International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 323-327
Published Online June - July 2016 in IJEAST (http://www.ijeast.com)

PERFORMANCE ANALYSIS OF POMDP AND MDP FOR COGNITIVE RADIO NETWORK

Pallavi K. Jadhav
ME (E &TC) student,
D.Y.Patil college of Engg. &Tech.
Kolhapur, Maharashtra, India

Prof. Dr. S.V.Sankpal
H.O.D. Dept. of Electronics Engineering,
D. Y. Patil college of Engg. & Tech.
Kolhapur, Maharashtra, India

Abstract: In cognitive radio network tcp good put is one of the measure issues to improve the performance of the CR network. However most research work concentrated on performance improvement of tcp has weaknesses as the underlying parameters are only considered to increase the TCP goodput, keeping the transport layer parameter unchanged. The second is formulated as the markov decision process in which the complete knowledge of the network is to be known. Hence to solve the above problem a POMDP base algorithm is proposed in this paper. In this proposed work each CRN users autonomously decides modulation type and power to be transmitted in PHY layer, channels which is to be selected in MAC layer to get best TCP Good put. As the channel is free space and the environment has perception error, this issue is formulated as Partial Observable Markov Decision Process (POMDP).This paper gives the comparison of the MDP with POMDP techniques. Simulation result shows the performance and comparison of the two techniques to achieve the optimal policy to improve the TCP good put in Cognitive Radio Network.

Keywords- CRNs; TCP throughput; MDP; Policy; optimal parameters

I. INTRODUCTION

The spectrum is being diverse and there is lot of challenges in front of the diverse wireless technology and stream traffic services due to the overcrowded unlicensed bands and the Underutilized license bands.

Cognitive radio is an intelligent wireless radio which has a knowledge of the Surrounding environment and which uses the schemes of understanding the behavior of the environment, Then from this environment learning the behavior and according to the environment statistical variations adapting according to the incoming RF stimuli. By changing the operating parameters the adaptation is done such as the modulation, transmitted power and carrier frequency in real time provided, there must be highly reliable communication when needed and also the radio spectrum must be used efficiently. For this parameter configuration a Markov decision process based decision technique algorithm is used.

If the system transits to the next state, depending on the current state only than such system is called Markovian system, in that if the system arrives at the same state twice, the behavior of the system is same. Hence there is no need of the memory uses for the agent operating in Markovian system. It is sufficient to observe the current state of the system in order to predict the system's future behavior. In a controlled Markovian system the agent influences the environment through its actions, yet the effect of an action depends solely on the current state. To choose the next action optimally the agent needs to only consider the current world state. This decision technique will efficiently bring change in underlying physical parameter according to the adaption with the environment and also calculate optimal parameter thereby increasing the throughput of TCP.

II. MDP FRAMEWORK

a. MDP Algorithm:-

This section will give the formal definition of MDP algorithm and the description of value iteration. Then we will describe the Q-learning MDP Algorithm.

A. Description:-
1. s Є S A finite state space.
2. a Є A a finite set of action.
3. T(s,a,s’) Transition function
4. R(s,a) a reward function

Where the transition function specify the probability of taking an action in state s and reaching in state s’ and the reward function specifies the reward the agent will receive after performing action in state s and transition in state s’.

MDP framework assumes that the agent has full knowledge of environment and treats time and set S and action A as discrete. For reinforcement learning algorithm, the MDP does not have to be known. The Markov property says that the state of the environment and the reward the agent receive at time t+1 is stochastically determined by the state of the agent at time t and the action the agent takes. This is the first order Markov process.

$$P(st, rt|s0, a0, \ldots, S_{t-1}, A_{t-1}) = P(st, rt| S_{t-1}, A_{t-1}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (1)$$

Long term reward is maximized by the task of agent. A mapping is required from states to actions as the problem is stochastic. We call such a mapping a policy and denote it as π(s).The long term reward intake is maximized by the optimal policy π*. Certain value is assigned to the agent to compute the optimal policy for being in a state or performing some action in a state.

b. **Value Functions** :

The return R_t of a state is defined as the cumulative reward the agent can expect to receive after reaching the given state at time step t. The sum of all reward the agent received is written as Rt for each time step mathematically weighted by a discount factor γ, where $0 < \gamma < 1$:

$$R_t=r_t+1, \quad Y_{rt}+2, \gamma 2rt+3+\ldots=\sum_0^{\alpha} \Upsilon krt+k+1\ldots\ldots\ldots\ldots\ldots\ldots.(2)$$

There are two purposes for introducing a discount factor (1) it models the preference of the agent to immediate rewards as opposed to those received in the future, and (2) ensures the infinite sum is finite as long as $\gamma < 1$ and the rewards are bounded. When the discount factor is set close to 1, the agent will value future rewards greatly, whereas one close to 0 will make the agent focus on immediate rewards and value the future less. The expected discount cumulative reward is defined as the value of state s under policy π and is given by

$$(V)^{\pi}=E\ [\sum_{k=0}^{\infty} \gamma^k\, r_t + k+1|s_t=s]\ldots\ldots\ldots\ldots\ldots\ldots\ldots...(3)$$

In most situations it is desired to have knowledge of the value of an action in a certain state, we call this the Q-value, with Q(s, a) providing the value of taking a in s, it is defined as:

$$Q^{\pi}(s.a) =E\ [\sum_{k=0}^{\infty} \gamma^k\, r_t + k+1|s_t=s,a_t=a]\ldots..(4)$$

Assuming the values of all successor states s′ are known to the agent, “Eq. (4)” can be rewritten as the reward the agent receives plus the discounted value of s′, weighted by the probability of ending in s′, after taking action a in s:

$$Q^{\pi}\ (s,a) = \sum_{s'} T(s,a,s^{'})[R(s,a) + \gamma V\pi(s^{'})]\ldots\ldots\ldots..(5)$$

This formula is a form of the Bellman equation named after Richard Bellman, who introduced it in 1957 [3]. With this function, we can iteratively update the value of all states, until it reaches a convergence criterion, resulting in an optimal state-value function V*(s), from which we can derive an optimal state action-value function Q*(s, a). Knowing the value of all states, the agent can select the action with the highest utility in every state, which will lead to an optimal policy. Value iteration is an algorithm that uses this concept.

III. POMDP FRAMEWORK

Sequential decision making is provided by the natural model of POMDP under uncertainty. A framework of MDP is augmented by this model to the situation where the secondary user cannot reliably identify the underlying environment of spectrum occupancy state. The important characteristics which keep the POMDP apart from different models are that the state is not directly observable. Instead the agent can only perceive observations which convey incomplete information about the world’s state [6]. It is very important tool which increases the application of MDP to many realistic problems. POMDP is characterized by seven distinct quantities namely states (S), actions (A), observations (Θ), reward (R) and the three probability distributions namely transition probabilities (P), initial belief (b0), and observation probabilities (θ) [6].All of this items together describes the probabilistic system model that underlies each POMDP. In this work we have not studied deep about the development and analysis of the POMDP solution, instead we will make use of available POMDP solution in our paper to achieve optimal parameters.

Let s denotes the instantaneous state of the system, so that the finite set is denoted by $S = \{s_1, s_2 \ldots..s_n\}$ and the nth channel state is denoted by $S_n(t)$. As Under the POMDP frame work the state of the system is not directly observable by the CR users so the CR user can calculate only the belief state over the state space. CR nodes take the sensor measurement result regarding the information of the belief state. Φ denotes the sensor measurement result such that $\Phi= \{\Phi 1, \Phi_2, \Phi_3\ldots.\}$ and the nth channel observation is denoted by Φn(t). Hence due to the spectrum sensing error Φn(t)is a incomplete projection of n^{th} channel at time t. The POMDP framework can be defined precisely only by specifying the state transition and observations by probabilistic law. This law includes the initial probability distribution (b_0) which gives the probability of the system to be in state s at time t=0, provided that this distribution is defined over all states in S.

The topology used is distributed topology hence CR users can get only the part state information of the whole network. Therefore, the dynamic parameters configuration is formulated as a POMDP which can be defined as a tuple (*S*, *A*, P, *R*, *Z*, *O*).In this model, S defines the all possible state space, A defines the array of action to be performed in state s: *R*:*S*×*A*→R represents the reward function that gives reward for the action perform in state s; P : S×A×S→S is the state transition probability for the state transition from s to s' ; *Z* stands for the set of observable history information which will give history base on the observations; O : S×A×Z→O depicts the observation function, which can calculate the potential observation of next state after an action. The brief information of each element in POMDP is given bellow.

a. System State:-

System state defines the set of possible states. It gives channel gain as s^n=(G^n) in POMDP as system state, where G^n is a matrix of channel gain and $gcl^n=g^n(c,l)$ is the channel gain of *c*.

b. System Action:-

Let A^n=(Pow^n, mod^n, X^n) depict the action space of n^{th} slot ,where $A^n=A_1{}^n\times... \times A_L{}^n$, $A_l{}^n$ is the action space of *l*. Pow^n and $mod^n\in R^L$ represent transmission power vector and modulation type vector, respectively. $X^n=(X_1{}^n,...,X_L{}^n)$stands for the channel allocation vector, which meets the condition $X_l{}^n=\{x\in\{0,1\}C$, $x\cdot Y^n=0\}$, in other words, the channels employed by CRN cannot conflict with registered network's, and the channels selected by each CR user should be less than or equal to m_l.

c. Reward Function:-

For the packet transmission, action $a_l{}^n$ is performed by the CRs users. Corresponding to the action perform TCP goodput is taken as the reward in acknowledgement stage of each slot.

$$r(s^n,a^n) =\frac{\sum_{Xl}n(c)=1\mathrm{Th}^n(c,l)}{\sum_{Xl}n(c)=1\ \mathrm{Band}(c)}........(6)$$

Where band(c) is the bandwidth of channel c. The product of TCP good put for each CR user is expressed as the average network utility, which is presented by equation.

$$R\ (S^n,a^n)=\prod_{i=1}^{L} r(S^n,a_l^n)..............(7)$$

d. Observation History and Observation Function:-

Let z^n represent the history information collection of past n slots, where $z^n=\{s^0, a^0, r^0, \ldots, s^n, a^n, r^n\}$ .This includes three elements such as state, action and reward function. O is the confidence probability which represents the distribution function of system states from s^n to s^{n+1} after action a^n. This transition takes place base on the history observation information which is express by $o(s^{n+1},a^n,z^n)=Pr(z^n|s^{n+1}, a^n)$.

IV. SIMULATION RESULTS ANALYSIS

Here we have taken NS2 platform. Under this platform we have assume 75 CR users are randomly distributed in a square area of 600mx500m and they can access 5 wireless channels. Each channel occupancy is given by pu for the registered users. The TCP packet length *Ltcp* is set to be 1500 bytes and the maximum number of retransmission *Nre* is 5. The maximum congestion window *cwnd* is given by 6000 bytes and initialize timeout *T0* is 2*s*. The ARQ protocol is selected in MAC layer and the maximum frame retransmission *Nfr* is 10, of which header length *Lfrh* is set to be 20 bits. The ACK frame length *Lack* is 24 bits and the bandwidth is assumed to be 1MHz. This paper assumes that each CR user can either be a sender or a receiver in a certain slot, while all of them are working abidingly.
After the simulation of 30 slots, the fig 4 calculated the delay for the packet transmission for the different nodes. According to the delay calculated, the throughput is calculated as shown in the figure 6. From figure 7 and 8 the optimal goodput and the average goodput is been calculated for different nodes by the simulation using the POMDP algorithm as shown in the graph below.

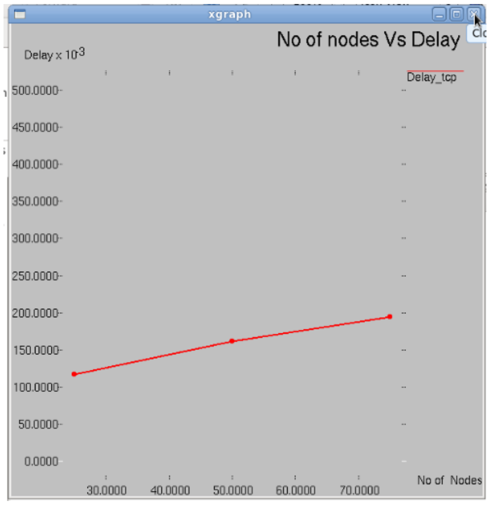

Figure 1: Graph of no. of nodes vs delay for mdp

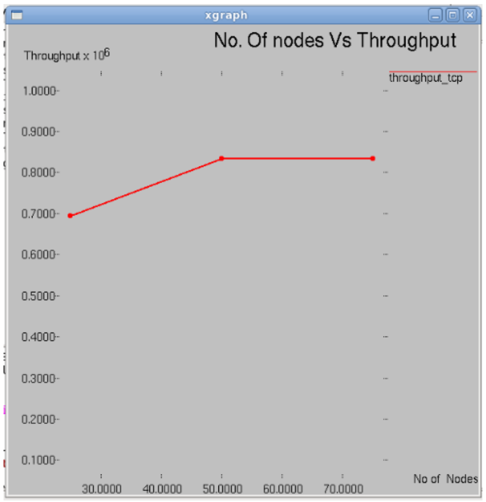

Figure **2**: Graph of no. of nodes vs throughput for mdp

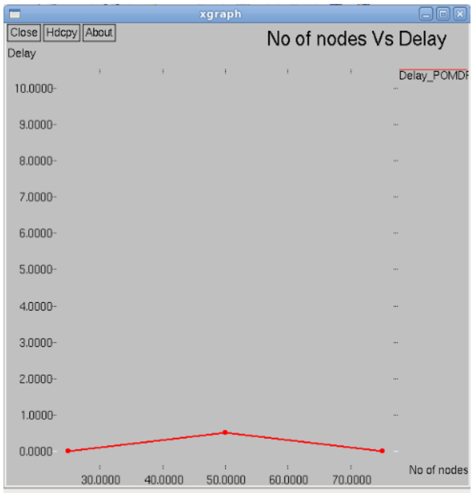

Fig 3:- Graph of no. of nodes vs delay for pomdp

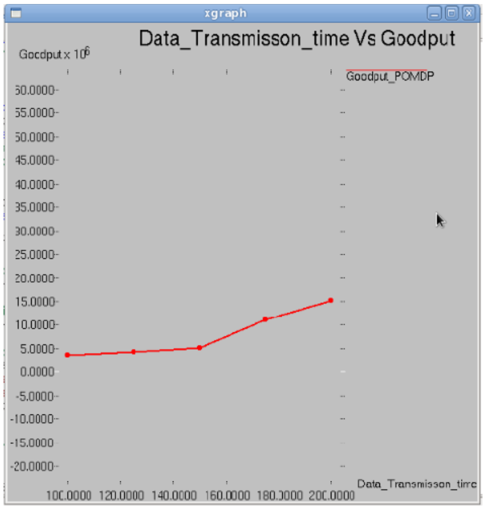

Fig 4:- Graph of no. of nodes vs Throughput for pomdp

V. CONCLUSION

In the distributed network, end to end TCP performance is one of the main criteria to measure the network performance. This paper present a MDP and POMDP base optimal parameter configuration schemes in CRN. Where MDP requires the complete knowledge of the network and POMDP works on the partial information about the network. Hence this paper Analysis the two techniques and the simulation graph shows the difference in the performance of the two techniques towards the enhancement of the TCP throughput. The simulation results show that the POMDP can find the optimal parameter in environment of perception error as compared with the traditional MDP algorithm.

VI. REFERENCES

[1] S. Hykin, "Cognitive Radio: brain-empowered wireless communications," *IEEE J.Sel. Areas* Commun, Vol. 23, no. 2, pp. 201-220, 2005.

[2] Hsien-Po Shiang and Mihaela van der Schaar, "Multi-User Video Streaming Over Multi-Hop Wireless Networks: A Distributed, Cross-Layer Approach Based on Priority Queuing, "IEEE Journal on Selected Areas in Communications, vol. 25(4), 2007.

[3] Fangwen Fu, Mihaela van der Schaar, "Decomposition Principles and Online Learning in Cross-Layer Optimization for Delay-Sensitive Applications, "IEEE Transactions on Signal Processing, vol. 58(3),2010.

[4] Yu Zhang, Fangwen Fu and Mihaela van der Schaar, “On-Line Learning and Optimization for Wireless Video Transmission, ” IEEE Transactions on Signal Processing, vol. 58(6), 2010.

[5] Zhichu Lin, Mihaela van der Schaar, “Autonomic and Distributed Joint Routing and Power Control for Delay-Sensitive Applications in Multi-Hop Wireless Networks,” IEEE Transactions on Wireless Communications, vol(1), 2011.

[6] J. Pineau, Tractable planning under uncertainty: Exploiting structure, Ph.D. Dissertation, Rutgers University, 2004.

[7] Q. Zhang, S.A. Kassam, Finite-state Markov model for Rayleigh fading channels, IEEE Transactions on Communications47(11)(199)1688-1692.