THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Tianze Deng
Department of Statistics
The University of Chicago

"Queueing System Design with General Customer Abandonment: Priority Queues versus Dedicated Queues in Moderate Overload"

May 7, 2026, at 1:30 PM
Jones 111, 5747 S. Ellis Avenue

Abstract

The comparison between pooled and dedicated queue structures is a fundamental question in service system design. Cao et al.~\cite{cao} establish, using a fluid model for moderately overloaded systems with exponential customer abandonment, that the dedicated structure combined with join-the-shortest-queue routing can outperform the pooled structure on certain performance measures, such as the probability of meeting a delay target. A natural and practically important question is whether this conclusion persists under general, non-exponential patience distributions --- since empirical patience data from real systems rarely conform to the exponential assumption.

This paper addresses that question. We adopt the fluid modeling framework of Cao et al.~\cite{cao}, extended to general abandonment distributions in the moderately overloaded regime, where the system operates above capacity, but abandonment prevents queues from growing without bound. We establish existence and uniqueness of the fluid model and its invariant state. Our central finding is that the relative performance of the two structures depends on the convexity of the patience distribution CDF F: the DQ-JSQ yields a smaller expected offered waiting time than the PQ when F is convex, and the PQ is preferable when F is concave. These results extend the practical guidance of Cao et al.~\cite{cao} to a significantly broader class of service systems.
