Fitting Time-Varying Queueing Models to Service System Data: Accounting for Dependence in the Arrival and Service Processes

Ward Whitt, Columbia University

Public Abstract

The research objective of this award is to develop methods to fit time-varying non-Markovian multi-server queueing models to data from healthcare systems and customer contact centers, so that the models can be used to analyze and improve the performance; e.g., to set staffing levels. Complex models are needed because: the arrival rates typically vary significantly over time; there is often customer abandonment, which significantly alters performance; service times and customer patience times are usually not exponentially distributed; finally, service systems often need to be represented as a network of queues. Special attention will be given to stochastic variability caused by dependence in the arrival and service processes as well as non-exponential distributions and the time-varying arrival rates. Such dependence can be caused by network structure, including overflows from one queue to another when the first is congested, as occurs in the different wards of a hospital. The dependence ("burstiness") in overflow processes has been found to have a big impact upon performance in telecommunication networks. Estimation will be guided by many-server heavy-traffic limits, revealing key partial characterizations of the model needed to analyze and improve the performance.

If successful, the results of this research will provide the basis for new methods and tools to fit stochastic models to service system data. These models will provide the basis for new methods and tools to design and manage large-scale service systems, such as healthcare systems and customer contact centers, for which there is a recognized need for new methods to improve capacity planning and operational control. The models are also relevant for many other applications, e.g., web server farms processing web-page requests, public housing authorities providing apartments to low-income tenants and financial back offices processing loan applications. Successful results can contribute to better ways to select capacity levels in design decisions and better ways to select staffing levels and to route and schedule customers in operational decisions. Nonstandard methods for treating time-varying arrivals are especially important when service times are relatively long, as in healthcare.