Reinforcement Learning for Channel and Radio Allocations to Wireless Backhaul Links

Juliana Liman and Li-Hsing Yen
Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.
Email: ushipei.ee08@nycu.edu.tw, lhyen@nctu.edu.tw

***Abstract*—WMN's mesh access points (MAPs) are linked through a wireless backhaul network that consist of mesh points (MPs) that is equipped with multiple radios that use multiple non-overlapping channels in parallel. MPs will establish designated links that should satisfy both common channel constraint and interference constraint which are conflicting in nature. Yen and Dai proposed game-theoretic radio resources allocation in WMN which is better than centralized and greedy approach if only two radios are available at each node, but when there are more than two radios per node, centralized and greedy approach perform better. So, this study would like to utilize reinforcement learning to improve previous research so the approach is also effective if there are more than two radios available per node. This study attempts to maximize the number of operative designated links in the backhaul networks subject to common channel constraint and interference constraint. We use multi-agent deep Q-learning to tackle this problem. We conduct simulations to compare the proposed approach with game based approach. The results of our experiments show that the proposed deep Q-learning algorithm performs better than game-theoretic approach in dense network where there are more than two in each MP, while the game-theoretic approach performs better than our proposed DQL algorithm in sparse network.**

***Index Terms*—wireless mesh network, reinforcement learning, deep Q-learning, radio resource, multi agent reinforcement learning**

I. Introduction

WMN connects several nodes, extending the coverage area of wireless signal, minimizing dead zones and thus providing better Internet connectivity [1]. WMN's mesh access points (MAPs) are linked through a wireless backhaul network that consist of mesh points (MPs) that is equipped with multiple radios that use multiple non-overlapping channels in parallel. MPs will establish designated links that should satisfy both common channel constraint and interference constraint which are conflicting in nature.

This study attempts to maximize the number of operative designated links in the backhaul networks by allocating channels to radios/links such that the common channel constraint is met while minimizing the co-channel interference of links. This study is limited to co-channel interference and interference from adjacent channels are not considered. Our approach is based on reinforcement learning, which is a type of machine learning where the learner or agent will be able to discover which action yields the highest rewards by trial-and-error search and delayed rewards [2]. We conduct experiments to compare our proposed reinforcement learning approach with the game theoretic approach. The results of our experiments show that the proposed deep Q-learning algorithm performs better than game-theoretic approach in dense network, while the game-theoretic approach performs better than our proposed DQL algorithm in sparse network where there are two radios in each MP.

The remainder of the paper is organized as follows. Section II reviews related works. Section III introduces our system model and problem formulation. Section IV details our system implementation. Our performance evaluations are presented in Section V. Finally, Section VI concludes this work and discusses several potential future works.

II. Related Work

A. Channel Allocation with Reinforcement Learning

The most commonly used reinforcement learning method for channel allocation is Q-Learning [3] [4] [5] [6]. Q-learning has an advantage where agents need no prior knowledge about their actions' effect on the environment [5], but Q-learning will take too much memory for the Q-table if the environment is complex or there are numerous agents, so some works [7] use Deep Q-learning instead. Deep Q-learning is better than Q-learning if the environment is constantly changing or more complex, as deep Q-learning will use loss function to predict Q-value instead of using Q-table, making it more memory efficient. However, deep-Q-learning requires more training time and computation resources than Q-learning, so in a small and simple environment, Q-learning is enough. Some works [8] [9] use Double Deep Q-learning (DDQN) to avoid overestimation [10].

B. Channel Allocation with Game Theoretic Approach

Yen and Dai [11] did similar research for efficient radio resources allocation in WMN using game-theoretic approach. They guarantee the common-channel constraint by using Pigeonhole Principle to limit the number of channels available for the radios to select. They attempt to minimize the co-channel interference by using two-stage non-cooperative game-theoretic approach with the consideration of physical-layer interference. The first stage assigns channels to radios while the second stage assigns the radio-channel pairs to links. They define utility to ensure the quality of their proposed approach to avoid co-channel interference. If the radios in the same MP use the same channel, then the utility of that assignment will be much lower than if the radios in the same MP use different channels. This approach results in more operative links than other similar approaches such as centralized and greedy approach if only two radios are available at each node, but when there are more than two radios per node, centralized and greedy approach perform better.

This study will to compare the result of the game-theoretic approach by Yen and Dai [11] with the proposed approach based on reinforcement learning in solving the channel allocation problem in the WMN environment.

III. System Model and Problem Formulation

The WMN environment for the reinforcement learning follow the same settings proposed by Yen and Dai [11].

A. Channel Allocation to Radios

Let $N = \{1, 2, ..., n\}$ denote the set of MPs that are deployed in the backhaul network, with r_i radio interfaces available in every MP. Let $P = \{p_1, p_2, ..., p_m\}$ denote the set of radios that will act as agents in reinforcement learning. The placement of the MPs can be represented by the MPs' coordinates, will be static, so the distance between any pair of MPs remain the same. The agents will choose a channel to be used for it. The initial channel assignment for the radios is arbitrary. The agents (i.e., radios) can change their channels throughout the entire learning phase. We set the learning termination condition to be if the step already reaches the maximum step or if there is no any change in the joint reward. To determine an agent's utility when it selects a channel, we will use the utility function proposed by Yen and Dai [11]. When agent p_i chooses channel c_i, the utility is defined as

$$u_i(c_i) = \sum_{j \neq i} f(c_i, c_j), \tag{1}$$

where $f(c_i, c_j)$ is the intensity of the interference experienced by p_i and p_j when they choose c_i and c_j, respectively. More specifically,

$$f(c_i, c_j) = \begin{cases} 0 & \text{if } c_i \neq c_j \\ -\frac{1}{d_{i,j}^{\alpha}} & \text{if } c_i = c_j \\ -\beta & \text{if } c_i = c_j \text{ and } MP(p_i) = MP(p_j), \end{cases} \tag{2}$$

where $d_{i,j}$ is the physical distance between p_i and p_j, α is path loss exponent constant ranging from 2 to 4, and β is a constant that is much larger than $\frac{1}{d_{i,j}^{\alpha}}$, which represents the severe self-interference that will happen if two radios in the same MP choose the same channel.

B. Channel Allocation to Links

After the assignment of channels to radios, it is possible that two or more radio-channel pairs can be assigned to a designated link. For this reason, we will assign the radio-channel pairs selected by agents to the designated links such that each link is assigned a specific radio-channel pair. In case like this, we need to choose one radio-channel pair for the link that will cause the lowest interference. We use the second stage of the game based approach proposed by Yen and Dai [11] to allocate radio-channel pairs to links.

C. Operative Link Ratio

Designated links that satisfy both common channel constraint and interference constraint are called operative links. Common channel constraint means operative links must have one common channel for communication and interference constraint means the Signal-to-interference-plus-noise ratio (SINR) on both nodes of the link exceed a threshold. In this study, we sill set the threshold to be 1 dB. Operative Link Ratio (OLR) is defined as the total number of operative links divided by the total number of designated links [12].

IV. Proposed Mechanism

A. Cooperative Multi-Agent Reinforcement Learning

Our agents are radios in WMN. All agents will learn at the same time, but all of agents must cooperatively maximize a single joint reward, so this study is an application of cooperative multi-agent reinforcement learning with single joint reward [13]. The state, reward and action in the proposed algorithm are defined as follows:

1) State: Our state space consists of the vector of channels used by each radio (i.e., the vector of all actions), which is defined as $\mathbf{S} = [c_1, c_2, ..., c_m]$, where c_i is the channel radio i chooses in a certain time.

2) Action: The action space for every agent consists of the channels that the agent can choose, which is defined as $A = \{c_1, ..., c_k\}$, then the action vector is defined as $\mathbf{A} = [a_1, a_2, ..., a_m]$. An agent can only choose a channel in a time.

3) Rewards: The reward ranges from -30 to +30. Let us define designated links that satisfy common channel constraint as $D' = \{(u_1, v_1), (u_2, v_2), ..., (u_m, v_m)\}$. If the number of D' increases after all agents choose new channels, then the agents will receive a joint reward 30, but if the number of D' decreases instead, then the agents will receive joint penalty -30 instead. If the number of D' remains unchanged, then the joint rewards is determined by the summation of the individual rewards of the agents $\sum r_t^n$ based on their utilities.

The agent will receive +5 points if the new channel the agent chooses yield a significant utility improvement. If the utility improvement is not significant, then the agent will receive +1 point. If the utility decreases instead, then agent will get 5 points penalty if the decrease is significant and 1 point penalty if the decrease is not significant. If the utility remains unchanged, then agent will not receive any reward or penalty. The rewards are defined as:

$$r_t = \begin{cases} 30 & \text{if the number of } D' \text{ increased} \\ -30 & \text{if the number of } D' \text{ decreased} \\ \sum r_t^n & \text{if if the number of } D' \text{ stays the same,} \end{cases} \tag{3}$$

with the reward for individual agent r_t^n defined as:

$$r_t^n = \begin{cases} 5 & \text{if } u_t^n - u_{t-1}^n \geq \beta \\ 1 & \text{if } u_t^n > u_{t-1}^n \text{ and } u_t^n - u_{t-1}^n < \beta \\ 0 & \text{if } u_t^n = u_{t-1}^n \\ -1 & \text{if } u_t^n < u_{t-1}^n \text{ and } u_t^n - u_{t-1}^n > -\beta \\ -5 & \text{if } u_t^n < u_{t-1}^n \text{ and } u_t^n - u_{t-1}^n \leq -\beta \end{cases} \tag{4}$$

V. Experiment Results

We will evaluate the performance of our proposed approach using three types of WMS network environment: sparse network, regular network, and dense network. We will compare the effect of different learning rates in our study. We will compare the proposed approach with another approach, that is the game-theoretic approach by Yen and Dai [11]. The basic experiment settings are listed in Table I. We calculate the maximum possible OLR in a network with exhaustive search. We measure the time-averaged regrets as performance metric, which is defined as

$$\frac{1}{T} * \sum_{i=1}^{T} (\text{Maximum possible OLR} - \text{OLR at episode } i) \tag{5}$$

TABLE I: Basic Experiment Settings

Parameter	Value
CPU	12th Gen Intel(R) Core(TM) i7 - 12700KF
GPU	NVIDIA GeForce RTX 3070
RAM	32 GB, 4400 MHz
Discount rate	0.95
Learning rate	0.001, 0.1, 0.7
Total episodes	300
Maximum step per episode	100
Nodes	10

A. Sparse Network

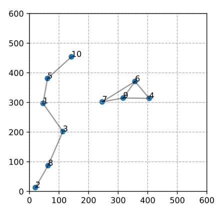


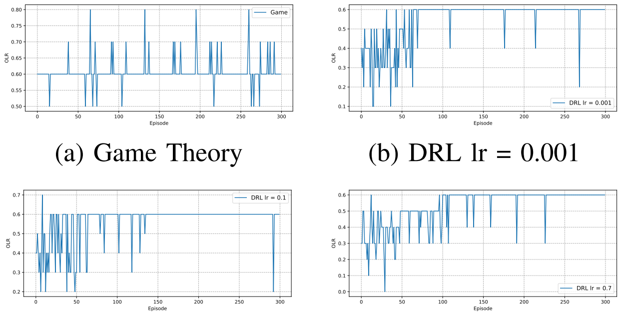
Fig. 1: Sparse Network Topology

TABLE II: Sparse Network Experiment Settings

Parameter	Value
Number of designated links	10
Radios per node	2
Transmission range	150 m
Number of channels	6

Figure 1 is the topology for sparse network. The settings for this topology is listed in Table II.

Figure 3 is the time-averaged regrets of sparse topology from game based approach and Deep Q-learning with learning



(a) Game Theory (b) DRL lr = 0.001 (c) DRL lr = 0.1 (d) DRL lr = 0.7

Fig. 2: The OLR comparison between game based approach and DRL approach in sparse network

rate 0.001, 0.1, and 0.7. In this topology, game based approach performs better than deep Q-learning approach because game base approach have lower time-averaged regrets. As we can see from Figure 2, the lower time-averaged regrets of game-theoretic approach is because game-theoretic approach can achieve higher OLR, while the proposed DRL algorithm approach converged at 0.6. However, even though the initial time-averaged regrets of proposed DRL approach is a lot higher than game-theoretic approach, the proposed DRL algorithm's time-averaged regrets is steadily going down. This is because at the beginning, the proposed DRL approach is going through trial and error before finally converging to a certain value.

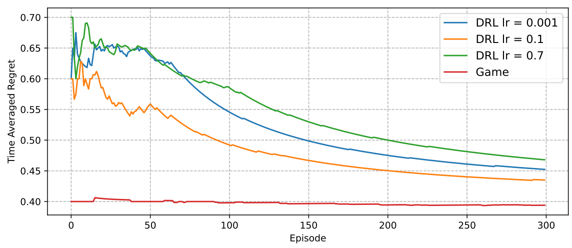


Fig. 3: The time-averaged regrets of game based approach and deep Q-learning approach with learning rate = 0.001, 0.1, and 0.7 in sparse network

B. Dense Network

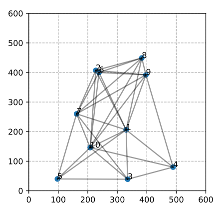


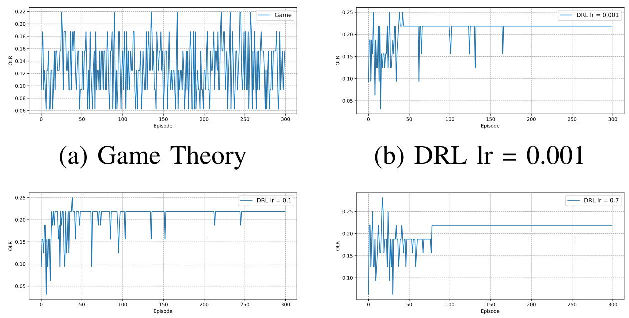
Fig. 4: Dense Network Topology

Figure 4 is the topology for dense network in this study. The settings for this topology is listed in Table III.

Figure 6 shows the time-averaged regrets of game based approach and deep Q-learning in dense network with 3 radios

TABLE III: Dense Network Experiment Settings

Parameter	Value
Number of designated links	32
Radios per node	3
Transmission range	350 m
Number of channels	12



(a) Game Theory (b) DRL lr = 0.001

(c) DRL lr = 0.1 (d) DRL lr = 0.7

Fig. 5: The OLR comparison between game based approach and DRL approach in dense network with 3 radios per node

per node. In this setting, the proposed DQL algorithm has lower time-averaged regrets. In dense network, both game-theoretic approach and proposed DQL algorithm have a hard time to reach high OLR, but as we can see in Figure 5, because game theoretic approach doesn't converge, it reached low OLR more often than proposed DQL algorithm. In this setting, all learning rates have similar time-averaged regrets.

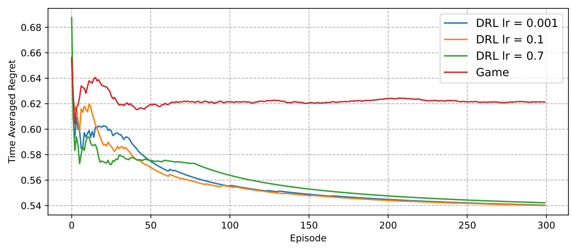


Fig. 6: The time-averaged regrets of game based approach and deep Q-learning approach with learning rate = 0.001, 0.1, and 0.7 in dense network with 3 radios per node

VI. Conclusions

In this work, we model the channel allocation problem in wireless mesh network to maximize the operative link ratio (OLR) using reinforcement learning. We use multi-agent deep Q-learning to solve this problem. The radios can work cooperatively to maximize the utilities gained when choosing the channel so the network can maximize its operative link ratio (OLR). We use reinforcement learning to allocate channels to radios, then use game theory to allocate the radio-channel pairs to links. After assigning the radio-channel pairs to links, we calculate the OLR to determine how good our proposed approach is.

We conducted simulations determine how effective our proposed approach is. We conducted the simulations with 2 different environments, sparse network and dense network. We compare the results of the simulations of these 3 environments with the result of the game based approach in the same environments.

The result of our experiments show that the proposed deep Q-learning algorithm performs better than game-theoretic approach in dense network, while game-theoretic approach performs better than our proposed DQL algorithm in sparse network because the number of channels is of no problem in sparse network, but it's critical in dense network as game theoretic approach in dense network does not have enough channels for allocation due to the Pigeonhole principle.

This proposed approach still has room for improvements. The time needed to run the algorithm is too long and need considerable amount of computing power, so more efficient algorithm like double deep Q-learning might be better. This proposed approach also don't have any dynamics because the radios are stationary and the initial channel assignments are arbitrary, so it will improve the algorithm flexibility if the future works incorporate more dynamics.

References

[1] J. Shin, R. Kumar, Y. Shin, and T. F. L. Porta, "Multi-hop wireless relay networks of mesh clients," in *2008 IEEE Wireless Communications and Networking Conference*, 2008, pp. 2717–2722.

[2] R. S. Sutton and A. G. Barto, *Introduction*, ser. Reinforcement Learning, second edition: An Introduction. MITP, 2018.

[3] J. Nie and S. Haykin, "A dynamic channel assignment policy through q-learning," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, p. 2531–2544, 1999.

[4] S.-M. Senouci and G. Pujoile, "Dynamic channel assignment in cellular networks: a reinforcement learning solution," in *10th International Conference on Telecommunications (ICT)*, 2003, pp. 302–309.

[5] A. E. Araqi and B. Mahboobi, "Joint channel assignment and multicast routing in multi-channel multi-radio wireless mesh networks based on q-learning," in *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, 2019, pp. 1–6.

[6] H. Jiang, H. He, L. Liu, and Y. Yi, "Q-learning for non-cooperative channel access game of cognitive radio networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–7.

[7] Y. P. Wang, K. X. Zheng, D. X. Tian, X. T. Duan, and J. S. Zhou, "Cooperative channel assignment for vanets based on multiagent reinforcement learning," *Frontiers of Information Technology Electronic Engineering*, vol. 21, no. 7, p. 1047–1058, 2020.

[8] K. Nakashima, S. Kamiya, K. Ohtsu, K. Yamamoto, T. Nishio, and M. Morikura, "Deep reinforcement learning-based channel allocation for wireless lans with graph convolutional networks," *IEEE Access*, vol. 8, pp. 31 823 – 31 834, 2020.

[9] H. V. Vu, Z. Liu, D. H. N. Nguyen, R. Morawski, and T. Le-Ngoc, "Multi-agent reinforcement learning for joint channel assignment and power allocation in platoon-based c-v2x systems," 2020. [Online]. Available: https://arxiv.org/abs/2011.04555

[10] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," *CoRR*, vol. abs/1509.06461, 2015. [Online]. Available: http://arxiv.org/abs/1509.06461

[11] L.-H. Yen and Y.-K. Dai, "A two-stage game for allocating channels and radios to links in wireless backhaul networks," *Wireless Networks*, vol. 21, no. 8, p. 2531–2544, 2015.

[12] L.-H. Yen and K.-W. Huang, "Link-preserving interference-minimisation channel assignment in multi-radio wireless mesh networks," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 18, no. 4, pp. 222–233, 2015.

[13] K. Tuyls, G. Weiss, and K. Ciosek, "Multiagent learning: basics, challenges, and prospects," *AI Magazine*, vol. 33, no. 3, p. 41–52, 2014.