Maximum Throughput in the Unlicensed Band under 3GPP Fairness

Jianbin Xiao*, Haowei Xu*, Xinghua Sun*, Feifan Luo*, and Wen Zhan*

*School of Electronics and Communication Engineering, Shenzhen Campus of Sun Yat-sen University
{xiaojb7, xuhw3, luoff6}@mail2.sysu.edu.cn, {sunxinghua, zhanw6}@mail.sysu.edu.cn

2022 IEEE/CIC International Conference on Communications in China (ICCC) | 978-1-6654-8480-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCC55456.2022.9880665

Abstract—**Since unlicensed band already hosts well-established technologies such as Wi-Fi, the transmissions from another network systems in unlicensed spectrum could make Wi-Fi suffer from severe performance degradation without a fair coexistence mechanism. 3GPP has proposed their definition of fairness criterion, referred to as "3GPP fairness", which restricts the Wi-Fi performance not to be affected by other unlicensed nodes. Thus it is of paramount importance to establish an access mechanism that guarantees the 3GPP fairness. This paper investigates how to maintain the 3GPP fairness between Wi-Fi and other unlicensed nodes, in the meanwhile, obtain the maximum of total throughput. We first obtain a benchmark by solving the optimization problem of total throughput under the 3GPP fairness constraint. Then we propose a deep reinforcement learning (DRL) mechanism based on Double Deep Q-network (DDQN) to help unlicensed nodes make access decisions while coexisting with Wi-Fi, so that they can learn to maximize total throughput without against the 3GPP fairness. Extensive simulations reveal that the DRL mechanism can approach the benchmark and therefore provides an approach for the coexistence of unlicensed nodes and Wi-Fi nodes.**

Index Terms—**3GPP fairness, DRL mechanism, unlicensed spectrum, Wi-Fi.**

I. INTRODUCTION

Due to the scarcity problem in the licensed band, the sharing among unlicensed band has caused wide public concern over the recent years. US and European regulators have opened up additional 1.2 GHz spectrum in the 6 GHz bands for unlicensed radio access technologies (RATs) [1], including Wi-Fi [2] and recently announced 5G New Radio Unlicensed (5G NR-U) [3]. Thus it has gained significant attestipulatesntion to guarantee the efficient and fair coexistence for both Wi-Fi and other unlicensed nodes.

Since the performance of Wi-Fi can be easily affected by other licensed nodes, the 3rd Generation Partnership Project (3GPP) has proposed their definition of fairness in Release 16 for 5G NR-U similar to that of LTE-Licensed Assisted Access (LTE-LAA) [4], [5], which stipulates that Wi-Fi performance is not degraded due to the coexistence of NR-U (or LTE-LAA). Towards this purpose, 3GPP fairness is attained between LTE-U and Wi-Fi by dividing the channel airtime into two orthogonal airtimes for LTE-U and Wi-Fi [6]. In [7], authors have demonstrated LTE-U and Wi-Fi coexistence under 3GPP fairness by appropriately regulating the duty cycle. However, the optimal parameter in duty cycle needs to be selected carefully based on the number of nodes in the Wi-Fi and LTE-U networks; otherwise 3GPP fairness will no longer be satisfied. In [8], authors consider LTE-LAA implementations with reservation signal duration limited by three OFDM symbols and the simulation results show improved coexistence between Wi-Fi and LTE-LAA. Moreover, In [9], the optimal backoff window sizes of the LTE-LAA and Wi-Fi are given in expressions to obtain fair and efficient coexistence of Wi-Fi and LTE under 3GPP fairness.

Despite the significant efforts that have been made towards coexistence under 3GPP fairness, the optimal performance under 3GPP fairness remains largely unknown. To address this issue, we formulate an optimization problem to maximize the total throughput under 3GPP fairness. A benchmark in terms of throughput of unlicensed nodes and the Wi-Fi nodes is derived in closed form, which is determined by the number of nodes, backoff parameters, and the durations of one successful transmission or collision. Motivated by [10], we propose a Deep Reinforcement Learning (DRL) mechanism based on Double Deep Q-network (DDQN) for unlicensed nodes to learn the access strategy on 3GPP fairness coexistence with Wi-Fi. Related works, such as [11], proposed Carrier-sense Deep-reinforcement Learning Multiple Access (CS-DLMA) protocol to achieve fairness objective when coexisting with Wi-Fi. However, the theoretical benchmark of considered problem was not analyzed in [11], and thus the performance of CS-DLMA lacks adequate evaluations. In contrast, extensive simulation show that the proposed DRL mechanism can approach the benchmark of 3GPP fairness under different numbers of nodes. Overall, our contributions can be summarized as follows:

- We formulate an optimization problem about maximizing the total throughput of Wi-Fi nodes and the other unlicensed nodes under 3GPP fairness without specification of access pattern or policy. A theoretical benchmark is then derived in closed form by solving the optimization problem.
- We propose a DRL mechanism for unlicensed nodes to achieve the maximum total throughput while coexisting Wi-Fi and DRL networks under 3GPP fairness. Extensive simulations results show that the benchmark can be approached under different experiment configurations.

The remainder of this paper is organized as follows. Section II presents the system model and the benchmark of throughput

---

This work was supported in part by the National Natural Science Foundation of China under Grant 62001524, and in part by Shenzhen Science and Technology Program (No. 2021A04 and No. RCBS20210706092408010)

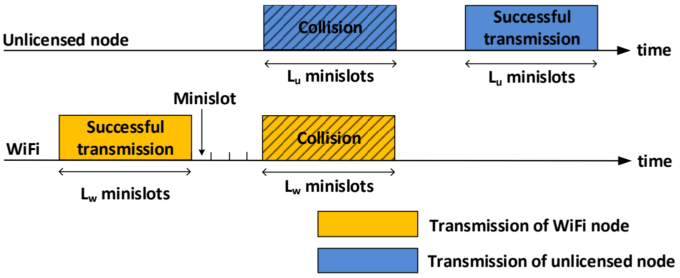


Fig. 1. Transmission and collision model when the unlicensed nodes coexist with Wi-Fi. Note that if Wi-Fi nodes sense the channel is idle and transmit, the unlicensed nodes probably access the channel in one of the following L_w minislots, which will cause a collision.

for DRL and WiFi nodes under the 3GPP fairness. The DDQN-based access protocol for the unlicensed while coexisting Wi-Fi nodes is expounded in Section III. Simulation results are presented in Section V, followed by the conclusion in Section VI.

II. SYSTEM MODEL AND BENCHMARK

A. System Model

Consider a time-slotted heterogeneous wireless network in which M unlicensed nodes and $N-M$ Wi-Fi nodes transmit packets to their associated Access Point (AP) via a shared wireless channel. We assume that the packets are always enough for all nodes, i.e., they are saturated. As illustrated in Fig. 1, the time is slotted as 1 to n minislots. We assume a node can only begin its transmission at the beginning of minislot and must finish its transmission at the end of minislot. Due to all nodes sharing a common wireless channel, a packet can be successfully received by the AP only if there is one node transmitting in the channel during the whole packet length. After each successful transmission, AP will broadcast an acknowledgment (ACK). On the contrary, if the transmissions overlap, i.e., there are more than one node transmitting packets meanwhile, then a collision occurs and a NACK will be broadcasted. Since the ACK/NACK can only be received after the transmission is finished, the whole packet will be regarded as failure if a collision occurs. We assume the duration of packet for both unlicensed nodes and Wi-Fi nodes consists of multiple minislots, which are denoted as L_w and L_u[1], respectively.

We consider that the Wi-Fi nodes adopt the Distributed Coordination Function (DCF) with the basic access mechanism. For each Wi-Fi node, it first generates a random value $\omega \in [0, 2^k W-1]$, where the initial value of k is 0 and W is the initial window size. Then it waits for ω idle minislots before transmitting a packet. Every time a Wi-Fi node encounters a collision, the value of k is incremented by 1, until it meets the cutoff phase K. It will be reset to 0 after a successful transmission. For fair coexistence purpose, we assume that the parameters are identical in Wi-Fi nodes.

The unlicensed nodes, on the other hand, can adopt any feasible mechanism. We assume that they adopt an identical access strategy. Thus the throughput of each node is also identical.

B. 3GPP Fairness and Benchmark

Let T denote the total operation time of the network. Without specification of access pattern or policy, the unlicensed nodes and Wi-Fi nodes successfully transmit in T_1 and T_2 minislots, respectively, where $T_1+T_2 \le T$. Then the throughput of each unlicensed node and Wi-Fi node are given by $Th_u^{(i)}$ and $Th_w^{(j)}$ as follows:

$$Th_u^{(i)} = \frac{T_1}{MT},\ Th_w^{(j)} = \frac{T_2}{(N-M)T}. \tag{1}$$

We suppose that $\lambda_{out}^{'}$ represents the throughput of single node in the network with N Wi-Fi nodes. 3GPP fairness is defined as that other network has less or equal impact on an existing Wi-Fi network compared with an additional Wi-Fi network on the same unlicensed channel. Then 3GPP fairness can be expressed as

$$\frac{T_2}{T} \ge (N-M)\lambda_{out}^{'}, \tag{2}$$

where $\frac{T_2}{T}$ and $(N-M)\lambda_{out}^{'}$ represent the total throughput of $N-M$ Wi-Fi nodes with the M unlicensed nodes and with other M Wi-Fi nodes, respectively. And $\lambda_{out}^{'}$ can be written as [12]

$$\lambda_{out}^{'} = \frac{-L_w p^{'} \ln p^{'}}{N(1+L_w-L_w p^{'})}, \tag{3}$$

where $p^{'}$ denotes the probability of successful transmission of each Wi-Fi node. With the initial window size W and cutoff phase K, $p^{'}$ is given as the single nonzero root of

$$p^{'} = \exp\left\{-\frac{2N}{1+W\left(\frac{p^{'}}{2p^{'}-1}-\left(\frac{p^{'}}{2p^{'}-1}-1\right)\cdot(2-2p^{'})^K\right)}\right\}. \tag{4}$$

Furthermore, we suppose that λ_{out} represents the throughput of single node in the network with $N-M$ Wi-Fi nodes. Since the coexistence of the unlicensed nodes may lead to performance degradation of Wi-Fi, we have

$$\frac{T_2}{T-T_1} \le (N-M)\lambda_{out}, \tag{5}$$

where $(N-M)\lambda_{out}$ represents the total throughput of $N-M$ Wi-Fi nodes without the unlicensed nodes and $\frac{T_2}{T-T_1}$ represents the relative total throughput of $N-M$ Wi-Fi nodes in $T-T_1$ minislots with the unlicensed nodes. Equality in (5) is satisfied only if there is no conflict between the unlicensed nodes and Wi-Fi nodes. With the duration of transmission L_w and number of Wi-Fi nodes $N-M$, λ_{out} can be written as

$$\lambda_{out} = \frac{-L_w p \ln p}{(N-M)(1+L_w-L_w p)}, \tag{6}$$

where p is similar to the definition in (4) as:

$$p = \exp\left\{-\frac{2(N-M)}{1+W\left(\frac{p}{2p-1}-\left(\frac{p}{2p-1}-1\right)\cdot(2-2p)^K\right)}\right\}. \tag{7}$$

---

[1] Here the duration of ACK or NACK is included in L_w and L_u.

To maximize the total throughput of Wi-Fi nodes and unlicensed nodes, we have the following optimization problem:

$$\begin{aligned} \max_{\pi_D} \quad & \frac{T_1}{T}+\frac{T_2}{T} \\ s.t. \quad & T_1,\ T_2 \geq 0 \\ & T_1+T_2 \leq T \\ & \frac{T_2}{T-T_1} \leq (N-M)\lambda_{out} \\ & \frac{T_2}{T} \geq (N-M)\lambda_{out}^{'}. \end{aligned} \tag{8}$$

The following theorem provides the throughput of each node by solving (8).

Theorem 1. *To maximize the total throughput under 3GPP fairness, the throughputs of each unlicensed node and each Wi-Fi node are given by*

$$Th_u^{(i)} = \frac{1-\frac{\lambda_{out}^{'}}{\lambda_{out}}}{M} \tag{9}$$

and

$$Th_w{}^{(j)} = \lambda_{out}^{'}, \tag{10}$$

respectively.

Proof. See Appendix A.

$\square$

III. DRL MECHANISM

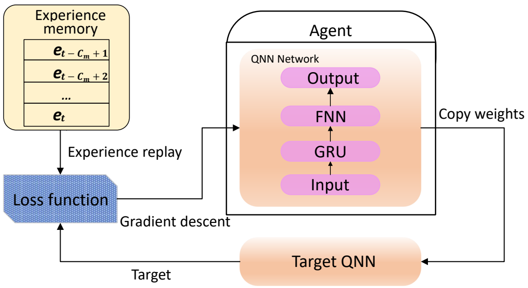
This section puts forth a DDQN-based access protocol for the unlicensed nodes (which we consider as DRL nodes) to coexist with Wi-Fi.

First, we need to define the corresponding *action*, *state*, and *reward*.

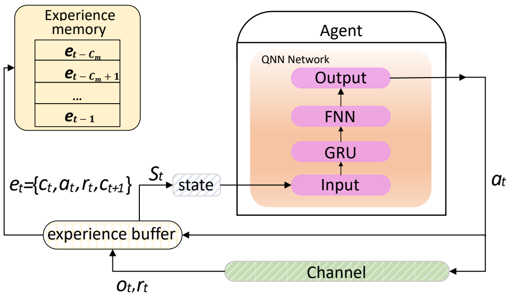
The *action* selected by the DRL node in time slot t is $a_t \in \{TRANSMIT, SENSE\}$, where $TRANSMIT$ means the DRL node transmits packet for the duration L_d, and $SENSE$ means it performs carrier sensing for one time slot. If $a_t = TRANSMIT$, after the transmission is completed, the DRL node will receive an observation $o_t \in \{SUCCESSFUL, COLLIDED\}$, indicating whether the packet is transmitted successfully or not; If $a_t = SENSE$, after one time slot, the DRL node can get an observation $o_t \in \{BUSY, IDLE\}$, indicating whether the channel is occupied or not by other nodes.

Since the duration of one action-observation pair $\{a_t, o_t\}$ may be one or multiple time slots, we define l_t to represent the time slots that it lasts. Then the channel state observed by the DRL node can be denoted as $c_{t+1} = \{a_t, o_t, l_t\}$, which reflects the system state in one time step t. The time step t can only move to $t+1$ when the action or observation changes. We define the *state* in time slot t to be $s_t = \{c_{t-L+1}, c_{t-L+2}, ..., c_t\}$, where L is the history length of channel state sequence.

Recall that the optimization objective is to maximize the total throughput of Wi-Fi nodes and unlicensed nodes while ensuring 3GPP fairness. We assume the DRL node can obtain a positive reward if Wi-Fi nodes successfully transmit one packet and can receive all ACKs and NACKs (regardless of whether the corresponding packet is transmitted by itself or the Wi-Fi nodes), which contain a *reward* vector $\boldsymbol{r}_t = \{r_t^{(u)}, r_t^{(w)}\}$. If Wi-Fi nodes have successfully transmitted a packet, then $\boldsymbol{r}_t = \{0, 1\}$. More details about the reward are given as follows:



(a) Training



(b) Execution

Fig. 2. Training and execution procedure of the DRL agent.

- $a_t = TRANSMIT$: (1) the obtained observation is $o_t = SUCCESSFUL$, if $\sum_{j=1}^{N-M} Th_w^{(j)} \geq (N-M)\lambda_{out}^{'}$, then it can get a reward vector $\boldsymbol{r}_t = \{1, 0\}$; otherwise, the reward vector is negative and obtained as $\boldsymbol{r}_t = \{-0.1, 0\}$, which can punish the behavior of over transmitting even if the transmission is successful; (2) the obtained observation is $o_t = COLLIDED$, then it can get a reward vector $\boldsymbol{r}_t = \{0, 0\}$.
- $a_t = SENSE$: (1) the obtained observation is $o_t = IDLE$, then it can get a reward vector $\boldsymbol{r}_t = \{0, 0\}$; (2) the observation is $BUSY$, then the reward vector can only be obtained from ACK or NACK at the end of the last slot of transmission.

The framework of the training and execution procedure is shown in Fig. 2. For training, a first-in-first-out (FIFO) experience memory $E = \{e_{t-C_m+1}, \ldots, e_t\}$ is adopted to store experience tuple $e_t = \{c_t, a_t, \boldsymbol{r}_t, c_{t+1}\}$ with a fixed capacity C_m. Then in DDQN [13], for every step, a random mini-batches of experience tuples are sampled from the experience memory to compute the loss function as following:

$$L(\theta) = \frac{1}{N_E} \sum_{e \in E} \|\mathbf{Y}_t^{DDQN} - \boldsymbol{Q}(s_t, a_t, \theta)\|_2^2. \tag{11}$$

where N_E denotes the batch size. To take different durations of action-observation pairs into account, $\boldsymbol{Y}_t^{DDQN}$ is given as

$$\boldsymbol{Y}_t^{DDQN} = \boldsymbol{r}_t \cdot (1+\gamma+\gamma^2+...+\gamma^{l_t-1}) + \gamma^{l_t}\boldsymbol{Q}\left(s_{t+1}, \arg\max_{a'} \sum_{i\in\{u,w\}} Q^{(i)}(s_{t+1}, a', \theta), \theta^-\right). \tag{12}$$

The weights θ of QNN are updated by gradient decent to minimize the loss function (11). Since the input of QNN is $s_t = \{c_{t-L+1}, c_{t-L+2}, ..., c_t\}$, we sample continuous experiences to extract L channel states for a s_t. To process the inputs s_t sequentially, we employ a three-layer QNN, in which two hidden layers are gated recurrent unit (GRU) layer and one hidden layer is feedforward neural network (FNN). As for the target QNN, its weights θ^- is periodically copied from θ every F training iterations.

For execution, we input the current state $s_t = \{c_{t-L+1}, c_{t-L+2}, ..., c_t\}$ into QNN network, and the output is an approximated value vector $\boldsymbol{Q}(s_t, a, \theta) = \{Q^{(u)}(s_t, a, \theta), Q^{(w)}(s_t, a, \theta)\}$ for two different networks. Note that the exact throughput $Th_u^{(i)}$ and $Th_w^{(j)}$ in (1) is unknown for themselves, $\boldsymbol{Q}(s_t, a, \theta)$ can be used to help the agent select actions to optimize the 3GPP fairness function as following:

$$a_t = \arg\max_{a} \sum_{i\in\{u,w\}} Q^{(i)}(s_t, a, \theta). \tag{13}$$

Note that one transmission from Wi-Fi lasts for L_w time slots, thus the DRL node needs to select the correct action L_w times consecutively to avoid a collision with the Wi-Fi nodes. Otherwise, the packets from both of DRL and Wi-Fi nodes will collide, and at most $(L_w + L_d - 1)$ time slots will be wasted. In practical wireless network, the duration of a transmission in a Wi-Fi network usually exceeds 100 time slots. To avoid a large number of collision slots during the training phase, the agent can only select $a_t = TRANSMIT$ if the last observation $o_{t-1} = IDLE$. The duration of one transmission of the DRL node is also set to be more than 100 time slots. Therefore the idle time slot before each transmission has a negligible impact on the throughput of the DRL node.

Finally, an $\epsilon-$greedy algorithm is adopted, i.e., the agent selects an optimal action as (13) with probability $1-\epsilon$, or a random action with probability ϵ, if the last observation $o_{t-1} = IDLE$. The value of ϵ decays every time slots until reaching a minimum value min-ϵ.

IV. PERFORMANCE EVALUATION

This section first presents the experiment setup. Then the performance of the proposed DRL mechanism is evaluated by simulations with different configurations including the number of Wi-Fi nodes and DRL nodes, the initial window size W and cutoff phase K.

A. Experiment Setup

In the simulation, we consider deploying the proposed DRL mechanism in a gateway node which is associated with all DRL nodes in the network. After each transmission decision to transmit, the gateway node will select one of the DRL nodes to transmit to the AP in a circular rotation method.

TABLE I
HYPER-PARAMETERS OF THE DRL NODE

Hyperparameters	Value
State length L	10
Duration of transmissions L_d	120
Size of experience memory C_m	500
Target QNN update frequency F	100
Batch size in experience replay N_E	32
Discount factor γ	0.995
Range of ϵ in $\epsilon-$greedy algorithm	1 to 0.05
Decay rate of ϵ	0.9995

As shown in Fig. 2, there are three hidden layers in the QNN architecture, including two GRU layers and a FC layer. Both of the hidden layers have 64 neurons and adopt the leaky ReLU functions as the activation functions. The RMSprop algorithm is used for batch gradient descent of the QNN. Other hyperparameters of the DRL agent are summarized in Table I. Wi-Fi nodes adopt the Binary Exponential Backoff (BEB). Each transmission of both DRL and Wi-Fi nodes lasts for 120 time slots. In order to diversify λ_{out} and $\lambda_{out}^{'}$ in values when numbers of DRL nodes and Wi-Fi nodes remain the same, the initial window size W and cutoff phase K are joined as pair $(W, K) \in \{(16, 2), (16, 4), (16, 6), (32, 4)\}$. We also consider other combinations, but the results did not change significantly, so only the results of these four combinations are shown.

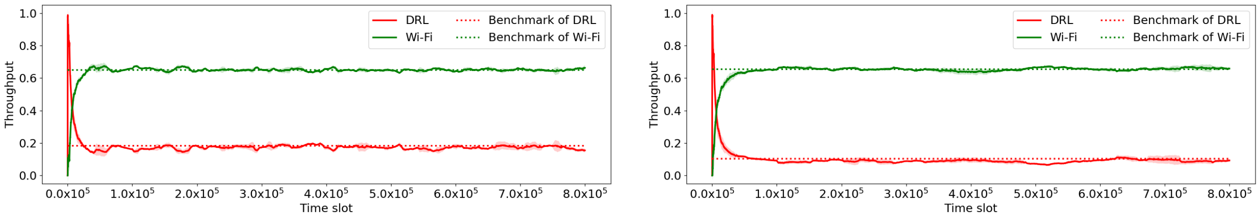
The simulated throughput is obtained as the ratio of successful transmissions over the last T time slots. Specifically, the value of T is set to be 100000 in the figures of "Throughput vs. Time slot".

B. Simulation Results

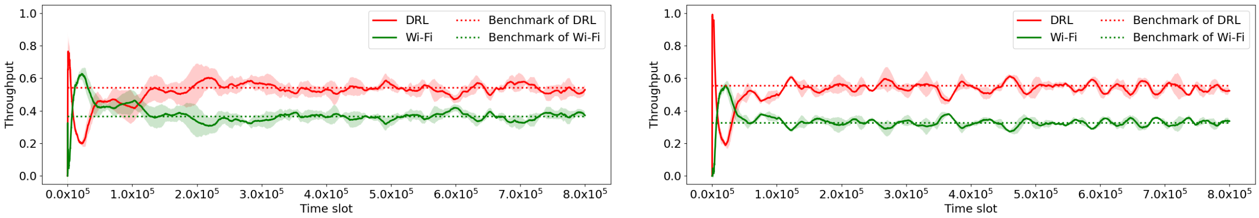
To examine the performance of the proposed DRL mechanism, we present the throughput of both the DRL and Wi-Fi nodes in different configurations including the number of nodes, initial window size W and cutoff phase K under the 3GPP fairness.

Fig. 3 presents the experimental results with different numbers of DRL nodes and Wi-Fi nodes to achieve 3GPP fairness. In the experiment, we set $W = 16$, and $K = 4$. Fig. 3 displays that both the DRL nodes and Wi-Fi nodes can approach the benchmark of throughput no matter whether there are one or multiple DRL nodes, i.e., the DRL agent has learned a near-optimal policy in all the cases to achieve 3GPP fairness when it coexists with the Wi-Fi nodes.

We next consider the coexistence of 10 DRL nodes and 10 Wi-Fi nodes where (W, K) is selected from $\{(16, 2), (16, 4), (16, 6), (32, 4)\}$. The simulation results are presented in Fig. 4. From the experiment we can see that the gap between benchmark with actual throughput is nearly zero w.r.t the choices of K and the initial window size W.

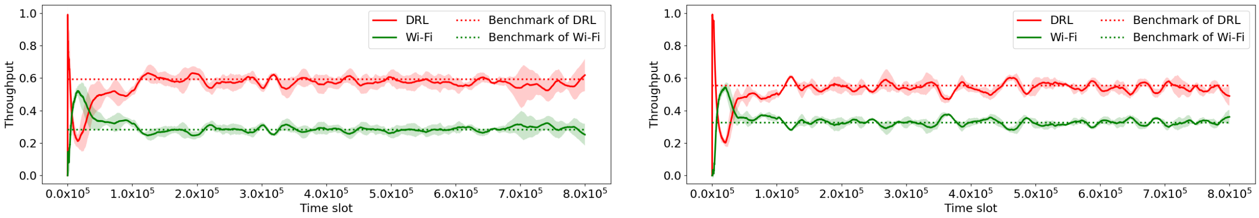


(a) 1 DRL node and 5 Wi-Fi nodes (b) 1 DRL node and 10 Wi-Fi nodes

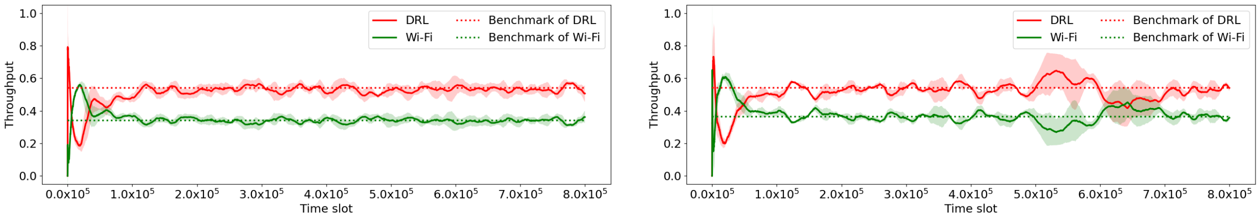


(c) 5 DRL nodes and 5 Wi-Fi nodes (d) 10 DRL nodes and 10 Wi-Fi nodes

Fig. 3. Throughput performance under the coexistence of different numbers of DRL and Wi-Fi nodes to achieve 3GPP fairness. Each curve is averaged over 5 different runs.



(a) $W = 16, K = 2$ (b) $W = 16, K = 4$



(c) $W = 16, K = 6$ (d) $W = 32, K = 4$

Fig. 4. Throughput performance under the coexistence of 10 DRL nodes and 10 Wi-Fi nodes to maximize the 3GPP fairness objective with different values of W and K. Each curve is averaged over 5 different runs.

V. CONCLUSION

In this paper, we formulated the problem of achieving 3GPP fairness for Wi-Fi nodes and unlicensed as an optimization problem. Then we derive a theoretical benchmark by solving the formulated problem. To approach the benchmark, We proposed a DDQN based access protocol and demonstrated that DRL nodes can coexist harmoniously with Wi-Fi in heterogeneous environments without knowing the parameters of Wi-Fi. Simulations indicated that the proposed DDQN mechanism can approach the benchmark under various network parameter settings including the number of Wi-Fi nodes and DRL nodes, the initial window size and cutoff phase.

APPENDIX A
PROOF OF THEOREM 1

Proof. Note that (8) $\frac{T_1}{T} + \frac{T_2}{T}$ is a monotonically increasing function of both T_1 and T_2. Since $0 \le (N-M)\lambda_{out} \le 1$, we have from (8) that

$$T_1 \le T - \frac{T_2}{(N-M)\lambda_{out}} \le T - T_2. \tag{14}$$

Therefore, in order to maximize $\frac{T_1}{T} + \frac{T_2}{T}$, there must be $T_1 = T - \frac{T_2}{(N-M)\lambda_{out}}$. Due to 3GPP fairness, we can have

$$(N-M)\lambda_{out}^{'} T \le T_2 \le (N-M)\lambda_{out} T. \tag{15}$$

Therefore, equation (8) is thus equivalent to

$$\begin{aligned} \max \quad & \frac{T - \frac{T_2}{(N-M)\lambda_{out}} + T_2}{T} \\ s.t. \quad & (N-M)\lambda_{out}^{'} T \le T_2 \le (N-M)\lambda_{out} T. \end{aligned} \tag{16}$$

It is clear that $\frac{T - \frac{T_2}{(N-M)\lambda_{out}} + T_2}{T}$ is a monotonically decreasing function of T_2. Therefore, the total throughput is maximized when $T_2^* = (N-M)\lambda_{out}^{'} T$. Thus we have

$$T_1^* = T - \frac{T_2^*}{(N-M)\lambda_{out}} = (1 - \frac{\lambda_{out}^{'}}{\lambda_{out}})T. \tag{17}$$

Based on (1) and (17), $Th_d^{(i)}$ and ${Th_w}^{(j)}$ can be finally given by (9) and (10), respectively.

□

REFERENCES

[1] *Report and Order and Further Notice of Proposed Rulemaking; In the Matter of Unlicensed Use of the 6 GHz band (ET Docket No. 18- 295); Expanding Flexible Use in Mid-Band Spectrum Between 3.7 and 24 GHz (GN Docket No. 17-183)*, FCC, April 2020.

[2] *IEEE Std. 802.11-2012 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE, Mar. 2012.

[3] *Technical Specification Group Radio Access Network; Physical layer procedures for shared spectrum channel access (Release 16), 3GPP TS 37.213 V16.4.0*, 3GPP, July 2020.

[4] 3GPP, "Study on NR-based access to unlicensed spectrum; study on NR-based access to unlicensed spectrum (Release 16)," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.889, Dec. 2018, V16.0.0.

[5] —, "Technical specification group radio access network; study on licensed-assisted access to unlicensed spectrum (Release 13)," 3rd Generation Partnership Project (3GPP), Tech. Rep. 36.889, Jun. 2015, V13.0.0.

[6] V. Valls, A. Garcia-Saavedra, X. Costa, and D. J. Leith, "Maximizing LTE capacity in unlicensed bands LTE-U/LAA while fairly coexisting with 802.11 WLANs," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1219-1222, 2016.

[7] S. Fang, Y. Gao, C. Zhang, and X. Hei, "Achieving 3GPP fairness for LTE-U and WiFi coexisting networks in unlicensed spectrum," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2019, pp. 1-2.

[8] M. Cierny, T. Nihtila, T. Huovinen, M. Kuusela, F. Chernogorov, K. Hooli, and A. Toskala, "Fairness vs. performance in rel-13 LTE licensed assisted access," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 133-139, 2017.

[9] X. Sun and L. Dai, "Towards Fair and Efficient Spectrum Sharing between LTE and WiFi in Unlicensed Bands: Fairness-constrained Throughput Maximization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2713-2727, Apr. 2020.

[10] Z. Guo, Z. Chen, P. Liu, J. Luo, X. Yang and X. Sun, "Multi-Agent Reinforcement Learning-Based Distributed Channel Access for Next Generation Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1587-1599, May 2022.

[11] Y. Yu, S. C. Liew and T. Wang, "Non-Uniform Time-Step Deep Q Network for Carrier-Sense Multiple Access in Heterogeneous Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 20, no. 9, pp. 2848-2861, Sept. 2021.

[12] L. Dai and X. Sun, ”A Unified Analysis of IEEE 802.11 DCF Networks: Stability, Throughput, and Delay” *IEEE Trans. Mob. Comput.*, vol. 12, no. 8, pp.1558-1572, Aug. 2013.

[13] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. 30th AAAI Conf. Artificial Intell.*, Phoenix, AZ, USA, 2016, pp. 2094-2100.