

Approximations for multi-server queues with abandonments

Ward Whitt

Department of Industrial Engineering &

Operations Research

Columbia University

New York, NY 10027-6699, USA

Abstract

Two approximations will be discussed. The first approximation is for the $M/GI/s+GI$ model, having general service-time and abandon-time distributions. The model is approximated by the purely Markovian $M/M/s + M(n)$ model, having state-dependent abandonment rates. Further approximations and numerical transform inversion are used to calculate steady-state performance measures, such as the waiting-time distribution and the conditional waiting-time distribution, given that the customer is served (or that the customer abandons).

The second approximation is for the general $G/GI/s + GI$ model. That model is approximated by a deterministic fluid model, which is especially interesting in the Efficiency-Driven (ED) limiting regime, where the traffic intensity ρ is held fixed with $\rho > 1$ as the arrival rate and number of servers increase. Because of the abandonments, the system is stable in the ED regime. Indeed, the ED regime can be realistic for call centers providing low-to-moderate quality of service, which often occurs in service-oriented (non-revenue-generating) call centers. The fluid approximation is supported by a stochastic-process limit in a discrete-time framework.