Co-Operation based Resource Selection in Cognitive Radio Network via Potential Games

K. Poongodi*, Hiran Kumar Singh and Dhananjay Kumar
Department of Information Technology, Anna University, MIT Campus, Chennai, Tamil Nadu, India; dk.poongodi@gmail.com, hksbxr@gmail.com, dhananjay@annauniv.edu

Abstract
Cognitive Radio is a forthcoming technique to improvise the utilization of radio frequency spectrum in wireless network. However, Cognitive Radio Network has several challenges such as channel selection, efficient spectrum sharing, network throughput, etc. This paper presents a resource (channel) selection strategy by the Secondary Users (SUs) in a dynamic environment via game theoretic approach namely potential games. A distributed, Stochastic Learning based Resource Selection (SLRS) and Q-Learning based Resource Selection (QLRS) are the two different algorithms proposed here. The strategy followed by the SU is based on its own action-reward history, even without knowing the actions in other SUs. The simulation results prove that the QLRS algorithm achieves higher throughput and fairness performances than the SLRS algorithm.

Keywords: Cognitive Radio, Channel Selection, Q-Learning, Stochastic Learning, Spectrum Sharing

1. Introduction
The expeditious increase of wireless applications has accomplished the single-network wireless system is inadequate in meeting the traffic demands due to the incompetent spectrum utilization. The supreme responsibility of the cognitive radio is the intellectual reasoning in which choice of a set of actions that lead to efficient decision making. In addition to this, the process of learning must be powerful enough to enrich the knowledge base in the subsequent reasoning\(^1\). But how are potential games related to our discussion of learning in networks of cognitive radios? Because potential games have desirable properties in terms of existence of Nash equilibrium and it should be converged through simple adaptations.

For instance, all finite potential games have at least one Nash equilibrium in pure strategies i.e., a finite game is a game where the player and action sets are finite. More generally, if the strategy space for the game is compact and the potential function is continuous, then the game has at least one pure strategy of Nash equilibrium. Just as importantly, from the point of view of learning, is that the players are guaranteed to reach these equilibria through best response and better response dynamics.

Some of the seminal work in applying potential games to cognitive radio problems was done by Neel\(^2\). A number of problems in multi-channel communications can be modeled as potential games. For example, when the utility function of each radio considers the social welfare of the network as a potential function naturally emerges. This is the case in the work on channel selection by Nie et al.,\(^3\). A distributed hybrid learning for 4G heterogeneous networks and the convergence of NE was demonstrated without indicating the whether or not the achieved strategy profile is an equilibrium point by Khan et al.,\(^4\).

The paper is organized as follows: Section II provides the Literature review. Section III includes a model of the system. Section IV explains the detailed description of proposed system. Section V explores the numerical results of the implementation and result analysis of proposed work. Section VI gives conclusion and future work of this paper.

*Author for correspondence
2. Related Work

This section includes the survey in the field of cognitive radio network on spectrum sharing using potential games. It includes some of the existing spectrum sharing techniques which could provide an efficient spectral utilization of the network. These schemes serve as a foundation towards the implementation of the proposed system.

The self-organizing mechanism for the small cell networks to maximize the spectral efficiency by using the reinforcement learning by Mehdi Bennis et al. They used the potential games to formulate the game model in which the convergence made towards the epsilon Nash equilibrium. The proposed learning algorithm has been validated using two utility functions, any well defined utility function with a finite set of transmission strategies can be considered. The Q-learning based channel selection approaches to improve the fairness among cognitive nodes. They proved that learning based channel selection algorithm converge to a Nash equilibrium point for nodes having unbalanced arrival packet rate in multi party multi agent stochastic games by Amiotosh Ghosh et al. However, this scheme does not support to improve the throughput of secondary users.

A competitive spectrum sharing scheme is proposed by Dusit Niyato et al. based on potential games. They have modeled this spectrum sharing as an oligopoly market and a game has been used to obtain the Nash equilibrium for the optimal allocated spectrum size for the Secondary Users (SUs). This improves the marginal profit of the primary users and at the same time it improves the SUs spectrum usage based on learning rate and channel quality information. Nonetheless, this scheme does not support for the higher throughput and it maximizes only the revenue of the primary users.

Xingqin Lin et al. proposed the framework for distributed wireless information flow allocation problem in multiple access network to minimize the power consumption while satisfying each end users minimum data rate requirement. They modeled the flow allocation problem as a game which is proved to be a best response potential game. The uniqueness of Nash equilibrium in the formulated potential game is actually having the globally optimal solution with better convergence rate for two proposed algorithms namely D-SBRA and P-SBRA. However, this proposed scheme does not take the SINR in to the account to improve the data rate of the end users. The game is shown to be the ordinal potential game (OPG) using the utility function proposed by Sastry et al. It poses the fewer constraints on the design of utility function using the exact potential game (EPG). A stochastic learning algorithm (SoNS) is proposed to perform the network selection independently at each SU based on its own action-reward history; without even knowing the knowledge of other SUs by Li-Chuan Tseng et al.

In this paper, we consider the problem of resource selection in cognitive radio network (CRN). Specifically, we consider the secondary network access scenario to access the licensed band which is served by the primary network namely secondary provider (SP). We designed the problem in such a way that there is a time varying demands of the number of available channels for the SUs is denoted as the external state. Therefore, the distributed algorithms such as Stochastic Learning based Resource Selection (SLRS) and Q-Learning based Resource Selection (QLRS) is proposed to address the resource management problems efficiently in cognitive radio network.

3. System Model

In this paper, we consider a cognitive radio network with \( M_{sp} \) and \( N_{su} \). The sets of SPs and SUs are denoted by \( \mathcal{M} \) and \( \mathcal{N} \), respectively. SPm holds \( C_m(t) \) unused channels at time instant t, after resource allocation for primary users (PUs) that can be used to serve the SUs. Figure 1 represents the typical cognitive radio network environment where the unused channels are allocated to the SUs.

![Figure 1. The typical cognitive radio network environment.](image-url)
The following are the assumptions to be made to construct the system model which reflects the cognitive radio network.

1) Each SU can prefer only one SP at a given time.
2) The data of the number of available channels owned by each SP are fixed but it is unknown to the SUs.
3) Each SU can select the SPs individually.
4) The number of SUs in the system, \( \mathbb{N} \) is known.

Distinctly, the information available for decision making is the action-reward history of the individual players called SUs.

The flow diagram in Figure 2 shows the sequence of actions that should be followed while allocating the resources to the SUs.

Let \( \mathbb{N}_m(t) = \{ j \in \mathbb{N} | a_j(t) = m \} \) be the set of SUs associated with SP \( m \) at time \( t \), where \( j \) is the SU, \( a_j(t) \) is the action (i.e., resource selection) of SU at time \( t \). Here, we define the two different algorithms for allocating the resources. In both the algorithms, the game model followed is potential game to analyze the Nash equilibrium. Now, we are considering the SUs are of the same priority so that the number of available channels is equally divided to them.

If the SU ‘\( j \)’ selects an action \( a_j(t) \) at time \( t \), then the throughput is calculated using (1).

\[
    r_j(t) = \frac{C_m(t) D_{m,j}}{n_m(t)}, \quad \forall j \in \mathbb{N}_m(t) \tag{1}
\]

Where \( r_j(t) \) is the throughput for the secondary user \( j \) at time \( t \), \( n_m(t) = |\mathbb{N}_m(t)| \), \( D_{m,j} \) is the data rate for the SU ‘\( j \)’ to be associated with the provider \( m \). The data rate may vary according to the modulation order in which the channel to be adopted at that time. We discarded the timing dependence variables to avoid the confusions and for the purpose of crispness in equations.

4. Stochastic Learning and Q-Learning based Resource Selections (SLRS & QLRS)

In this section, we present the game-theoretic formulation of the stochastic learning based resource selection and Q-learning based resource selection algorithm for resource allocation problem. The symbols used in the formulation are summarized in Table 1.

4.1 Game Model

We model the resource selection problem as a non-cooperative game where the SUs are the players, and the numbers of unused channels are considered as the external state. Then, the game is represented as follows [10]:

\[
    G = (C, \mathbb{N}, \{A_j\}, \{u_j\}), \quad \forall j \in \mathbb{N}
\]

Where \( C \) is the external state space, \( \mathbb{N} \) is the set of players, \( A_j \) is the set of actions that player \( j \) can take, and \( u_j \) is the

| Symbol | Meaning |
|--------|---------|
| \( \mathbb{N} \) | The set of SUs |
| \( \mathbb{M} \) | The set of SPs |
| \( C \) | Channel availability |
| \( C_m(t) \) | Number of available channels of SP \( m \) at time \( t \) |
| \( A_j \subseteq \mathbb{M} \) | The set of actions of player \( j \) |
| \( f_j \in A_j \) | An element of \( A_j \) |
| \( a_j(t) \in A_j \) | The action (selecting the idle channel) of player \( j \) at time \( t \) |
| \( a_j(t) \in A_j \) | The action of player except for \( j \) at time \( t \) |
| \( \mathbb{P}_j = (A_j) \) | The set of probability distribution over \( A_j \) |
| \( \mathbb{V}_j(t) \in \mathbb{P}_j \) | Mixed Strategy of player \( j \) at time \( t \) |
| \( \mathbb{R}_j(t) \in \mathbb{R} \) | Instantaneous reward of player \( j \) at time \( t \) |
utility function for the player \( j \) that depends on the action which is chosen by the players.

Each and every secondary user is selfish and they are individually tries to maximize their throughput. Thus, we define the utility function as the expected reward of the player \( j \) over the availability of the channels in the secondary provider is given by

\[
u_j(a_j, a_{-j}) = E_{C_{a_j}} \left( r_j \mid (a_j, a_{-j}) \right)
\]

\[
u_j(a_j, a_{-j}) = \frac{1}{n_{a_j}} \sum_{k=1}^{k_{a_j}} P_{a_j,k} C_{a_j,k} D_{a_j,k} - \sum_{k=1}^{k_{a_j}} n_{a_j}
\]

Where, \( n_{a_j} = P_j T_j \) Joules, \( P_j \) is the Power dissipation for switching, \( T_j \) is the Channel switching time, is the number of players taking the action \( a_j \), \( P_{a_j,k} \) is the probability of accessing the channels with \( \sum_{k=1}^{k_{a_j}} P_{a_j,k} = 1 \), \( C_{a_j,k} \) is the number of unused channels at that time and \( D_{a_j,k} \) is the data rate for that channel \( k \) corresponding to the action \( a_j \).

Formally, the game can be formulated as

\[
G: \max_{a_j \in A_j} \nu_j(a_j, a_{-j}), \forall j \in \mathbb{N}
\]

### 4.2 Analysis of Nash Equilibrium

With the utility function defined in equation (2), there is an existence of the NE point for the considered potential game. Then, the game \( G \) is an Ordinal Potential Game (OPG). By considering the OPG function as \( \Phi: x \in \mathbb{R}^n \rightarrow \mathbb{R} \) [10]. The average number of channels allocated to each of the SUs by the SPs, when \( y \) SUs are associated with the SP’s before the change. If the action is changed by the player \( j \), then it improves the utility function \( u_j \) using (2) and (4), we have the OPG to be represented as follows:

\[
u_j(a_j, a_{-j}) \succ \nu_j(a_j, a_{-j}) \Leftrightarrow \\
Z_{a_j}(n_{a_j} + 1) \cdot D_{a_j,j} \succ Z_{a_j}(n_{a_j}) \cdot D_{a_j,j}
\]

If there is a change in the player \( j \)’s action can merely affects the other player’s resource allocation by the SPs of that particular action by \( j \), then the change in the potential function \( \Phi \) caused by the player \( j \) is given by

\[
\Phi(a_j, a_{-j}) = Z_{a_j}(n_{a_j} + 1) \cdot D_{a_j,j} - Z_{a_j}(n_{a_j}) \cdot D_{a_j,j}
\]

Using (7) and (8) we find that the variations in the utility function \( u_j \) and the potential function \( \Phi \) due to the change in the player \( j \)’s action have the same sign since the player \( j \) is deviated unilaterally. Thus the function is represented as follows:

\[
u_j(a_j, a_{-j}) - \nu_j(a_j, a_{-j}) > 0 \\
\Leftrightarrow \Phi(a_j, a_{-j}) - \Phi(a_j, a_{-j}) > 0
\]

Therefore \( G \) is an OPG with the potential function \( \Phi \). The presence of a pure strategy NE is always guaranteed and it should concur with the local maximum of the potential function. Thus the analysis of NE shows that every potential function must have an NE point at one extent shows that there is a feasible allocation is possible.

### 4.3 Learning Procedure

For Algorithm 1, the NE is obtained by using the stochastic learning. From time to time there is a change in the channel availability and the actions selected by each SU may be different. The actions can be taken simultaneously and independently by each SU in the system may not be applicable so we propose the algorithm called Stochastic Learning based Resource Selection (SLRS), by which the SUs can learn the system towards the equilibrium strategy profile from their individual action-reward history. In SLRS, each SU updates the resource selection probability value is given by

\[
P_{j_{-i}}(t+1) = P_{j_{-i}j}(t) + b \cdot r_j(t)(e - P_{j_{-i}j}(t))
\]

where \( 0 < b < 1 \) is the learning rate, \( e \) is the unit probability function, and \( r_j(t) \) is the normalized reward.

The proposed SLRS algorithm is represented as follows:

**Algorithm 1 Stochastic Learning based Resource Selection (SLRS)**

**Input:** 1. Total Number of Available channels, \( \{C_{a}(j)\} \)
2. Set of Actions \( \{A\} \)

**Output:** Utility and Reward Function of the players \( j \in \{1, 2, \ldots, N\} \) is calculated.
Initialize the resource selection probability measure is
\[ p_{j,s}(t) = \frac{1}{|A_j|}, \forall j \in N, s_j \in A_j, \text{ at time } t = 0. \]
for each time, \( t \in \{0, 1, 2, \ldots, l\} \) do
begin
for each player, \( j \in \{1, 2, \ldots, N\} \) do
begin
Select an action \( a_j(t) \) as the outcome of the probabilistic measure based on \( p_j(t) \).
The instantaneous reward for SU is calculated from \( r_j(t) \) using (1).
if \( r_j(t) \geq r'_j(t) \)
Calculate the utility function using (2).
end
for each SU ‘j’ updates the resource selection probability measure using (8).
end
end.

The SLRS algorithm demonstrates that the Nash equilibrium occurs only in the mixed strategy since the strategy chosen by the SUs are different. Now the mixed strategy is defined as

\[ \mathcal{P}_j(t) = \{ p_{j,1}(t), \ldots, p_{j,M}(t) \} \text{ is the channel selection probability vector of player } j \text{ at time } t. \]

For Algorithm 2, the NE is obtained by using the Q-learning. The aim of the SU is to learn a policy (channel to be selected) for choosing the next action \( a_j(t) \) based on its current probability value \( p_j(t) \) that gives the maximum reward. A reward function is obtained using (1) in which the maximum reward is achieved by taking the action according to the greedy exploration strategy. At each time step each secondary user uses the Q learning scheme to update its probability value based on the strategy selected by them. The Probability value is calculated as follows:

\[ p_{j,s}(t+1) = p_{j,s}(t) + b \cdot (r_j(t+1) - \gamma \cdot \max_{s_j} p_{j,s_j}(t+1) - p_{j,s_j}(t))) \]

where ‘b’ is the learning rate, ‘\( \gamma \)’ is the discount factor, in which both lies between 0 and 1, \( r_j(t+1) \) is the normalized reward function. The optimal future value is predicted using \( \max_{s} p_{j,s}(t+1) \). These all variables constitute the learned value.

More generally, the Q learning does not require any adaptations to handle the problems with the random transitions and rewards. The optimal policy is followed by learning the action-value function that gives the expected utility of the SUs. When such an action-value function is learned, the optimal policy can be constructed by simply selecting the action \( a_j \) with the highest value in every time period.

The main advantage of Q-learning is able to compare the expected utility of the available actions without requiring a model of the environment. The QLRS algorithm is designed as follows:

**Algorithm 2 Q-Learning based Resource Selection (QLRS)**

**Input:**
1. Total Number of Available channels, \( |C_m(j)| \)
2. Set of Actions \( \{A_j\} \)

**Output:** Utility and Reward Function of the players \( j \in \{1, 2, \ldots, N\} \) is calculated.

Initialize the resource selection probability measure is
\[ p_{j,s}(t) = \frac{1}{|A_j|}, \forall j \in N, s_j \in A_j, \text{ at time } t = 0. \]
for each time, \( t \in \{0, 1, 2, \ldots, l\} \) do
begin
for each player, \( j \in \{1, 2, \ldots, N\} \) do
begin
Select an action \( a_j(t) \) as the outcome of the probabilistic measure based on \( p_j(t) \).
The instantaneous reward for SU is calculated from \( r_j(t) \) using (1).
if \( r_j(t) \geq r'_j(t) \)
Calculate the utility function using (2).
end
for each SU ‘j’ updates the resource selection probability measure using (9).
end
end.

If the learning rate ‘b’ is 0, then there is no learning about anything, while the value grows towards 1 would make the decision consider only most recent information.

The instantaneous reward (throughput) serves as a reinforcement signal so that a high reward brings a high probability in the next strategy updates in both the algorithms. The selection of resources is considered to be an action which is based on the probabilistic measure might result in handover between different networks in the beginning of the learning procedure.
Nevertheless, a stable long-term resource selection strategy will be yielded after the learning period using the theorem 3.1 in\(^9\), and the learning rate converges to the NE in a period of time.

If the learning rate \( b \) is 0, then the convergence of NE is very slower. The learning rate value is high that is near to 1 means it takes more time to attain the NE. In both the cases, the convergence of NE is guaranteed. An appropriate value of \( b \) is taken for the practical implementation in the network has a tradeoff between the accuracy and convergence rate of NE.

### 5. Numerical Results

In order to evaluate the performance of the proposed scheme in two different scenarios, we conduct a series of simulations. The simulation is carried out using LTE simulator which gives the complete environment of CRN. The simulation environment consists of SPs who owns the channels and SUs. The modulation orders like 16 QAM are adopted by the SUs depends on the Channel Quality Information (CQI) values under different RSS conditions. The Total Transmission Interval (TTI) for selecting the resources by the SUs takes 10 TTIs. The remaining parameters in the CRN are already inherited from the LTE simulator in-built functions.

The average SU throughput graph is generated for proposed scenarios with respect to the defined function \( F(x) \) shown in Figure 3. The following are the assumptions to be made to carry out the results. Let the learning rate \( b = 0.5 \), switching overhead \( \eta \) would takes the values by combining the power dissipation for switching and time needed for switching the channel is implemented in LTE simulator and the values are observed. The CQI values ranges between 0 and 15 in which each value adopts different QAMs. It achieves the maximum throughput in both the scenarios by unilateral deviation, and thus the resulting strategy is a mixed strategy of NE point. Each SU takes different strategies in the DRS and MQ algorithms.

The function \( F(x) \) in Figure 3 denotes the throughput ECDF (Empirical Cumulative Distribution Function) for the average SU throughput. The resource selection probability of unilateral deviation results in the minimum throughput for all players, insisting that the NE point is obtained by the learning algorithm of two different algorithms considers the learning rate to be 0.5 in Figure 4.

The comparison between two proposed algorithms namely SLRS and QLRS is shown in the Table II.

The performance of the two different algorithms to address the resource selection problem is evaluated by the fairness, Average and peak throughput, mean RB occupancy for the SUs. The fairness is measured by the Jain’s fairness index (JFI),

\[
J = \frac{\left( \sum_{j=1}^{N} x_j \right)^2}{N \sum_{j=1}^{N} x_j^2}
\]

we observe

![Figure 3. Reward for the SUs using SLRS and QLRS based on NE.](image)

![Figure 4. SUs probabilities based on different actions with respect to iterations for SLRS and QLRS.](image)

| S.NO | Parameters                  | Algorithm |
|------|-----------------------------|-----------|
| 1    | Fairness                    | 0.67      | SLRS 0.71 |
| 2    | Average UE Throughput(bit/Cu)| 1.36      | QLRS 1.52 |
| 3    | Peak UE Throughput(Mb/s)    | 3.05      | 3.26     |
| 4    | Mean RB Occupancy           | 78.60     | 79.56    |

Table 2. Achievable throughput & Fairness
that the efficiency of the learned NE strategy is above 75% for both algorithms. The QLRS scheme achieves higher throughput and fairness when compared to the SLRS scheme. Thus, the results show the advantage of the proposed method in which the throughput, fairness can be maintained through different learning approaches.

6. Conclusion and Future

The problem of resource selection in cognitive radio networks with dynamic channel availability with respect to time was formulated and analyzed. We designed the resource selection problem as an ordinal potential game in which the Nash equilibrium point is attained for mixed strategy. A distributed, Stochastic Learning based Resource Selection and Q-Learning based Resource Selection algorithm namely SLRS and QLRS was proposed. Simulation results have demonstrated that it achieves higher throughput for QLRS algorithm when compared to the SLRS algorithm. There is a need to incorporate suitable advanced learning that improves the dynamic resource selection in CRN, which could be the future work.

7. References

1. Gavrilovska L, Atanasovski V, Macaluso I, Luiz A, Dasilva LA. Learning and Reasoning in Cognitive Radio Networks. IEEE Communication Surveys and Tutorials. 2013; 15.(4).
2. Neel J, Reed J, Gilles R. Convergence of Cognitive Radio Networks. IEEE (WCNC). 2004: 4:2250–5.
3. Nie N, Comaniciu C. Adaptive Channel Allocation Spectrum etiquette for Cognitive Radio Networks. IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN); 2005. p. 269–78.
4. Khan M, Tembine H, Vasilakos A. Game Dynamics and Cost of Learning in Heterogeneous 4G Networks. IEEE J Sel Area Comm. 2012 Jan ; 30(1):198–213.
5. Bennis M, Perlaza SM, Blasco P, Han Z, Poor V. Self Organization in Small Cell Networks: A Reinforcement Learning Approach. IEEE Trans Wireless Comm. Jul 2013; 12(7).
6. Ghosh A, Hamouda W. Channel Selection for Heterogeneous Nodes in Cognitive Networks. IEEE Wireless Networking Symposium; 2013.
7. Niyato D, Hossain E. Competitive Spectrum Sharing in Cognitive Radio Networks: A Dynamic Game Approach. IEEE Trans Wireless Comm. 2008 Jul; 7(7).
8. Lin X, MingLok T. Distributed Wireless Information flow allocation in multiple access Networks. EURASIP JWCN. 2011. doi:10.1186/1687-1499-2011-14.
9. Sastry P, Phansalkar V, Thathacha M. Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information. IEEE Trans Syst Man Cybern, 1994 May; 24(5): 769–77.
10. Tseng L, Chien F, Zhang D, Chang RY, Chung W, Huang CY. Network Selection in Cognitive Heterogeneous Networks Using Stochastic Learning. IEEE Communication Letters. 2013 Dec; 17(12).
11. Monderer D, Shapley L. Potential Games. Games and Economic Behavior. 1996; 14:124–43.
12. Available from: www.nt.tuwien.ac.at/ltessimulator.
13. Xu Y, Wu Q, Shen L, Wang J, Alagan Anpalagan A. Opportunistic Spectrum Access with Spatial Reuse: Graphical Game and Uncoupled Learning Solutions. IEEE Trans Wireless Comm. 2013 Oct; 12(10).
14. Jain RK, Chiu DMW, Hawe WR. A Quantitative Measure of Fairness & Discrimination for Resource Allocation in Shared Computer System. ACM Trans Comput Syst), 1984 Sep.