QoS Provisioning Transmission Obtained by a Stochastic Game in Cognitive Radio Networks

Jerzy Martyna
Faculty of Mathematics and Computer Science
Institute of Computer Science, Jagiellonian University
ul. Prof. S. Łojasiewicza 6, 30-348 Cracow, Poland
Email: jerzy.martyna@ii.uj.edu.pl

Abstract—In this paper, we propose a stochastic game guaranteeing the QoS provisioning transmission in cognitive radio networks. Our approach is able to improve spectrum utilisation and bring monetary gains for secondary users. At each stage of the game, secondary users observe the spectrum availability, channel quality and the strategy of the QoS provisioning transmission for all players. These players can also include secondary users belonging to multiple classes. According to this observation, they will decide how many channels should be reserved for transmitting data within the required QoS parameters. By using the Q-learning algorithm, secondary users can learn the optimal policy that maximises the expected payoff sum. It is shown that performance gains through the stochastic game can be used as the method of QoS provisioning in cognitive radio networks.

I. INTRODUCTION

Cognitive radio (CR) networks belong to systems that increase spectrum utilisation efficiency. The computers in these networks allow, among others, switching between radio access technologies and transmitting different portions of the radio spectrum as unused frequency band slots [1], [2]. This dynamic spectrum access increases spectrum utilisation efficiency, which determines parameters such as optimal signal format, frequency band and modulation. It utilises idle resources that change by geographic location and time without any interferences.

A number of papers proposed the study of spectrum sensing, spectrum pooling, dynamic frequency hopping and transmission power control in cognitive radio networks. These include the paper by Z. Quan et al. [3], which introduced a novel wideband spectrum sensing technique called multiband joint detection, which jointly detects signal energy levels over multiple frequency bands rather than considering one band at a time. The authors of the paper [4] studied the problem of designing a sensing duration to maximise the achievable throughput for secondary networks under the constraint that primary users are sufficiently protected. They used an energy detection-sensing scheme to prove that the formulated problem indeed has one optimal sensing time that yields the highest throughput for the secondary network. The problem of opportunistic access to parallel channels occupied by primary users under a continuous-time Markov chain modelling of channel occupancy by primary users was analysed by Q. Zhao et al. [5]. As a result, the authors proposed a slotted transmission strategy for secondary users using a periodic sensing strategy with optimal dynamic access to the spectrum.

As given above, spectrum sharing and spectrum sensing in cognitive radio networks have been extensively studied over the past few years. However, the problem of providing quality of service (QoS) guarantees over cognitive radio channels has not been sufficiently considered. It is obvious that providing certain QoS assurances is crucial in cognitive radio networks. Note that in many situations, the primary users determine transmission at lower power levels. Thus, the performance of cognitive radio networks under QoS constraints depends on the buffer constraints and limiting the delay.

Stochastic game theory is an essential tool for cognitive radio networks. Among the game-theoretic approaches to addressing resource management in these networks, stochastic game is able to exploit the correlated channels in analysing decentralised behaviours of cognitive radios. Tembino et al. [6] demonstrate the existence of equilibriums and conditions for evolutionarily stable strategies under good and bad weather conditions based on a stochastic game for modelling the remaining energy of the battery for each radio device. The secondary use rate adaptation problem in cognitive radio networks with a constrained general-sum switching control Markovian dynamic game as the original problem, as considered by J.W. Huang and V. Krishnamurthy [7], has been transformed and solved using the Nash equilibrium policy.

In this study, we propose a stochastic game framework for the QoS of service provisioning over cognitive radio channels. Secondary users perform channel sensing to detect the activity of primary users. Depending on the presence or absence of active primary users, the secondary users transmit the data at two average power levels. The contributions of this paper are the following:

1) Formulating a state-transition model for QoS service provisioning over cognitive radio channels.
2) Providing a stochastic game for QoS service provisioning.
3) Secondary users can learn the optimal policy maximising the expected sum of discounted payoffs (defined as spectrum-efficient throughput) by using the Q-learning algorithm.
4) Incorporating the power and rate adaptation into the model by considering different assumptions on the availability of channel side information (CSI) at the transmitter.
In Section 3, a stochastic game is formulated by defining the states, actions and objective functions as well as the state transition rules. In Section 4, the optimal policy of the secondary user using the $Q$-learning algorithm is given. In Section 5, the simulation results are presented. Finally, the conclusion is given in Section 6.

II. SYSTEM MODEL

In this section, we present the model of the secondary user network, the cognitive transmission under QoS constraints and the effective capacity concept.

A. Secondary User Network

We assume that there is a secondary base station in the CR network that coordinates the spectrum usage for all secondary users and secondary users of second type (class). In our approach, the secondary network is a time-slotted system. To avoid interference with primary users, the secondary and secondary users of second type need to listen to the spectrum before their transmissions. Thus, the secondary and secondary users of second type can be omitted from the occupied slots. We assume a perfect sensing of the currently unused licensed spectrum and vacating the spectrum.

To achieve efficient spectrum utilisation, control messages are exchanged between the secondary base station and the secondary users through control channels. The control messages are associated with situations such as channel assignment, spectrum handoff, etc. Similarly, control messages are also exchanged in similar situations between secrecy users and the secondary base station through the control channels. If the control messages are not correctly received, the characteristics of some functions can be violated.

B. Cognitive Channel Model

In this paper, we consider a cognitive radio channel model in which a secondary transmitter attempts to send data to a secondary receiver with primary users present. The secondary users test channel activity. If the secondary transmitter selects its transmission when the channel is busy, the average power is $P_1$ and the rate is $r_1$. When the channel is idle, the average power is $P_2$ and the rate is $r_2$. We assume that $P_1 = 0$ represents the stoppage of the secondary transmission in the presence of an active primary user. Both transmission rates, $r_1$ and $r_2$, can be fixed or time-variant depending on whether the transmitter has channel side information or not. In general, we assume that $P_1 < P_2$. In the above model, the discrete-time channel input-output relation in the absence of the channel of the primary users is given by

$$y(i) = h(i)x(i) + n(i), \quad i = 1, 2, \ldots$$

where $i$ is the symbol duration. If primary users are present in the channel, the discrete-time channel input-output relation is given by

$$y(i) = h(i)x(i) + s_p(i) + n(i), \quad i = 1, 2, \ldots$$

where $s_p(i)$ represents the sum of the active primary users’ faded signals arriving at the secondary receiver $n(i)$ is the additive thermal noise at the receiver and is zero-mean, circularly symmetric, complex Gaussian random variable with variance $E\{\left|n(i)\right|^2\} = \sigma_n^2$ for all $i$.

We assume that the receiver knows the instantaneous lambda values $\{h(i)\}$, while the transmitter has no such knowledge. We construct a state-transition model for cognitive transmission by considering the cases in which the fixed transmission rates are lesser or greater than the instantaneous channel capacity values. In particular, the ON state is achieved if the fixed rate is smaller than the instantaneous channel capacity. Otherwise, the OFF state occurs.

Thus, we have the following four possible scenarios associated with the decision of channel sensing, namely [8]

1) channel is busy, detected as busy (correct detection),
2) channel is busy, detected as idle (miss-detection),
3) channel is idle, detected as busy (false alarm),
4) channel is idle, detected as idle (correct detection).

If the channel is detected as busy, the secondary transmitter sends with power $P_1$. Otherwise, it transmits with a larger power, $P_2$. In the above four scenarios, we have the instantaneous channel capacity, namely

$$C_1 = B \log_2(1 + SNR_1 \cdot z(i))$$  \quad \text{channel is busy, detected as busy (3)}

$$C_2 = B \log_2(1 + SNR_2 \cdot z(i))$$  \quad \text{channel is busy, detected as idle (4)}

$$C_3 = B \log_2(1 + SNR_3 \cdot z(i))$$  \quad \text{channel is idle, detected as busy (5)}

$$C_4 = B \log_2(1 + SNR_4 \cdot z(i))$$  \quad \text{channel is idle, detected as idle (6)}

where $z(i) = [h(i)]^2$, $SNR_i$ for $i = 1, \ldots, 4$ denotes the average signal-to-noise ratio (SNR) values in each possible scenario.

The cognitive transmission is associated with the ON state in scenarios 1 and 3, when the fixed rates are below the instantaneous capacity values ($r_1 < C_1$ or $r_2 < C_2$). Otherwise, reliable communication is not obtained when the transmission is in the OFF state in scenarios 2 and 4. Thus, the fixed rates above are the instantaneous capacity values ($r_1 \geq C_1$ or $r_2 \geq C_2$). The above channel model has 8
A secondary user can utilise unused spectrum bands belonging to \( L \) primary users. We assume that the bandwidth of licensed bands may be different, and each licensed band is partitioned into a set of adjacent channels with the same bandwidth. Thus, we can denote \( N_l \) channels in the primary user \( l \)‘s band. In our approach, when the primary user is active at time \( t \) in the \( l \)-th band, this is denoted by \( P_l^t = 1 \). Otherwise, the state is defined as \( P_l^t = 0 \).

Since the channel is modelled as a finite-state Markov chain (FSMC), the channel quality in terms of SNR of the 1st band can be expressed by FSMC. Thus, the achievable gain of the licensed band depends on the primary users's status \((P_l^t = 1\) when the primary user uses the 1st band at any time \( t \), otherwise \( P_l^t = 0 \).) Thus, each state of the FSMC is jointly modelled by the pair \((P_l^t, q_l^t)\), where \( q_l^t \) is the channel quality. The channel quality can take any value from a set of discrete values, i.e. \( q_l^t \in \{SNR_1,\ldots,SNR_R\} \).

Consider the scenario with a two type of secondary users belonging to two classes. The actions of the secondary users from the first class of the secondary users can be defined as \( a^1_l = (a^1_{l,D},a^1_{l,C},a^1_{l,D_1},a^1_{l,C_1}) \). The action \( a^1_{l,D} \) (or \( a^1_{l,C} \)) denotes that the secondary network will transmit data (control) messages at channels uniformly selected at time slot \( t \). Next, the action \( a^2_{l,D_2} \) (or \( a^2_{l,C_2} \)) indicates that the secondary network will transmit data (control) messages in the \( a^1_{l,D_2} \) (or \( a^1_{l,C_2} \)) channel selected from the previously used channels without success.

Similarly, the action of the secondary users belonging to the second type of the secondary users is defined \( a^2_{l} = (a^2_{l,D_1},a^2_{l,K_1},a^2_{l,D},a^2_{l,K_1}) \), where the action \( a^2_{l,D} \) (or \( a^2_{l,C_1} \)) denotes the secondary network will transmit data (control) messages at channels uniformly distributed at time slot \( t \). Analogous, the action \( a^2_{l,D_2} \) (or \( a^2_{l,C_2} \)) denotes that the secondary users will transmit the data (control) messages in the \( a^2_{l,D_2} \) (or \( a^2_{l,C_2} \)) channel from previously used channel without success.

After defining the state at each stage, we may provide the state transition rule, namely assuming that secondary users should observe which channel has been occupied by secondary users. Based on these observations, the secondary users can define the pair \((S_{l,D}^t,S_{l,C}^t)\), where \( S_{l,D}^t \) and \( S_{l,C}^t \) denote data and control channel numbers being used by secondary users of the second class in the 1st band observed at time slot \( t \). We assume that the secondary users cannot be informed as to whether an idle channel is occupied or not by the secondary users from the second class. Thus, the number of idle channels that are not being engaged by the secondary users of the second class is not an observation by the secondary users from the first class.

Thus, at every time slot time \( t \), the state of the stochastic game \( G \) is defined by \( s_l^t = \{s_{l,D}^t,s_{l,C}^t,\ldots,s_{l,C}^t\} \) where \( a^1_{l} = (P_l^t,q_l^t,S_{l,D}^t,S_{l,C}^t) \) indicates the state associated with band \( l \) (\( l \in \{1,\ldots,L\} \)).

After defining the state at each stage, we may provide the state transition rule, namely
probability of the number of secondary users of second type
t is sufficient for coordinating the spectrum management in
the selected channels and the secondary users belonging to
the secondary users will transmit data and control messages
secondary users. After the all players choose their actions,
of the primary user status and the channel conditions.

We will now consider the scenario with a two type of
users to obtain the optimal policy of the stochastic game is
independent of time, the policy is said to be stationary.

is said to be Markov. If the policy
independent of the states and actions in all previous states and
actions. Then, the policy \( \pi \) is said to be Markov. If the policy
is equal to \( 0 \) and the probability of detection is equal to 1

\[
p(s^t, a^t, a^t_S) = \prod_{i=1}^L p(s^{t+1}_i | s^t_i, a^t_i, a^t_{i,S}) \tag{8}
\]

\( p(s^t_i | s^t_i, a^t_i, a^t_{i,S}) \) can be further expressed by

\[
p(s^{t+1}_i | s^t_i, a^t_i, a^t_{i,S}) = p(s^{t+1}_{i,D}, s^{t+1}_{i,C} | s^t_{i,D}, s^t_{i,C}, a^t_i, a^t_{i,S}) \times p(P^{t+1}_i, g^{t+1}_i | P_t^i, g_t^i) \tag{9}
\]

where the first term on the right side represents the transition
probability of the number of secondary users of second type and
data channels, and the second term denotes the transition of
the primary user status and the channel conditions.

We will now consider the scenario with a two type of
users. After the all players choose their actions,
the secondary users will transmit data and control messages
in the selected channels and the secondary users belonging to
the second type will intercept their channels. We assume that
the same control messages are transmitted in all the control
channels, and one correct copy of control information at time
in the next time slot.

We assume that the stage payoff of the secondary users
maximizes the spectrum gain, namely

\[
r(s^t, a^t, a^t_S) = T(s^t, a^t, a^t_S) \times (1 - p^{\text{block}}(s^t, a^t, a^t_S)) \tag{10}
\]

where \( T(s^t, a^t, a^t_S) \) indicates the expected spectrum
when not all control channels get intercept and
\( p^{\text{block}}(s^t, a^t, a^t_S) \) denotes the probability that all control
channels in all \( L \) bands are intercepted.

IV. THE MINIMAX-Q LEARNING TO OBTAIN THE
OPTIMAL POLICY OF THE STOCHASTIC GAME

In this section, the minimax-Q learning for the secondary
users to obtain the optimal policy of the stochastic game is
presented.

In general, the secondary users treat the payoff in different
stages differently. Then, the secondary users’ objective is
find an optimal policy that maximizes the expected sum of payoffs

\[
\max E\{\sum_{t=0}^{\infty} \beta^t r(s^t, a^t, a^t_S)\} \tag{11}
\]

where \( \beta \) is the discount factor of the secondary user. In our
approach, the policy of the secondary network is expressed by
\( \pi : S \rightarrow \mathcal{PD} (A) \) and the policy of the secondary users
of second type \( \pi_S : S \rightarrow \mathcal{PD} (A_S) \), where \( s^t \in S, a^t \in A, a^t_S \in A_S. \) It is noticeable that the policy \( \pi^t \) at time \( t \)
is independent of the states and actions in all previous states and
actions. Then, the policy \( \pi \) is said to be Markov. If the policy
is independent of time, the policy is said to be stationary.

In the stochastic game between the secondary users and
the secondary users of second type is a zero-sum game, the
equilibrium of each stage is the minimax equilibrium. Tosolve
the game, we can use the minimax-Q learning method [10],
[11]. The \( Q \)-function of stage \( t \) is defined as the expected
discounted payoffs when the secondary users take action \( a^t \)
and the secondary users of second type take the action \( a^t_S \).
Then the \( Q \)-value in the minimax-Q learning of the game can
be expressed as

\[
Q(S^t, a^t, a^t_S) = r(s^t, a^t, a^t_S) + \beta \sum_{s} p(s^{t+1} | s^t, a^t, a^t_S) V(s^{t+1}) \tag{12}
\]

where \( V(s^{t+1}) \) is the value of a state in the game of
secondary users of second type.

V. SIMULATION RESULTS

We conduct simulations to evaluate performance in QoS
provisioning by a stochastic game. Firstly, we check the
convergence of the minimax-Q learning algorithm and analyze
the strategy of all secondary users for several stages.

We assumed that the observation time is equal to 1 sec
and the channel bandwidth is equal to 100 kHz. Moreover,
we assumed that the QoS exponent is \( \theta = 0.01 \) and the
average SNR values when the channel is detected correctly
are \( S N R_{1} = 0 \) dB and \( S N R_{1} = 10 \) dB for busy and idle
channels, respectively. In Fig. 2, we plot the effective capacity
as a function of the detection threshold value \( \lambda \). As we see in
Fig. 2 the effective capacity is increasing with increasing the
detection threshold value \( \lambda \).

In Fig. 3, we plot the effective capacity as a function of
the QoS exponent obtained for both classes of the secondary
users under the assumption that the probability of false alarm
is equal to 0 and the probability of detection is equal to 1

Fig. 2. Effective capacity as a function of the detection threshold value
for the secondary users of both classes.
VI. CONCLUSIONS

In this paper, we studied the stochastic game in cognitive radio networks with multiple classes of secondary users. Considering the spectrum environment as time-varying and that each group of secondary users is able to use an adaptive strategy, the provisioning of the QoS parameters is identified by finding the effective capacity of the cognitive radio channel. Simulation results show that an optimal policy can be obtained by using the minimax-$Q$ learning algorithm in a stochastic game. In particular, it is visible in the OFF state in cases of misdetection where the channel is detected as idle but is actually busy. Thus, the stochastic game provides a higher value of detection probability and also decreases the probability of false alarms in comparison to other methods. We also observed that, in the stochastic game, the exponent $\theta$ increases due to adapting rates and diminishing power, and thus the QoS constraints are more stringent.

REFERENCES