THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 61 | May–June 1982 | Number 5

Integral Representations and Asymptotic Expansions for Closed Markovian Queueing Networks: Normal Usage

By J. McKENNA and D. MITRA

(Manuscript received October 28, 1982)

In designing a computer system, it is vitally important to be able to predict the performance of the system. Often, quantities such as throughput, processor utilization, and response time can be predicted from a closed queueing network model. However, until now the computations involved were not feasible for large systems which call for models with many processing centers and many jobs distributed over many classes. We give a radically new approach for handling such large networks—an approach that begins with a representation of the quantities of interest as ratios of integrals. These integrals contain a large parameter reflecting the size of the network. Next, expansions of the integrals in inverse powers of this large parameter are derived. For cases in which the number of processing centers is greater than one, this is the only technique we know of that yields the complete asymptotic expansion. Our method for computing the terms of the expansion can be interpreted as decomposing the original network into a large number of small "pseudonetworks." Our technique also yields easily computed error bounds when only the first few terms of the expansion are used. This method has been implemented in a software package with which we can analyze systems larger by several orders of magnitude than was previously possible.

I. INTRODUCTION

Closed Markovian queueing networks, which are tractable in having the product form (or separability) in their stationary distribution, continue to have a profound influence on computer communication, computer systems analysis, and traffic theory.[1–4] The closed networks have been used to model multiple-resource computer systems,[2,5] multiprogrammed computer systems,[2,6–8] time-sharing,[2] and window flow control in computer communication networks;[9,10] networks with blocking[11,12] require the analysis of a large number of closed networks. Not surprisingly, considerable effort has gone into devising efficient procedures for computing the partition function,[13–17] an element of the product form solution requiring significant computation. More recently, mainly spurred by parallel technological development in computer communication, there has been a focusing of effort on *large* closed networks[11,15,18–21] with many classes of jobs and transactions and large populations in each class. The point of departure of this effort is the realization that the earlier recursive techniques for computing the partition function are severely limited in terms of computing time, memory storage, and attainable accuracy when it comes to the large networks presently demanding analysis.

In an earlier paper,[2] we introduced a new approach to calculating the partition function. We showed there that the partition function could be represented as an integral containing a large parameter which, in some sense, reflected the large size of the network. In general, the partition function is represented by a multiple integral. However, in the special case where there is only one node at which queueing can occur (a node of type 1, 2, or 4 in the terminology of Ref. 4), the partition function is represented by a single integral. In Ref. 21 we applied standard theory to obtain asymptotic expansions of the integral representation. The standard techniques, however, cannot be extended directly to multiple integrals. In this paper, we make use of the special properties of our integral representation and obtain a method for generating asymptotic expansions of our integrals which works equally well for single or multiple integrals. For our single integrals, the techniques developed in this paper are easier to apply than the standard techniques we used in Ref. 21. For our multiple integrals, our technique is the only one we know of which can be used to obtain higher order terms in the asymptotic expansion.

The computational effort of solving a large network with p classes and with populations in each class of the order of magnitude of 100 is here reduced to be roughly as complex as solving by older techniques for the partition functions of (p choose 4) networks where each of these networks has a total population of, at most, seven allocated over, at most, four classes. Thus, we have reduced the problem to the solution of a large number of small problems. One consequence is that for p large enough, even our technique will be computationally intractable. Our results nevertheless allow large, previously intractable networks to be solved with vastly reduced computational effort and with great accuracy. For example, a network with 20 classes, each class having a population of 100, can be handled easily by our technique. Two noteworthy aspects of this work are (*i*) the built-in notion of depth of accuracy: by computing more terms of the asymptotic expansion, it is possible to match computational effort to desired degree of accuracy, and (*ii*) a comprehensive error analysis that allows estimates to be accompanied by sharp error bounds with little incremental effort. Let us elaborate.

Section III recapitulates and extends the results in Ref. 21 to obtain representations as integrals of the partition function of most, but not all, closed product-form networks.[4] The class-by-class breakdown of the utilization of each processor, itself simply related to mean response time and throughput, is given in terms of a ratio of two integrals. These are multiple integrals with multiplicity equal to the number of centers in the network where queues may form.

The asymptotic expansions are in powers of $(1/N)$ where N is a parameter designed to reflect network size. Our computational experience has been that five terms in the mean value expansions are generally adequate both for large networks, for which our asymptotic technique is particularly well suited, and, to a surprising extent, for small networks as well.[21] Section IV gives the procedure for generating the general coefficient of the expansion, while the leading coefficients are explicitly derived.

A remarkable feature of the composition of the coefficients make their computation amenable to various techniques. As shown in Section 4.4, the coefficients turn out to be very simply related to the partition function of a certain hypothetical network which we call the pseudonetwork. The topology is related but not identical to that of the given network and the processing rates are quite different. Most importantly, however, to compute the leading expansion coefficients, it is necessary to compute the partition function for only small populations in the pseudonetwork. Thus, to compute five terms of the utilization expansion, we need only consider the total population over all classes to be at most seven in the pseudonetwork. Small population in the pseudonetwork has the consequence of its partition function being solvable by existing recursive techniques of proven efficacy.

Section V proves that the series in $(1/N)$ given in Section IV is endowed with properties substantially more desirable than those possessed by asymptotic expansions in general. It is shown in Section 5.2 that the truncation error is numerically less than the first neglected term of the expansion, and has the same sign. Thus, except for the effort in computing an additional term, all that is generally needed for an error analysis is already available in the basic series that is computed.

We note here that while he mathematical literature on single integrals is extensive,[22–24] th re is little on asymptotic expansions of multiple integrals. Two notable investigations along lines different from here are Bleistein's[25] and Skinner's.[26] Both are marked by extreme complexity. Bleistein gives the leading coefficient, while Skinner obtains the second term. Both terms are quite complicated.

Let us now elaborate on some limitations of the paper. An important conclusion of Ref. 21 is that qualitatively different expansions of the integrals exist depending on whether usage is "normal," "high," or "very high." This is even more true in the present context of multiple processing centers. Therefore, this paper is devoted exclusively to the case of normal usage. We propose to consider the remaining usage conditions in the future. Exactly what is meant by normal usage is explained in Section 4.1.

The paper also assumes that, for each class of jobs, the routing through the network contains at least one infinite server (IS) center. It turns out that for networks in which this is not true, the asymptotic expansions are more appropriately derived in the context of either high or very high usage conditions. However, certain basic results on the integral representations of partition functions and mean values are derived in Section III regardless of whether IS centers exist for all classes. This paper does *not* allow load dependent service rates in the first-come-first-serve centers.

The results in this paper have been incorporated in a large software package and this will be reported elsewhere. No results on moments of individual queue lengths are given here. However, their integral representations and asymptotic expansions are very similar,[21] and the details will be published elsewhere.

II. PRODUCT FORM IN STATIONARY DISTRIBUTIONS: PRELIMINARIES

2.1 Product form

We recapitulate some of the well-known results[4] concerning product form in stochastic networks and present them in the form that will be used later.

Let p be the number of classes of jobs and reserve the symbol j for indexing class. Hence, when the index for summation or multiplication is omitted, it is understood that the missing index is j, where $1 \le j \le p$. A total of s service centers are allowed. We will find it natural to distinguish the centers of types 1, 2, and 4 which have queueing from the remaining centers of type 3 which do not. (The definition of type 1 to 4 centers is given in Ref. 4.) Thus, centers 1 through q will be the queueing centers, while $(q + 1)$ through s will be the type 3 centers, which have also been called think nodes and IS nodes. We reserve the symbol i for indexing centers. Also, whenever class and center indices appear together, the first always refers to class.

Let the equilibrium probability of finding n_{ji} jobs of class j at center i, $1 \le j \le p$, $1 \le i \le s$, be $\pi(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_s)$, where

$$\mathbf{y}_i \triangleq (n_{1i}, n_{2i}, \cdots, n_{pi}), \qquad 1 \le i \le s. \tag{1}$$

Closed networks are characterized by conservation of jobs in each class. That is, the population of jobs of the jth class is constant at K_j, say. The well-known results on closed networks with the product form in its stationary distribution may be given in the following form:

$$\pi(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \frac{1}{G} \prod_{i=1}^{s} \pi_i(\mathbf{y}_i), \tag{2}$$

where

$$\begin{aligned} \pi_i(\mathbf{y}_i) &= \left(\sum n_{ji}\right)! \prod \left(\frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}\right), \qquad 1 \le i \le q, \\ &= \prod \left(\frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}\right), \qquad (q + 1) \le i \le s. \end{aligned} \tag{3}$$

In the above formulas, we have taken into account the previously stated assumption; namely, for the first-come-first-served discipline in type 1 centers, the service rate is independent of the number of jobs in queue. Also, in (3),

$$\rho_{ji} = \frac{\text{expected number of visits of class } j \text{ jobs to center } i}{\text{service rate of class } j \text{ jobs in center } i}, \tag{4}$$

where the numerator is obtained from the given routing matrix by solving for the eigenvector corresponding to the eigenvalue at 1.

In (2), G is the partition function, and it is explicitly

$$G(\mathbf{K}) = \sum_{\mathbf{1}'\mathbf{n}_1 = K_1} \cdots \sum_{\mathbf{1}'\mathbf{n}_p = K_1} \prod_{i=1}^{s} \pi_i(\mathbf{y}_i), \tag{5}$$

where we have written $\mathbf{1}'\mathbf{n}_j$ for $\sum_{i=1}^{s} n_{ji}$ and the condition $\mathbf{1}'\mathbf{n}_j = \boldsymbol{K}_j$ to indicate the conservation of jobs in each class. Thus,

$$G(\mathbf{K}) = \sum \cdots \sum \left[\prod_{i=1}^{q} \left\{ \left(\sum n_{ji}\right)! \prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!} \right\} \right] \left[\prod_{i=q+1}^{s} \left\{ \prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!} \right\} \right]. \tag{6}$$

In Section III, we will refer to this expression for the partition function.

2.2 *Asymptotic expansions*

A series

$$\sum_{k=0}^{\infty} A_k/N^k$$

is said to be an asymptotic expansion[22-24] of a function $I(N)$ if

$$I(N) - \sum_{k=0}^{m-1} A_k/N^k = 0(N^{-m}) \text{ as } N \to \infty \tag{7}$$

for every $m = 1, 2, \cdots$. We write

$$I(N) \sim \sum_{k=0}^{\infty} A_k/N^k.$$

The series may be either convergent or divergent.

III. INTEGRAL REPRESENTATIONS

As the representations presented here are basic to the subsequent development, we have allowed some duplication in Section 3.1 with Section 10 of Reference 21.

3.1 *Partition function*

We start with Euler's integral representation for the factorial,

$$n! = \int_0^{\infty} e^{-u} u^n du. \tag{8}$$

Returning to (6), we use this representation to write

$$(\sum n_{ji})! = \int_0^{\infty} e^{-u_i} \prod u_i^{n_{ji}} du_i, \qquad i = 1, 2, \cdots, q. \tag{9}$$

Substitution in (6) gives

$$G = \int_0^{\infty} \cdots \int_0^{\infty} \exp\left(-\sum_{i=1}^{q} u_i\right) \sum_{1'\mathbf{n}_1 = K_1} \cdots \sum_{1'\mathbf{n}_p = K_p} \left[\prod_{i=1}^{q} \left\{\prod \frac{(\rho_{ji} u_i)^{n_{ji}}}{n_{ji}!}\right\}\right] \cdot \left[\prod_{i=q+1}^{s} \left\{\prod \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}\right\}\right] du_1 \cdots du_q. \tag{10}$$

Now by the multinomial theorem,

$$G = (\prod K_j!)^{-1} \int_0^{\infty} \cdots \int_0^{\infty} \exp\left(-\sum_{i=1}^{q} u_i\right) \cdot \prod \left\{\sum_{i=1}^{q} \rho_{ji} u_i + \sum_{i=q+1}^{s} \rho_{ji}\right\}^{K_j} du_1 \cdots du_q. \tag{11}$$

It is noteworthy but not surprising that the parameters ρ_{ji} for all the type 3 centers appear lumped together. Hence, we may simplify the notation by introducing ρ_{j0}, where

$$\rho_{j0} = \sum_{i=q+1}^{s} \rho_{ji}, \qquad j = 1, 2, \cdots, p. \tag{12}$$

Another consequence of the notation is that the center index i may henceforth be understood to range over the processing centers only, i.e., $1 \leq i \leq q$.

The new quantity ρ_{j0} has the physical significance of being the weighted combination of all the mean think times of the IS centers in the routing of the jth class. In particular, if the routing of the jth class contains at least one IS center, then $\rho_{j0} > 0$ and otherwise $\rho_{j0} = 0$. Let I be the collection of indices of classes of the former type and let I^* be the complementary collection, i.e.,

$$j \in I \Leftrightarrow \rho_{j0} > 0 \quad \text{and} \quad j \in I^* \Leftrightarrow \rho_{j0} = 0. \tag{13}$$

With this notation,

$$G = \left[\prod_{j \in I} \rho_{j0}^{K_j} \Big/ \prod_j K_j!\right] \int_0^\infty \cdots \int_0^\infty \exp - \left(\sum u_i\right) \times \prod_{j \in I} \left\{1 + \sum_i \frac{\rho_{ji}}{\rho_{j0}} u_i\right\}^{K_j} \prod_{j \in I^*} \left\{\sum_i \rho_{ji} u_i\right\}^{K_j} du_1 \cdots du_q. \tag{14}$$

In vector notation, which we shall use widely, this reduces to

$$G = \left[\prod_{j \in I} \rho_{j0}^{K_j} \Big/ \prod_j K_j!\right] \int_{Q+} e^{-\mathbf{1}'\mathbf{u}} \prod_j (\delta_{jI} + \mathbf{r}'_j \mathbf{u})^{K_j} d\mathbf{u}, \tag{15}$$

where*

$$\begin{aligned}
\mathbf{u} &= (u_1, u_2, \cdots, u_q)' \\
\mathbf{1} &= (1, 1, \cdots, 1)' \\
\mathbf{r}_j &= (r_{j1}, r_{j2}, \cdots, r_{jq})', \qquad 1 \leq j \leq p \\
r_{ji} &= \rho_{ji}/\rho_{j0} \quad \text{if} \quad j \in I \\
&= \rho_{ji} \quad \text{if} \quad j \in I^* \\
\delta_{jI} &= 1 \quad \text{if} \quad j \in I \\
&= 0 \quad \text{if} \quad j \in I^* \\
Q^+ &= \{\mathbf{u} \mid \mathbf{u} \geq \mathbf{o}\}.
\end{aligned} \tag{16}$$

* Unfortunately, r_{ji} here is defined to be the reciprocal of the natural extension of r_j in Ref. 21.

We now introduce the large parameter N and define

$$\beta_j \triangleq K_j/N, \qquad 1 \le j \le p, \tag{17}$$

$$\mathbf{\Gamma}_j \triangleq N\mathbf{r}_j, \qquad 1 \le j \le p. \tag{18}$$

The suggestion in the notation is that in the generic large network $\{\beta_j\}$ and $\{\Gamma_{ji}\}$ are 0(1). That is, the ratio of processing time to think time is, in order of magnitude estimation, proportionately less for increased populations. There is great latitude in the choice of the large parameter. The guiding principle in choosing it should be that the resulting values of $\{\beta_j\}$ and $\{\Gamma_{ji}\}$ are as uniformly close to 1 as possible. In practice, we have used

$$N = \max_{ij} \left\{ \frac{1}{r_{ji}} \right\}. \tag{19}$$

On substituting (17) and (18) into (15) and after the change of variables $\mathbf{z} = \mathbf{u}/N$, we obtain from (15) another useful integral representation of the partition function which is distinguished by its dependence on N. Summarizing for future reference, we have

Proposition 1:

$$G(\mathbf{K}) = \left[\prod_{j \in I} \rho_{j0}^{K_j} \Big/ \prod_j K_j! \right] \int_{Q^+} e^{-\mathbf{1}'\mathbf{u}} \prod_j (\delta_{jI} + \mathbf{r'}_j\mathbf{u})^{K_j} d\mathbf{u} \tag{20}$$

$$= \left[N^q \prod_{j \in I} \rho_{j0}^{K_j} \Big/ \prod_j K_j! \right] \int_{Q^+} e^{-Nf(\mathbf{z})} d\mathbf{z}, \tag{21}$$

where

$$f(\mathbf{z}) \triangleq \mathbf{1'z} - \sum_{k=1}^{p} \beta_j \log(\delta_{jI} + \mathbf{\Gamma'}_j\mathbf{z}). \tag{22}$$

□

3.2 Mean values

We restrict our attention to the mean value $u_{\sigma i}(\mathbf{K})$ which gives the utilization of the ith processor by jobs of the σth class for a population distribution by class in the network denoted by $\mathbf{K} = (K_1, K_2, \cdots, K_p)$. Other interesting mean performance indices, such as throughput and mean response time are known to be simply related to $\{u_{\sigma i}(\mathbf{K})\}$ and the interested reader may consult Ref. 17.

In Ref. 17,

$$u_{\sigma i}(\mathbf{K}) = \rho_{\sigma i} \frac{G(\mathbf{K} - \mathbf{e}_\sigma)}{G(\mathbf{K})}, \qquad 1 \le \sigma \le p, 1 \le i \le q, \tag{23}$$

where $\mathbf{e}_\sigma$ is our notation for the vector with the σth component unity and all other components zero. Thus, the value for the partition

function is needed for the given population distribution and also for the population in the σth class reduced by 1.
Now from (20)

$$G(\mathbf{K}+\mathbf{e}_\sigma) = \frac{\rho_{\sigma 0} \prod_{j\in I} \rho_{j0}^{K_j}}{(K_\sigma + 1) \prod K_j!} \int_{Q^+} e^{-\mathbf{1}'\mathbf{u}}(1 + \mathbf{r}_\sigma'\mathbf{u}) \times \prod_j (\delta_{jI} + \mathbf{r}_j'\mathbf{u})^{K_j} d\mathbf{u} \quad \text{if} \quad \sigma \in I \tag{24}$$

$$= \frac{\prod_{j\in I} \rho_{j0}^{K_j}}{(K_\sigma + 1) \prod K_j!} \int_{Q^+} e^{-\mathbf{1}'\mathbf{u}}(\mathbf{r}_\sigma'\mathbf{u}) \times \prod_j (\delta_{jI} + \mathbf{r}_j'\mathbf{u})^{K_j} d\mathbf{u} \quad \text{if} \quad \sigma \in I^*. \tag{25}$$

From (23) to (25) and the same change of variables, namely $\mathbf{z} = \mathbf{u}/N$, employed in transforming (20) to (21) we obtain the following representation of the utilization in terms of integrals,

Proposition 2: For class index σ, $1 \le \sigma \le p$, and center index i, $1 \le i \le q$,

$$u_{\sigma i}(\mathbf{K}+\mathbf{e}_\sigma)^{-1} = \left\{\frac{1}{r_{\sigma i}(K_\sigma + 1)}\right\} \left[\delta_{\sigma I} + \frac{\int_{Q^+} (\mathbf{\Gamma}_\sigma'\mathbf{z}) e^{-Nf(\mathbf{z})} d\mathbf{z}}{\int_{Q^+} e^{-Nf(\mathbf{z})} d\mathbf{z}}\right]. \tag{26}$$

□

Some digressionary comments are as follows. Note that the above conceals that $r_{\sigma i}$ is normalized differently, but not unexpectedly, depending on whether $\sigma \in I$ or otherwise [see (16)]. Also note that in the typical large network for normal operating conditions we always expect $r_{\sigma i}$ to be $0(1/N)$, precisely because of the normalization used, and K_σ to be $0(N)$ so that the term in braces in (26) is then $0(1)$.

IV. ASYMPTOTIC EXPANSIONS

We henceforth consider only networks in which the route for each class always contains an infinite server center. Specifically,

$$\rho_{j0} > 0, \qquad j = 1, 2, \cdots, p, \tag{27}$$

and the set I^* is empty.

4.1 The assumption of "normal usage"

Define

$$\boldsymbol{\alpha} \triangleq \mathbf{1} - \sum \beta_j \mathbf{\Gamma}_j \tag{28}$$

so that in terms of the original network parameters

$$\alpha_i = 1 - \sum K_j \frac{\rho_{ji}}{\rho_{j0}}, \qquad i = 1, 2, \cdots, q. \tag{29}$$

It is important to note that α is independent of the choice of N. The parameter $\alpha_i(-\infty < \alpha_i < 1)$ is an indicator of the unutilized processing capability of the ith center. Positive values of α_i correspond to less than 100 percent utilization of the processor and negative values which, of course, can occur to very high utilizations.* Normal usage in large networks will almost certainly require $\alpha_i > 0$, and in all likelihood α_i will not be close to 0 for all i.

Hereafter, we assume

$$\alpha_i > 0, \qquad i = 1, 2, \cdots, q, \tag{30}$$

which condition we refer to as normal usage. Moreover, as α_i for some i comes close to 0, the expansions given here are not as efficient as those derived specifically for such conditions and which we propose to give in the future.

A justification of the usage interpretation that we have given to $\boldsymbol{\alpha}$ is provided by a result obtained later (see below Proposition 6) which states that, asymptotic with network size,

$$u_i = \text{utilization of } i\text{th processor} \sim 1 - \alpha_i. \tag{31}$$

An obvious caveat is that this result is derived for the assumption in (30). However, as the utilization in (31) can come close to unity even while (30) is satisfied, (31) suggests that for large networks normal usage will not extend beyond the range $\boldsymbol{\alpha} > \mathbf{0}$.

Observe that

$$f(\mathbf{0}) = 0 \tag{32}$$

$$\text{and} \quad \nabla f(\mathbf{z}) = \mathbf{1} - \sum_j \frac{\beta_j}{1 + \boldsymbol{\Gamma}_j'\mathbf{z}} \boldsymbol{\Gamma}_j \tag{33}$$

so that

$$\boldsymbol{\alpha} = \nabla f(\mathbf{0}). \tag{34}$$

The assumption of $\boldsymbol{\alpha} > \mathbf{0}$ and the form in (33) ensures that the function f has no stationary points in Q^+ since

$$\nabla f(\mathbf{z}) \geq \nabla f(\mathbf{0}) > \mathbf{0}, \qquad \mathbf{z} \in Q^+. \tag{35}$$

Also observe that

---

* Unfortunately, $\boldsymbol{\alpha}$ here has an opposite sign from the natural extension of α as defined in Ref. 21.

$$\left\{\frac{\partial^2 f}{\partial z_{i_1}\partial z_{i_2}}\right\} = \sum \frac{\beta_j}{(1+\mathbf{\Gamma}_j'\mathbf{z})^2}\mathbf{\Gamma}_j\mathbf{\Gamma}_j' \tag{36}$$

from which we note that the Hessian is positive semidefinite.

To conclude, with normal usage, f is a convex function, with its minimum in Q^+ attained at 0 and with no point in Q^+ where its gradient vanishes.

4.2 Transformations on integrals exploiting normal usage

Consider the following transformations on the basic integral:

$$\begin{aligned}
\int_{Q^+} e^{-Nf(\mathbf{z})}d\mathbf{z} &= \int_{Q^+} e^{-Nf(\mathbf{z})+N\sum\beta_j(\mathbf{\Gamma}_j'\mathbf{z})-N\sum\beta_j(\mathbf{\Gamma}_j'\mathbf{z})}d\mathbf{z} \\
&= \int_{Q^+} e^{-N\alpha'\mathbf{z}}\exp \\
&\quad -\left[N\sum_j \beta_j\{\mathbf{\Gamma}_j'\mathbf{z} - \log(1+\mathbf{\Gamma}_j'\mathbf{z})\}\right]d\mathbf{z} \\
&= N^{-q}\int_{Q^+} e^{-\alpha'\mathbf{u}}\exp \\
&\quad -\left[\sum_j \beta_j\left\{\mathbf{\Gamma}_j'\mathbf{u} - N\log\left(1+\frac{1}{N}\mathbf{\Gamma}_j'\mathbf{u}\right)\right\}\right]d\mathbf{u},
\end{aligned} \tag{37}$$

where $\mathbf{u} = N\mathbf{z}$. Now make the following change of variables,

$$v_i \triangleq \alpha_i u_i, \qquad 1 \le i \le q \tag{38}$$

and normalize the system parameters with respect to α, thus,

$$\tilde{\Gamma}_{ji} \triangleq \Gamma_{ji}/\alpha_i, \qquad 1 \le j \le p,\ 1 \le i \le q. \tag{39}$$

Observe that in particular

$$\mathbf{\Gamma}_j'\mathbf{u} = \tilde{\mathbf{\Gamma}}_j'\mathbf{v}. \tag{40}$$

From (37),

$$\int_{Q^+} e^{-Nf(\mathbf{z})}d\mathbf{z} = \frac{N^{-q}}{(\Pi\alpha_i)}\int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}H(N^{-1},\mathbf{v})d\mathbf{v}, \tag{41}$$

where

$$H(N^{-1},\mathbf{v}) \triangleq e^{s(N^{-1},\mathbf{v})}, \tag{42}$$

$$s(N^{-1},\mathbf{v}) \triangleq -\sum_{j=1}^{p}\beta_j\left\{\tilde{\mathbf{\Gamma}}_j'\mathbf{v} - N\log\left(1+\frac{1}{N}\tilde{\mathbf{\Gamma}}_j'\mathbf{v}\right)\right\}. \tag{43}$$

Our notation here suggesting N^{-1} as the independent variable may be perplexing at this stage, but it provides a clue to the direction of the analysis.

Here are two further digressionary comments: the above transformations are meaningful only in the context of normal usage, i.e., $\boldsymbol{\alpha} > \mathbf{0}$. In a similar vein, an interpretation of $\{\tilde{\Gamma}_{ji}\}$ as renormalized $\{\Gamma_{ji}\}$ is only meaningful for $\boldsymbol{\alpha} > \mathbf{0}$. It is noteworthy that hereafter we shall be dealing exclusively with $\{\tilde{\Gamma}_{ji}\}$ and not at all with $\{\Gamma_{ji}\}$.

We need to repeat the transformations given above for the integral $\int(\boldsymbol{\Gamma}'_\sigma\mathbf{z})e^{-Nf(\mathbf{z})}d\mathbf{z}$. The result may be combined with (42) and (43) to give the following compact representation:

$$\int_{Q^+}(\boldsymbol{\Gamma}'_\sigma\mathbf{z})^m e^{-Nf(\mathbf{z})}d\mathbf{z} = \frac{N^{-(q+m)}}{(\Pi\ \alpha_i)}\int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}(\tilde{\boldsymbol{\Gamma}}'_\sigma\mathbf{v})^m H(N^{-1}, \mathbf{v})d\mathbf{v}, \quad (44)$$

$$m = 0, 1, 2, \cdots. \qquad \square$$

We may now use these expressions in Proposition 2 to obtain for $1 \leq \sigma \leq p$ and $1 \leq i \leq q$,

$$u_{\sigma i}(\mathbf{K} + \mathbf{e}_\sigma)^{-1} = \left\{\frac{\rho_{\sigma 0}}{\rho_{\sigma i}(K_\sigma + 1)}\right\} \cdot \left[1 + \frac{1}{N}\frac{\int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}(\tilde{\boldsymbol{\Gamma}}'_\sigma\mathbf{v})H(N^{-1}, \mathbf{v})d\mathbf{v}}{\int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}H(N^{-1}, \mathbf{v})d\mathbf{v}}\right]. \quad (45)$$

Note that the bracketed term is independent of the processing center index i.

Let us agree to call

$$I(N) \triangleq \int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}H(N^{-1}, \mathbf{v})d\mathbf{v} \quad (46)$$

and

$$I_\sigma^{(1)}(N) \triangleq \int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}(\tilde{\boldsymbol{\Gamma}}'_\sigma\mathbf{v})H(N^{-1}, \mathbf{v})d\mathbf{v}, \quad (47)$$

where the superscript (1) is a mnemonic for first moment. In this notation, we have for future reference

Proposition 3:

$$u_{\sigma i}(\mathbf{K} + \mathbf{e}_\sigma)^{-1} = \left\{\frac{\rho_{\sigma 0}}{\rho_{\sigma i}(K_\sigma + 1)}\right\}\left[1 + \frac{1}{N}\frac{I_\sigma^{(1)}(N)}{I(N)}\right]. \quad (48)$$

$\square$

The asymptotic expansions of $I(N)$ and $I_\sigma^{(1)}(N)$ are considered next.

4.3 Asymptotic expansions

This section will outline the procedure for obtaining the asymptotic expansions together with plausibility arguments. The proofs of the assertions concerning asymptoticity will follow in Section V. Moreover, we defer till Section 4.4 certain observations which make feasible the efficient computation of the coefficients of the asymptotic expansions.

Our procedure for $I(N)$ is to first obtain a power series in N^{-1} of $H(N^{-1}, v)$ and then to integrate term by term. Thus, we let

$$H(N^{-1}, \mathbf{v}) = \sum_{k=0}^{\infty} \frac{h_k(\mathbf{v})}{N^k} \tag{49}$$

and

$$A_k \triangleq \int_{Q^+} e^{-1'\mathbf{v}} h_k(\mathbf{v}) d\mathbf{v} \tag{50}$$

and claim that

$$I(N) \sim \sum_{k=0}^{\infty} \frac{A_k}{N^k}. \tag{51}$$

Let us elaborate on the coefficients $\{h_k(\mathbf{v})\}$ in (49).

$$h_k(\mathbf{v}) = \frac{1}{k!} \frac{\partial^k}{\partial(1/N)^k} H(0, \mathbf{v}), \qquad k = 0, 1, 2, \cdots. \tag{52}$$

To make these explicit, we need to first introduce

$$f_k(\mathbf{v}) \triangleq \frac{(-1)^k}{k} \sum_j \beta_j (\tilde{\mathbf{\Gamma}}_j' \mathbf{v})^k, \qquad k = 1, 2, \cdots. \tag{53}$$

Their role becomes clear if we recall (42):

$$H(N^{-1}, \mathbf{v}) = e^{s(N^{-1}, \mathbf{v})}, \tag{54}$$

and note from (43) that for fixed $\mathbf{v} \in Q^+$, $s(N^{-1}, \mathbf{v})$ and, hence, $H(N^{-1}, \mathbf{v})$ are functions of N^{-1}, analytic in $\mathrm{Re}(N^{-1}) > \epsilon(\mathbf{v})$, where $\epsilon(\mathbf{v}) < 0$. Then,

$$s^{(k)}(0, \mathbf{v}) = -k! f_{k+1}(\mathbf{v}), \qquad k = 1, 2, \cdots. \tag{55}$$

To proceed now to the derivatives of $H(\cdot, \mathbf{v})$ itself, we will find useful the following expression in which it is understood that all derivatives are with respect to N^{-1}:

$$H^{(k+1)}(N^{-1}, \mathbf{v}) = \sum_{m=0}^{k} \binom{k}{m} s^{(k+1-m)}(N^{-1}, \mathbf{v}) H^{(m)}(N^{-1}, \mathbf{v}),$$
$$k = 0, 1, \cdots. \tag{56}$$

From (52), (55) and the above, it is easy to see that there exists the following recursive scheme for generating $\{h_k(\mathbf{v})\}$:

$$\left.\begin{aligned} h_0(\mathbf{v}) &\equiv 1 \\ h_{k+1}(\mathbf{v}) &= -\frac{1}{k+1}\sum_{m=0}^{k}(k+1-m)f_{k+2-m}(\mathbf{v})h_m(\mathbf{v}), \\ &\cdot k = 0, 1, 2, \cdots \end{aligned}\right\}. \tag{57}$$

In particular, the leading elements are

$$\begin{aligned} h_k(\mathbf{v}) &\equiv 1, \qquad k=0 \\ &= -f_2(\mathbf{v}), \qquad k=1 \\ &= -f_3(\mathbf{v}) + \frac{1}{2}f_2^2(\mathbf{v}), \qquad k=2 \\ &= -f_4(\mathbf{v}) + f_2(\mathbf{v})f_3(\mathbf{v}) - \frac{1}{6}f_2^3(\mathbf{v}), \qquad k=3. \end{aligned} \tag{58}$$

To summarize the steps discussed so far in the generation of the asymptotic expansion, we have

Proposition 4:

$$I(N) \sim \sum_{k=0}^{\infty}\frac{A_k}{N^k}, \tag{59}$$

where $A_k = \int e^{-\mathbf{1}'\mathbf{v}}h_k(\mathbf{v})d\mathbf{v}$, $\{h_k(\mathbf{v})\}$ is obtained recursively from (57) with leading elements exhibited in (58), and $\{f_k(\mathbf{v})\}$ is as in (53). □

The aspect of the above asymptotic expansion of the integral $I(N)$, which consists of decomposing the integrand into the product of an exponential and a function, the expansion of the latter function in a power series, and the final term-by-term integration is like the procedure which, in the context of single integrals, is justified by Watson's Lemma[23] under certain conditions. Our contribution has been to show that a generalization of this fundamental result exists for the multiple integrals of interest in stochastic networks.

Our procedure for obtaining the asymptotic expansion for $I_\sigma^{(1)}(N)$ is very similar and consists of obtaining a power series in N^{-1} of $(\tilde{\mathbf{\Gamma}}'_\sigma\mathbf{v})H(N^{-1},\mathbf{v})$ and integrating term by term. However, we notice the simplifying fact that

$$(\tilde{\mathbf{\Gamma}}'_\sigma\mathbf{v})H(N^{-1}, \mathbf{v}) = \sum_{k=0}^{\infty}\frac{(\tilde{\mathbf{\Gamma}}'_\sigma\mathbf{v})h_k(\mathbf{v})}{N^k}. \tag{60}$$

Thus, the procedure for this integral is as follows:

Proposition 5:

$$I_\sigma^{(1)}(N) \sim \sum_{k=0}^{\infty} \frac{A_{\sigma,k}^{(1)}}{N^k}, \tag{61}$$

where

$$A_{\sigma,k}^{(1)} \triangleq \int_{Q^+} e^{-\mathbf{1}'\mathbf{v}}(\tilde{\Gamma}_\sigma'\mathbf{v})h_k(\mathbf{v})\mathbf{dv}, \qquad k = 0, 1, 2, \cdots. \tag{62}$$

□

As observed in Ref. 21, the asymptotic expansion for the integrals may be used to generate asymptotic expansions for their ratios on account of powers of N^{-1} forming a multiplicative sequence.[22] Thus, the coefficients of an asymptotic expansion for $u_{\sigma i}(\mathbf{K} + \mathbf{e}_\sigma)^{-1}$ may be obtained from formal substitution in Proposition 2 of the expansions in Propositions 4 and 5. This gives

Proposition 6:

$$\left\{\frac{\rho_{\sigma i}(K_\sigma + 1)}{\rho_{\sigma 0}}\right\} u_{\sigma i}(\mathbf{K} + \mathbf{e}_\sigma)^{-1} \sim 1 + \frac{1}{N}\sum_{k=0}^{\infty} \frac{B_{\sigma,k}}{N^k}, \tag{63}$$

$$\begin{aligned} B_{\sigma,k} &= \frac{A_{\sigma,k}^{(1)}}{A_0}, \quad k = 0 \\ &= \frac{1}{A_0}\left[A_{\sigma,k}^{(1)} - \sum_{m=1}^{k} A_m B_{\sigma,k-m}\right], \qquad k = 1, 2, \cdots. \end{aligned} \tag{64}$$

□

With the above proposition we may generate $(k + 1)$ terms of the expansion for $u_{\sigma i}$ from k terms of $I(N)$ and $I_\sigma^{(1)}(N)$.

An immediate corollary to the above proposition is

$$u_{\sigma i}(\mathbf{K}) \sim \frac{\rho_{\sigma i} K_\sigma}{\rho_{\sigma 0}}$$

and

$$\begin{aligned} u_i(\mathbf{K}) &= \text{utilization of } i\text{th processor} = \sum_j u_{ji}(\mathbf{K}) \\ &\sim 1 - \alpha_i, \qquad i = 1, 2, \cdots, q, \end{aligned} \tag{65}$$

which was claimed earlier in Section 4.1 in the course of giving physical meaning to the parameters $\{\alpha_i\}$.

This corollary illustrates the important point that the terms in the asymptotic expansions A_k/N^k, $A_{\sigma,k}^{(1)}/N^k$, and $B_{\sigma,k}/N^{k+1}$ are all independent of N and depend only on the network parameters. The dummy parameter N serves to show how to group the terms of the same magnitude. This independence from N follows from the fact that α is independent of N, as noted earlier, and from the fact $f_k(v)/N^{k-1}$ is easily seen to be independent of N. Once this is noted, the result follows easily from the definitions of the terms. The choice of N is important numerically, which will be discussed in a subsequent paper.

4.4 Pseudonetworks and the computation of expansion coefficients

Here, we consider the compositions of coefficients, $\{A_k\}$ and $\{A_{\sigma,k}^{(1)}\}$, and find that quite remarkably they are related intimately to the partition function of a certain hypothetical network which we call the pseudonetwork. It turns out that to compute the leading elements of $\{A_k\}$ and $\{A_{\sigma,k}^{(1)}\}$, we need to consider the pseudonetwork with only small populations. Thus, existing techniques known to be effective for computing partition functions for small populations may be used to compute the leading coefficients of our asymptotic expansions.

An example will prove useful. From Proposition 4 and eq. (58) we see that

$$A_3 = \int e^{-\mathbf{1}'\mathbf{v}}\left\{-f_4(\mathbf{v}) + f_2(\mathbf{v})f_3(\mathbf{v}) - \frac{1}{6}f_2^3(\mathbf{v})\right\}d\mathbf{v}. \tag{66}$$

Now consider only the third term after denoting it by A_{33}. Using the expression for $f_2(\mathbf{v})$ as given by (53), we obtain

$$\begin{aligned} A_{33} = &-\frac{1}{48}\sum_j \beta_j^3 \int e^{-\mathbf{1}'\mathbf{v}}(\tilde{\mathbf{\Gamma}}_j'\mathbf{v})^6 d\mathbf{v} \\ &-\frac{1}{16}\sum_{j\neq k} \beta_j^2\beta_k \int e^{-\mathbf{1}'\mathbf{v}}(\tilde{\mathbf{\Gamma}}_j'\mathbf{v})^4(\tilde{\mathbf{\Gamma}}_k'\mathbf{v})^2 d\mathbf{v} \\ &-\frac{1}{48}\sum_{j\neq k\neq l\neq j} \beta_j\beta_k\beta_l \int e^{-\mathbf{1}'\mathbf{v}}(\tilde{\mathbf{\Gamma}}_j'\mathbf{v})^2(\tilde{\mathbf{\Gamma}}_k'\mathbf{v})^2(\tilde{\mathbf{\Gamma}}_l'\mathbf{v})^2 d\mathbf{v}, \end{aligned} \tag{67}$$

where the subscripts j, k, and l are class indices ranging over $[1, p]$. We now make the observation that the generic integral in the composition of the asymptotic expansion coefficients is within a multiplicative constant of

$$g(\mathbf{m}) = g(m_1, m_2, \cdots, m_p) \triangleq \frac{1}{\left(\prod_j m_j!\right)} \int_{Q^+} e^{-\mathbf{1}'\mathbf{u}} \prod_j (\tilde{\mathbf{\Gamma}}_j'\mathbf{u})^{m_j} d\mathbf{u}. \tag{68}$$

The above is an important form for we may now identify it with quantities previously encountered.

We first give the following equivalent expression for $g(\mathbf{m})$.

$$g(\mathbf{m}) = \sum_{\mathbf{1}'\mathbf{n}_1 = m_1} \cdots \sum_{\mathbf{1}'\mathbf{n}_p = m_p} \prod_{i=1}^{q} \left\{\left(\sum_j n_{ji}\right)! \prod_j \frac{\tilde{\Gamma}_{ji}^{n_{ji}}}{n_{ji}!}\right\}. \tag{68'}$$

Recall the expression in Proposition 1, eq. (20), for the integral representation of the partition function. Specialize the expression there to a network with no infinite server centers, i.e., set I is empty, and find that it reduces to

$$G(\mathbf{K}) = \frac{1}{\left(\prod_j K_j!\right)} \int_{Q^+} e^{-\mathbf{1}'\mathbf{u}} \prod_j (\mathbf{r}_j'\mathbf{u})^{K_j} d\mathbf{u}. \tag{69}$$

On comparing (68) and (69), or (68′) and (6), we may conclude that $g(\mathbf{m})$ is the partition function of a certain network.

Call this hypothetical network the pseudonetwork. What characterizes the pseudonetwork? To begin with, it is closed and lacks IS centers. There are, as in the original network, exactly q processing centers and p classes of jobs. The processing rate of jobs from the jth class in the ith center of the pseudonetwork is $\tilde{\Gamma}_{ji}$, where, you will recall, $\tilde{\Gamma}_{ji} = \Gamma_{ji}/\alpha_i$. In agreement with past convention, $(m_1, m_2, \cdots, m_p)'$ denotes in vector form the population distribution by class in the network.

We may follow the procedure outlined in the example concerning A_{33} above to express all the leading coefficients A_k, $k = 0, 1, 2, 3$ in terms of the partition function of the pseudonetwork. This gives

$$A_0 = 1$$

$$A_1 = -\sum_j \beta_j g(2\mathbf{e}_j)$$

$$A_2 = 2\sum_j \beta_j g(3\mathbf{e}_j) + 3\sum_j \beta_j^2 g(4\mathbf{e}_j) + \frac{1}{2}\sum_{j\neq k} \beta_j\beta_k g(2\mathbf{e}_j + 2\mathbf{e}_k)$$

$$\begin{aligned} A_3 = &-6\sum_j \beta_j g(4\mathbf{e}_j) - 20\sum_j \beta_j^2 g(5\mathbf{e}_j) - 15\sum_j \beta_j^3 g(6\mathbf{e}_j) \\ &- 2\sum_{j\neq k} \beta_j\beta_k g(2\mathbf{e}_j + 3\mathbf{e}_k) \\ &- 3\sum_{j\neq k} \beta_j^2\beta_k g(4\mathbf{e}_j + 2\mathbf{e}_k) \\ &- \frac{1}{6}\sum_{j\neq k\neq l\neq j} \beta_j\beta_k\beta_l g(2\mathbf{e}_j + 2\mathbf{e}_k + 2\mathbf{e}_l). \end{aligned} \tag{70}$$

In these expressions, j, k, and l are class indices each with range $[1, p]$.

Notice that in the computation of A_k, $k = 0, 1, 2, 3$, the population distribution $\mathbf{m}$ that appears in $g(\mathbf{m})$ may be characterized as being quite small. The total population in the pseudonetwork over all classes, $\sum m_j$, is at most 6. A further simplifying condition is that we need consider only distributions with three classes at most with nonzero populations—the extreme distribution arises in $g(2\mathbf{e}_j + 2\mathbf{e}_k + 2\mathbf{e}_l)$.

Since we are interested in population distributions where all but a small number of classes have no members at all, we may equivalently choose to view the pseudonetwork as a collection of smaller networks each with a small (less than p) number of classes of jobs. For example, $g(2\mathbf{e}_j + 3\mathbf{e}_k)$ may be viewed either as the partition function of the pseudonetwork in which all but classes j and k have zero population, or as originating from a network with only two classes with population distribution (2, 3) but one which is otherwise unchanged. Such distinctions, while not material to the procedures given here, may be consequential in the efficiency of the computations.

The coefficients $\{A^{(1)}_{\sigma,k}\}$ of the expansion for $I^{(1)}_{\sigma}(N)$ may also be expressed in terms of the partition function of the pseudonetwork. As the derivation is similar, it will suffice to give the results, which we do in the Appendix.

It will be observed that to compute $\{A^{(1)}_{\sigma,k}\}$, $k = 0, 1, 2, 3$ we need to consider various allocations to classes of a total population in the pseudonetwork of at most 7. Thus, to compute the leading five terms of the utilization $u_{\sigma i}$, we need A_k and $\{A^{(1)}_{\sigma,k}\}$ for $k = 0, 1, 2, 3$, and these are computed from the values of the partition function of the pseudonetwork for various allocations to classes of a total network population of at most 7. This is an elaboration of a claim made in the Introduction.

V. ERROR ANALYSIS AND PROOF OF ASYMPTOTICITY

Equations (59) and (61) contain claims requiring proof on the asymptoticity of the expansions of $I(N)$ and $I^{(1)}_{\sigma}(N)$. This is provided here as a corollary to a complete error analysis. In fact we show that the expansions given earlier have properties more attractive than that required of asymptotic expansions. For instance, asymptoticity requires that errors incurred in the estimation of the integrals from the use of, say, m leading terms is of the same *order* as the $(m + 1)$th term as $N \to \infty$. We prove that the expansions derived have the stronger property that the truncation error is bounded by the $(m + 1)$th term. The practical benefit of this analysis is that with very little incremental effort we can accompany our estimates of the mean values with sharp estimates of the estimation error.

5.1 *Completely monotonic functions* [24]

We need the following definition: for any nonnegative R, let $Q^+(R)$ be the set of nonnegative vectors with norm bounded by R, i.e., $Q^+(R) = \{\mathbf{v} \mid \mathbf{v} \geqq 0 \text{ and } \|\mathbf{v}\| \le R\}$. Note that $Q^+(R) \to Q^+$ as $R \to \infty$.

The proposition below (cf. Ref. 21) states a remarkable property of the function $H(N^{-1}, \mathbf{v})$ which is a key to much of the error analysis.

Proposition 7: For all $\mathbf{v} \in Q^+(R)$, $R < \infty$, $H(N^{-1}, \mathbf{v})$ is a completely monotonic (or alternating) function of N^{-1}. That is,

$$(-1)^k \frac{\partial^k}{\partial(1/N)^k} H(N^{-1}, \mathbf{v}) \geq 0 \quad \begin{array}{ll} \text{for} & k = 0, 1, 2, \cdots \\ \text{and} & 0 \leq N^{-1} < \infty. \end{array} \tag{71}$$

Proof: Consider the form for $H(N^{-1}, \mathbf{v})$ from (42 to 43), Section 4.2:

$$H(N^{-1}, \mathbf{v}) = \prod_j \left\{ e^{-\beta_j(\tilde{\boldsymbol{\Gamma}}_j'\mathbf{v})} \left(1 + \frac{1}{N} \tilde{\boldsymbol{\Gamma}}_j'\mathbf{v}\right)^{\beta_j N} \right\}. \tag{72}$$

Because products of completely monotonic functions are also completely monotonic, it suffices to show that

$$\left(1 + \frac{1}{N} \tilde{\boldsymbol{\Gamma}}_j'\mathbf{v}\right)^{\beta_j N}$$

is a completely monotonic function of N^{-1}. Let us write

$$\left(1 + \frac{1}{N} \tilde{\boldsymbol{\Gamma}}_j'\mathbf{v}\right)^{\beta_j N} = e^{t(w)}, \tag{73}$$

where

$$t(w) = \frac{\beta_j}{w} \log(1 + aw), \tag{74}$$

by identifying $w = 1/N$ and $a = \tilde{\boldsymbol{\Gamma}}_j'\mathbf{v}$. We note that $0 \leq w < \infty$ and that a is nonnegative and bounded, with the latter property being ensured by the restriction of $\mathbf{v}$ to $Q^+(R)$, $R < \infty$. Thus, all derivatives of $\log(1 + aw)$, and consequently of $t(w)$, exist and are continuous for $0 \leq w < \infty$. Since

$$\frac{d^{k+1}\{e^{t(w)}\}}{dw^{k+1}} = \sum_{m=0}^{k} \binom{k}{m} \left\{ \frac{d^{k+1-m}}{dw^{k+1-m}} t(w) \right\} \left[\frac{d^m}{dw^m} \{e^{t(w)}\} \right], \tag{75}$$

we may conclude from a simple inductive argument that

$$\begin{array}{ll} \text{if} & t(w) \text{ is a completely monotonic function of } w, \\ \text{then } e^{t(w)} & \text{is a completely monotonic function of } w. \end{array} \tag{76}$$

Finally, to show that $t(w)$ is a completely monotonic function of w is to show that $\{\log(1 + x)\}/x$ is completely monotonic. This is true, but we omit the proof. □

We need the following analogous property in connection with the integrand of $I_\sigma^{(1)}(N)$:

$$\left. \begin{array}{l} \text{For all } \mathbf{v} \in Q^+(R), R < \infty, (\tilde{\boldsymbol{\Gamma}}_\sigma'\mathbf{v}) H(N^{-1}, \mathbf{v}) \text{ is completely} \\ \text{monotonic in } N^{-1}. \end{array} \right\} \tag{77}$$

The proof is immediate from the preceding proposition since the additional factor $(\tilde{\boldsymbol{\Gamma}}_\sigma'\mathbf{v})$ does not depend on N^{-1}.

5.2 Error bounds

Proposition 8: For all N, $0 < N \leq \infty$

$$\frac{A_m}{N^m} < I(N) - \sum_{k=0}^{m-1} \frac{A_k}{N^k} < 0, \qquad m = 1, 3, 5, \cdots$$

$$0 < I(N) - \sum_{k=0}^{m-1} \frac{A_k}{N^k} < \frac{A_m}{N^m}, \qquad m = 2, 4, 6, \cdots . \tag{78}$$

Proof: We initially require $\mathbf{v} \in Q^+(R)$, $R < \infty$ so that the preceding proposition is applicable. Viewing $H(N^{-1}, \mathbf{v})$ as a function of N^{-1}, we may use a version of Taylor's theorem[24] that is accompanied by an estimate of the truncation error for the series to obtain

$$H(N^{-1}, \mathbf{v}) = \sum_{k=0}^{m-1} \frac{h_k(\mathbf{v})}{N^k} + \frac{1}{N^m} \frac{1}{m!} \frac{\partial^m}{\partial(1/N)^m} H(\xi, \mathbf{v}), \tag{79}$$

where

$$\xi \in [0, N^{-1}].$$

Consider first the case of m odd. From the preceding proposition, the mth derivative of $H(N^{-1}, \mathbf{v})$ is a nonpositive and monotonically nondecreasing function of N^{-1}. Hence,

$$\frac{\partial^m}{\partial(1/N)^m} H(0, \mathbf{v}) \leq \frac{\partial^m}{\partial(1/N)^m} H(\xi, \mathbf{v}) \leq 0, \qquad \xi \geq 0. \tag{80}$$

Substituting in (79),

$$\frac{h_m(\mathbf{v})}{N^m} \leq H(N^{-1}, \mathbf{v}) - \sum_{k=0}^{m-1} \frac{h_k(\mathbf{v})}{N^k} \leq 0. \tag{81}$$

Hence,

$$\int_{Q^+(R)} e^{-\mathbf{1}'\mathbf{v}} H(N^{-1}, \mathbf{v}) d\mathbf{v} - \sum_{k=0}^{m-1} \frac{1}{N^k} \int_{Q^+(R)} e^{-\mathbf{1}'\mathbf{v}} h_k(\mathbf{v}) d\mathbf{v}$$

$$\leq 0$$

$$\geq \frac{1}{N^m} \int_{Q^+(R)} e^{-\mathbf{1}'\mathbf{v}} h_m(\mathbf{v}) d\mathbf{v}. \tag{82}$$

The pair of bounds holds uniformly in R. Consequently, we may let $R \to \infty$ and drop the restriction on R to obtain (78).

The proof for m even is very similar with the starting point being the following replacement for (80),

$$0 \leq \frac{\partial^m}{\partial(1/N)^m} H(\xi, \mathbf{v}) \leq \frac{\partial^m}{\partial(1/N)^m} H(0, \mathbf{v}). \tag{83}$$

□

The rest of the proof is omitted.

The above proposition states that the error incurred from using only a certain number of leading terms of the expansion for $I(N)$ is numerically less than the first neglected term of the series and has the same sign.

Another implication that can be quite useful in practice is that the estimate with an odd number of terms is an upper bound on the integral and an even number of terms gives a lower bound. Thus, the error sequence alternates in sign. (It is also true but less consequential that the terms of the expansion also alternate in sign.)

In particular the above proposition proves the asymptoticity of the expansion in Section 4.3.

By a matching argument and with recourse to the complete monotonicity of $(\tilde{\Gamma}'_\sigma \mathbf{v})H(N^{-1}, \mathbf{v})$, see (77), we also have

Proposition 9: For any class index σ and all N, $0 \le N \le \infty$,

$$\frac{A^{(1)}_{\sigma,m}}{N^m} \le I^{(1)}_\sigma(N) - \sum_{k=0}^{m-1} \frac{A^{(1)}_{\sigma,k}}{N^k} \le 0, \qquad m = 1, 3, 5, \cdots$$

$$0 \le I^{(1)}_\sigma(N) - \sum_{k=0}^{m-1} \frac{A^{(1)}_{\sigma,k}}{N^k} \le \frac{A^{(1)}_{\sigma,m}}{N^m}, \qquad m = 2, 4, 6, \cdots. \tag{84}$$

With error estimates available for both $I(N)$ and $I^{(1)}_\sigma(N)$, it is straightforward to use these to obtain an error estimate for the mean value given in Proposition 3, eq. (48).

APPENDIX A

The Coefficients $\{A^{(1)}_{\sigma k}\}$ in Terms of the Partition Function of the Pseudo-Network

Here σ is a given fixed class index, while j, k, and l are also class indices each ranging over $[1, p]$. It is also understood that j, k, l, and σ are all distinct.

$$A^{(1)}_{\sigma,0} = g(\mathbf{e}_\sigma)$$

$$A^{(1)}_{\sigma,1} = -3\beta_\sigma g(3\mathbf{e}_\sigma) - \sum_j \beta_j g(\mathbf{e}_\sigma + 2\mathbf{e}_j)$$

$$\begin{aligned} A^{(1)}_{\sigma,2} = {} & 8\beta_\sigma g(4\mathbf{e}_\sigma) + 15\beta_\sigma^2 g(5\mathbf{e}_\sigma) \\ & + \sum_j [2\beta_j g(3\mathbf{e}_j + \mathbf{e}_\sigma) + 3\beta_j^2 g(4\mathbf{e}_j + \mathbf{e}_\sigma) + 3\beta_\sigma\beta_j g(3\mathbf{e}_\sigma + 2\mathbf{e}_j)] \\ & + \frac{1}{2}\sum_{j,k} \beta_j\beta_k g(2\mathbf{e}_j + 2\mathbf{e}_k + \mathbf{e}_\sigma). \end{aligned}$$

$$\begin{aligned} A^{(1)}_{\sigma,3} = {} & -30\beta_\sigma g(5\mathbf{e}_\sigma) - 120\beta_\sigma^2 g(6\mathbf{e}_\sigma) - 105\beta_\sigma^3 g(7\mathbf{e}_\sigma) \\ & - \sum_j [6\beta_j g(4\mathbf{e}_j + \mathbf{e}_\sigma) + 20\beta_j^2 g(\mathbf{e}_\sigma + 5\mathbf{e}_j) + 6\beta_\sigma\beta_j g(3\mathbf{e}_\sigma + 3\mathbf{e}_j) \\ & \quad + 2\beta_\sigma\beta_j g(4\mathbf{e}_\sigma + 2\mathbf{e}_j) \end{aligned}$$

$$+ 15\beta_j^3 g(\mathbf{e}_\sigma + 6\mathbf{e}_j) + 15\beta_\sigma^2\beta_j g(5\mathbf{e}_\sigma + 2\mathbf{e}_j) + 9\beta_\sigma\beta_j^2 g(3\mathbf{e}_\sigma + 4\mathbf{e}_j)]$$

$$- \sum_{j,k} \left[2\beta_j\beta_k g(\mathbf{e}_\sigma + 2\mathbf{e}_j + 3\mathbf{e}_k) + 3\beta_j^2\beta_k g(\mathbf{e}_\sigma + 4\mathbf{e}_j + 2\mathbf{e}_k) \right.$$

$$\left. + \frac{3}{2}\beta_\sigma\beta_j\beta_k g(3\mathbf{e}_\sigma + 2\mathbf{e}_j + 2\mathbf{e}_k) \right]$$

$$- \frac{1}{6} \sum_{j,k,l} \beta_j\beta_k\beta_l g(2\mathbf{e}_j + 2\mathbf{e}_k + 2\mathbf{e}_l + \mathbf{e}_\sigma).$$

REFERENCES

1. F. P. Kelly, "Reversibility and Stochastic Networks," New York: John Wiley, 1980, Chapter 3.
2. L. Kleinrock, "Queueing Systems. Vol II: Computer Applications," New York: John Wiley, 1976, Chapter 4.
3. M. Schwartz, "Computer-Communication Network Design and Analysis," Englewood Cliffs, N.J.: Prentice-Hall, 1977, Chapter 11.
4. F. Baskett et al., "Open, Closed and Mixed Networks of Queues with Different Classes of Customers," J. of the ACM, *22*, No. 2 (April 1975), pp. 248–60.
5. F. R. Moore, "Computational Model of a Closed Queueing Network with Exponential Servers," IBM J. of Res. and Dev. (1972), pp. 567–72.
6. J. Buzen, "Queueing Network Models of Multiprogramming," Ph.D. Thesis, Div. of Eng. and App. Sci., Harvard University, Cambridge Mass., 1971.
7. R. R. Muntz and J. Wong, "Asymptotic Properties of Closed Queueing Network Models," Proc. 8th Annual Princeton Conf. on Info. Sci. and Systems (March 1974), pp. 348–52.
8. D. P. Gaver and G. S. Shedler, "Approximate Models for Processor Utilization in Multiprogrammed Computer Systems," SIAM J. on Computing, 2 (1973), pp. 183–92.
9. M. Reiser, "A Queueing Network Analysis of Computer Communication Networks with Flow Control," IEEE Trans. Comm., *COM-27* No. 8 (August 1979), pp. 1199–209.
10. S. S. Lam and M. Reiser, "Congestion Control of Store-and-Forward Networks by Input Buffer Limits-An Analysis," IEEE Trans. Comm., *COM-27*, No. 1 (January 1979), pp. 127–34.
11. B. Pittel, "Closed Exponential Networks of Queues with Saturation: The Jackson-Type Stationary Distribution and its Asymptotic Analysis," Math. of Oper. Res., *4*, No. 4 (November 1979) pp. 357–78.
12. E. Arthurs and B. W. Stuck, "A Theoretical Performance Analysis of a Markovian Switching Node," IEEE Trans. Comm., *COM-26*, No. 11 (Nov. 1978), pp. 1779–84.
13. J. P. Buzen, "Computational Algorithms for Closed Queueing Networks with Exponential Servers," Commun. of the ACM, *16* (1973), pp. 527–31.
14. M. Reiser and H. Kobayashi, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms," IBM J. of Res. and Dev., *19*, No. 3 (May 1975), pp. 283–94.
15. M. Reiser and S. S. Lavenberg, "Mean Value Analysis of Closed Multichain Queueing Networks," J. Assoc. Comput. Mach., *27* (April 1980).
16. C. H. Sauer and K. M. Chandy, unpublished work.
17. S. C. Bruell and G. Balbo, "Computational Algorithms for Closed Queueing Networks," New York: North-Holland, 1980.
18. S. S. Lavenberg, "Closed Multichain Product Form Queueing Networks with Large Population Sizes," IBM Research Report, RC 8496 (No. 36973), September 1980.
19. J. Zahorjan, "The Approximate Solution of Large Queueing Network Models," Tech. Report CSRG-122, University of Toronto, August 1980.
20. Y. Bard, "Some Extensions to Multiclass Queueing Network Analysis," 4th Int. Symp. Model. and Perf. Eval. of Comp. Systems, Vienna (February 1979).
21. J. McKenna, D. Mitra, and K. G. Ramakrishnan, "A Class of Closed Markovian Queueing Networks: Integral Representations, Asymptotic Expansions, Generalizations," B.S.T.J., *60*, No. 5 (May–June 1981), pp. 599–641.

22. A. Erdelyi, "Asymptotic Expansions," New York: Dover Publications, 1956.
23. E. T. Copson, "Asymptotic Expansions," Cambridge, U.K.: Cambridge Univ. Press, 1965, Chapters 5 and 6.
24. F. W. J. Olver, "Introduction to Asymptotics and Special Functions," New York: Academic Press, 1974, Chapter 3.
25. N. Bleistein and R. A. Handelsman, "*Asymptotic Expansions of Integrals*," N.Y.: Holt, Rinehart, and Winston, 1975, Chapter 8.
26. L. A. Skinner, "Note on the Asymptotic Behavior of Multidimensional Laplace Integrals," SIAM J. Math. Annal., *11*, No. 5 (September 1980), pp. 911–7.