Reinforcement Learning-Based Trust and Reputation Model for Spectrum Leasing in Cognitive Radio Networks

Mee Hong Ling
Computer Science and Networked System
Sunway University
Petaling Jaya, Malaysia
Email: mhling@sunway.edu.my

Kok-Lim Alvin Yau
Computer Science and Networked System
Sunway University
Petaling Jaya, Malaysia
Email: koklim@sunway.edu.my

***Abstract*—Cognitive Radio (CR), which is the next generation wireless communication system, enables unlicensed users or Secondary Users (SUs) to exploit underutilized spectrum (called white spaces) owned by the licensed users or Primary Users (PUs) so that bandwidth availability improves at the SUs, which helps to improve the overall spectrum utilization. Collaboration, which has been adopted in various schemes such distributed channel sensing and channel access, is an intrinsic characteristic of CR to improve network performance. However, the requirement to collaborate has inevitably open doors to various forms of attacks by malicious SUs, and this can be addressed using Trust and Reputation Management (TRM). Generally speaking, TRM detects malicious SUs including honest SUs that turn malicious. To achieve a more efficient detection, we advocate the use of Reinforcement Learning (RL), which is known to be flexible and adaptable to the changes in operating environment in order to achieve optimal network performance. Its ability to learn and re-learn throughout the duration of its existence provides intelligence to the proposed TRM model, and so the focus on RL-based TRM model in this paper. Our preliminary results show that the detection performance of RL-based TRM model has an improvement of 15% over the traditional TRM in a centralized cognitive radio network. The investigation in the paper serves as an important foundation for future work in this research field.**

Keywords—Security; trust; reputation; reinforcement learning; cognitive radio

I. INTRODUCTION

With the advent of wireless communication, the demand for radio spectrum has placed great challenges on the traditional spectrum allocation policy, in which the licensed spectrum is mostly underutilized. Cognitive Radio (CR) [1] enables unlicensed users or Secondary Users (SUs) to sense for and opportunistically utilize white spaces without interfering with the licensed users or Primary Users' (PUs') activities. PUs' activities may or may not exist in underutilized channels at a particular time instance. In the absence of PUs' activities, a block of radio resource (e.g. a transmission opportunity) in an underutilized channel is regarded as a white space. Hence, CR promotes flexibility of channel access through reconfiguration of transmission parameters, particularly the operating channels.

An intrinsic characteristic of CR is collaboration, in which SUs collaborate with each other (e.g. message exchange) to improve network-wide performance. For instance, collaborative channel sensing enables SUs to collaborate with each other through exchanges of channel sensing outcomes in order to improve accuracy of the detection of white spaces. The inaccuracy of channel sensing outcomes is caused by multipath and shadowing, and so a very robust and accurate sensing capability on each SU is necessary if collaboration is not implemented. Hence collaboration helps to achieve robustness without imposing radical requirements on individual SU [3]. However, collaboration has inevitably laid SUs open to attacks. In Cognitive Radio Networks (CRNs), each SU may be potentially malicious. A single false sensing outcome from a malicious SU to a SU Base Station (SU BS), which serves as a fusion center or to clusterhead respectively, may cause inaccuracy in the final decision. Consequently, SU network performance will be affected. Moreover, there are many types of SUs competing to use the underutilized channels, including honest, faulty, selfish, and malicious SUs, as well as SUs that launch collusion attacks. The faulty SUs are malfunctioning devices, or they may be located in fading or shadowing zones, and so they send inaccurate sensing outcomes to the fusion center or their respective neighbors. The selfish and malicious SUs are motivated by their specific intentions to monopolize and interfere with either the SUs or the PUs, respectively. The SUs that launch collusion attacks are either malicious or selfish, and they jointly launch attacks to interfere with PUs or break the rules of CR for either selfish gain or malicious intention.

To address the above problem, Trust and Reputation Management (TRM) has been adopted to identify the misbehaving SUs (e.g. faulty, selfish and malicious) among collaborating SUs. Reinforcement Learning (RL) is applied to TRM in order to dynamically tackle honest SUs behavior which may turn malicious (e.g. the SUs may not relay packets for PUs as agreed) as time progresses. In [14], RL based auction algorithm for dynamic spectrum access has been applied to CRNs. However, to the best of our knowledge, RL-

978-1-4799-2845-3/13/$31.00 ©2013 IEEE

based TRM has not been applied to spectrum leasing. In this paper, we propose a novel RL-based TRM to be applied to spectrum leasing in CRNs, specifically Q-learning algorithm which aims to increase the detection efficiency of malicious SUs. The organization of this article is as follows. Section II presents related work of Reinforcement Learning (RL) and TRM and its application in CRNs. Section III discusses the proposed RL-based TRM model and simulation results. Section IV provides conclusion, and Section V presents future work.

II. RELATED WORK

A. Reinforcement Learning

RL is an unsupervised artificial intelligent approach that enables an agent to observe and learn about the static or dynamic operating environment in the absence of guidance, feedback or the expected response from supervisors or experienced entities, and subsequently make decisions on action selection in order to achieve optimal or near-optimal system performance.

The following sub-sections provide an overview of RL and its application to security.

1) Q-learning and its representations

Q-learning [12,13], is a single-agent on-line algorithm in RL. The on-line learning is real-time and continues through the entire life of an agent where it observes, learns and acts simultaneously. Fig. 1 shows an abstract view of RL in a centralized setting. RL consists of three main elements, namely *state*, *action* and *reward*. The *state*, which is observed from the operating environment, represents factors that affect the way in which an agent makes a decision. The *reward* represents the performance metrics to be maximized such as higher detection rate. The *action* represents a choice taken by an agent in order to maximize its reward.

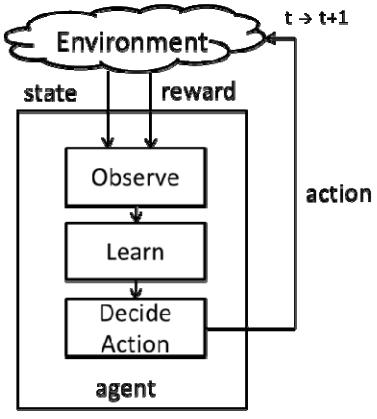

Fig. 1. Abstract view of a RL agent in its operating environment

2) Exploration and exploitation

One of the intrinsic characteristics of RL is the ability to explore and exploit an operating environment. To explore, an agent takes random actions. This may result in gaining lesser reward, but on the other hand, it may also lead to discovering an action that yields better rewards in the future. To exploit, the agent will continuously choose an action that has been found to provide the highest reward based on its past experience. To achieve an optimal or near-optimal performance, a balanced of trade-off between exploration and exploitation is required. Hence, RL method is suitable for any problem that requires an agent to learn and re-learn from an uncertain or changing operating environment in order to achieve a given goal.

3) Reinforcement Learning for Security Enhancements

The following are some of the benefits of applying RL to TRM in order to enhance security in CRNs:

- RL enables a node to learn and adapt to its dynamic and uncertain operating environment. For example, an honest or collaborating SU may turn malicious as time progresses. RL is capable of identifying malicious SUs in such operating environment [15].
- RL uses a simple modeling approach that reduces the complexity involved in the modeling of operating environment [17]. For instance, in CRN, a SU that selects its honest neighbors for collaboration would only need to either observe its neighbors' performance and behavior or jointly observe with other SUs, and exchanges its sensing outcome in order to compute a reputation value of its neighbors. The simplicity of this approach enables SU to focus on the actual subject of interest, in this case the malicious SUs. In our paper, we use state $s_{j,t}^{i}$ to represent the reputation value of the SU neighbor j.
- RL enables a node to tackle its security problem in which the solution is dependent on a series of decisions, with the notion of maximizing discounted reward [17]. This can be a more effective approach especially when the malicious SUs' behavior is dynamic. In our paper, we use reward $r_{j,t+1}^{i}$ to represent a gain when SU neighbor j relays packets to the correct destination.

RL schemes have been developed to tackle the following security challenges.

- *Dynamicity of the nodes' behavior*. Nodes may change from honest to malicious as time progresses, and vice-versa. RL model with suitability values for TRM has been applied to provide a dynamic and intelligent mechanism to monitor the nodes' behavior at all times [18].
- *Dynamicity of attack strategies*. Malicious nodes' strategies may change dynamically and it is a challenge to keep track of the way (e.g. frequency) they attack. For instance, the malicious nodes may continually change their jamming strategies causing the honest nodes to dynamically and strategically change their channel access in order to avoid the jammed channels. RL has been applied to provide a dynamic and intelligent mechanism to learn attackers' strategies as time progresses [20,21].

B. *Trust and reputation management in cognitive radio networks*

TRM for CRNs is a framework to identify malicious SUs and verify data authenticity. The trust of an entity represents its reputation value, which is calculated based on the entity's action or information (e.g. the expected delivery of packets for PUs as agreed) in different time period. Higher reputation value indicates greater trust of the entity among its community. The main objective of TRM is to promote trust amongst the SUs in order to lessen the negative impact of mistrust SUs. Specifically, it detects mistrust SUs or false sensing outcomes, and reduces the effects of mistrust, such as false positive and false negative. TRM aims to ameliorate the effects of attacks on tasks that require collaboration among entities through the detection of malicious SUs or manipulated information.

The TRM schemes in CRNs involve the detection of various kinds of entities (e.g. malicious SUs that generate false sensing outcomes, and those that interfere with PUs), and events (e.g. PUs existence). For instance, TRM schemes in [4]-[10] detects malicious SUs that intentionally or unintentionally generate false sensing outcomes by ignoring their false sensing outcomes, which may be used by a fusion center to make final decisions, while [11] detect malicious SUs that disobey the channel access rules and continue to access the channels in the presence of PUs activities.

C. *Application of RL-based TRM model in Cogntive Radio Networks*

To facilitate the opportunistic access of white spaces, CRNs offers Dynamic Spectrum Access (DSA) [22], which enables SUs to intelligently share local available channels among themselves based on their instantaneous demands. While DSA helps to improve resource utilization without regulating spectrum demand and access behaviors, it may face significant degradation in access reliability [23]. To ameliorate such negative effects, a DSA approach called spectrum leasing has been proposed [24]. This approach allows PUs to temporarily transfer and trade the licensed spectrum usage rights of the white spaces to SUs in exchange for monetary compensation [25-27] or packets relaying [28, 29]. To the best of our knowledge, the security features in spectrum leasing in CRNs is lacking and further security enhancements would need to be developed in order to provide a better measure of security.

While attacks on CRNs can be perceived in many forms such as unintentional, intentional, single, and collusion attacks, it is worth to note that all these attacks are operating in a dynamic environment. To increase the defense in spectrum leasing, TRM would need to be intelligent enough to learn and re-learn from the dynamic environment in order to be always ahead of the attackers so as to counter their attacks. This paper presents a preliminary investigation on the application of RL-based TRM to spectrum leasing in CRNs in order to enhance its security measures.

III. PROPOSED RL-BASED TRM MODEL FOR SPECTRUM LEASING

In CRNs, SUs can interact with PUs to negotiate and collaborate on leasing the licensed channels in return for channel access. The objective of the proposed model is to improve PUs transmission rate, throughput, end-to-end delay performances and energy efficiency through selecting the honest SUs to collaborate. In our model, SUs form an alternative route and offer their services as an intermediate relay node in order to enhance PUs. To reciprocate, SUs can 'piggy back' some of its own data while acting as a relay.

We model the environment of spectrum leasing via auction mechanism as follows:

- Step 1: The PU BS determines the cost and duration of the white spaces.
- Step 2: The PU BS broadcasts the cooperation information (e.g. spectral bands, SNR and cost) to SU BS.
- Step 3: The SU BS broadcasts the cooperation information to its SU hosts.
- Step 4: The SU hosts determine optimum transmission and relaying strategies using the cooperation information while the SU hosts determine bid values.
- Step 5: The SU hosts send their respective strategies and bids to SU BS.
- Step 6: Depending on the SUs' reputation values, SU BS decides to lease or not the channels to SU
- Step 7: The SU BS sends its decisions to PU BS.
- Step 8: The PU BS decides to lease or not, and select suitable SUs as relays.
- Step 9: The PU BS transmits packets; and the SU BS divides the spectral band and allocates orthogonal sub-bands, as well as the access time, to each SU, and the SUs transmit packets.

A general equation to calculate reputation value is as follows:

$$r_i(t) = r_i(t-1) + (-1)^{d_i(t)+D(t)} \quad (1)$$

where $d_i(t)$ is the sensing outcome of SU i and $D(t)$ is the final decision given by a fusion center [16]. In our spectrum leasing model, $r_i(t)$ increases when SU i has successfully relay the packets for PU, where $d_i(t)$ is the expected packet relay destination of SU i and $D(t)$ is the packet relay destination requested by PU.

In CRNs, choosing honest SUs to collaborate in spectrum leasing may increase the overall network performance. However, due to the possibility of attacks from malicious SUs, selecting the right SUs to collaborate may be a challenge. In addition, there is no guarantee that the chosen SUs will

continue to remain honest throughout the duration of spectrum leasing. Hence, it is critical to develop a TRM model that is robust and adaptable to the operating environment in order to increase the detection efficiency of malicious SUs.

Generally speaking, as seen from Fig. 2, the Q-learning algorithm works by estimating the values of state-action pairs. For each state-action pair, the agent observes its short-term reward (or delayed reward), and learns its long-term reward (or discounted reward) as time progresses. The state-action pairs and their respective discounted rewards are represented by Q-values, which are kept in a two-dimensional Q-table. The delayed reward $r_{t+1}(s_{t+1})$ is received after the agent has taken the action at time t; while the discounted reward $\gamma \max_{a\in A} Q_t(s_{t+1}, a)$ represents the cumulative rewards received by the agent in future. To re-iterate, the Q-value $Q_t(s_t, a_t)$ update is represented in (2).

$$Q_{t+1}(s_t, a_t) \leftarrow (1-\alpha)Q_t(s_t, a_t) + \alpha[r_{t+1}(s_{t+1}) + \gamma \max_{a\in A} Q_t(s_{t+1}, a)] \quad (2)$$

For each random initial state (s_t, a_t), initialize Q-table entry $Q_t(s_t, a_t) \leftarrow 0$

Observe current state s_t

For each time step t :

Select an action a_t and execute it

Receive delayed reward $r_{t+1}(s_{t+1})$

Observe the new state (s_{t+1})

Update the Q-table entry for $Q_t(s_t, a_t)$ as in (1)

$t \leftarrow t + 1$

End for

Fig. 2. Q-Learning Algorithm

The discount factor γ emphasizes on the importance of future rewards. If $\gamma = 1$, the agent considers the same weightage for both delayed and discounted rewards. If $\gamma = 0$, the agent only considers the delayed reward, and in this case, it is called a myopic approach as compared to the most far-sighted approach which has $\gamma = 1$.

As shown in the Q-learning algorithm, an agent will continue to learn to improve its learning experience until an optimal policy is found. When an agent selects an appropriate action for a state-action pair, the respective Q-value increases, and vice-versa. Hence, in order to maximize cumulative reward over a period of time, the agent learns to take the optimal or near-optimal policy π^* (or a series of actions), which has the optimal Q-value given a particular state as shown in (3).

$$V^{\pi^*}(s_t) = \max_{a\in A} Q_t(s_t, a) \quad (3)$$

The learning rate α determines to what extent the newly acquired knowledge overrides the previously learnt Q-value. If $\alpha = 1$, the agent considers the most current Q-value. The higher the learning rate value α, the greater the current learnt knowledge overrides the old. Higher learning rate speeds up the learning process and this may lead to faster convergence; however, it also causes the agent to respond more drastically to each reward update, which may destabilize the learning process and subsequently, the agent may not converge. On the other hand, a lower learning rate provides a smooth and predictable learning behavior of an agent, and the time taken to converge may be longer

The RL-based TRM model is shown in Table I. State s_t^i represents the SU neighbor nodes or reputation values. Action a_t^i represents the selection of a neighbor node j which has the highest reputation value for collaboration. Reward of a state-action pair $r_t^i(s_t^i, a_t^i)$ represents the cost incurred in collaboration with neighbor node j.

TABLE I STATE, ACTION AND REWARD FOR RL-BASED TRM MODEL

State	State $s_t^i = (p_{j,t}^i) \in S, \mathrm{p}_{j,t}^i \in \{1, 2, \dots, P\}$, where $p_{j,t}^i$ represents the reputation values of neighbor node $j \in J$, where J indicates all neighbor nodes of node i.
Action	Action $a_t^i \in A$ represents the selection of a neighbor node j to collaborate.
Reward	Reward $r_{t+1}^i(s_t^i, a_t^i) \in \{-1, 1\}$ represents a constant value to be rewarded to all collaborating nodes. Value 1 indicates the packets have been sent to the expected destination.

The following sections discuss the simulation scenario, simulation parameters, and preliminary results. As a preliminary investigation, a myopic approach of RL is used in the simulation.

A. Simulation Scenario

The RL-based TRM model is deployed at the SU BS. We consider a static network in a time-slotted environment, with SU BS located at the center of *M* SU nodes (Fig. 1). As an initial investigation, we consider SU BS interacts with PU directly. When unutilized channel is available, PU notifies SU BS, which then broadcasts the cooperation information to its SU hosts. Upon receiving the information, SU hosts send their respective bids to SU BS. Using the RL-based TRM model, SU BS filters out the malicious SUs from collaboration. During the collaboration process i.e. relaying packets for PU, there are *N* SUs that turn malicious and perform random attacks. A traditional TRM and RL-based TRM are used to detect malicious SUs. During each round of packets relay, random attack is launched.

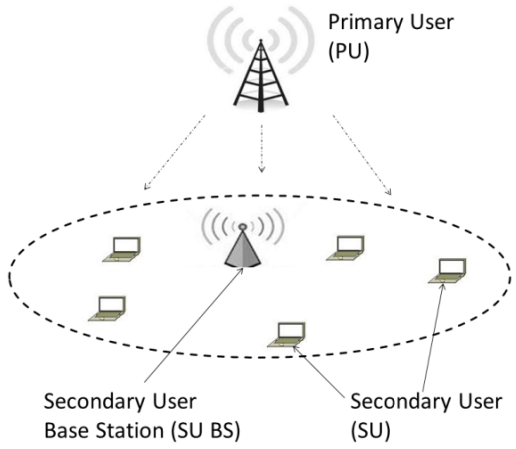

Fig. 1. Topology for collaborative channel leasing in CRNs

B. Simulation Parameters

The simulation parameters are shown in Table II. In this scenario, learning rate $\alpha = 0.2$ is chosen. As a preliminary investigation, we use a myopic approach with $\gamma = 0$.

TABLE II NOTATIONS AND PARAMETERS SETTINGS USED IN THE SIMULATION

Category	Symbol	Details	Values
Initialization	N	Number of collaborating nodes (SUs)	20
	M	Number of malicious SUs	1 to 7
Q-learning	α	Learning rate	0.2
	γ	Discount factor	0

C. Simulation results and discussions

Fig. 2 shows the comparison of the detection performance of traditional TRM and RL-based TRM models when the percentage of the number of malicious SUs in the network increases. From Fig. 2, we observe that the average number of iterations increases as the percentage of malicious SUs increases. Higher percentage of malicious SUs indicates more iterations are required for the detection. The simulation result shows that the detection performance of RL-based TRM model has an improvement of 15% over the traditional model. This happens over the common range from 5 to 30 per cent of malicious SUs. The increased in detection performance is due the capacity of RL-based TRM to learn and re-learn from its dynamic operating environment. However, we noted that there is only slight performance improvement in RL-based TRM when the percentage of malicious SUs is low (< 5%) or high (> 35%). When the number of malicious SUs is low, the RL-based TRM would have lesser previous learnt knowledge of each node behavior, hence its performance is similar to that of the traditional TRM. Similarly, when the number of malicious SUs is high, the learning process of RL-based TRM would mimic the traditional TRM since higher number of malicious SUs indicates easier detection.

From the analysis, further investigations could be carried out to refine the RL algorithm to improve the performance for high percentage (> 35%) of malicious SUs [30]. The low percentage (< 5%) of malicious SUs may not have much impact to the overall system performance [31].

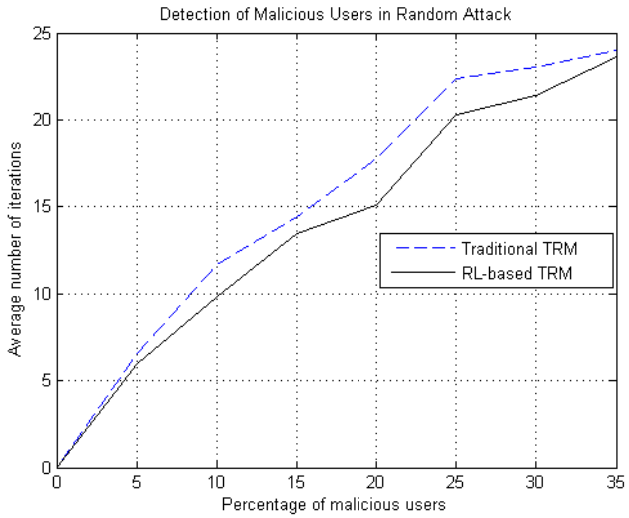

Fig. 2. RL-based TRM achieves lower average number of iterations in detection

IV. CONCLUSION

This paper advocates the use of RL to achieve a more efficient method to detect malicious SUs. Generally speaking, RL enable SUs to observe, learn and respond accordingly in a dynamic operating environment without having to strictly obey a predefined set of rules. This intrinsic characteristic of RL provides a competitive edge in a hostile environment, where the SUs may not be necessary honest or the honest SUs may turn malicious as time progresses. RL-based TRM shows that it is capable of achieving higher detection rate as compared to that of the traditional TRM. As it has shown to be an effective approach, the existing TRM schemes can be further enhanced to incorporate RL to improve PUs' network performance.

V. FUTURE WORK

The preliminary simulation and analysis show the effectiveness of the RL-based TRM model. However, several possible improvements can be carried out as follows:

- to introduce more than one PUs and CRNs in the environment in order to provide a comprehensive scenario.
- to incorporate discount factor γ, and exploration and exploitation into the RL algorithm to provide a more efficient detection mechanism. Investigation can be

done on the RL algorithm to find a trade-off between exploration and exploitation in order to yield higher rewards.

- to perform thorough security measures through testing the RL-based TRM under various attacks scenarios such as sybil attacks [2] and collusion attacks [30,19].
- to analyze and determining learning rate to reduce false positive.
- to decrease the probability of malicious SUs in collaboration, hence increasing the overall spectrum utilization, and PUs and SUs quality of service performance.

REFERENCES

[1] J. Mitola and G. Q. Maquire, "Cognitive radio: making software radios more personal. IEEE Personal Communications," 6(4), 13 – 18 (1999).

[2] S. Li, H. Zhu, B. Yang, C. Chen, and X. Guan, "Believe yourself: A user-centric misbehavior detection scheme for secure collaborative spectrum sensing," in Proceedings of the IEEE International Conference on Communications, 1– 5 (2011).

[3] S. M. Mishra, A. Sahai, and R. W. Brodersen, Cooperative sensing among cognitive radios, in Proceedings of IEEE International Conference on Communications. 4:1658 – 1663 (2006).

[4] R. Chen, J. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in Proceedings of the 27th IEEE International Conference on Computer Communications, 1876 – 1884 (2008).

[5] S. Xu, Y. Shang, and H. Wang, "Double thresholds based cooperative spectrum sensing against untrusted secondary users in cognitive radio networks," in Proceedings of the IEEE 69th Vehicular Technology Conference, 1 – 5 (2009).

[6] H. Li and Z. Han, "Catching attacker(s) for collaborative spectrum sensing in cognitive radio systems: an abnormality detection approach," IEEE Transactions, Wireless Communications, 9(11), 3554 – 3565 (2010).

[7] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," IEEE Transactions on Wireless Communications, 9(8) 2488 – 2497 (2010).

[8] M. S. Al-Kahtani and H. T. Mouftah, "A stable clustering formation infrastructure protocol in mobile ad hoc networks," in Proceedings of the Wireless and Mobile Computing, Networking And Communications, Vol. 3, 406 – 413 (2005).

[9] G. Zhang, R. Ding, and L. Huang, "Using trust to establish cooperative spectrum sensing framework," Procedia Engineering, 15(11), 1361 – 1365 (2011).

[10] W. Wang, H. Li, Y. Sun, and Z. Han, "Securing collaborative spectrum sensing against untrustworthy secondary users in cognitive radio networks," EURASIP Journal on Advances in Signal Processing, Hindawi (2010).

[11] Q. Pei, R. Liang, and H. Li, "A trust management model in centralized cognitive radio network," in Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 491 – 496 (2011).

[12] C. J. C. H. Watkins, Learning from Delayed Rewards. Cambridge University, (1989).

[13] R. S. Sutton and A, G, Barto, Reinforcement Learning, Cambridge, MA: MIT Press, (1998).

[14] Y. Teng, Y. Zhang, F. Niu, C. Dai, M. Song, "Reinforcement Learning Based Auction Algorithm for Dynamic Spectrum Access in Cognitive Radio Networks," in Proceedings of the IEEE 72nd Vehicular Technology Conference, 1 – 5 (2010).

[15] T. C. Clancy and N. Goergen, "Security in Cognitive Radio Networks: Threats and Mitigation," 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), (2008).

[16] H. Li, Q. Pei, X. Jiang, R. Liang, and P. Geng, A sub-spectrum sensing scheme based on reputation in cognitive radio networks, in Proceedings of the International Conference on Computational Intelligence and Security, 478 – 482 (2010).

[17] KL. A. Yau, P. Komisarczuk, and P. D. Teal, "Reinforcement Learning for context awareness and intelligence in wireless networks: Review, new features and open issues," Journal of Network and Computer Applications, 35(1), 253-267 (2012).

[18] N. Vučević, I. F. Akyildiz, and J. Pérez-Romero, "Dynamic cooperator selection in cognitive radio networks. Ad Hoc Networks," 10(5), 789– 802 (2012).

[19] H. Yu, S. Liu, A. C. Kot, C. Miao, and C. Leung, "Dynamic witness selection for trustworthy distributed cooperative sensing in cognitive radio networks," in Proceedings IEEE 13th International Conference Communication Technology (ICCT), 1 – 6 (2011).

[20] B. Wang, Y Wu, K. J. R. Liu, and T. C. Clancy, "An Anti-Jamming Stochastic Game for Cognitive Radio Networks", IEEE J.Sel. A. Commun., 29(4), 877– 889 (2011).

[21] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Anti-Jamming Games in Multi-Channel Cognitive Radio Networks," IEEE Journal on Selected Areas in Communications, 30(1), 4– 15 (2012).

[22] T. C. Clancy III, Dynamic spectrum access in cognitive radio networks Doctoral dissertation, University of Maryland, (2006).

[23] L. Cao and H. Zheng, "Balancing reliability and utilization in dynamic spectrum access," IEEE/ACM Transactions on Networking (TON), Vol. 20, 651– 61 (2012).

[24] S. K. Jayaweera and T. Li, "Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games," IEEE Transactions on Wireless Communications, Vol. 8, 3300– 10 (2009).

[25] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness Y, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," IEEE Journal on Selected Areas in Communications, Vol. 26, 203– 13 (2008).

[26] J. M. Chapin and W. H. Lehr, "Cognitive radios for dynamic spectrum access-the path to market success for dynamic spectrum access technology," IEEE Communications Magazine, Vol. 45, 96– 103 (2007).

[27] J. M. Chapin and W. H. Lehr, "Time-limited leases in radio systems (Topics in Radio Communications)," IEEE Communications Magazine, Vol. 45, 76– 82 (2007).

[28] H. Song, and X. Lin, "A leasing oriented MAC protocol for high spectrum usage in cognitive radio networks," in Proceedings of IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 173–178 (2009).

[29] Y. Yi, J. Zhang, Q. Zhang, T. Jiang, and J. Zhang, "Cooperative communication-aware spectrum leasing in cognitive radio networks," in Proceedings of IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN) 1–11, (2010).

[30] A. S. Rawat, P. Anand, and H. Chen, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," IEEE Transactions on Signal Processing, 59(2), 774 – 786 (2011).

[31] K. Zeng, P. Pawelczak, and D. Cabric, "Reputation-based cooperative spectrum sensing with trusted notes assistance," IEEE Communications Letters, 14(3), 226 – 228 (2010).