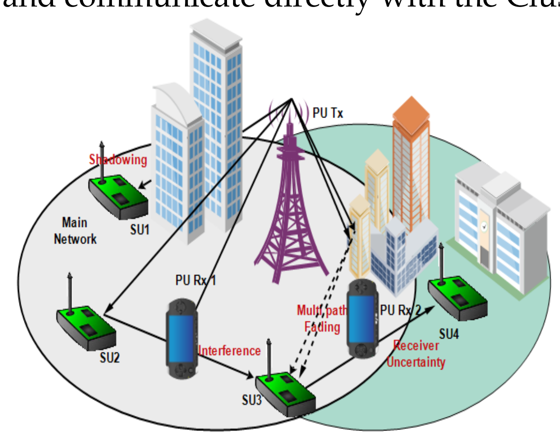

Fig 1: Illustration of Propagation Impairments

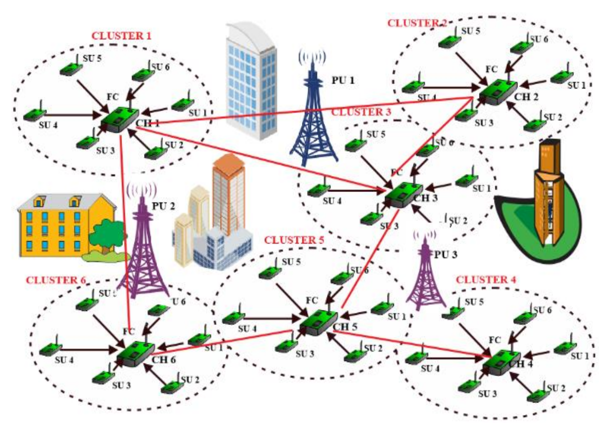

Fig. 2. Clustered cooperative channel sensing.

Reinforcement learning is a machine learning technique that allows an agent to interact with its operating environment and learn an optimal policy that maximizes cumulative rewards (Gosavi, 2011). Although RL is a well-known machine learning technique and has been extensively applied to many fields but its application in clustering algorithm is still at the infancy stage.

The RL-CCCS involves three main phases namely, initialization, set-up and maintenance phases. The initialization phase deals with neighbor discovery and determination of node eligibility to contest for Clusterhead role. In this phase, each SU_i identifies its one-hop neighbors, senses the spectrum to detect vacant channels and computes its eligibility probability $P_{ch,i}$based on node degree nd_i, number of vacant channels detected cv_i, residual energy er_i, distance to sink dc_i, and percentage of Clusterheads ψ_i (e.g., 6%) for the network. The main goal is to ensure that all the SUs in the network are covered by a set of CHs at the initial stage. The set-up phase involves election of Clusterheads based on eligibility probability $P_{ch,i}$ and establishment of clusters where MNs identify their respective cluster and associate with the CH in the cluster. During the Clusterhead election, SUs with highest eligibility probability $P_{ch,i}$ among the neighboring eligible SU would be emerged as Clusterheads while the others are likely to be member nodes. The maintenance phase allows re-election of Clusterhead whenever the residual energy of the incumbent CH depletes to a level below a threshold or when the CH disconnected from network.

2.1 Energy Consumption Model for Cognitive Radio Wireless Sensor Node

The main components of a cognitive radio sensor node are event sensing, data processing, communication and cognitive radio units. The event sensing unit monitors the environment and generates signal traffic whenever an event is occurred. The processing unit processes the data while the communication unit transmits the data to desire sink over a free licensed channel and receives data. The cognitive radio module detects set of unused licensed channels and then accesses the most suitable channel to communicate the data. The energy consumption for data transmission comprises of energy cost for intra-cluster and inter-cluster data communication. The energy consumption for intra-cluster communication $E_{Intr}(i,j)$ comprises of total energy consumption $\mathrm{E}_{tM}(B_{dt}, d_{i,j})$ of all MNs for transmitting B_{dt} bits of data packet over a distance $d_{i,j}$ to their respective Clusterheads CH_j and total energy consumption $\mathrm{E}_{rM}(B_{dt})$ for receiving the data by the Clusterhead CH_j. In inter-cluster data communications, each Clusterhead CH_j forwards its aggregated data packets B_{pk} to the BS either through intermediate Clusterhead CH_g $(if\ d_{j,g} > R_{max,j})$ or direct to the BS without any intermediate Clusterhead CH_g $(if\ d_{j,g} \leq R_{max,j})$. The total energy consumption for inter-cluster data communications E_{int} is the energy consumption for aggregating the received data $E_{ap}(B_{ap}, N_{cy})$ and energy consumption $\mathrm{E}_{tH}(B_{pk}, d_{j,g})$ for forwarding the data packet to the BS. Thus, the total energy consumption for data communications can be expressed as (Mustapha et al., 2015a):

$$E_{total}^{dt} = E_{Intr}(i,j) + E_{int}$$

$$= \sum_{j=1}^{q}(\sum_{i=1}^{m_n}(\underbrace{E_{tM}(B_{dt}, d_{i,j})}_{\substack{Data\\Transmitting}} + \underbrace{E_{rM}(B_{dt})}_{\substack{Data\\Receiving}}) + \underbrace{E_{ap}(B_{ap}, N_{cy})}_{\substack{Data\\Aggregation}} + \underbrace{E_{tH}(B_{pk}, d_{j,g})}_{\substack{Data\\Transmitting}}) \qquad 1$$

If $m_{n,j}$ denotes the number of member nodes in cluster cl_j and q is the number of clusters in the network which is also equal to the number of Clusterheads in the network, then the number of cognitive radio sensor nodes in the network $N = \sum_{j=1}^{q}(m_{n,j} + 1)$. Therefore, total energy consumed by member node MN_i in cluster cl_j is given as (Mustapha *et al.*, 2015a):

$$E_{MN}(i,j) = \underbrace{E_{ss}(B_{ss})}_{\substack{Events\\Sensing}} + \underbrace{E_{log}(B_{lg})}_{\substack{Data\\Logging}} + \underbrace{E_{cs}(n_z, T_{cs})}_{\substack{Channel\\sensing}} + \underbrace{E_{rp}(B_{ld}, d_{i,j})}_{\substack{Decision\\Reporting}} + \underbrace{E_{rx}(B_{ld})}_{\substack{Coop.Decision\\Receiving}} + \underbrace{E_{tM}(B_{dt}, d_{i,j})}_{\substack{Data\\Transmitting}} \qquad 2$$

Similarly, total energy consumption of Clusterhead CH_j in cluster cl_j is given as (Mustapha et al., 2015a):

$$E_{CH}(j) = \underbrace{E_{ss}(B_{ss})}_{\substack{Events\\Sensing}} + \underbrace{E_{log}(B_{lg})}_{\substack{Data\\Logging}} + \underbrace{E_{cs}(n_z, T_{cs})}_{\substack{Channel\\sensing}} + \sum_{i=1}^{m_n}(\underbrace{E_{rx}(B_{ld})}_{\substack{Decisions\\Receiving}} + \underbrace{E_{rM}(B_{dt})}_{\substack{Data\\Receiving}}) + \sum_{i=1}^{m_n+1}(\underbrace{E_{dp}(B_{dp})}_{\substack{Decisions\\Processing}} + \underbrace{E_{ap}(B_{ap}, N_{cy})}_{\substack{Data\\Aggregation}}) + \underbrace{E_{tM}(B_{dt}, d_{i,j})}_{\substack{Data\\Transmitting}} \qquad 3$$

Thus, total energy consumption for the entire network is given as (Mustapha et al., 2015a):

$$E_{net} = \sum_{j=1}^{q}\left(E_{CH}(j) + \sum_{i=1}^{m_{n,j}} E_{MN}(i,j)\right) \qquad 4$$

2.2 RL Model for Clustering and Cooperative Channel Sensing

The RL based clustering and cooperative channel sensing decision process are modelled as a Markov Decision Process (MDP) in which $\mathcal{S}$ denotes set of states which represents the model of the environment, $\mathcal{A}$ denotes set of actions and $\mathcal{R}$ denotes reward function in a given state. In the clustering scheme, the agent initializes the CH selection by selecting action $a_k^e = 0$ at stage $k = 0$ and $s_k^e = 0$ such that indicator function $I_A\{\overline{w}\} = 0$. In every episode e the agent selects an action $a_k^e \in \mathcal{A}$, $a_k^e = j | j = 1,2,3, \ldots n_{hg}$ in state $s_k^e = \bar{v} \in S$ which leads to selecting a Clusterhead $CH_j | j = 1, 2, 3, \ldots n_{hg}$ among the neighboring CHs. The agent computes energy cost and evaluates cooperative cost for the chosen CH and then adopts a policy π that maximizes the cumulative reward r_{k+1}^e as presented in Table 1. In the cooperative channel sensing, agent i chooses an action $a_k^e = z | z = 1,2, \ldots n_z$ in state $s_k^e = u \in S$ to sense a channel ch_z and computes metric functions based on local decisions $LD_{k,1}^{z,e}$ and cooperative decision $CD_k^{z,e}$ as presented in Table 1 (Mustapha *et al.*, 2016). The agent learns optimal set of the channels by adopting a policy π that maximizes the cumulative reward r_{k+1}^e received from exploitation of unknown states and experiences received from the known states.

Table 1: MDP Representation for the RL based Clustering and Cooperative Channel Sensing

	Description for RL Based Clustering	Description for Cooperative Channel Sensing
State	Finite set of neighboring clusters $\mathcal{S} = \{0,1,2,,..n_{hg}\}$. A state $s_k^e \in \mathcal{S}$ represents neighboring Clusterhead $CH_j \| j = 1, 2, 3, \dots n_{hg}$ that can be selected. $s_k^e = \bar{v}$ $I_A\{a_k^e = \text{j} \in \mathcal{A}, k \neq 0\}$, $I_A\{\bar{w}\} \neq 0$.	Finite set of licensed channels $\mathcal{S} = \{0,1,2,.n_z\}$ A state $s_k^e \in \mathcal{S}$ represents licensed channel $Ch_z \| z = 1,2,3 \dots .n_z$ that can be selected for sensing. Its value is given as $s_k^e =$ u $I_A\{a_k^e = \text{z} \in \mathcal{A}, k \neq 0\}$, $I_A\{\bar{a}\} \neq 0$.
Action	Set of actions in state $s_k^e = \bar{v}$, $\mathcal{A} = \{0,1,2,.n_{hg}\}$ An action $a_k^e \in \mathcal{A}$, $a_k^e = j \| j = 1,2,3, \dots n_{hg}$ selects cluster CH_j and computes immediate reward $r_k^e(s_k^e, a_k^e) \in \mathcal{R}$.	Set of actions in state $s_k^e = u$, $\mathcal{A} = \{0,1,2,3,..n_z\}$ An action $a_k^e \in \mathcal{A}$, $a_k^e = z \| z = 1,2,3, \dots n_z$ denotes task in a given state to sense channel Ch_z, compute the metric functions and immediate reward $r_k^e(s_k^e, a_k^e) \in \mathcal{R}$
Reward	Reward$r_{k+1}^e\ (s_{k+1}^e, a_{k+1}^e) = R_{wt,k+a}^e$. The cumulative reward $r_{k+1}^e = R_{wt,k+a}^e$ is weighted rewards for energy cost $rw_{E,k+1}^e$ and cooperative cost $rw_{C,k+1}^e$ for the selected Clusterhead.	Cumulative reward $r_{k+1}^e\ (s_{k+1}^e, a_{k+1}^e) = A_{w,k+a}^e$ The cumulative reward r_{k+1}^e is the weighted average of three metric functions, namely local decision success D_{k+1}^e, channel availability Av_{k+1}^e and energy cost E_{k+1}^e for sensing the channel.

3.0 Results and Discussion

The simulation is based on simulation parameters used in (Mustapha et al., 2015a) and (Mustapha et al., 2016) in which a low power wireless sensor nodes was considered in computing the energy dissipations. Q-learning and SARSA algorithms were implemented in MATLAB to evaluate the performance of the algorithm over $E_{ps} = 5000$ episodes as shown in Figure 3. A step size for the exploration of state-action pairs and for learning rate update is set to $\alpha^k = a/(b+k)$ in which $a = 10, b = 100$, while the discount factor is set to $\gamma = 0.9$. The result indicates that both the Q-learning and SARSA algorithms converged to the optimal solution but after different numbers of episodes. The SARSA learning algorithm converged to the optimal value after $E_{psd} = 3020$ episodes and achieved a maximum average expected cumulative reward value of $r_{Av}^t = 0.52$ which is much higher than that of the Q-learning algorithm. On the other hand, the Q-learning algorithm converged to an optimal solution at $E_{psd} = 2020$ which is much lower than SARSA and achieved a maximum cumulative reward value of $r_{Av}^t = 0.44$. This suggests that the Q-learning algorithm converges to optimal solution in a relatively shorter learning period because of its reliance on action selection strategy rather than cluster exploration to update its policy. In contrast, cluster explorations while updating the Q-value slows the convergence of the SARSA algorithm due to the extension of its learning period, but this of course yields a better cumulative reward. It can be concluded that learning period has a significant impact on the convergence of the algorithms.

To evaluate the performance of RL-CCCS, optimal channel sensing sequence for the other approaches i.e exhaustive search, heuristic search and greedy search algorithms are obtained through simulations (Mustapha et al., 2016). Then their performance is compared with the RL-CCCS approach as shown in Figure 4. Based on the optimal channel sensing order of each approach, a channel sensing energy cost is computed to determine average energy cost for sequential sensing of the channels. The result shows that heuristic search approach recorded much higher average energy costs, the RL-CCCS achieved minimum average energy cost compared to the other approaches. Although Greedy search approach achieved the least average energy cost among the three approaches, its average energy cost is much higher than the RL-CCCS approach. For example, average energy cost for RL based at sensing instance of 100 is 15.17% lower than average sensing energy cost for the Greedy search approach at the same sensing instance. It is evident that the RL based approach tends to learn the channels behavior through exploration and selects the best channels.

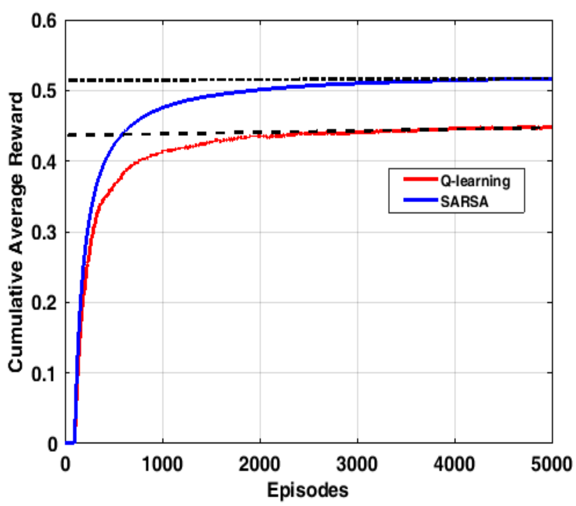

Figure 3. Convergence of RL-CCCS.

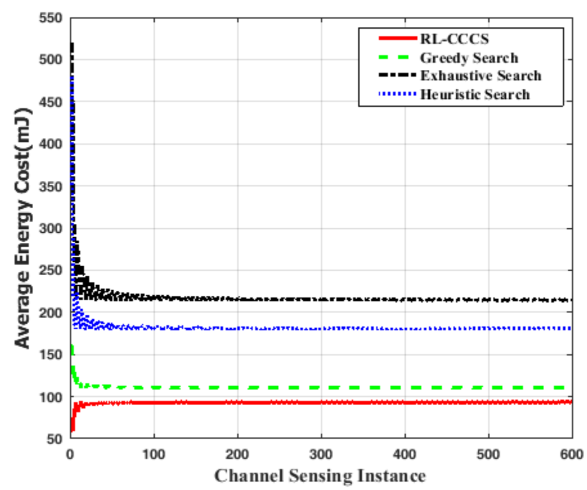

Figure 4. Comparison of average sensing energy cost of RL- CCCS and other approaches.

4.0 Conclusions

Energy efficient communications and efficient dynamic spectrum access are two nontrivial issues in cognitive radio sensor networks. The increasing demand for spectrum access and intrinsic features of CR sensor node which include constrained energy and computing resources necessitate the need to address these issues. In this paper, a RL based clustering and cooperative channel sensing algorithm that minimizes network energy consumption and enhances channel sensing in cognitive radio sensor networks is presented. MATLAB environment has been used as a simulation tool in which Monte Carlo simulation has been extensively used for the energy detection based performance evaluation. Performance comparison of the RL-CCCS with other schemes have shown viability and performance improvement of the RL based approach in terms of sensing energy cost and PU.

References

Ahmad, A., Ahmad, S., Rehmani, M. H. & Hassan, N. U. 2015. A Survey On Radio Resource Allocation In Cognitive Radio Sensor Networks. *Communications Surveys & Tutorials, Ieee,* 17**,** 888-917.

Akan, O., Karli, O. & Ergul, O. 2009. Cognitive Radio Sensor Networks. *Network, Ieee,* 23**,** 34-40.

Anastasi, G., Conti, M., Di Francesco, M. & Passarella, A. 2009. Energy Conservation In Wireless Sensor Networks: A Survey. *Ad Hoc Networks,* 7**,** 537-568.

FCC December, 2003. Docket No 03-222 Notice Of Proposed Rule Making And Order. Federal Communications Commission, Washington, Usa.

Gosavi, A. 2011. A Tutorial For Reinforcement Learning. *Department Of Engineering Management And Systems Engineering*.

Ibrahim Mustapha, B. A., Aduwati Sali, Mohd. Fadlee A. Rasid. Energy-Aware Cluster Based Cooperative Spectrum Sensing For Cognitive Radio Sensor Networks. International Symposium On Telecommunication Technologies 24-26 November, 2014 2014 Langkawi Island, Malaysia.

Lo, B. F. & Akyildiz, I. F. 2013. Reinforcement Learning For Cooperative Sensing Gain In Cognitive Radio Ad Hoc Networks. *Wireless Networks,* 19**,** 1237-1250.

Mitola, J. 2000. *Cognitive Radio: An Integrated Agent Architecture For Software Defined Radio.* Ph.D. Dissertation, Kth Royal Institute Of Technology, Stock-

Holm, Sweden, 2000.

Mustapha, I., Ali, B. M., Rasid, M. F. A., Sali, A. & Mohamad, H. 2015a. An Energy-Efficient Spectrum-Aware Reinforcement Learning-Based Clustering Algorithm For Cognitive Radio Sensor Networks. *Sensors,* 15**,** 19783-19818.

Mustapha, I., Ali, B. M., Rasid, M. F. A., Sali, A. & Mohamad, H. A Weighted Hard Combination Scheme For Cooperative Spectrum Sensing In Cognitive Radio Sensor Networks. 2015 Ieee 12th Malaysia International Conference On Communications (Micc), 23-25 Nov. 2015 2015b. 12-17.

Mustapha, I., Ali, B. M., Sali, A., Rasid, M. & Mohamad, H. 2016. An Energy Efficient Reinforcement Learning Based Cooperative Channel Sensing For Cognitive Radio Sensor Networks. *Pervasive And Mobile Computing*.

Singh, A., Bhatnagar, M. R. & Mallik, R. K. 2012. Cooperative Spectrum Sensing In Multiple Antenna Based Cognitive Radio Network Using An Improved Energy Detector. *Communications Letters, Ieee,* 16**,** 64-67.