Learning-Based Algorithms for Channel Allocations in Wireless Mesh Network

Chien-Liang Kuo, Jin-Wei Kuo, Xuan-Zhe Chen, and Li-Hsing Yen
Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu, Taiwan.
Email: {sonicokuo.cs08, weiiiii77777.cs08, xuanzhe.cs08, lhyen}@nycu.edu.tw

***Abstract*—Many studies have been devoted to channel allocation for backhaul links in wireless mesh networks. Among them, a game-theoretic approach proposed by Yen and Dai is promising for the ability to self-stabilize to a valid solution in a decentralized manner. However, game-based solutions are generally not optimal. Furthermore, Yen and Dai's approach did not fully utilize all available channels, wasting scarce bandwidth resource. In this paper, we propose two learning-based approaches to enhance the prior work. One uses Spatial Adaptive Play (SAP) for agents to learn best probability distributions on their possible channel selections. The other based on multi-agent reinforcement learning (MARL) algorithm allows each agent to find out its best selection over time. Simulation results reveal that the proposed approaches do improve the game-based solutions in terms of the number of operative links after channel allocations.**

I. Introduction

An IEEE 802.11 access point (AP) provides access service to clients within its signal coverage. A way to extend the service coverage is to deploy multiple APs dispersed across a region and connect these APs with each other to switch and forward client's data traffic, forming a wireless mesh network (WMN). A WMN provides two types of wireless links. One called access links are for client's connections to respective APs. The other is for APs to connect to each other or connect to a relay node for data forwarding. Theses links are called *backhaul links*. The operations of backhaul links demand *channel allocation* which configures a channel for each backhaul link. The channel allocation problem in WMN is to perform channel allocation that maximizes the number of operative backhaul links.

Whether a link is operative after channel allocation depends on several factors. First, devices at both ends of a wireless link should operate on the same channel. Because nowadays WMN devices (referred to as *nodes* hereafter) are equipped with multiple transceivers (i.e., radios), this condition implies that two nodes of a link should have at least one radio that is tuned to a common channel. We refer to this condition as the *common channel constraint*. Second, both radios of a link should experience sufficiently low co-channel interference, which is referred to as the *interference constraint*. The problem to maximize operative links is also subject to limited supply of resource. For example, we have a limited number of channels (channel constraint) for allocation and a node may have fewer radios than the number of backhaul links to build for this node (radio constraint).

Many approaches to channel allocation in WMN have been proposed. These approaches differ in their interference models, objectives, and methods. This study is a follow-up to the work by Yen and Dai [1]. They proposed a game-theoretic approach which models each radio as an autonomous agent that could independently select its channel. As agents do not coordinate to meet the common channel constraint, it is possible that two radios of a link do not operate on the same channel. To preclude this possibility, their approach applies the Pigeonhole Principle to confine the set of channels that could be selected by each agent. As a way to minimize co-channel interference, this approach defines a utility function for each agent which favors a channel with the minimal possible co-channel interference. The approach stochastically converges to a solution by agent's unilateral best-response or better-response dynamics.

The prior work [1] leaves some space for improvement. First, the results derived by best- or better-response dynamics are generally not optimal. We may devise another approach to enhance the results. Second, the constraint by the Pigeonhole Principle implies that the approach does not fully utilize all available channels. We may devise another approach that fully utilize all available channels while still meeting the common channel constraint.

In this paper, we propose two approaches to channel allocations in WMN. One uses Spatial Adaptive Play (SAP) [2] for agents to learn probability distributions on their possible channel selections so as to derive better game results on the basis of Yen and Dai's work [1]. The other uses Q-learning as a multi-agent reinforcement learning (MARL) algorithm which allows each agent to find out its best selection over time. The MARL-based approach can also fully utilize all available channels. We conducted simulations to investigate the performance of the proposed approaches. We then compare through simulations the proposed approaches with Yen and Dai's work [1] in terms of operative link ratio (OLR) [3]. The results reveal that the proposed approaches can outperform their counterparts.

The remainder of the paper is organized as follows. Sec. II presents background information and reviews related works. Sec. III elaborates the proposed approaches. Sec. IV shows our simulation results. The last session concludes the paper.

II. Background and Related Work

We assume that each node is equipped with one or more radios. Each radio is to be statically configured with a channel. If a node A has to build a backhual link with another node B (such a link is a *designated link*), A and B must have at least one of their radios tuned to a common channel to communicate. Such links are *committed links* [3]. However, the number of available channels for allocations is limited. This limitation forces some backhaul links to reuse a channel, which may cause co-channel interference among links. If a link experiences severe interference, even a committed link may become inoperative [3].

In the literature, protocol model and physical model are two ways to estimate interference [4]. The protocol model is simpler but might have a gap from reality [5]. Therefore, we adopt the physical model and use signal-to-interference-plus-noise ratio (SINR) as a gauge of the impact of interference. The same model was also adopted by other studies [1], [3].

Not all studies considered the common channel constraint. Some researchers just focused on interference minimization and ignored the common channel constraint [6], [7]. A simple way to meet the constraint is to allocate the i-th channel to the i-th radio in every node for every i [8]. This straightforward strategy performs poorly in networks with complex topology. Some researchers used the Pigeonhole Principle to confine the range of channels for allocations [1], [3] so the result definitely conforms to the common channel constraint. Another way to trivially meet the common channel constraint is to allocate channels to links [9], [10] rather than radios, but each node needs to figure out how to meet the radio constraint.

Channel allocations have been modeled as non-cooperative games, where nodes or links are agents who compete with each other in channel usages for their own interests [9], [1], [11]. Even though the agents are self-interested, a system-wide goal can still be achieved through a well-designed utility function that motivates agents to move toward the system goal. Yen and Dai [1] proposed a two-stage game for channel allocation. The first stage is a non-cooperative game where radios act as agents. When the first stage ends, each radio is assigned a particular channel. Then, in the second-stage game, each link acting as an self-interested agent selects its radio-channel pair.

Before channel allocation, they use the Pigeonhole Principle to confine the set of channels for allocations as a way to meet the common channel constraint. Consequently, the number of channels available for allocations can be significantly reduced, wasting scarce bandwidth resource.

Another downside of [1] is the way to derive the game result. In [1], agents use either better or best response policy to unilaterally make their decisions. Such a policy can lead to some suboptimal game result. There exist some learning-based approaches that could lead to better game results.

III. The Proposed Approach

We propose two learning-based approaches as enhancements to the work by Yen and Dai [1]. One is based on SAP and the other is based on MARL. These approaches are detailed in this section.

A. SAP-Based Approach

Agents in the first approach uses SAP for channel selections. With SAP, agent's selections are no longer pure strategies but a probability distribution on the set of pure strategies.

Let $\{p_1, p_2, \cdots, p_m\}$ be the set of agents, one for each radio. Let S_i denote the pure strategy set of agent p_i, which is the set of all channels available to p_i. A strategy profile is an m-tuple $C = (c_1, c_2, \cdots, c_m)$, where $c_i \in S_i$ represents agent p_i's channel selection. C can also be expressed as (c_i, C_{-i}). For a strategy profile C, the utility of each agent p_i is $u_i(c_i, C_{-i}) = -\sum_{j \neq i} f(c_i, c_j)$, where $f(c_i, c_j)$ is a function that estimates the cost of co-channel interference between channels c_i and c_j, respectively.

In [1], agents take either best response or better response to make decisions. For best response, agent p_i checks if its current selection is in $\mathit{BR}_i(C_{-i})$ defined as

$$\mathit{BR}_i(C_{-i}) = \{c_i \in S_i \mid \forall c_i' \in S_i, u_i(c_i, C_{-i}) \geq u_i(c_i', C_{-i})\}. \tag{1}$$

If it is not, p_i selects an arbitrary channel from BR_i. For better response, p_i can select an arbitrary channel from S_i as long as the new channel gives p_i a higher utility.

Agents do not follow a specific order to make their decisions, so the game result is non-deterministic. It can be proved that this channel allocation game is an exact potential game (EPG) [12] with potential function $\phi(C) = \frac{1}{2}\sum_i u_i(C)$. For this reason, both best response and better response dynamics eventually lead to an Nash equilibrium and thus guarantee stability.

SAP helps agents learn their best strategies in a stochastic way. Like the better and best response policies, SAP favors channels with high utilities. However, SAP also gives agents a small probability to select a channel with low utility (which is impossible with the better/best response policy). Therefore, SAP has the potential to explore more game results [13]. In SAP, each agent p_i calculates $P_i(c_i \mid C_{-i})$, the probability for p_i to select channel $c_i \in S_i$, as follows.

$$P_i(c_i \mid C_{-i}) = \frac{\exp\left[\beta u_i(c_i, C_{-i})\right]}{\sum_{c_i' \in S_i} \exp\left[\beta u_i(c_i', C_{-i})\right]}, \tag{2}$$

where $u_i(c_i, C_{-i})$ is p_i's utility and β is the *temperature* for SAP ($0 < \beta < \infty$). It has been proved [2] that in a finite EPG with potential function ϕ, SAP has a unique stationary distribution of strategy profiles.

The temperature β in SAP is critical for controlling agent's behaviors. If β approaches 0, agents tend to select channels at random, which explores more possible game results. On the other hand, if β is large, agents behave like they use the best response policy, which is to exploit a channel with the highest utility value.

To strike a balance between exploration and exploitation, a dynamic setting of β might be a good idea. For example, if $\beta = \sqrt{t}$, where t is the total number of movements till now, agents explore more in the beginning of a game and tend to exploit channels with higher utility later on. We will study several possible ways to dynamically adjust β in the next section.

B. MARL-Based Approach

The SAP-based approach is still constrained by the Pigeonhole Principle. In contrast, the second approach based on cooperative MARL [14] utilizes all available channels. We use Q-learning to train a learning model for each agent. Each agent takes an action (selecting a particular channel) and then receives a reward from the environment. The reward reflects the quality of the action with the consideration of the joint actions of all other agents. The reward is used to update the agent's leaning model so the agent can gradually improve the quality of its action-taking process.

The proposed approach is decentralized in the sense that each agent p_i maintains its own Q-table Q_i. A Q-table is a two-dimensional array which keeps a Q-value for each possible state-action pair. A state in our design corresponds to a particular strategy profile, i.e., an m-tuple of actions $C = (c_1, c_2, \cdots, c_m)$, where $c_i \in S_i$ for all i. All Q-values are initialized to 0's.

For an agent p_i to act, we use epsilon-greedy selection to balance exploration and exploitation. More specifically, p_i has a probability of $1 - \epsilon$ to exploit its Q-table Q_i by selecting an action with the highest Q-value in Q_i (w.r.t. the current state C). The agent also has a probability of ϵ to explore the state space by selecting one action at random. In that case, the agent needs to update its Q-value based on feedback from the environment. Suppose that agent p_i causes a state transition from C to C' by taking a random action $a \in S_i$. It updates the Q-value associated with (C, a) as follows.

$$Q_i(C,a) \leftarrow (1-\alpha)\cdot Q_i(C,a) + \alpha[R_i(C,a) + \gamma \max_{a'} Q_i(C',a')], \quad (3)$$

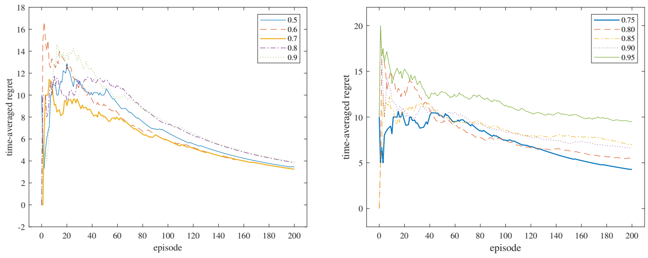
where α is the learning rate ($0 < \alpha \leq 1$), $R_i(C, a)$ is a function that returns the reward of p_i's action a in state C, and γ is the discount factor ($0 \leq \gamma \leq 1$).

We design the reward function $R_i(C, a)$ such that it awards or punishes p_i depending on how the action a in state C changes the number of committed links. Therefore, we can hopefully maximize the number of committed links without the Pigeonhole Principle. Even if the action does not increase the number of committed links, we still award the action if it causes a utility gain. The action is severely punished if it causes self-interference.

To simulate the sequential movements of agents in the original game, only one agent at a time is admitted to act. The admission is enforced by a random action selector. We also rule that no agent can act twice successively. A training episode ends when every agent has acted at least once and no agent has ever changed its action in the last m movements. The whole training process ends after a fixed number of episodes.

TABLE I: Performance results for game-based and SAP-based approaches (t denotes the number of movements)

Approach	Avg. Num. of Movements	Final Utility
Best Response	19.2	-0.1390
Better Response	32.0	-0.1385
SAP ($\beta = \log(t+1)$)	10000.0	-0.2135
SAP ($\beta = \sqrt{t}$)	10000.0	-0.1603
SAP ($\beta = t$)	2692.5	-0.1385
SAP ($\beta = t^2$)	75.3	-0.1384



(a) $\gamma = 0.95$ and $\epsilon_{\text{max}} = 1$

(b) $\alpha = 0.7$ and $\epsilon_{\text{max}} = 1$

Fig. 1: The time-averaged regret of the proposed MARL-based approach with different (a) α values and (b) γ values.

IV. Numerical Results

We conducted simulations for performance comparisons among several approaches, including the one in [1], the proposed SAP-based approach and the proposed MARL-based approach. All these approaches allocate channels to radios. We used the approach proposed in [1] to allocate radio-channel pairs to designated links.

A. SAP-Based and Game-Based Approaches

We first compare the game-based approaches [1] with the SAP-based approach with different fixed temperature values (β). We used a topology consisting of 14 dispersed nodes each equipped with 5 radios. The initial channels of radios were randomly allocated. We played the game for 10 times, each with at most 10000 movements. The results show that SAP with fixed β values performed worse than Best Response and Better Response.

We then investigated the performance of SAP with dynamic settings of β values. Here the topology consists of 10 nodes, each was equipped with four radios. The simulation results are listed in Table I. Observe that SAPs with low growth rates of β ($\beta = \log(t + 1)$ and $\beta = \sqrt{t}$) did not converge before the maximum movements. These results are not satisfying. In contrast, SAPs with high growth rates of β did converge earlier and had slightly higher utility values than Best Response and Better Response.

B. MARL-based Approach

We next tested the performance of the proposed MARL-based approach. The WMN consists of five nodes, each equipped with two radios. Three channels in total were available for allocations.

The key performance index for reinforcement learning is regret, which is the difference between the highest possible utility value and the actual utility value after each movement. To calculate regret, we did exhaustive search to find out the highest possible utility value for this topology. To study

TABLE II: Performance results using the same topology in Sec. IV-B

Approach	Avg. Number of Movements	Final Utility	OLR
Best Response	6.61	-0.0188	0.500
Better Response	7.58	-0.0188	0.500
SAP ($\beta = t^2$)	39.97	-0.0190	0.52625
MARL (3 channels)	150036.71	-0.0189	0.51125
MARL (4 channels)	60113.29	-0.0183	0.53375

TABLE III: Performance results using a new topology

Approach	Avg. Number of Movements	Final Utility	OLR
Best Response	4.14	-0.0355	0.500
Better Response	5.32	-0.0355	0.500
SAP ($\beta = t^2$)	23.72	-0.0360	0.505
MARL (3 channels)	20026.30	-0.0359	0.516
MARL (4 channels)	116.00	-0.0361	0.533

the dynamics of learning over time, we took time-averaged regret (the accumulated regret value divided by the number of movements taken) as our main performance metrics.

Figures 1a and 1b show how the time-averaged regrets varied over time with different settings of learning rate α and discount factor γ, respectively. In almost all settings, the time-averaged regret rose sharply in the beginning and then declined gradually, which confirms the effectiveness of the reinforcement learning. In Fig. 1a, the setting of $\alpha = 0.7$ had the best performance. This result can be justified as other settings of α may need more training episodes to converge to better results. The same reason also explains why the setting of $\gamma = 0.75$ had the best performance in Fig. 1b.

C. Overall Performance Comparisons

For a fair comparison among different approaches, we measured the operative link ratio (OLR) as the main performance metrics as it directly gauges the effectiveness of channel allocation. OLR is defined as the ratio of the total number of operative links to the total number of designated links [3].

We first reused the topology in Sec. IV-B for testing. All other settings followed [1]. For the game-based and SAP-based approaches, we randomly allocated initial channels to radios and played the game for 100 times. For the MARL-based approach, we trained the models for 200 episodes, and exploited the final Q-tables to perform channel allocations for 100 times. We test the MARL-based approach with two different numbers of channels. Table II shows the average number of movements, the final utility and the OLR of each approach.

Although the game-based approaches had higher utility, these approaches were inferior to the proposed approaches in terms of OLR. In particular, the MARL-based approach yielded the highest OLR with four channels. This is due to its ability to fully utilize all available channels. In contrast, the game-based and SAP-based approaches were constrained by the Pigeonhole Principle and thus could use only three out of four channels. When the number of channels was reduced to three, the MARL-based approach performed slightly worse than the SAP-based approach with dynamic β setting. On the other hand, the game-based approaches needed the fewest movements to converge, followed by the SAP-based approach and then the MARL-based approach.

We then used a new topology for another test. Here each node had four designated links and two radios. Table III lists the results.

In general, the MARL-based approach performed the best in terms of OLR, followed by the SAP-based approach and then the game-based approaches. The superiority of the MARL-based approach becomes clear when more channels can be utilized: it not only achieved the highest OLR, but also yielded a much fewer movements compared to its result with only three channels.

V. Conclusions

We have proposed two approaches to channel allocation in WMN. One takes SAP as a game-playing strategy for an existing game-theoretic solution. The other takes a multi-agent reinforcement learning approach. We have conducted extensive simulations to study the performance of the proposed approaches. The results show that, in various scenarios, the proposed approaches outperform the prior game-based approach in terms of operative link ratio. This research provides a new perspective for the channel allocation problem, which might be helpful to future researchers.

References

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

[2] H. P. Young, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press, 1998.

[3] L.-H. Yen and K.-W. Huang, "Link-preserving interference-minimization channel assignment in multi-radio wireless mesh networks," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 18, no. 4, pp. 222–233, Apr. 2015.

[4] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 384–404, Mar. 2000.

[5] Y. Shi, Y. T. Hou, J. Liu, and S. Kompella, "Bridging the gap between protocol and physical models for wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1404–1416, Jul. 2013.

[6] J. Xiao, N. Xiong, L. T. Yang, and Y. He, "A joint selfish routing and channel assignment game in wireless mesh networks," *Comput. Commun.*, vol. 31, no. 7, pp. 1447–1459, 2008.

[7] X. Chen, J. Xu, W. Yuan, W. Liu, and W. Cheng, "Channel assignment in heterogeneous multi-radio multi-channel wireless networks: A game theoretic approach," *Comput. Netw.*, vol. 57, pp. 3291–3299, 2013.

[8] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wirel. Netw.*, vol. 11, no. 4, pp. 471–487, 2005.

[9] D. Yang, X. Fang, and G. Xue, "Channel allocation in non-cooperative multi-radio multi-channel wireless networks," in *Proc. IEEE INFOCOM*, Orlando, Florida USA, Mar. 2012.

[10] B. hong Ma, T. Liang, L. Zhu, and H. Zheng, "A wireless mesh network channel assignment method based on potential game," *Applied Mechanics and Materials*, vol. 696, pp. 191–200, Nov. 2014.

[11] L.-H. Yen and B.-R. Ye, "Link-preserving channel assignment game for wireless mesh networks," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 31, no. 1, pp. 13–22, May 2019.

[12] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.

[13] K. Yamamoto, "A comprehensive survey of potential game approaches to wireless networks," *IEICE Trans. Commun.*, vol. E98.B, no. 9, pp. 1804–1823, Jan. 2015.

[14] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1226–1252, 2021.