Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 48 ***(4D)*** *151-156 (2016)*

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Cooperative and Reinforcement Learning in Energy Efficient Dual Hop Clustered Networks

A. F. RAMLI, Y. H. BASARUDIN, M. I. SULAIMAN, F. I. ADAM, D. GRACE*

University Kuala Lumpur British Malaysian Institute, Batu 8 Jalan Sungai Pusu, Gombak, Selangor, Malaysia.

Received 28rd August 2016 and Revised 17th October 2016

Abstract: This paper examines the application of distributed Reinforcement Learning RL to improve the spectral efficiency in high data rate applications for clustered networks. With RL, cluster members can learn to identify set of channels, which have the highest success rate. It is shown that RL can minimize the dual hop clustered networks interference as the uplink delay is reduced by up to 30% and improve the network energy efficiency of by up to 10% compared to a random channel allocation. However, distributed RL has a very poor convergence time. In this paper, we present two methodologies on how through cooperative learning and RL, cluster members can exchange channel historical information to facilitate learning. The proposed cooperative learning methods enable cluster members to enter exploitation stage by a factor of 3 times faster compared to distribute RL. Furthermore, the proposed methods allows each cluster to adapt the number of required preferred channel size depending upon the local area density and traffic. The results shows that the adaptability reduces variation in uplink delay between cluster members by 45% compared to distributed *RL* making the system more equal.

Keywords: Clustered Networks, Reinforcement Learning, Cooperative Learning, channel allocation, Energy efficient communication

1. INTRODUCTION

Advancement in processor technology has enables the introduction of smart mobile phones which has led to the soaring demand for high data rate wireless applications. Numerous researchers and wireless networks providers are coming up with novel solutions to optimize the performance of wireless networks in order to cope with the exponential growth of wireless traffic. Some of the proposed methodologies to increase the network throughput add complexity at the detriment of the network energy consumption. It is estimated that in 2012, the Information and Communication Technology (ICT) sector contributed 3% of the total global energy consumption (Hasan *et. al*. 2011). However, this figure is projected to increase due to the ever increasing demand for wireless applications by the public and the growing trend amongst industries toward adopting Wireless Sensor Networks (WSN) and Internet of Things (IOT).Greater awareness amongst the public on the adverse effect of carbon emission has on climate change has pressured the wireless network providers and regulatory bodies such as the ITU to provide not only high capacity wireless networks but energy efficient. Future wireless networks needs to minimize energy consumption whilst maintaining quality of service QOS.

As such, concept such as 'Cell Zooming' has been proposed by (Zhisheng, 2010) for cost efficient Green Cellular Networks. By allowing base stations to adaptively vary the cell coverage depending on the cell traffic load and channel conditions, energy consumption of the networks can be reduced. Others, such as Son (2011), developed a practical algorithm for the deployment of Hierarchical Cell Structure HCS. Through extensive simulations, the algorithm developed by Son (2014) were shown to be able to minimize the total energy consumption while satisfying the area spectral efficiency requirement. (Heinzelman *et. al*. 2002), it was noted that by hierarchical architecture namely clustering in wireless sensor networks WSN, the network lifetime can be significantly prolonged compared to multi-hop routing protocols. Unlike that of (Heinzelman *et. al*. 2002) which was designed for low data rates WSN, the applications hierarchical architecture in the form of dual hop clustered networks as a mean to achieve energy efficient communication (balancing the network energy consumption and throughput) for high data rate application were investigated in (Ramli *et. al*. 2015).

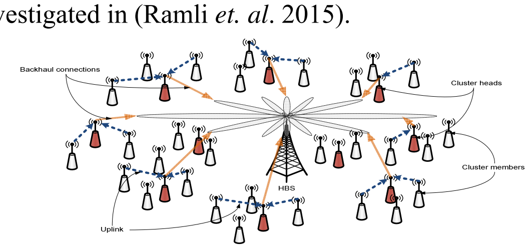

Fig-1 Dual hop Clustered Networks with directional antennas

++Corresponding author: Aizat Ramli email: aizatfaiz@unikl.edu.my
*Department of Electronics, University of York, York YO10 5DD, UK

(Jiang, 2011) demonstrated the application of fully distributed Reinforcement Learning RL scheme for channel selection and spectrum sharing in cognitive radio. Through RL, a node can learn to identify a set of channels that have highest probability of successful transmissions. A node therefore node need to sense certain number of channels prior to transmissions or directly transmit to these channels without sensing. It is demonstrated that the fully distributed RL could improve the cognitive radio network performance by reducing the probability of blocking and dropping. Although there were no direct figures on the energy consumption, Jiang claimed that reducing the need for sensing with RL could reduce the total network energy consumption by 5% compared to no learning scheme. The main drawback of a fully distributed RL as noted by (Jiang, 2011) is the poor convergence.

The aim of this paper is to explore the implementation of RL for channel allocations as a mean to achieve an energy efficient communication in dual hop clustered networks presented in (Ramli *et. al.* 2015). The motivation behind this study is because it was found that reducing the dual-hop clustered networks interference can increase its energy efficiency in Joules/bit by 50%. A novel learning algorithm will be proposed that utilizes the concept of cognitive cooperative in order to reduce the learning duration of RL. The cooperative algorithm differs from a distributed approach as information can be shared with surrounding nodes to facilitate learning process

2. REINFORCEMENT LEARNING

Reinforcement Learning RL is a sub-class of machine learning techniques in which the learner learns about its environment by taking random (trial and error) actions *A*. Specifically, before a learner undertake a set of possible Actions *A* at time *t*, a learner will first senses the state s_t of its environment *S*, $s_t \in S$. According to s_t, the learner will choose the appropriate action $a_t \in A(s_t)$ which will result in a change state of state at s_{t+1} and receive a reward $r_t \in R$. The size of r_t depends upon the favorability of s_{t+1} towards the learner. The objective of RL is to develop a policy π that can maximize the reward, $\pi : S \to A$.

(Jiang, 2011), developed a fully distributed RL for spectrum sharing in cognitive radio by having policy π which maps weights *W* associated for each channel into action *A* (channel selection), $\pi : W \to A$. The algorithm developed by Jiang can be adapted for uplink channel allocation problem in a dual hop clustered networks environment in order to reduce the network interference. It is expected that reducing the network interference will not only improve the system throughput but it will also reduce the networks energy consumption. One critical drawback of a fully distributed scheme as noted by (Jiang, 2011), (Giupponi *et. al.* and Moreno 2011) is the extremely poor learning rate.

3. NETWORK MODELLING

As highlighted by (Ramli *et. al.* 2015), the throughput of the dual hop clustered network is restricted due to the relaying constrained experienced at the cluster heads. The bottleneck is induced to the limited spectrum availability since the total channel pool *Q* has to split between cluster heads n^h and cluster members n^c . The spectrum efficiency can be markedly improved by applying directional antennas at the Hub Base Station HBS. The approach was adopted by FP7 BuNGee also with a two hop architecture but with nodes deployed in orderly placement.

In this paper, we use the same network model as that presented in Ramli *et. al.* (2015) but with incorporation 12 beams directional antenna deployed at the HBS as depicted in **(Fig-1)**. Then network compose of the following n^h cluster heads, n^c cluster members with Q^u available uplink channels and Q^b back haul channels. The directional antennas radiation pattern follows that of Bungee Deliverable: D1.2 (2010) and have a 30 degree separation on the main lobe. Simulation results presented in **(Fig-2)** indicated that the deployment of directional antennas are able on dual hop clustered networks are able alleviate the relaying constrained at backhaul since the uplink and throughput performance is identical.

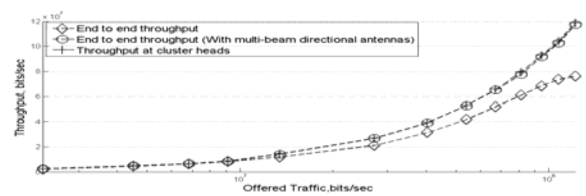

Fig-2 Dual hop Clustered Networks throughput performance with directional antennas

4. CHANNEL ALLOCATION SCHEMES IN CLUSTERED NETWORKS

In this section we discussed the implementation of distributed RL channel allocation in clustered networks and proposed a novel learning scheme through cooperation amongst clusters

4.1 Distributed RL channel allocation

A fully distributed RL can be applied to a dual hop clustered as follows; consider a clustered networks with a total of n^h cluster heads, n^c cluster members with Q^u available uplink channels. During exploration stage i.e. learning period, each *i-th* cluster member ($i \in \{1,\ n^c\}$) will accumulate and update the weight W_q^i for each *q* uplink channel, $q \in \{1,\ Q^u\}$. A cluster member n^c will have equal probability of access to Q^u via the random channel allocation scheme (trial and error). The weight W_q^i vector can be summarized by (1) and will be updated by n^c for each transmission at time *t* using the function given in (2).

$$W_q^i = \{W_1^i, W_2^i .., W_q^i\},\ i \in \{1,\ n^c\},\ q \in \{1,\ Q^u\} \quad \textbf{(1)}$$

$$W_q^i(t+1) = W_q^i(t) + r \quad \textbf{(2)}$$

A reward of r =1 will be assigned to the *q-th* channel for every successful transmission attempt by *i-th* cluster member to its cluster head. A reward of r =-1 will be assigned to the *q-th* channel if the channel is occupied by other n^c transmission or the transmission is interrupted due the Signal to Interference Noise Ratio *SINR* is less that the $SINR_{threshold}$. The weight W_q^i for each *q-th* channel for a particular *i-th* cluster member will continuously be updated until W_q^i exceed the weight threshold $W_{threshold}$. Once an *i-th* cluster member has a certain number of channels whose $W_q^i \geq W_{threshold}$, the channels will be classified a preferred channel set *P* and the *i-th* cluster member enters exploitation stage. During exploration stage, the *i-th* cluster member transmission will only take place on a random *P* and update W_q^i accordingly. As in Jiang (2011), the number of preferred channel set *P* equal to 5. If during exploitation stage *P<5,* then the *i-th* cluster member will revert to exploitation stage.

4.2 Cooperative RL channel allocation

The implementation of a cooperative RL channel allocation scheme in a clustered networks require cluster members belonging to the same cluster to cooperate and share the weight associated with each Q^u. The sharing of the information will enable cluster members to quickly accumulate weight and identify *P*. Since there is no communication amongst cluster members, a facilitator or a teacher is required. Cluster heads n^h can become facilitators by allowing it to have the ability to learn by accumulating the weight for each Q^u and share this information with its cluster members. The weight vector W_q^k on each *q-th* channel for *k-th* cluster heads can be described as follows

$$W_q^k = \{W_1^k, W_2^k .., W_q^k\},\ k \in \{1,\ n^h\},\ q \in \{1,\ Q^u\} \quad \textbf{(3)}$$

The following describes the operation of cooperative RL during exploration and exploitation stage

Exploration stage: For each successful transmission by *i*-th cluster member, the *k-th* cluster head will update W_q^k using the function given in (2) by giving a reward of r =1 for the *q-th* channel. Unlike distributed RL, the cluster heads has no information on block transmissions therefore no punishment will be applied. If an on ongoing transmission is interrupted due to the $SINR$< $SINR_{threshold}$, then a punishment or a reward of r =-1 will be assigned to the *q-th* channel

Exploitation stage: An *i-th* cluster member enters this stage when any of its *q-th* channel has a $W_q^k \geq W_{threshold}$. A channel which meets the requirement $W_q^k \geq W_{threshold}$ is classified as a preferred channel *P*. At this stage, a cluster members will first attempt to transmit onto a random *P*, if all the *P* are occupied then it will revert to the exploration stage. Enabling a cluster member to enter exploration stage due is insufficient *P* allow highly populated clusters to have a higher number of *P* size than those less densely populated clusters. **(Fig-3)** presents the flowchart for cluster members to transmit on to Q^u through cooperative RL.

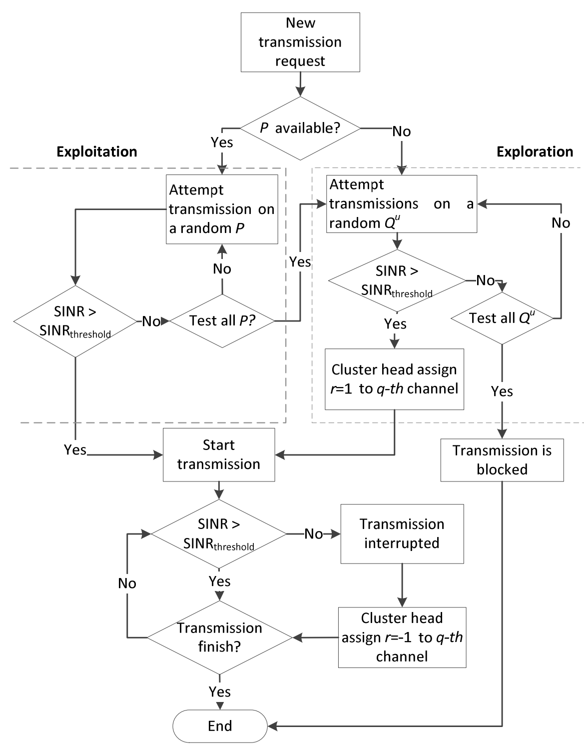

Fig-3 Flow chart of Cooperative RL channel allocation scheme.

As mentioned earlier, the weight W_q^k of the uplink channels Q^u is updated by the cluster heads needs to be shared with its cluster members. This process is crucial so that cluster members can determine the availability of *P*. Two schemes known as *Force Update* and *Delay Update* are proposed and discussed below for the process of Q^u information sharing between cluster heads and its cluster members.

Force Update: In this scheme, an *i-th* cluster member update its W_q^k at the start of transmission request. A cluster member will first send a Request to Send RTS on to a random Q^u to inform intention to transmit. A successful RTS transmissions will be acknowledged by the cluster head through sharing of W_q^k database information on to the same *q-th* channel. The cluster member will then re-transmit RTS based upon the updated W_q^k. The major drawback of this scheme is the excessive communication required before a transmission can start.

Delay Update: In this scheme, the updated W_q^k is sent along with Clear to Send CTS frame. An *i-th* cluster member will attempt transmission based on $W_q^k(t-1)$ which were updated from previous transmission. Although the decision is based upon a delayed W_q^k, this scheme eliminates unnecessary overhead of Force Update.

5. SIMULATIONS AND RESULTS

This section assess the performance of distributed RL and cooperative RL for channel allocation in a clustered networks. As stated earlier, the energy consumption and network modeling and parameters will follow that of (Ramli *et. al*. 2015) with the addition of 12 beam directional antennas and implementation of RL for channels allocation. The nodes were clustered using sum RSSI clustering protocol developed by (Ramli and Grace 2013) with nodes transmission range set at 200m. It is assumed that power is only consumed when the nodes in the network are in transmission or reception mode, otherwise they are assumed to be low powered state such as sleep mode in which the power consumption is considered negligible. The RL schemes will be compared to the random (also known as no learning scheme) and Least Interfere. Note that the Least Interfere scheme requires node to sense all the available spectrum prior to transmission and therefore limits its practicality. **(Table-1)** provides the summary on the parameters used in the simulations.

Table 1. System Parameters

Parameters	Value
Size of Network layout	1,000m×1,000m
Number of Nodes	100
Centre Frequency	2.1 GHz
Carrier Bandwidth	1MHz
Maximum Radiated Transmit Power	0dBW
Node Antenna Gain (*Gt,Gr*)	0dBi
Noise figure	5 dB
SINRthreshold	5 dB
SNIRmax	21 dB
Noise floor	-134dBW
File Length f_h	45Mb
Nodes antenna heights	25 m
Cmax	4.5bps/Hz
The total number of available channels *Q*	40
Uplink Channels Q^u	20
Traffic model	Poisson

Simulation result presented in Fig.5 shows the comparison on average end to end delay performance of several channel allocation schemes. The result provide an insight on the level of interference experienced by cluster members n^c. The random channel allocation experience the highest level of interference thus increases the probability of blocking and interrupted transmissions. High level of interference also reduces channel capacity since $capacity \propto 1 + Log_2(SINR)$ which in turns result in high end to end delay. The ability of RL schemes (distributed RL and cooperative RL) to learn enables the n^c to make an inform decision based on Q^u rate of success. With RL, nodes in the network learn to avoid transmissions on to the same spectrum as its immediate neighbour by prioritizing to transmit on to *P*. This result in improvement in spectrum efficiency and improve delay performance by up to 30%. The result in **(Fig. 5)** also shows that there is negligible difference in the performance between RL and random allocation scheme at an offered traffic of less than 30Mb/s. At low traffic loads, the occurrence of blocking and transmission interruptions is very low, therefore RL is unable to discern between the channels that provide the highest capacity. Although marginal, Delay Update scheme has a slightly poor delay performance compared to distributed RL and Force Weight scheme. The additional delay is caused by n^c utilising obsolete channel state information to decide on channel selection.

Although all the RL schemes have a comparable average delay performance, results shown in **(Fig. 6)** highlights that cooperative RL schemes are able to reduce the delay performance disparity amongst n^c by up to 45% compared to distributed RL. Unlike distributed RL which employed a fix *P*, cooperative RL has a flexible *P* by enabling n^c to re-enter exploration and find additional *P* if the current *P* set is insufficient. Therefore, highly densely populated clusters may choose to have a larger *P* set since it will experience a much higher level of interference than the sparsely populated clusters. The flexibility in *P* creates a more evenly distributed level of interference throughout the network and thus reduces delay variation between n^c.

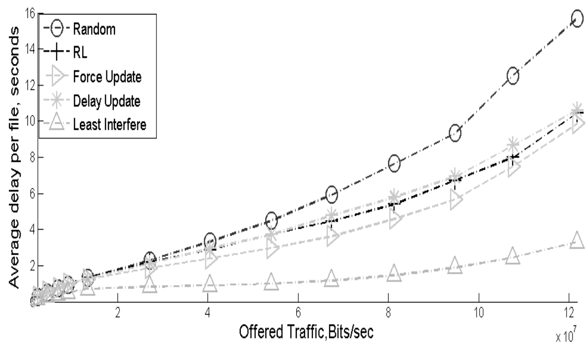

Fig-5 Cluster members average delay per file

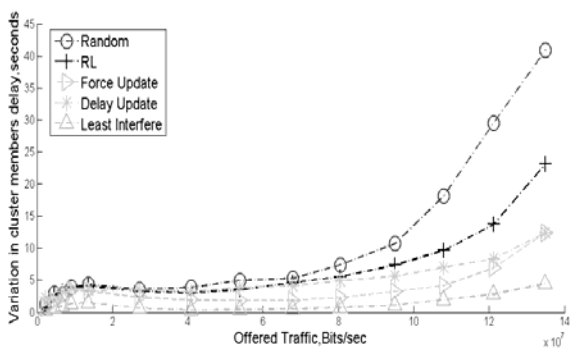

Fig-6 Variation in delay performance amongst cluster members

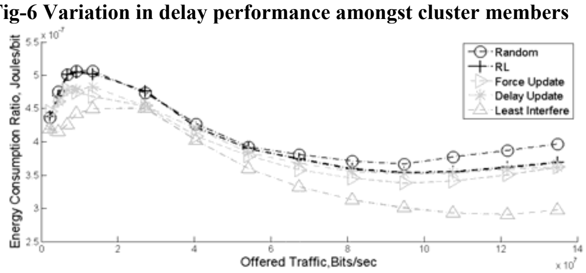

Fig-7 Clustered networks ECR performance

(Fig-7) presents the overall energy efficiency in Joules/Bit (also known as Energy Consumption Rating ECR) of clustered networks employing various channel allocation schemes. The results show that RL schemes at are able to improve the network energy efficiency compared to random channel allocation at offered traffic greater than 30Mb/s. By surprising the level of interference and improve network capacity, RL schemes are able to reduce the duration in which the cluster heads and cluster members to be transmission and reception mode which has consumes large amount of energy. The disparity in energy efficiency becomes more prominent at higher offered traffic. Result presented in **(Fig. 8)** quantify the amount of energy efficiency (Energy Reduction Gain ERG) that could be obtained. It is shown that Force Update can improve the efficiency by 10% and while the distributed RL and Delay Update have 8% in comparison with Random channel allocation.

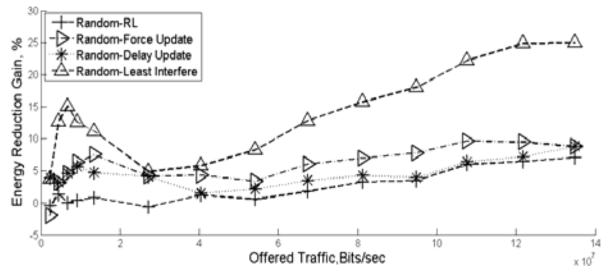

Fig-8 ERG of RL schemes against random allocation scheme.

The rate of learning of RL can be analysed by measuring the number of transmissions required for n^c to enter exploitation stage and transmit on *P*. As shown in **(Fig. 9),** 90% of n^c employing cooperative RL were able to exploit *P* at 2000-2500 transmissions whereas distributed RL require 6000 transmissions. The cooperative RL is able to achieve a learning rate of 3 times faster compared to distributed RL since it enables n^c belonging to the same cluster to cooperate and share the channel state information with clusters heads as facilitator or a teacher. Despite 100% of n^c obtaining *P* in the cooperative scheme, 5-10% of n^c re-enter exploration due to insufficient *P*.

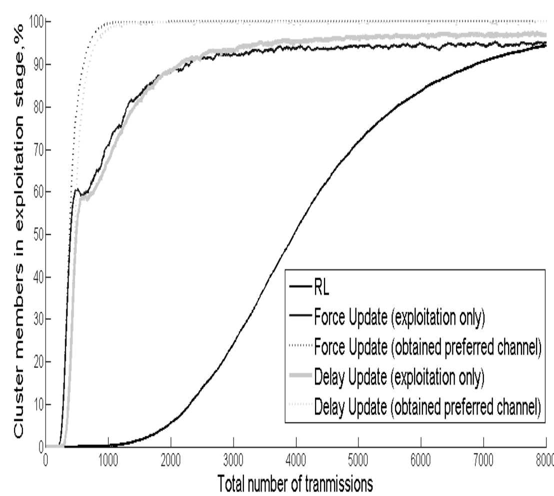

Fig-9 RL learning rate

6. CONCLUSION

This paper introduces RL channel allocation scheme as a mean to reduce the level of interference in clustered networks. The learning ability of RL enables it to learn and distinguish channels and prioritise on to channels that have high success rate. The result indicates that RL can reduce the interference level by 30%. Subsequently, the energy efficiency of the clustered networks is improved by 10%. With the ever increasing complexity and mobility of wireless communication system the extremely poor convergence time limits viability of distributed RL deployment.

Through the proposed cooperative RL channel allocation, the system can learn at a rate 3 times faster than distributed RL. Furthermore, the proposed schemes is flexible on the number of priority channels. This reduces the variation in the level of performance amongst nodes in the network and evenly distribute interference level throughout the networks making the networks more equal.

REFERENCES:

Giupponi, L., A. Galindo-Serrano, P. Blasco M. Dohler, (2010). "Docitive networks: an emerging paradigm for dynamic spectrum management [Dynamic Spectrum Management]," in IEEE Wireless Communications, vol. 17, no. 4, 47-54.

Heinzelman, W. R., A. Chandrakasan, H. Balakrishnan, (2002). "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," IEEE Transactions on Wireless Communications, vol. 1, no. 4, 660–670.

Hasan, Z., H. Boostanimehr, V. K. Bhargava, (2011). "Green Cellular Networks: A Survey, Some Research Issues and Challenges," Communications Surveys & Tutorials, IEEE, vol.13, no.4, 524,540, Fourth Quarter.

Jiang, T., (2011). Distributed Spectrum Sharing for Cognitive Radio Using Reinforcement Learning. PhD thesis, University of York.

Jiang, T., D. Grace, P. D. Mitchell, (2011). "Efficient Exploration in Reinforcement Learning Based Cognitive Radio Spectrum Sharing", IET Communications, Vol. 5, 10, 1309-17.

Moreno, P., (2011). Docitive Networks − A Step Beyond Cognition. Master thesis, Universitat Politècnica de Catalunya. "BuNGee Deliverable: D1.2, Baseline BuNGee Architecture,".

Niu, Z., Y. Wu; J. Gong, Z. Yang, (2010). "Cell zooming for cost-efficient green cellular networks," Communications Magazine, IEEE , vol. 48, no.11, 74,79.

Ramli, A., H. Basrudin, M. Yaakop, D. Grace, (2015). Energy Efficiency of a Dual Hop Clustered Networks in High Data Rate Applications, International Conference on Engineering Technologies and Entrepreneurship (ICETE-2015), Kuala Lumpur.

Ramli, A., and D. Grace, (2015). Reinforcement Learning Based Clustering Protocols for a Self-Organizing Cognitive Radio Network, Transactions on Emerging Telecommunication Technologies, Published online 7 October, doi: 10.1002/ett.2989.

Son, K., E. Oh, B. Krishnamachari, (2011). "Energy-aware hierarchical cell configuration: From deployment to operation", Proc. IEEE INFOCOM - GCN Workshop Shanghai China. 289-294.