日本オペレーションズ・リサーチ学会
2004年秋季研究発表会

A Fast Approximation for General Closed Queueing Networks

01105381 北海道大学 *木村 俊一 KIMURA Toshikazu
北海道大学 森岡 和行 MORIOKA Kazuyuki

1 Introduction

Queueing networks are important models in the performance analysis of complex systems such as computer/communication networks and flexible manufacturing systems. Efficient algorithms have been developed for computing performance measures of separable networks that have a *product-form* solution. Unfortunately, most of real systems are not separable, which makes approximate analysis based on decomposition a practical necessity.

The idea of decomposition approximation has been successfully applied to open queueing networks with general service times. For closed queueing networks, however, the situation becomes more complex due to the constant population constraint in the network. There are two well-known methods for non-separable closed queueing networks, *i.e.*, the *aggregation method* [1] and the *exponentialization approach* [2, 3]. The common idea of these methods is to transform the original network into an approximately equivalent exponential network, where each station has exponential service times with state-dependent rates. In the aggregation method, this transformation depends only on the mean service time of each station, so that it fails for networks with general service times. On the other hand, the idea of the exponentialization approach is to analyze each station with a state-dependent arrival rate and a service rate equal to conditional throughput, which enables us to take account of general service times. It has been known that this approach provides sufficiently accurate results, but that the computational load is very heavy.

In this paper, using a diffusion approximation for state-dependent queues with finite capacity, we modify the exponentialization approach to develop a fast approximation for a general class of closed queueing networks.

2 General Closed Queueing Networks

Consider a closed queueing network in equilibrium with M service stations and N customers, in which we assume that

1. the routing probability p_{ij} that a customer leaving station i enters station j is independent of the state of the system $(i, j = 1, \ldots, M)$;
2. customers are served under the first-come first-served discipline at all stations;
3. service times of customers at station i are iid with a general distribution and independent of the arrival process at station i $(i = 1, \ldots, M)$;
4. station i has s_i (≥ 1) identical servers in parallel and a limited local buffer with capacity r_i (≥ 0) $(i = 1, \ldots, M)$, and hence the maximum number of customers allowed at station i is $n_i \equiv \min(s_i + r_i, N)$, where $\sum_{i=1}^{M} n_i > N > \max_i s_i$.

The system is further specified by the following notations: For each server at station i $(i = 1, \ldots, M)$, let F_i be the service-time cdf with finite mean μ_i^{-1} and let c_i^2 be the squared coefficients of variation of F_i. Let N_i denote the number of customers at station i and let $p_i(n) = P(N_i = n)$ $(i = 1, \ldots, M,\ n = 0, \ldots, n_i)$. The problem we focus on here is to obtain all of the marginal distributions $\{p_i(n)\}$.

3 The Exponentialization Approach

Let $\nu_i(n)$ be *equivalent* service rate at station i in the equivalent exponential network when $N_i = n$ $(i = 1, \ldots, M,\ n = 0, \ldots, n_i)$, where $\nu_i(0) = 0$ for all i. Also let $p_i^*(n)$ denote the marginal probability of having n customers at station i in the exponential network characterized by the set of service rates $\{\nu_i(n)\}$. Obviously, the success of the exponentialization approach strongly depends on how to approximate $\{\nu_i(n)\}$ for which $p_i(n) = p_i^*(n)$ $(i = 1, \ldots, M,\ n = 0, \ldots, n_i)$. The exponentialization approach can be summarized by the following algorithm:

Step 0 For $i = 1, \ldots, M$ and $n = 0, \ldots, n_i$, set $\mu_i(n) := \min(n, s_i)\mu_i$.

Step 1 Solve the exponential network characterized by the set of service rates $\{\mu_i(n)\}$ and the routing matrix $P = (p_{ij})$ to obtain the marginal distribution $\{p_i^*(n)\}$. For $i = 1, \ldots, M$, set

$$\lambda_i(n) := \begin{cases} 0, & n = n_i \\ \dfrac{p_i^*(n+1)}{p_i^*(n)}\,\mu_i(n+1), & \\ & 0 \le n \le n_i - 1. \end{cases}$$

Step 2 For $i = 1, \ldots, M$, analyze station i as an isolated $M(n)/G/s_i/n_i$ queue having Poisson arrivals with state-dependent arrival rates $\{\lambda_i(n)\}$ and the service-time cdf F_i, obtaining its steady-state distribution $\{\pi_i(n); n = 0, \ldots, n_i\}$ as an approximation for $\{p_i(n)\}$.

Step 3 For $i = 1, \ldots, M$, set

$$\nu_i(n) := \begin{cases} 0, & n = 0 \\ \dfrac{\pi_i(n-1)}{\pi_i(n)}\,\lambda_i(n-1), & 1 \le n \le n_i. \end{cases}$$

If $\max_{i,n} |\mu_i(n) - \nu_i(n)| < \varepsilon$ for a given error bound $\varepsilon > 0$, then $p_i(n) := \pi_i(n)$ for all i and n, and stop; otherwise set $\mu_i(n) := \nu_i(n)$ for all i and n, and go to Step 1.

In each iteration of Step 2, Marie [2] proposed to calculate $\{\pi_i(n)\}$ exactly by using the Coxian service-time distribution. Clearly, this increases the computational time significantly. In this paper, we will simplify the calculation in Steps 2 and 3 by using a diffusion approximation for the $M(n)/G/s$ queue with finite capacity. The simplification makes the exponentialization approach be more tractable as a quick modeling tool for performance evaluation.

4 Diffusion Approximation for $\{\nu_i(n)\}$

For $i = 1, \ldots, M$ and $n = 1, \ldots, n_i$, let

$$a_i(n) = \lambda_i(n-1) + \min(n, s_i)\mu_i d_i^2(n),$$

$$b_i(n) = \lambda_i(n-1) - \min(n, s_i)\mu_i,$$

$$d_i^2(n) = 1 + \mathbf{1}_{\{n \ge s_i\}}(n)(1 - p_i^*(0))(c_i^2 - 1),$$

$$\gamma_i(n) = \exp\left\{\frac{2b_i(n)}{a_i(n)}\right\},$$

$$\xi_i(n) = (\gamma_i(1) - 1)\prod_{k=2}^{n} \gamma_i(k),$$

Also, for $i = 1, \ldots, M$, let

$$\alpha_i = \lambda_i(0) \quad \text{and} \quad \beta_i = s_i\mu_i.$$

Then, the diffusion approximation for $\{\pi_i(n)\}$ is given by

$$\pi_i(n) = \begin{cases} \pi_i(0)\dfrac{\alpha_i}{b_i(1)}\xi_i(n), & 1 \le n \le n_i - 1 \\ \pi_i(0)\dfrac{\alpha_i}{\beta_i}\dfrac{b_i(n_i)}{b_i(1)}\dfrac{\xi_i(n_i)}{\gamma_i(n_i) - 1}, & n = n_i, \end{cases}.$$

from which we can explicitly obtain $\nu_i(n)$ in Step 3 as

$$\nu_i(n) = \begin{cases} 0, & n = 0 \\ \dfrac{b_i(1)}{\gamma_i(1) - 1}, & n = 1 \\ \dfrac{\lambda_i(n-1)}{\gamma_i(n)}, & 2 \le n \le n_i - 1 \\ \dfrac{\beta_i}{b_i(n_i)}\dfrac{\gamma_i(n_i) - 1}{\gamma_i(n_i)}\lambda_i(n_i - 1), & n = n_i. \end{cases}$$

References

[1] Avi-Itzhak, B. and Heyman, D.P., "Approximate queueing models for multiprogramming computer systems," *Operations Research*, **21** (1973) 1212–1230.

[2] Marie, R.A., "An approximate analytical method for general queueing networks," *IEEE Transactions on Software Engineering*, **5** (1979) 530–538.

[3] Yao, D.D. and Buzacott, J.A., "The exponentialization approach to flexible manufacturing system models with general processing times," *European Journal of Operational Research*, **24** (1986) 410–416.