Achieving Context Awareness and Intelligence in Cognitive Radio Networks using Reinforcement Learning for Stateful Applications

A Technical Report submitted to the School of Engineering and Computer Science, Victoria University of Wellington, New Zealand, Jan 2010
Technical Report: ECSTR10-01

Kok-Lim Alvin Yau, Peter Komisarczuk and Paul D. Teal
Communications and Networking Research Group
School of Engineering and Computer Science
Victoria University of Wellington
P.O. Box 600, Wellington 6140, New Zealand
{kok-lim.yau, peter.komisarczuk, paul.teal}@ecs.vuw.ac.nz

***Abstract*-The tremendous growth in ubiquitous low-cost wireless applications that utilize the unlicensed spectrum bands has laid increasing stress on the limited and scarce radio spectrum resources. Given that the licensed or Primary Users (PUs) are oblivious to the presence of unlicensed or Secondary Users (SUs), Cognitive Radio (CR) is a new paradigm in wireless communication that allows the SUs to detect and use the underutilized licensed spectrum opportunistically and temporarily. Context awareness and intelligence are key characteristics of CR to enable the SU to sense for and use the underutilized licensed spectrum in an efficient manner. In this technical report, we advocate the application of Reinforcement Learning (RL) for achieving context awareness and intelligence, including application schemes that require state representation, or stateful application schemes. In RL, the state encompasses the condition of the operating environment that are relevant to decision making in the application scheme. We investigate the use of RL for stateful applications with respect to Dynamic Channel Selection (DCS) scheme that helps SU Base Station (BS) to select channel adaptively for data transmission to different SU hosts in centralized static and mobile CR networks. The purpose is to enhance Quality of Service (QoS), particularly throughput and delay, and in terms of minimising number of channel switchings. Channel heterogeneity is considered in this paper. Our contribution in this paper, in comparison to our previous work, is the extension of state representation into the DCS application scheme so that the DCS is aware of the changes in the operating environment. Simulation results reveal that the proposed scheme achieves very good performance, and similar trends are observed in our previous work.**

Keywords–Cognitive radio networks; dynamic channel selection; context awareness; intelligence; reinforcement learning

I. INTRODUCTION

The rapid proliferation of low-cost wireless applications in unlicensed spectrum bands has resulted in spectrum scarcity among those bands. However, studies sponsored by the Federal Communications Commission (FCC) discovered that the current static spectrum allocation has led to overall low spectrum utilization where up to 70% of the allocated licensed spectrum remains unused (called white space) at any one time even in a crowded area [1]. Consequently, Dynamic Spectrum Access (DSA) has been proposed so that unlicensed spectrum users or Secondary Users (SU)s are allowed to use the white space of licensed users' or Primary Users (PU)s' spectrum conditional on the interference to the PU being below an acceptable level. This function is realized using Cognitive Radio (CR) technology that enables an SU to change its transmission and reception parameters including operating frequencies.

Context awareness enables each SU to be aware of its operating environment; while intelligence enables each SU to make the right decision at the right time to achieve optimum performance. Generally speaking, context awareness and intelligence enable each SU to *sense*, *learn*, and *respond* in an efficient manner with respect to its operating environment without adhering to a *strict and static* self-defined policy.

To achieve context awareness and intelligence in CR networks, this paper investigates the use of Reinforcement Learning (RL) including application schemes that require state presentation as an extension to our work in [2]. In RL, the state encompasses the condition of the operating environment that are relevant to decision making in an application scheme. More discussion about RL and its state will be provided in subsequent sections. We investigate the use of RL for stateful application with respect to Dynamic Channel Selection (DCS) scheme that helps SU Base Station (BS) to select heterogeneous channels adaptively for data transmission to different SU hosts in centralized static and mobile CR networks. By considering heterogeneous channels, the channels may have different characteristics such as transmission range, packet error rate and interference behaviours due to varying carrier frequencies, time-varying channel conditions, nodal mobility, and neighbour interference. Additionally, each channel has different levels of PU activity. Our purpose is to enhance Quality of Service (QoS), particularly throughput and delay,

in terms of number of channel switchings. Our previous work [2] assumes there are only two SUs, namely a BS and an SU host, in a network. In this paper, this assumption is relaxed with the introduction of an extra SU and the application of state to represent different SU host is necessary.

Using the RL technique, our DCS determines how an SU BS chooses its next operating channel for data transmission for different SUs during channel switching. The SU switches its channel when PU activity level becomes high or the channel quality degrades in a particular channel. The rest of the paper is organized as follows. Section II reviews RL. Section III presents the context-aware and intelligent DCS scheme. Section IV shows simulation experiments, results and discussions. Section V presents our conclusions.

II. An Overview of Reinforcement Learning

Q-learning [3], [4] is an on-line algorithm in RL that determines an optimal policy without detailed modeling of the operating environment. In Q-learning, the learnt action value or Q-value, $Q(s,a)$ indicates the appropriateness of choosing action a in state s. At time $t+1$, the Q-value of a chosen action in state s at time t is updated as follows:

$$Q_{t+1}(s_t,a_t) \leftarrow (1-\alpha)Q_t(s_t,a_t) + \alpha(r_{t+1}(s_{t+1}) + \gamma \max_{a \in A} Q_t(s_{t+1},a)) \quad (1)$$

where $0 \leqslant \alpha \leqslant 1$ is the learning rate, $0 \leqslant \gamma \leqslant 1$ is the discount factor, and $r_{t+1}(s_{t+1})$ is the delayed reward, which is the reward received at time $t+1$ for the action taken at time t. An optimal policy is being searched for that maximizes the value function $V^{\pi}(s_t)$ as shown in equation (2):

$$V^{\pi}(s_t) = \max_{a \in A}(Q_t(s_t,a)) \quad (2)$$

The update of the Q-value in (1) does not cater for the actions that are not chosen. Choosing the best overall action according to (2), or the greedy action, at all times is termed exploitation. To improve the estimates of the other Q-values, the other actions are chosen once in a while though they are not known to be the best choice. Through this, better actions may be discovered, which is a procedure called exploration. One approach to exploration is the ε-greedy approach [3], where an agent chooses the greedy action as its next action with probability $1-\varepsilon$, and random action with a small probability ε.

Applying Q-learning in DCS provides several advantages. It helps an SU to adapt to its dynamic and uncertain operating environment. Using a simple modeling approach, the complexity involved in modeling the environment and channel heterogeneity can be minimized. For instance, an SU that selects a channel for data transmission does not model the channel behavior, which is characterized by channel selective fading, path loss, PU interference and others that affect an SU's performance in a complex manner. Rather than addressing a single factor at a time, an RL agent (or decision maker) observes the relevant factors in the decision making procedure as state and optimizes a general goal as a whole, such as throughput, with regard to the state through maximizing the value function $V^{\pi}(s_t)$.

Game theory [5] has been the driving impetus for studying the interaction of multiple agents, whose objective is to maximize their individual rewards, in CR neworkst. To date, research has been focusing on static or stateless (also called single-state) one-shot or repetitive games, such as the matrix game and potential game. Using the stateless game model, the agents are not adaptive to the dynamics in the operating environment. The RL approach, which represents the operating environment as the state, is an eminently suitable solution in CR networks. This paper investigates the network performance of DCS scheme that applies RL model with state representation.

Figure 1 shows the flowchart of the RL model. Generally speaking, the RL requires the following inputs: 1) state information; 2) a set of actions; and 3) a reward function.

III. Context-Aware and Intelligent Dynamic Channel Selection Scheme

A problem arises as to what is the best strategy to select an available channel among the licensed channels for data transmission from an SU BS to different SU hosts given that the objective is to maximize overall throughput and minimizing delay, in terms of number of channel switchings, in the presence of different levels of PU utilization levels (PUL) and Packet Error Rate (PER) in the licensed channels, as well as nodal mobility. Higher levels of PUL indicates higher levels of PU activity, and hence smaller amount of white spaces. Higher levels of PER indicates higher levels of packet drop due to interference, path loss, and other factors. Due to channel heterogeneity, a channel with low PUL does not imply a good channel if it has a high PER.

The DCS scheme is modeled using the RL approach. In [2], we have shown the RL model for the DCS scheme and it is briefly described here, as shown in Table 1. The RL learning engine is embedded in the SU BS, which is the decision maker that determines the channel to use for data transmission to the SU hosts. The state S_C has one component, which is node i's neighbour nodes j, with $N_n = |Nbr(i)|$ as the cardinality of node i's neighbour nodes. Since the RL learning engine is embedded in the SU BS, node i is the BS, and node j is the SU host. The condition of the state changes with time, for instance,

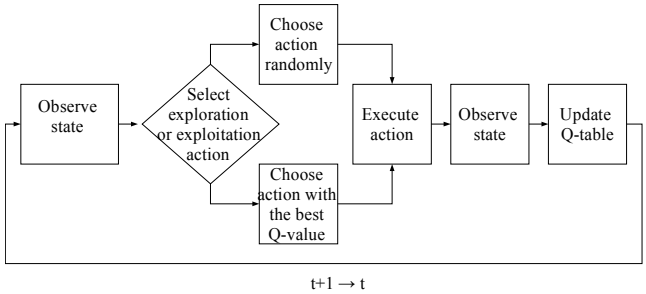

Figure 1. Flowchart of the RL model.

distance between the SU BS and SU hosts, and PER at each SUs. The probability of a successful data packet transmission is dependent on the channel PUL, PER and collision. The action A_C is to choose a channel for data transmission from K available channels set $\{a_C = c_1, c_2, \ldots, c_K\}$. For every successful data packet transmission, there is a reward with positive constant value $+RW$, otherwise a cost with negative constant value $-CT$ is incurred. In practice, the value of RW and CT are based on the amount of revenue and cost that a network operator earns or incurs for each successful or unsuccessful data packet transmission. Data packet transmission is successful when a link-layer acknowledgment is received for the data packet sent, else the transmission is unsuccessful. In addition, if a chosen channel is reoccupied by PU immediately before a data packet is to be sent, it is considered an unsuccessful transmission because the sensing mechanism detects the channel in use.

TABLE I
REINFORCEMENT LEARNING MODEL EMBEDDED IN THE SECONDARY USER BASE STATION FOR DYNAMIC CHANNEL SELECTION

RL Element	Dynamic Chanel Selection Model	
	Description	Representation
State	Set of node *i's* neighbour nodes *j*.	$S_C = \{s_C = j\}$, $j = \{n_1, n_2, .., n_{N_n}\}$
Action	Available channels for data transmission.	$A_C = \{a_C = c_1, c_2, \ldots, c_K\}$
Reward	Constant value to be rewarded/incurred for successful/unsuccessful data packet transmission.	$R_C = \{r(s,a)\} = \begin{cases} +RW & \text{if successful} \\ -CT & \text{otherwise} \end{cases}$

At each attempt to transmit a data packet, SU BS node i chooses a channel for data transmission. It keeps track of Q-value, $Q_t(a)$ for all its possible actions in a Q-table with $|A|$ entries. Equation (1) is rewritten as follows:

$$Q^i_{t+1}(s^i_t, a^i_t) \leftarrow (1-\alpha) Q^i_t(s^i_t, a^i_t) + \alpha r^i_{t+1}(s^i_t) \qquad (3)$$

with the $\max_{a \in A} Q_t(s_{t+1}, a)$ in (1) being omitted to indicate no dependency on the future discounted rewards. Note that the SU BS does not change its state or neighbour for data transmission while it is transmitting a data packet, hence $r_{t+1}(s_{t+1}) = r_{t+1}(s_t)$, and the discounted rewarded is omitted in (3). The greedy action is chosen as follows:

$$a^i_{C,t} = \underset{a^i_C \in A_C}{argmax}(Q^i_t(s^i_t, a^i_C)) \qquad (4)$$

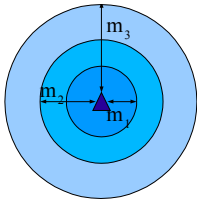

Figure 2. An SU and its transmission ranges using different channels. Another SU is located within the maximum transmission range. Each channel is licensed.

At the beginning of every attempt to transmit a data packet, the SU BS node i chooses to either continue or change an action or channel. In order to reduce the number of channel switchings, it does not switch channel unless the Q-value of the other action is better than the current one, or during exploration.

IV. SIMULATION EXPERIMENTS, RESULTS AND DISCUSSIONS

In this section, the simulation model, assumptions and its parameters are presented. Simulations were preformed using the INET framework for OMNeT++ [6]. We have implemented a CR-enabled environment in OMNeT++. Simulation parameters are shown in Table II. Both static and mobile networks are investigated.

A. Simulation Model, Assumptions and Parameters

We consider a centralized CR network with an SU BS, and $N_n = 2$ SU hosts, namely SU1 and SU2, in all scenarios, and this is sufficient to show how RL with state representation is applied to DCS. Each state represents an SU host. The condition of the state may change with time, for instance, the distance between the SU BS and SU host changes, or the PER for a particular channel at an SU host changes. The RL learning engine is embedded in the SU BS. The SU BS is static, while the SU hosts could be static or mobile.

Due to the limited sensing capability at each SU, there are K available channels. Each channel has its own characteristics including maximum transmission range m_i, PUL L_i^{PU}, and PER, P_i^E where $1 \leqslant i \leqslant K$.

We assume that for each channel, the channel utilization pattern of PU follows independent and identically distributed (i.i.d.) stochastic model. Each PU accesses one of the channels, while the SUs can access any one of the K available channels for data transmission. There are K PUs, each PU uses one channel and broadcasts packets according to the PU traffic model in Table II throughout the entire simulation area. The SUs transmit using a fixed transmission power at different channels; hence the transmission range for each channel varies as shown in Figure 2. In general, lower transmission frequency provides larger transmission range. From Table II, the values of Q-learning parameters, RW and CT, were chosen empirically to optimize simulation performance. In mobile networks, the SU nodes move in a random manner with their speed and direction following a normal distribution with mean and standard deviation as shown in Table II. In static network, the SU BS could communicate with each SU host using all $K = 3$ channels; while in mobile network,

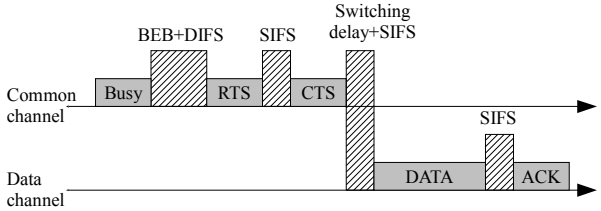

Figure 3. Illustration of cognitive MAC protocol.

some of the $K=3$ channels may be out of range, however, the SUs must be located within the maximum transmission range from the SU BS or m_3 in Figure 2.

TABLE II
NOTATIONS AND DEFAULT PARAMETER SETTINGS IN SIMULATION

Category	Symbol	Details	Values
Initialization	N	Number of SU	3 (one SU BS and two SU hosts)
	K	Number of available channels	3
	F	Center frequencies of available channels	{400MHz , 800MHz, 5.7GHz}
	P_i^E	PER of each available channel	[0.05, 0.3] Default value = 0.1
	T	Total simulation time	500s
Mac Protocol	t_{SIFS}	SIFS packet duration	10μs
	t_{DIFS}	DIFS packet duration	5μs
	t_{RTS}	RTS packet duration	272μs
	t_{CTS}	CTS packet duration	248μs
	t_{ACK}	ACK packet duration	248μs
	t_{ex}	Data packet expiration timer	5.798ms
	D	Data rate	2Mbps
Mobile Networks	μ_S	Mean of speed	20m/s
	σ_S	Standard deviation of speed	8m/s
Secondary user		Secondary user traffic model	Always backlogged.
	$t_{DATA,SU}$	Data packet duration	5.44ms
	$T_{SU,w}$	Switching delay	100μs
Primary user		Primary user traffic model	Stochastic channels with exponentially distributed ON and OFF times.
	$t_{DATA,PU}$	Data packet duration	5.44ms
	L_i^{PU}	PUL of each PU at each available channel	[0.1, 0.9] Default value = 0.1
Q-learning	α	Learning rate of Q-learning	0.2
	ε	Trade-off between exploration and exploitation	0.1
		Initial Q-value	1
	RW	Reward	15
	CT	Cost	5

An illustration of the cognitive MAC is shown in Figure 3. In the figure, switching delay may be ignored if channel switching is not necessary. For the cognitive MAC to operate, each SU has two transceivers, one is tuned to a common control channel, which is free of PU, at all times for control message exchange; while the other one is tuned to any other available channels for data transmissions. The common control channel has the largest transmission range compared to the available channels for data transmission. Since the last spectrum sensing indicates the PU occupancy in a particular channel, the channel, which was free, may become busy within a Short Inter-Frame Spacing (SIFS) interval right immediately prior to data transmission. In this case, the SU restarts its data transmission cycle again with RTS-CTS handshaking, and may reassign its channel. The RTS has channel switching information determined by the SU BS according to its Q-table. Synchronization between PUs and SUs, as well as among the SUs, are not necessary. The MAC protocol timing is shown in Table II. In the table, t_{ex} is the data packet expiration timer, which is initiated after sending a data packet and is reset upon receiving its ACK packet. The expiration of the t_{ex} indicates failed data transmission.

B. Performance Metrics

Our goal is to maximize throughput and minimize number of channel switchings, which causes non-negligible delay for data transmission, over different heterogeneous channels with different transmission range, PUL and PER. The mean amount of throughput and number of channel switchings of the RL-based DCS is compared with that of Random DCS where an available channel is chosen for next data transmission in a random manner. Graphs are presented with PUL and PER as ordinate respectively. When PUL is ordinate, each simulation result of mean throughput or mean number of channel switchings is for all possible combinations of PUL with the other parameter values remain constant as shown in Table II. As an example, a PUL of 0.8 for $K=3$ available channels may indicate [0.8,0.8,0.8], [0.8,0.7,0.9], and [0.9,0.9,0.6]. In the case of mobile networks, each set of PUL such as [0.8,0.7,0.9] is applied to various channels with different frequencies for each simulation run. When PER is ordinate, since the range is small within [0.05,0.3], each simulation result of mean throughput or mean number of channel switchings is average value of 50 runs using different levels of PERs, which are generated randomly, across the channels. For instance, a PUL level of 0.2 may indicate the PUL of [0.025,0.248,0.327] or [0.163,0.402,0.035] in the channels.

C. Simulation Results and Discussions

Four types of networks are simulated: static network with identical PER at each SU (NW1); mobile network with identical PER at each SU (NW2); static network with non-identical PER at each SU (NW3); and mobile network with non-identical PER at each SU (NW4). In NW1, both SU1 and SU2 observe the similar PER and PUL for all channels, hence the SU BS would perceive similar network performance for using a particular channel to transmit to both SUs. However, in NW2, NW3 and NW4, both SU1 and SU2 observe non-identical PER or some channels could not be reached from the SU BS due to their different distance from the SU BS, hence the SU BS would perceive different network performance for using a particular channel to transmit to both SUs. We first compare the performance

between RL and Random; followed by investigation into the effects of RL parameters, namely α and ε, on the network performance.

Figure 4 shows the throughput achieved by SU1 and SU2 using the RL and Random scheme for various levels of PUL. The PER for all channels are set to 0.1 to show the effectiveness of the RL method in choosing a channel with low level of PUL for data transmission. The RL scheme outperforms the Random for all levels of PUL. Both SU1 and SU2 achieve approximately the similar individual network performance. Throughput enhancement provided by the RL scheme is up to 2.3 times at 0.8 PUL in a static network; while it is up to 3.2 times at 0.8 PUL in a mobile network. Hence, the RL scheme learns well and helps the SU BS to choose a channel with low PUL such that the successful packet transmission rate is high, hence providing higher throughput. At 0.1 PUL, performance improvement provided by the RL scheme is not significant in a static network. This is due to small differences among the Q-values across the channels, or less differences in PUL across the available channels. However, this is not the case at 0.1 PUL in a mobile network where the RL outperforms the Random up to 1.7 times. This is because the RL helps the SU to choose the channels with suitable transmission range for data transmission.

Figure 5 shows the number of channel switchings achieved by SU1 and SU2 using the RL and Random scheme for various levels of PUL. The PER for all channels are set to 0.1. The RL scheme outperforms the Random for all levels of PUL to provide lower number of channel switchings. Both SU1 and SU2 achieve approximately the similar individual network performance. The RL scheme attains rather stable number of channel switchings because the ε is kept constant at 0.1 throughout the simulation; however the number of channel switchings increases at 0.7 PUL in both static and mobile networks. The effect of the number of channel switchings does not affect the throughput significantly in this simulation due to the small 100μs channel switching delay; however, this is dependent on the hardware performance in practice. For the Random, the number of channel switchings decreases with PUL, indicating decreasing number of attempts for the SU BS to transmit data packets. The reason is that failed data packet transmission incurs longer delay while waiting for data packet expiration timer t_{ex} to expire; and this happens more often with increasing PUL. The RL outperforms the Random up to 4.2 times at 0.5 PUL in a static network and up to 4.3 times at 0.3 PUL in a mobile network. For the RL scheme, the number of channel switchings is lower in a mobile network compared to a static network. The reason is that, as the SU nodes move further apart from each other, the number of channels that fulfill the transmission range requirement decreases, hence the occurrence of channel switching reduces. For the Random, the number of channel switchings is higher for a static network compared to a mobile network, indicating a larger number of attempts to transmit data packets by the SU sender in a static network. With a smaller number of channel switchings, the RL incurs less delay.

Figure 6 and 7 shows the equivalent Figure 4 and 5 respectively with linear combination of all the local performance at SU1 and SU2 to provide mean network-wide performance. Only network-wide performance is shown henceforth due to the similarity among nodal performance at SU1 and SU2.

Figure 8 shows the network-wide throughput using the RL and Random scheme for various levels of PER. The PUL for all channels are set to 0.1 to show the effectiveness of the RL method in choosing a channel with low level of PER for data transmission. The RL scheme outperforms the Random for all levels of PER. Network-wide throughput enhancement provided by the RL scheme is up to 1.1 times at 0.15 PER in a static network; while it is up to 1.6 times at 0.05 to 0.2 PER in a mobile network. Hence, the RL scheme learns well and helps the SU BS to choose a channel with low PER such that the successful packet transmission rate is high, hence providing higher throughput.

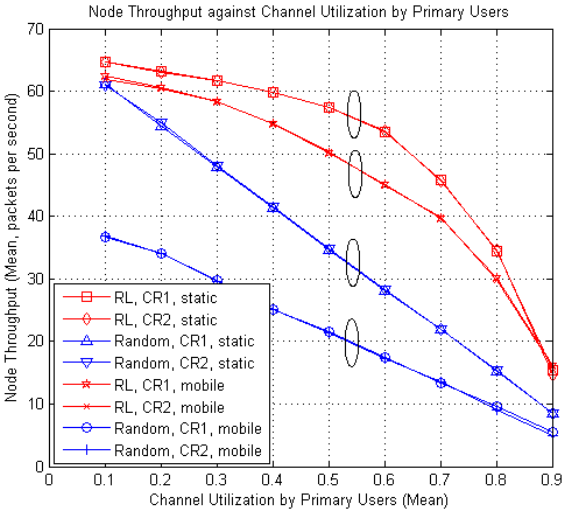

Figure 4. The mean throughput at each CR node against mean of channel utilization by PU for RL in static and mobile network. PER for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

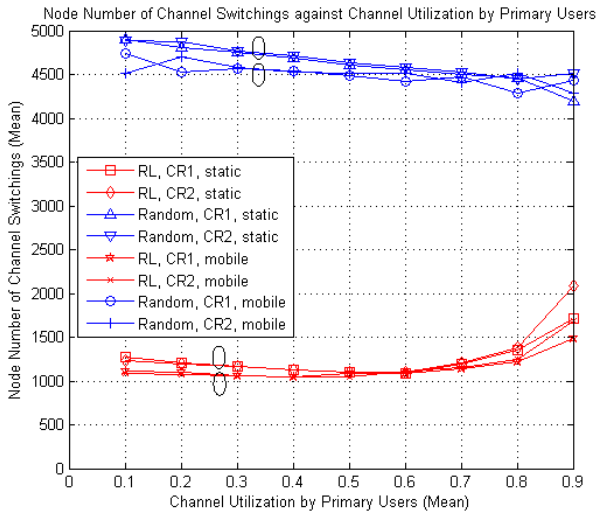

Figure 5. The mean number of channel switchings at each CR node against mean of channel utilization by PU for RL in static and mobile network. PER for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

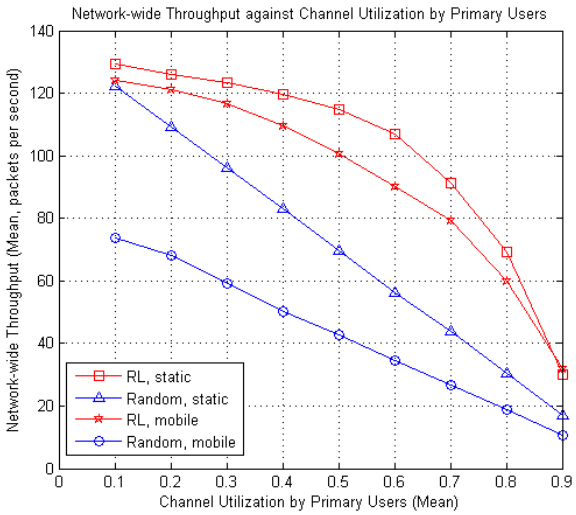

Figure 6. The mean network-wide throughput against mean of channel utilization by PU for RL in static and mobile network. PER for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

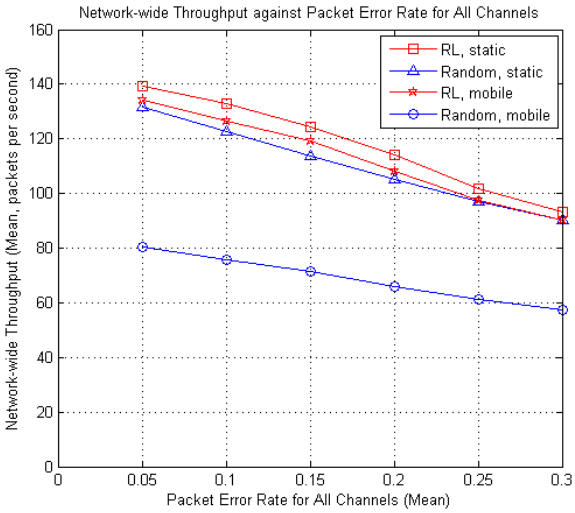

Figure 8. The mean network-wide throughput against mean of PER for all channels for RL in static and mobile network. PUL for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

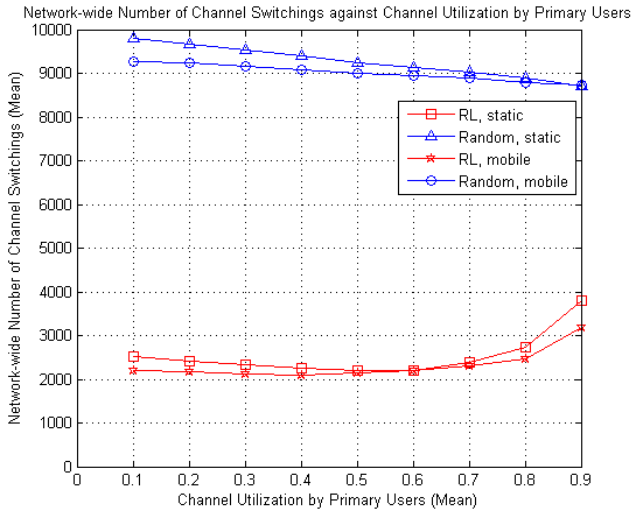

Figure 7. The mean network-wide number of channel switchings against mean of channel utilization by PU for RL in static and mobile network. PER for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

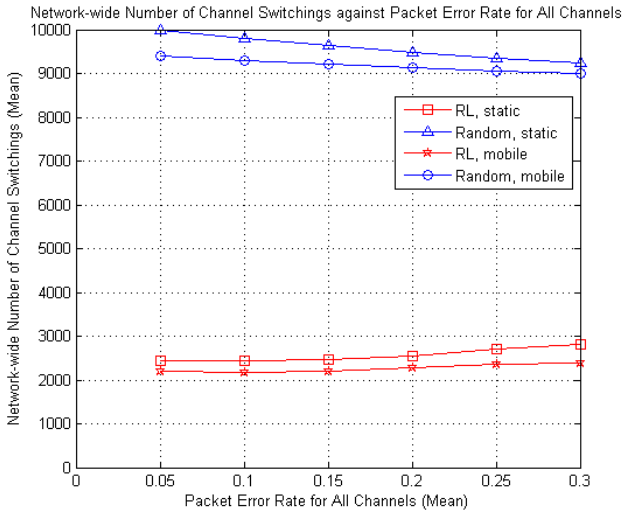

Figure 9. The mean network-wide number of channel switchings against mean of PER for all channels for RL in static and mobile network. PUL for all channels are set to 0.1; For RL, α is set to 0.2; and ε is set to 0.1.

Figure 9 shows the network-wide number of channel switchings using the RL and Random scheme for various levels of PER. The PUL for all channels are set to 0.1. The RL scheme outperforms the Random for all levels of PUL to provide lower number of channel switchings. The RL scheme attains rather stable number of channel switchings because the ε is kept constant at 0.1 throughout the simulation. The Random shares the similar trend in Figure 7. The RL outperforms the Random up to 4.0 times at 0.05 and 0.1 PER in a static network and up to 4.2 times at 0.05 to 0.15 PER in a mobile network. With a smaller number of channel switchings, the RL incurs less delay.

The next four subsections show the effects of RL parameters including α and ε on network-wide performance in the four types of networks.

Effects of α and ε on Static Network with PUL as Ordinate (NW1). The throughput and number of channel switchings achieved by the RL scheme are investigated for various levels of PUL in static network. The PER for all channels are set to 0.1. With PUL as ordinate, Figure 10 shows the effect of α on throughput; Figure 11 shows the effect of α on number of channel switchings; Figure 12 shows the effect of ε on throughput; and Figure 13 shows the effect of ε on number of channel switchings. In Figure 10, the value of α does not have significant effect on throughput. In Figure 11, for each α, the number of channel switchings reaches the lowest value at about 0.6 PUL because the standard deviation between the Q-values is higher at 0.6 PUL. The standard deviation for the PUL is best explained using an example. At 0.2, the PULs across

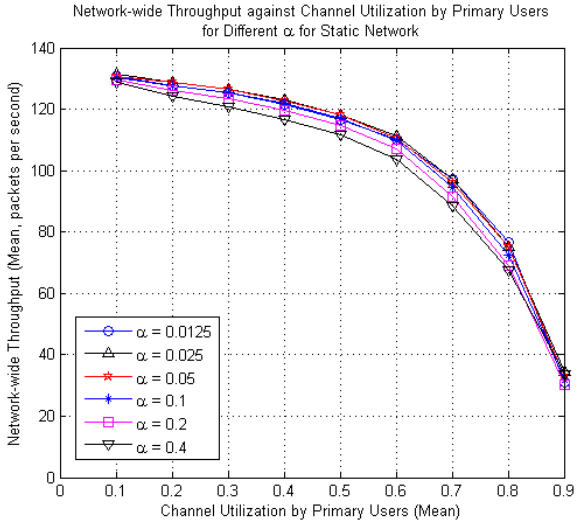

Figure 10. NW1. The mean network-wide throughput against mean of channel utilization by PU for RL with different α values in static network. PER for all channels are set to 0.1; ε is set to 0.1.

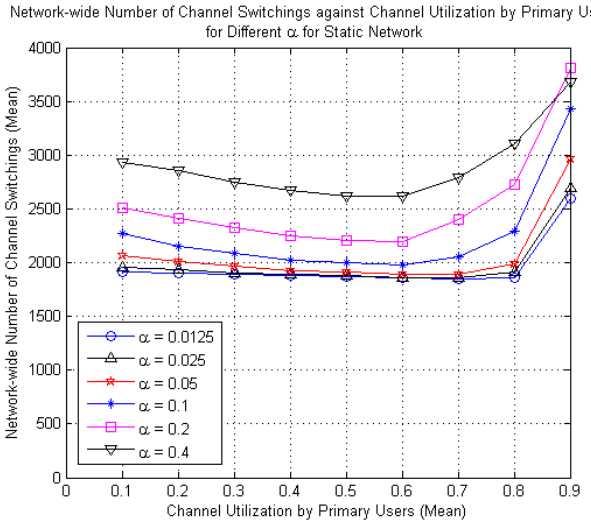

Figure 11. NW1. The mean network-wide number of channel switchings against mean of channel utilization by PU for RL with different α values in static network. PER for all channels are set to 0.1; ε is set to 0.1.

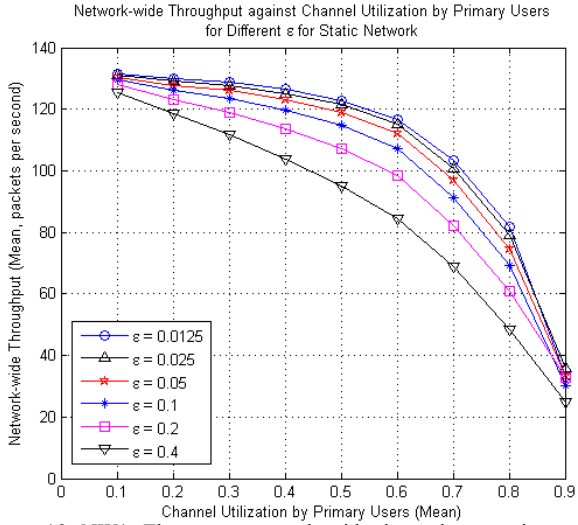

Figure 12. NW1. The mean network-wide throughput against mean of channel utilization by PU for RL with different ε values in static network. PER for all channels are set to 0.1; α is set to 0.2.

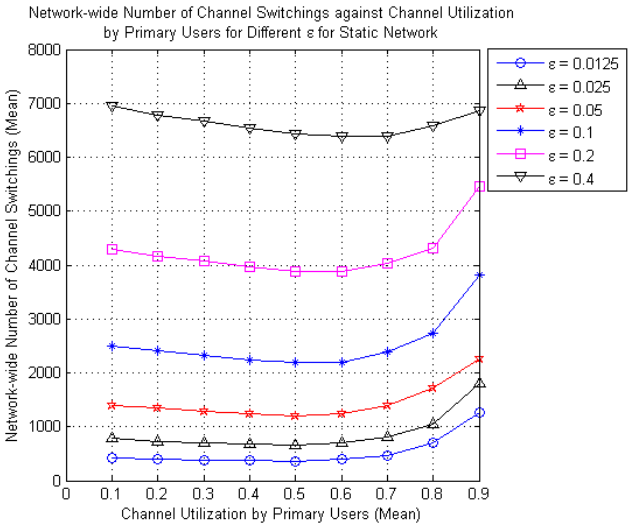

Figure 13. NW1. The mean network-wide number of channel switchings against mean of channel utilization by PU for RL with different ε values in static network. PER for all channels are set to 0.1; α is set to 0.2.

the three channels, sorted by increasing standard deviation, are [0.2,0.2,0.2], [0.2,0.3,0.1], [0.3,0,0.3], [0.4,0.1,0.1], [0.2,0,0.4], [0.5,0,0.1], and [0.6,0,0]. At 0.6, higher standard deviation is possible, for instance, [0,0.9,0.9]. Higher standard deviation of PUL leads to more obvious choice of channel selection, for instance, the SU BS chooses channel 1 with no PU activity when the PUL across the channels is [0,0.9,0.9]. In general, lower value of α provides lower number of channel switchings in this scenario. In Figure 12, the throughput increases as the ε converges to the lowest value or the least exploration. In Figure 13, the number of channel switchings shares the similar trend in Figure 11, though the ε results in larger range. Thus, the ε has greater effect on network performance compared to α.

Effects of α and ε on Mobile Network with PUL as Ordinate (NW2). The throughput and number of channel switchings achieved by the RL scheme are investigated for various levels of PUL in mobile network. The PER for all channels are set to 0.1. With PUL as ordinate, Figure 14 shows the effect of α on throughput, Figure 15 shows the effect of α on number of channel switchings, Figure 16 shows the effect of ε on throughput, and Figure 17 shows the effect of ε on number of channel switchings. Similar trends were observed in the case of static network in Figure 10-13 although lower throughput and number of channel switchings are observed in mobile networks.

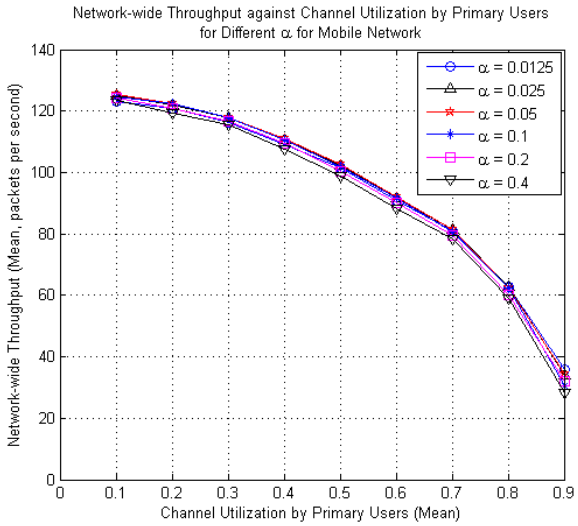

Figure 14. NW2. The mean network-wide throughput against mean of channel utilization by PU for RL with different α values in mobile network. PER for all channels are set to 0.1; ε is set to 0.1.

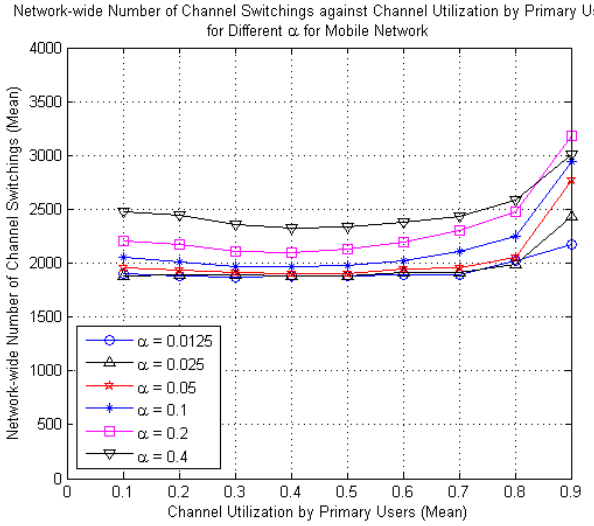

Figure 15. NW2. The mean network-wide number of channel switchings against mean of channel utilization by PU for RL with different α values in mobile network. PER for all channels are set to 0.1; ε is set to 0.1.

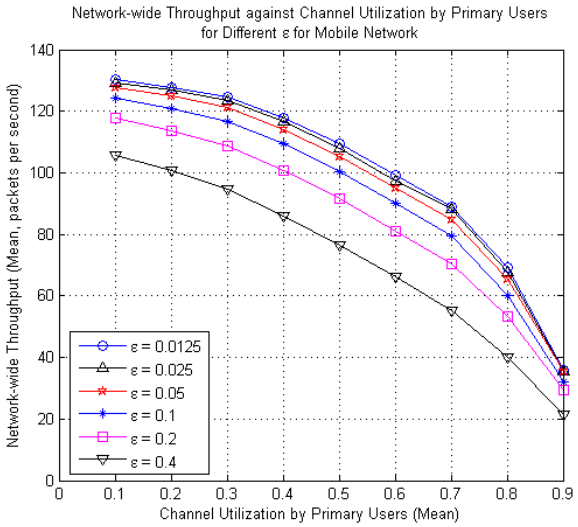

Figure 16. NW2. The mean network-wide throughput against mean of channel utilization by PU for RL with different ε values in mobile network. PER for all channels are set to 0.1; α is set to 0.2.

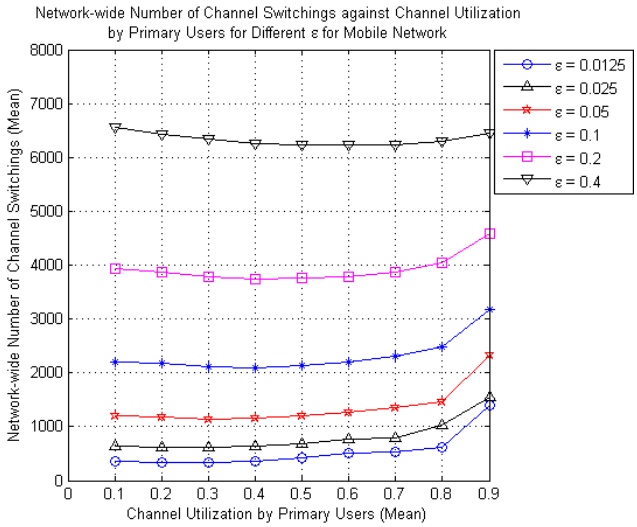

Figure 17. NW2. The mean network-wide number of channel switchings against mean of channel utilization by PU for RL with different ε values in mobile network. PER for all channels are set to 0.1; α is set to 0.2.

Effects of α and ε on Static Network with PER as Ordinate (NW3). The throughput and number of channel switchings achieved by the RL scheme are investigated for various levels of PER in static network. The PUL for all channels are set to 0.1. With PER as ordinate, Figure 18 shows the effect of α on throughput, Figure 19 shows the effect of α on number of channel switchings, Figure 20 shows the effect of ε on throughput, and Figure 21 shows the effect of ε on number of channel switchings. Similar trends were observed in the case of static network in Figure 10-13.

Effects of α and ε on Mobile Network with PER as Ordinate (NW4). The throughput and number of channel switchings achieved by the RL scheme are investigated for various levels of PER in mobile network. The PUL for all channels are set to 0.1. With PER as ordinate, Figure 22 shows the effect of α on throughput, Figure 23 shows the effect of α on number of channel switchings, Figure 24 shows the effect of ε on throughput, and Figure 25 shows the effect of ε on number of channel switchings. Similar trends were observed in the case of static network in Figure 18-21 although lower throughput and number of channel switchings.

Conclusions

In this paper, we advocate the use of Reinforcement Learning (RL) to achieve context awareness and intelligence in static and mobile CR networks including stateful applications. In RL, the state represents the condition of the operating environment. The RL is applied in a Dynamic

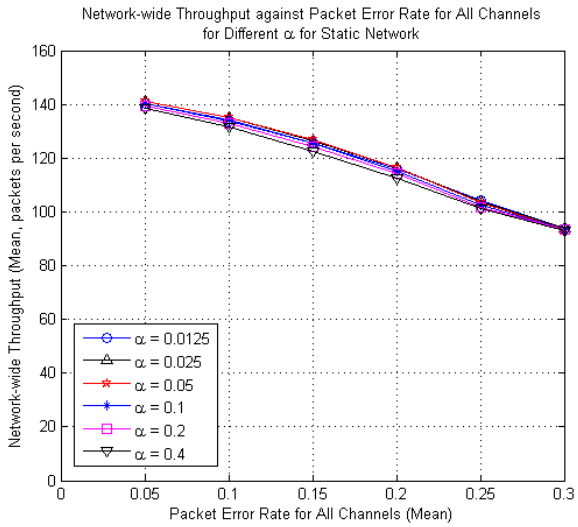

Figure 18. NW3. The mean network-wide throughput against mean of PER for all channels for RL with different α values in static network. PUL for all channels are set to 0.1; ε is set to 0.1.

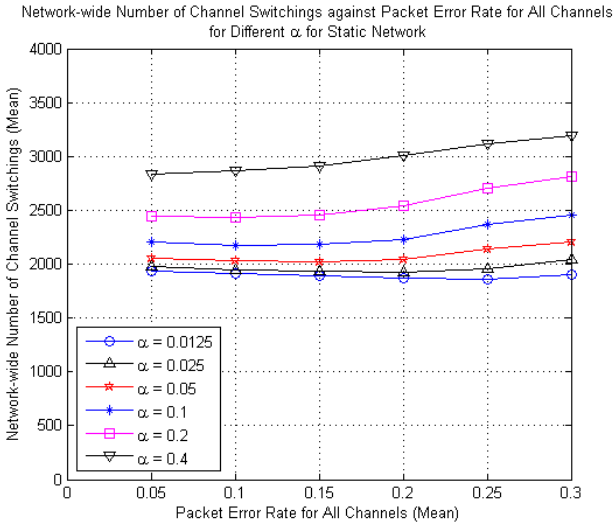

Figure 19. NW3. The mean network-wide number of channel switchings against mean of PER for all channels for RL with different α values in static network. PUL for all channels are set to 0.1; ε is set to 0.1.

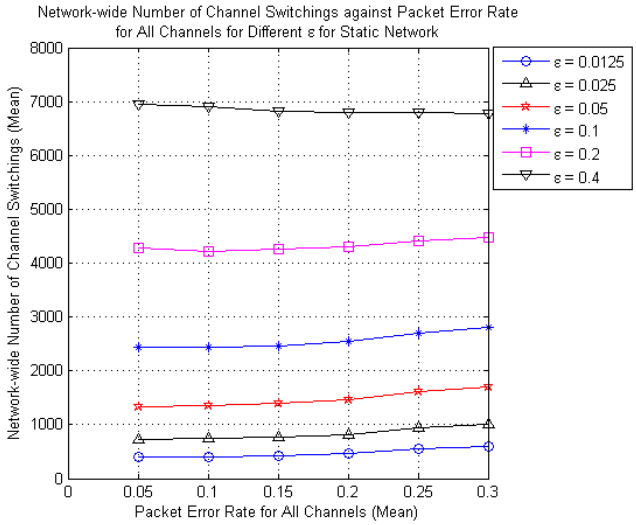

Figure 20. NW3. The mean network-wide throughput against mean of PER for all channels for RL with different ε values in static network. PUL for all channels are set to 0.1; α is set to 0.2.

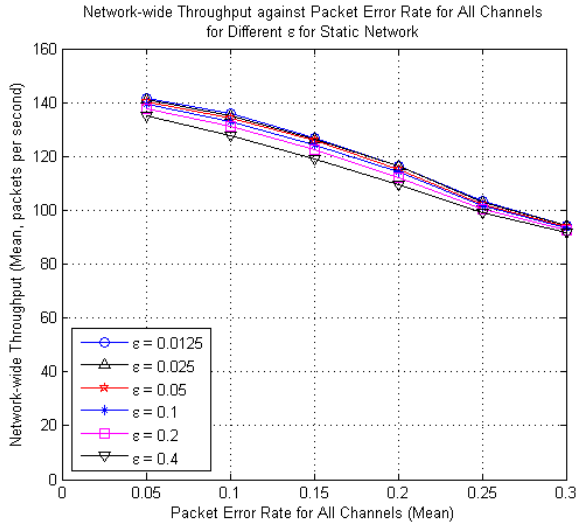

Figure 21. NW3. The mean network-wide number of channel switchings against mean of PER for all channels for RL with different ε values in static network. PUL for all channels are set to 0.1; α is set to 0.2.

Channel Selection (DCS) scheme for centralized CR network comprised of a single Secondary User (SU) base station and two SU hosts. The state represents the SU hosts and its condition changes with the distance between the SU Base Stateion (BS) and SU hosts, as well as the channel quality or Packet Error Rate (PER) at each SU. The RL has proven its superiority to the Random approach in most of the cases. The Random approach chooses an available channel for data transmission in a uniformly distributed random manner. The effects of RL parameters on network-wide performance were also investigated including the learning rate α and tradeoff between exploration and exploitation ε. Generally speaking, the throughput and number of channel switchings achieves its optimal performance when α and ε converge to a lower value; and that ε has greater effects on network-wide performance than does α.

References

[1] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," *Federal Communications Commission, Technical Report 02-155*, November 2002.

[2] K.-L. A. Yau, P. Komisarczuk, and P. D. Teal, "A context-aware and intelligent dynamic channel selection scheme for cognitive radio networks," *Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'09)*, June 2009.

[3] R. S. Sutton and A. G. Barto, Reinforcement learning: an introduction. Cambridge MA, MIT Press, 1998.

[4] C. Watkins, "Learning from delayed rewards," PhD thesis, University of Cambridge, UK, 1989.

[5] Z. Ji, and K. J. R. Liu, "Dynamic spectrum sharing: a game theoretical overview," *IEEE Com. Mg.*, 45(5), pp. 88-94, May 2007.

[6] INET Framework for OMNet++/OMNEST release 2006-10-12. http://www.omnetpp.org/doc/INET/.

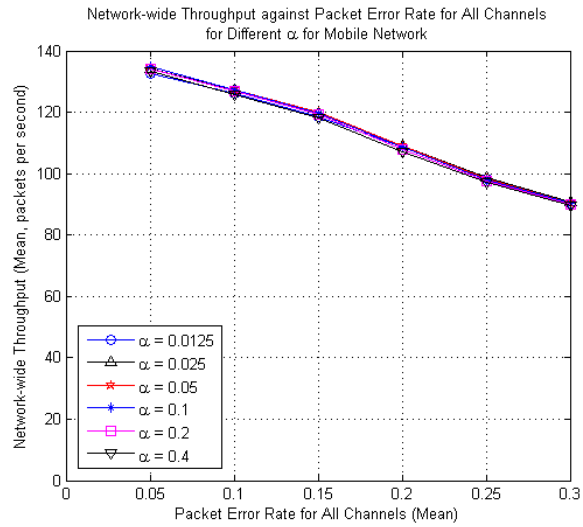

Figure 22. NW4. The mean network-wide throughput against mean of PER for all channels for RL with different α values in mobile network. PUL for all channels are set to 0.1; ε is set to 0.1.

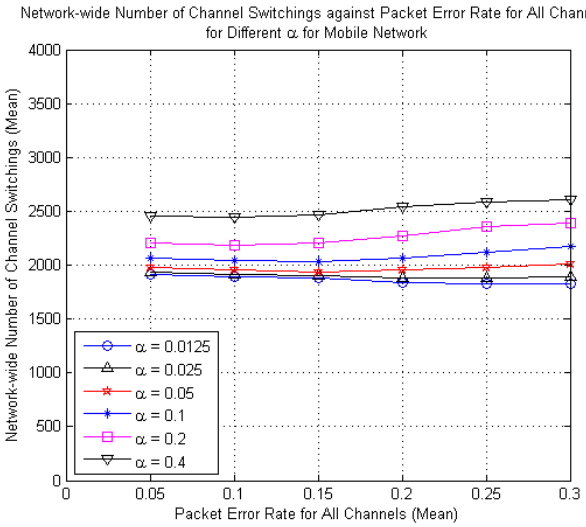

Figure 23. NW4. The mean network-wide number of channel switchings against mean of PER for all channels for RL with different α values in static network. PUL for all channels are set to 0.1; ε is set to 0.1.

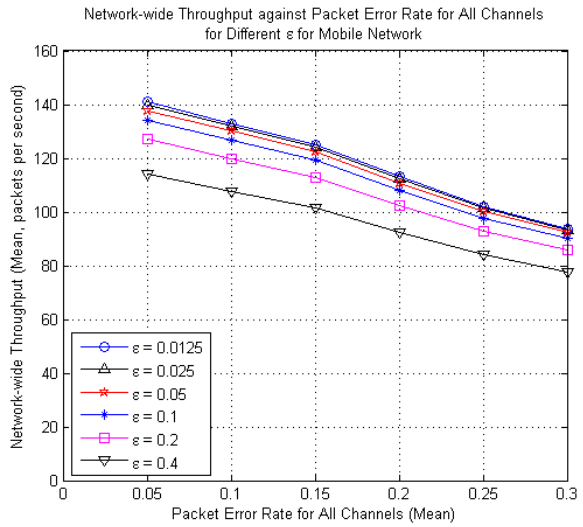

Figure 24. NW4. The mean network-wide throughput against mean of PER for all channels for RL with different ε values in mobile network. PUL for all channels are set to 0.1; α is set to 0.2.

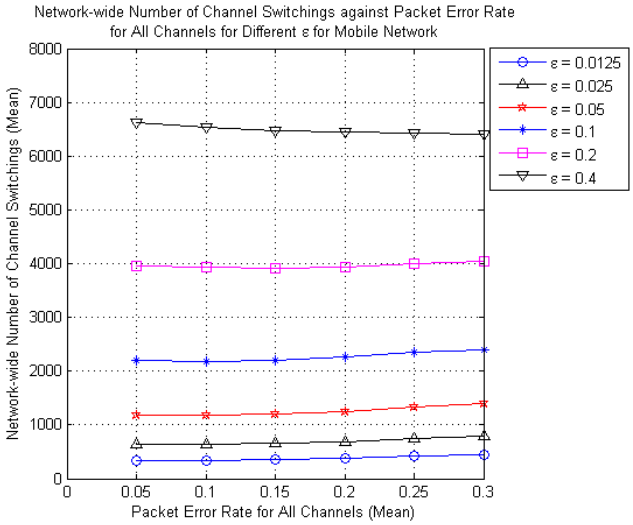

Figure 25. NW4. The mean network-wide number of channel switchings against mean of PER for all channels for RL with different ε values in mobile network. PUL for all channels are set to 0.1; α is set to 0.2.