

# Reinforcement learning and regret bounds for admission control

*Tuesday, June 18, 2024 3:00 PM (20 minutes)*

The expected regret of any reinforcement learning algorithm is lower bounded by  $\Omega(\sqrt{DXAT})$  for undiscounted returns, where  $D$  is the diameter of the Markov decision process,  $X$  the size of the state space,  $A$  the size of the action space and  $T$  the number of time steps. However, this lower bound is general. A smaller regret can be obtained by taking into account some specific knowledge of the problem structure. In this article, we consider an admission control problem to an  $M/M/c/S$  queue with  $m$  job classes and class-dependent rewards and holding costs.

Queuing systems often have a diameter that is at least exponential in the buffer size  $S$ , making the previous lower bound prohibitive for any practical use. We propose an algorithm inspired by UCRL2, and use the structure of the problem to upper bound the expected total regret by  $O(S \log T + \sqrt{mT \log T})$  in the finite server case. In the infinite server case, we prove that the dependence of the regret on  $S$  disappears.

**Presenter:** WEBER, Lucas (Inria)

**Session Classification:** Poster session