

累积优势

累计(Reward_RL - Reward_Baseline)

300
250
200
150
100
50
0

- vs. Always Wait
- vs. Always Reroute

0 50 100 150 200 250 300 350

决策步

