Queueing Systems Modeling with Operational Statistics

J. George Shanthikumar

Richard E. Dauch Chair in Manufacturing and Operations Management
4017 Rawls Hall, 100 S. Grant Street
Krannert School of Management
Purdue University West Lafayette, IN 47907
Phone: (765) 494-4468
Email: shanthikumar@purdue.edu

Abstract

Classical modeling of queueing systems starts with a set of statistical assumptions such as independent and identically distributed (*i.i.d*) inter-arrival times, *i.i.d* service times and the assumption that these two sequences are mutually independent. Based on further assumptions such as exponentially distributed inter-arrival times, or in the absence of further assumptions, exact or approximate performance measures are obtained. For the G/G/c queueing system, for example, the following approximation for the mean waiting time can be found in the standard literature.

$$\mathbb{E}[W_q] \approx \frac{\rho^{\sqrt{2c-1}}}{1-\rho} \times \frac{c_a^2 + c_s^2}{2} \times \mathbb{E}[S].$$

When applying these results to real queueing systems, one would estimate the mean and variance of the inter-arrival times and the effective service times (acounting for the possibility of service interuptions due to machine failures, maintenance etc.) or fit distributions for the inter-arrival and service times. Often however application of the above formula, for example, with the estimated parameters or simulating the G/G/c system with the fitted distribution functions provide results that substantially over-estimate the observed average waiting time. Some of the emperical evidence points to the fact that such approximations are better for queueing systems that are poorly run. In a way, it can be said that if the formulas or simulation give a good estimate of the actual performance, then it must be a very poorly run queueing system.

Using ideas from *operational analysis* of queueing systems, we will discuss the different ways we can use the actual data to develop models for real queueing systems. This will be accomplished using Operational Statistics. The difference between the estimates obtained using classical queueing models and the actual performance will be expalined through *queue management efficiency* and *coordination efficiency*. We will illustrate the applications of this type of modeling to "distorted," real data from a semiconductor manufacturing system.