A. E. Ferdinand

A Statistical Mechanical Approach to Systems Analysis

Abstract: The maximum entropy principle is used as the criterion for calculating the equilibrium state probabilities of a queuing or network system in which service rates are exponentially distributed. A configuration-independent partition function is given as the solution to this network problem; from this function the important properties of the system may be derived. Simple and well known examples are used to illustrate the method. A phenomenon similar to the phase transition of statistical mechanics is observed in a queuing model.

Introduction

The computer systems analyst and the traffic engineer are often faced with the analysis of complex systems that are difficult to characterize uniquely and can best be described by statistical methods. The physicist also deals with complex systems, which for the most part comprise many particles, in various states, that may or may not interact; statistical techniques have led to very satisfactory results in his case. In this paper the methods of statistical mechanics, usually applied by the physicist, are shown to be useful in the solution of queuing and network problems.

The basic mathematical expression in statistical mechanics is the partition function, which provides a link between the mechanical properties of a system and its thermodynamic properties. The partition function tells how, in the equilibrium distribution, the system is partitioned or divided among different energy levels. The important measurable properties of the statistical mechanical system are expressed in terms of derivatives of the logarithm of the partition function.

A function similar to the partition function of statistical mechanics can be developed for queuing and network systems, and important averages can be obtained from this function and its logarithmic derivatives. The statistical mechanical model of a queuing or network system is based on the single postulate that the equilibrium probabilities of the states of a network or queuing system maximize the entropy functional of that system. A high value of the entropy functional is a measure of a low degree of information. Thus one may say that the maximum entropy principle further postulates that an equilibrium distribution corresponds to a condition of maximum ignorance for a given average number of elements enqueued or distributed in the system.

The maximum entropy principle leads to a function that is similar in structure to the partition function of statistical mechanics and, for this reason, will also be called a partition function. In fact, partition functions are actually just multivariate generating functions.[1]

The general solution, for systems whose service rates are exponentially distributed, is configuration independent; that is, the solution is independent of the number of components (servers, terminals, channels, tasks, etc.) comprised by the system, and also independent of the spatial relationship of these components with one another. Futhermore, it provides the capability of addressing problems that are not easily solved by queuing theory techniques. For example, bidirectional queues, i.e., queues in which the elements can go in one of two directions after service, are readily addressed by the statistical mechanical approach. In the next section, the consequences of the maximum entropy principle are derived.

A similar approach has been taken by Beneš,[2] in which he demonstrates some of the underlying similarities between statistical mechanical systems and connecting networks.

The maximum entropy principle

In the approach usually taken in queuing and network problems, equilibrium in a system is assumed to correspond to some *particular* distribution of elements throughout the range of permissible states. This distribution is taken to be that which is most likely to occur for a very large number of elements, under assumptions with regard to state dependencies that lead to a Markov process as a model of the system. Based on those assumptions, an equation

The author is located at the IBM Systems Development Division Laboratory, Kingston, New York 12401.

is derived according to an invariance principle,[3] from which the state probabilities can be calculated.

In the approach to be described here, as in that of statistical mechanics, equilibrium is assumed to correspond not to some particular distribution, but to a condition of maximum entropy. The well known expression for entropy in terms of the probability vector is maximized to *obtain* an equilibrium distribution, and the equilibrium state probabilities can then be calculated according to a variational principle. The distribution is thus uniquely determined. The operative requirement is to defend the assumption that maximum entropy occurs at equilibrium in queuing systems and networks, and to establish the meaning of "entropy" and "energy" in the queuing/network context. It is shown that the probability vector obtained through the maximum entropy principle is the same as that obtained from a reversible Markov process in the same system.

The following lemma will first be proved:

Lemma:

Let S be a set of permissible states in a system, and $r_k(x)(k = 1, 2, \cdots, n)$ with $x \in S$ be a set of non-negative functions defined on S. Then the maximum of

$$H(q) = -\sum_{x \in S} q_x \ln q_x, \tag{1}$$

subject to the conditions

$$q_x \geq 0, \tag{2}$$

$$\sum_{x \in S} q_x = 1 \tag{3}$$

and

$$\sum_{x \in S} r_k(x) q_x = m_k \qquad (k = 1, 2, \cdots, n) \tag{4}$$

where the m_k are given positive numbers, is

$$H_{\max} = \ln Z(w_1, w_2, \cdots, w_n) - \sum_{k=1}^{n} m_k \ln w_k, \tag{5}$$

with

$$Z(y_1, y_2, \cdots, y_n) = \sum_{x \in S} \prod_{k=1}^{n} y_k^{r_k(x)} \tag{6}$$

Here w_k is the unique positive solution of

$$w_k \frac{d}{dw_k} [\ln Z(w_1, w_2, \cdots, w_n)] = m_k > 0. \tag{7}$$

This maximum, Eq. (5), is achieved by choosing

$$q_x = \left(\prod_{k=1}^{n} w_k^{r_k(x)} \right) / Z(w_1, w_2, \cdots, w_n)$$

$$= \exp \left[-(a+1) - \sum_{k=1}^{n} b_k r_k(x) \right], \tag{8}$$

where $a, b_1, b_2, \cdots, b_n$ are values of the Lagrange multipliers determined by

$$a = -1 + \ln Z(e^{-b_1}, e^{-b_2}, \cdots, e^{-b_n}); \tag{9}$$

$$m_j = \sum_{x \in S} r_j(x) \exp \left[-(a+1) - \sum_{k=1}^{n} b_k r_k(x) \right]; \tag{10}$$

$$w_k = \exp(-b_k). \tag{11}$$

In this Lemma one may consider $\mathbf{q} = [q_x]$ as a probability vector, in which case $H(\mathbf{q})$ is the entropy of the system. The variable y_k in $Z(y_1, y_2, \cdots, y_n)$ is related to a service rate in a system via the relation

$$y_k = v_k,$$

where v_k is the mean of the service rate (assumed to be exponentially distributed).

Proof:

With the Lagrange multipliers $a, b_1, b_2, \cdots, b_n$, define

$$h = -\sum_{x \in S} q_x \ln q_x - a \sum_{x \in S} q_x - \sum_{k=1}^{n} b_k \sum_{x \in S} r_k(x) q_x. \tag{12}$$

Differentiating with respect to q_x and setting the resulting derivatives to zero gives the equation

$$1 + a + \ln q_x + \sum_{k=1}^{n} b_k r_k(x) = 0 \qquad \text{for all } x \in S. \tag{13}$$

The Lagrange multipliers are determined by the conditions

$$\sum_{x \in S} q_x = 1; \tag{14}$$

$$\sum_{x \in S} r_k(x) q_x = m_k \qquad \text{for all } k. \tag{15}$$

Condition (14) gives

$$e^{-(a+1)} \sum_{x \in S} \prod_{k=1}^{n} \exp[-b_k r_k(x)] = 1, \tag{16}$$

whereas condition (15) gives

$$m_j = e^{-(a+1)} \sum_{x \in S} r_j(x) \prod_{k=1}^{n} \exp[-b_k r_k(x)] \tag{17}$$

$$= \left\{ \sum_{x \in S} r_j(x) \prod_{k=1}^{n} \exp[-b_k r_k(x)] \right\} \Big/ \left\{ \sum_{x \in S} \prod_{k=1}^{n} \exp[-b_k r_k(x)] \right\}. \tag{18}$$

By now defining

$$w_k = e^{-b_k}, \tag{19}$$

it is easily seen that w_k is a solution of

$$w_k \frac{d}{dw_k} [\ln Z(w_1, w_2, \cdots, w_n)] = m_k > 0. \tag{20}$$

But since

$$\frac{d^2}{db_i^2} \ln Z(e^{-b_1}, e^{-b_2}, \cdots, e^{-b_n}) > 0, \tag{21}$$

there is only one solution of Eq. (20), and that solution w_k is positive. Furthermore, since

$$\begin{aligned} \frac{\partial^2 H}{\partial q_x \, \partial q_y} &= 0 \qquad \text{if} \quad x \neq y; \\ &= -\frac{1}{q_x} \qquad \text{if} \quad x = y, \end{aligned} \tag{22}$$

the matrix of second derivatives of $H(q)$ is negative definite and therefore maximized.

In this Lemma, let

$$r_k(x) = r_k \qquad (k = 1, 2, \cdots, n) \tag{23}$$

be the number of elements enqueued in or distributed throughout the different parts of a queuing or network system in state x.

The following theorem about that system may now be stated:

Theorem:

Let $m_1, m_2, \cdots, m_n > 0$; furthermore, let

$$Z(y_1, y_2, \cdots, y_n) = \sum_{x \in S} \prod_{k=1}^{n} y_k^{r_k} \tag{24}$$

and y_k be the unique (positive) root of

$$m_k = y_k \frac{d}{dy_k} [\ln Z(y_1, y_2, \cdots, y_n)] \qquad (k = 1, 2, \cdots, n). \tag{25}$$

Then the maximum of

$$H(\mathbf{q}) = -\sum_{x \in S} q_x \ln q_x, \tag{26}$$

subject to the conditions that $\mathbf{q}$ is a probability vector over S (so that $H(\mathbf{q})$ is the entropy functional) and

$$\sum_{x \in S} r_k q_x = m_k, \tag{27}$$

is

$$H_{\max} = \ln Z(y_1, y_2, \cdots, y_n) - \sum_{k=1}^{n} m_k \ln y_k, \tag{28}$$

and is obtained by the vector $\mathbf{q}$ with components

$$q_x = \left(\prod_{k=1}^{n} y_k^{r_k} \right) \Big/ Z(y_1, y_2, \cdots, y_n). \tag{29}$$

The probability distribution q_x over S is precisely that which is determined uniquely by the maximum entropy principle. The function $Z(y_1, y_2, \cdots, y_n)$ is called the partition function of the queuing or network system, and a reader familiar with statistical mechanics will immediately recognize its similarity to that of statistical mechanics. The quantity m_k uniquely determines y_k and vice versa. The constraint introduced through the m_k in the derivation of the probability vector is analogous to the energy constraint used in statistical mechanics to derive the partition function of a statistical mechanical system. This energy constraint implies a dependence on a new variable, the temperature, which is of fundamental importance in describing a statistical mechanical system. The constraint imposed through the m_k in queuing or network problems similarly gives rise to parameters, analogous to temperature in statistical mechanics, that can be correspondingly used to describe the queuing or network system. If a basic identification is to be made between a queuing or network model and a statistical mechanical system, it will be through the variables y_k and temperature as

$$\ln y_k \propto 1/(k_{\mathrm{B}} T) \tag{30}$$

where k_{B} is the Boltzman constant and T is the temperature.

The analogy does not imply any physical equivalence between temperature (in thermodynamics) and the rate or availability of service (in a queuing system or network). Instead it implies that a change in the service rate variable for one element in a network has an effect on the relationship among elements that is mathematically analogous to that of a change in the temperature (or energy) variable for a particle in a statistical mechanical system. Since the temperature variable $k_{\mathrm{B}}T$ has the dimensions of energy in statistical mechanics, it seems appropriate to designate $\ln v_k$ as "energy" in queuing or network theory. It is, however, only for convenience in nomenclature and the analogy is only mathematical.

A reversible Markov process

In this section we follow the method of Beneš[2] and describe an ergodic reversible Markov process θ_t which takes values in the set S of states and has the property that the distribution over S is precisely the canonical distribution derived through the maximum entropy principle.

Let $x \in S$ be a state of the system and, furthermore, let the elements in the state x be of n types with r_k elements of type k, so that $\sum_{k=1}^{n} r_k$ is the total number of elements in the state x.

Let A_x and B_x be the sets of states adjacent to x, i.e., accessible from x by adding an element to those already in state x and removing an element from those in state x, respectively. The states A_x accessible by adding an element are said to lie "above" x; the states B_x accessible by removing an element are said to lie "below" x. The sets A_x and B_x are further subdivided into n subsets as follows:

A_{kx} is the set of states immediately above x and accessible from x by adding an element of kth type to those of that type in state x.

B_{kx} is the set of states immediately below x and accessible from x by removing an element of the kth type from those in state x. With these definitions, one may write

$$A_x = \bigcup_{k=1}^{n} A_{kx}; \tag{31}$$

$$B_x = \bigcup_{k=1}^{n} B_{kx}. \tag{32}$$

Let $|X|$ denote the number of elements in the set X. A process θ_t is defined as follows:

If $\theta_t = x$, then θ_* is moving to each $y \in A_{kx}$ at a rate $v_k > 0$, to each $y \in B_{kx}$ at a rate $v_k = 1$ and to any other state at a rate zero. The transition rate matrix $\mathbf{M}$ is, for this process, given by

$$\begin{aligned} m_{xy} &= -\sum_{k=1}^{n} [|B_{kx}| + v_k\,|A_{kx}|] \quad \text{for} \quad y = x, \\ &= 1 \quad \text{for} \quad y \in B_{kx} \quad \text{and all } k, \\ &= v_k \quad \text{for} \quad y \in A_{kx} \quad \text{and all } k, \quad \text{and} \\ &= 0 \quad \text{for} \quad y \notin A_x \cup B_x; \qquad y \neq x. \end{aligned} \tag{33}$$

In probabilistic terms, the transition rate matrix $\mathbf{M}$ can be interpreted as meaning that if $\theta_t = x$, then as $h \to 0$, (h being a small increment in time), there are conditional probabilities

$v_k h + o(h)$ that $\theta_{t+h} = y \in A_{kx}$ for all k;

$h + o(h)$ that $\theta_{t+h} = y \in B_{kx}$ for all k;

$1 - h \sum_{k=1}^{n} [|B_{kx}| + v_k\,|A_{kx}|] + o(h)$ that $\theta_{t+h} = x$;

and

$o(h)$ that $\theta_{t+h} = y \in A_x \cup B_x$; $y \neq x$.

The v_k may be considered as the mean of a holding or service rate that has a negative exponential distribution.

The statistical equilibrium equation of the process θ_t can be put in the vector form

$$\mathbf{Mq} = 0. \tag{34}$$

Alternatively, it may be written as

$$\left[\sum_{k=1}^{n} (|B_{kx}| + v_k\,|A_{kx}|)\right] q_x = \sum_{k=1}^{n} \Big(\sum_{y \in A_{kx}} q_y + \sum_{y \in B_{kx}} v_k q_y \Big) \quad \text{for all } x \in S. \tag{35}$$

This equation is satisfied by choosing

$$q_x \approx \prod_{k=1}^{n} v_k^{r_k}, \tag{36}$$

so that a normalized solution of the statistical equilibrium equation is

$$\begin{aligned} q_x &= \left(\prod_{k=1}^{n} v_k^{r_k}\right) \Big/ \sum_{x \in S} \prod_{k=1}^{n} v_k^{r_k} \\ &= \left(\prod_{k=1}^{n} v_k^{r_k}\right) \Big/ Z(v_1, v_2, \cdots, v_n). \end{aligned} \tag{37}$$

The canonical distribution of probability over S, which is the unique solution obtained through the maximum entropy principle, is also a solution of the statistical equilibrium equation. Furthermore, the components of the probability vector $\mathbf{q}$ satisfy the condition of reversibility, namely

$$q_x m_{xy} = q_y m_{yx}; \qquad x, y \in S, \tag{38}$$

which is analogous to the principle of detailed balance in statistical mechanics.

An alternative representation of the partition function

In its present form the partition function, $Z(v_1, v_2, \cdots, v_n)$, expressed as a summation over all permissible states, may become quite unmanageable when dealing with a complex system in which there are a vast number of states. To facilitate computation it is prudent to perform a partial summation and express the partition function in the more convenient form

$$Z(v_1, v_2, \cdots, v_n) = \sum_{r_1, r_2, \cdots, r_n} g(r_1, r_2, \cdots, r_n) \times \exp[-E(r_1, r_2, \cdots, r_n)], \tag{39}$$

where

$$\exp[-E(r_1, r_2, \cdots, r_n)] = \prod_{k=1}^{n} v_k^{r_k} \tag{40}$$

and $g(r_1, r_2, \cdots, r_n)$ is a combinatorial term that represents the total number of states in S that are characterized by the vector $(r_1, r_2, \cdots, r_n)$.

The probability $p(r_1, r_2, \cdots, r_n)$ of being in a state characterized by the vector $(r_1, r_2, \cdots, r_n)$ is given by

$$p(r_1, r_2, \cdots, r_n) = \frac{g(r_1, r_2, \cdots, r_n) \exp[-E(r_1, r_2, \cdots, r_n)]}{Z(v_1, v_2, \cdots, v_n)}. \tag{41}$$

The mean value $\langle r_k \rangle$ of variable r_k is again

$$\langle r_k \rangle = v_k \frac{d}{dv_k} \ln Z(v_1, v_2, \cdots, v_n). \tag{42}$$

The representation of the partition function in the form (39) leads to some most interesting possibilities, which to the author's knowledge have not been attempted in queuing or network theory. Observe that the term $\exp[-E(r_1, r_2, \cdots, r_n)]$, which is henceforth referred to as the energy term, is an easy one to derive even for a very complex network. As a consequence, calculation of the partition function (and of equilibrium probabilities) always reduces to a combinatorial problem in which the main task is the derivation of the combinatorial term $g(r_1, r_2, \cdots, r_n)$. However, this may not always be a simple problem. The combinatorial term will mirror the combinatorial properties of the network, the queuing disciplines, the type of queue (ordered, random, etc.), any constraints, and any limits on the queue size, etc. One can easily see that consideration of such constraints may result in very difficult, but not insoluble, combinatorial problems.

In cases where constraints, routing and queuing disciplines make it difficult to derive the combinatorial coefficient in closed form, one may still be able to derive the explicit numerical value of the combinatorial term corresponding to a given energy by means of numerical techniques on a computer. In so doing, one can develop an exact series expansion for the partition function and consequently obtain, again in the form of an exact series expansion, all the desired properties of the system under consideration.

Simple applications of the general solution

In this section, simple examples are given to demonstrate the use of the generalized partition function. In the ensuing discussion a service rate v should be interpreted as meaning that the service rate is exponentially distributed and has a mean value v.

Earlier in this paper, it has been stated that if a basic identification is to be made between a statistical mechanical system and a queuing system, it will be through the service rate v and the temperature T, via the relation

$$\ln v \propto 1/(k_B T).$$

The energy term in the partition function may easily be derived by first taking the system in some initial or ground state, and then considering deviations from that ground state. The choice of the ground state is an arbitrary one. Elements leaving the ground state (creating "vacancies") result in decrements in energy. Elements in an "excited" state (away from the ground state) give rise to increments in energy. This implies that it is possible to have an "excited" state in which there is no change in energy, but in which the presence of a new set of vacancies is enough to make the new state different from the ground state. Defined in this way, $E(r_1, r_2, \cdots, r_n)$ is the energy of those states that are characterized by the vector $(r_1, r_2, \cdots, r_n)$.

• *A simple switching network*

Consider a switching network comprising three switches, in which each switch can be either opened or closed. Suppose that the mean time during which any switch is open is the same as the mean time during which any switch is closed. This is, of course, tantamount to saying that there is no change in energy due to opening or closing a switch, because the energy of the open state is equal to that of the closed state. (The energy of work done in the switching process is not associated with the energy of states we are discussing here.) Describe each switch by a binary variable $d_k (k = 1, 2, 3)$, where

$$d_k = 1 \quad \text{if the } k\text{th switch is open;}$$
$$= 0 \quad \text{if the } k\text{th switch is closed.} \tag{43}$$

It is known that for such a system, the mean value of d_k is $\frac{1}{2}$, and the probability of any state is equal to $\frac{1}{8}$, i.e.

$$\langle d_k \rangle = \tfrac{1}{2} \qquad (k = 1, 2, 3); \tag{44}$$

$$p(d_1, d_2, d_3) = \tfrac{1}{8} \qquad \text{for all } d_1, d_2, d_3\,. \tag{45}$$

These results will be proved directly by use of our general result. Since there is no change of energy due to the opening or closing of a switch, the energy of the system is given by

$$E(d_1, d_2, d_3) = 0, \tag{46}$$

so that

$$\exp[-E(d_1, d_2, d_3)] = 1. \tag{47}$$

The combinatorial term is trivial, since there is only one state characterized by the vector (d_1, d_2, d_3); i.e.,

$$g(d_1, d_2, d_3) = 1. \tag{48}$$

On substituting into the general formula (39) we obtain

$$Z = \sum_{d_1=0}^{1} \sum_{d_2=0}^{1} \sum_{d_3=0}^{1} 1 = 2^3 = 8, \tag{49}$$

and from Eq. (41), the probability of being in a state characterized by the vector (d_1, d_2, d_3) is simply

$$p(d_1, d_2, d_3) = 1/Z = \tfrac{1}{8}. \tag{50}$$

The mean value of the variable d_1 is

$$\langle d_1 \rangle = \left(\sum_{d_1=0}^{1} d_1 \sum_{d_2=0}^{1} \sum_{d_3=0}^{1} 1 \right) \Big/ Z$$
$$= 2^2/Z = \tfrac{1}{2}. \tag{51}$$

The solution to a more general problem can be readily obtained. Consider a switching network consisting of n switches. Describe variables u_k and v_k as follows:

If the kth switch is closed, it has a probability $v_k dt$ of being opened in time dt, and if it is open, it has a probability $u_k dt$ of being closed in time dt. Describe the initial state of the system by the vector $(e_1, e_2, \cdots, e_n)$

with ($e_k = 0$ or 1). In a state characterized by the vector $(d_1, d_2, \cdots, d_n)$, the change in energy due to the kth switch is

$$\begin{aligned}
&-(e_k - d_k) \ln (u_k/v_k) \\
&\quad = 0 \qquad \text{if} \quad d_k = e_k; \\
&\quad = \ln (u_k/v_k) \qquad \text{if} \quad e_k = 0, \quad d_k = 1; \\
&\quad = \ln (v_k/u_k) \qquad \text{if} \quad e_k = 1, \quad d_k = 0.
\end{aligned} \tag{52}$$

The energy of the state characterized by the vector $(d_1, d_2, \cdots, d_n)$ is then

$$E(d_1, d_2, \cdots, d_n) = -\sum_{k=1}^{n} (e_k - d_k) \ln (u_k/v_k). \tag{53}$$

Also, since there is again only one state described by the vector $(d_1, d_2, \cdots, d_n)$, we have

$$g(d_1, d_2, \cdots, d_n) = 1. \tag{54}$$

On substituting into Eq. (39), one obtains the partition function Z for the switching network as

$$Z = \sum_{d_1} \sum_{d_2} \cdots \sum_{d_n} \prod_{k=1}^{n} (u_k/v_k)^{e_k - d_k}, \tag{55}$$

and the probability of being in the state described by $(d_1, d_2, \cdots, d_n)$ as

$$p(d_1, d_2, \cdots, d_n) = \left[\prod_{k=1}^{n} (u_k/v_k)^{e_k - d_k}\right] \Big/ Z, \tag{56}$$

with Z given by Eq. (55).

• *The machine interference model*

The machine interference model[3] comprises n identical machines and a single repairman. When a machine breaks down it is repaired by the repairman and put back into operation. If the repairman is busy, a broken machine has to wait for service, so that a queue builds up in front of the repairman.

Let u and v be the breakdown and repair rates of a machine respectively and, furthermore, let the initial or ground state of the system be that in which all machines are operating, i.e., in which there are no machines being or waiting to be repaired. Consider a situation in which k machines are broken. The energy $E(k)$ of states consisting of k broken machines is given by

$$\begin{aligned}
E(k) &= -k \ln u + k \ln v \\
&= -k \ln (u/v) = -\ln (u/v)^k,
\end{aligned} \tag{57}$$

where $-k \ln u$ is the energy due to the k "vacancies" (created by the breakdown of k machines) and $k \ln v$ is the energy due to the k elements in the "excited" state, i.e., the k machines awaiting repair. From Eq. (57), we have

$$\exp [-E(k)] = (u/v)^k. \tag{58}$$

The combinatorial term $g(k)$ is the number of distinct states that have energy $E(k)$. There are $\binom{n}{k}$ ways in which k machines could be broken. For each way in which k machines could be broken, there are $k!$ distinct states of the system since there are no restrictions on the order of the broken machines in the queue. Hence we have

$$g(k) = \binom{n}{k} k!, \tag{59}$$

and the partition function is given by

$$Z = \sum_{k=0}^{n} \binom{n}{k} k! \, (u/v)^k. \tag{60}$$

The probability of having k broken machines is then given by

$$p(k) = \binom{n}{k} k! \left(\frac{u}{v}\right)^k \Big/ \sum_{k=0}^{n} \binom{n}{k} k! \left(\frac{u}{v}\right)^k, \tag{61}$$

which is the well-known solution of the machine interference model.

• *The generalized machine interference model*

Here again, as in the machine interference model, we place no restrictions on the order of elements in the queue before the repairman. The system consists of n machines and a single repairman. The kth machine is characterized by a breakdown rate u_k and a repair rate v_k. We define an occupational variable d_k as

$$\begin{aligned}
d_k &= 0 \qquad \text{if the } k\text{th machine is broken and is awaiting repair or being repaired,} \\
&= 1 \qquad \text{if the } k\text{th machine is operational.}
\end{aligned} \tag{62}$$

The model is shown in Fig. 1. Consider the ground state as being that in which all occupational variables have the value 1. The energy of a state of the system characterized by the vector $(d_1, d_2, \cdots, d_n)$ is given by

$$E(d_1, d_2, \cdots, d_n) = -\sum_{k=1}^{n} (1 - d_k) \ln (u_k/v_k), \tag{63}$$

so that

$$\exp [-E(d_1, d_2, \cdots, d_n)] = \prod_{k=1}^{n} (u_k/v_k)^{1-d_k}. \tag{64}$$

If there are no restrictions or constraints on the order of elements (broken machines) in the queue before the repairman, then the number of distinct states characterized by the vector $(d_1, d_2, \cdots, d_n)$ is given by

$$g(d_1, d_2, \cdots, d_n) = \left[\sum_{k=1}^{n} (1 - d_k)\right]!. \tag{65}$$

Hence the partition function for this system is

$$Z_n = \sum_{d_1} \sum_{d_2} \cdots \sum_{d_n} \left[\sum_{k=1}^{n} (1 - d_k)\right]! \prod_{k=1}^{n} (u_k/v_k)^{1-d_k}, \tag{66}$$

and the probability of being in a state characterized by the vector $(d_1, d_2, \cdots, d_n)$ is

$$p(d_1, d_2, \cdots, d_n) = \left[\sum_{k=1}^{n} (1 - d_k)\right]! \prod_{k=1}^{n} (u_k/v_k)^{1-d_k}/Z_n \qquad (67)$$

with Z_n given in Eq. (66). When all machines have the same characteristics, that is, when both the breakdown rate and the repair rate are the same for all machines, the solution of the machine interference model, Eq. (60), is obtained by performing the summation in Eq. (66).

- *The cyclic queuing model*

The cyclic queuing model was described by Koenigsberg.[4] It consists of m sequential stages in a loop with each stage acting as a single server. The system serves N indistinguishable units; each of these units goes through all stages in succession and continuously repeats the process.

Let v_k be the rate of servicing a request at the kth stage and, furthermore, let the initial conditions be such that all N units are at the first stage. In one of the states characterized by the vector $(n_1, n_2, \cdots, n_m)$, where n_k is the number of units at the kth stage, the number of vacancies at the first stage is $N - n_1$, and the energy due to each vacancy is $-\ln v_1$. Thus the energy of a state characterized by a vector $(n_1, n_2, \cdots, n_m)$ is given by

$$E(n_1, n_2, \cdots, n_m) = -(N - n_1) \ln v_1 + \sum_{k=2}^{m} n_k \ln v_k, \qquad (68)$$

the summand being the contribution to the energy due to the elements at the kth $(k = 2, 3, \cdots, m)$ stage, a unit at the kth stage having energy $\ln v_k$. Hence

$$\exp[-E(n_1, n_2, \cdots, n_m)] = v_1^N \prod_{k=1}^{m} (1/v_k)^{n_k}. \qquad (69)$$

Since the units are indistinguishable, for a given $(n_1, n_2, \cdots, n_m)$, rearrangement of units gives rise to no new distinct configurations and therefore to no new states. In this case we have

$$g(n_1, n_2, \cdots, n_m) = 1. \qquad (70)$$

The partition function for this model is therefore

$$Z_N = v_1^N \sum_{n_1, n_2, \cdots, n_m} \prod_{k=1}^{m} (1/v_k)^{n_k}, \qquad (71)$$

with

$$\sum_{k=1}^{m} n_k = N.$$

The probability of being in a state characterized by the vector $(n_1, n_2, \cdots, n_m)$ is

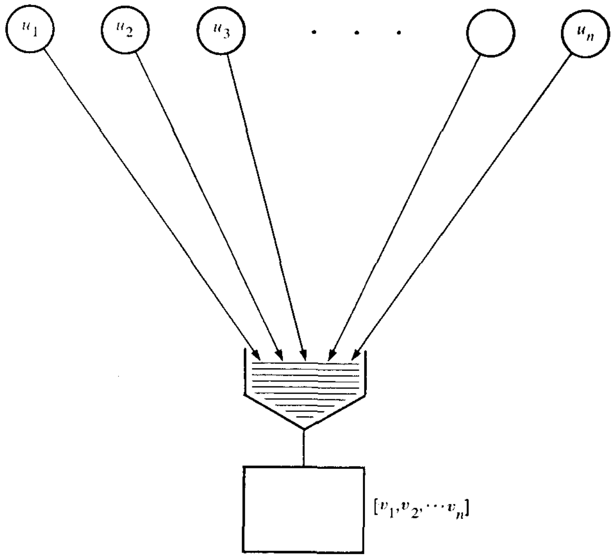

Figure 1 The generalized machine interference model.

$$p(n_1, n_2, \cdots, n_m) = \left[v_1^N \prod_{k=1}^{m} (1/v_k)^{n_k}\right] \Big/ Z_N, \qquad (72)$$

which is the Koenigsberg solution. For more complex models, which are readily solved using the generalized configuration-independent partition function, see Reference 5.

Further considerations

The ideas put forward in this section are familiar to and frequently used in statistical mechanics. Since such techniques are not commonly used in queuing or network analysis, it is worthwhile to mention them because of their potential applicability in systems analysis.

The use of exact series expansions in the analysis of complex networks has already been suggested. If the total number of elements in a system is large, the summation in the partition function, Eq. (39), may require a significant amount of computation time. In this case one can resort to the use of a truncated series expansion in which one uses only the lower-order terms of the series represented by the partition function. This can be easily achieved by restricting the summation to only those terms of a polynomial of some predefined degree that is less than the degree of the partition function itself. Of course the use of the truncated series implies an assumption that the total contribution to the partition function (of all terms of order q, for example) is a strictly monotonic decreasing function of q. This condition can very often be achieved by a judicious choice of initial conditions. By comparing results obtained through the use of a truncated series of degree q, for example, with those

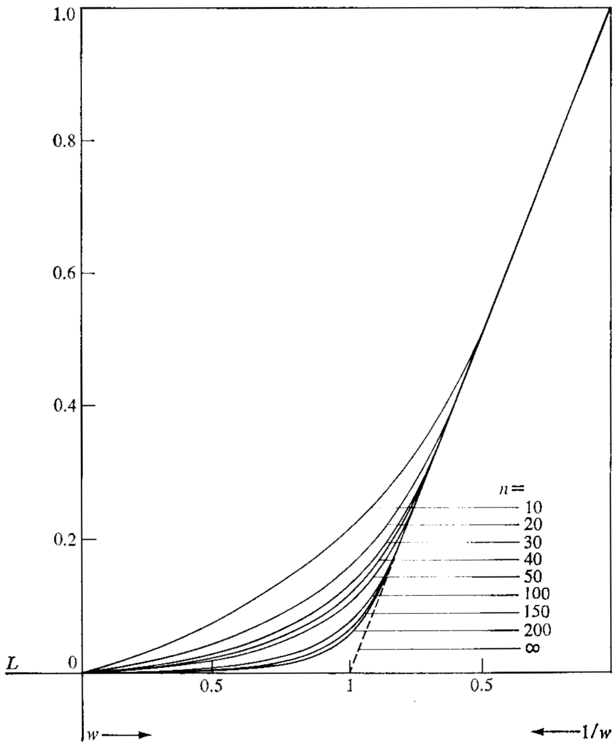

Figure 2 The normalized mean length of the queue of the machine interference model plotted against $w = nu/v$ for $w \leq 1$ and against $1/w$ for $w > 1$.

obtained through the use of a similar series of degree $q + 1$, one can always decide whether the result will converge, at least to the accuracy one desires. In today's queuing systems (computers, traffic, etc.) one is generally confronted with heavily loaded systems for which computation of the partition function by digital computers may be expensive. The most promising approach to this problem would seem to be that of determining the asymptotic behavior of the partition function as a function of total system size (number of elements), and then using the asymptotic formula to calculate the properties of the system. This technique is now seldom exploited in computer systems analysis, even for queuing models as simple as the machine interference model. (The asymptotic properties of this model are given in Ref. 5.) The use of asymptotic formulae not only provides a better understanding of the properties of a system, but also results in considerable saving in time and human effort in the analysis of a large system.

In present day analysis of complex systems, a simulation technique is often employed. However, a simulation model and a mathematical model of the same system (if indeed such a mathematical model exists) often give different results in a certain region. This difference, in some instances, is probably due to a phenomenon similar to that of phase transitions in statistical mechanics. (By way of definition, transitions in which the first or higher derivatives of the logarithm of the partition function diverge or change discontinuously are called phase transitions.) The machine interference model exhibits such a phenomenon. Figure 2 illustrates the normalized mean length of the queue, i.e.,

$$L = \frac{1}{n} x \frac{d}{dx} \ln Z, \qquad (73)$$

with $x = u/v$ and Z given in Eq. (60). In the limit as $n \to \infty$, L develops a mathematical singularity at the point $w = w_c = 1$, the "critical point" of the model. This is amply illustrated in Fig. 2, by the kink at $w_c = 1$ when $n \to \infty$. The derivative of queue length with respect to the variable x (this is analogous to the specific heat in statistical mechanics) diverges to infinity at the point $w = 1$ as $n \to \infty$.

For finite n, dL/dx goes through a maximum at a point $w_{\max}$ where

$$w_{\max} = 1 + g(n), \qquad (w = nx); \qquad (74)$$

$$\frac{dL}{dx} \approx Cn \quad \text{at} \quad w_{\max}, \qquad (75)$$

$g(n)$ being a positive decreasing function of n, and C a constant. The point $w_{\max}$ can be called the unstable point of the system, and it tends to the critical point as $n \to \infty$. Because of the large fluctuations in the vicinity of $w_{\max}$, one finds in performing simulation experiments around the point $w_{\max}$ that the time taken to attain equilibrium is longer in the vicinity of $w_{\max}$ than it is in regions away from $w_{\max}$. Furthermore, one also finds that this difficulty in attaining equilibrium (around $w_{\max}$) increases with increasing n. It is sometimes also found that the mathematical model and its simulation counterpart give different results in the neighborhood of $w_{\max}$. This is presumably because the mathematical model of Eq. (60) is an equilibrium solution, and the simulation model may not have attained equilibrium, even after a presumably long settling time.

Conclusions

The techniques described in this paper have their roots in statistical mechanics, a science that has successfully predicted the behavior of macroscopic bodies and systems composed of a large number of microscopic elements. It would seem feasible that such statistical mechanical techniques should be applicable to the large and complex systems envisioned by today's industry. Indeed, such an approach has been considered by Beneš,[2] but no further work has been performed in this direction because of a certain amount of skepticism surrounding the difficulty of evaluating the partition function.

Herein, this difficulty is partially overcome by expressing the partition function, not as a sum over all possible distinct states as in its representation by Beneš,[2] but instead as a sum over all distinct energy states. This latter representation of the partition function simplifies the problem considerably in that it is now reduced to finding a combinatorial term representing the number of distinct configurations having a particular energy. This combinatorial term can be obtained, if not by simple combinatorial considerations, by computational analysis.

The technique provides a uniform way of treating systems of varying degrees of complexity and generality, the geometry of the system configuration appearing only through the combinatorial term. It has been demonstrated that the partition function can be used to overcome some of the inherent difficulties associated with queuing theory.

Acknowledgments

The author expresses his appreciation to E. M. Thomas for introducing him to this exciting problem and for the interest he showed and maintained throughout the course of this work.

He also expresses his gratitude to G. R. Holzman for his interest, encouragement, and stimulating discussions and for his efforts to promote interest in this endeavor.

Sincere thanks are due also to R. Landauer and B. Meister for making the author aware of the work of Beneš, and for their interest in this effort.

References

1. J. Riordan, *An Introduction to Combinatorial Analysis,* J. Wiley and Sons, Inc., New York 1958, Chapter 2.
2. V. E. Beneš, *Mathematical Theory of Connecting Networks and Telephone Traffic,* Academic Press, Inc., New York and London 1965, Chapter 7.
3. W. Feller, *An Introduction to Probability Theory and Its Applications,* 2nd Edition, J. Wiley and Sons, Inc., New York 1957, pp. 416–418.
4. E. Koenigsberg, "Cyclic Queues," *Operations Research Quarterly* **9**, 22 (1958).
5. A. E. Ferdinand, "A Statistical Mechanics Approach to Systems Analysis," *IBM TR 21-348,* 1969.

Received October 14, 1969