

Average Reward

- Baseline ($\alpha = 0.5$)
- Ours ($\alpha \in (0, 1)$)

160k
140k
120k
100k
80k
60k
40k
20k
0

64.98x

DB

5.11x

SUBTB

289.60x

TB

2.24x

FL-DB

1.49x

FL-SUBTB

Objective

