SRC TR 86-44

Information Theoretic Analysis For A General Queueing System At Equilibrium With Application To Queues In Tandem

by

J. Cantor, A. Ephremides and D. Horton

INFORMATION THEORETIC ANALYSIS FOR A GENERAL QUEUEING SYSTEM AT EQUILIBRIUM WITH APPLICATION TO QUEUES IN TANDEM

by

J. Cantor
Business and Technological Systems, Inc.
10210 Greenbelt Road
Suite 440
Seabrook, Maryland 20706

A. Ephremides
Electrical Engineering Department
University of Maryland
College Park, Maryland 20742

D. Horton
Formerly of: Westinghouse Defense & Electronics Center
Baltimore, Maryland
Now with: Martin Marietta Aerospace
Orlando, Florida 32822

INFORMATION THEORETIC ANALYSIS FOR A GENERAL QUEUEING SYSTEM AT EQUILIBRIUM WITH APPLICATION TO QUEUES IN TANDEM*

ABSTRACT

In this paper, information theoretic inference metholology for system modeling is applied to estimate the probability distribution for the number of customers in a general, single server queueing system with infinite capacity utilized by an infinite customer population. Limited to knowledge of only the mean number of customers and system equilibrium, entropy maximization is used to obtain an approximation for the number of customers in the G|G|1 queue. This maximum entropy approximation is exact for the case of G = M, i.e., the M|M|1 queue. Subject to both independent and dependent information, an estimate for the joint customer distribution for queueing systems in tandem is presented. Based on the simulation of two queues in tandem, numerical comparisons of the joint maximum entropy distribution is given. These results serve to establish the validity of the inference technique and as an introduction to information theoretic approximation to queueing networks.

*This work was supported under a Naval Research Laboratory Fellowship under Grant N00014-83G-0203 and under an ONR Grant N00014-84K-0614.

I. INTRODUCTION

Classical queueing theory [1,2,3] has proven to be quite successful in modeling both communication networks [4,5] and computer systems [7]. In most cases, rather unrealistic assumptions about either the underlying arrival process (e.g., Poisson) or service distribution (e.g., negative exponential) must be employed to obtain results as it is under these assumptions that queueing theory most easily yields solutions. Because the processes of "real world" systems generally do not obey this exponential structure, the success of the models has remained in doubt.

Shore [8] has proposed an explanation based on information theoretical system modeling and the princiiple of maximum entropy or, more generally, the principle of relative-entropy [9]. In Shore's approach, an abstract system model consisting of various "states" is introduced. The probability of the occurrence of a certain state is estimated by the maximum entropy distribution subject to known information in the form of expected values of functions of the states. A relationship between the abstract system and the actual system is thereby established and estimates of desirable probability distributions of the model are obtained. Shore utilizes this technique to derive both the equilibrium and time-dependent probability distribution for the number of customers (number of jobs, number of calls, etc.) in $M|M|\infty|\ |N$ and $M|M|\infty$ queueing systems.

In this paper, information theoretic system modeling is applied to estimate the probability distribution for the number of customers in a general, $G|G|1$, queueing system. The extension to queueing networks is considered through the approximation of N systems in tandem.

Beneš [10], it appears, first proposed use of maximum entropy and statistical mechanical analysis of large-scale communication systems. For a telephone system in which only the expected number of calls in progress is known, Beneš shows that the maximum entropy distribution is precisely that which is obtained as the equilibrium distribution of an ergodic, reversible birth-death Markov process with constant birth and death rates. A direct implication of this result is that constrained only by the expected number of customers in the system, the maximum entropy approximation to a single service queueing system is the celebrated M|M|1 queue. The maximum entropy distribution is, of course, obtained independently of the underlying stochastic processes and under moderately few technical assumptions [9].

Ferdinand [11] uses the principles of statistical mechanics to derive the soution to the M|M|1|N (finite capacity) queue. In later work, Shore [12] establishes maximum entropy (termed information theoretic) approximations for a number of "performance distributions" of M|G|1 and G|G|1 queues at equilibrium. These performance distributions, such as the number of customers in the system, a customer's waiting time, or the number served in a busy period, are estimated subject to moments of the interarrival and service time distributions. It is demonstrated that, using relatively few moments, that maximum entropy provides good approximations to a variety of M|G|1 systems. Further, for many of the distributions, the approximations yield exact results when G = M. Using an approach similar to that of Shore, El-Affendi and Kouvatsos [13] independently establish a maximum entropy approximation to the number of customers in a M|G|1 system as well as the service distribution corresponding to the estimate. Further, an approximation to a specific G|M|1 queue is determined.

In Section II of this paper, the general relative entropy formalism and technique for solution are presented, and the specialized case of entropy maximization is discussed. These techniques are applied in Section III to derive the maximum entropy approximations to the G|G|1 queueing system. In Section IV, utilizing the G|G|1 approximation, the maximum entropy approximation to the joint distribution of the number of customers in N queues in tandem is established. A discussion of maximum entropy "product form" solutions subject to information on the marginal and joint distribution is included. In Section V, numerical comparisons of the maximum entropy approximation for known or simulated distributions for two queues in tandem are given. The paper is concluded by a general discussion in Section VI.

II. PROBLEM STATEMENT AND RELATIVE-ENTROPY MINIMIZATION

Consider a system that has a countable set S of possible states with

$$p(S_i) > 0, \quad S_i \varepsilon S \ , \quad i = 1,2,\dots \ . \tag{1}$$

$$\sum_i p(S_i) = 1 \tag{2}$$

where $p(S_i)$ is the probability of the occurrence of the state S_i. Assume that there exists a "true" distribution, $q^+ \varepsilon D$ which is unknown. It is desirable to estimate this distribution q^+ based on incomplete information. Let p be the current estimate or initial value distribution of q^+.

Suppose new information about q^+ becomes available in the form of expected values of known functions, f_ℓ; $\ell = 1,2,\dots,M$, of the states as follows:

$$\sum_i q^+(S_i) f_\ell(S_i) = \langle f_\ell \rangle \ . \tag{3}$$

The constraints (1)-(3) do not precisely identify q^+. Indeed, besides the true distribution, q^+, there exists a subset of distributions D' of D which also satisfies all constraints. One way of uniquely choosing an estimate for q^+, well-accepted in the literature [14,15,9], is the method of minimizing the relative entropy (also known as cross-entropy, Kullback-Leiber number, directed divergence, or discrimination information), namely, choose $q \varepsilon D$ so that $H[q,p]$ defined by

$$H[q,p] = \sum_i q(S_i)\log(\frac{q(S_i)}{p(S_i)}) \tag{4}$$

is minimized. The choice of q as above is called the final value distribution. For estimating probability distributions, relative-entropy minimization has been shown to be self-consistent and uniquely correct [9], and, therefore, the estimates are sometimes called information theoretic approximations. It turns out [15,16,17] that if there exists a solution to (4) such that the constraints (1)-(3) are satisfied, then that solution has the form

$$q(s_i) = p(S_i)\exp(-\beta_o - \sum_{\ell=1}^{M} \beta_\ell f_\ell(S_i)) \tag{5}$$

at all states, except possibly on a set for which q is identically zero [16]. In (5), β_ℓ, $\ell = 0,1,\ldots,M$ are Lagrangian multipliers. Further if β_ℓ can be determined such that the constraints (1)-(3) are satisfied, then a solution to (4) exists* and is given by (5). From (2) and (5), the following partition function $\exp(\beta_o)$ given by

$$\exp(\beta_o) = \sum_i^{M} \exp(-\sum_{\ell=1} \beta_\ell f_\ell(S_i)) \tag{6}$$

or

$$\beta_o = \log \sum_i \exp(-\sum_{\ell=1}^{M} \beta_\ell f_\ell(S_i)) \tag{7}$$

can be defined, noting that β_o is a function of the other multipliers. If the sum in (7) converges and, in particular, assumes a closed form solution, then the multipliers may be determined via the following relations:

$$-\frac{\partial\beta_o}{\partial\beta_\ell} = \langle f_\ell \rangle \, . \tag{8}$$

*The issue of existence of a solution has been studied by Csiszár [16].

In the event that no closed form is obtainable or the system (8) is inconsistent, the multipliers must be approximated by numerical techniques [18]. For systems that have an uncountable number of system states, the sums are replaced by integrals in the usual way.

Entropy Maximization

When little is known *a priori* about the distribution to be estimated, a natural choice for the initial value distribution is one in which all states are given equal weight, i.e., the *uniform* on S. In this case, the solution q in (4) is said to *maximize* the entropy H given by

$$H[q] = -\sum_i q(S_i)\log(q(S_i)) \ . \tag{9}$$

When a solution exists, q is denoted the *maximum entropy* approximation to q^+ and is given by

$$q(S_i) = \exp(-\beta_0 - \sum_{\ell=1}^{M} \beta_\ell f_\ell(S_i)) \ . \tag{10}$$

Note that (10) is identical to (5) with the initial value distribution deleted.

A large value of the entropy functional (9) corresponds to a high degree of uncertainty. The maximum entropy distribution can thus be interpreted as the probability distribution that reflects maximum uncertainty while utilizing all available information. In this sense, it is the "least biased" or "most conservative" distribution one can propose which satisfies all constraints.

In the following section, the methodology of relative-entropy minimization is applied to obtain approximations to the number of customers in a $G|G|1$ system at equilibrium.

III. MAXIMUM ENTROPY APPROXIMATION TO GENERAL QUEUEING SYSTEMS.

Consider a queueing system consisting of an infinite customer population from which individual customers arrive singly with interarrival times identically distributed according to a general distribution $A(t)$. The customers (possibly) wait in an infinite capacity buffer, then are served individually by a server according to a general service time distribution $B(t)$, and finally return to the arriving portion of the customer population. The queueing system is assumed to be at equilibrium with the steady state distribution to the number of customers in the system denoted by q^+.

Available information in a queueing system is often in the form of moments of the interarrival distribution

$$a_m = \int t^m dA(t) \qquad m = 1,2,\ldots$$

and the service time distribution

$$s_m = \int t^m dB(t) \qquad m = 1,2,\ldots .$$

The ratio of the first moments, $\rho = s_1/a_1$, is denoted the utilization factor, and it determines the traffic intensity or "loading" on the system [2].

Based on the expected number of customers in the system, $\langle K\rangle$, the maximum entropy approximation to $q^+(K)$ can be obtained subject to

$$\sum_{k=0}^{\infty} q(K = k) = 1 \tag{11}$$

and

$$\sum_{k=0}^{\infty} kq(K = k) = \langle K\rangle . \tag{12}$$

From (6)-(8) and (10), it follows that (see [12] for details)

$$q(K = k) = \frac{1}{1 + \langle K \rangle} \left(\frac{\langle K \rangle}{1 + \langle K \rangle}\right)^k . \tag{13}$$

The distribution, $q(K)$, has nothing specifically to do with a queueing system but is the maximum entropy distribution obtained from a single moment constraint and normalization and (13) is applicable to <u>any</u> system whether or not the system is at equilibrium when only the mean of the distribution to be estimated is available. What is of interest here is that (13) is the formula to the steady state distribution of the $M|M|1$ queueing system [2]. This can be seen by noting that for the $M|M|1$ queue, $q^+(K)$, is given by

$$q^+(K = k) = (1 - \rho)(\rho)^k \tag{14}$$

and

$$\langle K \rangle = \frac{\rho}{1 - \rho} \tag{15}$$

so, putting (15) into (14), (13) follows. The implication of the result is that the maximum entropy approximation to a $G|G|1$ system subject to only the mean number of customers is the $M|M|1$ queue. It turns out that this approximation is sometimes a satisfactory estimate for the $M|G|1$ system [12], especially when the service distribution is close to exponential, e.g., the $M|H_2|1$ system, but as a general $G|G|1$ approximation, there is no reason to assume (13) will be "close."

To specify the queueing system, we impose additional constraints. Namely, as is well-known [1,2], all single server systems at steady state satisfy the equilibrium condition

$$\frac{1 - q(0)}{s_1} = \frac{1}{a_1} \tag{16}$$

which has the interpretation that the average rate of arrivals to the system is equal to the average rate of departures. Now, subject to (11)-(12) and rewriting (16) as

$$q(0) = 1 - \rho \tag{17}$$

where $\rho = s_1/a_1$, the following queueing system approximation can be considered. The constraint (17) can be posed as

$$\sum_{k=0}^{\infty} I(k)q(K = k) = 1 - \rho$$

where

$$I(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k > 0 \end{cases} \quad . \tag{18}$$

The maximum entropy solution has the form

$$q(K = k) = \exp(-\beta_o - \beta_1 k - \beta_2 I(k)) \; . \tag{19}$$

For notational convenience, pose

$$x: = \exp(-\beta_0) \qquad y: = \exp(-\beta_1) \qquad z: = \exp(-\beta_2)$$

and observe from (19) that

$$q(K = k) = \begin{cases} xz & k = 0 \\ xy^k & k > 0 \end{cases} \quad . \tag{20}$$

From (11), (12), (17), and straightforward manipulations, we obtain:

$$y = \frac{\langle K \rangle - \rho}{\langle K \rangle}$$

$$z = \frac{y(1 - \rho)}{\rho(1 - y)} = (1 - \rho)\,\frac{(\langle K \rangle - \rho)}{\rho^2} \; .$$

$$x = \frac{(1 - y)}{z(1 - y) + y} = \frac{\rho^2}{\langle K \rangle - \rho} \; .$$

Substituting the values for x,y,z in (20), the approximation to the G|G|1 queueing system is given by

$$q(K = k) = \begin{cases} 1 - \rho & k = 0 \\ \left(\frac{\rho^2}{\langle K\rangle - \rho}\right)\left(\frac{\langle K\rangle - \rho}{\langle K\rangle}\right)^k & k > 0 \end{cases} \quad . \tag{21}$$

For the particular case of $\rho = \langle K\rangle/(1 + \langle K\rangle)$, (21) reduces to $q(K = k) = (1 - \rho)(\rho)^k$ the exact M|M|1 formula as expected. It is noteworthy that in [12] (17) is applied not as a constraint but as a condition on the solution, (13). This results in a G|G|1 approximation given by (14) which is clearly less general than (21).

In the following section, utilizing the approach which led to the G|G|1 approximation, maximum entropy approximations to the joint distribution of the number of customers in more than one queueing system is examined and the approximation to queues in tandem is presented.

IV. MAXIMUM ENTROPY APPROXIMATION TO QUEUEING SYSTEMS IN TANDEM

Under certain conditions, the relative-entropy joint distribution, or, specifically, the maximum entropy distribution, is equal to the product of relative-entropy distributions. The concept of product form distribution is stated precisely below.

Definition: A joint probability distribution $q^+(K_1,K_2,\ldots,K_N)$ is of the product form if

$$q^+(K_1,K_2,\ldots,K_N) = \prod_{i=1}^{N} q^+(K_i) ,$$

i.e., the joint distributrion is equal to the product of the marginal distributions $q^+(K_i)$, $1 \leq i \leq N$.

The above definition is, of course, equivalent to the random variables $K_1, K_2, \ldots, K_N$ being statistically independent. Much of this section will be devoted to the discussion of when the approxiation is of the product form.

Maximum Entropy Approximation to N General Queueing Systems.

Subject to constraints only on the mean number of customers in each system and equilibrium, it is now shown that the maximum entropy approximation to the joint distribution of the number of customers in N queueing systems is a "product form" solution. The equilibrium condition for each system is as before; the average arrival rate is equal to the average departure rate. For notational simplicity, let

$\langle K_i \rangle :=$ the expected number of customers in system i

and

$$\rho_i := \frac{s_1^{(i)}}{a_1^{(i)}} = \text{the utilization factor for system } i$$

where $s_1^{(i)}$ and $a_1^{(i)}$ are the first interarrival and service moments of system i. Under the above assumptions on the constraints of the joint distribution, the following result can be shown.

Proposition 1: The maximum entropy approximation to the distribution of the number of customers in N G|G|1 queueing systems at equilibrium is of the product form, i.e.,

$$q(K_1, K_2, \ldots, K_N) = \prod_{i=1}^{N} q(K_i) \tag{22}$$

where

$$q(K_i = k_i) = \begin{cases} 1 - \rho_i & k_i = 0 \\ \left(\dfrac{\rho_i^2}{\langle K_i \rangle - \rho_i}\right)\left(\dfrac{\langle K_i \rangle - \rho_i}{\langle K_i \rangle}\right)^k & k_i > 0 \end{cases} \tag{23}$$

Proof:

Because there are constraints only on the marginal distributions, the joint maximum entropy distribution has the form

$$q(K_1,K_2,\ldots,K_N) = \exp(-\beta_o - \sum_{i=1}^{2N} \beta_i f_i(K_1,K_2,\ldots,K_N))$$

where, by assumption, $f_i(K_1,K_2,\ldots,K_N)$ is a function of only K_i. Specifically,

$$\exp(-\beta_o - \sum_{i=1}^{2N} \beta_i f_i(K_1,K_2,\ldots,K_N)) = \exp(-\beta_o) \prod_{i=1}^{N} \exp(-\beta_i K_i - \lambda_i I(K_i)$$

by simply relabeling the multipliers $\beta_{2i} = \lambda_i$. $I(K_i)$ is defined by (18). Under the assumption that the marginal distributions satisfy the normalization constraint, the product form follows. By (22), it is clear that the maximum entropy approximation to each marginal distribution is exactly the G|G|1 approximation given by (21) and (23) is obtained. Q.E.D.

It should be emphasized that Proposition 1 is not a result on queueing networks. A queueing network is an inter-connected group of queueing systems where customers can enter an individual "node" (queueing system), extract service, and then either depart the network completely or go to another node and extract more service. In order to discuss approximating distributions to such systems, many new parameters must be considered, such as the network topology or the possible transitions between systems. To utilize information theoretic analysis, it is necessary to capture the system interaction through appropriate equilibrium constraints. In the certain instances where one has knowledge of rate balance equations for each individual node of a network as well as system equilibrium rate balance equations, one possible approach follows from relative-entropy minimization subject to fully decomposable subset and aggregate constraints [19].

This technique is not considered here; instead, it turns out that the approach utilized for the single queueing system is applicable to an elementary queueing network, namely, queueing systems in tandem.

<u>Maximum Entropy Approximation to Tandem Queues at Equilibrium.</u>

A simple queueing network topology is N queues in tandem. The tandem queueing network of consideration consists of N queueing systems in which customers departing node i immediately enters node i+1. It is assumed that once a customer enters the first node, he must extract service at each node before departing the network and reentering the arriving customer population. The arrival rate for node i+1 is therefore the departure rate for node i. It is further assumed that the system as a whole is at equilibrium, so, by the topology, it follows that

$$\frac{1}{a_1^{(i)}} = \frac{1 - q(K_i = 0)}{s_1^{(i)}} = \frac{1}{a_1^{(i+1)}} \tag{24}$$

i.e., the average arrival rate to all nodes is identical and therefore must be equal to the arrival rate to the first node; $1/a_1^{(1)}$. Hence, subject to (24), equilibrium, and the expected number of customers in each node, (22)-(23), is the maximum entropy approximation to the distribution of the number of customers in a network of tandem queues where ρ_i has the particular form

$$\rho_i = \frac{s_1^{(i)}}{a_1^{(1)}} .$$

Although the interconnection of the tandem queues can be addressed by (24), the <u>form</u> of the approximation (22) is generally incorrect. The only known tandem network with a product form distribution for the number of customers in the network is one with Poisson arrivals to the first node and where customer service times in each node are distributed according to an exponential distribution.

Moreover, the service times for a specific customer are independent from node to node. The product form follows directly from Burke's Theorem [2], and it turns out that each node is an M|M|1 queue [2]. Thus, in order to obtain a maximum entropy approximation to the tandem network which is not of the product form, it is necessary to repose the inference problem and obtain the approximation subject to "dependent" information. This notion of dependent information is now made precise.

Maximum Entropy Approximation Subject to Joint Information.

Subject to information only on the marginal distribution, the maximum entropy approximation to a joint probability distribution is of the product form [17]. Conversely, if there is any hope of capturing the possible dependence of random variables, it is necessary to use constraints which give information on two or more random variables simultaneously, i.e.,

$$\sum_{K_1,K_2,\ldots,K_N} f(K_1,K_2,\ldots,K_N)q(K_1,K_2,\ldots,K_N) = \langle f(K_1,K_2,\ldots,K_N)\rangle \ .$$

Constraints of this form are on the joint distribution and will be termed joint constraints. A special case of interest occurs when the function of the random variables, $f(K_1,K_2,\ldots,K_N)$ is of the form

$$f(K_1,K_2,\ldots,K_N) = K_1^i\ K_2^j\ \ldots\ K_N^k$$

where i, j, and k are integers, then the joint constraint is a joint moment of the true distribution.

There is nothing in the maximum entropy formalism which disallows obtaining the maximum entropy solution subject to constraints only on the joint distribution. Under the assumption that the problem is well-defined, it turns out that all marginal distributions are identical. It appears, however, that in general, nothing can be stated about whether the approximation to the joint distribution is of the product form.

The relative-entropy approximation subject to constraints on <u>both</u> the marginal and joint distribution can still be of the product form. Under the assumption that the initial value distribution is of the product form, necessary and sufficient conditions are presented for the special case of moment constraints. The proof of the following proposition is given in the Appendix.

<u>Proposition 2</u>: The relative-entropy approximation to $q^{+}(K_1,K_2,\ldots,K_N)$ subject to both marginal and joint moment constraints is of the product form if and only if each joint constraint is equal to the product of marginal constraints satisfied by the approximation.

If the random variables K_i are correlated, then joint constraints can be used to force probabilistic dependence of the approximation. In the present application where the random variables represent quantities of individual systems, joint constraints thereby provide a means for the modeling and eventual analysis of system interaction.

As a special application of the preceding result, consider the maximum entropy approximation to the joint distribution $q^{+}(K_1,K_2)$ subject to normalization and the <u>covariance</u> of the random variables K_1 and K_2, i.e.,

$$\mathrm{Cov}(K_1K_2) := \sum_{K_1,K_2} (K_1 - \langle K_1\rangle)(K_2 - \langle K_2\rangle)q(K_1,K_2)$$

This is not a valid constraint as the covariance cannot be written as the expected value of a single function of the random variables. However, if the problem is posed as the determination of the maximum entropy distribution subject to $\langle K_1\rangle$, $\langle K_2\rangle$, and $\langle K_1K_2\rangle$, the problem is generally well-defined and the solution will automatically satisfy the covariance, $\mathrm{Cov}(K_1K_2)$. Noting that

$$\mathrm{Cov}(K_1K_2\rangle = \langle K_1K_2\rangle - \langle K_1\rangle\langle K_2\rangle$$

and under the assumption of zero covariance, the joint constraint is equal to the product of marginal constraints. Hence, subject to constraints only on the means and the first joint moment, the maximum entropy approximation is of the product form if and only if the random variables are uncorrelated.

Maximum Entropy Approximation to Two Queues in Tandem.

Using the preceding ideas, the maximum entropy approximation to the number of customers in a two-node tandem queueing network subject to both marginal and joint constraints is now presented. Two separate approximations to $q^+(K_1,K_2)$ are considered.

The first approximation is obtained subject to normalization, and the independent information: $\langle K_1\rangle$, ρ_1, and $\langle K_2\rangle$, ρ_2 where $\rho_i = s_i^{(i)}/a_i^{(1)}$. As already noted, the maximum entropy solution is a special case of Proposition 1. Let $q_1(K_1,K_2)$ denote this approximation which is given by

$$q_1(K_1,K_2) = q_1(K_1)q_1(K_2)$$

where

$$q_1(K_i = k_i) = \begin{cases} 1-\rho_i & k_i = 0 \\ \left(\dfrac{\rho_i^2}{\langle K_i\rangle - \rho_i}\right)\left(\dfrac{\langle K_i\rangle - \rho_i}{\langle K_i\rangle}\right)^{k_i} & k_i > 0 \end{cases} \quad i = 1,2 \tag{25}$$

The second approximation is obtained subject to the same constraints on $q_1(K_1,K_2)$ and the first joint moment, $\langle K_1K_2\rangle$. The form of the approximation denoted, $q_2(K_1,K_2)$, is given by

$$q_2(K_1,K_2) = \exp(-\beta_0 - \beta_1K_1 - \beta_2I(K_1) - \beta_3K_2 - \beta_4I(K_2) - \beta_5K_1K_2). \tag{26}$$

In this case, no analytic form could be found and, for the forthcoming numerical results, (26) will be estimated by the APL function of Johnson [18], which computes estimates of relative-entropy approximations given an arbitrary constraint matrix. These approximations will be compared in the following section for a variety of two-node queueing networks.

V. NUMERICAL RESULTS

As far as can be determined, there are essentially no known distributions for the number of customers in a network of tandem queues except for a tandem network of $M|M|1$ queues. In order to compare the approxiations, a simulation of two queues in tandem is performed using an APL function which computes the joint distribution. The simulated distribution will be denoted $q^s(K_1,K_2)$.

To verify the accuracy of the simulation, a chi-square goodness of fit test [20] was performed at the five percent level for the $M|M|1 \rightarrow M|1$ system. After performing the test, it was determined that the null hypothesis; the distribution observed is the joint distribution for the number of customers is a Jacksonian tandem network [2] of $M|M|1$ queues could not be rejected.

Two different networks were simulated, one in which customers require service according to the same general distribution in each node but have independent service times and one in which a customer has the identical service time in both nodes. One reason for considering these networks is that for the simulation distributions, $Cov(K_1,K_2)$ was generally larger than for systems with independent, identically distributed service times. In both networks, customers arrive to the first node according to a Poisson process, and the first node is therefore an $M|G|1$ queue. A comprehensive analysis of the network where customers have identical service times in both nodes under the assumption of Poisson arrivals to the first queue is given in Boxma [21], the particular case where the first node is an $M|M|1$ queue is considered in Pinedo and Wolff [22].

As stated, the aproximations (25)-(26) are for tandem networks with general interarrival and service distributions. The forthcoming examples are restricted, however, to systems with Poisson arrivals to the first system. There are two reasons for this: one is the aforementioned interest of such networks; the other reason is that these results serve to extend examples presented by Shore [12] for single $M|G|1$ systems. Moreover, in the first three examples, the departure process from the first node is not Poisson and so the second system is truly a $G|G|1$ queue, and thus the numerical results give some insight on maximum entropy approximations to general queueing systems.

Numerical results are presented in terms of the joint distribution of the simulated distribution and the two maximum entropy approximations. Several examples are given based on various service distributions. The statistics $\langle K_1\rangle$, $\langle K_2\rangle$, and $\langle K_1K_2\rangle$ are computed from $q^s(K_1,K_2)$, the constraints, $1 - \rho_1$ and $1 - \rho_2$ are derived from $q^+(K_1,K_2)$. The approximations are compared via the following measures: the relative entropy between $q_i(K_1,K_2)$ and $q^s(K_1,K_2)$

$$H[q_i,q^s] = \sum_{K_1,K_2} q_i(K_1,K_2)\log\frac{q_i(K_1,K_2)}{q^s(K_1,K_1)} \qquad i = 1,2$$

the sum of the square of the pointwise difference between $q_i(K_1,K_2)$ and $q^s(K_1,K_2)$,

$$\sum_{K_1,K_2} (q_i(K_1,K_2) - q^s(K_1,K_2))^2 \qquad i = 1,2$$

and the maximum absolute pointwise difference between $q_i(K_1,K_2)$ and $q^s(K_1,K_2)$

$$\max_{(K_1,K_2)} |q_i(K_1,K_2) - q^s(K_1,K_2)| \qquad i = 1,2$$

A "small" value for each measure implies the distributions are "close." In the case where $q_i(K_1 = k_1, K_2 = k_2)$ is close to $q^s(K_1 = k_1, K_2 = k_2)$ for each pair k_1, k_2, the relative-entropy, $H[q_i, q^+]$, can be considered the average percentage of relative difference between the distributions. This interpretation is valid, for example, when the maximum absolute pointwise difference is a very small value and the ratio $q_i(K_1,K_2)/q^s(K_1,K_2)$ is thereby close to one. It then follows that

$$\sum q_i(K_1,K_2)\log \frac{q_i(K_1,K_2)}{q^s(K_1,K_2)} \approx \sum q_i(K_1,K_2)\left(\frac{q_i(K_1,K_2)}{q^s(K_1,K_2)} - 1\right) .$$

Example 1: $M|H_2|1 \rightarrow H_2|1$

In this first example, customers arrive to the first node according to a Poisson process with rate $1/a_1^{(1)} = 1$, then proceed to a service area consisting of two parallel servers. The service distribution is the hyperexponential distribution. The first node is therefore the $M|H_2|1$ system. The service density, $b(t)$, is given by

$$b(t) = \frac{1}{4}\exp(-t) + \frac{3}{2}\exp(-2t) \tag{27}$$

from which it follows that $s_1^{(1)} = 5/8$. Hence, $\rho_1 = 5/8$. By (24), the arrival rate to the second node is also equal to one, even though the departure process from the first node is not Poisson. The customers receive service in the second node with density given by (27), but each customer has a service time length independent of his service time in the first system. Regardless of this independence condition, the utilization factor of the second system is given by $\rho_2 = 5/8$. The simulation was run for 5,016 customers, and from $q^s(K_1,K_2)$, the following statistics were computed: $\langle K_1\rangle = 1.77$, $\langle K_2\rangle = 1.83$, and $\langle K_1K_2\rangle = 3.14$. From the Pollaczeck-Khinchin mean value formula [2] (cf.

page 187), $\langle K_1 \rangle = 1.79$ when computed from the true distribution, $q^+(K_1)$, so the simulation value is extremely close. Note that $\mathrm{Cov}(K_1,K_2) = -0.1$; thus, the simulated distribution is only slightly correlated.

Example 2: $M|H_2|1 \to I|1$

All discussion of example 1 applies to this second example, except now customers have identical service times in each node. The service density in node one is again given by (27), so the first system is an $M|H_2|1$ queue and the network will be denoted by $M|H_2|1 \to I|1$ where "I" indicates identical service times. From example 1, it follows that $\rho_1 = 5/8 = \rho_2$. The simulation was run for 2,536 customers, and the following statistics were computed: $\langle K_1 \rangle = 1.70 = \langle K_2 \rangle$ and $\langle K_1K_2 \rangle = 4.1$. Hence, $\mathrm{Cov}(K_1,K_2) = 1.22$, and customers having identical service times in each node significantly increases the correlation of the random variables K_1 and K_2 as expected.

Example 3: $M|D|1 \to I|1$

In the third example, customers again arrive to the first node according to a Poisson process with rate equal to one. Each customer then demands a fixed amount of service of length equal to 0.5. The first node is therefore an $M|D|1$ system with service time density given by

$$b(t) = \delta(t - 0.5) \tag{28}$$

where δ is the usual Dirac delta function. From (28), it is clear that $s_1 = 0.5$, and by a similar argument as in the preceding examples, the utilization for both systems is equal to 0.5, i.e., $\rho_1 = 0.5 = \rho_2$. The simulation was run for 1,024 customers, and the following statistics were computed: $\langle K_1 \rangle = 0.765$, $\langle K_2 \rangle = 0.512$, and $\langle K_1K_2 \rangle = 0.75$. From the Pollaczeck-Khinchin formula $\langle K_1 \rangle = 0.75$, when computed for the true distribution, so, again, the simulation statistic is very close. For the simulation distribution, $Cov(K_1,K_2) = 0.13$, so, as in example 1, the random variables are essentially uncorrelated. The small correlation likely results from the light load on the first system; this is only conjecture, however, and more investigation is necessary to establish any connection between "loading" and correlation.

Example 4: $M|M|1 \to I|1$

In the fourth example, customers arrive to the first node with rate equal to 0.8 and demand service according to a negative exponential distribution. Hence, the first node is an $M|M|1$ queue. The service density is given by

$$b(t) = \exp(-t)$$

so $s_1^{(1)} = 1$. Each customer has identical service times in each node, and thus this network differs from a tandem network of $M|M|1$ queues discussed in connection with Jackson's and Burke's theorems. The utilization factor for each node is given by $\rho_1 = 0.8 = \rho_2$. The simulation was run for 4,840 customers, and the following statistics were computed: $\langle K_1 \rangle = 3.36$, $\langle K_2 \rangle = 2{,}83$, and $\langle K_1K_2 \rangle = 13.4$. From the Pollaczeck-Khinchin formula, $\langle K_1 \rangle = 4$ when computed from the true distribution. For the simulation distribution $Cov(K_1,K_2) = 3.9$, which is the largest correlation among the examples.

For each example, the results of the comparisons are given in Tables 1-4.

Due to the generally small probabilities of any given state, both approximations $q_1(K_1,K_2)$ and $q_2(K_1,K_2)$ are "close" to the simulated distribution under the measures of maximum pointwise difference and the sum of the square of pointwise difference. Hence, the relative-entropy measure (average percentage of relative difference) is the most revealing. By the relative-entropy measure, both approximations in examples one and two differ from the simulation by less than one percent. The addition of the joint constraint results in a 2.4 percent

Example 1: $q'(K_1,K_2)$
Simulation Distribution for
$M|H_2|1 \rightarrow H_2$ Tandem Network

		K_2					
		0	1	2	3	4	5
	0	.1390	.0861	.0522	.0364	.0216	.0144
	1	.0831	.0466	.0316	.0184	.0130	.0096
	2	.0476	.0311	.0163	.0114	.0087	.0055
$K1$	3	.0340	.0020	.0150	.0087	.0057	.0035
	4	.0261	.0132	.0095	.0053	.0041	.0021
	5	.0165	.0078	.0047	.0027	.0026	.0014

$q_1(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 1.77, <K_2> = 1.83.$

		K_2					
		0	1	2	3	4	5
	0	.1410	.0799	.0527	.0347	.0229	.0151
	1	.0826	.0469	.0309	.0204	.0135	.0088
	2	.0545	.0304	.0200	.0132	.0087	.0057
$K1$	3	.0346	.0197	.0130	.0085	.0056	.0037
	4	.0224	.0127	.0084	.0055	.0036	.0024
	5	.0145	.0083	.0054	.0036	.0024	.0016

$q_2(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 1.77, <K_2> = 1.83, <K_1K_2> = 3.14$

		K_2					
		0	1	2	3	4	5
	0	.1390	.0795	.0527	.0350	.0232	.0154
	1	.0822	.0469	.0310	.0205	.0135	.0090
	2	.0536	.0304	.0020	.0132	.0087	.0057
$K1$	3	.0349	.0198	.0130	.0085	.0056	.0037
	4	.0228	.0128	.0084	.0055	.0036	.0023
	5	.0148	.0083	.0054	.0035	.0023	.0015

Example 2: $q^s(K_1,K_2)$
Simulation Distribution for
M|H$_2$|1 → I|1 Tandem Network

K_2

		0	1	2	3	4	5
	0	.2020	.0764	.0387	.0249	.0131	.0083
	1	.1020	.0490	.0257	.0209	.0120	.0083
	2	.0418	.0265	.0225	.0173	.0115	.0086
$K1$	3	.0199	.0197	.0166	.0141	.0085	.0046
	4	.0095	.0142	.0164	.0122	.0056	.0047
	5	.0031	.0081	.0070	.0038	.0036	.0038

$q_1(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 1.70, <K_2> = 1.70.$

K_2

		0	1	2	3	4	5
	0	.1410	.0855	.0544	.0346	.0220	.0140
	1	.0855	.0520	.0331	.0210	.0134	.0085
	2	.0054	.0331	.0210	.0134	.0085	.0054
$K1$	3	.0346	.0210	.0134	.0085	.0054	.0034
	4	.0220	.0134	.0085	.0054	.0034	.0022
	5	.0140	.0085	.0054	.0034	.0022	.0014

$q_2(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 1.70, <K_2> = 1.70, <K_1K_2> = 4.1$

K_2

		0	1	2	3	4	5
	0	.1560	.0925	.0534	.0303	.0178	.0103
	1	.0925	.0569	.0342	.0205	.0123	.0074
	2	.0534	.0342	.0214	.0134	.0084	.0052
$K1$	3	.0308	.0205	.0134	.0087	.0057	.0037
	4	.0178	.0123	.0084	.0057	.0038	.0026
	5	.0103	.0074	.0052	.0037	.0026	.0018

Example 3: $q^*(K_1,K_2)$
Simulation Distribution for
M|D|1 → I|1 Tandem Network

		K_2					
		0	1	2	3	4	5
$K1$	0	.2930	.1950	0	0	0	0
	1	.1530	.1780	0	0	0	0
	2	.0363	.0916	0	0	0	0
	3	.0049	.0345	0	0	0	0
	4	.0008	.0101	0	0	0	0
	5	0	.0021	0	0	0	0

$q_1(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = .765, <K_2> = .512,$

		K_2					
		0	1	2	3	4	5
$K1$	0	.2500	.2440	.0057	1.34×10^{-4}	3.14×10^{-6}	7.37×10^{-8}
	1	.1630	.1600	.0037	8.77×10^{-5}	2.05×10^{-6}	4.81×10^{-8}
	2	.0056	.0055	.0013	3.04×10^{-5}	7.12×10^{-7}	1.67×10^{-8}
	3	.0196	.0019	.0004	1.05×10^{-5}	2.47×10^{-7}	5.78×10^{-9}
	4	.0067	.0066	.0002	3.64×10^{-6}	8.84×10^{-8}	8.54×10^{-8}
	5	.0023	.0023	.0005	1.26×10^{-6}	2.96×10^{-8}	$.6930\times10^{-9}$

$q_2(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = .765, <K_2> = .512, <K_1K_2> = .521$

		K_2					
		0	1	2	3	4	5
$K1$	0	.2820	.2150	.0025	2.95×10^{-5}	3.46×10^{-7}	4.05×10^{-9}
	1	.1580	.1700	.0028	4.94×10^{-5}	7.96×10^{-7}	1.26×10^{-8}
	2	.0043	.0065	.0015	3.54×10^{-5}	8.25×10^{-7}	1.92×10^{-8}
	3	.0118	.0025	.0008	2.70×10^{-5}	8.98×10^{-7}	2.92×10^{-8}
	4	.0032	.0096	.0004	2.07×10^{-5}	9.56×10^{-7}	4.43×10^{-8}
	5	.0009	.0037	.0002	1.58×10^{-5}	9.56×10^{-7}	6.72×10^{-8}

Example 4: $q^s(K_1,K_2)$
Simulation Distribution for
M|M|1 → I|1 Tandem Network

K_1 \ K_2	0	1	2	3	4	5
0	.0982	.0382	.0253	.0175	.0099	.0077
1	.0527	.0318	.0247	.0167	.0125	.0083
2	.0290	.0242	.0261	.0186	.0152	.0079
3	.0169	.0189	.0218	.0214	.0152	.0076
4	.0082	.0114	.0175	.0217	.0107	.0095
5	.0040	.0072	.0116	.0179	.0090	.0080

$q_1(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 3.36, <K_2> = 2.83,$

K_1 \ K_2	0	1	2	3	4	5
0	.0400	.0451	.0324	.0233	.0167	.0120
1	.0378	.0427	.0306	.0220	.0158	.0113
2	.0289	.0326	.0234	.0168	.0121	.0090
3	.0221	.0249	.0179	.0128	.0092	.0066
4	.0169	.0190	.0137	.0098	.0070	.0050
5	.0129	.0145	.0104	.0075	.0053	.0039

$q_2(K_1,K_2)$
Maximum Entropy Distribution Subject to Utilization,
$<K_1> = 3.36, <K_2> = 2.83, <K_1K_2> = 13.4$

K_1 \ K_2	0	1	2	3	4	5
0	.0467	.0526	.0346	.0227	.0149	.0097
1	.0443	.0509	.0341	.0228	.0153	.0102
2	.0315	.0369	.0251	.0172	.0117	.0080
3	.0224	.0267	.0186	.0129	.0090	.0062
4	.0159	.0193	.0137	.0097	.0070	.0049
5	.0113	.0140	.0101	.0073	.0052	.0038

Table 1: Comparison of Approximations for Example 1

$M|H_2|1 \rightarrow H_2$ Tandem Network

Constraints: $<K_1> = 1.77$, $<K_2> = 1.83$, $<K_1K_2> = 3.14$			
	$H[q_i, q^*]$	$\sum(q_i(K_1,K_2) - q^*(K_1,K_2))^2$,	$\max \mid q_i(K_1,K_2) - q^*(K_1,K_2) \mid$
q_1	6.24×10^{-3}	1.43×10^{-4}	6.20×10^{-3}
q_2	1.60×10^{-3}	1.44×10^{-4}	6.60×10^{-3}

Table 2: Comparison of Approximations for Example 2

$M|H_2|1 \rightarrow I|1$ Tandem Network

Constraints: $<K_1> = 1.77$, $<K_2> = 1.83$, $<K_1K_2> = 14$			
	$H[q_i, q^*]$	$\sum(q_i(K_1,K_2) - q^*(K_1,K_2))^2$,	$\max \mid q_i(K_1,K_2) - q^*(K_1,K_2) \mid$
q_1	1.01×10^{-1}	5.76×10^{-3}	6.17×10^{-2}
q_2	7.38×10^{-2}	3.60×10^{-3}	4.60×10^{-2}

Table 3: Comparison of Approximations for Example 3

M|D|1 → I|1 Tandem Network

Constraints: $<K_1> = 1.77$, $<K_2> = 1.83$, $<K_1K_2> = 14$			
	$H[q_i, q^*]$	$\sum(q_i(K_1,K_2) - q^*(K_1,K_2))^2$,	$\max \mid q_i(K_1,K_2) - q^*(K_1,K_2) \mid$
q_1	5.81×10^{-2}	7.05×10^{-3}	4.94×10^{-2}
q_2	2.39×10^{-2}	1.53×10^{-3}	2.63×10^{-2}

Table 4: Comparison of Approximations for Example 4

M|M|1 → I|1 Tandem Network

Constraints: $<K_1> = 3.36$, $<K_2> = 2.83$, $<K_1K_2> = 13.14$			
	$H[q_i, q^*]$	$\sum(q_i(K_1,K_2) - q^*(K_1,K_2))^2$,	$\max \mid q_i(K_1,K_2) - q^*(K_1,K_2) \mid$
q_1	2.33×10^{-1}	5.56×10^{-3}	5.82×10^{-2}
q_2	1.91×10^{-1}	5.03×10^{-3}	5.15×10^{-2}

increase in accuracy in example 3 and a four percent increase in example 4, the most correlated among the examples.

VI. DISCUSSION

In this paper, information theoretic analysis was applied to obtain an explicit maximum entropy distribution for the number of customers in a G|G|1 system to "minimal" information. The approach was then extended to multiple simultaneous G|G|1 systems where the approximation was obtained subject to independent information resulting in an independent (product form) distribution. The result is of interest primarily because it sparked a general discussion of product form approximations where it was determined that dependency of random variables is captured only by imposing constraints on both the marginal and joint distributions. Moreover, although not directly applicble, the result serves to stimulate interest in the use of this inference technique for networks of queueing systems.

The difficulty in applying maximum entropy analysis to a network of queues lies in capturing the interconnection of the systems. As a first step, a very simple network, queues in tandem, was considered. For this network, the interconnection of the queues was captured by the equilibrium constraint, which led to an explicit formula for the number of customers in the tandem system subject to the mean number of customers in each system. A secondary problem, that of the form of the approximation, was addressed by adding an additional constraint on the first joint moment. This result is interesting because it identifies minimal information for a joint system that addresses the possible dependency of the random variables.

The maximum entropy approximation could prove useful in applications in which knowledge of the full probability distribution is necessary. One example is queueing network control. By strainghtforward techniques (measurement or prediction), it is common to know (or have unbiased estimates of) average flows and average service rates and thereby average queue sizes. Higher order moments such as variances are needed, however, to apply modern distributed routing algorithms that utilize estimates of flow derivatives. The more sophisticated versions of these algorithms require the entire distribution to predict the optimal control function. It should be noted that the maximum entropy approximation has been shown to be continuous as a function of the constraints [23]. Thus, if only good estimates of moments are available (for example, through repeated sampling), the approximation will be close to the approximation subject to the true (unknown) moments.

The approximation could also be used as an additional tool in conventional operational analysis [24-26]. Indeed, in the present approach, standard performance measures are utilized as constraints for the distribution. In any case, if the customer population approximation is close to the true distribution in some precise sense, then it can provide a more extensive analysis of system interaction. For example, knowledge of the joint customer distribution readily provides the distribution for the number of customers in the network as well as marginal and conditional customer distributions. From the conditional distributions, conditional moments can be computed and, under appropriate assumptions, estimates of conditional expected delays and sojourn (response) times become available.

Tandem queues were selected to introduce the approach to networks. Because of the special topological structure, much is known [21-22,27] about such systems. The focus here is not the particular study of tandem queues but an initial consideration of the issues involving information theoretic analysis for queueing networks. It seems clear that the approach applied to the tandem network can be extended to more general networks through imposed marginal and joint constraints along with analogous equilibrium constraints that reflect the topological structure.

To assess the quality of the two maximum entropy distributions as approximations to the customer distribution of a tandem network, several different examples were presented. The particular case of a two-node network was considered where customers extracted service according to a variety of distributions and had either identical or independent and identically distributed service times in each node. Based on these examples, it appears that minimal information results in "good" approximations. In particular, for small correlation, it appears that the product form approximation is quite accurate. This approximation is appealing as it has closed form solution. When, however, the random variables are strongly correlated, the joint constraint should become more important.

A question that arises in comparing probability distributions is what constitutes a good measure of closeness. For distributions such that the probability of any particular state is small, for example, if the mass is spread "uniformly" over a large state space, then the maximum absolute pointwise difference of the sum or the square of the pointwise difference of the two distributions is expected to be small. In this case, it appears that the relative-entropy between the two distributions is the best among the three proposed measures as it relates the two on a "micro" level. Emphasizing this measure for the tandem examples, it is reasonable to say that the approximations, in particular those obtained with the joint constraint, are all close. Of course, far more analytical and numerical work is needed to say anything precise about the approximations, but as an inference technique for single system queues and possibly networks of queues, the results are promising.

VII. ACKNOWLEDGEMENT

The authors thank J. Shore for reading an earlier version of this paper and providing many helpful suggestions, and R. Johnson, A. Makowski, and B. Hughes for many helpful and stimulating discussions.

APPENDIX

Proof of Proposition 2.

For notational simplicity, let

$$K: = (K_1, K_2, \ldots, K_N) .$$

By system independence [17], if the initial value distribution is of the product form, the relative-entropy approximation to q (K) subject only to constraints on the marginal distribution, i.e.,

$$\sum_K f_j(K)q(K) = \sum_K f_j(K_i)q(K) = \langle f_j(K_i)\rangle \qquad \begin{matrix} i = 1,2,\ldots,L \\ j = 1,2,\ldots,M \end{matrix}$$

is of the product form,

$$q(K) = \prod_{i=1}^{N} q(K_i) . \tag{29}$$

Now, (29) also satisfies the constrained problem of the relative-entropy distribution subject to

$$\sum_K f_j(K_i)q(K) = \langle f_j(K_i)\rangle \qquad j = 1,2,\ldots,M \tag{30}$$

and

$$\sum_K f_\ell(K)q(K) = \langle f_\ell(K)\rangle = \prod_{i=1}^{N} \langle f_{\ell i}(K_i)\rangle \qquad \ell = M+1, M+2, \ldots, N \tag{31}$$

where $\langle f_{\ell i}(K_i)\rangle$ is an element of the set of marginal constraints given by (30). The relative-entropy solution subject to (30)-(31) and normalization is given by

$$q(K) = p(K_1)p(K_2) \ldots p(K_N)\exp(-\beta_0 - \sum_{i=1}^{M} \beta_i f_i - \sum_{\ell=M+1}^{N} \beta_\ell f_\ell) .$$

However, by uniqueness of the relative-entropy solution, $\beta_\ell = 0$, $M + 1 \leq \ell \leq N$

and the product form solution follows.

Conversely, if the relative-entropy approximation is of the product form, and every joint constraint is a joint moment satisfied by the approximation, then it follows that

$$\sum_K f(K)q(K) = \langle f(K)\rangle = \sum_K K_1^i \, K_2^j \ldots K_N^k q(K)$$

$$= \sum_{K_1} K_1^i q(K_1) \sum_{K_2} K_2^j q(K_2) \ldots \sum_{K_N} K_N^k q(K_N) = \langle K_1^i\rangle \, \langle K_2^j\rangle \ldots \langle K_N^k\rangle \, .$$

Hence, each joint constraint is equal to the product of marginal constraints satisfied by the marginal relative entropy distributions. Q.E.D.

REFERENCES

1. J. W. Cohen, The Single Server Queue, North Holland, Amsterdam, 1982.

2. L. Kleinrock, Queueing Systems, Volume 1: Theory, John Wiley and Sons, New York, 1975.

3. R. B. Cooper, Introduction to Queueing Theory, North Holland, New York, 1981.

4. R. Syski, Introduction to Congestion Theory in Telephone Systems, Oliver and Boyd, London, 1960.

5. L. Kleinrock, Communication Nets, McGraw-Hill, New York, 1964.

6. A. S. Tanenbaum, Computer Networks, Prentice-Hall, New Jersey, 1981.

7. K. S. Trivedi, Probability and Statistics with Reliability, Queueing and Computer Science Applications, Prentice-Hall, New Jersey, 1982.

8. J. E. Shore, "Derivation of equilibrium and time-dependent solutions of queueing systems using entropy maximization," Proceedings 1978 National Computer Conference, pp. 483-487, AFIPS, 1978.

9. J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," IEEE Trans. Inform. Theory, vol. IT-26, pp. 26-37, Jan. 1980. (See also comments and corrections in IEEE Trans. Inform. Theory, IT-29, Nov. 1983.)

10. V. E. Benes, Mathematical Theory of Connecting Networks and Telephone Traffic, Academic Press, New York, 1965.

11. A. E. Ferdinand, "A Statistical Mechanics Approach to Systems Analysis," IBM J. Res. Dev., pp. 539-547, 1970.

12. J. E. Shore, "Information Theoretic Approximations for M|G|1 and G|G|1 Queueing Systems," Acta Informatica, vol. 17, pp. 43-61, 1982.

13. M. A. El-Affendi and D. D. Kouvatos, "A Maximum Entropy Analysis of the M|G|1 and G|M|1 Queueing Systems at Equilibrium," Acta Informatica, vol. 19, pp. 339-355, 1983.

14. E. T. Jaynes, "Information theory and statistical mechanics I," Phys. Rev., vol. 106, pp. 620-630, 1957.

15. S. Kullback, Information Theory and Statistics, Dover, New York, 1969, Wiley, New York, 1959.

16. I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," Ann. Prob., vol. 3, pp. 146-158, 1975.

17. J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," IEEE Trans. Iform. Theory, vol IT-27, pp. 472-482, July 1981.

18. R. W. Johnson, "Determining probability distributions by maximum entropy and minimum cross-entropy," APL79 Conference Proceedings, pp. 24-29, ACM 0-89791-005, May 1979.

19. J. E. Shore, "Cross-Entropy Minimization Given Fully Decomposable Subset and Aggregate Constraints," IEEE Trans. on Information Theory, Vol IT-28, No. 6, pp. 956-961, Nov. 1982.

20. W. Mendenhall and R. L. Scheaffer, Mathematical Statistics with Applica--tions, Duxbury Press, North Scituate, Mass., 1974.

21. O. J. Boxma, "On a Tandem Queueing Model with Identical Service Times at Both Counters I," Adv. Appl. Prob., vol. 11, pp. 616-643, 1979.

22. M. Pinedo and R. W. Wolff, "A Comparison Between Tandem Queues with Dependent and Independent Service Times," Operations Research, Vol. 30, No. 3, pp. 464-479, 1982.

23. J. L. Cantor, "Information-Theoretic Analyses for a Multi-Server Queueing System at Equilibrium with Application to Queues in Tandem," M.SC. Thesis, Electrical Engineering Dept., Univ. of Maryland, College Park, Maryland (1984).

24. P. J. Denning, F. P. Buzen, "The Operational Analysis of Queueing Network Models," Computing Surveys 10, 225-261 (1978).

25. J. Zahorjan, "The Approximate Analysis of Large Queueing Network Models," Tech. Rep. CSR6-122 (Ph.D. Thesis), Univ. of Toronto, Toronto, Ont. (1980).

26. W. Whitt, "The Queueing Network Analyzer," B.S.T.J., Vol. 62, No. 9, pp. 2279-2815 (1983).

27. S. V. Tembe, R. W. Wolff, "The Optimal Order of Service in Tandem Queues," Opns. Res. 22, pp. 824-832 (1974).