Optimal Control of Parallel Queues with Batch Service

C. Xia, G. Michalaidis, N. Bambos, and P. W. Glynn

Probability in the Engineering and Informational Sciences. Vol. 16(3), 289-307 (2002)

- XiaMichailidisBambosG02.pdf

We consider the problem of dynamic allocation of a single server with *batch processing* capability to a set of parallel queues. Jobs from different classes cannot be processed together in the same batch. The arrival processes are mutually independent Poisson flows with equal rates. Batches have independent and identically distributed exponentially distributed service times, independent of the batch size and the arrival processes+ It is shown that for the case of infinite buffers, allocating the server to the longest queue, stochastically maximizes the aggregate throughput of the system. For the case of equal-size finite buffers the same policy stochastically minimizes the loss of jobs due to buffer overflows. Finally, for the case of *unequal*-size buffers, a threshold-type policy is identified through an extensive simulation study and shown to consistently outperform other conventional policies. The good performance of the proposed threshold policy is confirmed in the heavy-traffic regime using a fluid model.