White Rose Research Online
eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/75108/

Version: Published Version

Conference or Workshop Item:
Jiang, Tao, Grace, D. and Liu, Yiming (2008) Cognitive radio spectrum sharing schemes with reduced spectrum sensing requirements: IET Seminar on Cognitive Radio and Software Defined Radios: Technologies and Techniques, Septemeber 2008. In: UNSPECIFIED.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Cognitive Radio Spectrum Sharing Schemes with Reduced Spectrum Sensing Requirements

Tao Jiang, David Grace, Yiming Liu

Communication Research Group, Department of Electronics
University of York, York, YO10 5DD, United Kingdom
Email: {tj511|dg| yl127} @ohm.york.ac.uk

Keywords: Cognitive radio, Reinforcement learning, Spectrum sharing, Spectrum sensing, Pre-Play.

Abstract

In this paper, we present a novel distributed spectrum sharing scheme for cognitive radio which can effectively reduce the need for spectrum sensing. This is achieved by utilizing the experience of reinforcement learning. Instead of sensing all of the available spectrum arbitrarily, the scheme is designed to share the spectrum based on an optimum spectrum sharing strategy which is discovered by the agents from their interaction with the wireless communication environment. It shows that reinforcement learning enables an efficient approach of spectrum sensing. The performance of the reinforcement learning scheme is investigated and comparisons with a no learning scheme are given to illustrate the benefits of our scheme.

1 Introduction

Radio spectrum is the 'lifeblood' of wireless communication. Conventional licensed spectrum allocation strategy by radio regulatory bodies can be overly restrictive, making a large part of radio spectrum underutilized [1, 2]. According to Federal Communications Commission (FCC), 15% to 85% assigned spectrum is used with large temporal and geographical variations. Efficient utilization of the radio spectrum has attracted significant attention because of the limited physical spectrum resource and the inefficient usage of spectrum. Cognitive radio (CR), a new paradigm of wireless communication, has been considered as a potential way to solve the conflict between the scarcity of spectrum and the inefficient usage of this physical resource [3-5].

The definition of cognitive radio used in this paper is suggested in [6] as: 'a radio that is aware of and can sense its environment, learn from its environment and adjust its operation according to some objective function'. One distinguishing feature of cognitive radio is the ability of learning. Reinforcement learning uses a mathematical way to define the success level of the interaction between an agent and its environment [7, 8]. Its emphasis on individual learning from direct interaction with the environment makes it perfectly suited to distributed spectrum sharing scenarios [5, 9]. In this paper, we implement reinforcement learning by using a reward function and reward values. Based on the results of the reward function, the action strategy of the agent is modified accordingly. In other words, agents adjust their operation according to the reward function feedback.

The awareness of the state of environment is another vital element of cognitive radio. Spectrum sensing, used as a way to detect unused radio resources and to estimate the interference level, is a power-intensive and time-consuming process [2, 5]. The purpose of this paper is to introduce our reinforcement learning based distributed spectrum sharing scheme which can limit the need for sensing when cognitive radio users find a set of preferred spectrum holes based on their past experience. In our scheme, a weight is assigned to the used resource which indicates the importance of the resource for a CR user, and the weight is updated after every communication action. Once users are 'mature' enough to choose a suitable spectrum for communication by themselves, they are allowed to set up wireless links without sensing the target resource beforehand. We investigate and compare the system performance of three different schemes: (1) a full sensing scheme which CR users scan the target spectrum at the beginning of each activation; (2) a restricted sensing scheme that users only sense the spectrum in their ideal resource set; and (3) a minimum sensing scheme where users directly use their preferred spectrum holes to communicate without sensing. The time and power consumption of these schemes is also shown to illustrate the benefits of our scheme.

This paper is organized as follows. First, we introduce our spectrum sharing scheme and the objective function used for the scheme. Then simulation results are discussed. After that, a brief discussion of the potential work is given. Finally, conclusions are drawn.

2 Reinforcement learning based spectrum sharing with pre-play

The ultimate goal of cognitive radio is to communicate in the best available channel. This is accomplished by exploiting its cognitive capability. Spectrum sensing, the first step in the cognitive cycle of cognitive radio, is designed to monitor and detect available spectrum bands [5]. Since the process of spectrum sensing is time-consuming and power-intensive [2, 5], it is reasonable to reduce the requirement for spectrum

sensing appropriately. Reinforcement learning, a computational approach to learn from interaction, provides an ideal method to efficiently sense and share the spectrum holes. CR users in our distributed spectrum sharing scheme will access the communication resource according to the result of the reinforcement learning. The success level of a particular action, which is whether the target spectrum is suitable for the considered communication request, is assessed by the CR user. Based on the assessment, a reward is assigned in order to reinforce the weight of the physical resource. The weight is practically a number which is attached to an available resource and this number reflects the importance and priority of the channel to a certain CR user.

By using the word 'pre-play' we define a stage that distributed CR users are searching for optimum resources and learning from the experience of searching. In the pre-play stage, players explore the available spectrum pool by accessing all physical resources with equal probability. The weights of the used resources for these users will be modified after every activation. According to the reward function, the weight of the successfully used spectrum is increased by a certain weighting factor. Otherwise, the weight is reduced. By playing the game repeatedly, CR users learn how to choose appropriate channels to communicate. The pre-play stage is effectively the convergence period of our learning algorithm. Once a user converges to an ideal state of spectrum sharing which in our case is to find a set of good resources, it will either directly choose a channel from the restricted resource set without sensing (minimum sensing scheme) or sense the good resources with higher priority (restricted sensing scheme). Unlike sensing, learning will never stop and the weights of these preferred spectrum holes are updated by learning constantly. Meanwhile, users who have already obtained their restricted resource set will only move back to the pre-play stage again when the weight of any ideal resource has decreased under a specific threshold. In other words, if the ideal spectrum is no longer good enough to communicate, the user will again search for a new optimum resource set.

2.1 Spectrum sharing algorithm

Fig.1 is an example layout of the nodes in this paper. We consider the CR users are a set of transmitting-receiving pairs of nodes, denoted as U, uniformly distributed in a square area and all the pairs $U_i \in U$ are spatially fixed. The steps of our algorithm are given as follows.

- **Step 1: State evaluation.** In this step, U_i evaluates its own local system state. In our case, it is whether U_i has found its preferred resource set. We define a preferred channel weight threshold (W_{thr}) which in this paper equals 5 (every time a channel is used successfully its weight increases). U_i compares the weight of the used channel with W_{thr} at every communication request. If the weight is above W_{thr}, U_i considers the channel as a preferred channel and this channel is selected into the preferred channel set. The preferred channel of U_i is effectively the most successful channel used in the past. If the preferred channel set of U_i has been filled with suitable channels, U_i will be considered in an ideal state and allowed to move to next stage, the limited sensing stage.
- **Step 2: Spectrum sensing.** Depending on the result of the evaluation in step 1, there are three different rules in this step:
  - If U_i is still in pre-play stage, it chooses a channel randomly from the available spectrum set. U_i senses the interference level on that channel. If the interference level I of the channel is below the interference threshold I_{thr}, U_i is activated. Otherwise the weight of the spectrum is decreased and U_i starts with a new channel again.
  - If U_i is in the limited sensing stage.
    - Restricted sensing scheme: U_i senses the spectrum in their ideal resource set randomly.
    - Minimum sensing scheme: U_i directly accesses the spectrum in the preferred channel set without sensing.
- **Step 3: *SINR* measuring.** After step 2, the existing users within the same channel can measure the Signal-to-Interference-plus-Noise Ratio (*SINR*) at their receivers. The purpose of measuring *SINR* is to maintain the communication quality of the channels. We set up a *SINR* threshold $SINR_{thr}$. If the *SINR* of the activated pair U_i is greater than the threshold ($SINR_i > SINR_{thr}$), U_i successfully uses the spectrum and the weight of the channel will be increased by a weighting factor f. If $SINR_i < SINR_{thr}$, U_i is blocked by the channel and the weight is updated with a punishment. In addition, according to the measurement of *SINR* of the existing users, the existing users whose *SINR* is decreased below the *SINR* threshold are dropped and the weights of the channel for these users are also decreased accordingly.

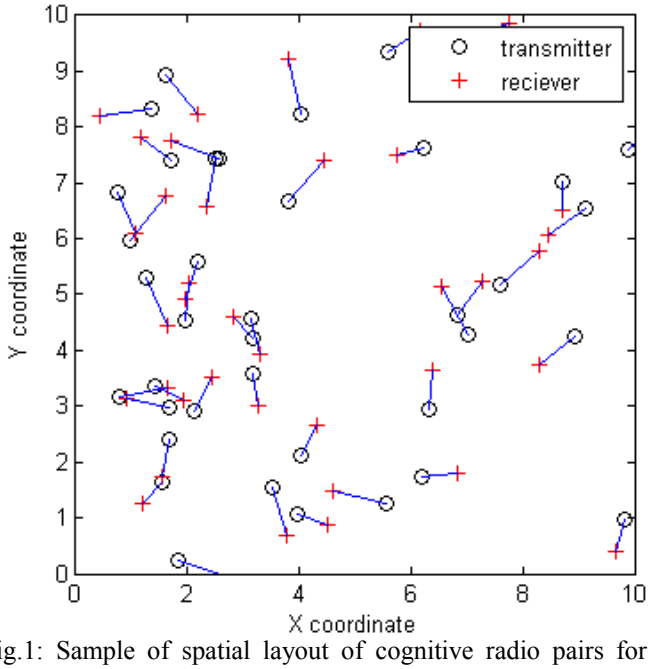

Fig.1: Sample of spatial layout of cognitive radio pairs for simulation

2.2 Objective function

Reinforcement learning is a computational approach to learn from interaction rather than from a known teacher. It is well suited to problems which include a long-term versus short-term reward trade-off [7]. A key element of reinforcement learning is the reward function. A CR user updates its action strategy based on the feedback of the reward function. In other words, the CR user adjusts its operation according to the function. From this point of view, the reward function in reinforcement learning is also the objective function of cognitive radio in our scenario. The following linear function is used as the objective function to update the spectrum sharing strategy in this paper:

$$W_t = f_1 \cdot W_{t-1} + f_2 \quad (1)$$

where W_{t-1} is the weight of a channel at time *t-1*, and W_t is the weight at time t according to previous weight W_{t-1} and the updated feedback from system. f_1 and f_2 are the weighting factors that have different values depending on the localized judgment of current system states and the environment. f is effectively the reward value in function (1). In order to map situations to actions, either a reward value or a punishment values are assigned to f based on the evaluation of the success level of CR users' action.

3 Simulation and results

A basic transmitter-receiver pair communication system model is used because we try to focus on the behavior of the CR users and consequently achieve a deep understanding of such behavior. We believe the technique is widely applicable for other system models. The Okumura-Hata propagation model [10] is used along with log-normal shadowing with a standard deviation of 8dB.

1000 cognitive radio pairs are uniformly distributed on a square service area of 1000km^2. An event-based scenario is employed in our work, at each event a random subset of pairs are activated. A number of 400 is assigned to define the maximum size of the subset. 100 channels are available for communication. The size of preferred channel set is set to 5 which is 5% of the available resources.

The wireless link length is uniformly distributed between 1km and 2km. A carrier frequency of 300MHz is used and the transmitter antenna height is set to 30m. The transmit power is fixed at 1watt and no further power control policy is applied. The gains of the transmit and receive antennas are both fixed at 0dBi. An interference threshold is fixed at -40dBm. The *SINR* threshold is set to 10dB. A noise floor of -124dBm is used, which corresponds to a noise bandwidth of 5MHz and a receiver noise temperature of 300K.

A set of weighting factor values are used which is shown in Table 1. Based on the degree of success, either a reward or a punishment is assigned to the weight of the used spectrum. After each activation, the weight of the successfully used spectrum for a user is increased by a reward. When the attempt fails, the weight is reduced by a punishment. It can be seen in Table 1 that the absolute values of the reward value and the punishment value are equal. In other words the weight is increased or decreased by the same step size.

f_1		f_2	
Reward	**Punishment**	**Reward**	**Punishment**
1	1	1	-1

Table 1: Weighting factor values

The performance of schemes which we discussed above is shown in Fig.2 – Fig.5. We measure the blocking probability at regular points in the service area and a Cumulative Distribution Function (CDF) of system blocking probability at these points is derived. Since we use the information of system dropping along with blocking to adjust the spectrum sharing strategy of CR user, the performance of system interruption is improved. In order to illustrate such improvement, a CDF of dropping probability is also calculated at the same time. An important requirement in our simulation is that all parameters of user are exactly the same for each scheme evaluation. Different system performance is caused only by different spectrum sharing schemes.

Fig.2 illustrates the CDF of system blocking probability of the three schemes which we discussed before. About 70% users' blocking probability in the minimum sensing scheme are below 0.04. But in the full sensing scheme and the restricted sensing scheme, it is about 87% and 95% respectively. Comparing with the red dotted line which is the CDF of the full sensing scheme, the blocking probability of the minimum sensing scheme is higher. It is reasonable that a scheme which always chooses a free channel to operate performs better than a scheme occasionally picks a channel without sensing. It is not expected that the minimum sensing scheme can show its advantages from this point of view. On the contrary, the restricted sensing scheme achieves a better performance compared to the full sensing scheme. This is because the user in the restricted sensing scheme is able to sense the channels which have higher probability to success according to prior experience. This is particularly important because communication can still be dropped.

It can be seen that in every scheme there are about 2% of users whose blocking probability is above 0.2. The blocking probability of these users is difficult to improve no matter which scheme is applied. This is because these users are located either at an extremely high user density area or at a place suffering significant shadowing. The opportunity for these users to successfully set up a communication link is limited.

Fig.3 shows the CDF of dropping probability which illustrates the level of system interruption. Since the information of system dropping is also used to update the spectrum sharing strategy, the performance of the restricted sensing scheme is better than the scheme without learning. But just like the

performance of blocking probability, the dropping probability of the minimum sensing is also higher than the full sensing scheme. A scheme which stops sensing to some extent can not operate better than the full sensing scheme in the aspect of communication quality. However, it can be seen that the overall performance of the minimum scheme is acceptable that the gap between the minimum sensing scheme and others is not huge. The genuine benefit of the limited sensing schemes is discussed in the following paragraphs.

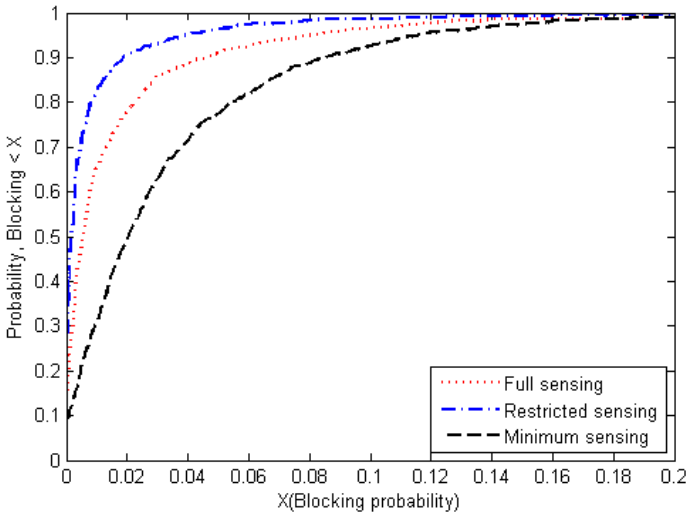

Fig.2: Cumulative distribution function of system blocking probability at discrete points over the service area

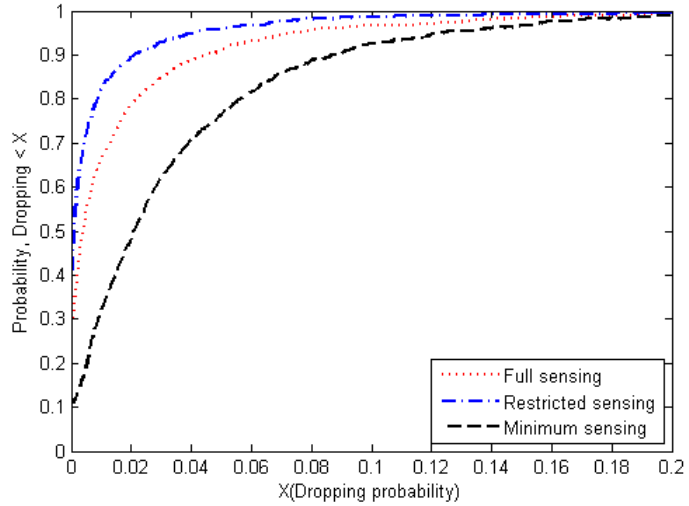

Fig.3: Cumulative distribution function of system dropping probability at discrete points over the service area

Fig.4 shows the average number of channels that CR users must sense in each event. The advantage of our reinforcement learning scheme can be clearly seen. The number of sensed channels effectively represents the time and energy consumption of spectrum sensing. Since the nodes in the full sensing scheme never stop sensing and choose the spectrum on a random basis, the red line with cross maintains its position at about 1.15 throughout the simulation. The blue line with upward-pointing triangle is converging towards to 1 which represents the ideal state of the restricted sensing scheme. In the optimal state, all the transmission requests are accepted on the first tested channel. The average number of sensed channels of the restricted sensing scheme in each event cannot be less than one, because like the full sensing scheme this scheme also never stops sensing.

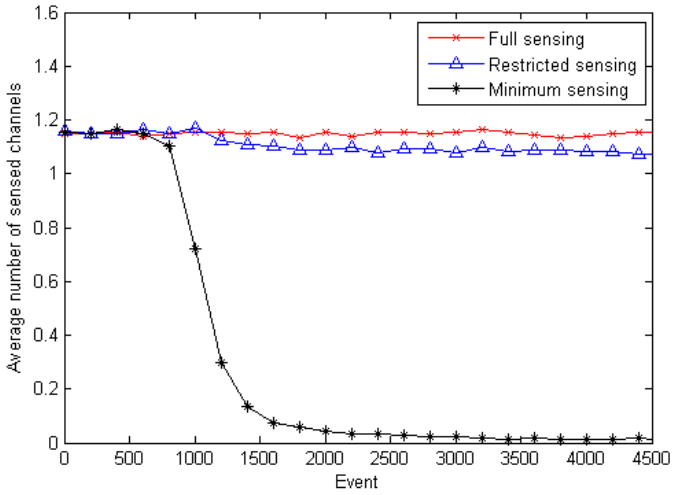

Fig 4: Average number of sensed channels

The behavior of the minimum sensing scheme can be divided into three periods. From the first event to about event 600 is the first period. Users in this period are all in the pre-play stage. It means users are searching for their optimum resources during this period. The second period is from event 600 to event 2000. The needs for spectrum sensing are dramatically reduced in this period. After a certain simulation time, a spectrum sharing equilibrium is established by the application of the reinforcement learning algorithm. CR users start to directly access the spectrum in the preferred channel set without sensing. In the third period, the black line with asterisk remains at the value of 0.03 which means the state of the system is stable. After the spectrum sharing equilibrium is established, the CR users are able to avoid collisions by utilizing their experience from learning rather than spectrum sensing. In this way, the requirements for spectrum detection are greatly reduced. Compared with the full sensing scheme, the time and energy consumption of the restricted sensing scheme is 5% lower. When it comes to the minimum sensing scheme, the overall average number of sensed channels is about 23% of the full sensing scheme. If we only compare this figure after event 2000, it is only about 1.72% of the full sensing scheme. The needs for spectrum sensing are almost eliminated by reinforcement learning.

Fig. 4 also shows the convergence behavior of our learning scheme. Like other learning algorithms for dynamic channel assignment[11, 12], our scheme needs a sufficiently high number of stages to converge to its optimal state. From the start of the simulation to event 2000, our learning scheme was converging to its ideal spectrum sharing strategy. CR users found their preferred resource set gradually. After event 2000, the learning scheme finally arrived at its spectrum sharing

equilibrium which practically means CR users' preferred resource sets are fully occupied by good channels. The user is able to avoid improper channels by using its prior experience. Though the node is designed to move back to the pre-play stage if only one of its preferred channels is no longer good to communicate, the state of the learning scheme is extremely stable. Obviously, the CR users in our scheme have the potential to share spectrum in a 'polite' way even if they do not sense beforehand.

Nevertheless, the differences between our learning and other DCA learning algorithm are also clear. Unlike the centralized Q-learning approach proposed in [11] and the no regret learning investigated in [12], the nodes in our scheme do not examine all available spectrum by playing all possible actions. It is possible that the CR users only explore a small part of the available spectrum pool before they find their good resource set. As long as the preferred channel set is full, no new channels will be chosen. In addition, our learning scheme only updates the weight of the strategy currently performed. From this point of view, the complexity of our learning scheme is lower.

In order to illustrate the system performance in more detail, we record the number of sensed channels in each activation and plot the CDF of it in Fig. 5. It can be seen that about 77% of the transmission activations in the minimum sensing scheme succeed without sensing the target spectrum. The restricted sensing scheme performs slightly better than the full sensing scheme. About 90% of the communication requests in the restricted sensing scheme succeed before the user tests the third channel, but in the full sensing scheme only 85% users are able to meet this requirement. Fig.5 also shows that about 99% requests are accomplished before sensing four channels.

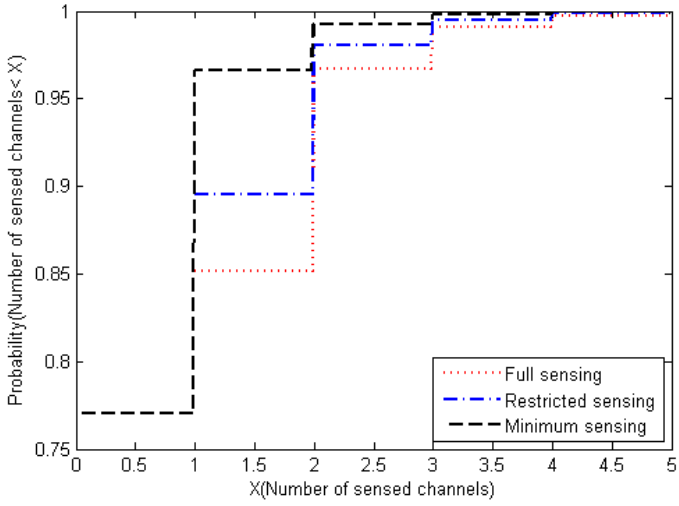

Fig 5: Cumulative distribution function of the number of sensed channels in each communication activation

4 Discussion

The idea that the cognitive radio user keeps a set of preferred resource makes our scheme particularly suitable for the OFDM based cognitive radio system where multiple spectrum bands can be simultaneously used for the transmission. By choosing multiple channels from the preferred spectrum set, our reinforcement learning scheme has the potential to achieve a better system performance. This argument is helpful to push our work forward.

5 Conclusions

In this paper, we proposed a reinforcement learning based spectrum sharing scheme for cognitive radio which has the potential to reduce the need for spectrum sensing. By utilizing the ability of learning, cognitive agents can remember their preferred communication resources, and this learning ability enables an efficient approach to spectrum sensing and sharing. The advantages of our scheme can be clearly seen from the simulation results. By utilizing reinforcement learning, the need for spectrum sensing is significantly reduced. The overall time and energy consumption of spectrum sensing in the minimum sensing scheme is about 23% of the full sensing scheme. After the minimum sensing scheme converged to its spectrum sharing equilibrium, this figure is only 1.72%. The restricted sensing scheme improves the system performance in two aspects: the sensing consumption is 5% lower than the full sensing scheme. On the other hand, the blocking and dropping probability is also the lowest of the three schemes. Since time and power efficiency are critical issues in real time communication, the advantages of our learning scheme is definite.

References

[1] FCC, "Notice of proposed rule making and order," ET Docket No 03-222, December 2003.

[2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, pp. 2127-2159, 2006.

[3] J. Mitola, "Cognitive radio: Making software radios more personal," *IEEE Personal Communication*, vol. 6, pp. 13-18, Aug. 1999.

[4] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal on selected areas in communications*, vol. 23, pp. 201-220, Feb. 2005

[5] B. Fette, *Cognitive Radio Technology*: Newnes, 2006.

[6] L. Dasilva and A. Mackenzie, "Cognitive Networks: Tutorial," in *CrownCom* Orlando, FL, July 2007.

[7] R. S. Sutton and A. G. Barto, *Reinforcement learning : an introduction*: The MIT Press, 1998.

[8] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement Learning: A Survey," *Journal of artificial intelligence Research*, vol. 4, pp. 237-285, May. 1996.

[9] M. Bublin, J. Pan, I. Kambourov, and P. Slanina, "Distributed spectrum sharing by reinforcement and game theory," presented at 5th Karlsruhe workshop on software radio, Karlsruhe, Germany, March. 2008.

[10] S. R.Saunders, *Antennas and propagation for wireless communication systems*: Wiley, 1999.

[11] J. Nie and S. Haykin, "A Dynamic Channel Assignment Policy Through Q-Learning," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1443-1455, NOV. 1999.

[12] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Networks and Applications*, vol. 11, pp. 779-797, December, 2006.