QUEUES WITH DELAY SENSITIVE SERVICE TIMES

Carri W. Chan, Columbia Business School, USA, cwchan@columbia.edu
Vivek F. Farias, Massachusetts Institute of Technology, USA, vivekf@mit.edu
Gabriel Escobar, Kaiser Permanente, USA, gabriel.escobar@kp.org

Mainstream queueing models are frequently employed in modeling healthcare delivery in a number of settings, and further used in making operational decisions for the same. The vast majority of these queueing models assume that the service requirements of a job are independent of the state of the queue upon its arrival. In a healthcare setting, this assumption is equivalent to ignoring the effects of delay experienced by a patient awaiting care. However, it is only natural to conjecture that long delays may have adverse effects on patient outcomes and can potentially lead to longer lengths of stay (LOS) when the patient ultimately does receive care. This work sets out to understand these delay issues from an operational perspective. In particular, we empirically measure how congestion in the Intensive Care Unit (ICU) can lead to delays ICU admission and measure the impact on the patient's ICU LOS. Next, we consider how to incorporate these measured delayed effects into a queueing model and characterize approximations to various quantities of interest when the service time of a job is adversely impacted by the delay experienced by that job. Our findings suggest that this delay effect can be substantial and ignoring it when using queueing models to model healthcare delivery systems may result in significant under-provisioning.