

HEAVY-TRAFFIC APPROXIMATIONS FOR A LAYERED NETWORK
WITH LIMITED RESOURCES

Angelos Aveklouris
Maria Vlasiou
Jiheng Zhang
Bert Zwart

Abstract: Motivated by a web-server model, we present a queueing network consisting of two layers. The first layer incorporates the arrival of customers at a network of two single-server nodes. We assume that the interarrival and the service times have general distributions. Customers are served according to their arrival order at each node and after finishing their service they can re-enter at nodes several times for another service. At the second layer, active servers act as jobs that are served by a single server working at speed one in a processor-sharing fashion. We further assume that the degree of resource sharing is limited by choice, leading to a limited processor-sharing discipline. Our main result is a diffusion approximation for the process describing the number of customers in the system. Assuming a single bottleneck node and studying the system as it approaches heavy traffic, we prove a state-space collapse property.

2010 AMS Mathematics Subject Classification: Primary: 60K25, 90B15; Secondary: 68M20.

Keywords and phrases: Layered queueing network, limited processor sharing, fluid model, diffusion approximation, heavy traffic.

THE FULL TEXT IS AVAILABLE [HERE](#)