Research Article

Optimizing Spectrum Trading in Cognitive Mesh Network Using Machine Learning

Ayoub Alsarhan[1] and Anjali Agarwal[2]

[1] *Department of Computer Information System, Hashemite University, Zarqa 13115, Jordan*
[2] *Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada H3G2W1*

Correspondence should be addressed to Ayoub Alsarhan, ayoubm@hu.edu.jo

In a cognitive wireless mesh network, licensed users (primary users, PUs) may rent surplus spectrum to unlicensed users (secondary users, SUs) for getting some revenue. For such spectrum sharing paradigm, maximizing the revenue is the key objective of the PUs while that of the SUs is to meet their requirements. These complex contradicting objectives are embedded in our reinforcement learning (RL) model that is developed and implemented as shown in this paper. The objective function is defined as the net revenue gained by PUs from renting some of their spectrum. RL is used to extract the optimal control policy that maximizes the PUs' profit continuously over time. The extracted policy is used by PUs to manage renting the spectrum to SUs and it helps PUs to adapt to the changing network conditions. Performance evaluation of the proposed spectrum trading approach shows that it is able to find the optimal size and price of spectrum for each primary user under different conditions. Moreover, the approach constitutes a framework for studying, synthesizing and optimizing other schemes. Another contribution is proposing a new distributed algorithm to manage spectrum sharing among PUs. In our scheme, PUs exchange channels dynamically based on the availability of neighbor's idle channels. In our cooperative scheme, the objective of spectrum sharing is to maximize the total revenue and utilize spectrum efficiently. Compared to the poverty-line heuristic that does not consider the availability of unused spectrum, our scheme has the advantage of utilizing spectrum efficiently.

1. Introduction

With the explosion of the number of emerging wireless applications for mobile users, the frequency spectrum has become congested to support the dramatic increase in the demand for the limited spectrum. Moreover, traditional spectrum management policies have contributed significantly in spectrum scarcity crisis [1]. In such schemes, the licensed spectrum is used only by the owner of license; other users are prevented from utilizing the unused spectrum. Consequently, spectrum owners are prevented from real-time interaction with radio environment and from determining appropriate communication parameters and adapting to the changes in the radio environment. For example, to increase data transfer rate and avoid interference, the wireless system may detect and switch to another lightly crowded band.

Fixed spectrum assignment policies prevent users from dynamically utilizing unused allocated spectrum; hence poor utilization and spectrum holes will be resulted. Moreover, the owner loses the profits from renting the unused spectrum. In wireless technology, another challenge is guaranteeing the QoS of applications that require huge bandwidth resources and service continuity protection [1]. To overcome the spectrum scarcity problem, the Federal Communications Commission (FCC) allows SUs to use unutilized spectrum if they do not interfere with PUs [1–3]. Recently, wireless mesh networks (WMNs) have emerged as a significant new technology that can provide ease of installation, low-cost means for flexible, and fast deployment of Internet-based services in diverse environments [1–3]. In order to become a mature technology, WMNs need to offer multimedia and emergency services that require more bandwidth resources.

Dynamic spectrum access (DSA) is proposed to mitigate spectrum scarcity through utilizing spectrum efficiently. It also enables users to adjust communication parameters (such as operating frequency, transmission power, and modulation scheme) in response to the changes in the radio environment [1–3]. DSA enables implementation of cognitive radio (CR) that brings a promise to increase spectrum at a minimum cost by using licensed spectrum whenever spectrum owners do not use it. This approach provides up to 85% of the unused spectrum [1]. CR also enhances the capability of WMNs to support broadband systems. CR encourages implementing new more flexible spectrum sharing paradigms. These sharing paradigms include the use of trading spectrum access on secondary market where PUs can rent unused spectrum to SUs and generate more revenue [3]. Despite of obvious advantages of using CR in WMNs, there are still several issues that require more investigation such as economic factors that include PUs revenues and SUs satisfaction. Spectrum trading also presents the challenge of sharing spectrum among primary users. This paper addresses when and how spectrum is shared among PUs and between primary and secondary users. Spectrum is shared between PUs and SUs based on our economic model and under dynamic traffic load conditions. Our economic model includes the costs and revenues associated with renting a spectrum. The cost of renting spectrum is a reduction of spectrum for PUs in favor of increasing revenue.

In our work, PUs borrow channels from other PUs. Our design objective is to improve spectrum utilization (among PUs) and maximize revenue for spectrum owners (spectrum trading), while meeting some defined constraints. In order to develop an intelligent radio that is able to deal with conflicting objectives in radio environment, we propose to use reinforcement learning which is an effective tool to deal with rational entities that make decisions to maximize their benefits with whatever little information they have [4]. It provides a mathematical framework for modeling decision-making in situations where the decision maker is not sure about the outcome.

In our work, reinforcement learning (RL) is used as a means for extracting an optimal policy that helps a PU to adapt to the changing radio environment conditions. PUs employ the extracted optimal policy to solve the following dilemma. When a request for spectrum arrives, the PU recognizes that it should give part of its spectrum to gain the revenue from rent. However, the QoS for PU might be degraded due to renting the spectrum. The PU might reject serving because it needs this spectrum and loses the reward. As a result, the PU waits for its demand for the spectrum to subside before renting spectrum. Consequently, the likelihood of losing a reward of serving SUs increases, which pushes the PU to become more spectrum-demanded in order to reduce its loss. Under the emerging secondary market spectrum policy, when renting available spectrum to other parties (i.e., PU, SU), the PUs need to consider the economic factors, such as the spectrum price and the operation revenue obtained. We formulate this spectrum trading problem as a revenue maximization problem. Such a

formulation allows RL to optimize the trading problem. The contributions of our paper are as follows.

(i) A new spectrum-sharing scheme among PUs is proposed.

(ii) How the concept of RL can be used to obtain a computationally feasible solution to the considered spectrum trading problem is described.

(iii) An extensive numerical evaluation, based on analysis and simulation, of the RL-based method for spectrum trading is presented.

The rest of this paper is organized as follows. Firstly, we present previous work in spectrum sharing and trading, followed by our assumptions and work environment. We then describe our spectrum sharing scheme and formulate the spectrum trading problem. In the next section, we describe our model for solving the problem using RL, and illustrate its implementation and how we optimize obtained revenues using the RL algorithm. Next, we present some of the tests performed and show the behavior of the implemented system under different conditions. Finally, our last section concludes the paper.

2. Related Work of Spectrum Trading Using CRs

In a cognitive network, PUs can rent their unused spectrum to SUs. The problem of spectrum trading was considered in [5] where each node charges other nodes for relaying its traffic. The objective function is defined as the revenue obtained from transmitting the node traffic plus other nodes charges minus the price paid for other nodes along the route to the destination. In [6], multiple PUs sell unused spectrum resources to SUs to get monetary gains while SUs try to get permissions from PUs for accessing the rented spectrum. In order to maximize the payoffs of both primary and secondary users, game theory is used to coordinate the spectrum allocation among primary and secondary users through a trading process. The payoff of a PU is defined as the difference between the price of the sold spectrum and the cost of buying spectrum. However, the model does not consider the QoS of PUs.

In the framework proposed in [7], a PU may lease the owned spectrum to SUs in exchange for cooperation in the form of distributed space-time coding. For the PU, the main concern is maximizing its quality of service in terms of either rate or probability of outage, accounting for the possible contribution from cooperation. However, SUs compete among themselves for transmission within the leased time slot following a distributed power control mechanism. PU charges SUs for the leased spectrum in [8]. The problem is formulated as an oligopoly market competition and a noncooperative game is used to obtain the spectrum allocation for SUs. Nash equilibrium is considered as the solution of this game. In [9], it is extended to multiple PUs selling the spectrum to SUs. The model considers the behavior of other PUs to specify the price of spectrum. In [10], the advantages of employing market forces to address the issues of wireless spectrum congestion and the allocation

of spectrum are addressed. It is shown that when unlicensed spectrum is assigned to all competing SUs during periods of excess demand an inefficient outcome is likely to result. PUs compete to sell a spectrum to a set of buyers in [11]. Game theoretic approach is proposed to obtain the selling quantities and bidding price.

Several studies tackle the issue of spectrum sharing among PUs. In [12], PUs compete with each other to get the spectrum. To analyze the dynamic spectrum allocation of the unused spectrum bands to PUs, an auction theory was used. The problem was formulated as a multiunit sealed-bid sequential and concurrent auction. In [13], PUs dynamically compete for portions of available spectrum. PUs are charged by the spectrum policy server for the amount of bandwidth they use in their services. The competition problem is formulated as a non-cooperative game and new iterative bidding scheme that achieves Nash equilibrium of the operator game is proposed. In the proposed system in [14], two spectrum brokers offer a spectrum for a group of PUs. The broker wants to maximize its own revenue. Brokers' revenues are modeled as the payoffs that they gain from the game. On the other hand, PUs want to maximize its own QoS satisfaction at minimum expense.

Centralized regional spectrum broker distributes a spectrum among PUs in [15]. PUs do not own any spectrum; instead they obtain time bound rights from a regional spectrum broker to part of the spectrum and configure it to offer the network service. In [16], users adjust their spectrum usage based on a defined threshold called poverty line. A PU can borrow from its neighbors if the neighbors have number of idle channels greater than a poverty line. However, this scheme does not consider the availability of channels and the load of PU. It is possible that the neighbors have a number of idle channels less than their poverty line and these channels will be unused. Moreover, none of these schemes consider what follows.

(i) Utilizing spectrum efficiently: spectrum owners compete for spectrum to maximize their revenues regardless of efficient spectrum utilization.

(ii) Maximizing total revenues of PUs through utilizing the whole spectrum: the cooperation between PUs to maximize total revenues is neglected in these schemes.

(iii) Learning PUs a control policy to adapt the offered size of spectrum and spectrum price based on the changes in the radio environment such as traffic load, cost of services, and spectrum price.

Using simulations, we show the ability of our scheme to utilize spectrum efficiently by comparing its performance with the poverty-line scheme. Moreover, we conduct some experiments to show how our scheme can adapt to different network conditions such as traffic load and spectrum cost.

3. Network Overview

In this section, we present our assumptions. The network consists of two types of nodes: mesh routers (MRs) and mesh clients (MCs). A wireless mesh network has several MRs that jointly form a cluster [17]. Each cluster is a WLAN, where MRs play the role of access point and the MCs act as nodes served by them. The algorithm proposed in [17] is used to form and maintain clusters. Moreover, the proposed signaling protocol in [17] is used to manage communication among the PUs and the SUs. MRs have fixed locations, whereas MCs are moving and changing their places arbitrarily. The spectrum is divided into non-overlapping channels which are the basic unit of allocation. The network consists of W PUs and N SUs. We define a PU as a spectrum owner that may rent a spectrum to other users. Each PU has K channels assigned to it in advance. Each PU offers an adaptable number of channels to MRs (SUs). MRs use the rented spectrum to serve MCs. We assume that spectrum-request arrival follows Poisson distribution with arrival rate λ (the mean number of requests arriving per unit time). The service rate for incoming request is assumed to be exponentially distributed with service rate μ. These assumptions capture some reality of wireless applications such as phone call traffic.

4. Spectrum Trading Model

In this section, we formulate a theoretical model that is used to describe the general spectrum trading problem between PUs and SUs. Next we describe our on-demand-based spectrum sharing scheme and define the constraints of borrowing a spectrum among PUs.

4.1. Spectrum Trading Problem Formulation. In our model, we define the components for primary user $y(\text{PU}_y)$ as follows.

(i) Spectrum allocation vector SP_y:

$\text{SP}_y = \{\text{SP}_y(m) \mid \text{SP}_y(m) \in \{0, 1\}\}$ is a vector of spectrum status. If $\text{SP}_y(m) = 1$, channel m is not available currently. Spectrum status changes over time according to the spectrum demand.

(ii) Interference vector I_y:

$I_y = \{I_y(i) \mid I_y(i) \in \{0, 1\}\}$ is a vector that represents the interference among PU_y and other PUs; if $I_y(i) = 1$, then PU_y and PU_i cannot use the same channel at the same time because they would interfere with each other.

(iii) Channel reward vector R_y:

$R_y = \{R_y(m) \mid R_y(m) \in \{0, \infty\}\}$ is a channel reward vector, which describes the reward that PU_y gets by successfully renting channels to SUs. $R_y(m)$ is the reward that PU_y gets from renting channel m. It is computed as follows:

$$R_y(m) = p\mu\omega_m, \tag{1}$$

where p is spectrum price for renting a channel m and ω_m represents the quality of wireless transmission for channel m and is computed as follows:

$$\omega_m = \frac{\mathcal{C}\{m\}}{\arg \max_z \mathcal{C}\{z\}}, \tag{2}$$

where $\mathcal{C}\{m\}$ is the capacity of channel m and is computed using Shannon's formula. To fit the reward function in (1), channel m's capacity is normalized by the largest capacity among all channels. It is clear from (2) the channel with higher capacity provides high-quality communication and it should get higher reward than others. The average reward for a PU is computed mathematically as follows:

$$\overline{R} = r\overline{\lambda}, \tag{3}$$

where r is the reward of serving one request, and $\overline{\lambda}$ is the average rate of accepting SUs, defined as

$$\overline{\lambda} = \frac{A_c}{T_r}, \tag{4}$$

where A_c is the number of accepted requests, and T_r is the total number of requests. Equation (3) is used to compute analytical reward for a PU. The total reward TR_y is the following:

$$TR_y = \text{SP}_y \cdot R_y^t, \tag{5}$$

where R_y^t is transpose of the channel reward vector.

(iv) Borrowable channel set BC_y:

our scheme allows two neighbors to exchange channels to maximize their reward while complying with conflict constraint from set of the neighbors. We define that two PUs are neighbors if their transmission coverage area is overlapped with each other. The set of channels that PU_y can borrow from PU_j should not interfere with PU_y neighbors. We refer to these channels as $\text{BC}_y(\text{PU}_y, \text{PU}_j)$:

$$\text{BC}_y\left(\text{PU}_y, \text{PU}_j\right) = \frac{L\left(\text{PU}_j\right)}{L\left(G\left(\text{PU}_y\right) \setminus \text{PU}_j\right)}, \tag{6}$$

where L gives the set of channels assigned to the given user(s) (e.g., $L(\text{PU}_j)$ represents the list of PU_j channels); $G(\text{PU}_y)$ is a list of neighbors of a primary user PU_y.

4.2. On-Demand-Based Spectrum Sharing Scheme. In our scheme, PUs can exchange channels if the borrowed channels do not interfere with the channels of neighbors. After serving a request, the PU returns back borrowed channels to the owner users. PUs adjust their spectrum usage based on demand. As a result, the PU decides to borrow channels if the spectrum is not available to accommodate SUs requests and it is profitable to serve new SUs in terms of revenue. In our scheme, spectrum is shared among PUs as follows.

Step 1. PU computes the revenue of serving new SUs.

Step 2. If the revenue is positive and worthy, a PU requests neighboring PUs for a spectrum through a "borrowing frame" that is broadcast to all neighbors. The request frame specifies the size of required spectrum.

Step 3. Each PU receives a "borrowing frame," checks its idle channel list, and if there are idle channels, the PU temporarily gives up a certain amount of idle spectrum and sends an "accept frame" that includes channel IDs. If all channels are busy then the request is ignored.

Step 4. After receiving "accept frame(s)," the PU specifies a borrowable channel set BC and ranks its elements based on their capacity. If the PU does not receive any "accept frame," it queues the requests.

Step 5. After selecting channels, the PU informs the owners of the selected channels.

Step 6. After the PU finishes serving SUs, it returns the borrowed channels.

Our scheme guarantees high utilization through using all system channels provided that the interference constraint is met.

5. Reinforcement Learning-Based Model

Reinforcement learning is a subarea of machine learning concerned with how a system administrator takes actions in different circumstances in a work environment to maximize long-term revenue [4]. Let $X = \{X_0, X_1, X_2, X_3, \ldots, X_t\}$ be the set of possible states an environment may be in, and let $A = \{a_0, a_1, a_2, \ldots, a_t\}$ be a set of actions a learning agent may take. In RL, a policy is any function: $\pi : X \rightarrow A$ that maps states to actions. Each policy gives a sequence of states when executed as follows: $X_0 \rightarrow X_1 \rightarrow X_2 \ldots$ where X_t represents the system state at time t and a_t is the action at time t. Given the state X_t, the learning agent interacts with the environment by choosing an action a_t, then the environment gives a reward r_t and the system transits to the new state X_{t+1} according to the transition probability $P_{X, X_{t+1}}$ and the process is repeated. The goal of agent is to find an optimal policy $\pi^*(X)$ which maximizes the total reward over time. In this section, we define RL model applicable to control the spectrum trading.

5.1. Basic Formulation of RL Model. For the basic formulation, we describe the elements that facilitate the definition of the RL. These elements are the events and states of the system. Each PU has one finite FIFO queue for SUs (MRs) requests. The PU uses extracted optimal control policy to decide whether it is worthy to increase the offered spectrum for a new request or queue it. The request is added to the tail of the queue if a spectrum is insufficient to accommodate it and a PU fails to borrow a spectrum from other PUs. The request is served if the PU has sufficient spectrum. However,

if a queue is full, the request is rejected. In our work, the agent is developed to be implemented at the PU level of WMN in a distributed manner. It provides the trading functionality for a single queue. Each agent uses its local information and makes a decision for the events occurring in the PU in which it is located.

In our model, we have an adaptable spectrum size, $f(X_t)$, according to the percentage of queue usage (traffic load) and the gained revenue. At time t, the state of the system X_t is the number of accepted requests. Accepted requests is served immediately if there is adequate spectrum or they might be placed in the queue. Let $\{X_t, t \geq 0\}$ denote a random variable which represents system states, X is the state space. At state X_t, spectrum size $f(X_t)$ is used to serve the queued requests with a service rate $f(X_t)\mu$. Transition from one state to another means a request arrival or the SU is served. All possible states are limited by the following constraints:

(i) $X_t \leq QS$, where QS is the maximum length of the queue.

(ii) $f(X_t) \leq KW$,

where K and W are defined in Section 3. From a state, the system cannot make a transition (arrive, depart) unless the constraints are met.

5.2. Spectrum Trading Agent and State Space.
In our system, an event can occur in a PU (agent) when a new request for spectrum arrives or a SU releases its assigned spectrum. These events are modeled as stochastic variables with appropriate probability distribution. At any time the PU is in a particular configuration defined by the size of offered spectrum for trading, the price of spectrum, and the number of admitted SUs.

In our case, each time a request for spectrum arrives one of the following decisions must be made: accept arrival request or reject the request. Upon serving the request, a PU has to decide the optimal offered-spectrum size for renting. The action space is given by

$$A = \{a : a \in \{0, 1\}\}, \tag{7}$$

where $a = 0$ denotes request rejection, $a = 1$ indicates that the PU has accepted the request and it might be placed in the queue if the spectrum is insufficient to accommodate it.

5.3. Model Optimization.
In our model, the value $f(X_t)$ indicates the optimal spectrum size offered for SUs at state X_t that maximizes the estimated mean value of revenue,

$$\overline{V}(\pi^*) = \overline{R} - C, \tag{8}$$

where $\overline{R}$ is the average reward given in (3) and C is the cost of renting spectrum to the SUs and is computed as follows:

$$C = f(X_t) * \delta, \tag{9}$$

where δ is the cost of renting one spectrum unit to SUs. Due to spectrum renting, the spectrum remaining for the primary user becomes smaller; hence its QoS is degraded. The rate of revenue at state X_t is computed as

$$q(X_t) = a_t(r * X_t * f(X_t)\mu - C), \tag{10}$$

where r is the reward for renting spectrum and is computed using (1). The actual mean value of the net revenue under policy π for PU_y is given by

$$\overline{A}_y(\pi) = \frac{\lim_{D \to \infty} \sum_{t=1}^{D} q(X_t)}{D}, \tag{11}$$

where $\overline{A}_y$ indicates the average actual value of the net revenue of PU_y when policy π is executed and D represents the time horizon. The state transition probability is given by

$$P_{X_t, X_{t+1}}(a) = \begin{cases} \lambda, & X_{t+1} = X_t + 1 \\ f(X_t)\mu, & X_{t+1} = X_t - 1 \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

where X_t represents current state and X_{t+1} is the next transited state. In our system, an event can occur in a PU (agent) when a new request for spectrum arrives or a SU releases its assigned spectrum. These events are modeled as stochastic variables with appropriate probability distribution. Hence, the state transition occurs when a request arrives or is served and this is shown in (12).

5.4. Optimal Policy.
The optimal policy gives the maximum net revenue when a PU adopts it. It specifies the optimal spectrum size and price for each state. Basically, in our model the optimal policy is specified according to the average revenue value obtained for each transition with the offered spectrum size. For each state, the revenue gained depends on the action reward, cost of spectrum, and the spectrum demand. When a new spectrum request arrives at the queue, the PU checks if it is worthy to increase the offered spectrum based on the revenue gained. It then either increases the offered spectrum or keeps it. When a SU departs from the system the PU may decrease the offered spectrum based on revenue. Although decreasing the size of spectrum decreases the customers' satisfaction—since their waiting time will increase accordingly—the PU always chooses the action that maximizes its revenue. In our work, a PU uses RL to choose a policy, $\pi : X \to A$, for deciding the next action a_t based on the current state X_t. We apply a value iteration algorithm to find an optimal policy. The value function [4] of policy π is given as

$$V^{\pi}(X_t) = q(X_t) + \alpha \sum_{X_{t+1} \in X} P_{X_t, X_{t+1}}(a) V^{\pi}(X_{t+1}), \tag{13}$$

where α is the discount revenue that satisfies $0 \leq \alpha < 1$ and starting with $t = 0$. The value function $V^{\pi}(X_t)$ can be considered as the expected revenue for policy π starting from state X_0. The optimal value function is given [4] as

$$V^*(X_t) = q(X_t) + \max_{a \in A} \alpha \sum_{X_{t+1} \in X} P_{X_t, X_{t+1}}(a) V^*(X_{t+1}). \tag{14}$$

The optimal policy is given as follows [4]:

$$\pi^*(X_t) = \arg\max_{a \in A} \sum_{X_{t+1} \in X} P_{X_t, X_{t+1}}(a) V^*(X_{t+1}). \quad (15)$$

We define an optimal policy π^* as follows:

$$H_y(\pi^*) \geq H_y(\pi), \quad (16)$$

where H_y indicates the total net revenue of PU$_y$ computed as follows:

$$H_y(\pi^*) = \lim_{D \to \infty} \sum_{t=1}^{D} q(X_t). \quad (17)$$

5.5. Analytical Model for Spectrum Trading. Network conditions are changing randomly. These conditions include traffic level, spectrum cost, and the size of unused spectrum. As a consequence, PUs should adapt to continue increasing the revenue. The principal parameters that PUs control are the price and the size of the offered spectrum. In our model, the PUs' revenues sensitivity to the number of the offered spectrum size ($f(X_t)$) can be derived from (8):

$$\frac{\partial \overline{V}}{\partial f(X_t)} = \left(\frac{\partial \overline{R}}{\partial f(X_t)}\right) - \left(\frac{\partial C}{\partial f(X_t)}\right) = \left(\frac{\partial \overline{R}}{\partial f(X_t)}\right) - \delta. \quad (18)$$

We assume the average reward sensitivity to the spectrum size can be approximated by the cost of accepting new SUs, u, which is calculated as follows:

$$u = r - o, \quad (19)$$

where o is the reward increment from accepting new requests. Substituting in (18), the PU's revenue is maximized when spectrum size equals the root of

$$\frac{\partial \overline{V}}{\partial f(X_t)} = u(f(X_t)) - \left(\frac{\partial C}{\partial f(X_t)}\right) = 0. \quad (20)$$

We used Newton's method of successive linear approximations to find the root of (20). The new spectrum size $f(X_t)_{n+1}$ at each iteration step n is computed as follows:

$$f(X_t)_{n+1} = f(X_t)_n - \frac{u_n - \delta}{\partial(u(f(X_t)) - \delta)/\partial f(X_t)}. \quad (21)$$

Approximating the derivative in (21) at step n:

$$\frac{\partial(u(f(X_t)) - \delta)}{\partial f(X_t)} = \frac{\partial u(f(X_t))}{\partial f(X_t)} \cong \frac{u_n - u_{n-1}}{f(X_t)_n - f(X_t)_{n-1}} \quad (22)$$

and substituting (22) in (21), the new spectrum size will be

$$f(X_t)_{n+1} = f(X_t)_n - \left(f(X_t)_n - f(X_t)_{n-1}\right) \frac{u_n - \delta}{u_n - u_{n-1}}. \quad (23)$$

Spectrum size adaptation is then realized using Algorithm 1, where ε is the tolerable error. The presented solution for revenue maximization does not take into account the QoS of PUs. The request of spectrum from the PU is blocked if it arrives while a PU is already using all of its spectrum. Therefore, the probability of blocking for PU$_y$ is computed as follows [18]:

$$B_y = \frac{\rho^K}{K!} \left(\sum_{k=0}^{K} \left(\frac{\rho^K}{K!}\right)^{-1}\right), \quad (24)$$

where ρ is computed as follows:

$$\rho = \frac{\lambda}{\mu}. \quad (25)$$

Although for optimal spectrum size and price, one can expect that standard blocking constraint of PUs will be met. However, in some scenarios the blocking probabilities may exceed the constraints. To cope with this constraint, we use a spectrum price for controlling the size of the offered spectrum and meeting the blocking probability for the PUs. It is clear when a PU increases the price of spectrum the arrival rate of SUs and the demand of spectrum will be decreased. The arrival rate depends on the offered price. The new arrival rate of SUs is calculated as follows [19]:

$$\lambda = \tau e^{-\varphi \acute{p}}, \quad (26)$$

where τ is the maximum number of users arriving to a PU, φ represents the rate of decrease of the arrival rate as spectrum price increases and is related to the degree of competition between the PUs, and $\acute{p}$ is the new price. Here, we assume φ is given a priori. There is an inverse relationship between the price and the demand of the spectrum. A PU has to meet its blocking probability constraint B_y^C. Blocking probability depends on the number of available channels and the traffic load. If the blocking probability for a PU exceeds the blocking constraint, a PU continues to increase the spectrum price till its blocking probability is met. Because of the inverse relationship between spectrum price and its demand, it can be easily shown that when a spectrum price is increased the available channel for a PU will be increased and therefore the blocking probability will be reduced. This feature indicates that if a PU$_y$ blocking constraint B_y^C is not met, $B_y > B_y^C$, the spectrum price is increased to fulfill the blocking probability constraint. However, we assume that the potential price increment should be minimized as possible as it can keep the demand for spectrum-high and maximize the PUs revenues. After increasing the spectrum price, the new revenue is computed as follows:

$$\Delta \acute{V} = \lambda(\acute{p} - p) - C. \quad (27)$$

```
AdaptSpectrumSize (u_n, f(X_t)_n, f(X_t)_{n-1}, δ, ε)
begin
if ((Abs(u_n− δ) < ε))
return f(X_t)_n, u_n;
else
        {
compute new value of u_n and f(X_t)_{n+1};
AdaptSpectrumSize (u_n, f(X_t)_{n+1}, f(X_t)_n, δ, ε);
        }
end;
```

ALGORITHM 1

This leads to the following problem formulation:

$$\max_{f(X_t)} \overline{V} = \lambda \acute{p} - C - \min_{\acute{p}} \lambda(\acute{p} - p)$$

$$\text{subject to} \quad \sum_{y=1}^{W} SP_y \le KW,$$

$$SP_y(c)SP_j(c)I_y(j) = 0, \quad (28)$$

$$B_y \le B_y^C,$$

$$\Delta V' = \lambda(\acute{p} - p) - C \ge 0.$$

In our proposed adaptation scheme the new values of spectrum prices reflect the amount of spectrum required by a PU. Due to the competition in the market, a price increment is limited due to the possibility of losing customers. If a blocking constraint for a PU is met it tries to meet the blocking constraint for SUs by increasing the offered spectrum size using the *AdaptSpectrumSize* algorithm.

6. Performance Evaluation

In this section, we show simulation results to demonstrate the ability of our spectrum scheme to adapt to different network conditions. The system of PUs and SUs is implemented as a discrete event simulation. The simulation is written by using Matlab. We uniformly distribute 10 PUs and each PU is randomly assigned 20 channels. For the mesh network, 100 MCs are distributed uniformly in the transmission region of the MRs. The results presented are for several system settings scenarios in order to show the effect of changing some of the control parameters. The network parameters chosen for evaluating the algorithm and the methodology of the simulation are shown in Table 1.

6.1. Impact of Spectrum Size and Number of Primary Users on Spectrum Borrowing among PUs. Simulations are done to explore the availability of channels that can be borrowed under different configurations of spectrum size and primary user deployment. We vary the number of PUs and the number of channels (spectrum size). We assume that two users interfere if the distance between them is less than 20 m and they use the same channel. Figure 1 shows the

TABLE 1: Simulation parameters.

Parameter	Value
Number of mesh routers	10
Number of clients	100
Number of primary users	10
Number of channels per a PU	20
Total number of channels	200
Number of messages per client	Random
Type of interface per node	802.11 b
MAC layer	IEEE 802.11 b
Transmission power	0.1 watt
Packet size	512
λ	1
Channel bandwidth	100 kHz
Blocking probably constraint for a PU	0.015
SNR	4 db

borrowing probability for different numbers of PUs. We calculate the probability of existing channels being available for borrowing. Simulations are done to investigate the effect of the number of PUs on the probability of channel borrowing. We can see that the possibility of adjusting spectrum based on borrowing is not guaranteed for a large number of PUs with a small size of spectrum. Moreover, it can be seen that increasing the number of PUs, the borrowing probability decreases due to the interference among users. The spectrum size is another factor that influences channel borrowing probability. Increasing the size of spectrum (i.e., increasing the number of channels in the system) reduces the likelihood of interference.

6.2. Performance of On-Demand Sharing Scheme. We compare the performance of our on-demand-based spectrum sharing scheme with the poverty-line heuristic [16] through simulations. For PU_y, the poverty line is computed as follows:
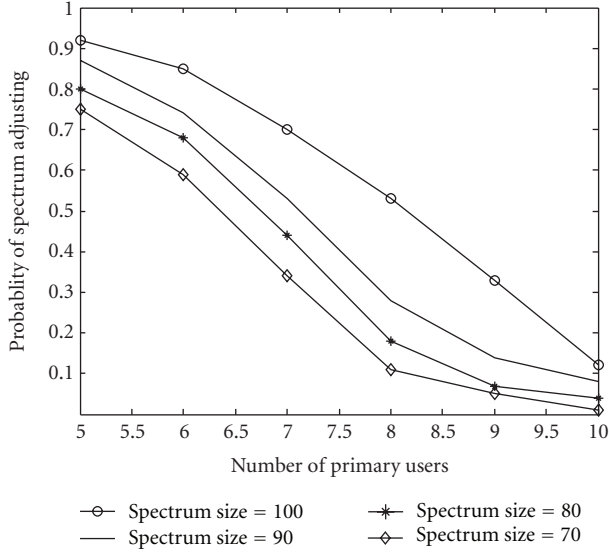
$$PL(y) = \frac{L(PU_y)}{NG(PU_y)}. \quad (29)$$
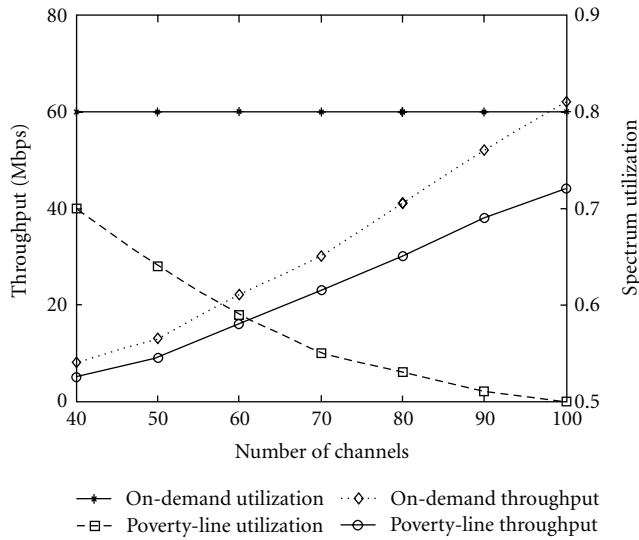
FIGURE 1: Probability of spectrum adjusting.

FIGURE 2: Throughput and spectrum utilization comparison.

The performance metrics considered are as follows.

(1) Throughput, which is the average rate of successful message delivery over a communication channel which can be expressed as follows:

$$\text{Throughput} = \frac{\text{total no. of bytes received}}{\text{simulation time}}. \quad (30)$$

(2) Spectrum utilization, S, which is the percentage of busy spectrum at time t and is computed as follows:

$$S = \frac{\sum_{w=1}^{W} \text{SP}_w}{KW}. \quad (31)$$

We examine the performance under different parameter settings. Throughput comparison of the two schemes is
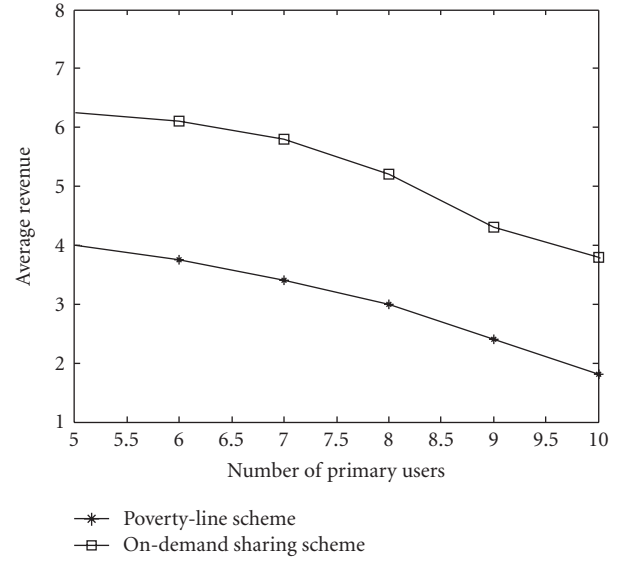
FIGURE 3: Average revenue sensitivity versus number of PUs in cognitive network.

shown in Figure 2. The figure shows that the throughput increases as the total number of channels increases. This is due to more spectrum that can be employed. Our scheme utilizes the unused spectrum resourcefully because there is no limit to channel borrowing among PUs. For poverty-line heuristic [16], a PU cannot exceed a certain number of channels that can be borrowed from its neighbors even if the neighbors have idle channels.

We further present the results of spectrum utilization with different spectrum sizes in Figure 2. Our scheme performs better than the poverty-line heuristic. Our scheme utilizes the whole spectrum because PUs can have access to neighbor's channels based on availability of channels and on-demand. This improves the cognitive network throughput and overall spectrum utilization. However, some unused spectrum is not utilized under poverty-line heuristic because of the threshold constraint. It is clear from Figure 2 that our scheme is not sensitive to the number of channels in the network. However, the only constraint that prevents our scheme from full utilization of spectrum is the interference factor. In the poverty-line-based scheme, spectrum sharing is limited by the poverty line that depends on the number of idle channels. From the figure, we can see that as the number of channels increases the utilization of channels decreases because of an increment in idle channels.

Figure 3 displays the result of spectrum trading. The result shows that our scheme achieves higher revenue than poverty-line scheme. The revenue decreases as the number of PUs increases, since in this case PUs are assigned less number of channels; therefore the size of offered spectrum will decrease. We also compare the performance of the two schemes under varying traffic load in Figure 4. The result shows that the revenue increases as the spectrum demand increases.
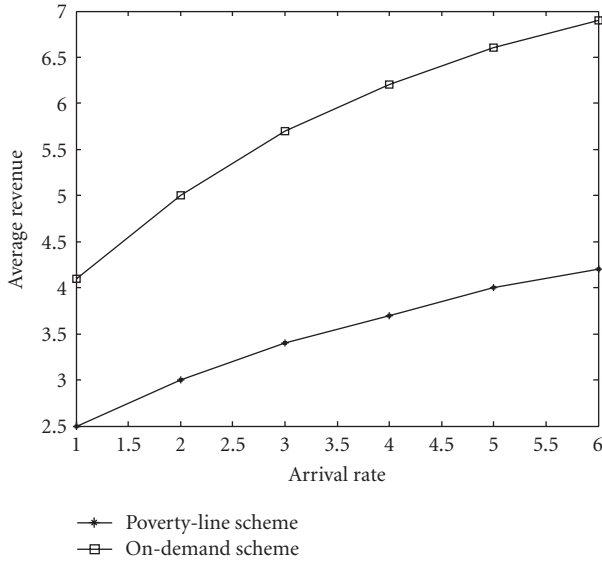
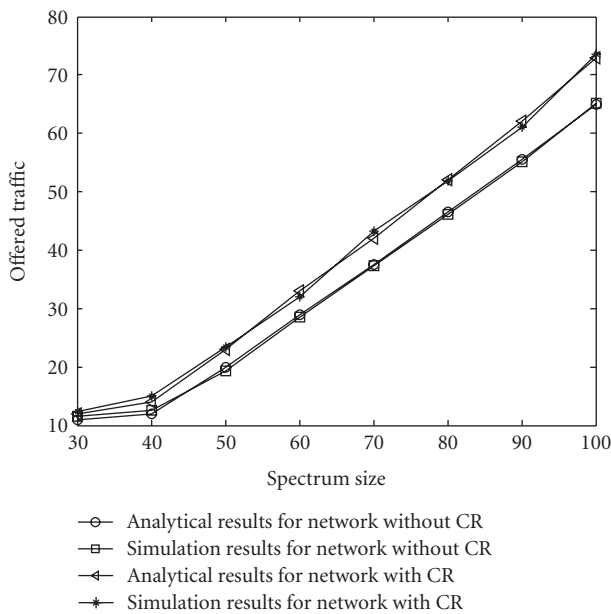FIGURE 4: Average revenue sensitivity versus arrival rate.

FIGURE 6: Adapting spectrum size to meet blocking probability constraint.

FIGURE 5: Offered traffic for WMNs with and without CR abilities (blocking probability).

have a priority over SUs. Each PU needs a spectrum for its usage and to support the maximum classic traffic for ($B_y^C \leq$ 1%) constraint. If an additional network overlays its traffic over the unused spectrum it should not affect the B_y^C of the PUs.

6.4. Spectrum Price Adaptation. A PU with well-dimensioned spectrum size and correctly chosen spectrum price provides the desired QoS and maintains blocking probabilities in an acceptable range. When the spectrum demand increases, blocking probabilities normally increase beyond their constraints. While our adaptation scheme tries to maximize PUs' revenues by increasing spectrum size when the spectrum demand increases, it maintains QoS by bringing blocking probabilities back to its constrained range by increasing the spectrum price. Our intelligent algorithm is converged after 4 steps. Figure 6 displays the offered spectrum size at PU$_y$ for different arrival rates. When spectrum arrival rate is increased and blocking probability does not surpass B_y^C, PU$_y$ adapts by increasing the size of the offered spectrum as shown in the figure to generate more revenue. However, when the demand decreases, PU$_y$ reduces the size of the offered spectrum to avoid a waste of spectrum.

We study the effect of spectrum adaptation on the gained revenue for different offered spectrum sizes in Figure 7. The results show that our algorithm increases the offered spectrum size to gain more revenue. When the offered spectrum becomes large the quality of service of PU may be degraded because of the reduction of its spectrum size. Therefore, the adaptation scheme stop increasing the offered spectrum. Figure 7 also shows the size of offered spectrum for different service costs. It is clear the adaptation scheme offers more spectrum when the cost of serving SUs is low. When a PU$_y$ offers large size of spectrum, its blocking probability B_y may surpass its blocking constraint B_y^C. The spectrum price adaptation is integrated in our adaptation

6.3. Supporting QoS for SUs in CRs. In this section, we explore the performance of WMNs with cognitive abilities. CRs take advantage of surplus spectrum by renting it to the SUs and getting profits. Figure 5 shows a comparison between the traffic for WMNs with CR abilities and WMNs without CR abilities. Clearly, the cognitive systems outperform the classic WMNs that do not use CR technology. The main disadvantage of CRs is the waiting time of flows. This is a direct consequence of the PUs requirement of not renting a surplus spectrum if there is no revenue. However, despite the PUs requirement, the overall performance is far better when CR is enabled. CRs cannot guarantee QoS because PUs flows
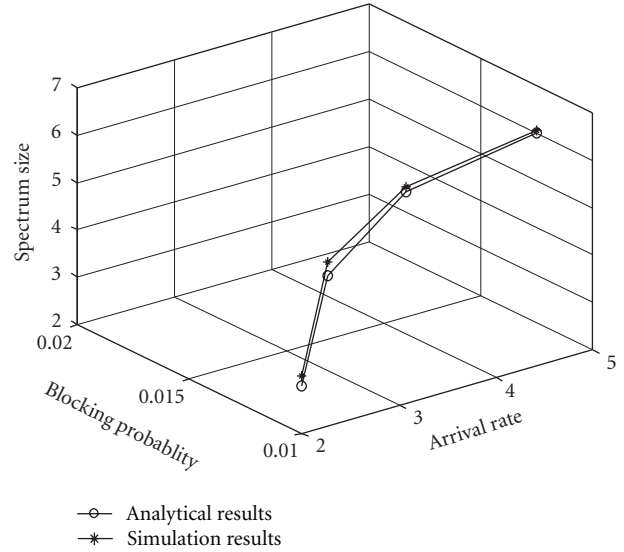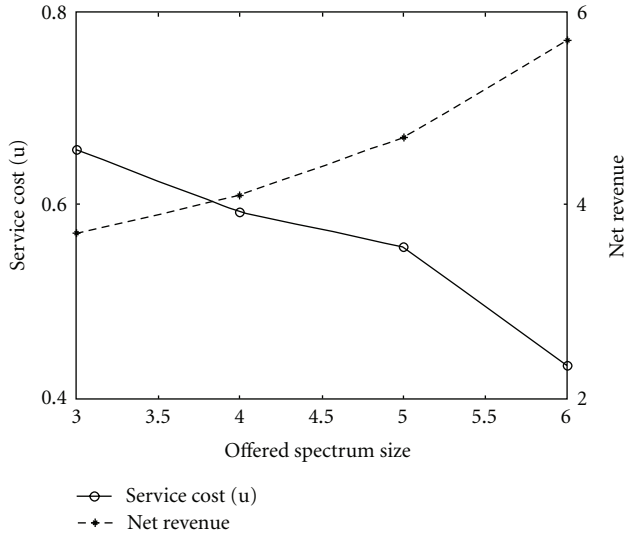
FIGURE 7: Adapting spectrum size for different service cost and the gained average revenue for the adaptation scheme.

FIGURE 9: Offered spectrum for different blocking network constrains.
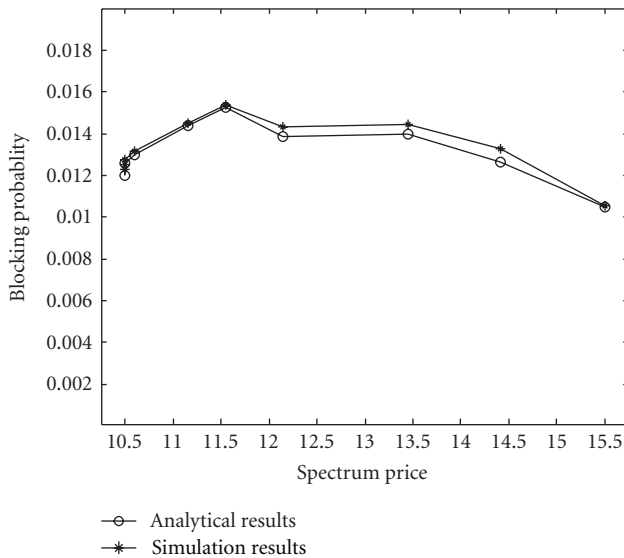

FIGURE 8: Adapting spectrum price to meet blocking probability constraint.

process to ensure it meets the blocking constraints. Figure 8 shows the spectrum price adaptation when the blocking probability surpasses blocking constraint. PU increases the price of spectrum to decrease the accepting rate for each SUs class and to maintain QoS for PUs. The results show our scheme's ability to bring blocking probabilities back to their constrained range by adapting spectrum price.

6.5. Tradeoffs between a PU Revenue and QoS Constraints.
Figure 9 plots the tradeoff between a PU revenue and its QoS. To show the relationship between the two, we vary the blocking probability constraint for a PU (the QoS requirement for a PU). Blocking constraint becomes stricter in such a way that more in-service primary users should be
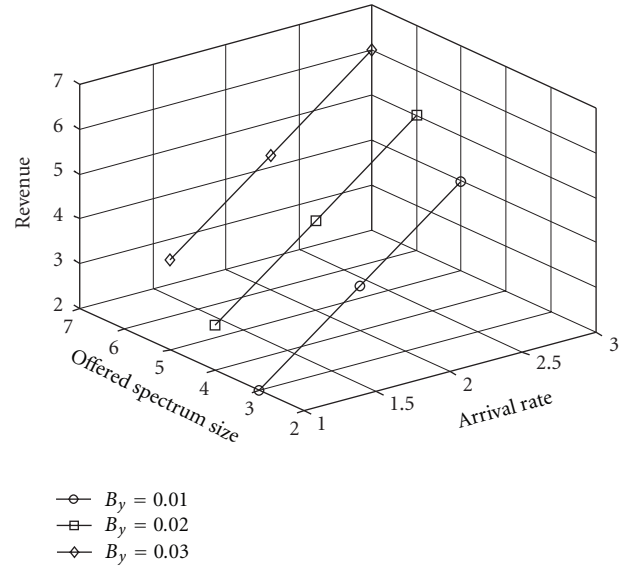
protected from channel eviction. For this, SUs arrivals must be blocked more often and the rejection ratio is increased. As a result, a PU cannot offer more spectrum for a small value of blocking probability. However, as this constraint is relaxed, a PU can serve more SUs and can offer more spectrum to generate more revenue. For large values of blocking probability, a PU can maintain a QoS for its applications and this can be observed from the figure. The revenue gained for large values of blocking probability is increased and a PU becomes less strict so that a lower number of SUs are rejected upon their arrival.

6.6. Optimal Policy as a Function of Spectrum Price, Cost, and Quality.
We simulate the behavior of the described system under different spectrum prices. Figure 10 displays the size of the offered spectrum for different service prices. From the figure, we can clearly see that even though spectrum prices are higher a PU may increase the offered spectrum size. There is a direct correlation between the offered spectrum size and the spectrum price, so the more reward we have (due to price) the more spectrum PU can offer SUs. However, a PU cannot further increase the price because it will affect SU's spectrum demand.

We compare the same system for different cost of service (δ) in Figure 11 for a fixed spectrum price. From the figure, we clearly notice how sensitive the optimal size of offered spectrum is to the spectrum cost, where the offered size drops as spectrum cost increases. Figure 12 shows the offered spectrum size as a function of spectrum quality (ω). It is clear as the spectrum quality improves, the PU will offer more spectrum increasing its reward.
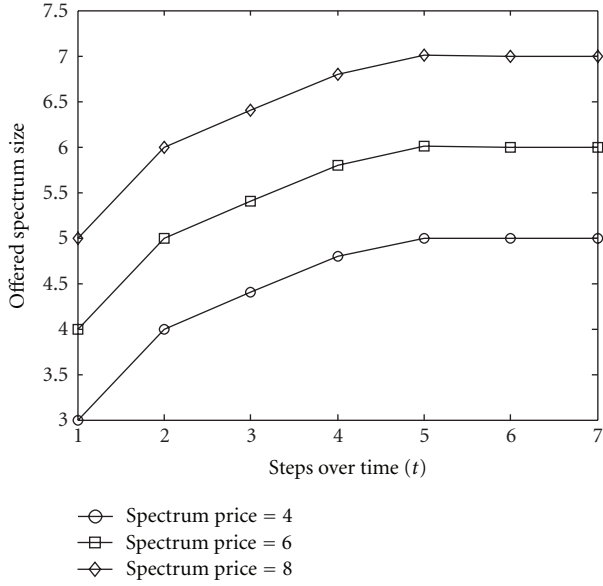
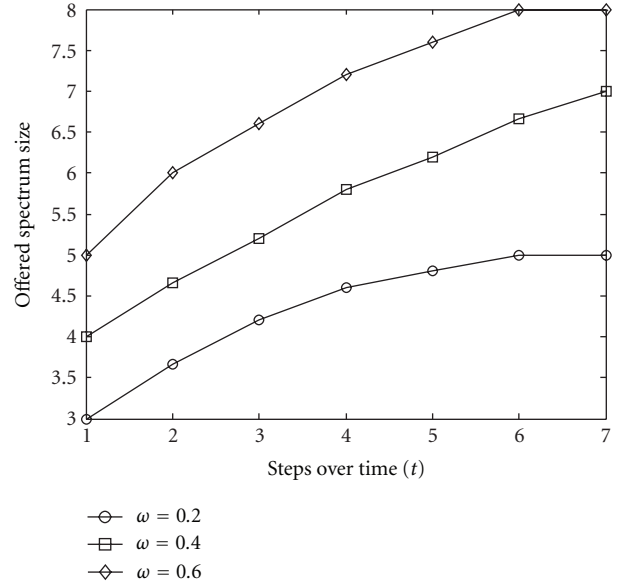FIGURE 10: Optimal policy as a function of spectrum prices.

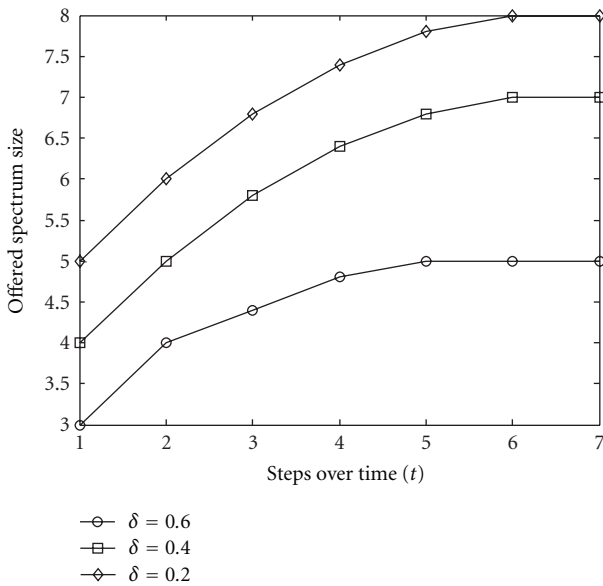FIGURE 12: Offered spectrum size as a function of spectrum quality.

FIGURE 11: Offered spectrum size for different spectrum cost.

7. Conclusion

In this paper, we present a novel machine-learning-based model to obtain an optimal policy for controlling spectrum trading in cognitive wireless networks. The proposed model has two contributions to cognitive networks. From the application side, the main contribution is developing a control policy that considers different requirements such as rewards for PUs, wireless requirement (channel inter-ference), the cost of spectrum renting, and PUs QoS. All basic functions are integrated and optimized into one homogenous, theoretically based model. From the modeling side, we formulate a spectrum trading problem as a reward maximization problem. Such a formulation allows RL to optimize the trading problem. The approach presents a general framework for studying, analyzing, and optimizing other resource management in cognitive mesh networks.

Another contribution is to propose a new scheme for the PUs to control spectrum trading for the emerging spectrum secondary market. PUs can employ the proposed scheme to choose the optimal price and size of the offered spectrum. The objective is to adapt the size and price of spectrum in order to continuously maximize PUs' net revenues while maintaining PUs' QoS. Simulations were also conducted and shown to closely agree with the analytical model. They demonstrated the ability of our algorithm to support SUs requirements and obtain the potential performance gains by applying cognitive radio. Moreover, the numerical results show that the proposed approach is able to find an efficient tradeoff between different rates of spectrum size and different costs of spectrum. The results show the ability of our scheme to find the optimal spectrum size for different spectrum prices. We vary system parameters to understand the behavior of the system under different scenarios. The results show a direct correlation between the reward rates and the spectrum price, and an inverse relationship between the spectrum cost and the allocated bandwidth. We also propose a new distributed spectrum sharing scheme among primary users. PUs share spectrum based on demand whereby they can borrow spectrum from their neighbors while complying with interference rules. The benchmark in our experiments is the poverty-line heuristic used in [16]. Because it can more efficiently employ limited spectrum resources compared to the poverty-line heuristic, our scheme achieves higher net revenues. The poverty-line heuristic restricts borrowing by a threshold called poverty line. Moreover, numerical results

show that our scheme is able to find an efficient tradeoff between PU revenues and SUs delay.

References

[1] 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*, vol. 50, no. 13, pp. 2127–2159, 2006.

[2] E. Hossain and V. K. Bhargava, *Cognitive Wireless Communication Networks*, Springer, New York, NY, USA, 2007.

[3] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks*, vol. 47, no. 4, pp. 445–487, 2005.

[4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, Mass, USA, 1998.

[5] R. K. Lam, J. C. S. Lui, and C. Dah-Ming, "On the access pricing issues of wireless mesh networks," in *Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS '06)*, pp. 61–61, 2006.

[6] Z. Ji and K. J. R. Liu, "Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users," in *Proceedings of the 3rd Annual IEEE Communications Society on Sensor and Ad hoc Communications and Networks (SECON '06)*, pp. 119–127, September 2006.

[7] O. Simeone, I. Stanojev, S. Savazzi, Y. Bar-Ness, U. Spagnolini, and R. Pickholtz, "Spectrum leasing to cooperating secondary ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 203–213, 2008.

[8] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: a dynamic game approach," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2651–2660, 2008.

[9] D. Niyato and E. Hossain, "Market-equilibrium, competitive, and cooperative pricing for spectrum sharing in cognitive radio networks: analysis and comparison," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4273–4283, 2008.

[10] M. M. Bykowsky, M. Olson, and W. W. Sharkey, "Efficiency gains from using a market approach to spectrum management," *Information Economics and Policy*, vol. 22, no. 1, pp. 73–90, 2010.

[11] H. Sartono, Y. H. Chew, W. H. Chin, and C. Yuen, "Joint demand and supply auction pricing strategy in dynamic spectrum sharing," in *Proceedings of the IEEE 20th Personal, Indoor and Mobile Radio Communications Symposium (PIMRC '09)*, pp. 833–837, September 2009.

[12] S. Sengupta and M. Chatterjee, "Sequential and concurrent auction mechanisms for dynamic spectrum access," in *Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom '07)*, pp. 448–455, August 2007.

[13] O. Ileri, D. Samardzija, T. Sizer, and N. B. Mandayam, "Demand responsive pricing and competitive spectrum allocation via a spectrum server," in *Proceedings of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN '05)*, pp. 194–202, November 2005.

[14] G. Işiklar and A. B. Bener, "Brokering and pricing architecture over cognitive radio wireless networks," in *Proceedings of the 5th IEEE Consumer Communications and Networking Conference (CCNC '08)*, pp. 1004–1008, January 2008.

[15] M. M. Buddhikot, P. Kolody, S. Miller, K. Ryan, and J. Evans, "DIMSUMNet: new directions in wireless networking using coordinated dynamic spectrum access," in *Proceedings of the IEEE WoWMoM*, pp. 78–85, 2005.

[16] L. Cao and H. Zheng, "Distributed rule-regulated spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 130–145, 2008.

[17] A. Alsarhan and A. Agarwal, "Cluster-based spectrum management using cognitive radios in wireless mesh network," in *Proceedings of the 18th International Conference on Computer Communications and Networks (ICCCN '09)*, San Francisco, Calif, USA, August 2009.

[18] P. Beckmann, *Elementary Queuing Theory and Telephone Traffic*, Series on Telephone Traffic, Lee's ABC of the Telephone, Geneva, Switzerland, 1977.

[19] G. Gallego and G. van Ryzin, "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science*, vol. 40, no. 8, pp. 999–1020, 1994.