RL-QN: A Reinforcement Learning Framework for Optimal Control of Queueing Systems

Bai Liu †, Qiaomin Xie‡, Eytan Modiano †
† Massachusetts Institute of Technology, ‡ Cornell University

Link to the full paper: https://arxiv.org/abs/2011.07401

The rapid growth of information technology has resulted in increasingly complex network systems and poses challenges in obtaining explicit knowledge of system dynamics. For instance, due to security or economic concerns, a number of network systems are built as overlay networks, e.g. caching overlays, routing overlays and security overlays. In these cases, only the overlay part is fully controllable by the network administrator, while the underlay part remains uncontrollable and/or unobservable. The "black box" components make network control policy design challenging.

In this work, we target at optimizing the average queue backlog of a general discrete-time queueing network system with unknown dynamics. We aim to design an optimal control policy that minimizes queue backlog, rather than stabilizing the system.

We consider a discrete time network with the general topology of a directed graph. Each node maintains one or more queues for undelivered packets, and each queue has an unbounded buffer. During each time slot, external data packets arrive at nodes where they are either processed and then depart the system or are relayed. For relayed packets, the communication links can be stochastic, i.e. data transmissions between nodes can fail. We assume the underlying dynamics are partially or fully unknown. This model captures a large class of queueing networks that involve routing, scheduling and switching.

Reinforcement learning is well-suited to learning the optimal control for a system with unknown parameters. We consider model-based reinforcement learning methods, which tend to be more analytically tractable. However, conventional model-based reinforcement learning methods can only be applied to systems with a finite number of states, whereas queueing systems usually have unbounded state space. Fortunately, a vast class of queueing systems have been shown to have stabilizing policies. Therefore, we only apply reinforcement learning techniques on a bounded state space, while simply apply a *known* stabilizing policy to the rest of the states.

We propose Reinforcement Learning for Queueing Networks (RL-QN) that operates in an episodic manner. Each episode conducts either exploration or exploitation, decided by a "coin toss" scheme similar to the decaying ϵ-greedy method. For an exploration episode, we apply a randomized policy that takes action uniformly to obtain samples for the estimation of state-transition probabilities of the queueing system. For an exploitation episode, we apply model-based reinforcement learning techniques. We first compute an estimated optimal policy for the queueing system with bounded state space using the estimated dynamics obtained from exploration episodes. We then apply the estimated optimal policy to the bounded states and the *known* stabilizing policy to the rest of the states throughout the episode.

We rigorously show that, under RL-QN, the asymptotic episodic average queue backlog is upper bounded as

$$\lim_{k\to\infty} \mathbb{E}\left[\frac{\sum_{t=t_{k-1}+1}^{t_k}\sum_i Q_i(t)}{L_k'}\right] \leq \rho^* + \mathcal{O}\left(\frac{U^{1+\max\{2\alpha,\gamma\}}}{\exp(U)}\right),$$

where ρ^* is the expected average queue backlog under the optimal policy, L_k' is the length of episode k, U is the buffer size of the queueing system with bounded state space, α and γ are some positive constants. This conclusion indicates that the long-term episodic average queue backlog approaches the optimum exponentially fast by choosing larger U's.