Case problem for Abacus: Cooperation in queueing networks

Judith Timmer
Stochastic Operations Research group
j.b.timmer@utwente.nl

15 January 2019

1 Introduction

The mission of the Stochastic Operations Research (SOR) group is to conduct mathematical education and research of high international standards in the areas of stochastic processes and mathematics of operations research, to contribute to the development of mathematics in a multidisciplinary engineering environment and to contribute to a better understanding and functioning of our increasingly complex society.

Research in the SOR group covers a wide range of topics. Typical research themes are communication systems (a.o. energy minimization in wireless networks), social networks and big data (a.o. making our society more efficient by acting upon social interactions), health care logistics (a.o. improving treatment and diagnostics planning) and production & logistics (a.o. optimizing inventory management).

This case is concerned with the performance of a queueing network, in which jobs arrive at several queues. You may think of applications like tasks in a cloud-computing environment, messages in a communication network or products in a logistics network. The goal of this case is to show that cooperation among the queues may improve the network's performance.

1. In the sequel, replace the terms 'queue' and 'job' by those of your preferred application: 'cloud' and 'task', 'computer' and 'message' or 'factory' and 'product', respectively.

1.1 Preliminaries: $M/M/1$ queues

For this case we need basic knowledge of $M/M/1$ queues, which is part of AM module 8. Therefore we start with a refresher or crash course on this matter.

An $M/M/1$ queue is illustrated in figure 1. A queue consists of a waiting line and a server. The

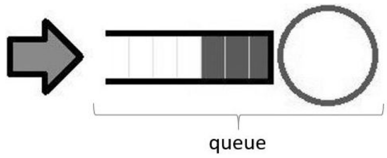

Figure 1: An $M/M/1$ queue.

server is represented by the circle and the rectangular shape in between the arrow and the circle represents the waiting line. Jobs arrive to a queue, as indicated by the arrow on the left. If another job is being served, the job joins the waiting line. Else, it immediately enters service. Jobs are served on a first come, first served basis.

Basic characteristics of such a queue are as follows.

- The interarrival times of jobs are independent exponentially distributed random variables with parameter λ (the arrival rate), $\lambda > 0$.
- The service times of jobs are independent exponentially distributed random variables with parameter μ (the service rate), $\mu > \lambda$.
- Let π_j be the steady state probability of j jobs being present in the $M/M/1$ queue (waiting in line or being served). It can be shown that $\pi_j = \rho^j(1-\rho)$ where $\rho = \lambda/\mu$, $j = 0, 1, 2, \ldots$.

2. Let $L = \sum_{j=0}^{\infty} j\pi_j$ be the average number of jobs present in the queue. Show that $L = \lambda/(\mu - \lambda)$.

2 Case: Cooperation in queueing networks

Consider a small queueing network with two queues. Let λ_i and μ_i be the arrival and service rate, respectively, of queue i, $i = 1, 2$. The cost of queue i is defined to be its average number of jobs present in euros, $\lambda_i/(\mu_i - \lambda_i)$.

The queues are allowed to cooperate to improve upon their joint cost $\sum_i \lambda_i/(\mu_i - \lambda_i)$. Cooperation means that the queues may share their service rates. This should be done in such a way that the joint cost after sharing service rates is as small as possible subject to the conditions that the total service rate should be used and each queue's arrival rate should be smaller than its service rate after sharing.

Once the optimal joint cost is obtained, these costs may be allocated to the queues. A fair cost allocation satisfies two criteria: (i) the optimal joint cost should be completely allocated, and (ii) the allocated cost to a queue should be less than or equal to its own cost.

The main questions are: Does cooperation lead to cost savings? Does there exist a fair cost allocation? How should the cost be allocated? In this case you will investigate these questions.

3. Pick your own parameter values λ_1, λ_2, μ_1 and μ_2 of the two queues such that the arrival rate falls below the service rate, $\lambda_i < \mu_i$ for all i.
4. What is the minimal joint cost of the two queues? How should the service rates be shared to arrive at this cost? Does cooperation save costs?
5. Does there exist a fair allocation of the optimal joint cost? If so, what is a fair allocation of this joint cost to the queues?
6. Now consider general arrival rates λ_i and service rates μ_i, $\mu_i > \lambda_i$, for $i = 1, 2$. Answer the questions 4-5 for this general setting.
7. Is it possible that a queue may get paid (is allocated negative cost) to cooperate? If so, provide an example. If not, explain why not.

If time permits, you may also try the questions below.

8. Extend your analysis of section 2 to a network with three or more queues. Repeat the questions. Extend the fairness-criterion for a cost allocation as follows: the cost allocated to any group of queues should be less than or equal to its optimal cost.
9. Repeat your analysis in section 2 with another objective function: minimize $\sum_i \lambda_i/\mu_i$ with λ_i/μ_i the traffic intensity of queue i. Compare the results. Discuss the differences and similarities.