

Applying Machine Learning Techniques For Spectrum Trading In Future LTE Based Cognitive Radio Systems

Hiran Kumar Singh

*Department of Information Technology
Anna University, MIT Campus
Chennai-600044, TamilNadu - INDIA
hksbxr@gmail.com*

Dhananjay Kumar

*Department of Information Technology
Anna University, MIT Campus
Chennai-600044, TamilNadu - INDIA
dhananjay@annauniv.edu*

Abstract

The spectrum trading has been advocated as an efficient method of spectrum sharing in future cognitive radio system. In real-time the spectrum trading becomes challenging because of complex nature of multi-operator environment of cognitive radio system. This necessitates the adoption of a suitable machine learning based game theory. In this paper, we explore the correlated equilibrium and reinforcement technique for the spectrum trading with multiple primary carrier providers (PCP) trading and hence leasing the spectrum to multiple secondary carrier providers (SCP) for a short period. We formulate optimum correlated equilibrium, Q-learning, actor critic learning, and conjecture based Q-learning in a LTE based cognitive radio system (CRS). The simulation results suggest a cumulative increase in throughput for the conjecture based Q-learning.

Keywords: Cognitive Radio, Spectrum trading, Resource utilization, Game theory, Learning.

Introduction

The broadband services and many other services through internet in wireless demands large bandwidth and hence the resource management in Long Term Evolution (LTE) and LTE-Advanced (LTE-A) need to be managed efficiently that not only to optimally share spectrum but also network infrastructure. The ITU-R report [1] has recognized the important features related to the usage of cognitive radio system (CRS), and these

systems adopt a method which approves to acquire the knowledge of its environment and adaptively adjust its system parameters and learn from the results. The CRS follows a cognitive cycle (Fig.1) to detect, position, plan, learn, resolve, and act [2].

A suitable machine learning technique in CRS need to be developed to learn and analyze various traffic patterns on various wireless channels over time and then estimate the idle channels [3-4]. In a multi-operator environment there could a situation when a carrier provider is in need of extra channels in a given cell/sector. On the other hand other carrier provider may have excess resources in same cell/sector and time. A smart sharing mechanism based on mutually agreed policy will help both operators to not only manage their resource efficiently but also meet the customer's demand for the quality of service (QoS). Further the sharing of spectrum and infrastructure can be devised based on revenue model which brings incentives to the concerned parties. This situation could be better understood, defined, formulated, and planned by adoption of economic model based on game theory. The game theory in CRS will provide a suitable mechanism to ensure coexistence of different carrier providers while optimally sharing the resources.

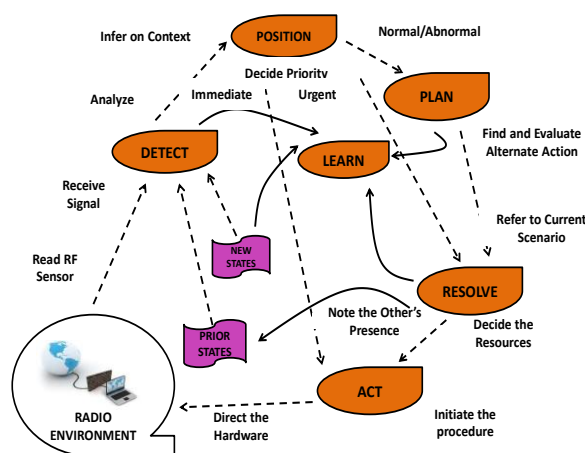


Figure 1: A Simple Cognitive Cycle

The dynamic spectrum access (DSA) by the SCPs is modeled in three groups – shared use, commons, and the exclusive use model. In the shared use, the SCPs can make use of the spectrum owned by PCPs without any price when not used by them. In the commons model, the spectrum is open for everyone to access (e.g. ISM bands). These two methods have some specific drawbacks. In the exclusive use model, the PCPs lease their vacant spectrum to SCPs and gain some revenue while the SCPs could have the assured access to the spectrum for a shorter or longer period of time as per the agreement made. The ITU report [3-4] explains the different prices and various techniques involved in sharing and evaluating the spectrum.

In a multi operator radio access network (Fig.2) [5] apart from spectrum sharing, RAN can also be rented which leads to formulation of a combined trading policy

which integrates cost of RAN into the price of spectrum. The main benefits of combining RAN along with the spectrum trading policy is that a carrier provider can easily facilitate the traded spectrum subjected to the agreement in trading policy. Further the billing and charging for the customer could be still carried out by the home carrier provider based on its own business policy. This may require implementation of additional signaling mechanism between two service provider's networks.

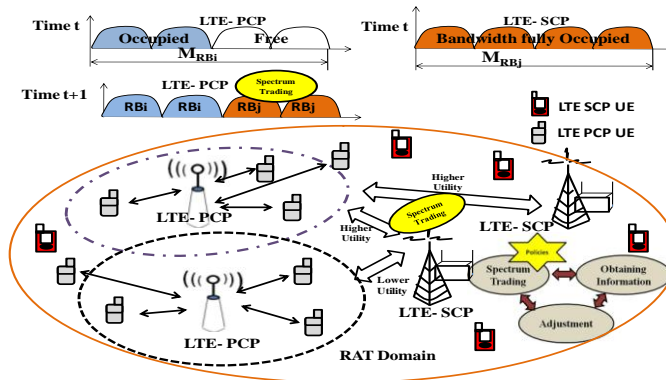


Figure 2: A Scenario of spectrum sharing between LTE-PCP and LTE-SCP [6]

The financial incentives to the parties involved becomes one of the major requirement in the spectrum sharing and hence the market driven spectrum trading is a desirable solution [7,8]. In multi-operator network the implementation challenges increases by many fold as the market based strategy needs to consider not only DSA but also the overall cost involved in operation and maintenance of the networks. Moreover, fairness issue cannot be neglected in any multiple access systems.

The research presented in this paper is extension of our previous work [6], which is focused on to devise an adaptive sharing mechanism of resource block (RB) for a future LTE based network (e.g., LTE-A) while considering major issue in trading, fairness, and implementation. In this paper the aim of our proposed algorithm to incorporate utility based learning method that maximizes the overall throughput of the networks involved. In earlier paper [6] the system was formulated on agreement model based on the utility maximization in trading the resources for short-term basis. The trading method included role reversal i.e., any carrier service provider can act as a buyer (seller) when the resources at a particular time period are deficient (surplus). Unlike our previous work here the main issue to incorporate the advanced machine learning technique in utility calculation of PCP/SCP.

In this paper we prefer optimal correlated equilibrium than Nash Equilibrium not just because of its generality but it permits each player to choose action according to the observation of environment. The reinforcement learning could be a basis of learning & interaction among the carrier provider and the prevailing surroundings is used by the PCP/SCP to optimize their utility without considering the models of the environmental dynamics. The three technique under reinforcement learning namely: Q-learning, Actor critic learning, and Conjecture based Q-learning is formulated and

implemented here. These three learning techniques are implemented and simulated on LTE simulator, and throughput performance is analyzed to find best method in multi-operator scenario.

The rest of the paper is organized as follows: Section 2 describes a literature survey of related works on spectrum trading in CR networks. The system model and optimization formulation are described in Section 3. The problem formulation for our three approaches along with the heuristic algorithm is presented in Section 4. Section 5 presents the simulation results under LTE simulation environment. The conclusion and future work are listed in Section 6.

Related Work

The challenges of applying game theory and machine learning for spectrum trading in cognitive radio system has topic of active research in recent year. In [8] authors discussed about the spectrum trading in order to maximize the revenue of primary users and also maximizing the satisfaction of secondary users. This work addressed the issue of pricing in a dynamic spectrum access environment. The problem of pricing negotiation between the operator and the service users were formulated as a multiunit sealed-bid auction. In an auction mechanism an optimization problem was formulated to maximize the revenue of the spectrum owner through pricing and spectrum assignment.

The game theory [9] to sense the unused spectrum (available channels) and help in assigning appropriate channel to the cognitive users. In this work the incentives of all players in the game can be expressed in one global function.

Anon-cooperative game [10] that achieves a steady state with desirable features. A non-cooperative game in a slotted ALOHA setting is designed, and the existence and uniqueness of the Nash Equilibrium (NE) solution is analyzed.

To achieve the [11] Nash Bargaining Solution (NBS) in the allocation of power, and evaluating it with respect to efficiency and fairness. It is shown that in a high interference environment with a finite number of channels, the utility space of the spectrum sharing game is non-convex. Non-convexity can lead to optimal operating points that require mixed strategies.

The problem of spectrum sharing among a primary user and multiple secondary users and centralized and distributed decision making scenarios are considered [12]. In the former case, each secondary user is assumed to be able to observe the strategies adopted by other users. In the latter case, the adaptation for spectrum sharing is performed in a distributed fashion based on communication between each of the secondary users and the primary user only. In [13] authors considered a limited cognitivity. The learning process of the parameters is performed by a set of rules that update the radio settings based on a random process named better reply dynamics (BRD).

A two-tier market for decentralized dynamic spectrum access is proposed [14]. In the proposed Tier-1 market, spectrum is traded from a primary user (PU) to secondary users (SUs) in a relatively large time scale to reduce signaling overhead. Observing the limitations of the auction market, a two-tier market structure based on the

decentralized bargain theory to enable distributed DSA is proposed. In [15] discussed the sensing of primary users in the licensed frequency bands. In this paper reinforcement learning algorithm is proposed. By using this the secondary user learns to find the optimal set of cooperating neighbours with minimal traffic and select independent users for cooperation under correlated shadowing. Cooperative sensing process is formulated as finite-horizon Markov decision process. The concept of harmonized Q-learning for the radio resource management in LTE based networks that manage its resource pool dynamically is introduced [16]. The multi operator system is modeled on the game theory based Q-Learning.

System Model & Optimization Framework

System Model

We consider a multi-operator LTE based cognitive radio systems (LTE-CRS) where channel utilization and hence demand varies as per the subscriber's location and need. At any given time an operator (SCP / PCP) estimate its excess resource requirement and explores the trading and acquiring mechanism for additional resource with other operator (PCP / SCP). The total available bandwidth (W) can be assumed to be distributed into (M)resource blocks (RBs) either contiguously or non-contiguously. Furthermore, the available free RBs that can be traded by the PCP at time t are assumed to be randomly distributed in time-resource grid. The role played by the LTE carrier provider with CRS capability could be interchangeable; call it as primary (LTE-PCP) or secondary (LTE-SCP). The LTE-SCP initiates the process of the resources acquisition from LTE-PCP for a short contract period τ . These resources fall back to PCP after a contract period or by the same SCP by renewing the contract. The major requirement for contract under trading process is the satisfying minimum utility of both primary and secondary carrier providers. Further, the contract may include radio access network (RAN) of PCP to facilitate the traded spectrum to the SCP and the cost of maintaining RAN for SCP is treated as an additional resource offered by the PCP. The optimization framework in framing the cost/utility expression need to be formulated based on trading mechanism, which need to consider some basic gaming principle.

Development of Heuristic Algorithm

There is a need to develop heuristic algorithms employing suitable machine learning technique in trading policy. The learning mechanism could be based correlated equilibrium where some optimization can be carried out or it could Q-learning method with some variation. Under Q-learning two commonly used algorithm namely Actor critic learning and Conjecture based Q-learning can adopted for the trading process. It is also assumed that both SCP and PCP are willing to participate and role change can happen i.e., the SCP may become PCP and vice versa in a given circumstances.

Correlated Equilibrium

In game theory, a correlated equilibrium is a solution concept that is more general than the well-known Nash equilibrium. The idea is that each player chooses his/her action according to his/her observation of the value of the same public signal. A strategy assigns an action to every possible observation a player can make. If no player would want to deviate from the recommended strategy (assuming the others don't deviate), the distribution is called a correlated equilibrium.

Equivalently, a correlated equilibrium is a probability distribution on N tuples of actions, which can be interpreted as the distribution-of-play instructions given to the players by some “device” or “referee.” Every player is given privately instructions for his play only; the joint distribution is known to all of them. Also, for every possible instruction that a player receives, the player realizes that the instruction provides a best response to the random estimated play of the other players assuming they all follow their instructions. The correlated equilibrium has several important advantages: It is a perfectly reasonable, simple, and plausible concept; it is guaranteed to always exist.

Optimum Correlated Equilibrium

The optimal correlated equilibrium can be formulated as:

$$\begin{aligned} & \max \sum_{i \in N} E_p(U_i) \\ \text{S.T.} & \left\{ \begin{aligned} & \sum_{s_{-i} \in \dot{U}_{-i}} p(s_i, s_{-i}) [U_i(s'_i, s_{-i}) - U_i(s_i, s_{-i})] \leq 0 \\ & \forall i \in N, \forall s_i, s'_i \in \dot{U}_i \end{aligned} \right. \end{aligned} \quad (1)$$

Where $E_p(\cdot)$ is the expectation over p . The constraint guarantee that the solution is within the correlated equilibrium set. The utility function used here is same as the one used above.

Suppose that the proposed game G is played repeatedly through time: $n = 1, 2, \dots$

1. Initialization:

At the initial time $n = 1$, each player initializes his/her strategy arbitrarily.

2. Iterative Update Process

- **Utility Update:** At the time n , each player i calculates the utility of the current strategy $S_i \in \Omega_i$ and the utility for choosing the different strategy $S'_i \in \Omega_i$
- **Average Regret Update:** If player i replaces strategy S_i , every time that it was played in the past, by the different strategy S'_i , the resulting difference in i 's average utility up to time n is

$$D_i^n(S_i, S'_i) = \frac{1}{n} \sum_{\Omega \leq n} [U_i(S'_i, S_{-i}^{(\hat{\theta})}) - U_i(S_i, S_{-i}^{(\hat{\theta})})] \quad (2)$$

$$R_i^n(S_i, S'_i) = \max \{ D_i^n(S_i, S'_i), 0 \} \quad (3)$$

where $R_i^n(S_i, S'_i)$ represents the average regret at time n for not having played, every time that S_i was played in the past, the different strategy S'_i .

Strategy Decision: Assuming $S_i \in \Omega_i$ is the strategy last chosen by player i , $S_i^n = S_i$. Then at time $n+1$, i updates his/her decision strategy according to the probability distribution:

$$\begin{cases} p_i^{n+1}(S'_i) = \frac{1}{i} R_i^n(S_i, S'_i), \forall S'_i \neq S_i \\ p_i^{n+1}(S_i) = 1 - \sum_{S'_i \neq S_i} p_i^{n+1}(S'_i) \end{cases} \quad (4)$$

In the proposed algorithm, each player does not need to be concerned about the individual strategies and utilities of other players, global system structure, etc. Each one just needs to know the effect of other players on its individual utility function. In addition, each player views its current actual strategy as a reference point, and makes a decision for next period according to propensities to depart from it. However, the change should bring the improvement in individual utility, relative to the current choice.

Reinforcement Learning

Learning schemes can be employed in cognitive radio systems to intelligently locate the spectrum holes with some knowledge about the operating environment. The activities of the Primary Users follow the Markovian process model. The opportunistic spectrum access environment follows a distinct feature which makes it too hard to fabricate models which predict the dynamics. So it. The reinforcement learning which is a basis of learning & interaction among the user and the environment is used by the secondary users to optimize their behaviour without considering the models of the environmental dynamics. And the comparison among Q-learning, actor critic learning and conjecture based Q-learning is done [18]. Consider the system model with p primary users and s secondary users comprising total of M users. Each user possesses N number of resource blocks. The objective of reinforcement learning is to ensure the achievement of minimum utility by every user.

The capacity of user k (C_k^t) at time t is given by

$$C_k^t = \sum_{i=1}^N w \cdot \log_2 \left[1 + \frac{p_k^t(i) \cdot s_k^t(i)}{I_k^t(i)} \right] \quad (5)$$

The received interference by resource block i at user k (I_k^t) is given by

$$I_k^t(i) = \sigma_k^2(i) + \sum_{k=1}^M W_{i,k}^t(i) \cdot p_i^t(i) \quad (6)$$

Where w is the bandwidth of the channel i .

$p_k^t(i)$ is the action chosen by user k

$s_k^t(i)$ is the channel state vector of user k

$\sigma_k^2(i)$ is noise covariance.

$W_{i,k}^t(i)$ is the interference received by channel i due to channel k

The global objective is chosen to maximize the performance of the worst off user such that the available resources are fairly allocated among all the users. Therefore,

given the action vector of all the users, the global utility function at time t is defined as follows.

$$U_G(\mathbf{P}^t) = \min_{k \in K} [\min(\frac{C_k^t}{d_k^t}, 1)] \text{-----} \quad (7)$$

Where d_k^t is the demand level of user k .

Here, the term $\min(\frac{C_k^t}{d_k^t}, 1)$ represents the satisfaction level of the user k . At each time t , the capacity of the user k C_k^t should not exceed its demand level d_k^t as it leaves less resource to other users.

$$U_G(\mathbf{p}^t) = -\frac{1}{d_k^t} (C_k^t - d_k^t) \text{-----} \quad (8)$$

The overall utility can be represented as

$$U_G(\mathbf{p}^t) = \min_{k \in K} [\min(\frac{C_k^t}{d_k^t}, 1)] + \frac{1}{d_k^t} (C_k^t - d_k^t) \text{-----} \quad (9)$$

Q-Learning

At each time step, each secondary user uses the Q-learning scheme to select a spectrum band among all the bands. The aim of the secondary user is to learn a policy (band to be switched over) $\pi: S \rightarrow A$ for choosing the next action a_i based on its current state s_i that gives the maximum reward. A function $Q: S \times A \rightarrow R$ is defined for each state-action (s_i, a_i) pair as the maximum reward that can be achieved when taking action a_i from state s_i according to the exploration strategy. Hence, with the help of the Q-function, the SUs can optimally select actions that maximize Q_{s_i, a_i} , at each state. The Q-learning algorithm works by selecting actions and observing the following state and the resulting reward. With this information, Q value is updated using the following equation.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha [R_{t+1} + \tilde{\alpha}(\max_a Q(s_{t+1}, a)) - Q(s_t, a_t)] \text{-----} \quad (10)$$

where, $0 < \alpha < 1$ is the learning rate of the secondary user group and $\tilde{\alpha}$ is the discount factor.

The reward value is calculated using the iterative process defined as follows. The predefined incremental value is added to the reward function for every time particular resource block is found to be free.

Q-Learning algorithm is given below.

1. Initialize the Q-table with state and action pair.
2. for each secondary user s_i
3. for time period $t=2$ to 10
4. for each free RB $i=1:10$
5. Select an action a_i
6. Calculate the reward perceived during action a_i
7. Update Q-table using (10)
8. end

9. end
10. end
11. Allocated resources using Q-table.
12. Calculate the utility using (9)

Actor critic learning

A function $V: S \rightarrow R$ is defined for each state–action (s_i, a_i) pair as the maximum reward that can be achieved when taking action a_i from state s_i according to the exploration strategy and a Preference table (P) which stores the preference of state dependent actions are defined for each SU. Actor critic learning scheme learns by selecting actions and observing the following state and the **resulting reward**. **With this information, V is updated via** the following equation,

$$v_{s_i}(t + 1) = v_s(t) + \alpha [r_{s_i, a_i} - v_{s_i}(t)] \quad \text{-----} \quad (11)$$

where, $0 < \alpha < 1$ is the learning rate of the secondary user group g and r_{s_i, a_i} is the associated reward function.

Actor critic learning uses state values to update the preference table.

$$P_{s_i, a_i}(t+1) = v_{s_i}(t) + \alpha [r_{s_i, a_i} - v_{s_i}(t)] + P_{s_i, a_i}(t) \quad \text{----} \quad (12)$$

Usage of state values speeds up the learning process. With the help of the P-table, the users can optimally select actions that have highest preference at each state. The following provides a algorithm of Actor-Critic Learning .

1. Initialize the V-table, P-table and probability vector
 2. for each secondary user s_i
 3. for time period $t=2$ to 10
 4. for each free RB $i=1:10$
 5. Select an action a_i
 6. Calculate the reward perceived
 7. Update V-table using (11)
 8. Update P-table using (12)
 9. end
 10. end
 11. end
 12. Allocated resources using P-table.
- Calculate utility using(9)

Conjecture based Q learning

Conjecture based Q-learning aims to design a simple noncooperative power allocation scheme that requires quite limited information exchanges among the users. The reached NE is based on the assumptions about what knowledge the SUs possess and assumes that every SU’s strategy will not change at the NE.

$$C_i^{t}(s, a_{-i}) = \bar{C}_i(s_i, a_{-i}) - w_i^{s_i, a_{-i}} [\pi_i^t(s_i, a_i) - \pi_i^{t-1}(s_i, a_i)] \quad \text{----} \quad (13)$$

Where $\pi_i^{t-1}(s_i, a_i)$ is the reference point for specific probability.

$\bar{C}_i(s_i, a_{-i})$ is the reference point for specific conjecture.

$w_i^{s_i, a_{-i}}$ is the relative weightage given.

Every user has beliefs concerning the way in which other SUs react are a dynamic version of conjecture.

The multiagent Q-learning updating is modified as

$$Q_i^{t+1}(s_i, a_i) = (1 - \alpha) \cdot Q_i^t(s_i, a_i) + \alpha \{ C_i^{t+1}(s_i, a_{-i}) \cdot R_i(s_i, a_i, a_{-i}) + \beta \max_{a_i} Q_i^t(s_i, a_i) \} \quad (14)$$

where α is the learning rate and β is the discount factor.

The SU updates its Q-values only with its own information during the stochastic learning process. To avoid observing the other SUs' private strategy information, the SU i conjectures about how its competitors' strategy decisions vary in response to its own actions.

Conjecture based learning algorithm is given below

1. for each strategy $S_i \in S, a_i \in A_i$
2. initialize transmission strategy
 $\pi_i^t(s_i, a_i), Q_i^t(s_i, a_i), c_i^t(s_i, a_i)$ and w_i
3. end
4. for time period $t=2$ to 10
5. choose action a_i according to $\pi_i^t(s_i)$
6. measure the received SINR \tilde{a}_i
7. if $\gamma_i > \gamma_i^*$
8. Calculate reward.
9. end
10. Update $Q_i^{(t+1)}(s_i, a_i)$ value based on $c_i^t(s_i, a_i)$ using (14).
11. Update strategy $\pi_i^t(s_i, a_i)$.
12. Update conjecture using (13)
13. end
14. Calculate utility using (9)

Results and Discussions

Simulation Setup

The simulation set up for the cognitive radio was carried out in cellular wireless environment with help of LTE simulator in MATLAB. The LTE simulator permitted us to create sector cell in which link access uses adaptive modulation and coding. The maximum bandwidth available is 20 MHz with the bandwidth of each sector being 180 KHz. We considered total of 19 cells with 3 sectors each. Each sector consists of 10 users. As per the facility available in simulator a tri sector tilted antenna model was used in physical layer.

Table 5.1: Main Simulation Parameters

Parameter	Values
Frequency band	2.14 GHz
Bandwidth	20 MHz
No: of RBs	100
Subcarriers/ RBs	12
Subcarrier spacing	15 KHz
Modulation & Coding levels	QPSK,16-QAM,64-QAM
Number of eNodeB Sector	57
UE per Sector	10
Learning Rate	0.8

Throughput Analysis

Each secondary user involved in the transaction calculates the utility of the current strategy and at the next transmission time interval replaces the previous strategy with other one and with no regret jumps to new strategy. The utility function is designed in such a way that effect of other players will be reflected upon any particular player. After establishment of correlated equilibrium the resource allocation of primary and secondary users is observed as cumulative distribution of throughput. The observed throughput is higher in case of adoption of correlated equilibrium than conjecture based Q-learning.

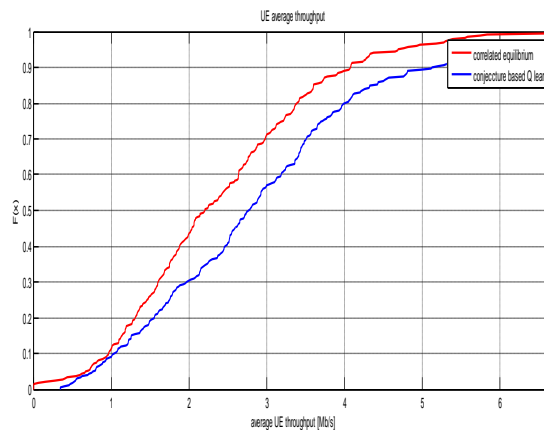


Figure 5.1: Throughput comparison between the system involving establishment of correlated equilibrium and conjecture based Q-learning.

The comparison between the establishment of correlated equilibrium and conjecture based equilibrium, in which every player makes assumption about other players' action is given in Fig 5.1

Comparing Q-learning, conjecture based Q-learning and actor critic learning

Here Q learning as a model free reinforcement technique learns the values of state/action pairs during simulation. Suppose $Q(s,a)$ is the expected discount return obtained by performing action a in state s which performs optimally thereafter. During run-time an optimal action in a state in any action try to maximize $Q(s,a)$.

In implementation of the conjecture based Q learning, the secondary user SU conjectures about how its competitors strategy decision varies in response with its own actions. On the other hand in actor critic learning actor a stochastic policy that maps states to action probability vectors and critic attempts to estimate the value of each state was implemented. The comparison of throughput among the Q-learning, conjecture based Q-learning and actor critic learning is shown in Fig 5.2.

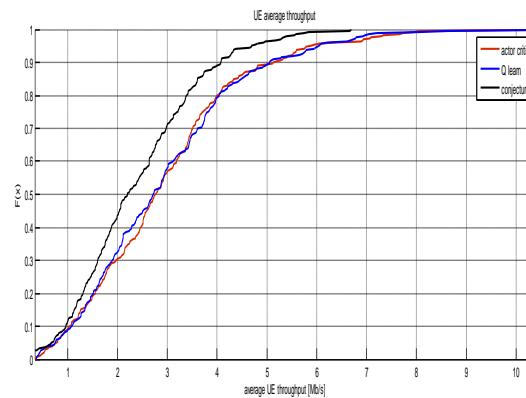


Figure 5.2: Comparison among three learning algorithms in terms of throughput

As shown in Fig.5.2 conjecture based Q learning yields more throughput than any other learning algorithms because it does take into account the assumptions made by other users too. Actor critic learning scheme and Q learning yields figuratively same throughput.

Conclusion

Efficient resource utilization can be achieved by effective use of available resources through new variation of learning called conjecture based Q-learning and overall efficiency and throughput of the system comprising multiple spectrum users tend to improve by a efficient spectrum trading process. In Conjecture based Q-learning, which is a variation of Q-learning, every user conjectures about how competitors' strategy decisions vary in response to its own actions. This added feature enables the user equipment to effectively identify the best available resource blocks for use. The simulation results showed that this learning technique provides the higher throughput than Q learning and actor critic learning. Resource sharing among the various users in spectrum trading environment is achieved by the proposition of three different algorithms. The correlated equilibrium is achieved in the environment where all the users are non-cooperative and it maximized the overall utility of the system.

As a future work the spectrum trading could be implemented in completely dynamic and non-cooperative environment in two stages. In first stage trading could occur between primary and secondary users, and in second stage market equilibrium could be achieved among all the secondary users by taking signal to interference noise ratio and power constraints into account thereby maximizing throughput in terms of both revenue and energy.

Acknowledgement

We thank the Institute of Telecommunication, Vienna University of Technology, Vienna for providing their LTE simulator in free download version.

References

- [1] ITU-R SM.2152 Y.2009, "Definitions of Software Defined Radio (SDR) and Cognitive Radio System (CRS)", Tech.Rep., Year 2009.
- [2] J Mitola, "Cognitive radio: An integrated agent architecture for software defined radio", Ph.D. Thesis, Royal Institute of Technology (KTH) year 2000.
- [3] ITU-R 241-2/5, "Cognitive Radio System (CRS) applications in the land mobile service", Annex 26 Doc.5A/306-E, Working Party 5A Chairman's Report, pp. 1-56, 3 June 2012.
- [4] ITU- R 241-2/5, "Cognitive Radio System applications in the land mobile service", Annex 22, Doc. 5A/198-E, Working Party 5A Chairman's Report, pp.1-55, 20 November 2012.
- [5] Network sharing MORAN and MOCN for 3G, Nokia Siemens Networks, May 2013.
- [6] HiranKumarSingh, Dhananjay Kumar andSrilakshmi R, "short Term spectrum Trading in Future LTE BasedcognitiveRadio Systems", *KSII Transactions on Internet and Information Systems Vol. 9, No. 1, Jan, 2015,pp. 34-49.*
- [7] Liang Qian, Feng Ye, Lin Gao, XiaoyingGan, Tian Chu, XiaohuaTian, Xinbing Wang and Mohsen Guizani, "Spectrum Trading in Cognitive Radio Networks: An Agent- Based Model under Demand Uncertainty", *IEEE Transactions on Communications*, Vol. 59,No. 11, pp. 3192-3203, November 2011.
- [8] DusitNiyato and EkramHossain (2009) "*Multiple-Seller and Multiple-Buyer Spectrum Trading using Game-Theoretic ModellingApproach*", *IEEETransactions on Mobile Computing*, Vol. 8, No. 8.pp. 1009-1022, August 2009.
- [9] Yenumula, B.(2011) "Potential Game Models For Efficient Resource Allocation in Wireless Networks", *Science Academy Transactions on Computer and Communication Networks* vol. 1, No.4.
- [10] YunusSarikaya and TansuAlpcan(2012) "*Dynamic Pricing and Queue*

- Stability in Wireless Random Access Games*”, IEEE Journal of Selected Topics in Signal Processing, vol. 6, No. 2.
- [11] Juan E. Suris, Luiz A. DaSilva, Zhu Han, Allen B. MacKenzie, and Ramakant S. Komali “Asymptotic Optimality for Distributed Spectrum Sharing using Bargaining Solutions” *IEEE Transactions on Wireless Communications* 8(10):5225-5237 (2009)
 - [12] Dusit Niyato and Ekram Hossain (2008) “Competitive Spectrum Sharing in Cognitive Radio Networks: A Dynamic Game Approach”, IEEE Transactions on Wireless Communication vol. 7, No. 7.
 - [13] Cattoni, A. F. & Friederichs, Karl-Joseph. “On the possible usage as LTE-Advanced as a building block for PHY/MAC in Cognitive Radio platforms” Joint workshop with COST2100, Bologna, Italy. Dec 2010.
 - [14] Dan Xu and Xin Liu (2013) “Decentralized Bargain: A Two-Tier Market for Efficient and Flexible Dynamic Spectrum Access”, IEEE Transactions on Mobile Computing, vol. 12, no. 9.
 - [15] Brandon F. Lo, I.F. Akyildiz, Reinforcement learning-based cooperative sensing in cognitive radio ad hoc networks, in: Proc. of IEEE PIMRC 2010, pp.2242–2247
 - [16] Dhananjay Kumar, Kanagaraj .N, .Srilakshmi (2013), “Harmonized Q Learning for Radio Resource Management in LTE based Networks”, in the proceedings of Building Sustainable communities, ITU- T Kaleidoscope , pp. 95-102.
 - [17] Adrian Foster, “Spectrum Sharing and Tariffs-Impact of Sharing on Prices”, *a Seminar on Economic and financial aspects of telecommunications Study Group 3(SG3RG-LAC)*, Y.2011.
 - [18] Ivo Grondman, Lucian Busoniu, Gabriel A. D. Lopes, and Robert Babuska “A Survey of Actor-Critic Reinforcement Learning Standard and Natural Policy Gradients” IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 42, No. 6, November 2012