Asymptotic Expansions for Closed Markovian Networks with State-Dependent Service Rates

DEBASIS MITRA AND J. MCKENNA

AT&T Bell Laboratories, Murray Hill, NJ

Abstract. A method is presented for calculating the partition function, and from it, performance measures, for closed Markovian stochastic networks with queuing centers in which the service or processing rate depends on the center's state or load. The analysis on which this method is based is new and a major extension of our earlier work on load-independent queuing networks. The method gives asymptotic expansions for the partition function in powers of $1/N$, where N is a parameter that reflects the size of the network. The expansions are particularly useful for large networks with many classes, each class having many customers. The end result is a decomposition by which expansion coefficients are obtained exactly by linear combinations of partition function values of small network constructs called pseudonetworks. Effectively computable bounds are given for errors arising from the use of a finite number of expansion terms. This method is important because load dependence is at once an essential element of sophisticated network models of computers, computer communications, and switching, teletraffic, and manufacturing systems, and the cause of very intensive computations in conventional techniques. With this method, very large load-dependent networks can be analyzed, whereas previously only small networks were computationally tractable.

Categories and Subject Descriptors: D.4.8 **[Operating System]**: Performance—*queuing theory*; G.m **[Miscellaneous]**: *queuing theory*

General Terms: Performance, Theory, Verification

Additional Key Words and Phrases: Asymptotics, integral representations, load dependence

1. *Introduction*

During the past two decades, Markovian queuing networks have emerged as one of the most important tools for modeling computer systems, computer communications, and switching, teletraffic, and manufacturing systems [12, 24]. This was in large measure because an important class of such networks that is analytically tractable, the so-called *local balance* [2], *quasi-reversible* [11], or, simply, *product-form networks*, was discovered. Tractability is preserved even in the presence of queuing or processing centers in which the service rate depends, with some restrictions, on the load or state of the center. Sophisticated models of practical systems quite often require such load-dependent processors [24]. For example, teletraffic models [7] frequently concern centers with multiple servers, perhaps the most common example of load-dependent queuing. Another major reason for interest in networks with load-dependent centers is that such networks, when

Preliminary results have appeared in *Proceedings of the 10th International Symposium on Computer Modelling, Measurement, and Evaluation* (Paris, France). North-Holland, New York, 1984.

Authors' address: AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974.

appropriately chosen [6], approximate the behavior of non-product-form networks. Unfortunately, the element of load dependence makes the analysis of such networks very computation intensive [3]. Therefore, their use has been confined to relatively small networks, since large networks have required either intractably large calculations or heuristics [1, 4, 8, 14, 23, 27], which, although useful, are handicapped by one or more of the following: errors of unknown magnitude, unknown range of applicability, questions of uniqueness, and problems with convergence [14].

In this paper we report on a new computational method, based on mathematical foundations, which should alleviate the situation. This method is particularly valuable when the network has many classes, each with a large population. The theory significantly extends the theory in [15]–[17] for *load-independent*, closed, product-form networks of a very general type. In [17] we reported the discovery of an integral representation for the partition function of such a network. This integral [15] contained a large parameter N, which in a natural way described the large size of the network. In [15] we were able to obtain an asymptotic expansion for this integral in inverse powers of N for multimode, multiclass networks in "normal usage." In [16] we introduced the notion of "moment partition functions" and showed that the most general moments of queue lengths can be given as ratios of these quantities. By first giving integral representations to the moment partition functions, we were then able to extend to them the complete treatment previously given to the partition function.

The theory given here departs in several fundamental respects from the earlier work. It is therefore quite remarkable that the end results, that is, the computational algorithms and error bounds, are conceptually similar to the corresponding results for load-independent networks. The common key to the calculations is the following decomposition: The terms of the asymptotic expansion may be calculated *exactly* by linearly combining many partition function values of small network constructs, called pseudonetworks. When the original network has a load-dependent center, then the pseudonetwork has a corresponding center that is load dependent. The load dependence in the construct is derived from the load dependence in the original network by means of an explicit formula. The error analysis gives explicit bounds for the error incurred from the use of only a finite number of terms of the asymptotic expansion. The form and efficacy of the bounds are similar regardless of the nature of the dependence on the load.

This paper assumes as before that the network is in "normal usage," that is, that none of the processors are utilized very heavily. However, the technical requirement associated with normal usage is considerably relaxed here. By allowing the pseudonetwork to be load dependent, even when the original network is load independent, it is *always* possible to satisfy this requirement. The combination of the relaxed normal usage requirement and load-dependent pseudonetwork has the effect of significantly extending the class of networks that is analyzable by the computational method given here. In particular, the method may now be applied to smaller networks where it requires fewer computations than the conventional recursions.

This paper only treats partition functions. The extension to moment partition functions and thus to the calculation of arbitrary moments of queue lengths follows along the lines of [16]. The extension to networks with open and closed classes is possible since it has been shown in an unpublished work [21] that performance measures of a mixed network follow directly from the analysis based on integral representations of an associated closed network construct. The detailed technical development in this paper is restricted to the case of a network with only one center which has general load dependence. However, the computational algorithm, in particular, the structure of the pseudonetwork, and the composition of the expansion coefficients from the pseudonetwork, for networks with more than one load-dependent center is clearly indicated in Section 7.

The computational methods derived from asymptotic expansions have been incorporated in a software package called PANACEA. An early version of PANACEA that handles only closed networks is described in [20]. It is noteworthy that PANACEA computes not only performance measures but also their lower and upper bounds.

1.1 Overview of Results

(1) We start with the convolutional representation

$$G_q(\mathbf{K}) = \sum_{0\le\mathbf{n}\le\mathbf{K}} G_{q-1}(\mathbf{K}-\mathbf{n})\pi_q(\mathbf{n}), \tag{1.1}$$

where G_q and G_{q-1} are, respectively, the partition functions for the network with and without the load-dependent node with index q.

(2) We show that $G_q(\mathbf{K})$ is simply and exactly related to an integral **I**. This representation is valid for small and large networks, with or without infinite server (IS) node and for all levels of usage.

(3) We introduce a parameter N, which in a natural way describes the size of the network. For large networks with IS centers, that is, networks with large populations and small values of ratios of mean service times to mean think times, the parameter N is large.

(4) It is shown that there exists an asymptotic expansion

$$I(N) \sim \sum_{n=0}^{\infty} \frac{A_n}{N^n}, \qquad \text{as} \quad N \to \infty, \tag{1.2}$$

where the expansion coefficients $\{A_n\}$ are exactly specified.

(5) Each expansion coefficient is shown to be exactly equivalent to a linear combination of many partition function values of small networks, which we call pseudonetworks. These constructs have the same topology as the original network, except that IS centers are not present. The pseudonetworks have at least as many load-dependent centers as the original network. These networks are small in the sense that, to calculate the expansion coefficients $A_0, A_1, \ldots, A_r$, we need to consider one pseudonetwork, which has a total network population of $2r$. In practice $r = 4$ has typically proved both necessary and adequate [20].

A simple and explicit formula is given for characterizing the load dependence of node q in the pseudonetwork in terms of the load dependence of node q in the original network.

(6) We prove the following, which establishes the asymptotic property of the expansion in (1.2) and also provides complete and effectively computable error bounds:

$$0 \le (-1)^r\left[I(N) - \sum_{n=0}^{r-1} \frac{A_n}{N^n}\right] \le (-1)^r \frac{A_r}{N^r}, \qquad r = 1, 2, \ldots. \tag{1.3}$$

That is, the error incurred from using only the leading r terms of the expansion in (1.2) alternates in sign and is bounded (in magnitude) by the $(r+1)$th term of the expansion.

(7) We assume that the processing centers in the network are in "normal usage." In this paper we are substantially relaxing this requirement over our earlier work

since a load-dependent center can always be considered in normal usage by truncating in a natural way a sequence that arises in the analysis. Truncation always gives load dependence in the corresponding center of the pseudonetwork, but also lower error bounds.

(8) In addition to the foregoing, we derive an "alternative" integral representation of $I(N)$ that yields the same pseudonetworks as the first representation does.

(9) Thus this paper contains two derivations that are somewhat independent, even though a deliberate effort is made to establish equivalences at various intermediate points. The aforementioned "alternative" representation is developed in Appendix A. Yet another, third, course based on distributions (the "delta function" and its derivatives) has great appeal; Appendix A provides some clues but a systematic development is not undertaken in this paper.

2. *Integral Representations for the Partition Function*

2.1 Product Form. We recapitulate some of the well-known results [2] 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 the missing index is j where $1 \leq j \leq p$. A total of s service centers are allowed. We find it natural to distinguish the centers of Types 1, 2, and 4, which have queuing from the remaining centers of Type 3, which do not. (The definition of Type 1 through 4 centers is given in [2].) Thus centers 1 through q will be queuing centers, while $(q + 1)$ through s will be Type 3 centers, which have also been called think nodes and infinite server (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 N_{ji} be the random variable denoting in steady state the number of jobs of class j at center i, and $\mathbf{N} = (N_{ji})$ the $p \times s$ matrix whose (j, i) element is the random variable N_{ji}. Let $\mathbf{n}$ denote a $p \times s$ matrix whose elements are nonnegative integers, and let $\mathbf{n}_i = (n_{1i}, n_{2i}, \ldots, n_{pi})'$ denote the ith column of $\mathbf{n}$. In this paper we consider closed networks where the population of jobs in class j is a constant, K_j. The state space of $\mathbf{N}$ is the set S of matrices $\mathbf{n}$, which have integer components, and satisfy the population constraints

$$S = \{\mathbf{n} \mid 0 \leq n_{ji}, \sum_{i=1}^{s} n_{ji} = K_j, \ 1 \leq j \leq p, \ 1 \leq i \leq s\}. \tag{2.1}$$

Then the well-known results on closed networks with product-form stationary distributions can be given in the following form:

$$\pi(\mathbf{n}) = \frac{1}{G} \prod_{i=1}^{s} \pi_i(\mathbf{n}_i). \tag{2.2}$$

As mentioned earlier, in this paper we initially allow only one queuing center, say the center indexed by q, to have general state-dependent service rates. Hence

$$\pi_q(\mathbf{n}_q) = \frac{n_q!}{\prod_{k=1}^{n_q} \mu_q(k)} \prod_{j=1}^{p} \frac{e_{jq}^{n_{jq}}}{n_{jq}!}, \tag{2.3}$$

where $n_q = \sum_j n_{jq}$, e_{jq} is the relative arrival rate of class j jobs to center q, and $\mu_q(k)$ is the service rate when there are k jobs queued. We assume the usual sufficient condition for product form, namely, the state-dependent service rates are independent of the class of the job being serviced.

For the other centers in the network we have

$$\pi_i(\mathbf{n}_i) = n_i! \prod_{j=1}^{p} \frac{\rho_{ji}^{n_{ij}}}{n_{ji}!}, \qquad 1 \le i \le q-1, \tag{2.4a}$$

$$= \prod_{j=1}^{p} \frac{\rho_{ji}^{n_{ji}}}{n_{ji}!}, \qquad q+1 \le i \le s, \tag{2.4b}$$

where $\rho_{ji} = e_{ji}/\mu_{ji}$.

For uniformity of notation, we write (2.3) in the form

$$\pi_q(\mathbf{n}_q) = f(n_q) \prod_{j=1}^{p} \frac{\rho_{jq}^{n_{jq}}}{n_{jq}!}, \tag{2.5a}$$

where $\{\rho_{jq}\}$ and $\{f(n)\}$ are defined as follows:

$$\rho_{jq} = \frac{e_{jq}}{\mu_q}, \qquad 1 \le j \le p,$$

$$f(n) = n! \frac{\mu_q^n}{\prod_{k=1}^{n} \mu_q(k)}, \qquad n \ge 1, \tag{2.5b}$$

and μ_q is any positive constant. In this case the expression for $\pi_q(\mathbf{n}_q)$ in (2.3) and (2.5) are equivalent.

2.2 Examples of Sequences Characterizing Load Dependence. We consider a few canonical examples of sequences $\{f(n)\}$, where the functional form of $f(n)$ is known for all $n \ge 1$. Note that in many other examples only the numerical values for a limited range of n may be available, such as when it is derived from measurements or simulations.

2.2.1 *Load Independence.* That is, in (2.3),

$$\mu_q(k) = \mu_q, \qquad k \ge 1. \tag{2.6}$$

We may take $\rho_{jq} = e_{jq}/\mu_q$ and $f(n) = n!$ to obtain the form in (2.5), and also, of course, the form in (2.4) for the load-independent queuing centers.

2.2.2 *Multiple Homogeneous Servers.* Let s denote the number of servers at center q, each capable of serving at rate μ_q. Then

$$\mu_q(k) = k\mu_q, \qquad 1 \le k \le s-1,$$

$$= s\mu_q, \qquad s \le k. \tag{2.7}$$

Again let $\rho_{jq} = e_{jq}/\mu_q$ and obtain

$$f(n) = 1, \qquad n \le s-1,$$

$$= \frac{n!}{s^n} \cdot \frac{s^s}{s!}, \qquad s \le n. \tag{2.8}$$

2.2.3 *Load Dependence Due to Heffes.* Heffes [10] has considered the following load dependence:

$$\mu_q(k) = \frac{\mu k}{a+k}, \qquad k \ge 1, \tag{2.9}$$

where μ and a are positive real constants. Let $\rho_{jq} = e_{jq}/\mu$, so that

$$f(n) = \frac{\Gamma(a+n+1)}{\Gamma(a+1)}, \qquad n \geq 1. \tag{2.10}$$

Notice in (2.9) that $\mu_q(k)$ is linear with k for small k and asymptotically, with large k, approaches a constant, a characterization somewhat similar to that of the multiple server.

2.2.4 *Norton Equivalent of Identical FCFS Centers.* It is known that a subnetwork of a quasi-reversible [11] or local balance [2] network may be exactly lumped into one load-dependent center, and this property is sometimes referred to as Norton's theorem [5] in analogy to a similar result in electrical circuit theory. We now consider the qth center to be the Norton equivalent of $(r+1)$ load-independent first-come, first-served (FCFS) centers $(r \geq 0)$ with common service rate μ.

It is quite easy to show that with $\rho_{jq} = e_{jq}/\mu$,

$$f(n) = \frac{(r+n)!}{r!} = \frac{\Gamma(r+n+1)}{\Gamma(r+1)}, \qquad n \geq 1. \tag{2.11}$$

The forms in (2.10) and (2.11) are identical for purposes of mathematical analysis.

2.2.5 *Norton Equivalent of FCFS Centers with Distinct Service Rates.* We consider the qth center of our network to be the Norton equivalent of r load-independent, FCFS centers with *distinct* service rates $\mu_1, \mu_2, \ldots, \mu_r$, (When the service rates are not all distinct, a slightly modified analysis applies.) It may be shown that with $\rho_{jq} = e_{jq}$,

$$f(n) = n! \sum_{m_1+\cdots+m_r=n} \frac{1}{\mu_1^{m_1} \cdots \mu_r^{m_r}}, \qquad n \geq 1, \tag{2.12}$$

$$= n! \sum_{k=1}^{r} \frac{a_k}{\mu_k^n}, \qquad n \geq 1, \tag{2.13}$$

where, $a_k = 1/\prod_{l \neq k} (1 - \mu_k/\mu_l)$, $1 \leq k \leq r$. Equation (2.13) is obtained from (2.12) by inverting a partial fraction representation of $F(z) = \sum (-z)^n f(n)/n!$.

2.3 Integral Representation for the Partition Function

In (2.2) G is the partition function, and it is explicitly

$$G_q(\mathbf{K}) = \sum_{\mathbf{n} \in S} \prod_{i=1}^{s} \pi_i(\mathbf{n}_i). \tag{2.14}$$

We have added the subscript q to specify the number of queuing centers, since we also need to consider $G_{q-1}(\mathbf{K})$, which is defined in an analogous manner but without the qth center. Thus, $G_{q-1}(\mathbf{K})$ is the partition function of a network without load-dependent centers. It is known [3] that the following convolutional representation relates the two partition functions:

$$G_q(\mathbf{K}) = \sum_{\mathbf{0} \leq \mathbf{n} \leq \mathbf{K}} G_{q-1}(\mathbf{K} - \mathbf{n}) \pi_q(\mathbf{n}), \tag{2.15}$$

and also [16] that

$$G_{q-1}(\mathbf{K}) = \left[\frac{1}{\prod_{j=1}^{p} K_j!}\right] \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} D_{q-1}(\mathbf{K}, \mathbf{u})\, d\mathbf{u}, \tag{2.16}$$

where,

$$\mathbf{u} = (u_1, u_2, \ldots, u_{q-1})',$$

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

$$\mathbf{K} = (K_1, K_2, \ldots, K_p)',$$

$$Q_{q-1}^{+} = \{\mathbf{u} \mid u_i \geq 0,\ 1 \leq i \leq q-1\},$$

$$D_{q-1}(\mathbf{K}, \mathbf{u}) = \prod_{j=1}^{p} (\rho_{j0} + \boldsymbol{\rho}_j'\mathbf{u})^{K_j},$$

$$\boldsymbol{\rho}_j = (\rho_{j1}, \rho_{j2}, \ldots, \rho_{j,q-1})',$$

$$\rho_{j0} = \sum_{i=q+1}^{s} \rho_{ji}.$$

Note that $\boldsymbol{\rho}_j$ and $\mathbf{u}$ are $(q-1)$-tuples, while $\mathbf{K}$ is a p-tuple. Observe too that, not surprisingly, the parameters ρ_{ji} for all the Type 3 centers appear lumped together in ρ_{j0}.

We may now combine (2.5), (2.15), and (2.16) to obtain

$$G_q(\mathbf{K}) = \sum_{m=0}^{K} f(m) \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{n_1+\cdots+n_p=m} \prod_j \frac{(\rho_{j0} + \boldsymbol{\rho}_j'\mathbf{u})^{K_j-n_j}}{(K_j - n_j)!} \frac{\rho_{jq}^{n_j}}{n_j!}\, d\mathbf{u}, \tag{2.17}$$

where $K = \sum K_j$. But an application of Leibniz's rule gives

$$\sum_{n_1+\cdots+n_p=m} \prod_j \frac{(\rho_{j0} + \boldsymbol{\rho}_j'\mathbf{u})^{K_j-n_j}}{(K_j - n_j)!} \frac{\rho_{jq}^{n_j}}{n_j!} = \frac{1}{m!} \frac{\partial^m}{\partial t^m} \prod_j \frac{(\rho_{j0} + \boldsymbol{\rho}_j'\mathbf{u} + \rho_{jq}t)^{K_j}}{K_j!}\Bigg|_{t=0},$$

and thus we arrive at the following:

PROPOSITION 1

$$G_q(\mathbf{K}) = \left[\frac{1}{\prod_j K_j!}\right] \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{m=0}^{K} \frac{f(m)}{m!} \frac{\partial^m D_q}{\partial t^m}\Bigg|_{t=0} d\mathbf{u}, \tag{2.18}$$

where,

$$D_q(\mathbf{K}, \mathbf{u}, t) = \prod_{j=1}^{p} (\rho_{j0} + \boldsymbol{\rho}_j'\mathbf{u} + \rho_{jq}t)^{K_j}. \tag{2.19}$$

Note that the representation holds regardless of whether IS, that is, Type 3, centers are visited ($\rho_{j0} \neq 0$) or not ($\rho_{j0} = 0$) by class j jobs.

In the rest of the paper we follow convention and write

$$\frac{\partial^m}{\partial t^m} D_q(\mathbf{K}, \mathbf{u}, 0) = \frac{\partial^m}{\partial t^m} D_q(\mathbf{K}, \mathbf{u}, t)\Bigg|_{t=0}.$$

Note that in (2.18) we may correctly extend the range of m to $0 \leq m < \infty$. This is because D_q is a polynomial in t of degree K, and consequently the mth derivative with respect to t is zero for $m > K$.

3. *Large Networks in Normal Usage*

We henceforth consider only networks in which the route for each class always contains an IS, that is, Type 3, center. Specifically,

$$\rho_{j0} > 0, \qquad j = 1, 2, \ldots, p. \tag{3.1}$$

3.1 Large Network Normalizations. For large networks, certainly, we expect the class populations K_j to be large. But, at the same time, we expect the ratio of normalized processing time to think time, ρ_{ji}/ρ_{j0}, to be smaller for increased class populations. Instead of working with large and small parameters simultaneously, let us introduce a large parameter N to reflect the size of the network and normalize network parameters to be about 1 in order-of-magnitude estimation.

Although there is great latitude in the selection of the large parameter, we have found the following to work well in practice:

$$N = \max_{i,j} \left\{ \frac{\rho_{j0}}{\rho_{ji}} \right\},$$

where the maximum is on the set of nonzero ρ_{ji}. Now define for $1 \le j \le p$

$$\beta_j \triangleq \frac{K_j}{N}, \qquad \boldsymbol{\Gamma}_j \triangleq \frac{N}{\rho_{j0}}\, \boldsymbol{\rho}_j, \qquad \text{and} \qquad \Gamma_{jq} \triangleq \frac{N}{\rho_{j0}}\, \rho_{jq}. \tag{3.2}$$

We expect all these newly defined quantities to be $O(1)$. We have also defined $K = \sum K_j$ and this quantity is given by βN, where $\beta = \sum \beta_j$.

Note that the function D_q in Proposition 1 may be expressed in terms of the newly introduced normalized parameters as

$$D_q(\mathbf{K}, \mathbf{u}, t) = \left[\prod_j \rho_{j0}^{K_j} \right] \prod_j \left\{ 1 + \frac{1}{N} (\boldsymbol{\Gamma}_j' \mathbf{u} + \Gamma_{jq} t) \right\}^{\beta_j N}. \tag{3.3}$$

3.2 Normal Usage and Its Consequences. Define

$$\left. \begin{aligned} \alpha_i &\triangleq 1 - \sum_{j=1}^{p} \frac{K_j \rho_{ji}}{\rho_{j0}}, \\ &= 1 - \sum_{j=1}^{p} \beta_j \Gamma_{ji}, \end{aligned} \right\} 1 \le i \le q. \tag{3.4}$$

We have previously [15, 16] defined the load-independent queuing center i to be in normal usage if the following condition is satisfied:

$$\alpha_i > 0.$$

We now give the condition for the load-dependent queuing center q to be in normal usage. From (2.5b) and (3.4),

$$\alpha_q = 1 - \frac{1}{\mu_q} \sum_{j=1}^{p} K_j \frac{e_{jq}}{\rho_{j0}} = 1 - \frac{\lambda_q}{\mu_q}, \tag{3.5a}$$

where

$$\lambda_q \triangleq \sum_{j=1}^{p} K_j \frac{e_{jq}}{\rho_{j0}}. \tag{3.5b}$$

However, since μ_q can be any positive number, it can always be chosen so that $\alpha_q > 0$. Instead, we say the load-dependent node q is in normal usage if

$$\tilde{\alpha}_q \triangleq 1 - \lambda_q c_q > 0, \tag{3.6a}$$

where

$$c_q \triangleq \overline{\lim_{n \to \infty}} \left\{ \frac{1}{\mu_q(n)} \right\}. \tag{3.6b}$$

(We always assume $c_q < \infty$, which includes all realistic cases we know of.) Thus we obtain $\bar{\alpha}_q$ from α_q by replacing $1/\mu_q$ in (3.5a) by $\overline{\lim}_{n\to\infty} \{1/\mu_q(n)\}$.

Although the condition $\bar{\alpha}_q > 0$ seems, at least superficially, to be a reasonable extension of the condition $\alpha_q > 0$ for load-independent queuing centers, it will be seen subsequently that it is a very natural condition with deep consequences. In the first place, (3.6a) can be written as $0 < \lambda_q < 1/c_q$, so it is always possible to pick a positive number μ_q such that $\lambda_q < \mu_q < 1/c_q$. In other words, if the load-dependent node q is in normal usage, it is possible to choose a positive number μ_q so that simultaneously $\alpha_q > 0$, where α_q is defined in (3.5), and $\mu_q c_q < 1$.

We now define a transformation of the sequence $\{f(n)\}$ into the sequence $\{\phi(n)\}$ by

$$\phi(n) \triangleq \sum_{m=0}^{\infty} f(n+m)\frac{(1-\alpha_q)^m}{m!}, \qquad n = 0, 1, 2, \ldots. \tag{3.7}$$

This new sequence, besides being of independent interest, is also basic to the computational procedure given in this paper since it will be shown later that it characterizes the load dependence of the corresponding center in the pseudonetwork.

If use is made of (2.5b), (3.4), (3.5), it can be seen that the series on the right-hand side of (3.7) can be written in the following three equivalent ways

$$\mu_q^n \sum_{m=0}^{\infty} \frac{(n+m)!}{m!} \frac{\lambda_q^m}{\prod_{k=1}^{n+m} \mu_q(k)}, \tag{3.8a}$$

$$\sum_{m_1=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} f(\sum m_j + n) \prod_j \frac{(\beta_j \Gamma_{jq})^{m_j}}{m_j!}, \tag{3.8b}$$

$$\sum_{m_1=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} f(\sum m_j + n) \prod_j \frac{(K_j \rho_{jq}/\rho_{j0})^{m_j}}{m_j!}. \tag{3.8c}$$

We now see from (3.8a) that the convergence of the series defining $\{\phi(n)\}$ is independent of the choice of μ_q, and that the condition of normal usage, (3.6), is sufficient to ensure the convergence of these series.

Examination of (3.8c) shows that $\phi(0)$ admits of a simple physical interpretation; it is the partition function of the queuing center in isolation. The convergence of the series for $\phi(0)$ is therefore equivalent to the stochastic stability of a hypothetical open network consisting only of the qth node subject to Poisson loads. The rate of this offered traffic for class j equals the rate of jobs departing another hypothetical IS node with a constant population K_j and mean service time ρ_{j0}. Similarly $\phi(n)$ for $n > 0$ may be interpreted as the partition function of this isolated hypothetical node conditioned on there being n jobs resident in the node.

It may also be seen from (3.6) that if center q is load independent with mean service rate μ_q, then $c_q = 1/\mu_q$, and hence from (3.6), $\bar{\alpha}_q$ is equal to α_q as defined in (3.4). Thus the general definition of normal usage given above reduces to the definition of normal usage previously given for load-independent queuing centers.

We assume that all the queuing centers in the network are in normal usage.

In many practical situations the value of $f(n)$ for $1 \le n \le K$, the total network population, will be known but not its functional form. In these cases, as Proposition 1 shows, we may take $f(n) = 0$, $n > K$. However, $f(n) = 0$, $n > K$, is equivalent to setting $\mu_q(k) = \infty$, $k > K$; this implies from (3.6) that $c_q = 0$ and hence $\bar{\alpha}_q = 1 > 0$ from (3.6). Then the *center q is in normal usage* and the series defining $\{\phi(n)\}$ obviously converge. In fact $\phi(n) = 0$, $n > K$.

In certain other situations in which the functional form of $\{f(n)\}$ is known for all n, such as in the load-independent and multiple homogeneous server cases, there may be practical advantages in following the other, also correct, course of not setting $f(n) = 0$, $n > K$. In the load-independent case, the main advantage is that the pseudonetworks are themselves load independent if $f(n) = n!$, $n \geq 1$; they are load dependent if $f(n) = n!$, $n \leq K$; $= 0$, $n > K$. The pseudonetwork calculations are obviously simpler if they are load independent. The errors bounds, however (see Section 6.4), will always be lower if the course involving truncation is followed.

To summarize, "normal usage" is guaranteed in the case most likely to occur out of necessity and choice, namely, the series $\{f(n)\}$ is truncated at $n = K < \infty$. Truncation always gives lower error bounds. The technical content of "normal usage" is almost entirely concerned with the case in which the series $\{f(n)\}$, $0 \leq n \leq \infty$, is used, when available, without truncation. The latter course was followed in our earlier work [15, 16]. The procedure, based on appropriate truncations, is new to this paper. Although the effect of relaxation is small for very large networks, the main gain is in connection with small- and medium-sized networks, where it makes the asymptotic method of calculations also attractive in comparison to alternative methods.

3.3 Examples of $\{\phi(n)\}$. The sequence $\{\phi(n)\}$ may always be numerically evaluated. However, for the canonical cases of load dependence stated in Section 2.2 we have obtained closed-form expressions. In all these cases, the functional form of $f(n)$ is known for all n and it is used to compute $\phi(n)$ without truncation.

3.3.1 *Load Independence.* Normal usage corresponds to $\alpha > 0$.

$$\phi(n) = \frac{n!}{\alpha^{n+1}}, \qquad n \geq 0. \tag{3.9}$$

This form is the reason that a load-independent center remains load independent in the pseudonetwork; see Section 5.

3.3.2 *Multiple(s) Homogeneous Servers.* Normal usage corresponds to $s - 1 + \alpha > 0$. Also,

$$\begin{aligned}\phi(n) &= \frac{s^s}{(s-1)!}\frac{n!}{(s-1+\alpha)^{n+1}} \\ &\quad - \sum_{m=0}^{s-2-n} \frac{(1-\alpha)^m}{m!}\left[\frac{(m+n)!s^s}{s!s^{m+n}} - 1\right], \qquad n \leq s-2, \\ &= \frac{s^s}{(s-1)!}\frac{n!}{(s-1+\alpha)^{n+1}}, \qquad n \geq s-1.\end{aligned} \tag{3.10}$$

3.3.3 *Load Dependence Due to Heffes.* For the sequence $\{f(n)\}$ given in (2.10), normal usage corresponds to $\alpha > 0$. Also,

$$\phi(n) = \frac{1}{\alpha^{n+\alpha+1}}\frac{\Gamma(n+a+1)}{\Gamma(a+1)}, \qquad n \geq 0. \tag{3.11}$$

3.3.4 *Norton Equivalent of $(r + 1)$ Identical FCFS Centers* (see (2.11)). Normal usage corresponds to $\alpha > 0$. Also,

$$\phi(n) = \frac{1}{\alpha^{n+r+1}}\frac{\Gamma(n+r+1)}{\Gamma(r+1)}, \qquad n \geq 0. \tag{3.12}$$

3.3.5 *Norton Equivalent of r FCFS Centers with Distinct Service Rates* (see (2.12)). Normal usage corresponds to

$$\left(\min_{1\le k\le r} \mu_k\right) - 1 + \alpha > 0.$$

Also,

$$\phi(n) = \sum_{k=1}^{r} a_k\mu_k \frac{n!}{(\mu_k - 1 + \alpha)^{n+1}}, \qquad n \ge 0. \tag{3.13}$$

3.4 TRANSFORMATION TO INTEGRAL REPRESENTING PARTITION FUNCTION. We here exploit normal usage in the load-independent centers, that is,

$$\alpha_i > 0, \qquad 1 \le i \le q - 1, \tag{3.14}$$

to make a simplifying transformation to the integral representation in Proposition 1. Make the change of variables

$$\alpha_i u_i \to u_i, \qquad 1 \le i \le q - 1, \tag{3.15}$$

and observe that on account of (3.14) the region of integration is unchanged. The parameters for center i, $1 \le i \le q - 1$, are renormalized with respect to α; thus

$$\tilde{\Gamma}_{ji} \triangleq \frac{\Gamma_{ji}}{\alpha_i}, \qquad 1 \le j \le p, \quad 1 \le i \le q - 1, \tag{3.16}$$

so that, in particular,

$$\mathbf{\Gamma}_j' \, \mathbf{u} \to \tilde{\mathbf{\Gamma}}_j' \mathbf{u}. \tag{3.17}$$

With these changes we have the following form for the partition function:

PROPOSITION 2

$$G_q(\mathbf{K}) = \left[\frac{\prod_{j=1}^{p} \rho_{j0}^{k_j}/K_j!}{\prod_{i=1}^{q-1} \alpha_i}\right] I(N), \tag{3.18}$$

where

$$I(N) = \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{m=0}^{\beta N} \frac{f(m)}{m!} \frac{\partial^m \theta}{\partial t^m}\bigg|_{t=0} d\mathbf{u}, \tag{3.19}$$

and

$$\theta(N^{-1}, \mathbf{u}, t) = \prod_{j=1}^{p} [\exp{(-\,\beta_j \tilde{\mathbf{\Gamma}}_j' \mathbf{u})}] \left\{1 + \frac{1}{N} (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)\right\}^{\beta_j N}. \tag{3.20}$$

The term in square brackets in (3.18) is never computed in practice since performance measures are given by ratios of partition function values. For this reason, from now on we are only concerned with the integral $I(N)$.

At times, we find it useful to view $\theta(N^{-1}, \mathbf{u}, t)$ as a function of a real variable N. At such times it is useful to keep in mind that the function is of interest at the discrete points where $\beta_j N$, $1 \le j \le p$, are integers. Note, in particular, with reference to (3.19) that

$$m > \beta N \text{ and } \beta_j N,\ 1 \le j \le p, \text{ are integers} \Rightarrow \frac{\partial^m \theta}{\partial t^m} = 0.$$

For this reason we may (i) take $f(m)$ to have *any* value we want in the range $m > K = \beta N$, and (ii) write (3.19) also as

$$I(N) = \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^m \theta}{\partial t^m}\bigg|_{t=0} d\mathbf{u}. \tag{3.21}$$

We take note of two items here. First, the sequence in which $f(m) = 0$ for $m > K$ is a special case of the infinite sequence $\{f(m)\}_{m=0}^{\infty}$ assumed in (3.21). Second, regardless of the manner in which the infinite sequence is defined, only the values of $f(m)$, $m \le K$, determine $I(N)$.

The asymptotic expansion in $1/N$ for $I(N)$ will however depend on the infinite sequence $\{f(m)\}$. In fact, as we see in Section 6.4, the special infinite sequence in which $f(m) = 0$, $m > K$, gives the smallest error bounds in our computational procedure.

4. *Asymptotic Expansions*

4.1 Expansions and Error Bounds. We return to the representation of the partition function in Proposition 2 and (3.21). We show in Proposition 8, Section 6, that, for all values of N such that $\beta_j N$, $1 \le j \le p$, are positive integers,

$$0 \le (-1)^r \left[\frac{\partial^m \theta}{\partial t^m}\bigg|_{t=0} - \sum_{n=0}^{r-1} \frac{1}{N^n} \left\{ \frac{1}{n!} \frac{\partial^n}{\partial (N^{-1})^n} \frac{\partial^m}{\partial t^m} \theta \bigg|_{\substack{1/N=0\\ t=0}} \right\} \right]$$

$$\le (-1)^r \frac{1}{N^r} \left\{ \frac{1}{r!} \frac{\partial^r}{\partial (N^{-1})^r} \frac{\partial^m}{\partial t^m} \theta \bigg|_{\substack{1/N=0\\ t=0}} \right\}, \tag{4.1}$$

for $r = 1, 2, \ldots$.

Now multiply (4.1) by $f(m)/m!$ and sum with respect to m for $0 \le m \le \infty$:

$$0 \le (-1)^r \left[\sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^m \theta}{\partial t^m}\bigg|_{t=0} - \sum_{n=0}^{r-1} \frac{1}{N^n} \left\{ \frac{1}{n!} \sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^n}{\partial (N^{-1})^n} \frac{\partial^m}{\partial t^m} \theta \bigg|_{\substack{1/N=0\\ t=0}} \right\} \right]$$

$$\le (-1)^r \frac{1}{N^r} \left\{ \frac{1}{r!} \sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^r}{\partial (N^{-1})^r} \frac{\partial^m}{\partial t^m} \theta \bigg|_{\substack{1/N=0\\ t=0}} \right\}, \tag{4.2}$$

for $r = 1, 2, \ldots$. Multiply (4.2) by $e^{-\mathbf{1}'\mathbf{u}}$ and integrate with respect to $\mathbf{u} \in Q_{q-1}^{+}$:

$$0 \le (-1)^r \left[I(N) - \sum_{n=0}^{r-1} \frac{A_n}{N^n} \right] \le (-1)^r \frac{A_r}{N^r}, \qquad r = 1, 2, \ldots, \tag{4.3a}$$

where

$$A_n \triangleq \frac{1}{n!} \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^n}{\partial (N^{-1})^n} \frac{\partial^m}{\partial t^m} \theta \bigg|_{\substack{1/N=0\\ t=0}} d\mathbf{u}, \qquad n = 0, 1, \ldots. \tag{4.3b}$$

In particular, therefore, we have for all values of N such that $\beta_j N$, $1 \le j \le p$, are positive integers,

$$I(N) \sim \sum_{n=0}^{\infty} \frac{A_n}{N^n}, \qquad \text{as} \quad N \to \infty, \tag{4.4}$$

as well as bounds on the error incurred by truncating the expansion.

In Section 4.2 we describe a simple transformation that reduces the infinite sum in (4.3b) to a finite sum. Section 5 uses the latter form to obtain an exact procedure for computing each of the expansion coefficients. This procedure uses the pseudonetwork and its partition function for the exact computation.

4.2 A SIMPLIFYING TRANSFORMATION FOR THE EXPANSION COEFFICIENTS. We show here that there is a function $H(N^{-1}, \mathbf{u}, t)$ simply related to $\theta(N^{-1}, \mathbf{u}, t)$, possessing an attractive algebraic structure and for which

$$\sum_{m=0}^{\infty} \frac{f(m)}{m!} \frac{\partial^n}{\partial(N^{-1})^n} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0)$$

$$= \sum_{m=0}^{2n} \frac{\phi(m)}{m!} \frac{\partial^m}{\partial t^m} \frac{\partial^n}{\partial(N^{-1})^n} H(0, \mathbf{u}, 0), \qquad n = 0, 1, \ldots, \tag{4.5}$$

where the sequence $\{\phi(m)\}$ is as defined in Section 3.3.

The function H is defined thus:

$$\theta(N^{-1}, \mathbf{u}, t) \triangleq \exp[(1 - \alpha_q)t]H(N^{-1}, \mathbf{u}, t); \tag{4.6}$$

that is, from (3.4) and (3.20),

$$H(N^{-1}, \mathbf{u}, t) = \prod_j \exp[-\beta_j(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq}t)]\left\{1 + \frac{1}{N}(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq}t)\right\}^{\beta_j N}. \tag{4.7}$$

The function H has been investigated before [15] and we give without proof a statement of the property that we will find useful.

PROPOSITION 3. *Let*

$$h_n(\mathbf{u}, t) \triangleq \frac{1}{n!} \frac{\partial^n}{\partial(N^{-1})^n} H(0, \mathbf{u}, t), \qquad n = 0, 1, 2, \ldots, \tag{4.8}$$

where $h_n(\mathbf{u}, t)$ is a multinomial of degree $2n$ in the p variables $(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq}t)$, $1 \le j \le p$, and a polynomial in t of degree $2n$.

To see (4.5), observe that from (4.6), for $n = 0, 1, \ldots$ and $m = 0, 1, 2, \ldots$,

$$\frac{1}{n!}\frac{1}{m!}\frac{\partial^n}{\partial(N^{-1})^n}\frac{\partial^m}{\partial t^m}\theta(0, \mathbf{u}, 0) = \frac{1}{n!}\sum_{k=0}^{m}\left\{\frac{(1-\alpha_q)^{m-k}}{(m-k)!}\right\}\left\{\frac{1}{k!}\frac{\partial^k}{\partial t^k}\frac{\partial^n}{\partial(N^{-1})^n}H(0, \mathbf{u}, 0)\right\}$$

$$= \sum_{k=0}^{2n}\left\{\frac{(1-\alpha_q)^{m-k}}{(m-k)!}\right\}\left\{\frac{1}{k!}\frac{\partial^k}{\partial t^k}h_n(\mathbf{u}, 0)\right\}, \tag{4.9}$$

since from Proposition 3, $\partial^k h_n/\partial t^k = 0$ for $k > 2n$.

Multiply (4.9) by $f(m)$ and sum with respect to m to obtain, after recalling the definition of $\{\phi(n)\}$ in (3.7) and exchanging the order of summations, for $n = 0, 1, \ldots$,

$$\frac{1}{n!}\sum_{m=0}^{\infty}\frac{f(m)}{m!}\frac{\partial^n}{\partial(N^{-1})^n}\frac{\partial^m}{\partial t^m}\theta(0, \mathbf{u}, 0) = \sum_{m=0}^{2n}\frac{\phi(m)}{m!}\frac{\partial^m}{\partial t^m}h_n(\mathbf{u}, 0), \tag{4.10}$$

which is equivalent to (4.5).

We take stock of the progress made toward computation of the expansion coefficients in the following, which combines (4.3), (4.4), and (4.10) and amply demonstrates the utility of the transformation just described.

PROPOSITION 4

$$I(N) \sim \sum_{n=0}^{\infty} \frac{A_n}{N^n}, \qquad \text{as} \quad N \to \infty, \tag{4.11}$$

where

$$A_n = \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \left[\sum_{m=0}^{2n} \frac{\phi(m)}{m!} \frac{\partial^m}{\partial t^m} h_n(\mathbf{u}, t) \bigg|_{t=0} \right] d\mathbf{u}, \qquad n = 0, 1, \ldots . \tag{4.12}$$

References [15] and [16] have given systematic, recursive procedures for generating the multinomials $h_n(\mathbf{u}, t)$. We reproduce here the leading elements:

$$\begin{aligned} h_n(\mathbf{u}, t) &= 1, && n = 0, \\ &= -\frac{1}{2} \sum_j \beta_j (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)^2, && n = 1, \\ &= \frac{1}{3} \sum_j \beta_j (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)^3 \\ &\quad + \frac{1}{8} \sum_j \beta_j^2 (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)^4 \\ &\quad + \frac{1}{8} \sum_{j_1 \neq j_2} \beta_{j_1} \beta_{j_2} (\tilde{\mathbf{\Gamma}}_{j_1}' \mathbf{u} + \Gamma_{j_1 q} t)^2 (\tilde{\mathbf{\Gamma}}_{j_2}' \mathbf{u} + \Gamma_{j_2 q} t)^2, && n = 2. \end{aligned} \tag{4.13}$$

We note that in our procedure for computing the sequence $\{A_n\}$, the explicit differentiation with respect to t indicated in (4.12) is not explicitly undertaken. Instead, by noting that such differentiations are inherent to the definition of partition functions of load-dependent networks (see Proposition 1), we represent A_n as a linear combination of partition functions of constructs called pseudonetworks. This is undertaken in Section 5.

5. *Computation of Expansion Coefficients*

5.1 PSEUDONETWORKS. For any $\mathbf{k} = (k_1, k_2, \ldots, k_p)$, *a* p-tuple of nonnegative integers, define

$$g(\mathbf{k}) = \sum_{n_{11}+\cdots+n_{1q}=k_1} \cdots \sum_{n_{p1}+\cdots+n_{pq}=k_p} \left[\prod_{i=1}^{q-1} n_i! \left\{ \prod_{j=1}^{p} \frac{\tilde{\Gamma}_{ji}^{n_{ji}}}{n_{ij}!} \right\} \right] \left[\phi(n_q) \prod_{j=1}^{p} \frac{\Gamma_{jq}^{n_{jq}}}{n_{jq}!} \right], \tag{5.1}$$

where, by convention $n_i = \sum_{j=1}^{p} n_{ji}$, $1 \leq i \leq q$.

A comparison with (2.14) in Section 2.3 indicates the $g(\mathbf{k})$ is a partition function of a certain closed network. This network lacks IS centers but contains, as does the original network, $(q - 1)$ load-independent queuing centers and one load-dependent queuing center. The load-independent centers have processing rates $\{\tilde{\Gamma}_{ji}\}$, where, see (3.16), $\tilde{\Gamma}_{ji} = \Gamma_{ji}/\alpha_i$. The load-dependence in the qth center is determined by the sequence $\{\phi(n)\}$. The number of classes in this network is p, as in the original network. The p-tuple $\mathbf{k}$ denotes the population distribution in the network.

We refer to the network underlying (5.1) as a pseudonetwork.

If all queuing centers in the original network are load independent, then from (3.7),

$$\phi(n) = \frac{n!}{\alpha_q^{n+1}}, \qquad n \geq 0,$$

and

$$g(\mathbf{k}) = \frac{1}{\alpha_q} \sum_{n_{11}+\cdots+n_{1q}=k_1} \cdots \sum_{n_{p1}+\cdots+n_{pq}=k_p} \prod_{i=1}^{p} n_i! \left\{ \prod_{j=1}^{q} \frac{\tilde{\Gamma}_{ji}^{n_{ji}}}{n_{ji}!} \right\}, \tag{5.2}$$

where

$$\tilde{\Gamma}_{jq} = \frac{\Gamma_{jq}}{\alpha_q}, \qquad 1 \leq j \leq p.$$

The form in (5.2) without the multiplicative constant $1/\alpha_q$ in the right-hand side has been used in [15] and [16] to define pseudonetworks.

From the result in Proposition 1 it also follows that

$$g(\mathbf{k}) = \left[\frac{1}{\prod_{j=1}^{p} k_j!}\right] \int_{Q_{q-1}^{+}} e^{-\mathbf{1}'\mathbf{u}} \sum_{m=0}^{k} \frac{\phi(m)}{m!} \frac{\partial^m}{\partial t^m} \prod_{j=1}^{p} (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)^{k_j} \bigg|_{t=0} d\mathbf{u}, \tag{5.3}$$

where

$$k = \sum_{j=1}^{p} k_j.$$

5.2 EXPANSION COEFFICIENTS IN TERMS OF PSEUDONETWORK'S PARTITION FUNCTION. Compare (5.3) with the expression for A_n in (4.12) after recalling Proposition 3. This Proposition states that $h_n(\mathbf{u}, t)$ is a multinomial in $(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)$, $1 \leq j \leq p$, of degree $2n$ and hence it may be expressed as a sum of terms, each term being within a multiplicative constant of $\prod_j (\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)^{k_j}$ for some $\{k_j\}$, $\sum k_j \leq 2n$. Therefore, the expansion coefficient A_n may be expressed as a linear combination of $g(\mathbf{k})$, $\sum k_j \leq 2n$.

The following expressions for the leading expansion coefficients illustrate the procedure. We have used the expressions in (4.13).

$$\begin{aligned} A_0 &= g(\mathbf{0}), \\ A_1 &= -\sum_j \beta_j \, g(2\mathbf{e}_j), \\ A_2 &= 2 \sum_j \beta_j \, g(3\mathbf{e}_j) + 3 \sum_j \beta_j^2 \, g(4\mathbf{e}_j) + \frac{1}{2} \sum_{j_1 \neq j_2} \beta_{j_1} \beta_{j_2} \, g(2\mathbf{e}_{j_1} + 2\mathbf{e}_{j_2}), \end{aligned} \tag{5.4}$$

where j, j_1, j_2 are class indices each with range $[1, p]$ and $\mathbf{e}_j$ is the p-tuple with the jth element unity and all other elements zero.

The important feature of these expressions, which holds in general, is that they are identical to the expressions given in [15] for networks with load-independent queuing centers. The sole caveat is that the value of A_0 is given as 1 in [15], which happens to be the value of $g(\mathbf{0})$ as defined there, while here $g(\mathbf{0}) = \phi(0)$.

On account of the above-mentioned feature, the enumerations and characterizations of the pseudonetworks given previously [15, 16] apply as well in the present load-dependent context. The computational efficiencies that stem from the decomposition principle underlying the use of pseudonetworks have also been previously documented [20]. We summarize in

PROPOSITION 5. *The expansion coefficients $A_0, A_1, \ldots, A_r (r \geq 0)$ are obtained by linearly combining the partition function values of the pseudonetwork in which the network population over all classes is $2r$.*

With small populations in the pseudonetworks, the recursive techniques given in [3] for calculating partition functions of load-dependent networks may be used to analyze the individual pseudonetworks.

6. *Proof of Asymptoticity Error Analysis*

6.1 A BASIC LEMMA. Define for $x \geq 0$, $s \geq 0$,

$$A(x, s) \triangleq (1 + xs)^{1/x}. \tag{6.1}$$

We have previously proved in [15] that for $n = 0, 1, 2, \ldots$

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} A(x, s) \leq (-1)^n \frac{\partial^n}{\partial x^n} A(0, s), \qquad 0 \leq x \leq \infty; \tag{6.2}$$

that is, $A(x, s)$ is a completely monotonic function [18] of x for all $x \geq 0$. We state in the following proposition that a similar property is also true for $\partial^m A(x, s)/\partial s^m$, $m \geq 1$, for all sufficiently small x. The proof is in Appendix B.

PROPOSITION 6

(*i*)
$$\frac{\partial^m}{\partial s^m} A(x, s) = P_m(x) Q_m(x, s), \qquad m \geq 1, \tag{6.3}$$

where

$$P_m(x) = \begin{cases} 1, & \text{if } m = 1, \\ (-1)^{m-1} \prod_{k=1}^{m-1} \left(x - \frac{1}{k}\right), & \text{if } m \geq 2, \end{cases} \tag{6.4}$$

and

$$Q_m(x, s) = (m - 1)! \frac{A(x, s)}{(1 + xs)^m}, \qquad m \geq 1. \tag{6.5}$$

(*ii*) *For $m = 1, 2, \ldots, n = 0, 1, 2, \ldots, s \geq 0$*

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} P_m(x) \leq (-1)^n \frac{\partial^n}{\partial x^n} P_m(0), \qquad 0 \leq x \leq \frac{1}{m - 1}; \tag{6.6}$$

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} Q_m(x) \leq (-1)^n \frac{\partial^n}{\partial x^n} Q_m(0), \qquad 0 \leq x \leq \infty. \tag{6.7}$$

(*iii*) *For $m = 1, 2, \ldots, n = 0, 1, 2, \ldots, s \geq 0$ and $0 \leq x \leq 1/(m - 1)$,*

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial s^m} A(x, s) \leq (-1)^n \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial s^m} A(0, s). \tag{6.8}$$

6.2 COMPLETE MONOTONICITY IN THE INTEGRAL. We use the proposition just stated to establish that the function θ and its derivatives with respect to t possess certain remarkable properties. Recall that we have previously shown that the partition function $G_q(\mathbf{K})$ is simply related to the integral $I(N)$, where

$$I(N) = \int_{Q^+_{q-1}} e^{-\mathbf{1}'\mathbf{u}} \left[\sum_{m=0}^{\beta N} \frac{f(m)}{m!} \frac{\partial^m \theta}{\partial t^m} \bigg|_{t=0} \right] d\mathbf{u}; \tag{6.9}$$

$$\theta(N^{-1}, \mathbf{u}, t) = \prod_{j=1}^{p} [\exp(-\beta_j \tilde{\mathbf{\Gamma}}'\mathbf{u})] \left(1 + \frac{\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t}{N}\right)^{\beta_j N}. \tag{6.10}$$

It is convenient to write

$$\theta(N^{-1}, \mathbf{u}, t) = \prod_{j=1}^{p} \theta_j(N^{-1}, \mathbf{u}, t), \tag{6.11a}$$

where

$$\theta_j(\mathrm{N}^{-1}, \mathbf{u}, t) \triangleq [\exp(-\beta_j \tilde{\mathbf{\Gamma}}_j' \mathbf{u})] \left(1 + \frac{\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t}{N}\right)^{\beta_j N}, \qquad 1 \leq j \leq p. \tag{6.11b}$$

Now note that for $0 \leq m \leq \beta N$,

$$\frac{1}{m!} \frac{\partial^m \theta}{\partial t^m} = \sum \prod_{j=1}^{p} \left\{ \frac{1}{m_j!} \frac{\partial^{m_j} \theta_j}{\partial t^{m_j}} \right\}, \tag{6.12a}$$

where the sum is taken over integral $m_1, m_2, \ldots, m_p$ such that

$$0 \leq m_j \leq \beta_j N, \qquad 1 \leq j \leq p, \qquad \text{and} \qquad m_1 + m_2 + \cdots + m_p = m. \tag{6.12b}$$

The first set of constraints follows from the fact that θ_j is a polynomial in t of degree $\beta_j N$.

At this stage it will be useful to view

$$\omega \triangleq \frac{1}{N} \tag{6.13}$$

as a nonnegative real variable. In particular, therefore, from (6.11),

$$\theta = \theta(\omega, \mathbf{u}, t) = \prod_{j=1}^{p} \theta_j(\omega, \mathbf{u}, t), \tag{6.14a}$$

where

$$\theta_j(\omega, \mathbf{u}, t) = \exp(-\beta_j \tilde{\mathbf{\Gamma}}'\mathbf{u}) \{1 + \omega(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t)\}^{\beta_j/\omega}. \tag{6.14b}$$

We claim that for $1 \leq j \leq p$, $m = 0, 1, 2, \ldots$, $n = 0, 1, 2, \ldots$, $\mathbf{u} \in Q^+_{q-1}$,

$$\begin{aligned} 0 &\leq (-1)^n \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta_j(\omega, \mathbf{u}, t) \\ &\leq (-1)^n \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta_j(0, \mathbf{u}, t), \qquad 0 \leq \omega \leq \frac{\beta_j}{m-1}, \end{aligned} \tag{6.15}$$

that is, $\partial^m \theta_j / \partial t^m$ is a completely monotonic function of $\omega \in [0, \beta_j/(m-1)]$. Note that the interval $0 \leq \omega \leq \beta_j/(m-1)$ subsumes the range $0 \leq m \leq \beta_j N$ in which $\partial^m \theta_j / \partial t^m$ is of interest in (6.12).

The proof of (6.15) follows from Proposition 6. For

$$[\exp(\beta_j \tilde{\mathbf{\Gamma}}'\mathbf{u})] \theta_j(\omega, \mathbf{u}, t) = A(x, s), \tag{6.16a}$$

where

$$\frac{\omega}{\beta_j} = x, \qquad \beta_j(\tilde{\mathbf{\Gamma}}_j' \mathbf{u} + \Gamma_{jq} t) = s. \tag{6.16b}$$

With this identification, (6.15) is equivalent to (6.8) in statement (iii) of the proposition.

By using (6.15) in (6.12), it is now straightforward to establish the complete monotonicity of $\partial^m \theta / \partial t^m$ as a function of ω for $0 \leq \omega \leq \beta / m$ since sums and products of completely monotonic functions are themselves completely monotonic [18]. In summary, we have

PROPOSITION 7. *For* $0 \leq m \leq \beta/\omega$, $n = 0, 1, 2, \ldots$, $\mathbf{u} \in Q_{q-1}^+$

$$0 \leq (-1)^n \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta(\omega, \mathbf{u}, t) \leq (-1)^n \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, t). \tag{6.17}$$

6.3 ERROR ANALYSIS. An immediate corollary to (6.17) is that for $0 \leq m \leq \beta/\omega$, $r = 1, 2, \ldots,$

$$\begin{aligned} 0 &\leq (-1)^r \left[\frac{\partial^m}{\partial t^m} \theta(\omega, \mathbf{u}, 0) - \sum_{n=0}^{r-1} \omega^n \left\{ \frac{1}{n!} \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0) \right\} \right] \\ &\leq (-1)^r \omega^r \left\{ \frac{1}{r!} \frac{\partial^r}{\partial \omega^r} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0) \right\}. \end{aligned} \tag{6.18}$$

For, by Taylor's theorem,

$$\begin{aligned} &(-1)^r \left[\frac{\partial^m}{\partial t^m} \theta(\omega, \mathbf{u}, 0) - \sum_{n=0}^{r-1} \omega^n \left\{ \frac{1}{n!} \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0) \right\} \right] \\ &\quad = (-1)^r \frac{\omega^r}{r!} \frac{\partial^r}{\partial \omega^r} \frac{\partial^m}{\partial t^m} \theta(\xi, \mathbf{u}, 0), \end{aligned} \tag{6.19}$$

where $0 < \xi < \omega$. But the right-hand side is subject to the lower and upper bounds given in Proposition 7, and these give (6.18).

With respect to the range of validity of (6.18), observe that for $\omega > \beta/m$ but such that β_j/ω, $1 \leq j \leq p$, are integers, (6.18) is also true; in fact, $\partial^m \theta / \partial t^m = 0$. This is apparent from consulting (6.11) and (6.12).

In summary we have

PROPOSITION 8. *For all values of* ω *where* β_j/ω, $1 \leq j \leq p$, *are positive integers*

$$\begin{aligned} 0 &\leq (-1)^r \left[\frac{\partial^m}{\partial t^m} \theta(\omega, \mathbf{u}, 0) - \sum_{n=0}^{r-1} \omega^n \left\{ \frac{1}{n!} \frac{\partial^n}{\partial \omega^n} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0) \right\} \right] \\ &\leq (-1)^r \omega^r \left\{ \frac{1}{r!} \frac{\partial^r}{\partial \omega^r} \frac{\partial^m}{\partial t^m} \theta(0, \mathbf{u}, 0) \right\} \end{aligned} \tag{6.20}$$

for $r = 1, 2, \ldots.$

This result, which is also given in (4.1), is the key to the error analysis in Section 4.1. For, by multiplying by $f(m)/m!$, summing with respect to m for $0 \leq m < \infty$, multiplying by $e^{-\mathbf{1}'\mathbf{u}}$, and, finally, integrating with respect to $\mathbf{u} \in Q_{q-1}^+$, we obtain the error bounds given in (4.3).

6.4 TRUNCATION ALWAYS GIVES LOWER ERROR BOUNDS. In Sections 1.1, 3.2, and 3.4, we have claimed that truncating the series $\{f(n)\}$ at $n = K$ always gives lower error bounds. This is seen to be true by recalling the following chain of facts. From (3.7) it is clear that, for each n, $\phi(n)$ is smaller with truncation. From this observation and (5.1) it follows that for each $\mathbf{k}$, $g(\mathbf{k})$ is smaller with truncation. As a consequence (see (5.4)), each expansion coefficient A_n is smaller in magnitude with truncation. The claim is substantiated, finally, by inspecting (4.3a) which gives the error bounds.

7. *Generalization to Networks with Several Load-Dependent Centers*

So far, our notational system has been directed at the asymmetric case in which all but one of the q queuing centers in the network are load independent. Let us first make a few small modifications to correct this. These changes are suggested from the point of view that *all* q queuing centers are load dependent with load independence as a special case. First let (2.4a) apply for $1 \le i \le q$ with $n_i!$ replaced by $f_i(n_i)$. In the same vein let the sequence $\{\phi_i(n)\}$ be defined as in (3.6) for $1 \le i \le q$. Also, let the partition function of the pseudonetwork given in (5.1) be replaced by the following symmetrical form:

$$g(\mathbf{k}) = \sum_{n_{11}+\cdots+n_{1q}=k_1} \cdots \sum_{n_{p1}+\cdots+n_{pq}=k_p} \prod_{i=1}^{q} \left\{ \phi_i(n_i) \prod_{j=1}^{p} \frac{\Gamma_{ji}^{n_{ji}}}{n_{ji}!} \right\}. \tag{7.1}$$

Note that the multiplicative constant $1/\prod_{i=1}^{q-1} \alpha_i$ applied to the expression in (5.1) gives the expression in (7.1).

With these modifications incorporated, it can be shown that

$$G_q(\mathbf{K}) = \left[\prod_{j=1}^{p} \frac{\rho_{j0}^{K_j}}{K_j!} \right] I(N), \tag{7.2}$$

where $I(N)$ has the expansion given in (4.4) with the error bounds in (4.3a) and the expansion coefficients $\{A_n\}$ are composed from partition function values exactly as specified in Section 5. For instance, the leading expansion terms are given by (5.4).

The reader should verify that the above is consistent with the detailed results given earlier for the case of only one center with general load dependence.

Also, with these changes we are free to exercise the option that has been provided for the load-dependent center, namely, appropriate truncation of the series $\{f.(\cdot)\}$ and $\{\phi.(\cdot)\}$.

8. *Conclusions*

A complete theory has been developed for generating the entire asymptotic expansion of the partition function of a general class of Markovian, closed queuing networks with centers in which the service rate depends on the load. The theory, which is not simple, yields a simple computational technique. Equation (5.4) encapsulates the procedure; it states that the leading coefficients of the expansion are exactly given as linear combinations of many values of the partition function of the pseudonetwork, a construct. The computation of these values may be undertaken by any conventional procedure since the class populations in the pseudonetwork are small, as attested to by (5.4). The computational technique is accompanied by an error bound that is explicit and effective. If the need for more accurate solutions exists, then typically this is satisfied by computing additional terms of the expansion. Finally, the technique is free of problems related to numerical stability, convergence, and uniqueness.

Appendix A. Alternative Integral Representation of Partition Function

We make an observation that is quite important, even though the approach that it suggests is not systematically developed in this paper. Given the sequence $\{f(n)\}$, we may ask for a kernel function $p(t)$ such that

$$\int_0^\infty p(t)t^n \, dt = f(n), \qquad n \geq 1. \tag{A1}$$

In fact our previous work [15, 16] on the load-independent case [$f(n) = n!$, $p(t) = e^{-t}$] relied on such a representation. For the examples in 2.2.3, 2.2.4, and 2.2.5, respectively, it is easy to verify the following solutions for the kernel function

$$p(t) = \frac{1}{\Gamma(a+1)} e^{-t}t^a, \qquad \frac{1}{r!} e^{-t}t^r, \qquad \sum_{k=1}^{r} a_k \mu_k e^{-\mu_k t}. \tag{A2}$$

For the case of s homogeneous servers (see 2.2.2), a kernel function is a distribution [9]:

$$p(t) = \frac{s^s}{(s-1)!} e^{-st} - \sum_{m=0}^{s-2} (-1)^m c_m \delta^{(m)}(t), \tag{A3}$$

where

$$c_m = \left(\frac{s^s}{s!\,s^m} - \frac{1}{m!}\right)$$

and $\delta^{(m)}(t)$ is the mth derivative of the unit impulse function.

It is also known that any sequence $\{f(n)\}$ has a representation

$$f(n) = \int_0^\infty t^n \, dm(t), \qquad n \geq 1, \tag{A4}$$

in which $m(t)$ is a function of bounded variation [26]. The function $m(t)$ is not unique, and, in general, it is a difficult task to determine an $m(t)$ given $\{f(n)\}$. It should be noted that the representations (A1) and (A4) are the same for the examples of $p(t)$ given in (A2). We need only set $m(t) = \int_0^t p(\tau)\, d\tau$ to see this. However, the two representations are different in the case of s homogeneous servers for $s \geq 3$. Note that, if (A1) holds, then

$$\int_0^\infty p(t)e^{(1-\alpha)t^n} \, dt = \phi(n), \qquad n \geq 0, \tag{A5}$$

where $\phi(n)$ is defined in (3.7).

A very useful representation of a class of sequences $\{f(n)\}$, which we make use of, can be obtained in terms of double integrals. It is seen that this class includes all the cases of interest in this paper. The representation is motivated by work of Rooney [22], but we give a self-contained derivation here.

Instead of $\{f(n)\}$, consider the sequence $\{g(n)\}$ defined by

$$g(0) = 1, \qquad g(n) = \frac{f(n)}{n!}, \qquad n \geq 1. \tag{A6}$$

Then, since

$$\overline{\lim_{n\to\infty}} \left|\frac{g(n+1)}{g(n)}\right| = \mu_q c_q < 1,$$

the condition of normal usage for the node guarantees absolute convergence of the series $\sum_{n=0}^{\infty} g(n)$ [13]. Next, define the function

$$m(w) = \sum_{k=0}^{\infty} g(k) e^{-w/2} L_k(w), \tag{A7}$$

where $L_k(w)$ is the Laguerre polynomial of order k [19]. Because of the inequality [26]

$$|e^{-w/2} L_k(w)| \le 1, \qquad 0 \le w, \qquad k = 0, 1, \dots, \tag{A8}$$

it follows that the series defining $m(w)$ is uniformly and absolutely convergent, and so $m(w)$ is well defined. Further, the series defining $m(w)$ can be multiplied by $e^{-1/2w} L_n(w)$ and integrated from 0 to ∞ term by term. Since [19]

$$\int_0^\infty e^{-w} L_k(w) L_n(w)\, dw = \delta_{n,k}, \tag{A9}$$

we have shown that

$$\int_0^\infty e^{-w/2} L_n(w) m(w)\, dw = g(n), \qquad n = 0, 1, \dots. \tag{A10}$$

The representation (A10) can now be combined with the standard integral representation of the Laguerre polynomial [19],

$$L_n(w) = \frac{e^w}{n!} \int_0^\infty e^{-v} v^n J_0(2\sqrt{vw})\, dv, \tag{A11}$$

to yield the desired representation.

$$f(n) = n!\, g(n) = \int_0^\infty \left[\int_0^\infty e^{-v} v^n J_0(2\sqrt{vw})\, dv\right] e^{w/2} m(w)\, dw. \tag{A12}$$

In (A12) $J_0(x)$ is the Bessel function of the first kind of order zero. Note that the order of integration in (A12) cannot be interchanged.

It will turn out that closed-form expressions for $m(w)$ will not be needed for computational purposes. However, as an example, the weight function for the homogeneous s-server queuing node of 2.2.2 is

$$m(w) = \sum_{n=0}^{s-2} \left(\frac{1}{n!} - \frac{1}{s^n}\frac{s^s}{s!}\right) \exp\left(-\frac{1}{2}w\right) L_n(w) + \frac{s^s}{s!}\frac{s}{s-1} \exp\left[-\frac{1}{2}\left(\frac{s+1}{s-1}\right) w\right]. \tag{A13}$$

We now use the double integral representation (A12) of $f(n)$ to obtain another representation of the partition function. Replace $f(m)$ in (2.17) by (A12), and interchange the order of integration and (finite) summation. Then, as in [16], it is straightforward to show that

$$G_q(K) = \left[\frac{1}{\prod_{j=1}^{p} K_j!}\right] \int_0^\infty \left[\int_{Q_q^+} e^{-\mathbf{1}'\mathbf{u}} D_q(\mathbf{K}, \mathbf{u}) J_0(2\sqrt{u_q x})\, d\mathbf{u}\right] e^{x/2} m(x)\, dx. \tag{A14}$$

In (A14),

$$\begin{aligned}
\mathbf{u} &= (u_1, u_2, \ldots, u_q)', \\
\rho_j &= (\rho_{j1}, \rho_{j2}, \ldots, \rho_{jq})', \\
\mathbf{1} &= (1, 1, \ldots, 1)', \\
Q_q^+ &= \{\mathbf{u} \mid u_i \geq 0,\ 1 \leq i \leq q\}, \\
D_q(\mathbf{K}, \mathbf{u}) &= \prod_{j=1}^{p} (\rho_{j0} + \rho_j' \mathbf{u})^{K_j},
\end{aligned} \tag{A15}$$

and ρ_{j0} and $\mathbf{K}$ are as in (2.16).

We next examine the relationship between the representation (2.18) and (A14) for $G(\mathbf{K})$. For any polynomial $p(t)$, the Hankel transform relation holds [25]:

$$e^{-t}p(t) = \int_0^\infty \left[\int_0^\infty e^{-u}p(u)J_0(2\sqrt{ux})\,du\right] J_0(2\sqrt{xt})\,dx. \tag{A16}$$

The following generating function for the Laguerre polynomials is known [19]:

$$J_0(2\sqrt{xt}) = \sum_{n=0}^{\infty} e^{-t}\frac{t^n}{n!}L_n(x). \tag{A17}$$

It is easy to show that, if (A17) is substituted into (A16), the order of summation and integration can be interchanged and the factor of e^{-t} can be canceled on both sides of the equation to yield

$$p(t) = \sum_{n=0}^{\infty}\left\{\int_0^\infty \left[\int_0^\infty e^{-u}p(u)J_0(2\sqrt{ux})\,du\right] L_n(x)\,dx\right\}\frac{t^n}{n!}. \tag{A18}$$

However, (A18) implies immediately that

$$\int_0^\infty \left[\int_0^\infty e^{-u}p(u)J_0(2\sqrt{ux})\,du\right] L_n(x)\,dx = \frac{d^n}{dt^n}p(t)\bigg|_{t=0}. \tag{A19}$$

If expression (A7) for $m(w)$ is now substituted into (A14), it is easy to see that the order of summation and integration can be interchanged, since $D_q(\mathbf{K}, \mathbf{u})$ is a polynomial in u_q. If this is done, and use is made of (A19) and the definition of $D_q(\mathbf{K}, \mathbf{u})$, it is seen that (A14) reduces to (2.18).

Since we assume that all the centers are in normal usage, from the remarks following (3.6), we can always choose μ_q so that $m(w)$ is well defined and $\alpha_q > 0$.

Hence

$$\alpha_i > 0, \qquad 1 \leq i \leq q.$$

Then if we make the change of variables

$$\alpha_i u_i \rightarrow u_i, \qquad 1 \leq i \leq q, \qquad x = \alpha_q v,$$

we obtain the following expression for $I(N)$ as defined in (3.18) and (3.19):

$$I(N) = \int_0^\infty \left[\int_{Q_q^+} e^{-\mathbf{1}'\mathbf{u}}H(N^{-1}, \mathbf{u})J_0(2\sqrt{u_q v})\,d\mathbf{u}\right] e^{v/2}\hat{m}(v)\,dv. \tag{A20}$$

Here

$$H(N^{-1}, \mathbf{u}) = H\left(N^{-1}, \mathbf{u}_{q-1}, \frac{u_q}{\alpha_q}\right), \tag{A21}$$

where we have partitioned the q vector $\mathbf{u}$ into a $q - 1$ vector, $\mathbf{u}_{q-1}$, and the last component u_q, $\mathbf{u} = (\mathbf{u}_{q-1}, u_q)$, and

$$\hat{m}(v) = \exp\left[-\frac{1}{2}(1 - \alpha_q)v\right]m(\alpha_q v) = \sum_{n=0}^{\infty} g(n)\exp\left(-\frac{1}{2}v\right)L_n(\alpha_q v). \tag{A22}$$

However [19],

$$L_n(\alpha_q v) = \sum_{j=0}^{n} \binom{n}{j} \alpha_q^j (1 - \alpha_q)^{n-j} L_j(v), \tag{A23}$$

and if we substitute this expression for $L_n(\alpha_q v)$ into (A22) and interchange the order of summation, we obtain

$$\hat{m}(v) = \sum_{j=0}^{\infty} \hat{g}(j) e^{-v/2} L_j(v), \tag{A24}$$

where

$$\hat{g}(j) = \alpha_q^j \sum_{n=0}^{\infty} \binom{n+j}{j} (1 - \alpha_q)^n g(n + j) = \frac{\alpha_q^j \phi(n)}{j!}, \tag{A25}$$

with $\phi(n)$ defined in (3.7). If we replace $H(N^{-1}, \mathbf{u})$ in (A20) by its expansion in powers $1/N$, we formally obtain an expansion of the form (4.4), but now the coefficients A_n are

$$A_n = \int_0^{\infty} \left[\int_{Q_q^+} e^{-\mathbf{1}'\mathbf{u}} h_n\left(\mathbf{u}_{q-1}, \frac{u_q}{\alpha_q}\right) J_0(2\sqrt{u_q v})\, d\mathbf{u}_q\right] e^{v/2} \hat{m}(v)\, dv. \tag{A26}$$

However, if expression (A24) for $\hat{m}(v)$ is substituted into (A26), the order of summation and integration are interchanged (valid since h_n is a polynomial), and use is made of (A17) and (A25), it is clear that the expressions (4.12) and (A26) for A_n are the same. Expression (A26) thus provides an alternative expression for the coefficients in the asymptotic expansion of $I(N)$.

Appendix B. Proof of Proposition 6

(i) Let the induction hypothesis be

$$\frac{\partial^r}{\partial s^r} A(x, s) = P_r(x) Q_r(x, s), \qquad r = 1, 2, \ldots, m, \tag{B1}$$

where the functions P_r and Q_r, $r \geq 1$, are as defined in (6.5) and (6.6). It may be directly verified that (B1) is true for $m = 1$.

Now, from (B1),

$$\begin{aligned} \frac{\partial^{m+1}}{\partial s^{m+1}} A(x, s) &= P_m(x) \frac{\partial}{\partial s} Q_m(x, s) \\ &= \left\{P_m(x)\left(\frac{1}{m} - x\right)\right\}\left\{m!\, \frac{A(x, s)}{(1 + xs)^{m+1}}\right\} \\ &= P_{m+1}(x) Q_{m+1}(x, s), \end{aligned} \tag{B2}$$

since the functions $\{P_m(x)\}$ satisfy the recursion

$$P_{m+1}(x) = -P_m(x)\left(x - \frac{1}{m}\right), \qquad m \geq 1. \tag{B3}$$

This completes the inductive proof.

(ii) To show (6.6) it is enough to prove that for the range of m, n, s, and x in the statement,

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} P_m(x). \tag{B4}$$

For, (B4) implies that $(-1)^n \partial^n P_m(x)/\partial x^n$ is a nonnegative, monotonic nonincreasing function of x, and, therefore, the right inequality in (6.6) follows.

The proof of (B4) is again by induction. The induction hypothesis is

$$0 \leq (-1)^n \frac{\partial^n}{\partial x^n} P_r(x), \qquad r = 1, 2, \ldots, m, \qquad n = 0, 1, 2, \ldots,$$

$$0 \leq x \leq \frac{2}{r-1}. \tag{B5}$$

Now, from (B3), it follows that for $n = 0, 1, 2, \ldots,$

$$\begin{aligned}(-1)^n \frac{\partial^n}{\partial x^n} P_{m+1}(x) &= \frac{(-1)^n}{m} \frac{\partial^n}{\partial x^n} P_m(x) - (-1)^n \frac{\partial^n}{\partial x^n}\left\{xP_m(x)\right\} \\ &= \left(\frac{1}{m} - x\right)(-1)^n \frac{\partial^n}{\partial x^n} P_m(x) + n(-1)^{n-1} \frac{\partial^{n-1}}{\partial x^{n-1}} P_m(x) \\ &\geqslant 0, \qquad \text{for} \quad 0 \leq x \leq \frac{1}{m}.\end{aligned} \tag{B6}$$

This completes the inductive proof of (6.6).

The proof of (6.7) is straightforward. Equation 6.3 shows that $A(x, s)$ is a completely monotonic function of x for all $x \geq 0$, and the same is easily verified to be true for $(1 + xs)^{-m}$. As the product of completely monotonic functions is completely monotonic [18], $Q_m(x, s)$ is also a completely monotonic function of x for all $x \geq 0$.

(iii) As $P_m(x)$ and $Q_m(x, s)$ have been shown in (ii) to be completely monotonic for m, n, s, and x in the range of interest here, it follows that their product is also completely monotonic, which is what (6.8) states. □

REFERENCES

1. BARD, Y. Some extensions of multiclass queueing network analysis. In *Performance of Computer Systems*, M. Arato, A. Butrimenko, and E. Gelenbe, Eds. North-Holland, New York, 1979.
2. BASKETT, F., CHANDY, K. M., MUNTZ, R. R., AND PALACIOS, F. Open, closed and mixed networks of queues with different classes of customers. *J. ACM 2*, 2 (Apr. 1975), 248-260.
3. BRUEL, S. C., AND BALBO, G. *Computational Algorithms for Closed Queueing Networks.* North-Holland, New York, 1980.
4. CHANDY, K. M., AND NEUSE, D. Linearizer: A heuristic algorithm for queueing network models of computing systems. *Commun. ACM 25*, 2 (Feb. 1982), 126-134.
5. CHANDY, K. M., HERZOG, U., AND WOO, L. S. Parametric analysis of queueing networks. *IBM J. Res. Develop. 19*, 1 (Jan. 1975), 43-49.
6. CHANDY, K. M., HERZOG, U., AND WOO, L. S. Approximate analysis of general queueing networks. *IBM J. Res. Develop. 19*, 1 (Jan. 1975), 50-57.

7. COOPER, R. B. *Introduction to Queueing Theory*. 2nd Ed. North-Holland, New York, 1981.

8. DE SOUZA E SILVA, E., LAVENBERG, S. S., AND MUNTZ, R. R. A perspective on iterative methods for the approximate analysis of closed queueing networks. In *Proceedings of the International Workshop on Applied Mathematics and Performance Reliability Models of Computer Communication Systems* (Pisa, Italy). North-Holland, New York, 1983, pp. 191–210.

9. DOETSCH, G. *Introduction to the Theory and Application of the Laplace Transformation*. Springer-Verlag, New York, 1974.

10. HEFFES, M. Moment formulae for a class of mixed multi-job-type queueing networks. *Bell Syst. Tech. J. 61*, 5 (May–June 1982), 709–745.

11. KELLY, F. P. *Reversibility and Stochastic Networks*. Wiley, New York, 1980.

12. KLEINROCK, L. *Queueing Systems*. Vol. II: *Computer Applications*. Wiley, New York, 1976.

13. KNOPP, K. *Theory and Application of Infinite Series*. Blackie, London, 1928.

14. LAVENBERG, S. S. Closed multichain product form queueing networks with large population sizes. In *Applied Probability—Computer Science The Interface*, vol. 1, R. L. Disney and T. J. Ott, Eds. Birkhauser, Boston, 1982, pp. 219–249.

15. MCKENNA, J., AND MITRA, D. Integral representations and asymptotic expansions for closed Markovian queueing networks: Normal usage. *Bell Syst. Tech. J. 61*, 5 (May–June 1982), 661–683.

16. MCKENNA, J., AND MITRA D. Asymptotic expansions and integral representations of moments of queue lengths in closed Markovian networks. *J. ACM 31*, 2 (Apr. 1984), 346–360.

17. MCKENNA, J., MITRA, D., AND RAMAKRISHNAN, K. G. A class of closed Markovian queueing networks: Integral representations, asymptotic expansions, generalizations. *Bell Syst. Tech. J. 60*, 5 (May–June 1981), 599–641.

18. OLVER, F. W. J. *Introduction to Asymptotics and Special Functions*. Academic Press, Orlando, Fla., 1974.

19. RAINVILLE, E. D. *Special Functions*. Chelsea, New York, 1960.

20. RAMAKRISHNAN, K. G., AND MITRA, D. An overview of PANACEA, A software package for analyzing Markovian queueing networks. *Bell Syst. Tech. J. 61*, 10 (Dec. 1982), 2849–2872.

21. RAMAKRISHNAN, K. G., AND MITRA, D. PANACEA 2.1: User's Manual and Theory. Bell Labs. Technical Memorandum, 1983.

22. ROONEY, P. G. On the Laguerre and Hermite coefficient problems. *J. Math. Anal. Appl. 4* (1962), 475–487.

23. SCHWEITZER, P. Approximate analysis of multiclass closed networks of queues. In *Proceedings of the International Conference on Stochastic Control and Optimization* (Amsterdam, The Netherlands). 1979.

24. SAUER, C. H., AND CHANDY, K. M. *Computer Systems Performance Modeling*. Prentice-Hall, Englewood Cliffs, N.J., 1981.

25. TITCHMARSH, E. C. *Fourier Integrals*. Oxford University Press, Oxford, England, 1937.

26. WIDDER, D. V. *The Laplace Transform*. Princeton University Press, Princeton, N.J., 1941.

27. ZAHORJAN, J., AND LAZOWSKA, E. D. Incorporating load dependent servers in approximate mean value analysis. In *Proceedings of the 1984 ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems* (Cambridge, Mass., Aug. 21–24). ACM, New York, 1984, pp. 52–62.

RECEIVED AUGUST 1984; REVISED AUGUST 1985; ACCEPTED SEPTEMBER 1985