Imitation Learning Backoff: Reinforcement Learning-based Channel Access for Guaranteeing Fairness (Student Abstract)

Taegyeom Lee, Ohyun Jo*

Chungbuk National University, 28644 Cheongju, Republic of Korea
taegyeom_1@chungbuk.ac.kr, ohyunjo@chungbuk.ac.kr

Abstract

This paper addresses contention window optimization for multi-access scenarios. Our investigation into state-of-the-art models revealed that a limited number of nodes dominate the communication channels. Such unfairness issue is critical in networks as they can lead to significant disruptions. To mitigate this problem, we propose an imitation learning-based backoff mechanism. The proposed model is a reinforcement learning-based contention window optimization method. It imitates the expert's policy to ensure fair policy convergence for the agent and includes opportunities for weight adjustment to boost performance. The proposed model shows a fairness improvement of approximately 20% to 41% across various scenarios.

Introduction

The UAV (Unmanned Aerial Vehicle) network considered in this work involves a competitive access environment where multiple UAVs share a single channel. In such an environment, packet collisions occur when two or more UAVs attempt to access the channel simultaneously. The collision issues can be addressed with sequential access methods; however, it is very challenging to predict in advance when and how many UAVs will attempt to access the channel and to assign priorities accordingly. Therefore, contention-based channel access models have been studied. These models adjust the size of the Contention Window (CW) based on state information modeled through Markov Decision Processes (MDP) (Zhang et al. 2017; Balador et al. 2017; Wu and Xu 2017; Yang et al. 2016). However, they face challenges in adapting to changes in network topology. We focus on pattern information which can be obtained from the UAV's trajectory, speed, and flight direction in the UAV networks.

We have previously investigated a LB (Learning Backoff) mechanism to optimize short backoff times based on the pattern information (Lee and Jo 2023). LB mitigates collision issues through a exploration and exploitation policy, where each LB-UAV can find an optimal policy without exchanging information among them. However, we discovered that policies of LB can lead to monopolization of the channel by a few UAVs. This paper presents ILB(Imitation Learning Backoff). The ILB model is a distributed learning model but enables fair decision-making. The fairness of ILB is achieved by imitating the behavior of experts.

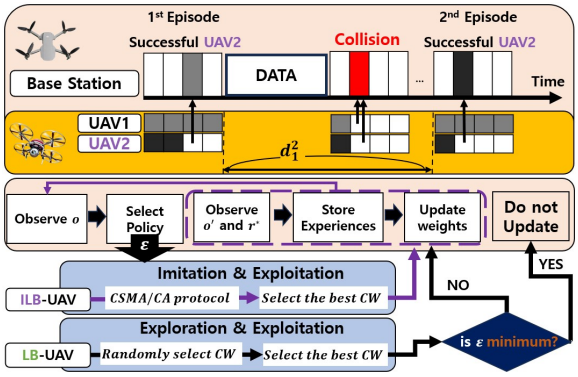

Figure 1: Structure of ILB and LB channel access model.

Imitation Learning Backoff Model

Non-stationarity Problem. The non-stationarity problem arises when the environment in a multi-agent system is dynamically changing due to the actions of multiple agents. In the considered multiple access scenario, UAVs select a CW to access the shared channel. The actions of the UAVs impact changes in channel traffic. The agent of UAV can converge to the optimal policy that adapts to the non-stationarity of environment. As our simulation results revealed, the LB model leads to a loss of fairness, as a small number of UAVs tend to monopolize the channel.

Policy Imitation and Weight Adjustment. To address the non-stationarity problem, which leads to unfairness, this paper proposes an approach based on policy imitation and WA (Weight Adjustment). Initially, the ILB model ensures that all agents imitate and learn the same policy to help UAVs anticipate each other's actions and overcome the problem. The policy is based on the CSMA/CA protocol (Colvin 1983). The CSMA/CA protocol is highly fair in wireless communication environments where predicting channel traffic is difficult. ε denotes the probability of selecting a policy. When ε reaches its minimum value, agents stop policy imitation and proceed weight adjustment. ILB-UAVs learn

*Corresponding author: Ohyun Jo

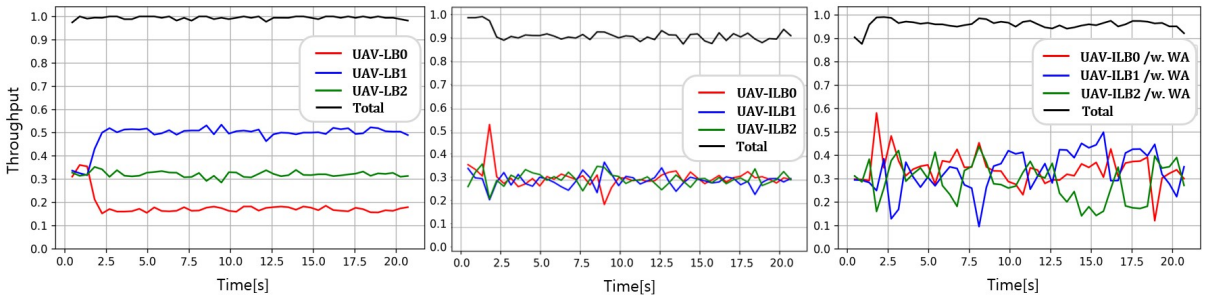

Figure 2: Comparison throughputs of LB, ILB and ILB-WA in 3-UAVs scenario.

from new experiences to better adapt to each other's policies. This process is conducted after every 100 successful transmissions.

ILB Model for CW optimization. Fig. 1 illustrates the structure of the ILB model and the LB model in the considered scenario. The ILB agent's action $a \in A$ determines the upper bound of the CW. a can be expressed as $CW_a = 2^{a+u} - 1,\ a \in 0, 1, 2, ...m$, where u can be adjusted according to the network environment.

The observation $o \in \Omega$ is observed at the end of an episode. o_n represents the observation obtained in the n-th episode. The o can be defined as follows. $o \triangleq [\frac{C_d}{A_d}, d, RSS]$ Here, A_d and C_d represent the number of transmission attempts and collisions during time d, respectively. As can be seen in Fig 1, d denotes the time from the agent's last successful transmission to the next packet's successful transmission. The RSS represents the received signal strength ratio from other UAVs. This observation provides a representative pattern of information obtainable from the considered scenario (Lee and Jo 2023). The reward function is designed to ensure fair channel utilization and collision avoidance. The reward function for fairness is defined as follows:

$$r_d = \begin{cases} \frac{d'}{d'+d} - \frac{d}{d'+d} & if\ successful, \\ 0 & collision. \end{cases} \quad (1)$$

Here, d is the same as in o, and d' represents the delay from the previous episode. The reward function for collision avoidance is defined as follows:

$$r_c = \begin{cases} 1 - \frac{C_d}{A_d} & if\ successful, \\ -1 & collision. \end{cases} \quad (2)$$

The agent gets the sum of the two reward functions as the total reward. We use the Deep Q-network (DQN) model as our reinforcement learning model. The loss function $L(w)$ used for weight updates is defined as follows. $L(w) = \sum_{bs}(r^* + \gamma \underset{a_t}{max} Q(o, a; w^-) - Q(o', a'; w))^2$. Here, r^* is the total reward, bs is the batch, γ is the discount factor, and w^- and w denote the network parameters of the DQN. o refers to the previous observation, and o' refers to the current observation.

Experimental Results

In this section, we conduct extensive experiments to evaluate the fairness, throughput, and delay of the proposed model. Fig. 2 illustrates the changes in individual throughput of UAVs and the total network throughput in a 3-UAV scenario. As shown in the figure, LB agents have priorities, with the highest-priority agent capturing most of the total throughput. In contrast, the proposed model exhibits more uniform throughput compared to the LB model.

	3-UAVs			4-UAVs			5-UAVs		
Models	**F**	**T**	**D(μs)**	**F**	**T**	**D (μs)**	**F**	**T**	**D(μs)**
CSMA/CA	0.99	0.89	0.0012	0.99	0.83	0.0017	0.99	0.78	0.0023
LB(Lee and Jo 2023)	0.82	0.98	0.0031	0.80	0.98	0.0038	0.70	0.97	0.0047
ILB	0.98	0.92	0.0015	0.99	0.84	0.0020	0.98	0.80	0.0027
ILB +WA	0.99	0.96	0.0014	0.99	0.92	0.0020	0.99	0.90	0.0025

Table 1: Comparison of performance in different scenarios

Table 1 presents a comparative analysis of the performance of four models using three metrics: throughput (T), fairness index (F), and delay (D). Throughput T refers to the transmission success rate of UAVs. The fairness index is first introduced to evaluate the fairness degree of resource allocation policies (Guo et al. 2014). JFI F is defined as $F = \frac{(\sum_{j=1}^{n} T_j)^2}{n \times \sum_{j=1}^{n} {T_j}^2}$, where T_j represents the throughput of the j-th UAV and n denotes the total number of UAVs.

Delay D is defined as $D = \frac{\sum_{n=1}^{N} \sum_{i=1}^{I_n} d_i^n}{\sum_{n=1}^{N} I_n}$, where d_i^n represents the delay in the i-th episode of the n-th UAV and I_n denotes the number of episodes conducted by the n-th UAV. The proposed models demonstrate high fairness indices. The ILB-WA model, with the addition of the WA process, shows dynamic throughput variations in Fig. 4 but maintains fair channel utilization. The higher throughput of LB compared to the proposed models is attributed to fixed priorities, as evidenced by LB's highest delay rate. Although the proposed model exhibits slightly worse delay performance, it significantly outperforms the conventional model in terms of throughput.

Conclusion

In this paper, we propose a channel access method based on imitation learning to ensure fairness. We have formulated an ILB model tailored to the CW optimization problem and introduced a novel DRL mechanism to enhance fairness. The proposed model has demonstrated, through extensive experiments, its ability to mitigate the monopolization issues of state-of-the-art models and enable stable and successful packet processing.

Acknowledgments

This work was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00165, Development of 5G+ Intelligent Basestation Software Modem). This work was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2021R1A2C2095289).

References

Balador, A.; Calafate, C.; Cano, J.-C.; and Manzoni, P. 2017. A density-based contention window control scheme for unicast communications in vehicular ad hoc networks. *International Journal of Ad Hoc and Ubiquitous Computing*, 24(1-2): 65–75.

Colvin, A. 1983. CSMA with collision avoidance. *Computer communications*, 6(5): 227–235.

Guo, C.; Sheng, M.; Wang, X.; and Zhang, Y. 2014. Throughput maximization with short-term and long-term Jain's index constraints in downlink OFDMA systems. *IEEE transactions on communications*, 62(5): 1503–1517.

Lee, T.; and Jo, O. 2023. Learning Backoff: Deep Reinforcement Learning-Based Wireless Channel Access. *IEEE Systems Journal*.

Wu, G.; and Xu, P. 2017. Improving performance by a dynamic adaptive success-collision backoff algorithm for contention-based vehicular network. *IEEE Access*, 6: 2496–2505.

Yang, B.; Wang, S.; Shi, X.; and Han, W. 2016. QoS-assurance-and-mobility-based EDCA mechanism for vehicular network access. In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, 2208–2212. IEEE.

Zhang, C.; Chen, P.; Ren, J.; Wang, X.; and Vasilakos, A. V. 2017. A backoff algorithm based on self-adaptive contention window update factor for IEEE 802.11 DCF. *Wireless networks*, 23: 749–758.