UNIVERSITEIT VAN AMSTERDAM

Queueing Systems With Nonstandard Input Processes
D.T. Koops

Summary

This thesis considers queueing systems with nonstandard arrival processes, ranging from doubly stochastic Poisson processes driven by shot noise (Chapter 3), and Hawkes processes (Chapter 4), to nonhomogeneous Poisson processes (Chapter 5) and Lévy-driven fluid queues (Chapters 6 and 7). The primary focus of this thesis is performance evaluation of the corresponding queueing systems. Chapter 5 stands out, as the problem considered there is a statistical inversion problem.

In Chapter 2 we introduce shot-noise processes. In a single-server queue with Poisson arrivals, where the service rate depends linearly on the amount of work in the system, a shot-noise process arises naturally in the form of a workload process. We derive some preliminary results on the number of customers in such a queue. We gradually increase the complexity of the analysis, from one-dimensional shot noise to Markov modulated shot-noise networks. The results are closely related to a series of papers of Kella and co-authors, who derived similar results using a different approach. The results and insights obtained in this chapter led to a paper [19] about rare-event simulation, and to Chapter 3.

In Chapter 3 we recognize that shot-noise processes are not only useful to model the workload in a server with workload-dependent service rate, but they can also play the role of a stochastic arrival rate. Specifically, we consider an infinite-server queue for which the arrival process is a Cox process (i.e., a doubly stochastic Poisson process), of which the arrival rate is given by shot noise. Such shot-noise Cox processes are natural models for arrival rates that tend to display unpredictable jumps at Poisson epochs, after which the arrival rate reverts to lower values. We perform transient analysis on the number of customers in an infinite-server queue, jointly with the value of the driving shot-noise process. Additionally, we derive heavy-traffic asymptotics for the number of customers in the system by using a linear scaling of the shot intensity and the number of jobs. First we focus on a one-dimensional setting in which there is a single infinite-server queue, which we then extend to a network setting.

In Chapter 4 we study the number of customers in infinite-server queues with a self-exciting (Hawkes) arrival process. The Hawkes arrival rate has a similar stochastic structure as a shot-noise process. The difference is that for shot-noise processes the jumps occur at Poisson epochs, whereas for the Hawkes

process the jumps occur simultaneously with arrivals to the queue. The different timing of jumps also leads to a different interpretation for the cause of temporary increases in demand: for shot-noise Cox processes the cause is exogeneous and unpredictable, whereas for the Hawkes process the current demand is a function of past demand. This endogeneity is also referred to as a feedback loop or self-excitation. We study a Markovian and non-Markovian case separately. If one assumes that service requirements are exponentially distributed and that the kernel of the Hawkes rate decreases exponentially, then the queueing process is Markovian. We obtain a system of differential equations that characterizes the joint distribution of the arrival intensity and the number of customers. Moreover, we provide a recursive procedure that results in explicit (transient and stationary) moments. Subsequently, we allow for non-Markovian Hawkes arrival processes and non-exponential service times. By viewing the Hawkes process as a branching process, we find that the probability generating function of the number of customers in the system can be expressed in terms of the solution of a functional fixed-point equation. We also include various asymptotic results: in case of heavy-tailed intensity jumps, we find that the number of customers is also heavy tailed with the same index. In addition, we consider a heavy-traffic regime. We conclude the paper by discussing how our results can be used computationally and by verifying the numerical results via simulations.

In Chapter 5 we consider an infinite server queue with general service times, now with a nonhomogeneous Poisson arrival process. This means that the arrival rate is allowed to vary deterministically in time. This chapter provides a mathematical framework for estimating linear functionals of the service time distribution, such as the distribution function of the service time in a point and the expected service time. The main difficulty is the partial information: we suppose that only the number of busy servers is observed over time, but we cannot distinguish the jobs. The problem is reduced to a statistical deconvolution problem, which is solved by using Laplace transform techniques and kernels for regularization. Upper bounds on the mean squared error of the proposed estimators are derived. Some concrete simulation experiments are performed to illustrate how the method can be applied and to provide insight in the actual performance.

In Chapter 6 we study the stationary workload distribution of a fluid tandem queue in heavy traffic. We consider different types of Lévy input, covering compound Poisson, α-stable Lévy motion (with $1 < \alpha < 2$), and Brownian motion. In our analysis, we separately deal with Lévy input processes with increments that have finite and infinite variance. A distinguishing feature of this chapter is that we do not only consider the usual heavy-traffic regime, in which the load at one of the nodes goes to unity, but also a regime in which we simultaneously let the load of both servers tend to one, which, as it turns out, leads to entirely different heavy-traffic asymptotics. Numerical experiments indicate that under specific conditions the resulting simultaneous heavy-traffic approximation significantly outperforms the usual heavy-traffic approximation.

Finally, in Chapter 7 we consider the problem of estimating transaction times in the Bitcoin blockchain. When making a Bitcoin transaction, the user may include a fee. Our aim is to estimate the distribution of the time until the transaction ends up in the blockchain (i.e., the confirmation time) as a function of the fee that is paid by the user. The quantification of the trade-off between fee and waiting time allows users to make a proper consideration. The mempool in a Bitcoin blockchain behaves similar to a queueing system with batch departures at Poisson epochs. We argue that the time until a Bitcoin transaction is confirmed can be modelled as a particular stochastic fluid queueing process (to be precise: a Cramér-Lundberg process). We approximate the queueing process in two different ways. The first approach leads to a lower bound on the confirmation probability, which becomes increasingly tight as traffic decreases. The second approach relies on a diffusion approximation with a continuity correction, which becomes increasingly accurate as traffic intensifies. The accuracy of the approximations under different traffic loads is evaluated in a simulation study.