DOI: https://doi.org/10.52756/ijerr.2024.v46.026 *Int. J. Exp. Res. Rev.*, Vol. 46: 326-341 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in

Proximal Policy Optimization for Efficient Channel Allocation with Quality of Service (QoS) in Cognitive Radio Networks

Kalyana Chakravarthy Chilukuri[1], N Chaitanya Kumar[2], T. Vidhyavathi[3], Regidi Suneetha[4], V Sita Rama Prasad[5], Badugu Samatha[6] and Mahanty Rashmita[7]*

[1]Department of Computer Science and Engineering, MVGR College of Engineering(A), Vizianagaram, Andhra Pradesh, India; [2]Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences(A), Visakhapatnam, Andhra Pradesh, India; [3]Department of Electronics and Communication Engineering, Anil Neerukonda Institute of Technology and Sciences(A), Visakhapatnam, Andhra Pradesh, India; [4]Department of Electronics and Communication Engineering, Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh, India; [5]Department of Computer science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India; [6]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh, India; [7]Department of Basic Sciences and Humanities, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

E-mail/Orcid Id:

KCC, kch.chilukuri@gmail.com, https://orcid.org/0000-0002-8137-7074; *NCK*, chaitanya.india9@gmail.com, https://orcid.org/0009-0006-0535-7174; *TV*, vidyavathi.ece@anits.edu.in, https://orcid.org/0000-0003-3609-729X; *RS*, suneetha.ece@svpec.edu.in, https://orcid.org/0000-0003-2204-5480;*VSRP*, vsrprasad45@gmail.com, https://orcid.org/0009-0002-4924-8712; *BS*, bsamatha@kluniversity.in, https://orcid.org/0000-0003-1353-2797; *MR*, rashmitamoon@gmail.com, https://orcid.org/0000-0001-9247-8295

Article History:

Received: 18th Jul., 2024
Accepted: 25th Dec., 2024
Published: 30th Dec., 2024

Keywords:

Reinforcement learning, Channel allocation, QoS, Q-Learning, Actor-critic, PPO

How to cite this Article:
Kalyana Chakravarthy Chilukuri, N Chaitanya Kumar, T. Vidhyavathi, Regidi Suneetha, V Sita Rama Prasad, Badugu Samatha and Mahanty Rashmita (2024). Proximal Policy Optimization for Efficient Channel Allocation with Quality of Service (QoS) in Cognitive Radio Networks. *International Journal of Experimental Research and Review*, *46*, *326-341*.
DOI:
https://doi.org/10.52756/ijerr.2024.v46.026

Abstract: A multi-variable relationship exists in Cognitive Radio Networks (CRNs) where factors such as Energy efficiency, Throughput, Delay and Signal Noise Ratio (SINR) are related. The SINR shows the quality of the signal and is defined as the total power of a specific signal over the total power of an inter signal plus noise. This work proposes an effective energy and delay-efficient channel allocation strategy for CRNs (Cognitive Radio Networks) using Q-Learning and actor-criticism algorithms that maximize rewards. We also propose a Proximal Policy Optimization (PPO) algorithm that uses clipping of surrogate objectives to prevent large policy changes and ensure that the other parameters remain stable over time. We study the tradeoff between rewards, energy efficiency and other parameters and compare the algorithms with respect to the same. Results show that the proposed PPO method, while using optimally increased energy consumption, significantly reduces the delay, improves the thought and reduces the packet loss ratio for efficient channel allocation. This is positive with our findings shown in the results section and by comparing the proposed method with other algorithms to identify improved throughput and channel utilization. As the simulation results indicate that the PPO algorithm has very high throughput and significantly minimizes the delay and packet loss, it is suitable for application in all sorts of services such as video, imaging or M2M. The results are also compared with two of the existing channel allocation schemes and they confirm that the proposed algorithm performs better in terms of throughput discussed in one scheme and channel efficiency in the other.

Introduction

Although the application of deep learning techniques for spectrum allocation in in CRNs has been widely studied, the problem of channel allocation, which is essential for routing, has not been dealt with extensively. Most of the existing algorithms are focused on efficient spectrum allocation. These techniques focus strictly on a few particular aspects, such as effective queuing of the secondary users or classifying the secondary users based on their priority levels. Some of these techniques focus

***Corresponding Author:** rashmitamoon@gmail.com

on assigning frequency bands to services based on the type and the Quality of Service requirements. Much of the work pertains to different strategies such as game-theory-based, centralised dynamic channel allocation using SDNs, Fuzzy logic-based allocation etc. At a deeper level, channel allocation consists of assigning specific frequency channels to the users within the identified spectrum.

Moreover, existing channel allocation algorithms using Deep or Reinforcement Leaning focus on optimizing the rate and transmission power allocation based on SINR, energy optimization, collision detection and a few other aspects. Very few algorithms have performed effective channel allocation based on the QoS parameters, viz., delay throughput and SINR, while optimizing the energy efficiency in Cognitive Radio Networks (Madhuri et al., 2022; Kumar et al., 2024; Varshney et al., 2024). There have only been a few algorithms that address two or three of these parameters at the most.

In cognitive radio networks (CRNs), the relationship between energy efficiency, throughput, delay, and signal-to-noise ratio (SINR) is complex and has interdependencies. The SINR indicates the signal quality and is the ratio of the power of a desired signal to the power of interference plus noise. Higher SINR enables better modulation schemes, resulting in an increased throughput. The total Energy consumption depends on the transmission, reception, and processing power of CRN nodes. Throughput refers to the amount of data transmitted successfully in a given time frame on a network (Nayani et al., 2021). Delay is the total time for packets to reach from a sender to a receiver. Both throughput and delay are affected by poor channel allocation.

To achieve a higher SINR, more energy is required. But again, too high energy would lead to reduced throughput due to failures. Thus, finding the right balance of energy in order to optimize the other Quality of Service (QoS) parameters remains a difficult challenge due to the dynamic nature of channel allocation, particularly to the Secondary Users(SUs) in CRNs. Efficient channel allocation in Cognitive Radio Networks results in proper spectrum utilization, interference mitigation, scalability and QoS, ensuring critical services operate without interruptions. This is a particularly necessary requirement with CRNs, where SUs need to vacate quickly as soon as the primary users (PUs) re-occupy the channels.

Related work

El-Toukhy et al. (2016) propose a Markovian-based queuing approach that reduces the number of states for priority Secondary Users by reducing their blocking and dropping probability and optimizing their throughput, both in case of the arrival of other higher-priority SUs or with increasing number of PUs. Results show that in the case of the arrival of more PUs, the blocking probability of SUs reduces, being the least for higher priority SUs. In the case of the' arrival, SUs' blocking probability of SUs increases, still being the least for higher priority SUs. In case of the arrival of more PUs, the dropping probability of SUs increases, again being the least for higher priority SUs. This results in better throughput of higher priority SUs with increasing PUs and other SUs. However, the throughput keeps reducing in the case of PU arrivals, while it increases in the case of SU arrivals. This is obvious since, with the proposed approach, secondary users need to vacate immediately irrespective of the PUs priority, whereas in the case of SU arrivals, only the lower priority SUs will have to vacate, resulting in improved throughput.

Azaly et al. (2020) suggest an SDN controller-based channel allocation scheme for CRNs using efficient spectrum allocation. They state that the dynamic channel allocation for SUs can be performed by the SUs themselves using distributed channel allocation without the need for a centralized controller channel or a common control channel that collects information from all the SUs. However, the authors focus on the centralized approach of using the SDN controller. Two separate algorithms are proposed for Dynamic Channel Reservation (DCR), one to increase the SU retainability by assigning a higher number of channels when the PU load increases. The SU retainability is obtained by subtracting the probability of an SU successfully completing its service request from 1. The second algorithm maximizes the channel availability of the SUs by lowering the blocking probability, i.e., releasing the reserved channel in case of increased SU load. The PU and the SU probabilities are calculated separately based on the steady-state probability matrix. The SU Handover probability is the probability that at least one SU will be holding one channel, which will be handed over to a PU upon its arrival. The throughput of PU is calculated based on the PU availability and the PU arrival rate. The throughput of the SU is calculated based on the SU arrival rate, SU availability, and SU retainability. SU throughput probability decreases with the SU arrival rate. The SU cost function is higher in case of a higher arrival

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

rate of SUs. The cost function for the SU is defined as one that minimizes the Retainability and the Availability and maximizes the Handover probability. Results show that the DCR is superior to Static Channel Reservation(SCR) in terms of availability for higher or lower secondary user arrival rates and particularly for high primary user arrival rates. Also, DCR is superior to Static Channel Reservation(SCR) in terms of attainability, particularly for lower primary user arrival rates where the second algorithm is recommended. Overall, the proposed DSA algorithms significantly enhance the system performance in terms of both availability and retainability.

Zhang et al. (2023) propose a dynamic channel allocation protocol DPrA and compare it the Maximum Throughput Allocation(MTA) protocol. After calculating the steady state probability vector based on the transition state probabilities of the SUs and the activity state of the PUs, performance metrics for queuing analysis, throughput, queue length and the packet rejection state are found to assess the SUs' performance. The proposed algorithm was observed to perform better than the MTA on all fronts, i.e., improved thoughput, reduced queue length and reduced packet rejection rate.

Scientists are measured user satisfaction based on the index of the preferred channel allocated. The channel might or might not be allocated to the user based on the utility offered for the channel. Authors compare their strategy with SMC-MAC where random allocation of channels takes place and PPDA, where allocation takes place based on the offered price for the channel. According to the proposed method, preferred channels are not always assigned to the same users; instead, every user gets a fair share of their prioritized channels, improving channel utilization and user satisfaction.

Azaly et al. (2020) define a SU cost function as a weighted sum of channel non-availability (blocking probability), forced termination due to PU arrival and SU handover probability. Authors propose and compare two algorithms, one of which SU retainability by assigning higher number of reserved channels when the PU arrival rate increases and the second maximizes the Sus's availability by lowering the number of reserved channels and making them available for other SUs. Results show that the introduction of reserved channels improves SU performance by reducing the cost of over-SU. Also, the service completion rate of the secondary users increases with an increase in the number of reserved channels.

Dey and Misra (2018) propose a channel allocation method in CRNs for video traffic by calculating the Channel Quality Index based on the packet error and the channel data rate. The probability of a false alarm and the probability of detection calculated earlier is used to determine the channel data rate and the collision rate(used to determine the packet error rate). The Quality of Experience is evaluated based on the difference in the mean opinion scores obtained for different video types. The proposed schemes in this work are compared with traditional random and uniform channel assignment schemes where the channel quality estimation and QoS requirements are not considered for channel allocation and it was observed that there was a huge improvement in the Quality of Experience of Gentle Motion and rapid Motion videos (Malik, 2020).

Nayak et al. (2020) propose an Energy detection-based channel allocation technique for Cognitive Radio Networks. Two registers are used, one to count the number of samples up to eight and the other to add the energy of all the cumulative samples, the operations synchronized by a Multiplexer through a counter. The signals are produced using the modulation techniques BPSK and QPSK and are checked for different probability detection techniques. It was observed that for a very signal-to-noise ratio, the proposed energy detection technique works much better than the matched filter or cyclostationary feature detection technique.

Ye et al. (2020) propose a Deep reinforcement learning-based technique for maximizing the rewards, determined by the SINR of both the primary and the secondary users on different channels, that is also regulated by the power constraints of PUs using an improved strategy. This strategy alleviates the energy consumption for frequent power adjustments, using the SINR prediction based on the current SINR, during which only the transmitted power has been adjusted. The proposed LSTM-DQN algorithm was compared with two existing algorithms, DQN and Priority Memory DQN(PM-DQN), for various transmit power ranges of Primary and Secondary users under SINR thresholds using a two-ground reflection. Five hidden layers, with the topmost being the LSTM layers, are employed. RELU activation function for three layers and tanh for the other layer were used. Adam optimizer is used to update the weight of the Nueral Network. Results show that the proposed method yields higher rewards compared to the other two methods, confirming the effectiveness of the joint rate and power control strategy.

Jang et al. (2019) propose an optimal method for band selection and channel selection in Cognitive Radio Adhoc-Networks (CRAHNs) using a Q-learning strategy. The reward function is designed to maximize the SU network average operation time, desired data rate, and

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

channel utilization by providing a fair allocation of channels between users. The reward function is readjusted based on the Q-Learning learning parameters, which in turn allows the Q-Learning module to determine the channel as well as it's band group such that the channel is efficiently utilized while ensuring it's data rate demand.

Xu et al. (2024) model channel assembling in CRNs by prioritizing elastic services into high, medium and low categories. Information validity is measured by the time that remains before the deadline. Message correlation measures the importance of information, message size measured as the proportion of the SUs message to the average length of messages. These are the three dynamic parameters to determine whether elastic services have higher priority over real-time services. Elastic services with lower priority are always scheduled after real-time services. Separate reserved queues are used for elastic and real-time traffic and whenever a Secondary User is interrupted by a Primary User, the SU is sent to a separate packet classifier to determine its priority mentioned above. To prevent starvation of this quote, another packer classifier is used in case of partial interruption of a high-priority SU in order to provide differentiated service. Results show that while the network capacity of both elastic and real-time SUs decreases as the PU arrival rate increases. However, this decrease is smaller for real-time SUs compared to elastic SUs. The spectrum utilization of the secondary network was found to increase with the increase in the queue capacity, and the blocking probability of high-priority elastic SUs reduced significantly compared to real time SUs, with an increase in the PU arrival rate.

Wang and Liao (2018) propose a centralized Fuzzy Inference based channel allocation scheme for SUs. The cumulative signal power is used to detect the presence of a PU if it falls above a threshold. In order to avoid incorrect detection due to hidden terminal problems, when a minimum of n SUs detect the presence of PU, then the channel will be unavailable for the SU. For the priority of allocation of the channel, a membership function is defined using four input variables – Spectrum utilization efficiency, distance, mobility and signal strength and 81 rules that classify the priority as High, Very High, Medium, Low and Very Low are defined in the rule base. Authors compare their scheme with the random and Kaniezhil's schemes and observe that SU throughput is better with increasing SNR.

Chakraborty and Misra (2020) have proposed a three-phase approach that uses a proactive, reactive target channel sequence determination followed by CCC allocation to minimize the handoff delay call drop probability and improve channel prediction.

Tlouyamma and Velempini (2021) provide an improved channel selection algorithm based on sensing probability. The probabilities of the on-off periods of the SUs are determined from the PDF.

Pal and Dahiya, (2020) use an objective function that minimizes the chance of a PU arrival. The consistency of level of channel occupancy, consistency of occupancy and the difference between current and previous channel capacities are used in the objective function to determine the rewards. Higher rewards refer to the arrival of PU, in which case the above factors are dynamically adjusted. Authors compare this Q-Learning-based approach with three other existing methods and observe that the throughput and the PDR reduce as the number of vehicles increases.

Hossen and Yoo (2019) have worked on a Q-Learning-based clustering method that performs spectrum sensing based on a reward function that finds the channel availability subject to the received signal's energy threshold. Further, the cluster objective function finds the fitness value based on the residual energy, fitness of channel associated with the node and the number of clusters that can be reached through the node. This method is compared with the k-means and the multi-channel-based clustering methods and was observed to yield an improved lifetime of clusters with reduced interference between clusters.

Srivastava et al. (2024) propose a Q-Learning based channel selection method that considers the channel capacity, interference power and the cumulative distance from the transmitters to the receiver. The reward function is updated based on the probability of channel availability or non-availability (i.e., increased or decreased). The packet deliver ratio. Throughput was observed to have been higher compared to random allocation and CCCA, while the delay and PU collision ratios were lower.

Jang et al. (2019) use a Q-Learning approach, where the reward function is designed to optimize the data rate, occupancy, band change and date rate efficiency. Results show an increase in data rate efficiency of the proposed scheme compared to random selection or Max-Q selection.

Pal and Dahiya (2022) and Pal et al. (2020) have improved channel selection in Cognitive Radio Adhoc Networks by using a weight function that includes channel occupancy and considering several parameters. The solution to this Linear Programming Problem allows the selection of optimal channels.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

Existing System

Zhao et al. (2019) have proposed a joint Power and interference mitigation-based channel allocation method(JPCRL). This method finds the SINR by subtracting the path loss, interference power from neighbouring transmitters and noise from the Direct Received Signal Strength(DRSS). If the interference and the noise are larger compared to the received signal strength, it would result in a low SINR, severely affecting the communication. Authors focus on maximizing the throughput subject to the SINR and other association constraints. The difference of the total throughputs between consecutive states measures the immediate rewards, while the long time rewards are measured by a cumulative rewards measured by a utility function and for the purpose of maximizing the sytem throughput, the network with the maximum average utility is chosen.

$$Qt+1(S,A)=Qt(S,A)+\alpha[R+\beta\cdot A'\max Qt(S',A')-Qt(S,A)]$$

According to this Q-Learning update policy, the Q-value is updated in the current state, considering the rewards obtained in the previous state and the expected rewards in the future. β is the discount factor that balances immediate and future rewards, and α is the learning rate that updates the Q-value based on the (predicted-observed) rewards. Although channel utilization is not directly measured in the method, the DRSS affects the SINR, which determines the channel utilization. Only when the SINR is above a threshold will the signal be transmitted successfully and the channel utilization be efficient. Results focus on the throughput with increasing users.

A Deep Q-learning-based algorithm was proposed by Pavan et al. (2024) for channel allocation in Cognitive Radio Networks. This is an extension of the above work. The reward function is defined as the ratio of SINR to the SNR in each state, and the optimal policy maximizes the rewards. The channel assignment to the SU is done based on the updated Q value, if it falls below the threshold value. Authors compare the utilization of the channels with the JPCRL and observe that their proposed method DRLCA outperforms the latter JPCRL in terms of both the maximum and the minimum channel utilization and with increasing thresholds as well as with the increasing number of iterations. However, the System throughput comparison is not done with the above work.

Materials and Methods

Q-Learning Algorithm

Q-Learning is a model-free reinforcement learning method that finds the optimal action-selection policy for an agent that interacts with an environment. The main element of Q-Learning is the Q-table, which stores the expected utility of taking a given action in a given state.

Following is the learning process:

#The agent will explore the environment and choose an action using a ε-greedy strategy (it explores random actions with some probability. It will not choose the best-known action).

#After taking the action, the agent receives some reward and proceeds to observe the next state.

#The Q-value will be updated as

$$Q(s,a)\leftarrow Q(s,a)+\alpha[r+\gamma\max_{a'} Q(s',a')-Q(s,a)]$$

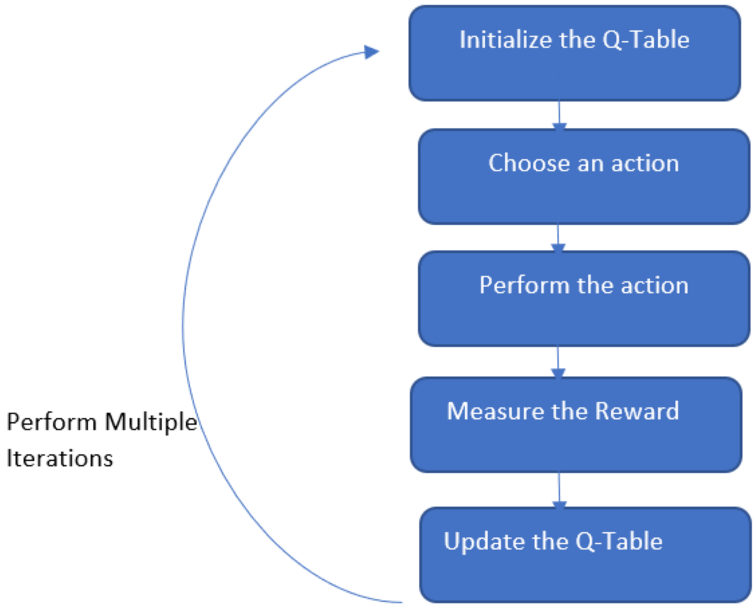

Figure 1. Q-Learning algorithm

Figure 1. xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

Here:

- $Q(s,a)$ is the Q-value for state s and action a
- α is the learning rate
- r is the received reward
- γ is the discount factor
- s' is the next state
- $\max_{a'} Q(s',a')$ is the Maximum Q-value for the next state s', among all possible actions a'

α, γ, ε and the number of episodes are the hyperparameters in Q-Learning algorithm.

Actor-Critic Algorithm

The Actor-Critic method offers the combined benefit of the value-based and the policy-based approaches. The actor will suggest actions based on the current policy, and the critic will evaluate the action taken by the actor, giving feedback on how good the action was.

Following is the learning process

- The actor will update the policy based on the critic's feedback.
- The critic will evaluate the action taken by using the Temporal Difference (TD) error

$\delta=r+\gamma V(s')-V(s)$

Here:

- δ is the Temporal difference error (TD error).
- $V(s)$ is the value function for the state s
- r is the reward received after taking an action in the state s
- $V(s')$ is the value estimation for the next state s'

The value function will be updated as:

$V(s)\leftarrow V(s)+\beta\delta$

Here, β is the learning rate of the value function.

During the learning process, the actor will update it's policy, using the Temporal Difference error (δ) to improve it's selection of it's future action as:

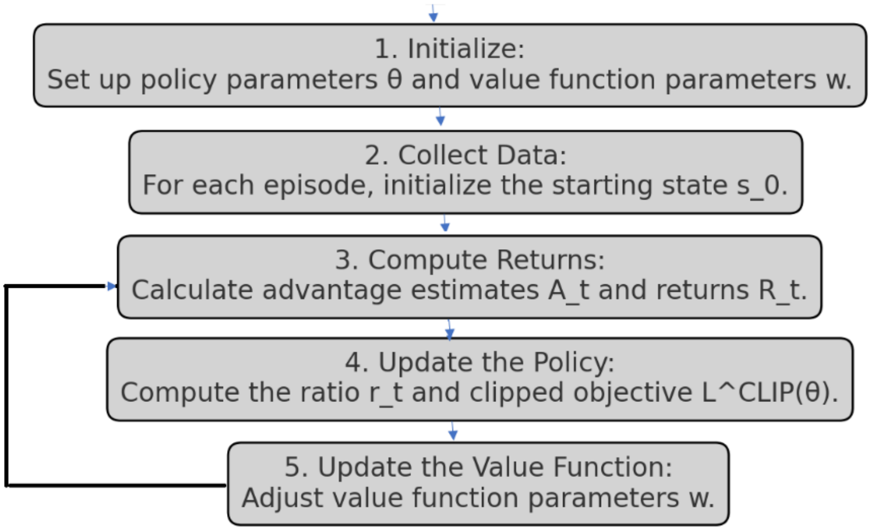
1. Initialize:
Set up policy parameters θ and value function parameters w.
2. Collect Data:
For each episode, initialize the starting state s_0.
3. Compute Returns:
Calculate advantage estimates A_t and returns R_t.
4. Update the Policy:
Compute the ratio r_t and clipped objective L^CLIP(θ).
5. Update the Value Function:
Adjust value function parameters w.

Figure 2. The PPO algorithm.

$\pi(a|s)\leftarrow\pi(a|s)+\alpha\delta$

Here:

- $\pi(a|s)$ is the probability of taking action a in state s.
- α is the Learning rate for the policy.

α , β, γ are the hyperparameters in Actor-Critic algorithm.

Proximal Policy Optimization (PPO)

PPO is a policy gradient method that aims to improve the stability and reliability of policy updates. It will restrict the change to policy updates by means of a surrogate objective function.

The advantage estimates are calculated as

$A_t=r_t+1+\gamma V_w(s_{t+1})-V_w(s_t)$

During the learning process, the objective function used is

$L^{clip}(\theta)=E_t[\min(r_tA_t,\mathrm{clip}(r_t,1-\epsilon,1+\epsilon)A_t)]$

Here:

- $r_t=\pi_\theta(a_t|s_t)\ /\ \pi_{\theta old}(a_t|s_t)$ is the probability ratio.
- A_t is the estimated advantage.
- ϵ is the hyperparameter that would control the allowed deviation of the new from the old policy.

Policy parameters θ are updated using gradient descent using

$\theta\leftarrow\theta+\alpha_\theta\nabla L^{clip}(\theta)$

Value function parameters are updated using

$w\leftarrow w+\alpha_w\nabla MSE(R_t-V_w(st))$

α_θ, α_w, clip ratio ϵ and number of epochs K are the hyperparameters to be set during initialization.

Reward Calculation

The reward is calculated by combining all the performance metrics into a reward function defined as $R=W_{throughput}.TPUT+W_{Energy}.ENER-W_{delay}.DEL-W_{packet\_loss}.PL+W_{SINR}.mean(SINR)$, Where R is the total reward,TPUT is the throughput achieved, ENER is the energy consumed, DEL is the delay, PL is the packet loss incurred.

$W_{throughput}$,W_{Energy},W_{delay},W_{packet_loss},W_{SINR} are the respective weights assigned to each performance metric based on their priority.

Let ENER represent the energy consumption, TPUT the throughput, and DEL the delay. In order to obtain the optimal energy, ENER needs to be maximized, subject to constraints on TPUT and DEL on the Langrangian function

$L(ENER,TPUT,DEL,\ \lambda_{TPUT},\ \lambda_{DEL})=ENER+\ \lambda_{TPUT}\ (TPUT_{min} - TPUT)+\ \lambda_{DEL}\ (\ DEL_{max} - DEL)$

Here, λ_{TPUT} and λ_{DEL} are the Lagrange multipliers corresponding to the throughput and delay constraints.

The solution for the optimal energy can be obtained by

$\partial L/\partial ENER = 0$, $\partial L/\partial TPUT = 0$ and $\partial L/\partial DEL = 0$

Which means for the optimal energy consumption ENER, the constraints if any, on the throughput and the delay must be optimized.

A similar approach might be adapted for throughput optimization

$L(TPUT,ENER,DEL,\lambda ENER,\lambda DEL)=-TPUT+\lambda ENER(ENERmax-ENER)+\lambda DEL(DELmax-DEL)$

Here, $\partial L/\partial TPUT$, $\partial L/\partial ENER$ and $\partial L/\partial DEL$ can be calculated respectively, which means for optimal throughput TPUT, the constraints, if any, energy consumption and delay must be optimized.

For an optimal solution to the relaxed objective function, the multi-optimization equation given below can be solved by

minimizing ENER subject to $TPUT \geq TPUTmin$, $DEL \leq DELmax$

$L_{Relaxed}(ENER,TPUT,DEL)=ENER+W_{TPUT}\cdot \max(0,\ TPUT_{min}-TPUT)+W_{DEL}\cdot \max(0,DEL-\ DEL_{max})$

Evaluation metrics

For the purpose of comparing the proposed method with the other methods, the following metrics were used.

#Throughput: The rate of successful transmission of data in the network

#Energy Consumption: The energy consumed by the nodes in the CRN.

#Delay: The time taken by the packets to reach the destination from a source.

#Packet Loss: The percentage of packets lost during the transmission.

#Channel utilization is measured by the number of active channels whose SINR is above the specified threshold.

#Reward: The maximum total reward obtained by optimizing the above QoS parameters, subject to the constraints.

The following are the simulation parameters, and the rewards are calculated using the reward function.

#Number available channels - n_channels: 5

#Number of SUs - n_users: Varies from 10 to 90

#The Signal to Noise Threshold-sinr_threshold: 15

#Number of iterations:env_steps: 500

Reward weights

#Throughput weight-w_throughput: 0.5

#Energy consumption weight-w_energy: -3

#Delay weight-w_delay: -1

#Packet Loss weight-w_packet_loss: -1

#SINR weight-w_SINR: 1

Three agent classes were created: the Q-LearningAgent, the ActorCriticAgent and the PPOAgent. The Q-LearningAgent uses Ɛ greedy strategy for action selection to balance exploration and exploitation.

Results and Discussion

The proposed method implements three reinforcement learning algorithms Q-Learning, Actor-Critic and Proximal Policy Optimization algorithms and highlights the advantage of PPO over other methods in terms of optimizing most of the QoS parameters in the reward function.

It can be observed from the results that while the PPO algorithm results in more Energy consumption for channel allocation compared to the other two methods, it far out performs the Q-Learning and the Actor-Critic algorithms in terms of delay, throughput as well as the packet loss. The reason is that the PPO algorithm favours exploration compared to other methods, owing to its policy updating mechanism that uses a clipped objective to ensure that policy updates do not diverge too much from the previous policy. This might result in suboptimal choices that increase energy consumption while trying to discover better strategies. However, PPO incorporates the Generalized Advantage Estimation (GAE) to estimate the advantage function, leading to more informative updates and improved learning signals. This is particularly beneficial in optimizing for throughput and minimizing delay and packet loss. In other words, since PPO updates its policy continuously based on new experiences, it explores more aggressive actions that will maximize throughput or reduce delay, even if they consume more energy. It can be seen that the PPO algorithm offers lesser rewards when compared to the Actor-Critic algorithm. This is because the clipping in PPO prevents the policy from changing too much during updates,

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

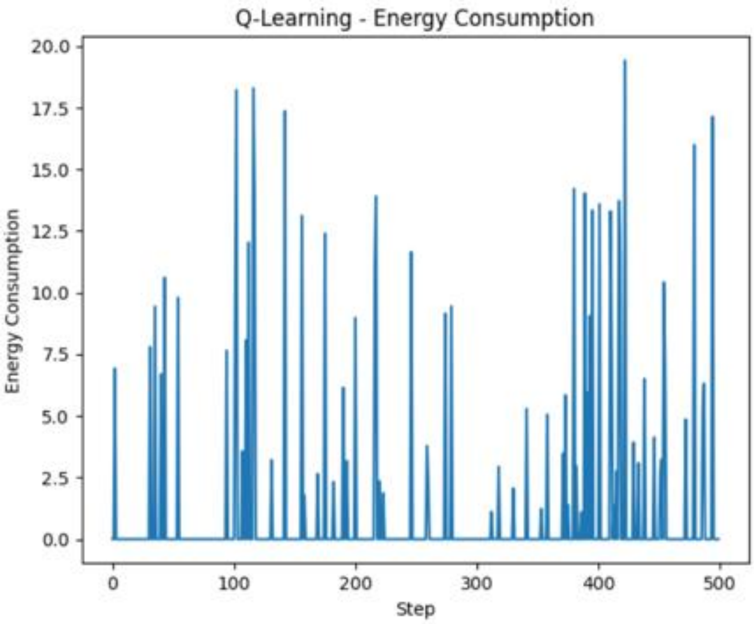

Figure 3. Energy Consumption graph for Q-Learning.

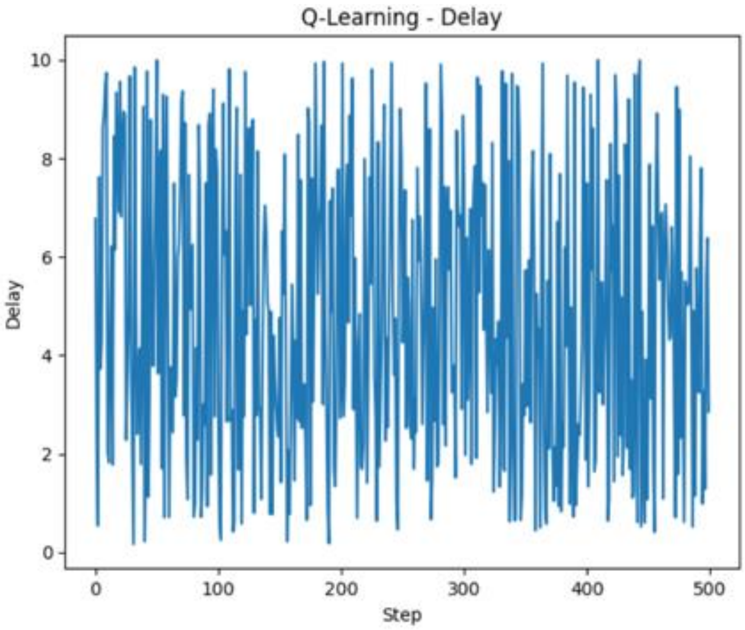

Figure 4. Delay graph for Q-Learning.

resulting in stable but slow exploration. On the other hand, Actor-Critic follows an aggressive policy adaptation, resulting in more rewards in complex environments. As the considered environment is particularly dynamic, with varying delay, throughput etc., PPO tends to be less aggressive. Q-Learning and Actor-Critic algorithms are able to gather higher rewards more quickly without reference to the policy changes. For a similar reason, Q-Learning's focus on exploiting energy-efficient policies means it overlooks other critical factors like throughput and delay. Since actions that save energy might not be optimal for maximizing throughput, Q-Learning tends to get stuck in the previously known optimal values. It over-prioritizes minimizing the energy without considering throughput and other values.

A few earlier methods have implemented Q-Learning and considered one or two QoS parameters in the reward function. Our proposed method also compares and evaluates PPO's throughput and channel utilization with two of these earlier methods in the existing work. For the purpose of evaluating the effectiveness of our proposed method against existing schemes, the results are compared with the system throughput and the channel utilization of the JPCRL and the DRLCA algorithms (Zhao et al., 2019; Pavan et al., 2024), respectively. Results also show a clear improvement in both the parameters for the proposed PPO method. For even more complex environments, such as larger channels with continuous state spaces, Deep Q-Learning or Double Deep Q-Learning (DDQN) would be required.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

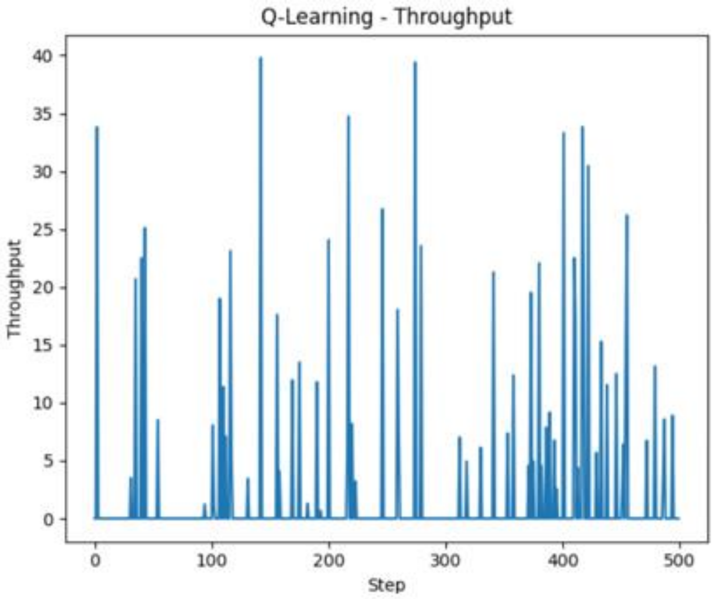

Figure 5. Throughput graph for Q-Learning.

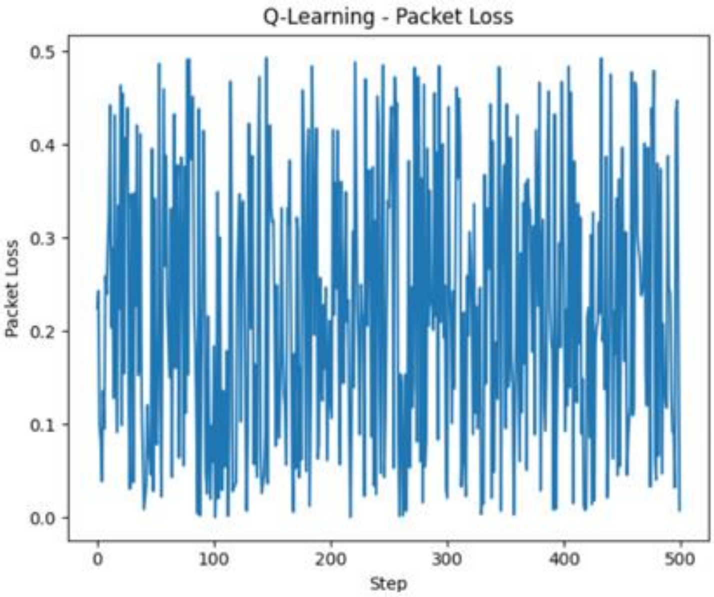

Figure 6. Packet Loss graph for Q-Learning.

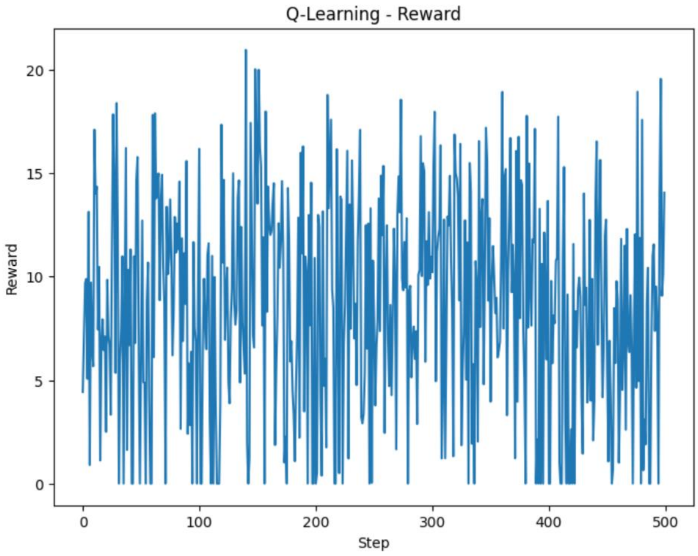

Figure 7. Reward graph for Q-Learning.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

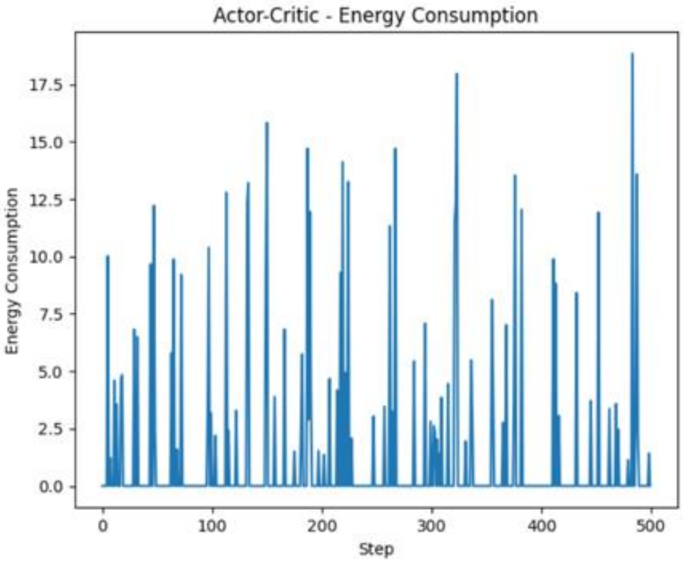

Figure 8. Energy Consumption graph for Actor-Critic.

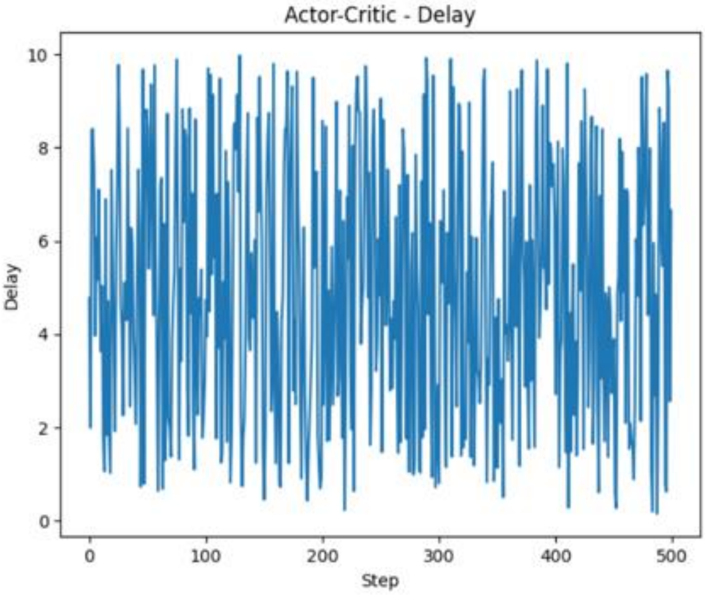

Figure 9. Delay graph for Actor-Critic.

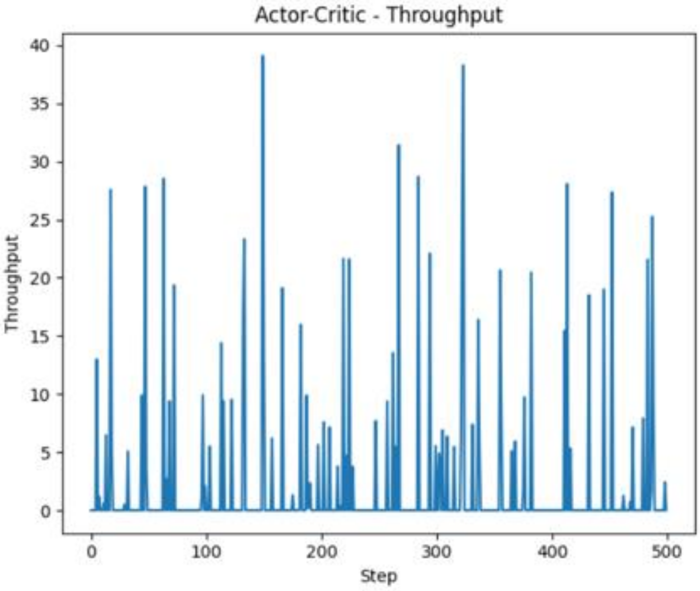

Figure 10. Throughtput graph for Actor-Critic.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

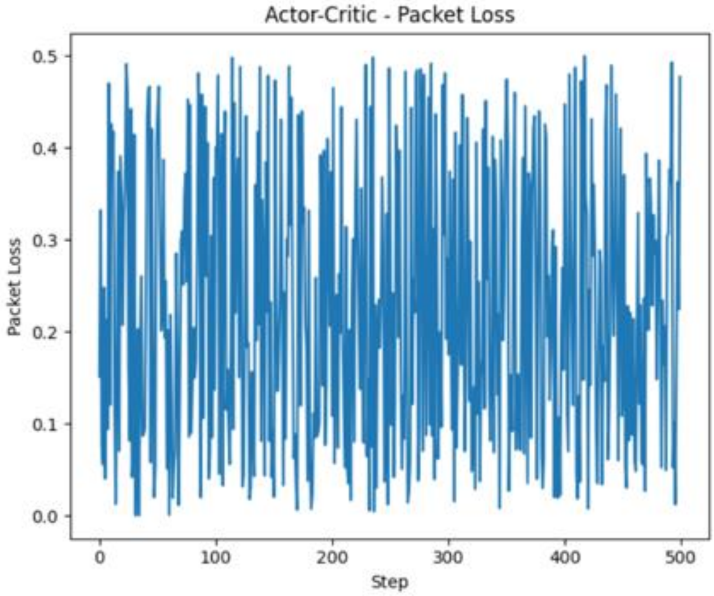

Figure 11. Packet Loss graph for Actor-Critic.

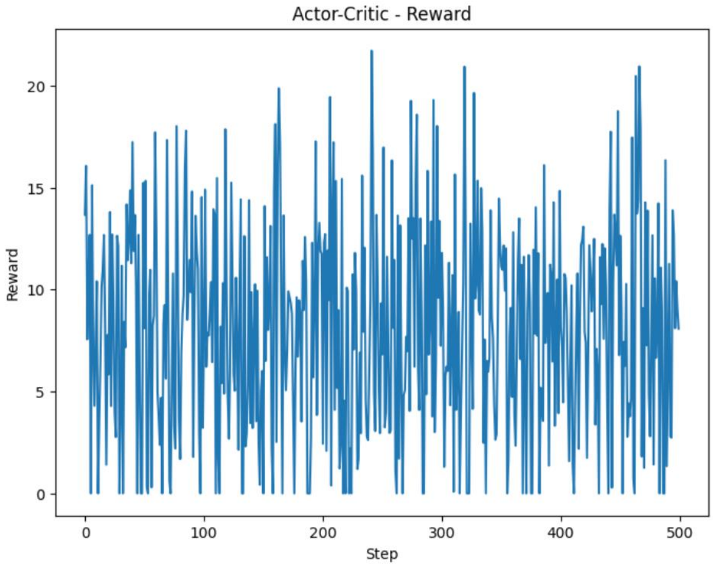

Figure 12. Reward graph for Actor-Critic.

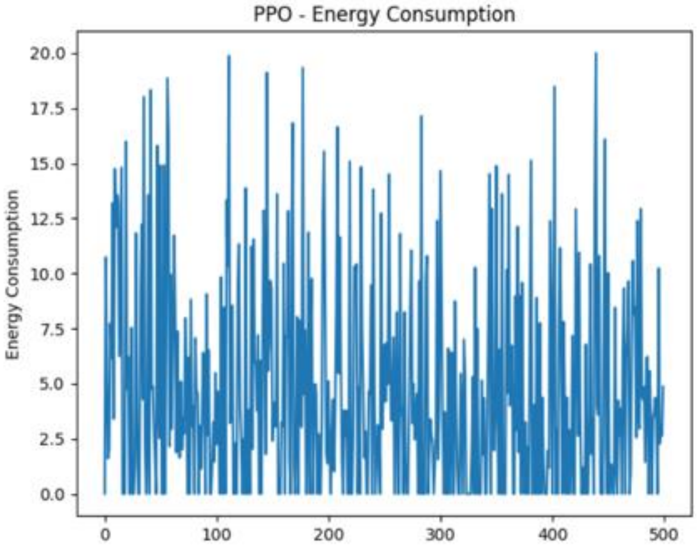

Figure 13. Energy Consumption graph for PPO

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

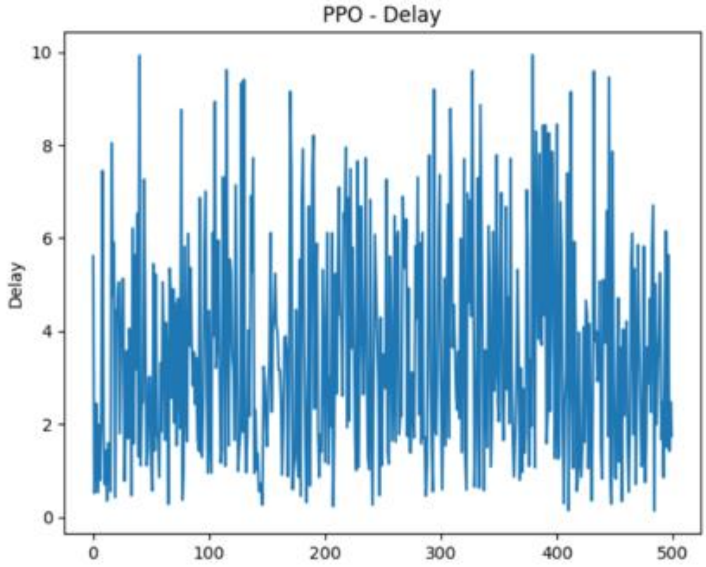

Figure 14. Delay graph for PPO

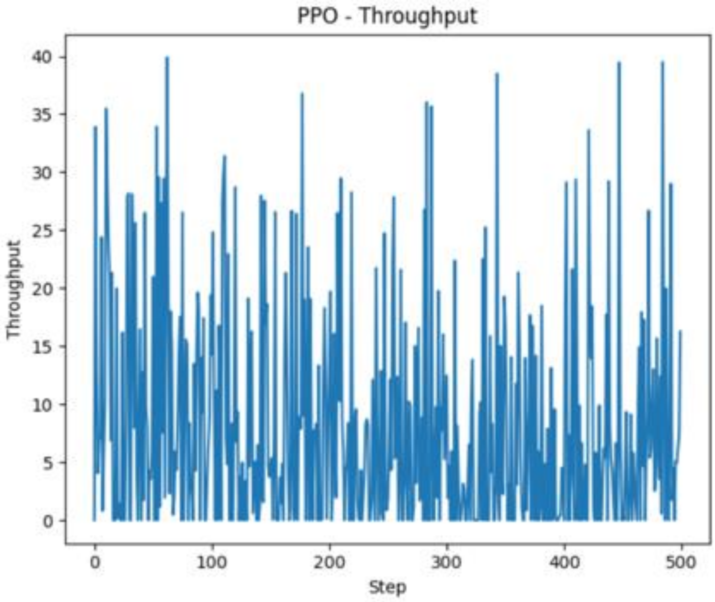

Figure 15. Throughput graph for PPO.

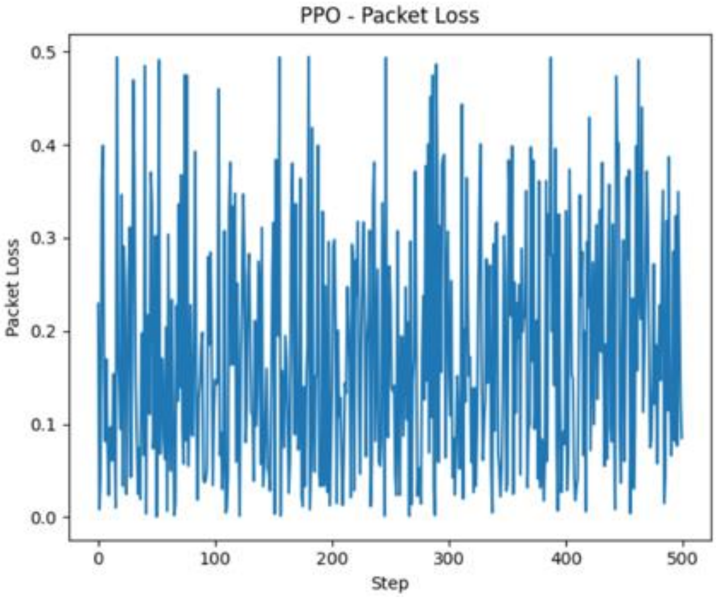

Figure 16. Packet Loss graph for PPO.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

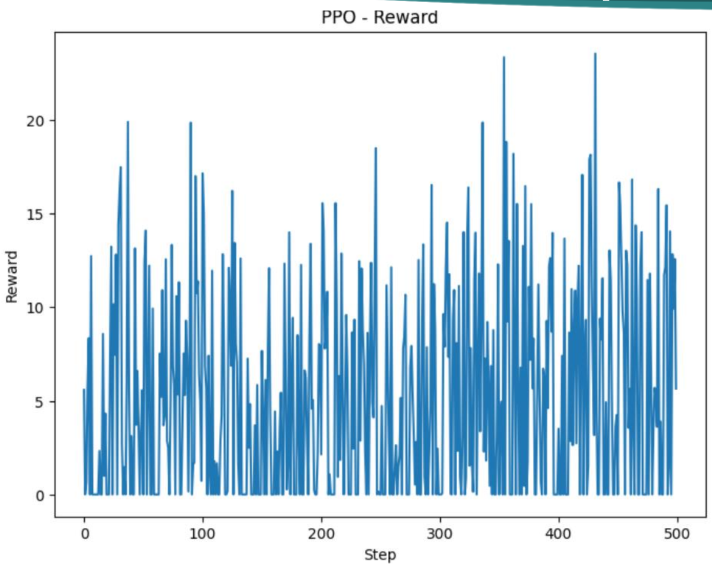

Figure 17. Reward graph for PPO.

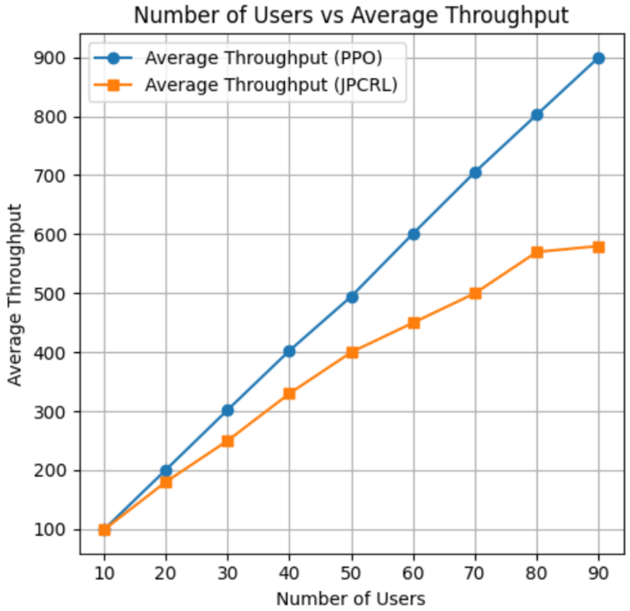

Figure 18. Throughput graph for PPO vs JPCRL.

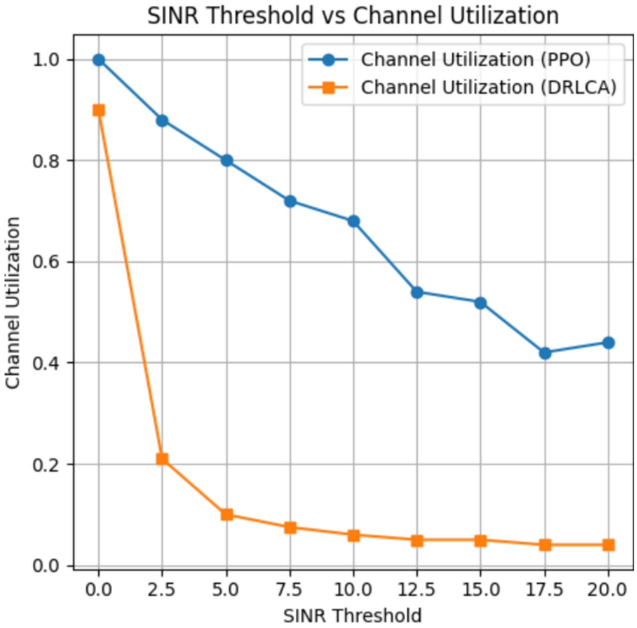

Figure 19. Channel Utilization graph for PPO vs DRLCA.

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

Table 1. Throughput Table for PPO vs JPCRL.

Number of Users	Average Throughput (PPO)	Average Throughput (JPCRL)
10.0	99.02	99
20.0	200.38	180
30.0	302.13	250
40.0	402.66	330
50.0	494.62	400
60.0	601.3	450
70.0	705.49	500
80.0	802.85	570
90.0	899.94	580

Table 2. Channel Utilization Table for PPO vs DRLCA.

SINR Threshold	Channel Utilization (PPO)	Channel Utilization (DRLCA)
0.0	1.0	0.9
2.5	0.88	0.21
5.0	0.8	0.1
7.5	0.72	0.075
10.0	0.68	0.06
12.5	0.54	0.05
15.0	0.52	0.05
17.5	0.42	0.04
20.0	0.44	0.04

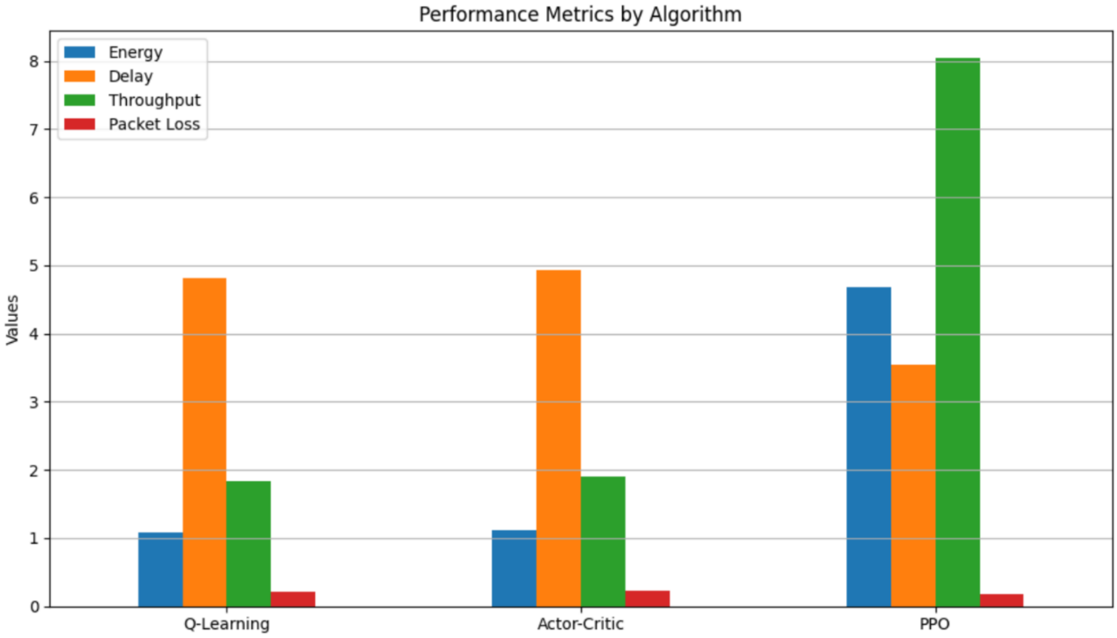

Figure 20. Performance metrics comparison for the three RL strategies.

Conclusion

In this work, we propose reinforcement learning strategies that can optimize energy consumption in some cases and throughput in other cases in a CRN. The QoS requirements guide the choice of strategy to use with CRNs for channel allocation. The effectiveness of a particular strategy for parameter optimization depends on the design of the reward function. In the case of a weighted reward function used here, the proper choice of weights would certainly impact the strategy, particularly the PPO, which can be used for optimizing the energy. Still, then the other parameters would be suboptimal. Put another way, PPO consumes more energy as it emphasizes finding a balance between all the performance metrics, not just energy efficiency, improving throughput, delay, and packet loss. Comparison of the proposed method with existing algorithms also confirms its effectiveness in terms of

DOI: https://doi.org/10.52756/ijerr.2024.v46.026

improved throughput and channel utilization. Since PPO algorithm performs exceedingly well in terms of throughput and substantially reduces the delay and packet loss, it is an ideal choice for applications such as video, imaging or M2M communications.

Conflict of interest

None

References

Azaly, N., Badran, E., Kheirallah, H., & Farag, H. (2020). Centralized Dynamic Channel Reservation Mechanism via SDN for CR Networks Spectrum Allocation. *IEEE Access, 8,* 192493-192505. https://doi.org/10.1109/ACCESS.2020.3032666

Chakraborty, T., & Misra, I. S. (2020). A novel three-phase target channel allocation scheme for multi-user Cognitive Radio Networks. *Computer Communications, 154,* 18–39. https://doi.org/10.1016/j.comcom.2020.02.026

Dey, S., & Misra, I. (2018). A Novel Content Aware Channel Allocation Scheme for Video Applications over CRN. *Wireless Personal Communications, 100*. https://doi.org/10.1007/s11277-018-5650-4

El-Toukhy, A., Tantawy, M., & Tarrad, I. (2016). QoS-driven channel allocation schemes based on secondary users' priority in cognitive radio networks. *International Journal of Wireless and Mobile Computing, 11*(1), 91. https://doi.org/10.1504/IJWMC.2016.080182

Hossen, M. A., & Yoo, S.J. (2019). Q-Learning Based Multi-Objective Clustering Algorithm for Cognitive Radio Ad Hoc Networks. *IEEE Access, 7*, 181959–181971. https://doi.org/10.1109/access.2019.2959313

Jang, S.J., Han, C.H., Lee, K.E., & Yoo, S.J. (2019). Reinforcement learning-based dynamic band and channel selection in cognitive radio ad-hoc networks. *EURASIP Journal on Wireless Communications and Networking, 2019*(1). https://doi.org/10.1186/s13638-019-1433-1

Kumar, A., Dutta, S., & Pranav, P. (2023). Prevention of VM Timing side-channel attack in a cloud environment using randomized timing approach in AES – 128. *Int. J. Exp. Res. Rev., 31*(Spl Volume), 131-140. https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.013

Madhuri, T. N. P., Rao, M. S., Santosh, P. S., Tejaswi, P., & Devendra, S. (2022). Data Communication Protocol using Elliptic Curve Cryptography for Wireless Body Area Network. *2022 6th International Conference on Computing Methodologies and Communication* (ICCMC), pp.133–139. https://doi.org/10.1109/iccmc53470.2022.9753898

Malik, S. A. (2020). Efficient channel allocation using matching theory for QoS provisioning in radio networks. *Sensors, 20*(7), 1872. https://doi.org/10.3390/s20071872

Nayak, D., Muduli, A., Hussain, M., Mirza, A., Gummadipudi, J., & Kumar, N. (2020). Channel allocation in cognitive radio networks using energy detection technique. *Materials Today: Proceedings, 33*. https://doi.org/10.1016/j.matpr.2020.06.491

Nayani, A. S. K., Sekhar, C., Rao, M. S., & Rao, K. V. (2021). Enhancing image resolution and denoising using autoencoder. *In Lecture Notes on Data Engineering and Communications Technologies*, pp. 649–659. https://doi.org/10.1007/978-981-15-8335-3_50

Pal, R., & Dahiya, S. (2022). Optimal Channel Selection algorithm for CRahNs. *Physical Communication, 54*, 101853. https://doi.org/10.1016/j.phycom.2022.101853

Pal, R., Gupta, N., Prakash, A., Tripathi, R., & Rodrigues, J. J. P. C. (2020). Deep reinforcement learning based optimal channel selection for cognitive radio vehicular ad-hoc network. *IET Communications, 14*(19), 3464–3471. Portico. https://doi.org/10.1049/iet-com.2020.0451

Pavan, M.N., Kumar, S., Nayak, G., & Narender, M. (2024). Deep Reinforcement Learning based channel allocation (DRLCA) in Cognitive Radio Networks. *Journal of Electrical Systems.*

Srivastava, A., Pal, R., Prakash, A., Tripathi, R., Gupta, N., & Alkhayyat, A. (2024). Optimal Channel Selection and Switching Using Q-Learning in Cognitive Radio Ad Hoc Networks. *IEEE Transactions on Consumer Electronics, 70*(3), 6314–6326. https://doi.org/10.1109/tce.2024.3413333

Tlouyamma, J., & Velempini, M. (2021). Channel Selection Algorithm Optimized for Improved Performance in Cognitive Radio Networks. Wireless Personal Communications.

Varshney, P., Singh, R. P., & Jain, R. K. (2024). Performance Analysis of Millimeter-Wave Propagation Characteristics for Various Channel Models in the Indoor Environment. *International Journal of Experimental Research and Review, 44*, 102–114. https://doi.org/10.52756/IJERR.2024.v44spl.009

Wang, Y.H., & Liao, S.L. (2018). Dynamic channel allocation scheme in CRN applied fuzzy-inference system. *Journal of Computers* (Taiwan). *29*, 141-155. http://dx.doi.org/10.3966/199115992018062903013.

Xu, Q., Li, S., Gaber, J., & Han, Y. (2024). Modelling Analysis of Channel Assembling in CRNs Based on Priority Scheduling Strategy with Reserved Queue. *Electronics*, *13*(15), 3051. https://doi.org/10.3390/electronics13153051

Ye, Z., Wang, Y., & Wan, P. (2020). Joint Channel Allocation and Power Control Based on Long Short-Term Memory Deep Q Network in Cognitive Radio Networks. *Complexity*, *2020*, 1-11. http://dx.doi.org/10.1155/2020/1628023

Zhang, M., Zhu, X., Jiang, H., Bian, T., & Yang, Y. (2023). A dynamic channel allocation protocol based on data traffic characterization for cognitive-radio wireless sensor networks. SSRN. http://dx.doi.org/10.2139/ssrn.4457362

Zhao, G., Li, Y., Xu, C., Han, Z., Xing, Y., & Yu, S. (2019). Joint Power Control and Channel Allocation for Interference Mitigation Based on Reinforcement Learning. *IEEE Access*, *7*, 177254–177265. https://doi.org/10.1109/access.2019.2937438

How to cite this Article:

Kalyana Chakravarthy Chilukuri, N Chaitanya Kumar, T. Vidhyavathi, Regidi Suneetha, V Sita Rama Prasad, Badugu Samatha and Mahanty Rashmita (2024). Proximal Policy Optimization for Efficient Channel Allocation with Quality of Service (QoS) in Cognitive Radio Networks. *International Journal of Experimental Research and Review, 46,* 326-341.

DOI : https://doi.org/10.52756/ijerr.2024.v46.026