Probability and Mathematical Finance Seminar

Control of Many Server Queueing Systems in Heavy Traffic

Gennady Shaikhet
Carnegie Mellon University

We consider a stochastic queueing model with several service stations and several customer classes, in which a system administrator dynamically controls scheduling and routing. Along with a cost, one faces a stochastic control problem, not necessary Markovian, which is impossible to solve in general. We study this model in a heavy traffic parametric regime, in which both the arrival rate and the number of servers in the system increase to infinity, while keeping, in an appropriate sense, a critically balanced system. Under the scaling limit the queueing model converges to a diffusion model with a singular control term. As it often occurs, solving the diffusion model helps obtaining asymptotic solutions to the queueing model.

In particular, we show that the singular control term can constrain the diffusion to the halfplane corresponding to all queues being empty - a phenomenon that we call null controllability. We show that an analogous, asymptotic result also holds for the original queueing model, by constructing control policies under which, for any given $0 < \varepsilon < T < \infty$, all queues in the system are kept empty on the time interval $[\varepsilon, T]$, with probability approaching one. This introduces a new, unusual heavy traffic 'behavior': On one hand the system is critically loaded; on the other hand, as far as queue lengths are concerned, the system behaves as if it is underloaded.

Monday, September 10, at 5 P.M. in Wean Hall, 6423.