# Multi-Agent Deep Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks with Imperfect Channels ---Benchmark

This document derives the benchmarks for different scenarios used in our paper: Multi-Agent Deep Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks with Imperfect Channels. We first assume both the uplink channels and downlink channels are perfect, i.e., $e_{up} = 0$ and $e_{down} = 0$. The derivations of the benchmarks are given below.

**Benchmark for "coexistence of one DLMA agent with one TDMA user and one ALOHA user"**

For the coexistence of one agent with one TDMA agent and one ALOHA use, we replace the agent with a model-aware user. The TDMA user transmits in the 2nd slot out of 5 slots within a TDMA frame. The ALOHA user transmits with a fixed probability $q = 0.2$ in each time slot. The model-aware knows the transmission pattern of the user and the transmission probability of the ALOHA user. In addition, we assume the feedback channel from the AP to the model-aware user is perfect. The optimal transmission strategy of the model-aware user can be derived as follows.

In the slots that are occupied by TDMA, the model-aware user just does not transmit. In the slots that are not occupied by TDMA, the model-aware user transmits with probability $p$. The value of $p$ is decided by the objective of the model-aware user and the transmission probability of ALOHA. In particular, the individual throughputs of the model-aware user, TDMA, and ALOHA are $0.8 \cdot p(1-q)$, $0.2 \cdot (1-q)$, and $0.8 \cdot (1-p)q$.

To maximize sum throughput (i.e., $\alpha = 0$ in the $\alpha$-fairness objective), the model-aware user needs to determine $p$ that maximizes

$$F(p) = 0.8 \cdot p(1-q) + 0.2 \cdot (1-q) + 0.8 \cdot (1-p)q.$$

We can determine the value of $p$ as follows: if $q < 0.5$, $p = 1$; if $q > 0.5$, $p = 0$; if $q = 0.5$, $p$ can any value in the range of $[0,1]$. Therefore, when $q = 0.2$, $p = 1$ and $F(p) = 0.8$, i.e., the sum throughput benchmark in Fig. 8 of our paper is 0.8.

To achieve proportional fairness (i.e., $\alpha = 1$ in the $\alpha$-fairness objective), the model-aware user needs to determine $p$ that maximizes

$$F(p) = \log\left(0.8 \cdot p(1-q)\right) + \log\left(0.2 \cdot (1-q)\right) + \log\left(0.2 \cdot (1-p)q\right).$$

We can find that when $p = 0.5$, the above equation is maximized. Therefore, the sum-log throughput benchmark in Fig. 9 of our paper is -5.5.

For other objectives (i.e., different values of $\alpha$ in the $\alpha$-fairness objective), we can also decide the value $p$ by maximizing

$$F(p) = f_\alpha\left(0.8 \cdot p(1-q)\right) + f_\alpha\left(0.2 \cdot (1-q)\right) + f_\alpha\left(0.8 \cdot (1-p)q\right),$$

where $f_\alpha(\cdot)$ is given by

$$f_\alpha\left(x^{(i)}\right) = \begin{cases} \left(x^{(i)}\right)^{1-\alpha} / (1-\alpha), & \alpha > 0 \ \& \ \alpha \neq 1 \\ \log\left(x^{(i)}\right), & \alpha = 1 \end{cases}.$$

**Benchmark for “coexistence of one Four agents with one TDMA user”**

In this case, we study the coexistence of four DLMA agents with one TDMA user. TDMA transmits in the 2nd slot out of 5 slots within a TDMA frame. We replace the four agents with four model-aware users. The model-aware users are aware of the transmission pattern of TDMA, and they have direct collaborations---no collisions among the model-aware users. The optimal benchmarks for different objectives are the same, i.e., the throughput of each agent and the TDMA user are both 0.2.

**Benchmark for “coexistence of one Five agents with two TDMA users and three ALOHA users”**

In this case, the TDMA user transmits in the 2nd slot out of 10 slots within a TDMA frame and the second TDMA user transmits in the 8th slot out of 10 slots within a TDMA frame. The transmission probability of each ALOHA user is $q = 0.1$. We replace the five agents with five model-aware users. We also assume the model-aware users are aware of transmission strategies of TDMA and ALOHA and the model-aware users have direct collaboration. To derive the benchmarks, we can regard the five model-aware users as one centralized model-aware user. The centralized model-aware user decides the transmissions of these five model-aware users, and assigns one of them to transmit when it decides to transmit. We denote the transmission probability of the centralized model-aware user by $p$. Then the individual throughputs of each model-aware user, each TDMA user, and each ALOHA user are $0.8\cdot(1-q)^3\cdot p/5$, $0.1\cdot(1-q)^3$, and $0.8\cdot q(1-q)^2(1-p)$.

To maximize sum throughput (i.e., $\alpha = 0$ in the $\alpha$-fairness objective), the model-aware user needs to determine $p$ that maximizes

$$F(p) = 5\left(0.8\cdot(1-q)^3\cdot p/5\right) + 2\left(0.1\cdot(1-q)^3\right) + 3\left(0.8\cdot q(1-q)^2(1-p)\right).$$

We can determine the value of $p$ as follows: if $q < 0.25$, $p = 1$; if $q > 0.25$, $p = 0$; if $q = 0.25$, $p$ can any value in the range of $[0,1]$. Therefore, when $q = 0.1$, $p = 1$ and $F(p) = 0.7288$, i.e., the sum throughput benchmark in Fig. 12 of our paper is 0.7288.

To achieve proportional fairness (i.e., $\alpha = 1$ in the $\alpha$-fairness objective), the model-aware user needs to determine $p$ that maximizes

$$F(p) = 5\log\left(0.8\cdot(1-q)^3\cdot p/5\right) + 2\log\left(0.1\cdot(1-q)^3\right) + 3\log\left(0.8\cdot q(1-q)^2(1-p)\right).$$

We can find that when $p = 5/8$, the above equation is maximized. Therefore, the sum-log throughput benchmark in Fig. 13 of our paper is -26.86.

For other objectives (i.e., different values of $\alpha$ in the $\alpha$-fairness objective), we can also decide the value $p$ by maximizing

$$F(p) = 5f_\alpha\left(0.8\cdot(1-q)^3\cdot p/5\right) + 2f_\alpha\left(0.1\cdot(1-q)^3\right) + 3f_\alpha\left(0.8\cdot q(1-q)^2(1-p)\right),$$

where $f_\alpha(\cdot)$ is given by

$$f_\alpha\left(x^{(i)}\right) = \begin{cases} \left(x^{(i)}\right)^{1-\alpha}/(1-\alpha), & \alpha > 0 \,\&\, \alpha \neq 1 \\ \log\left(x^{(i)}\right), & \alpha = 1 \end{cases}.$$

We now analyze the benchmarks when the channels are imperfect. When the downlink channels are imperfect, i.e., $e_{down} > 0$, we remark that the benchmarks are the same as the downlink channels are perfect since our feedback recovery mechanism can totally eliminate the detrimental effect of imperfect downlink channels.

When the unlink channels are imperfect, i.e., $e_{up} > 0$, as mentioned in our paper, the unlink channel errors are unavoidable, and the benchmarks for $e_{up} > 0$ is smaller then the corresponding benchmarks for $e_{up} = 0$. Particularly, for the case of "coexistence of one DLMA agent with one TDMA user and one ALOHA user", we can also replace the agent with a model-aware user, and the individual throughput of the model-aware user, TDMA, and ALOHA are $0.8 \cdot p(1-q) \cdot (1-e_{up})$, $0.2 \cdot (1-q) \cdot (1-e_{up})$, and $0.8 \cdot (1-p) q \cdot (1-e_{up})$.

We can find that the optimal strategy (i.e., the value of $p$) of the model-aware user when $e_{up} > 0$ is the same as the strategy when $e_{up} = 0$. The only difference is that the maximum value of $F(p)$ is smaller. Specifically, for maximizing sum throughput, when $q = 0.2$ and $e_{up} > 0$, $p = 1$ and the sum throughput benchmark in Fig. 10 of our paper is $0.8(1-e_{up})$ --- $0.8(1-e_{up}) = 0.72$ when $e_{up} = 0.1$ and $0.8(1-e_{up}) = 0.64$ when $e_{up} = 0.2$. For achieving proportional fairness, , when $q = 0.2$ and $e_{up} > 0$, $p = 0.5$ and the sum-log throughput benchmark in Fig. 11 of our paper is $-5.5 + 3\log(1-e_{up})$ --- $-5.5 + 3\log(1-e_{up}) = -5.81$ when $e_{up} = 0.1$ and $-5.5 + 3\log(1-e_{up}) = -6.17$ when $e_{up} = 0.2$.