Capacity and Efficiency Analysis of Multiserver-Queuing Systems
H. Zhang

Abstract

This thesis studies mean performances of various queuing systems with different number of servers and service rates. Specifically we investigate how mean queue length or mean waiting time performs between systems with one server but fast service rate and with two or more servers but slow service rate. We address this problem with simple M/M/c queuing systems and then study more general phase-type systems. Analytical analyses of mean performance and their numerical analyses are presented.

For the analysis of phase-type systems we discuss the theory of quasi-birth-and-death processes with matrix-geometric stationary distributions. In particular, we present a detailed analysis of the waiting time distribution in multi-server systems with phase type service distributions. As a canonical example we especially focus on the two-server model with two-phase service distribution and discuss how to generalize to multiple servers and more phases in the service distribution.

This thesis will in the end give the main conclusion that fast single-server queuing system is not always the best (compared with the slow multi-server) with more general service distributions than exponential. The result is of practical interest on making dimensioning decisions.