Asymptotic analysis for closed multiclass queueing networks in critical usage*

Jie Wang and Keith W. Ross

Department of Systems, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 1 June 1993; revised 26 October 1993

We consider a class of closed multiclass queueing networks containing First-Come-First-Serve (FCFS) and Infinite Server (IS) stations. These networks have a product-form solution for their equilibrium probabilities. We study these networks in an asymptotic regime for which the number of customers and the service rates at the FCFS stations go to infinity with the same order. We assume that the regime is in critical usage, whereby the utilizations of the FCFS servers slowly approach one. The asymptotic distribution of the normalized queue lengths is shown to be in many cases a truncated multivariate normal distribution. Traffic conditions for which the normalized queue lengths are *almost* asymptotically independent are determined. Asymptotic expansions of utilizations and expected queue lengths are presented. We show through an example how to obtain asymptotic expansions of performance measures when the networks are in mixed usage and how to apply the results to networks with finite data.

Keywords: Queueing networks, asymptotic analysis, critical usage.

1. Introduction

Unless the number of classes and stations are very small, closed multiclass product-form queueing networks are all but impossible to analyze by combinatorial methods. An alternative solution procedure is asymptotic analysis. Basically there are two types of asymptotic results for closed queueing networks in the existing literature: results for the asymptotic *distributions* of queue lengths and those for the asymptotic *expansions* of performance measures of the networks.

The first asymptotic study was carried out by Gordon and Newell [5], who explored the marginal distribution of queue lengths for a single-class queueing network as the number of customers approaches to infinity. Pittel [12] obtained an important result for the asymptotic distribution of the queue lengths and the free capacities for closed exponential networks with restricted queue capacities. He classified the service stations into two different groups – namely, those *rarely visited* and those *heavily loaded* – according to the solution of a nonlinear

* Supported partially by NSF grant NCR93-04601.

programming problem. These two groups correspond to *normal usage* and *heavy usage* studied by Mitra [10]. Pittel proved, under an asymptotic regime, that the limiting distribution of the population at a rarely visited service station coincides with the stationary distribution for an isolated queue; and that the normalized version of the queue lengths at heavily loaded stations has an asymptotic distribution of a truncated multivariate normal. However, his study does not cover *critical usage* because of the second part of Assumption 8 in Pittel [12]. Whitt [17] also studied the asymptotic distribution of the queue lengths for closed queueing networks. In particular, he considered a closed network consisting of FCFS stations and one IS station and showed that the stochastic process describing the subnetwork consisting of only the FCFS stations, under the *normal usage* condition, approaches that of an open network when the number of customers and the traffic intensity at the IS station tend to infinity with the same order. He proved the result for single-class networks and multiclass networks with a uniformly tightness restriction.

McKenna, Mitra and Ramakrishnan [7–9, 14] developed the asymptotic expansions of the normalization constant and performance measures for multiclass queueing network in *normal usage*. The basic idea is to express the normalization constants and performance measures as a power series of $1/N$, where N is a large parameter characterizing the network size; the coefficients of the series are easily calculated since they are the normalization constants of small networks. The software package PANACEA, developed at Bell Labs, is based on this asymptotic expansion theory and can solve many large networks for which the conventional combinatorial methods are inapplicable. However, the asymptotic expansions only work well when the *normal usage* condition is satisfied with some margin. In particular, the performance of the method deteriorates as the utilization of any processing node becomes close to unity as pointed out in Ross et al. [15]. Less progress has been made for asymptotic expansions for networks not in normal usage. Morrison and Mitra [11] obtained some results for the asymptotic expansions of the waiting time for a network in critical usage, but the network consists of only one PS station and one IS station. Another study done by Mitra [10] is for a simple closed network with cyclic routing, in which case the normalization constant can be expressed in terms of a one-dimensional integral; he developed asymptotic expansions for all three typical traffic conditions – normal usage, critical usage and heavy usage – and obtained the leading terms of the expansions for utilizations, throughputs and queue lengths.

Networks in critical usage have not been thoroughly studied in an asymptotic context. In an earlier paper, Ross and Wang [16], we studied the application of Monte Carlo integration to queueing networks in critical usage and found the asymptotically optimal importance sampling distribution.

In this paper, we study the asymptotic behavior of multiclass queueing networks in critical usage. The networks we consider consist of IS stations and single-server FCFS stations. We examine both the asymptotic distribution of queue lengths at the FCFS stations and the network performance measures when the parameter N, which characterizes the population sizes and processing rates, goes to infinity. One of our major tools is the integral representation for the normalization constant of the product-form queueing networks discovered by McKenna and Mitra [7].

This paper has basically three contributions. First we derive the asymptotic distribution of the normalized queue lengths for networks in critical usage and specify traffic patterns that lead to approximately asymptotic independence among the normalized queue lengths. Second, again under the critical usage condition, we determine the asymptotic expansions of performance measures for networks which are more general than those studied in Morrison and Mitra [11] and Mitra [10]. Third, we illustrate by an example how to obtain asymptotic expansions when networks are in mixed usage, i.e., some of the FCFS stations are in normal usage and the rest are in critical usage.

Although the networks in our model consist of FCFS stations and IS stations only, the results we have obtained are based on the product-form solution and integral representation, hence, as indicated by McKenna and Mitra [8], they are applicable to networks with IS stations and any types 1, 2, and 4 service stations.

The paper is organized as follows. In section 2 we briefly introduce the queueing network model studied, the asymptotic regime, and the notion of critical usage in the asymptotic regime. In section 3 we derive the asymptotic distribution for the queue lengths. In section 4 we establish the asymptotic expansions for normalization constants. We derive the asymptotic expansions for performance measures in section 5. In section 6 we discuss how to apply the asymptotic expansion results to networks with finite data. We present one example showing how one can derive parameters for the asymptotic regime from the data of a network and how approximations of performance measures can be obtained by applying asymptotic expansion results when the network is in mixed usage.

2. Preliminaries

We consider closed multiclass queueing networks consisting of First-Come-First-Serve (FCFS) stations and Infinite-Server (IS) station. We are mainly interested in the network behavior associated with the FCFS stations. Without loss of generality, we can assume that there is only one IS station, as noted by Mitra [13]. Each FCFS station has a single server. Let J denote the number of classes and L denote the number of FCFS stations. Let the population size for the jth class be denoted by N_j. The service times at the FCFS stations are exponentially distributed and do not depend on the class; denote $1/\mu_l$ for the mean service time at FCFS station l. The service times at the IS station have an arbitrary distribution; denote $1/\mu_{j0}$ for the mean service time of a class-j customer at the IS station. For each class j, let λ_{jl} be the relative visit ratio of a customer to service station l, which are the eigenvectors corresponding to eigenvalue 1 of routing matrix for each class. Let $\rho_{jl} := \lambda_{jl}/\mu_l$, $l = 1, \ldots, L$ and $\rho_{j0} := \lambda_{j0}/\mu_{j0}$. The state of the system is denoted by $\boldsymbol{n} := (n_{jl} : 1 \leq j \leq J, 0 \leq l \leq L)$, where n_{jl} denotes the number of class-j customers at service station l. The set of all possible states is given by

$$\Omega := \{\boldsymbol{n} : n_{j0} + n_{j1} +, \ldots, + n_{jL} = N_j, \quad j = 1, \ldots, J\}.$$

Let $n_l := n_{1l} + \cdots + n_{Jl}$ be the number of customers present at service station l. It is well known that the equilibrium probability being in state $\boldsymbol{n} \in \Omega$ is given by the following product form (see for example Baskett and et al. [1]):

$$\pi(\boldsymbol{n}) = \frac{f(\boldsymbol{n})}{g},$$

where

$$f(\boldsymbol{n}) = \left[\prod_{j=1}^{J} \frac{\rho_{j0}^{n_{j0}}}{n_{j0}!}\right] \left[\prod_{l=1}^{L} n_l! \prod_{j=1}^{J} \frac{\rho_{jl}^{n_{jl}}}{n_{jl}!}\right],$$

and

$$g = \sum_{\boldsymbol{n} \in \Omega} f(\boldsymbol{n}).$$

Many performance measures of interest can be expressed as simple functions of similar multidimensional sums. In order to be more specific, define $\Omega_{(j)}$ to be the state space associated with the queueing network that is identical to the original network except there is one less class-j customer. Further define

$$g^l := \sum_{\boldsymbol{n} \in \Omega} n_l f(\boldsymbol{n}),$$

$$g_j := \sum_{\boldsymbol{n} \in \Omega_{(j)}} f(\boldsymbol{n}),$$

$$g_j^l := \sum_{\boldsymbol{n} \in \Omega_{(j)}} n_l f(\boldsymbol{n}).$$

With these notations, we have the following simple expressions for the performance measures (see Bruell and Balbo [4], McKenna and Mitra [8]). The average utilization of the server by class-j customers at FCFS station l is

$$\mathrm{UTIL}_{jl} = \rho_{jl}\frac{g_j}{g};$$

the throughput of class-j customers at service station l is

$$TH_{jl} = \lambda_{jl}\frac{g_j}{g};$$

the average number of class-j customers at station l is

$$Q_{jl} = \rho_{jl}\frac{g_j^l + g_j}{g}.$$

Following Whitt [17], McKenna and Mitra [7], we now consider an infinite sequence of closed queueing networks indexed by N. Each of the networks has J classes, L single-server FCFS stations and one IS station. The population size and the relative loads depend on N. More specifically, for the Nth network, the populations and relative loads satisfy

$$N_j(N) = \beta_j N + \alpha_j\sqrt{N}, \quad j = 1, \ldots, J, \tag{1}$$

$$\rho_{j0}(N) > 0, \quad j = 1, \ldots, J, \tag{2}$$

$$\frac{\rho_{jl}(N)}{\rho_{j0}(N)} = \frac{\Gamma_{jl}}{N}, \quad j = 1, \ldots, J, l = 1, \ldots, L, \tag{3}$$

where the β_j's are positive, Γ_{jl}'s are nonnegative, and the α_j's are arbitrary constants. Without loss of generality, we assume that $N_j(N) = \beta_j N + \alpha_j\sqrt{N}$ is an integer, since it can always be replaced by its lower floor.

Thus as $N \to \infty$ the population sizes are increasing, but the relative loads at FCFS stations are decreasing with the same rate. We are interested in the asymptotic properties of the sequence of networks – that is, the properties associated with letting N tend to infinity. Under the above asymptotic regime, we always have

$$\lim_{N\to\infty} \sum_j N_j(N)\frac{\rho_{jl}(N)}{\rho_{j0}(N)} = \sum_{j=1}^{J} \beta_j \Gamma_{jl}.$$

Let

$$y_{jl} := \beta_j \Gamma_{jl}$$

and

$$y_l := \sum_{j=1}^{J} y_{jl}.$$

For a given asymptotic regime of the form (1)–(3), we can classify the FCFS stations into three different categories.

DEFINITION 1

The FCFS station l is said to be in normal usage, critical usage, or heavy usage if $y_l < 1$, $y_l = 1$, or $y_l > 1$, respectively.

Through section 5 we only study and present results for networks in which *all* the FCFS stations are in critical usage, i.e., $y_l = 1$, $l = 1, \ldots, L$. Note that this assumption implies

$$\sum_j N_j(N) \frac{\rho_{jl}(N)}{\rho_{j0}(N)} = 1 + O\left(\frac{1}{\sqrt{N}}\right), \quad l = 1, \ldots, L.$$

In the example in section 6, we illustrate how to generalize the asymptotic expansion results to networks in mixed usage where some of the FCFS stations are in normal usage while the rest are in critical usage.

McKenna and Mitra [7] derived the following integral representation of the normalization constant:

$$g = \frac{1}{\prod_{j=1}^{J} N_j!} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^{J} (\rho_{j0} + \boldsymbol{\rho}_j' \boldsymbol{u})^{N_j} \, \mathrm{d}\boldsymbol{u},$$

where

$$\boldsymbol{u} = (u_1, \ldots, u_L)'$$

$$\mathbf{1} = (1, \ldots, 1)'$$

$$\boldsymbol{\rho}_j = (\rho_{j1}, \ldots, \rho_{jL})'$$

$$\mathbf{Q}^+ = \{\boldsymbol{u} \in \mathbb{R}^L : u_l \geq 0, l = 1, \ldots, L\}.$$

It follows that

$$g_j = \frac{N_j}{\prod_{k=1}^J N_k!} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \frac{\prod_{k=1}^J (\rho_{k0} + \boldsymbol{\rho}_k'\boldsymbol{u})^{N_k}}{\rho_{j0} + \boldsymbol{\rho}_j'\boldsymbol{u}} \,\mathrm{d}\boldsymbol{u},$$

$$g^l = \frac{1}{\prod_{j=1}^J N_j!} \int_{\mathbf{Q}^+} (u_l - 1)\,\mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J (\rho_{j0} + \boldsymbol{\rho}_j'\boldsymbol{u})^{N_j} \,\mathrm{d}\boldsymbol{u},$$

$$g_j^l = \frac{N_j}{\prod_{k=1}^J N_k!} \int_{\mathbf{Q}^+} (u_l - 1)\,\mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \frac{\prod_{k=1}^J (\rho_{k0} + \boldsymbol{\rho}_k'\boldsymbol{u})^{N_k}}{\rho_{j0} + \boldsymbol{\rho}_j'\boldsymbol{u}} \,\mathrm{d}\boldsymbol{u}.$$

Let $g(N)$, $g_j(N)$, $g^l(N)$ and $g_j^l(N)$ denote the above normalization constants corresponding to the Nth network in the asymptotic regime. For simplicity, let N_j denote $N_j(N)$ defined in (1). The following formulas are obtained by using the scalings (1)–(3).

$$g(N) = \int_{\mathbf{Q}^+} f(\boldsymbol{u}, N^{-1/2}) \,\mathrm{d}\boldsymbol{u} \prod_{j=1}^J \frac{\rho_{j0}^{N_j}}{N_j!},$$

$$g_j(N) = \frac{N_j}{\rho_{j0}} \int_{\mathbf{Q}^+} f_j(\boldsymbol{u}, N^{-1/2}) \,\mathrm{d}\boldsymbol{u} \prod_{k=1}^J \frac{\rho_{k0}^{N_k}}{N_k!},$$

$$g^l(N) = \int_{\mathbf{Q}^+} (u_l - 1) f(\boldsymbol{u}, N^{-1/2}) \,\mathrm{d}\boldsymbol{u} \prod_{j=1}^J \frac{\rho_{j0}^{N_j}}{N_j!},$$

$$g_j^l(N) = \frac{N_j}{\rho_{j0}} \int_{\mathbf{Q}^+} (u_l - 1) f_j(\boldsymbol{u}, N^{-1/2}) \,\mathrm{d}\boldsymbol{u} \prod_{k=1}^J \frac{\rho_{k0}^{N_k}}{N_k!},$$

where

$$f(\boldsymbol{u}, t) := \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \sum_l \Gamma_{jl} u_l t^2\right)^{\beta_j/t^2 + \alpha_j/t},$$

$$f_j(\boldsymbol{u}, t) := \frac{f(\boldsymbol{u}, t)}{1 + \sum_l \Gamma_{jl} u_l t^2}.$$

3. Asymptotic distribution of queue lengths

Mitra [10] shows that the mean and standard deviation of the queue lengths at FCFS stations are $O(\sqrt{N})$ for a single-class cyclic network in critical usage. Here we will establish a more general result that gives the asymptotic distribution of normalized queue lengths at the FCFS stations for a *multiclass* network. Without loss of generality we assume that each FCFS station is visited by at least one class; therefore, $\sum_{j=1}^{J} \beta_j \Gamma_{jl}^2 > 0$ for all l.

Let $X_{jl}(N)$ denote the number of class-j customers at FCFS station l for Nth network. Further define

$$X_l(N) := \sum_j X_{jl}(N),$$

$$Y_{jl}(N) := \frac{X_{jl}(N)}{\sqrt{N}},$$

$$Y_l(N) := \sum_j Y_{jl}(N),$$

$$\boldsymbol{Y}(N) := (Y_1(N), Y_2(N), \ldots, Y_L(N))',$$

$$\boldsymbol{\Gamma}_j := (\Gamma_{j1}, \ldots, \Gamma_{jL})',$$

$$\boldsymbol{\Gamma} := \sum_j \beta_j \boldsymbol{\Gamma}_j \boldsymbol{\Gamma}_j',$$

$$z_{jl} := \alpha_j \Gamma_{jl},$$

$$z_l := \sum_j z_{jl},$$

$$\boldsymbol{z} := (z_1, \ldots, z_L)'.$$

Since $\Gamma_{jl} \geq 0$, for any $\boldsymbol{u} \in \mathbf{Q}^+$,

$$\begin{aligned} \boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u} &= \sum_{j=1}^{J} \beta_j \left(\sum_{l=1}^{L} \Gamma_{jl} u_l \right)^2 \\ &\geq \sum_{j=1}^{J} \beta_j \sum_{l=1}^{L} \Gamma_{jl}^2 u_l^2 \\ &= \sum_{l=1}^{L} \left(\sum_{j=1}^{J} \beta_j \Gamma_{jl}^2 \right) u_l^2; \end{aligned}$$

hence, $e^{-(1/2)\boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u} + \boldsymbol{z}'\boldsymbol{u}}$ is always integrable on $\mathbf{Q}^+$.

First we need the following technical result which can be easily proved from lemma 3 in Ross and Wang [16].

LEMMA 1

For any $\boldsymbol{x} = (x_1, x_2, \ldots, x_L)'$ such that $x_l > 0$ for $l = 1, \ldots, L$, and $\boldsymbol{w} = (w_1, \ldots, w_L)'$, let

$$\theta(t, \boldsymbol{x}, \boldsymbol{w}) := \mathrm{e}^{-t \sum_{l=1}^{L} x_l} \left(1 + \frac{\sum_{l=1}^{L} x_l \, \mathrm{e}^{w_l/t}}{t}\right)^{t^2}.$$

Then

(i)

$$\lim_{t \to \infty} \theta(t, \boldsymbol{x}, \boldsymbol{w}) = \mathrm{e}^{-\frac{1}{2}\left(\sum_{l=1}^{L} x_l\right)^2 + \sum_{l=1}^{L} x_l w_l};$$

(ii) For any $c > 0$, there exists t_0, such that when $t > t_0$,

$$0 < \theta(t, \boldsymbol{x}, \boldsymbol{w}) < \mathrm{e}^{-\frac{1}{2} c \sum_{l=1}^{L} x_l} \left(1 + \frac{\sum_{l=1}^{L} x_l}{2c}\right)^{c^2}.$$

THEOREM 1

Suppose all FCFS stations are in critical usage. Then

(i)

$$\lim_{N \to \infty} \boldsymbol{Y}(N) \overset{\text{dist.}}{=} \boldsymbol{Y},$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_L)'$ has probability density function

$$p(\boldsymbol{u}) = \frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u} + \boldsymbol{z}'\boldsymbol{u}}}{\int_{\mathbf{Q}^+} \mathrm{e}^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u} + \boldsymbol{z}'\boldsymbol{u}} \, \mathrm{d}\boldsymbol{u}}, \quad \boldsymbol{u} \in \mathbf{Q}^+. \tag{4}$$

(ii) Let $Y_{jl} := y_{jl} Y_l$. Then

$$\lim_{N \to \infty} (Y_{jl}(N), j = 1, \ldots, J; l = 1, \ldots, L) \overset{\text{dist.}}{=} (Y_{jl}, j = 1, \ldots, J; l = 1, \ldots, L).$$

Proof

From the product-form solution and integral representation we have

$$P(X_{jl}(N) = n_{jl}, j = 1, \ldots, J; l = 0, \ldots, L)$$
$$= \frac{1}{g(N)} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \left[\prod_l \prod_j \frac{(\rho_{jl} u_l)^{n_{jl}}}{n_{jl}!} \right] \prod_j \frac{\rho_{j0}^{n_{j0}}}{n_{j0}!} \, \mathrm{d}\boldsymbol{u}.$$

Define the following sets of vectors with nonnegative integer components:

$$\Lambda := \left\{ \boldsymbol{m} = (m_1, \ldots, m_L) : \sum_l m_l \le \sum_j N_j \right\},$$

and for any $\boldsymbol{m} \in \Lambda$,

$$\Omega(\boldsymbol{m}) := \left\{ \boldsymbol{n} \in \Omega : \sum_j n_{jl} = m_l, \text{for } l = 1, \ldots, L \right\}.$$

Let $n^j := \sum_{l=1}^L n_{jl}$, then we have

$$P(X_l(N) = m_l, l = 1, \ldots, L)$$
$$= \frac{1}{g(N)} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \sum_{\boldsymbol{n} \in \Omega(\boldsymbol{m})} \left[\prod_l \prod_j \frac{(\rho_{jl} u_l)^{n_{jl}}}{n_{jl}!} \right] \left[\prod_j \frac{\rho_{j0}^{N_j - n^j}}{(N_j - n^j)!} \right] \mathrm{d}\boldsymbol{u}.$$

Now the moment generating function of $(Y_1(N), \ldots, Y_L(N))$ is

$$\phi_N(t_1, \ldots, t_L) = E[\mathrm{e}^{\sum_l t_l Y_l(N)}]$$
$$= \sum_{\boldsymbol{m} \in \Lambda} \mathrm{e}^{\sum_l t_l m_l / \sqrt{N}} P(X_l(N) = m_l, l = 1, \ldots, L)$$
$$= \sum_{\boldsymbol{m} \in \Lambda} \mathrm{e}^{\sum_l t_l m_l / \sqrt{N}} \sum_{\boldsymbol{n} \in \Omega(\boldsymbol{m})} \frac{1}{g(N)}$$
$$\times \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \left[\prod_l \prod_j \frac{(\rho_{jl} u_l)^{n_{jl}}}{n_{jl}!} \right] \left[\prod_j \frac{\rho_{j0}^{N_j - nj}}{(N_j - n^j)!} \right] \mathrm{d}\boldsymbol{u}$$

$$= \sum_{n \in \Omega} \left[\prod_j \prod_l \mathrm{e}^{t_l n_{jl}/\sqrt{N}} \right] \frac{1}{g(N)}$$

$$\times \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \left[\prod_l \prod_j \frac{(\rho_{jl} u_l)^{n_{jl}}}{n_{jl}!} \right] \left[\prod_j \frac{\rho_{j0}^{N_j - n^j}}{(N_j - n^j)!} \right] \mathrm{d}\boldsymbol{u}$$

$$= \sum_{n \in \Omega} \frac{1}{g(N)} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \left[\prod_l \prod_j \frac{(\rho_{jl} u_l \mathrm{e}^{t_l/\sqrt{N}})^{n_{jl}}}{n_{jl}!} \right] \left[\prod_j \frac{\rho_{j0}^{N_j - n^j}}{(N_j - n^j)!} \right] \mathrm{d}\boldsymbol{u}$$

$$= \frac{1}{g(N) \prod_{j=1}^J N_j!} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(\rho_{j0} + \sum_l \rho_{jl} \mathrm{e}^{t_l/\sqrt{N}} u_l \right)^{N_j} \mathrm{d}\boldsymbol{u}$$

$$= \frac{\prod_j \rho_{j0}^{N_j}}{g(N) \prod_{j=1}^J N_j!} \int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \sum_l \frac{\rho_{jl}}{\rho_{j0}} \mathrm{e}^{t_l/\sqrt{N}} u_l \right)^{N_j} \mathrm{d}\boldsymbol{u}$$

$$= \frac{\int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \sum_l \frac{\rho_{jl}}{\rho_{j0}} \mathrm{e}^{t_l/\sqrt{N}} u_l \right)^{N_j} \mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \sum_l \frac{\rho_{jl}}{\rho_{j0}} u_l \right)^{N_j} \mathrm{d}\boldsymbol{u}}.$$

Recalling that

$$\frac{\rho_{jl}}{\rho_{j0}} = \frac{\rho_{jl}(N)}{\rho_{j0}(N)} = \frac{\Gamma_{jl}}{N},$$

changing variables $u_l/\sqrt{N} \to u_l$ for $l = 1, \ldots, L$, and applying $\sum_j \Gamma_{jl} \beta_j = 1$ for all l's, and invoking lemma 1, we have

$$\lim_{N \to \infty} \phi_N(t_1, \ldots, t_L)$$

$$= \lim_{N \to \infty} \frac{\int_{\mathbf{Q}^+} \mathrm{e}^{-\sqrt{N}\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \frac{\sum_l \Gamma_{jl} u_l \mathrm{e}^{t_l/\sqrt{N}}}{\sqrt{N}} \right)^{\beta_j N + \alpha_j \sqrt{N}} \mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} \mathrm{e}^{-\sqrt{N}\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^J \left(1 + \frac{\sum_l \Gamma_{jl} u_l}{\sqrt{N}} \right)^{\beta_j N + \alpha_j \sqrt{N}} \mathrm{d}\boldsymbol{u}}$$

$$= \lim_{N\to\infty} \frac{\displaystyle\int_{Q^+} \exp\left(\sum_j \beta_j\left[-\sqrt{N}\sum_l \Gamma_{jl}u_l + N\ln\left(1+\frac{\sum_l \Gamma_{jl}u_l e^{t_l/\sqrt{N}}}{\sqrt{N}}\right)\right] + \sum_j \alpha_j\sqrt{N}\ln\left(1+\frac{\sum_l \Gamma_{jl}u_l e^{t_l/\sqrt{N}}}{\sqrt{N}}\right)\right) \mathrm{d}\boldsymbol{u}}{\displaystyle\int_{Q^+} e^{-\sqrt{N}\mathbf{1}'\boldsymbol{u}} \prod_{j=1}^{J}\left(1+\frac{\sum_l \Gamma_{jl}u_l}{\sqrt{N}}\right)^{\beta_j N+\alpha_j\sqrt{N}} \mathrm{d}\boldsymbol{u}}$$

$$= \frac{\displaystyle\int_{Q^+} e^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u}+\boldsymbol{z}'\boldsymbol{u}+\boldsymbol{t}'\boldsymbol{u}}\,\mathrm{d}\boldsymbol{u}}{\displaystyle\int_{Q^+} e^{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{\Gamma}\boldsymbol{u}+\boldsymbol{z}'\boldsymbol{u}}\,\mathrm{d}\boldsymbol{u}}$$

$$= \phi(t_1,\ldots,t_L),$$

which is the moment generating function of $p(\boldsymbol{u})$. Since $\phi(t_1,\ldots,t_L)$ and, for any N, $\phi_N(t_1,\ldots,t_L)$ exist for all $t_1,\ldots,t_L$, the convergence of the moment generating function implies the convergence in distribution (see Billingsley [2]).

With similar procedures, we can obtain for any $t_{jl}, j=1,\ldots,J, l=1,\ldots,L$,

$$\lim_{N\to\infty} E[e^{\sum_j\sum_l t_{jl}Y_{jl}(N)}] = E[e^{\sum_j\sum_l t_{jl}\gamma_{jl}Y_l}]$$
$$= E[e^{\sum_j\sum_l t_{jl}Y_{jl}}],$$

which implies the weak convergence of $(Y_{jl}(N),\ j=1,\ldots,J,\ l=1,\ldots,L)$ to $(Y_{jl}, j=1,\ldots,J, l=1,\ldots,L)$. □

Remarks

(1) Although matrix $\boldsymbol{\Gamma}$ is always semi-positive definite, it is not necessarily positive definite. In particular, when $J<L$, $\boldsymbol{\Gamma}$ is always singular.

(2) When $\boldsymbol{\Gamma}$ is positive definite, the distribution in (4) is a truncated multivariate normal.

(3) In general, the components of $\boldsymbol{Y}$ are not independent.

A natural question is for what networks will $\boldsymbol{\Gamma}$ be positive definite? Since $\beta_j>0$, we have for any $\boldsymbol{x}=(x_1,\ldots,x_L)'\in\mathbb{R}^L$

$$\boldsymbol{x}'\boldsymbol{\Gamma}\boldsymbol{x} = \boldsymbol{x}'\left(\sum_j \beta_j\boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_j'\right)\boldsymbol{x} = \sum_j \beta_j(\boldsymbol{x}'\boldsymbol{\Gamma}_j)^2 = 0$$

if and only if $\boldsymbol{x}'\boldsymbol{\Gamma}_j=0$ for all $j=1,\ldots,J$. Therefore, $\boldsymbol{\Gamma}$ is positive definite if and only if $\mathrm{rank}\{\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_J\}=L$.

To further our discussion, we make the assumption in (3) more specific as follows. Let $\mu_l(N)$ denote the service rate of the server at the lth FCFS station in the Nth network. Assume

$$\mu_l(N) := N\mu_l^*,$$

where μ_l^* is a positive constant. Also assume that μ_{j0}, the service rate for class-j customers at the IS station, does not depend on N. Then

$$\Gamma_{jl} = \frac{\mu_{j0}}{\lambda_{j0}} \frac{\lambda_{jl}}{\mu_l^*}.$$

Therefore for any j,

$$\boldsymbol{x}'\boldsymbol{\Gamma}_j = \sum_l x_l \Gamma_{jl} = 0$$

if and only if

$$\sum_l \frac{x_l}{\mu_l^*} \lambda_{jl} = 0.$$

Let $\boldsymbol{P}$ be the $J \times L$ matrix whose components in the jth row are the visit ratios to the L FCFS stations of class j. It follows that $\boldsymbol{\Gamma}$ is positive definite if and only if $\boldsymbol{P}$ has rank L. When $J \gg L$, it is likely that $\boldsymbol{\Gamma}$ will be positive definite.

When the matrix $\boldsymbol{\Gamma}$ is diagonal, the distribution in (4) is simplified significantly: it is the product of L truncated univariate normal distributions. Unfortunately, $\boldsymbol{\Gamma}$ is diagonal only for trivial networks. Nevertheless, when the diagonal elements of $\boldsymbol{\Gamma}$ are much greater than the off diagonal elements, the components of $\boldsymbol{Y}$ are approximately independent random variables. Therefore, we are motivated to find the traffic conditions that will lead to such $\boldsymbol{\Gamma}$. The following two examples illustrate such traffic conditions.

EXAMPLE 3.1

Consider a network with one IS station and L FCFS stations. For each pair of FCFS stations there is exactly one class visiting them and the IS station. Hence there are a total of $J = \binom{L}{2}$ classes and each FCFS station is visited by exactly $L-1$ classes. Clearly, the corresponding matrix $\boldsymbol{\Gamma}$ is positive definite. Suppose, for each class, the traffic is distributed among the two FCFS stations such that if class j visits FCFS stations l and m, then $\Gamma_{jl} = \Gamma_{jm}$. For every l, let Θ_l be the set of classes that visit FCFS station l. Let $\Gamma(l,m)$ denote the (l,m) component of $\boldsymbol{\Gamma}$. Then

$$\Gamma(l,l) = \sum_{j \in \Theta_l} \beta_j \Gamma_{jl}^2.$$

On the other hand, for any $m \neq l$, there exists some $j \in \Theta_l$ such that $\Gamma(l,m) = \beta_j \Gamma_{jl} \Gamma_{jm} = \beta_j \Gamma_{jl}^2$; hence, for $m \neq l$, the ratio of the off-diagonal and diagonal terms is

$$\frac{\Gamma(l,m)}{\Gamma(l,l)} = \frac{\beta_j \Gamma_{jl}^2}{\sum_{j \in \Theta_l} \beta_j \Gamma_{jl}^2},$$

which is very small for all m if the $\beta_j \Gamma_{jl}^2$'s are not much different from each other and L is large.

EXAMPLE 3.2

Consider the following generalization of the above example. Now for each set of K FCFS stations there is one class visiting them and the IS station. The total number of classes is $J = \binom{L}{K}$. Each FCFS station is visited by $\binom{L-1}{K-1}$ classes. Each pair of FCFS stations is visited by $\binom{L-2}{K-2}$ classes. The corresponding matrix $\boldsymbol{\Gamma}$ is positive definite. Suppose for each class the traffic is evenly distributed among the K FCFS stations. More specifically, if class j visits FCFS stations $l_1, \ldots, l_K$, suppose $\Gamma_{jl_1} = \ldots = \Gamma_{jl_K}$. Clearly, for any l, $\Gamma(l,l)$ is a summation of $\binom{L-1}{K-1}$ terms, while for any $l \neq m$, $\Gamma(l,m)$ is a summation of $\binom{L-2}{K-2}$ terms. Since

$$\binom{L-1}{K-1} = \frac{L-1}{K-1} \binom{L-2}{K-2},$$

when $L \gg K$ and the summands are not mch different from each other, we have $\Gamma(l,m)/\Gamma(l,l) \ll 1$ for all l and m with $l \neq m$.

Summarizing the above discussion, when the traffic has a symmetric pattern and the number of FCFS stations is large, the normalized queue lengths are approximately independent. We note that a similar observation is made by Kelly [6] for loss networks.

4. Asymptotic expansions of normalization constants

Morrison and Mitra [11] and Mitra [10] obtained the asymptotic expansions of several performance measures for networks in critical usage. However, they only considered a family of networks for which the normalization constants can be expressed in terms of *one*-dimensional integrals and the coefficients of the expansions involve a class of functions closely related to the classical parabolic cylinder functions. In this section we develop asymptotic expansions for a much larger family of networks.

After the change of variable $u_l/\sqrt{N} \to u_l$, $l = 1, \ldots, L$ (see Ross and Wang [16]), we have the following expression for normalization constant of the Nth network,

$$g(N) = N^{L/2} H(N^{-1/2}) \prod_{j=1}^{J} \frac{\rho_{j0}^{N_j}}{N_j!},$$

where

$$H(t) := \int_{\mathbf{Q}^+} h(\boldsymbol{u}, t)\, \mathrm{d}\boldsymbol{u},$$

$$h(\boldsymbol{u}, t) := \mathrm{e}^{-\mathbf{1}'\boldsymbol{u}/t} \prod_j \left(1 + t \sum_l \Gamma_{jl} u_l\right)^{\beta_j/t^2 + \alpha_j/t}.$$

Let

$$s_1(\boldsymbol{u}, t) := \sum_{j=1}^{J} \beta_j \left[-\frac{1}{t} \sum_{l=1}^{L} \Gamma_{jl} u_l + \frac{1}{t^2} \ln\left(1 + t \sum_l \Gamma_{jl} u_l\right)\right],$$

$$s_2(\boldsymbol{u}, t) := \sum_{j=1}^{J} \alpha_j \frac{\ln\left(1 + t \sum_l \Gamma_{jl} u_l\right)}{t}.$$

When $\sum_{j=1}^{J} \beta_j \Gamma_{jl} = 1$, $l = 1, \ldots, L$,

$$h(\boldsymbol{u}, t) = \mathrm{e}^{s_1(\boldsymbol{u},t) + s_2(\boldsymbol{u},t)};$$

furthermore, $s_1(\boldsymbol{u}, t)$ and $s_2(\boldsymbol{u}, t)$ are infinitely differentiable with respect to t in a neighborhood of $t = 0$, hence so is $h(\boldsymbol{u}, t)$. The following lemma is easily obtained from the standard Laplace integral theory (see for example Bleistein and Handelsman [3]).

LEMMA 2

Suppose all FCFS stations are in critical usage.

(i) For every $\boldsymbol{u} \in \mathbf{Q}^+$, $h(\boldsymbol{u}, t)$ is infinitely differentiable in a neighborhood of $t = 0$.

(ii) For every $\boldsymbol{u} \in \mathbf{Q}^+$,

$$h(\boldsymbol{u}, t) = \sum_{k=0}^{n} \frac{t^k}{k!} \frac{\partial^k}{\partial t^k} h(\boldsymbol{u}, 0) + o(t^n) \quad \text{for all } n, \text{as } t \to 0.$$

(iii)

$$H(t) = \sum_{k=0}^{n} \frac{t^k}{k!} \int_{\mathbf{Q}^+} \frac{\partial^k}{\partial t^k} h(\boldsymbol{u}, 0)\, \mathrm{d}\boldsymbol{u} + o(t^n), \quad \text{for all } n, \text{as } t \to 0.$$

The following theorem is a direct result of the lemma.

THEOREM 2

As $N \to \infty$,

$$H(N^{-1/2}) = \sum_{k=0}^{n} \frac{1}{N^{k/2} k!} \int_{\mathbf{Q}^+} \frac{\partial^k}{\partial t^k} h(\boldsymbol{u}, 0)\, \mathrm{d}\boldsymbol{u} + o\left(\frac{1}{N^{n/2}}\right) \quad \text{for all } n.$$

To give some of the leading terms of the expansion of $H(N^{-1/2})$, we note that for any $\boldsymbol{u}$ when t is sufficiently small,

$$s_1(\boldsymbol{u}, t) = \sum_{k=0}^{\infty} (-1)^{k+1} \frac{t^k}{k+2} \sum_j \beta_j \left(\sum_l \Gamma_{jl} u_l\right)^{k+2};$$

$$s_2(\boldsymbol{u}, t) = \sum_{k=0}^{\infty} (-1)^{k} \frac{t^k}{k+1} \sum_j \alpha_j \left(\sum_l \Gamma_{jl} u_l\right)^{k+1}.$$

Hence

$$s_1(\boldsymbol{u}, 0) = -\frac{1}{2} \sum_j \beta_j \left(\sum_l \Gamma_{jl} u_l\right)^2,$$

$$s_2(\boldsymbol{u}, 0) = \sum_j \alpha_j \left(\sum_l \Gamma_{jl} u_l\right),$$

$$s_1^{(1)}(\boldsymbol{u}, 0) = \frac{1}{3} \sum_j \beta_j \left(\sum_l \Gamma_{jl} u_l\right)^3,$$

$$s_2^{(1)}(\boldsymbol{u}, 0) = -\frac{1}{2} \sum_j \alpha_j \left(\sum_l \Gamma_{jl} u_l\right)^2,$$

$$s_1^{(2)}(\boldsymbol{u}, 0) = -\frac{1}{2} \sum_j \beta_j \left(\sum_l \Gamma_{jl} u_l\right)^4,$$

$$s_2^{(2)}(\boldsymbol{u}, 0) = \frac{2}{3} \sum_j \alpha_j \left(\sum_l \Gamma_{jl} u_l\right)^3,$$

and

$$\begin{aligned} h(\boldsymbol{u},0) &= \mathrm{e}^{s_1(\boldsymbol{u},0)+s_2(\boldsymbol{u},0)} \\ &= \exp\left(-\frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^2 + \sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)\right), \\ h^{(1)}(\boldsymbol{u},0) &= h(\boldsymbol{u},0)[s_1^{(1)}(\boldsymbol{u},0) + s_2^{(1)}(\boldsymbol{u},0)] \\ &= \exp\left(-\frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^2 + \sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)\right) \\ &\quad \times \left\{\frac{1}{3}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^3 - \frac{1}{2}\sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)^2\right\}, \\ h^{(2)}(\boldsymbol{u},0) &= h(\boldsymbol{u},0)\{[s_1^{(1)}(\boldsymbol{u},0) + s_2^{(1)}(\boldsymbol{u},0)]^2 + s_1^{(2)}(\boldsymbol{u},0) + s_2^{(2)}(\boldsymbol{u},0)\} \\ &= \exp\left(-\frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^2 + \sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)\right) \\ &\quad \times \left\{\left[\frac{1}{3}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^3 - \frac{1}{2}\sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)^2\right]^2\right. \\ &\quad \left. - \frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^4 + \frac{2}{3}\sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)^3\right\}, \end{aligned}$$

and

$$h^{(k+1)}(\boldsymbol{u},0) = \sum_{m=0}^{k} \frac{k!}{m!(k-m)!} h^{(m)}(\boldsymbol{u},0)[s_1^{(k-m+1)}(\boldsymbol{u},0) + s_2^{(k-m+1)}(\boldsymbol{u},0)];$$

thus

$$\begin{aligned} H(N^{-1/2}) &= \int_{\mathbf{Q}^+} \exp\left(-\frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^2 + \sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)\right) \mathrm{d}\boldsymbol{u} \\ &\quad + \frac{1}{\sqrt{N}} \int_{\mathbf{Q}^+} \exp\left(-\frac{1}{2}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^2 + \sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)\right) \\ &\quad \times \left\{\frac{1}{3}\sum_j \beta_j\left(\sum_l \Gamma_{jl}u_l\right)^3 - \frac{1}{2}\sum_j \alpha_j\left(\sum_l \Gamma_{jl}u_l\right)^2\right\} \mathrm{d}\boldsymbol{u} \end{aligned}$$

$$+\frac{1}{2N}\int_{\mathbf{Q}^+}\exp\left(-\frac{1}{2}\sum_j\beta_j\left(\sum_l\Gamma_{jl}u_l\right)^2+\sum_j\alpha_j\left(\sum_l\Gamma_{jl}u_l\right)\right)$$

$$\times\left\{\left[\frac{1}{3}\sum_j\beta_j\left(\sum_l\Gamma_{jl}u_l\right)^3-\frac{1}{2}\sum_j\alpha_j\left(\sum_l\Gamma_{jl}u_l\right)^2\right]^2\right.$$

$$\left.-\frac{1}{2}\sum_j\beta_j\left(\sum_l\Gamma_{jl}u_l\right)^4+\frac{2}{3}\sum_j\alpha_j\left(\sum_l\Gamma_{jl}u_l\right)^3\right\}\mathrm{d}\boldsymbol{u}+O\left(\frac{1}{N^{3/2}}\right).$$

For the normalization constant corresponding to one less class-j customer, we have

$$g_j(N)=\frac{N_j}{\rho_{j0}}N^{L/2}H_j(N^{-1/2})\prod_{k=1}^{J}\frac{\rho_{k0}^{N_k}}{N_k!},$$

where

$$H_j(t):=\int_{\mathbf{Q}^+}\frac{h(\boldsymbol{u},t)}{1+t\sum_l\Gamma_{jl}u_l}\,\mathrm{d}\boldsymbol{u}.$$

The asymptotic expansion of $H_j(N^{-1/2})$ is obtained with the technique used for $H(N^{-1/2})$. As $N\to\infty$,

$$H_j(N^{-1/2})=\int_{\mathbf{Q}^+}h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}+\frac{1}{\sqrt{N}}\int_{\mathbf{Q}^+}[h^{(1)}(\boldsymbol{u},0)-\boldsymbol{\Gamma}_j'\boldsymbol{u}h(\boldsymbol{u},0)]\,\mathrm{d}\boldsymbol{u}+O\left(\frac{1}{N}\right).$$

Now for any l, we have

$$g^l(N)=N^{L/2}H^l(N^{-1/2})\prod_{j=1}^{J}\frac{\rho_{j0}^{N_j}}{N_j!},$$

$$g_j^l(N)=N^{L/2}H_j^l(N^{-1/2})\prod_{j=1}^{J}\frac{\rho_{j0}^{N_j}}{N_j!},$$

where

$$H^l(N^{-1/2})=\int_{\mathbf{Q}^+}(\sqrt{N}u_l-1)h(\boldsymbol{u},N^{-1/2})\,\mathrm{d}\boldsymbol{u},$$

$$H_j^l(N^{-1/2})=\int_{\mathbf{Q}^+}(\sqrt{N}u_l-1)\frac{h(\boldsymbol{u},N^{-1/2})}{1+N^{-1/2}\sum_l\Gamma_{jl}u_l}\,\mathrm{d}\boldsymbol{u}.$$

We then have

$$\frac{H^l(N^{-1/2})}{\sqrt{N}} = \int_{\mathbf{Q}^+} u_l h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u} + \frac{1}{\sqrt{N}} \int_{\mathbf{Q}^+} [u_l h^{(1)}(\boldsymbol{u},0) - h(\boldsymbol{u},0)]\,\mathrm{d}\boldsymbol{u} + O\left(\frac{1}{N}\right),$$

$$\frac{H_j^l(N^{-1/2})}{\sqrt{N}} = \int_{\mathbf{Q}^+} u_l h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u} + \frac{1}{\sqrt{N}} \times \int_{\mathbf{Q}^+} [u_l h^{(1)}(\boldsymbol{u},0) - h(\boldsymbol{u},0) - \boldsymbol{\Gamma}_j' \boldsymbol{u} h(\boldsymbol{u},0)]\,\mathrm{d}\boldsymbol{u} + O\left(\frac{1}{N}\right).$$

5. Asymptotic expansions of performance measures

With the above expansions for $H(N^{-1/2})$, $H^l(N^{-1/2})$ and $H_j^l(N^{-1/2})$, we can easily obtain the following expansions for performance measures. Again, the proof is straightforward and omitted.

THEOREM 3

$$\mathrm{UTIL}_{jl}(N) = y_{jl} + \frac{a_{jl}}{\sqrt{N}} + O\left(\frac{1}{N}\right),$$

$$\frac{\mathrm{Q}_{jl}(N)}{\sqrt{N}} = b_{jl} + \frac{c_{jl}}{\sqrt{N}} + O\left(\frac{1}{N}\right),$$

where

$$a_{jl} = z_{jl} - y_{jl} \frac{\int_{\mathbf{Q}^+} \boldsymbol{\Gamma}_j' \boldsymbol{u} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}},$$

$$b_{jl} = \frac{y_{jl} \int_{\mathbf{Q}^+} u_l h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}},$$

$$c_{jl} = \frac{z_{jl} \int_{\mathbf{Q}^+} u_l h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}} + \frac{y_{jl} \int_{\mathbf{Q}^+} [u_l h^{(1)}(\boldsymbol{u},0) - \boldsymbol{\Gamma}_j' \boldsymbol{u} h(\boldsymbol{u},0)]\,\mathrm{d}\boldsymbol{u}}{\int_{\mathbf{Q}^+} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}} + \frac{y_{jl} [\int_{\mathbf{Q}^+} h^{(1)}(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}][\int_{\mathbf{Q}^+} u_l h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}]}{[\int_{\mathbf{Q}^+} h(\boldsymbol{u},0)\,\mathrm{d}\boldsymbol{u}]^2}.$$

Remarks

(1) For critical usage, the asymptotic expansions are in terms of $1/\sqrt{N}$ instead of $1/N$. Mitra [10] obtained similar asymptotic expansions for a special network with cyclic and homogeneous routing.

(2) The evaluation of the above coefficients is still not easy. However, as theorem 3 shows, the expansions can give us some insight on how the performance measures change as the populations and relative processing rates increase. For example,

$$\mathrm{UTIL}_{jl} = y_{jl} + O\left(\frac{1}{\sqrt{N}}\right),$$

which is similar to the case of normal usage, except that now the error term is $O(1/\sqrt{N})$ instead of $O(1/N)$.

6. Application to networks with finite data

In sections 3 through 5 we presented an asymptotic analysis for networks in critical usage. At this juncture, it is natural to ask the following questions: (1) Can we generalize the results to networks in mixed usage, i.e., some FCFS stations in normal usage and the others in critical usage? (2) How do we apply the results to a network with finite data?

The answer to the first question is yes: With similar techniques we can easily extend the results to mixed usage; we will illustrate how to do so in the following example.

To apply the results to a network with finite data, we have to first determine what FCFS stations are in normal usage and critical usage and then derive parameters N, β_j's, α_j's and Γ_{jl}'s from the original network data, the N_j's and ρ_{jm}'s so that we can apply the asymptotic expansion results to approximate the network performance measures. McKenna and Mitra [7] proposed a rule for identifying stations in normal usage: if

$$\sum_j N_j \frac{\rho_{jl}}{\rho_{j0}} < 1, \tag{5}$$

then station l is considered as being in normal usage. Since in the asymptotic regime we have

$$\sum_j N_j(N) \frac{\rho_{jl}(N)}{\rho_{j0}(N)} = 1 + O\left(\frac{1}{\sqrt{N}}\right), \quad l = 1, \ldots, L,$$

it is natural to consider station l (of a network with finite data) as being in critical usage if

$$\left|\sum_j N_j \frac{\rho_{jl}}{\rho_{j0}} - 1\right| \ll 1. \tag{6}$$

However our numerical testing shows that if the inequality in (5) holds by too small of a margin, the asymptotic expansion under normal usage will be inefficient because too many terms will be required to achieve reasonable accuracy. On the other hand, if the left hand side in (6) is not small enough, the approximation obtained under critical usage will not be good either. Our numerical testing with PANACEA indicates that when (5) holds with a margin of 15%, the asymptotic expansion under normal usage can give reasonably good approximations and when the left hand side in (6) is less than 15%, the asymptotic expansion under critical usage condition is good. Hence we propose the following heuristic for determining the stations in normal usage and critical usage: if station l is such that

$$\sum_j N_j \frac{\rho_{jl}}{\rho_{j0}} < 0.85,$$

then we use asymptotic expansion under normal usage for station l; if station l is such that

$$0.85 \leq \sum_j N_j \frac{\rho_{jl}}{\rho_{j0}} \leq 1.15,$$

then we use asymptotic expansion under critical usage for station l. (More refined heuristics taking into account population sizes can be proposed).

We suggest the following rule for choosing the asymptotic parameters. Let $\mathbf{C}$ be the set of FCFS stations that are in critical usage:

(1) Pick the parameter N with either $N = \max_{j=1,\ldots,J} N_j$ or $N = \max_{\rho_{jl}>0} \rho_{j0}/\rho_{jl}$.
(2) Let $\Gamma_{jl} = (\rho_{jl}/\rho_{jl})N$ for $j = 1, \ldots, J$, $l = 1, \ldots, L$.
(3) Solve for $\beta_j > 0, j = 1, \ldots, J$ from

$$\sum_j \beta_j \Gamma_{jl} = 1, \quad l \in \mathbf{C}.$$

(4) Let $\alpha_j = (N_j - \beta_j N)/\sqrt{N}$, for $j = 1, \ldots, J$.

The following is an example illustrating how to obtain an asymptotic approximation of utilizations of FCFS stations in a network with finite data.

EXAMPLE 6.1

Consider a network consisting of one IS station, station 0, and two FCFS stations, stations 1 and 2. There are two classes of customers circulating in the network. Class 1 has $N_1 = 20$ customers and visits station 0 and 1 cyclically. Class 2 has $N_2 = 10$ customers and visits station 0, 1 and 2 cyclically. The service rates at the three stations are $\mu_0 = 0.35$, $\mu_1 = 10.0$ and $\mu_2 = 7.0$. Hence $\rho_{10} = \frac{20}{7}$, $\rho_{11} = \frac{1}{10}$, $\rho_{12} = 0$, $\rho_{20} = \frac{20}{7}$, $\rho_{21} = \frac{1}{10}$ and $\rho_{22} = \frac{1}{7}$.

Since

$$N_1 \frac{\rho_{11}}{\rho_{10}} + N_2 \frac{\rho_{21}}{\rho_{20}} = 1.05,$$

$$N_1 \frac{\rho_{12}}{\rho_{10}} + N_2 \frac{\rho_{22}}{\rho_{20}} = 0.5,$$

we consider station 1 as being in critical usage and station 2 in normal usage, thus the network is in mixed usage. Note that because $N_1(\rho_{11}/\rho_{10}) + N_2(\rho_{21}/\rho_{20}) > 1$, the asymptotic expansion for normal usage is not applicable.

We pick $N = \max\{N_1, N_2\} = 20$. Then

$$\Gamma_{11} = N\frac{\rho_{11}}{\rho_{10}} = \frac{7}{10}, \quad \Gamma_{12} = N\frac{\rho_{12}}{\rho_{10}} = 0.0,$$

$$\Gamma_{21} = N\frac{\rho_{21}}{\rho_{20}} = \frac{7}{10}, \quad \Gamma_{22} = N\frac{\rho_{22}}{\rho_{20}} = 1.0.$$

Solving for β_1 and β_2 from

$$\beta_1\Gamma_{11} + \beta_2\Gamma_{21} = 1,$$

we obtain $\beta_1 = 1$, $\beta_2 = \frac{3}{7}$ as one possible solution. (Intuitively, β_j should be chosen to be approximately proportional to N_j.) Then $\alpha_1 = 0.0$ and $\alpha_2 = 0.319$.

Now we derive asymptotic expansions for this network. After the change of variables $u_1/\sqrt{N} \to u_1$, $u_2 \to u_2$, we have the following expression for the normalization constant,

$$g = N^{1/2}H(N^{-1/2})\prod_{j=1}^{2}\frac{\rho_{j0}^{N_j}}{N_j!},$$

where

$$H(t) = \int_0^{+\infty}\int_0^{+\infty} h(u_1, u_2, t)\,\mathrm{d}u_1\,\mathrm{d}u_2,$$

$$h(u_1, u_2, t) = \mathrm{e}^{-u_1/t-u_2} \prod_j (1 + t\Gamma_{j1}u_1 + t^2\Gamma_{j2}u_2)^{\beta_j/t^2+\alpha_j/t}.$$

Let

$$s_1(u_1, u_2, t) = \sum_j \beta_j \Big[-\frac{\Gamma_{j1}u_1}{t} + \frac{1}{t^2}\ln(1 + t\Gamma_{j1}u_1 + t^2\Gamma_{j2}u_2) \Big],$$

$$s_2(u_1, u_2, t) = \sum_j \frac{\alpha_j}{t}\ln(1 + t\Gamma_{j1}u_1 + t^2\Gamma_{j2}u_2).$$

Since $\beta_1\Gamma_{11} + \beta_2\Gamma_{21} = 1$, we have

$$h(u_1, u_2, t) = \mathrm{e}^{s_1(u_1,u_2,t)+s_2(u_1,u_2,t)-u_2}.$$

By Taylor expansion, the following are obvious:

$$s_1(u_1, u_2, t) = \Big(\sum_j \beta_j\Gamma_{j2}\Big)u_2 - \frac{1}{2}\Big(\sum_j \beta_j\Gamma_{j1}^2\Big)u_1^2 - t\Big(\sum_j \beta_j\Gamma_{j1}\Gamma_{j2}\Big)u_1u_2 + \frac{t}{3}\Big(\sum_j \beta_j\Gamma_{j1}^3\Big)u_1^3 + o(t),$$

$$s_2(u_1, u_2, t) = \Big(\sum_j \alpha_j\Gamma_{j1}\Big)u_1 + t\Big(\sum_j \alpha_j\Gamma_{j2}\Big)u_2 - \frac{t}{2}\Big(\sum_j \alpha_j\Gamma_{j1}^2\Big)u_1^2 + o(t),$$

$$s_1(u_1, u_2, 0) = \Big(\sum_j \beta_j\Gamma_{j2}\Big)u_2 - \frac{1}{2}\Big(\sum_j \beta_j\Gamma_{j1}^2\Big)u_1^2,$$

$$s_2(u_1, u_2, 0) = \Big(\sum_j \alpha_j\Gamma_{j1}\Big)u_1.$$

Hence

$$h(u_1, u_2, 0) = \mathrm{e}^{-\left(1-\sum_j \beta_j\Gamma_{j2}\right)u_2+\left(\sum_j \alpha_j\Gamma_{j1}\right)u_1-\frac{1}{2}\left(\sum_j \beta_j\Gamma_{j1}^2\right)u_1^2}$$

$$= \mathrm{e}^{-0.571u_2-0.350u_1^2+0.223u_1}.$$

It is straightforward, under the mixed usage assumption, to obtain the formulas that parallel those in theorem 3. For instance, for $l = 1, 2$ and $j = 1, 2$,

$$\mathrm{UTIL}_{jl}(N) = y_{jl} + \frac{a_{jl}}{\sqrt{N}} + O\left(\frac{1}{N}\right),$$

where

$$a_{jl} = z_{jl} - y_{jl} \frac{\int_0^{+\infty}\int_0^{+\infty} \Gamma_{j1} u_1 h(u_1, u_2, 0)\, du_1\, du_2}{\int_0^{+\infty}\int_0^{+\infty} h(u_1, u_2, 0)\, du_1\, du_2}. \tag{7}$$

It should be pointed out that the a_{jl} above is not zero in general even for a station in normal usage. This is somehow surprising since the asymptotic expansions are in $1/N$ when *all* FCFS stations are in normal usage. Also notice that only the variable corresponding to the station in critical usage appears in the integrand in the numerator.

Recalling that $z_{jl} = \alpha_j \Gamma_{jl}$ and $y_{jl} = \beta_j \Gamma_{jl}$, we have

$$z_{11} = 0.0, \quad z_{12} = 0.0, \quad z_{21} = 0.223, \quad z_{22} = 0.319,$$

$$y_{11} = 0.70, \quad y_{12} = 0.0, \quad y_{21} = 0.30, \quad y_{22} = 0.429,$$

After some simple numerical calculation, we have

$$a_{11} = -0.528, \quad a_{21} = -0.003, \quad a_{12} = 0.0, \quad a_{22} = -0.004.$$

Therefore, using the first two leading terms in the asymptotic expansions, we have the following approximations for the utilizations,

$$\mathrm{UTIL}_{11} \approx 0.582, \quad \mathrm{UTIL}_{21} \approx 0.299, \quad \mathrm{UTIL}_{12} \approx 0.0, \quad \mathrm{UTIL}_{22} \approx 0.428,$$

which are close to the exact values 0.599, 0.281, 0.0 and 0.401. (The exact values are obtained with Monte Carlo integration, see Ross et al. [15].)

Note that because there is only one FCFS station in critical usage the evaluation of the integrals in (7) is straightforward. When there are more FCFS stations in critical usage, evaluating the multidimensional integrals on $\mathbf{Q}^+$ is not trivial. One possible approach, as indicated in Ross and Wang [16], is to use Monte Carlo integration. Because the exponential terms in all the integrands are the same, we can estimate all of them with one sequence of random samples.

References

[1] F. Baskett, M. Chandy, R. Muntz and J. Palacios, Open, closed and mixed networks of queues with different classes of customers, J. ACM 22 (1975) 248–260.

[2] P. Billingsley, *Probability and Measure* (Wiley, New York, 1979).
[3] N. Bleistein and R.A. Handelsman, *Asymptotic Expansions of Integrals* (Holt, Rinehart, and Winston, New York, 1975).
[4] P. Bratley, B.L. Fox and L.E. Schrage, *A Guide to Simulation* (Springer, New York, 1987).
[5] W.J. Gordon and G.F. Newell, Closed queueing systems with exponential servers, Oper. Res. 15 (1967) 254–265.
[6] F. Kelly, Loss networks, Ann. Appl. Prob. 1 (1991) 319–378.
[7] J. McKenna and D. Mitra, Integral representations and asymptotic expansions for closed Markovian queueing networks: Normal usage, Bell Syst. Tech. J. 61 (1982) 661–683.
[8] J. McKenna and D. Mitra, Asymptotic expansions and integral representations of moments of queue lengths in closed Markovian networks, J. ACM 31 (1984) 346–360.
[9] J. McKenna, D. Mitra and K.G. Ramakrishnan, A class of closed Markovian queueing networks: Integral representations, asymptotic expansions, and generalizations, Bell Syst. Tech. J. 60 (1981) 599–641.
[10] D. Mitra, Asymptotic optimal design of congestion control for high speed data networks, IEEE Trans. Commun. COM-40 (1992) 301–311.
[11] J.A. Morrison and D. Mitra, Heavy-usage asymptotic expansions for the waiting time in closed processor-sharing systems with multiple classes, Adv. Appl. Prob. 17 (1985) 163–185.
[12] B. Pittel, Closed exponential networks of queues with saturation: the Jackson-type stationary distribution and its asymptotic analysis, Math. Oper. Res. 4 (1979) 357–378.
[13] K.G. Ramakrishnan and D. Mitra, PANACEA: An integrated set of tools for performance analysis, Technical report, AT&T Bell Laboratories (1989).
[14] K.G. Ramakrishnan and D. Mitra, An overview of PANACEA, a software package for analyzing Markovian queueing networks, Bell Syst. Tech. J. 61 (1982) 568–592.
[15] K.W. Ross, D.H.K. Tsang and J. Wang, Monte Carlo summation and integration applied to multichain queueing networks, to appear in J. ACM.
[16] K.W. Ross and J. Wang, Asymptotically optimal importance sampling for product-form queueing networks, ACM Trans. Modeling Comp. Simul., to appear.
[17] W. Whitt, Open and closed models for networks of queues, Bell Syst. Tech. J. 63 (1984) 1911–1979.