Performance Evaluation 79 (2014) 146–159

Contents lists available at ScienceDirect

Performance Evaluation

journal homepage: www.elsevier.com/locate/peva

Perfect sampling for closed queueing networks

Anne Bouillard [a,*], Ana Bušić [b], Christelle Rovetta [b]

[a] ENS Paris, 45 rue d'Ulm 75005 Paris, France
[b] Inria, 23 avenue d'Italie, 75013 Paris, France

ARTICLE INFO

Article history:
Available online 10 July 2014

Keywords:
Queueing networks
Simulation
Coupling from the past

ABSTRACT

In this paper we investigate coupling from the past (CFTP) algorithms for closed queueing networks. The stationary distribution has a product form only in a very limited number of particular cases when queue capacity is finite, and numerical algorithms are intractable due to the cardinality of the state space. Moreover, closed networks do not exhibit any monotonic property enabling efficient CFTP. We derive a bounding chain for the CFTP algorithm for closed queueing networks. This bounding chain is based on a compact representation of sets of states that enables exact sampling from the stationary distribution without considering all initial conditions in the CFTP. The coupling time of the bounding chain is almost surely finite, and numerical experiments show that it is close to the coupling time of the exact chain.

1. Introduction

One reason for the popularity of Markovian representations of queueing networks is that they admit a product-form stationary distribution under general conditions. This structure is no longer guaranteed when the queues have a finite capacity, so that an exact analysis may not be computationally tractable. When the stationary distribution cannot be computed then we may turn to approximations or simulation.

This paper concerns simulation, with a focus on stopping criteria. The asymptotic variance appearing in the Central Limit Theorem has been the most common metric to devise stopping rules, while mixing times have become a standard alternative [1,2]. Unfortunately, there are no generic and tractable techniques to compute or bound either the asymptotic variance or the mixing time for non-reversible Markov chains.

In the 1990s, Propp and Wilson introduced a method for sampling a random variable according to the stationary distribution of a finite ergodic Markov chain [3]: the coupling from the past (CFTP) algorithm. The CFTP algorithm automatically detects and stops when the sample has the correct distribution. In this way it is possible to generate i.i.d. samples from the chain, and the asymptotic variance of the resulting simulator is the standard variance of the random variable whose mean we wish to estimate.

The number of steps required in the CFTP algorithm is proportional to the coupling time of the chain, but the time complexity strongly depends on the complexity of the one-step transition of the chain. In particular, for closed queueing networks, no efficient CFTP method has been proposed previously, with the exception of networks with a product form distribution [4].

Different techniques can be used to efficiently compute one step of the CFTP algorithm: the simplest solution, for monotone Markov chains, is to compute the minimal and maximal trajectories only [3]. For Markov chains with no monotone

* Corresponding author.
E-mail addresses: anne.bouillard@ens.fr (A. Bouillard), Ana.Busic@inria.fr (A. Bušić), christelle.rovetta@inria.fr (C. Rovetta).

http://dx.doi.org/10.1016/j.peva.2014.07.010
0166-5316/

representations, new techniques have been developed to approximate each step of the computation, at the cost of slightly increasing the number of iterations of the algorithm. Bounding chains have been constructed to detect coalescence for state spaces with lattice structure [5,6], or for models with short range local interactions, such as interacting particle systems [7]. For applications of [5] to queueing networks, see for instance [8,9].

The main difficulty with closed queueing networks is that the customer population is constant. This imposes a global constraint on the model, so the approach of [5] cannot be applied directly. Without monotonicity, the complexity of one iteration of the original CFTP algorithm by Propp and Wilson [3] depends on the cardinality of the state space, which is exponential in the number of queues.

In this paper, we develop an effective CFTP algorithm for closed queueing networks. Our main contribution is a new technique for constructing bounding chains, which is adapted to a large class of Markovian closed queueing networks. It is based on a compact representation for sets of states, using diagrams that we introduce in Section 3. We perform the CFTP algorithm on the space of the diagrams, for which the one-step transition is simpler to compute than using the original state space. For the diagrams, one iteration in the CFTP algorithm can be computed in $O(KM^2)$ time, as we discuss in Section 3.2.3.

The paper is organized as follows. In Section 2 we present the queueing model and discuss the solutions that have been proposed in the literature for special cases. The analysis is based on a *diagram representation* introduced in Section 3. This is used in Section 4 to prove that the CFTP algorithm terminates in finite time, almost surely. Finally, Section 5 contains results from numerical, comparing our algorithm with the classical CFTP in terms of the number of iterations. Note that classical CFTP can be used only for very small models, as its one-step transition depends on the cardinality of the state space. Final remarks and conclusions are contained in Section 6.

2. Model and background

2.1. Presentation of the model

We denote by $\mathbb{N}$ the set of non-negative integers and by $e_i \in \mathbb{N}^k$ the vector such that $(e_i)_j = 1$ if $i = j$ and 0 otherwise.

Consider a closed network of $\cdot/M/1/C$ queues with M customers. We denote by $Q = \{1, \dots, K\}$ the set of queues. For $i \in Q$, μ_i is the service rate and C_i the capacity of queue i. After a service in queue i, a customer is routed to queue j with probability $p_{i,j}$, independently of the current state and past evolution of the network. If queue j is full, the customer is blocked at queue i for another service time. Let $P = (p_{i,j})_{i,j\in Q}$ denote the routing probability matrix; for all $i, j \in Q$, $p_{i,j} \geq 0$ and for all $i \in Q$, $\sum_{j\in Q} p_{i,j} = 1$.

The evolution of this network follows a continuous time Markov chain on the state space

$$\mathcal{S} = \left\{ x = (x_1, x_2, \dots, x_K) \in \mathbb{N}^K \ : \ \sum_{i=1}^{K} x_i = M \quad \text{and} \quad 0 \leq x_i \leq C_i, \ \forall i \in Q \right\}.$$

The upper bound for $|\mathcal{S}|$ is given by $\binom{K+M-1}{K-1}$; this is the exact value for $|\mathcal{S}|$ if all queues have infinite capacity.

The topology of the network can be represented by a directed graph $G = (Q, R)$ where $R = \{(i,j) \ : \ p_{i,j} > 0\}$. As we consider closed networks, without loss of generality, we can assume that G is strongly connected.

For $(i,j) \in R$, we denote by $t_{i,j} \ : \ \mathcal{S} \to \mathcal{S}$ the function that describes routing of a customer from queue i to queue j:

$$t_{i,j}(x) = \begin{cases} x - e_i + e_j & \text{if } x_i > 0 \text{ and } x_j < C_j, \\ x & \text{otherwise.} \end{cases}$$

This transition does not affect the state if queue i is empty or if queue j is full.

We consider a discrete time Markov chain, obtained using uniformization with constant $\sum_{i=1}^{K} \mu_i$. This chain has the same state space as the continuous time chain.

A functional representation of this discrete time Markov chain can be given using functions $t_{i,j}$ and routing matrix P. Denote by $(U_n)_{n\geq 1}$ an i.i.d. sequence of random variables with distribution

$$P(U_1 = (i,j)) = \frac{\mu_i}{\sum_{j\in Q} \mu_j} p_{i,j}.$$

Define $F : \mathcal{S} \times R$ as

$$F(x, (i,j)) = t_{i,j}(x).$$

Let X_0 be the initial state, independent of $(U_n)_{n\geq 1}$, and $X_{n+1} = F(X_n, U_{n+1}), \ n \in \mathbb{N}$.

2.2. Coupling from the past algorithm

The coupling from the past (CFTP) algorithm has been first introduced in 1996 by Propp and Wilson [3]. The key idea is to sample a value at time 0 of a trajectory that starts arbitrary far in the past.