Contribution ID: **84** Type: **not specified**

Learning payoffs while routing in skill-based queues

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

We consider skill-based routing in queueing systems with heterogeneous customers and servers, where the quality of service is measured by customer-server dependent random rewards and the reward structure is a priori unknown to the system operator. We analyze routing policies that simultaneously learn the system parameters and optimize the reward accumulation, while satisfying queueing stability constraints. To this end, we introduce a model that integrates queueing dynamics and decision making. We use learning techniques from the multi-armed bandit (MAB) framework to propose a definition of regret against a suitable oracle reward and formulate an instance-dependent asymptotic regret lower bound. Since our lower bound is of the same order as results in the classical MAB setting, an asymptotically optimal learning algorithm must exploit the structure of the queueing system to learn as efficiently as in the classical setting, where decisions are not constrained by state space dynamics. We discuss approaches to overcome this by leveraging the analysis of the transient behavior of the queueing system.

Presenter: VAN KEMPEN, Sanne (TU/e)

Session Classification: Poster session