Division of Research
Graduate School of Business Administration
The University of Michigan

March 1983
Revised, December 1983

EQUIVALENT REPRESENTATIONS OF SYSTEM THROUGHPUT IN CLOSED QUEUEING NETWORK MODELS OF MULTISERVER QUEUES

Working Paper No. 324-B

Kathryn E. Stecke
The University of Michigan

and

Bruce W. Schmeiser
Purdue University

ABSTRACT

The steady-state expected production (throughput) of many closed queueing network models of multiserver queues is a function of the number of items and facilities in the system, the number of servers at each facility, and the workload allocated to each facility. With appropriate scaling, the expected production is given six representations, each with a corresponding practical interpretation. These representations are discussed and proved equivalent. The representations both provide insight and are useful in proving some results relating production to server grouping and workload allocation decisions.

1. INTRODUCTION

Systems in which parts or customers visit several different facilities for processing can, in some cases, be adequately modeled by a multiclass or multiserver network of queues. Multiclass models can depict sequential, fixed routing for some types of parts, while multiserver models are used when there can be more than one server at some facilities. The multiclass models have been used to evaluate the performance of both computer systems (see Kleinrock [1976], Rose [1976], Baskett et al. [1975], Reiser and Kobayashi [1975, 1976], and Bard [1979]) and flexible manufacturing systems (FMSs) (see Cavaillé and Dubois [1982]). An FMS consists of a set of numerically controlled machine tools, interconnected with automated material handling equipment, capable of the simultaneous and efficient manufacture of a variety of part types. Most real-time functions, such as actual machining operations, automatic tool interchange, and part movement are under the control of one or more computers (see Buzacott and Shanthikumar [1980] and Stecke [1977, 1983]). Multiserver models are well-suited for representing those FMSs where functionally similar machine tools can be pooled into a machine group (multiserver queue). These machine tools are then identically tooled so as to be able to perform the same operations during real-time control, as discussed, e.g., by Stecke and Solberg [1981a, 1981b, 1982]. Pooling machines into groups allows the machine redundancy required to automatically reroute in machine breakdown situations.

Closed networks of arbitrarily connected multiserver queues are considered in this paper. The results are therefore also applicable to the central-server model. Manufacturing, rather than computer, performance terminology is used, reflecting the motivating FMS application. The correspondence with queueing terminology is direct and meaningful: <u>expected production</u> is <u>throughput</u>, part is job or customer, part type is customer type, machine group is multiserver queue, and machine is device or server.

Using the closed queueing network (CQN) model, the expected production (throughput) of an FMS is defined in §2 as a function of several system parameters. As a result of our particular scaling of one of the parameters--the workload assigned to each machine group--several alternative interpretations of production, defined in §3, are proven equivalent in §4. Some of the representations are new; the equivalence of all six is new. These alternative representations help the user of queueing network models to better understand and to interpret the mathematical representation of expected production obtained from the CQN model. They also provide additional mathematical tools that can be used to discover and prove properties of optimal solutions of several associated performance optimization problems, as discussed in §5. Finally, they can motivate further research into the modeling of FMSs and their performance. A brief summary is provided in §6.

The current research was motivated by investigations of two particular FMS production planning problems, called the grouping problem (how to best partition m machines into g machine groups) and the loading problem (how to best allocate operations and associated tooling among groups). For both problems, the objective is to maximize expected production subject to FMS technological and capacity constraints (see Stecke and Morin [1982], Stecke and Solberg [1982], Stecke [1977, 1982, 1983]). The alternative representations are useful in providing insight into understanding the mathematical definition of expected production from CQN models of FMSs.

2. THE CLOSED QUEUEING NETWORK MODEL

Consider a single-class, closed system containing n parts, each of the same part type. The system consists of m machines (servers) that have been partitioned into g machine groups, with s_i machines in group i. The average processing time of an operation by a machine in group i is t_i, $i=1,\ldots,g$.

The routing is arbitrary, and can be described by the visit frequencies, or relative arrival rates, q_i, where q_i can be:

i) the probability that the next machine visited belongs to group i;

ii) the average number of times per some time period that a machine in group i is visited; or

iii) the mean number of operations per part at machine group i.

For example, to model realistic systems containing more than one part type, the values of q_i can be determined by

$$q_i = \sum_j a_j\, r_{ij},$$

$$= \text{the average number of operations per part at group } i,$$

where

a_j = production ratio of part type j relative to all part types currently being produced;
$a_j > 0$, for all j, and $\sum_j a_j = 1$; and

r_{ij} = number of operations on part type j performed by group i.

The q_i can also be solutions to the traffic equations, $q_i = \sum_j p_{ji} q_j$. The routing might also be first-order Markovian (defined by transition probabilities, p_{ij}), multiple-class Markovian (defined by transition probabilities for each part type k, $p_{ij}(k)$--see Reiser and Kobayashi [1975]), higher-order Markovian (for example, second order is defined by p_{ijk}, which is the probability that a part previously at i, now at j, goes next to k--see Kobayashi and Reiser [1975]), or fixed routes through the system (defined by routing vectors for each part type, $r(k) = (r(k,1), r(k,2),\ldots)$, where $r(k,j)$ is the index of the j'th machine group visited by a part of type k--see Kelly [1979]). All of these routing mechanisms produce the same values for certain output measures such as expected production, as does the q_i. For additional routing details, see Stecke and Solberg [1981a].

The queue discipline can be FCFS, infinite server, LCFS preempt-resume, processor sharing (see Baskett et al. [1975]), random selection, or an arbitrary distribution defined at each node (Kelly [1979]). The service time distribution is arbitrary, except for FCFS machine groups, which require exponential service times.

The usual measure of the relative workload assigned to group i is w_i, the product of visit frequency and average processing time, or $w_i = q_i t_i$, $i=1,\ldots,g$ (Buzen [1973], Reiser and Kobayashi [1975], Solberg [1977]). These workloads are relative in that the q_i's need not sum to one.

The state of the system is denoted by $\tilde{n} = (n_1, n_2, \ldots, n_g)$, where n_i is the number of parts at machine group i, both those waiting and those in process. For all i, n_i is an integer between zero and n, and $\sum_{i=1}^{g} n_i = n$. The steady-state probability of being in state $\tilde{n}$ is denoted by $p(\tilde{n}) = p(n_1, n_2, \ldots, n_g)$ which has the product form solution:

$$p(\tilde{n}) = \frac{1}{G(g,n;S,W)} \prod_{i=1}^{g} f_i(n_i), \tag{1}$$

where $S = (s_1, s_2, \ldots, s_g)$, $W = (w_1, w_2, \ldots, w_g)$, and the normalizing constant is:

$$G(g,n;S,W) = \sum_{\substack{n_1 \geq 0 \quad n_2 \geq 0 \quad \cdots \quad n_g \geq 0 \\ n_1 + n_2 + \cdots + n_g = n}} f_1(n_1) f_2(n_2) \cdots f_g(n_g) \tag{2}$$

and

$$f_i(n_i) = \begin{cases} \dfrac{w_i^{n_i}}{n_i!}, & n_i \leq s_i, \\[2ex] \dfrac{w_i^{n_i}}{s_i! \, s_i^{n_i - s_i}}, & n_i > s_i; \qquad i=1,\ldots,g. \end{cases} \tag{3}$$

Single-machine, multiple-machine, and infinite-machine groups correspond, respectively, to $s_i = 1$, $1 < s_i < n$, and $s_i \geq n$.

The expected production, which is the expected number of parts produced per unit time, is an important performance measure. For a system containing n parts, it is a function of G(g,n;S,W), which in turn is a function of assigned workload, w_i, and grouping, s_i. In fact, for a particular scaling of the q_i and t_i, which will be provided shortly, the production function, Pr(g,n;S,W), is given by (Reiser and Kobayashi [1975]):

$$Pr(g,n;S,W) = \frac{G(g,n-1;S,W)}{G(g,n;S,W)}. \tag{4}$$

3. RESCALING THE WORKLOAD AND PRODUCTION FUNCTION

For our purposes, w_i is scaled to provide our workload measure:

$$X_i = q_i t_i / [(\sum_{j=1}^{g} q_j t_j)/(\sum_{j=1}^{g} s_j)] \; . \tag{5}$$

Notice that the numerator is the usual definition of workload assigned to group i, and the denominator is the average workload per machine.

Our particular scaling is useful for several reasons:

i) For any given number of machines in the system, regardless of their grouping, the total amount of work to be allocated among groups always equals the total number of machines:

$$\sum_{i=1}^{g} X_i = \sum_{i=1}^{g} s_i = m.$$

ii) The workload is independent of any particular scaling of q_i and t_i;

iii) Alternative workloads can be compared with a balanced workload, since a balanced workload, regardless of system size or configuration, is

$$X_1/s_1 = X_2/s_2 = \dots = X_g/s_g = 1.$$

iv) As shall be seen in the following section, this scaling provides a normalized, dimensionless measure of production, whose values lie between zero and one.

v) As also seen in the following section, workload definition (5) allows some new, alternative, equivalent representations of expected production, which are useful in proving properties of the production function.

Note that equations (1), (2), (3), and (4) remain nearly the same when X_i is substituted for w_i. The only difference is that the functions $f_i(n_i)$ are replaced by $h_i(n_i)$, by substituting X_i for w_i to reflect the chosen scaling. These $h_i(n_i)$ are also valid factors of the product form solution, $p(\tilde{n})$.

4. ALTERNATIVE REPRESENTATIONS OF THE PRODUCTION FUNCTION

Prior to presenting the different representations of the production function, some preliminary definitions are required.

4.1 Preliminary Notation and Definitions

Def 1: Let $I_i(\tilde{n})$ denote the number of busy machines in group i when the system is in state $\tilde{n}$.

For example, for a system of seven groups with two machines in each group that is in state $\tilde{n} = (0,0,1,4,2,0,1)$,

$I_1(\tilde{n}) = I_2(\tilde{n}) = I_6(\tilde{n}) = 0$,

$I_3(\tilde{n}) = I_7(\tilde{n}) = 1$, and

$I_4(\tilde{n}) = I_5(\tilde{n}) = 2$.

Def 2: Let $U(\tilde{n})$ denote the fraction of machines that are busy in a particular state $\tilde{n}$. That is,

$$U(\tilde{n}) = [\sum_{i=1}^{g} I_i(\tilde{n})]/(\sum_{i=1}^{g} s_i).$$

$U(\tilde{n})$ is the ratio of busy machines when the system is in state $\tilde{n}$ and the total number of machines. Referring to the above example, $U(\tilde{n}) = 6/14 = 3/7$.

Def 3: Let Pr_I denote the expected ideal production rate of a manufacturing system, which is the expected number of parts per time unit produced if all machines are always busy.

Pr_I is a measure of the maximum capacity of the system.

Def 4: $Pr_A(\tilde{n})$ is the expected actual production rate when the system is in state $\tilde{n}$, which is the expected ideal production rate weighted by the utilization of machines in $\tilde{n}$. That is,

$$Pr_A(\tilde{n}) = Pr_I U(\tilde{n}).$$

Def 5: Let Pr_A denote the expected production rate, which is the actual steady-state expected number of parts produced per time unit. That is,

$$Pr_A = \sum_{\tilde{n} \in N_{g,n}} p(\tilde{n}) Pr_A(\tilde{n}),$$

where

$$N_{g,n} = \{\tilde{n} \mid \sum_{i=1}^{g} n_i = n \quad \text{and} \quad n_i \geq 0 \quad \text{for } i=1,\ldots,g\}.$$

Pr_A is the usual production function obtained by substituting w_i into equation (4).

4.2 The Alternative Representations

The six different characterizations of the production function are presented in the following theorem.

Theorem 1. The following alternative representations of the production function, $Pr(g,n;S,X)$, are equivalent.

A. $$\frac{\displaystyle\sum_{\tilde{n}\in N_{g,n-1}} \prod_{i=1}^{g} h_i(n_i)}{\displaystyle\sum_{\tilde{n}\in N_{g,n}} \prod_{i=1}^{g} h_i(n_i)}\ ;$$

B. $(1/m) \sum_{k=1}^{m} k \text{ Prob}\{k \text{ machines are busy}\};$

C. $\sum_{\tilde{n}\in N_{g,n}} p(\tilde{n})U(\tilde{n});$

D. The steady-state probability that a randomly selected machine is busy;

E. The ratio of the actual steady-state expected number of parts per time unit to the expected ideal number of parts per time unit which would be obtained if all of the machines were always busy:

$$Pr_A/Pr_I;$$

F. $$\frac{\displaystyle\sum_{\tilde{n}\in N_{g,n}} U(\tilde{n}) \prod_{i=1}^{g} h_i(n_i)}{\displaystyle\sum_{\tilde{n}\in N_{g,n}} \prod_{i=1}^{g} h_i(n_i)}\ .$$

The first representation, A, is that defined by equation (4), using $h_i(n_i)$ rather than $f_i(n_i)$. <u>Because of our particular scaling of workload as X, rather than w</u>, this usual definition of expected production is scaled to provide the remaining alternative definitions. In particular, representations B and C are both measures of system utilization: the expected fraction of machines busy. Representation D is a measure of single machine utilization. It is the fraction of time busy, a dimensionless quantity normalized to lie between zero and one. Representation E is an efficiency measure. Finally, representation F provides a relationship similar to the first; however, now the sums are over identical state spaces.

Notice that all representations are not defined on the same state space. The first, third, and sixth definitions include similar sums over the same state space, each sum involving $\binom{n+g-1}{g-1}$ terms. The summation of the second representation, which contains many fewer terms, is over all machines, m. The fourth is a probability and the fifth provides a transformation from the normalized production efficiency measure to the actual production rate, as measured in completed parts per time unit.

4.3 Equivalence of the Six Alternative Definitions

Prior to proving the equivalence of the six representations in Theorem 1, a preliminary result is required. This result mathematically defines Pr_I, the expected ideal production rate, which is obtained when all machines are always busy, or the maximum system capacity.

Lemma 2. $$Pr_I = (\sum_{i=1}^{g} s_i)/(\sum_{i=1}^{g} q_i t_i) .$$

Proof: Since

$$q_i t_i = \text{workload on machine group } i\text{, in time units per part,}$$

we have:

$$\sum_{i=1}^{g} q_i t_i = \text{total workload on all machine groups.}$$

Then

$$(\sum_{i=1}^{g} q_i t_i)/(\sum_{i=1}^{g} s_i) = \text{average workload per machine, given that the machines are always busy, in time units per part.}$$

Inverting, we have:

$$(\sum_{i=1}^{g} s_i)/(\sum_{i=1}^{g} q_i t_i) = \text{expected production rate of the system when all machines are always busy, in parts per time unit.}$$

$$= Pr_I. \quad ||$$

We now prove Theorem 1. The representations are proven equivalent in the following order: representation A is equivalent to E; E to C; C to B; C to D; and finally C to F.

Proof of Theorem 1:

i) Representation A is equivalent to E:

$$\frac{\sum_{\tilde{n}\varepsilon N_{g,n-1}} \prod_{i=1}^{g} h_i(n_i)}{\sum_{\tilde{n}\varepsilon N_{g,n}} \prod_{i=1}^{g} h_i(n_i)} \qquad \text{(by Definition A)}$$

$$= \frac{\sum_{\tilde{n}\varepsilon N_{g,n-1}} [\prod_{i=1}^{g} (\sum_{j=1}^{g} s_j / \sum_{j=1}^{g} q_j t_j)^{n_i} f_i(n_i)]}{\sum_{\tilde{n}\varepsilon N_{g,n}} [\prod_{i=1}^{g} (\sum_{j=1}^{g} s_j / \sum_{j=1}^{g} q_j t_j)^{n_i} f_i(n_i)]} \qquad \text{(by substitution)}$$

$$= \frac{(\sum_{j=1}^{g} s_j / \sum_{j=1}^{g} q_j t_j)^{n-1} \sum_{\tilde{n}\varepsilon N_{g,n-1}} [\prod_{i=1}^{g} f_i(n_i)]}{(\sum_{j=1}^{g} s_j / \sum_{j=1}^{g} q_j t_j)^{n} \sum_{\tilde{n}\varepsilon N_{g,n}} [\prod_{i=1}^{g} f_i(n_i)]} \qquad \text{(by simplification)}$$

$$= Pr_A / (\sum_{i=1}^{g} s_i / \sum_{i=1}^{g} q_i t_i) \qquad \text{(by equations (3), (2), and (4))}$$

$$= Pr_A/Pr_I \qquad \text{(by Lemma 2)}$$

which is definition E. $||$

ii) Representation E is equivalent to C:

$$\sum_{\tilde{n}\varepsilon N_{g,n}} p(\tilde{n})U(\tilde{n}) \qquad \text{(by Definition C)}$$

$$= \frac{\sum_{\tilde{n}\varepsilon N_{n,g}} p(\tilde{n})U(\tilde{n})Pr_I}{Pr_I}$$

$$= \frac{\sum_{\tilde{n} \in N_{g,n}} p(\tilde{n}) Pr_A(\tilde{n})}{Pr_I} \qquad \text{(by Definition 4)}$$

$$= Pr_A / Pr_I \qquad \text{(by Definition 5)}$$

which is definition E. ||

iii) Representation C is equivalent to B:

$$\sum_{\tilde{n} \in N_{g,n}} p(\tilde{n}) U(\tilde{n}) \qquad \text{(by Definition C)}$$

$$= \sum_{k=1}^{m} \sum_{\tilde{n} \in N_{g,n} \cap \{\tilde{n} \mid U(\tilde{n}) = k/m\}} p(\tilde{n})(k/m)$$

$$= \sum_{k=1}^{m} p(k \text{ machines are busy})(k/m)$$

which is definition B. ||

iv) Representation C is equivalent to D:

Prob{a randomly selected machine is busy} (by Definition D)

$$= \sum_{i=1}^{g} \sum_{k=1}^{s_i} \text{Prob\{machine k in machine group i is selected\}} \cdot \text{Prob\{machine k of group i is busy} \mid \text{machine k of group i is selected\}}$$

$$= \sum_{i=1}^{g} \sum_{k=1}^{s_i} (1 / \sum_{i=1}^{g} s_i) \text{ Prob\{machine k in group i is busy\}}$$

$$= \sum_{i=1}^{g} \text{E\{number of busy machines in group i\}} / (\sum_{i=1}^{g} s_i)$$

$$= (\sum_{i=1}^{g} [\sum_{\tilde{n} \in N_{g,n}} p(\tilde{n}) I_i(\tilde{n})]) / (\sum_{i=1}^{g} s_i) \qquad \text{(by Definition 1)}$$

$$= \left(\sum_{\tilde{n} \varepsilon N_{g,n}} p(\tilde{n}) \sum_{i=1}^{g} I_i(\tilde{n}) \right) / \left(\sum_{i=1}^{g} s_i \right)$$

$$= \sum_{\tilde{n} \varepsilon N_{g,n}} p(\tilde{n}) U(\tilde{n}) \qquad \text{(by Definition 2)}$$

which is definition C. $||$

v) Representation C is equivalent to F:

$$\sum_{\tilde{n} \varepsilon N_{g,n}} p(\tilde{n}) U(\tilde{n}) \qquad \text{(by Definition C)}$$

$$= \sum_{\tilde{n} \varepsilon N_{g,n}} [G^{-1}(g,n;S,X) \prod_{i=1}^{g} h_i(n_i)] U(\tilde{n})$$

(by Definition of $p(\tilde{n})$, equation (1))

$$= \frac{\sum_{\tilde{n} \varepsilon N_{g,n}} U(\tilde{n}) \prod_{i=1}^{g} h_i(n_i)}{\sum_{\tilde{n} \varepsilon N_{g,n}} \prod_{i=1}^{g} h_i(n_i)} \qquad \text{(by equation (2))}$$

which is definition F. $||$

The proof of Theorem 1 is now complete. The alternative representations are used in the following section to explore and demonstrate additional properties of optimal workloads and the production function.

5. APPLICATIONS OF THE ALTERNATIVE REPRESENTATIONS

Some properties of the production function are developed in this section by using the alternative definitions. §5.1 contains some preliminary results for networks of multiserver queues that are based on the equivalent definitions. These results are used in §5.2 to prove some results for balanced workloads and their optimality in closed networks of singleserver queues.

5.1 Preliminary Results for Networks of Multiserver Queues

The following results are valid for systems consisting of groups of machines with s_i machines in each group i.

Corollary 3. For any grouping S, the expected number of busy machines is given by:

$$Pr(g,n;S,X) \sum_{i=1}^{g} s_i .$$

Proof: The result follows from Representation B of Theorem 1. ||

The next theorem defines the number of states utilizing k out of m machines when n parts are in the system.

Theorem 4. For any S, the number of states, $\tilde{n}$, such that $U(\tilde{n}) = k/m$ is

$$\binom{m}{k}\binom{n-1}{n-k}, \text{ for } k = 0,1,2,\ldots,m.$$

Proof:

Consider any state, $\tilde{n}$, such that $U(\tilde{n}) = k/m$; that is, there are k out of m machines busy and m-k machines idle.

The number of ways to select k machines out of m machines is $\binom{m}{k}$.

This is the number of states utilizing k of m machines with k parts in the system.

Furthermore, the number of ways to distribute n indistinguishable parts among k busy machines is

$$\frac{(n-1)(n-2)\ldots(n-k+1)}{(k-1)(k-2)\ldots1} = \frac{(n-1)!}{(n-k)!(k-1)!} = \binom{n-1}{n-k}.$$

Therefore, the number of states utilizing k out of m machines with n parts in the system is

$$\binom{m}{k}\binom{n-1}{n-k}. \quad ||$$

As a direct consequence of applying Theorem 4 to representation C, we have:

Corollary 5. If $p(\tilde{n}) = p_k$ for all $\tilde{n}$ such that $U(\tilde{n}) = k/m$, then

$$Pr(g,n;S,X) = \sum_{k=1}^{m} p_k \binom{m}{k}\binom{n-1}{n-k}\frac{k}{m}.$$

Proof: The result follows directly from Theorem 4 and representation C. ||

Corollary 5 may also be seen to be true by noting that under the assumptions of Corollary 5,

$$p_k \binom{m}{k}\binom{n-1}{n-k} = \text{Prob}\{k \text{ machines are busy}\},$$

and applying representation B.

Corollary 5 is required in the proof of Theorem 9.

5.2 Results for Closed Networks of Singleserver Queues

The results in this section consider systems having only one machine in each group. In this case, $s_1 = \ldots = s_g = 1$, and $S = (1,\ldots,1) = \vec{1}$. An equal, balanced workload on each machine implies $X_1 = X_2 = \ldots = X_g = 1$, which is represented by $X = (X_1, X_2,\ldots,X_g) = \vec{1}$.

Theorem 6. $Pr(g,n;\vec{1},\vec{1}) = n/(n+g-1)$.

Proof: $Pr(g,n;\vec{1},\vec{1}) = G(g,n-1;\vec{1},\vec{1})/G(g,n;\vec{1},\vec{1})$ (by equation (4))

The result follows directly by substituting $X = \vec{1}$ into equations (3) and (4) and simplifying. ||

The following corollary to Theorem 6 states that production for a balanced workload monotonically increases in n, and approaches one in the limit as n approaches infinity. Since $Pr(g,n;S,X)$ is a probability, by representation D of Theorem 1, then balancing the workload is optimal for the limiting case of an infinite number of parts in the system.

Corollary 7. For $g > 1$, $d[Pr(g,n;\vec{1},\vec{1})]/dn > 0$ and $\lim_{n\to\infty} Pr(g,n;\vec{1},\vec{1}) = 1$. Hence balancing is optimal for $n = \infty$.

Proof: $d[Pr(g,n;\vec{1},\vec{1})]/dn = (g-1)/(g+n-1)^2$,

which is positive for all $g > 1$. Also,

$\lim_{n\to\infty} Pr(g,n;\vec{1},\vec{1}) = \lim_{n\to\infty} \frac{n}{n+g-1} = 1$, by ℓ'Hospital's Rule. ||

Theorem 6 and Corollary 7, in terms of a different workload scaling, can be found in Buzacott and Shanthikumar [1980]. Additional information on the optimality of balanced (unbalanced) workloads in certain networks of multiserver queues can be found in Stecke and Morin [1982] (Stecke and Solberg [1982]). The following theorem provides another result for balanced workloads.

Theorem 8. If $S = \vec{1}$, then $X = \vec{1}$ if and only if each state, $\tilde{n}$, is equally probable.

Proof:

(Sufficient):

Assume that $X_1 = X_2 = \ldots = X_g = 1$.

For all $\tilde{n}\varepsilon N_{g,n}$,

$$p(\tilde{n}) = G(g,n;\vec{1},X)^{-1} \prod_{i=1}^{g} X_i^{n_i}, \qquad \text{(equation (1))}$$

where

$$G(g,n;\vec{1},X) = \sum_{\tilde{n}\varepsilon N_{g,n}} \prod_{i=1}^{g} X_i^{n_i}.$$

Then, evaluating at $X = \vec{1}$,

$$p(\tilde{n}) = \Big(\sum_{\tilde{n}\varepsilon N_{g,n}} 1\Big)^{-1}$$

$$= \binom{n+g-1}{g-1}^{-1}, \text{ for all } \tilde{n}\varepsilon N_{g,n}.$$

Therefore each state is equally likely.

(Necessary):

Assume that $p(\tilde{n}) = 1/\binom{n+g-1}{g-1}$, for all $\tilde{n}\varepsilon N_{g,n}$.

Then

$$G(g,n;\vec{1},X)^{-1} \prod_{i=1}^{g} X_i^{n_i} = 1/\binom{n+g-1}{g-1}, \text{ for all } \tilde{n}\varepsilon N_{g,n}.$$

This provides $\binom{g+n-1}{g-1}$ equations in g unknowns.

Choose the g states, $\tilde{n}^i$, where all n parts are at machine i; i=1,...,g.

Then we have the g equations:

$$X_i^n = G(g,n;\vec{1},X)/\binom{n+g-1}{g-1} = C, \quad i=1,...,g.$$

Taking the single, positive, real root of $X_i^n = C$ yields:

$$X_i = [G(g,n;\vec{1},X)/\binom{n+g-1}{g-1}]^{1/n} = \text{Constant} = K, \quad i=1,...,g. \tag{6}$$

Since

$$\sum_{i=1}^{g} X_i = g \qquad \text{(by definition)}$$

$$= Kg, \qquad \text{(by equation (6))}$$

then

$$X_i = 1, \qquad i=1,...,g. \quad ||$$

Corollary 5, Theorem 6, and Theorem 1 are now used to prove Theorem 9, which claims that if all states, $\tilde{n}$, of a system of single machines are equally probable in steady-state, then the expected production achieved is the same as that for a balanced workload. Note that Theorem 9 actually follows directly from Theorem 8. However, additional insight can be obtained from the following direct proof of Theorem 9 that uses the alternative definitions.

Theorem 9. If $p(\tilde{n}) = 1/\binom{n+g-1}{g-1}$ for every state $\tilde{n}$ and $S = \vec{1}$, then

$$Pr(g,n;\vec{1},X) = Pr(g,n;\vec{1},\vec{1}).$$

Proof:

$\Pr(g,n;\vec{1},X)$

$$= \sum_{\tilde{n}\in N_{g,n}} p(\tilde{n})U(\tilde{n}) \qquad \text{(by Representation C of Theorem 1)}$$

$$= \sum_{\tilde{n}\in N_{g,n}} \binom{n+g-1}{g-1}^{-1} U(\tilde{n}) \qquad \text{(by assumption)}$$

$$= \sum_{k=1}^{g} \binom{n+g-1}{g-1}^{-1} \binom{g}{k}\binom{n-1}{n-k} \frac{k}{g} \qquad \text{(by Corollary 5 and m=g)}$$

$$= \binom{n+g-1}{g-1}^{-1} \sum_{k=1}^{g} \binom{g-1}{k-1}\binom{n-1}{n-k}$$

$$= \binom{n+g-1}{g-1}^{-1} \binom{n+g-2}{n-1}$$

$$= \frac{n}{n+g-1}$$

$$= \Pr(g,n;\vec{1},\vec{1}) \qquad \text{(by Theorem 6).} \qquad ||$$

The following Lemma provides a result on the unweighted (by $p(\tilde{n})$) sum of the utilizations of all states, $U(\tilde{n})$.

<u>Lemma 10</u>. $\sum_{\tilde{n}\in N_{g,n}} U(\tilde{n}) = \binom{n+m-2}{m-1}$, if $S = \vec{1}$.

Proof:

Equating the numerators of representations A and F of Theorem 1 yields

$$\sum_{\tilde{n}\in N_{g,n}} U(\tilde{n}) \prod_{i=1}^{g} h_i(n_i) = G(g,n-1;S,X), \text{ for all } S \text{ and } X.$$

If $S = \vec{1}$, then $h_i(n_i) = X_i^{\,n_i}$, for all i. In particular, when $X_1 = \ldots =$ $= X_g = 1$,

$$\sum_{\tilde{n}\in N_{g,n}} U(\tilde{n}) = G(g,n-1;\vec{1},\vec{1})$$

$$= \binom{n+g-2}{g-1}, \qquad \text{(by substitution in equations (2) and (3) and using Theorem 9)} \qquad ||$$

6. SUMMARY

As a result of a particular scaling of workload in a closed network of multiserver queues, six equivalent representations of an also scaled production function emerged. The new representations are useful to provide insight into understanding the production function. Some of the equivalent definitions of expected production were used to derive some additional relationships that are potentially useful for future research. In addition, some properties of balanced systems that may also prove useful in the study of both manufacturing and computer system performance are presented.

ACKNOWLEDGMENTS

The research reported here has been partially supported by the National Science Foundation under GRANT No. APR75 15256 and also by a grant from Ford Motor Company. The authors gratefully thank James J. Solberg of Purdue University for many useful discussions.

REFERENCES

BARD, YONATHAN, "Some Extensions to Multiclass Queueing Network Analysis," Performance of Computer Systems, M. Arato, A. Butrimenko, and E. Gelenbe (eds.), North Holland, Amsterdam (1979).

BASKETT, FOREST, CHANDY, K. MANI, MUNTZ, RICHARD R. and PALACIOS, FERNANDO G., "Open, Closed, and Mixed Networks of Queues with Different Classes of Customers," Journal of the Association for Computing Machinery, Vol. 22, No. 2, pp. 248-260 (April 1975).

BUZACOTT, JOHN A. and SHANTHIKUMAR, J. GEORGE, "Models for Understanding Flexible Manufacturing Systems," AIIE Transactions, Vol. 12, No. 4, pp. 339-349 (December 1980).

BUZEN, JEFFREY P., "Queueing Network Models of Multiprogramming," Ph.D. dissertation, Division of Engineering and Applied Physics, Harvard University, Cambridge MA (May 1971).

BUZEN, JEFFREY P., "Computational Algorithms for Closed Queueing Networks with Exponential Servers," Communications of the Association for Computing Machinery, Vol. 16, No. 9, pp. 527-531 (September 1973).

CAVAILLÉ, JEAN-BERNARD and DUBOIS, DIDIER, "Heuristic Methods Based on Mean-Value Analysis for Flexible Manufacturing Systems Performance Evaluation," Proceedings, IEEE Conference on Decision and Control, Orlando FL (December 1982).

KELLY, FRANK P., Reversibility and Stochastic Networks, John Wiley & Sons, New York NY (1979).

KLEINROCK, LEONARD, Queueing Systems, Volume 2: Computer Applications, John Wiley & Sons, New York NY (1976).

KOBAYASHI, HISASHI and REISER, MARTIN, "On Generalization of Job Routing Behavior in a Queueing Network Model," IBM Research Report RC 5252, IBM Thomas J. Watson Research Center, Yorktown Heights NY (February 1975).

REISER, MARTIN and KOBAYASHI, HISASHI, "Queueing Networks with Multiple Closed Chains: Theory and Computational Algorithms," IBM Journal of Research and Development, Vol. 19, No. 3, pp. 283-294 (May 1975).

REISER, MARTIN and KOBAYASHI, HISASHI, "Horner's Rule for the Evaluation of General Closed Queueing Networks" Communications of the Association for Computing Machinery, Vol. 18, No. 10, pp. 592-593 (October 1975).

REISER, MARTIN and KOBAYASHI, HISASHI, "On the Convolution Algorithm for Separable Queueing Networks," in Proceedings, International Symposium on Computer Performance Modeling, Measurement, and Evaluation, Harvard University, Cambridge MA, pp. 109-117 (1976).

ROSE, C. A., "Validation of a Queueing Model with Classes of Customers," in Proceedings, International Symposium on Computer Performance Modeling, Measurement, and Evaluation, Harvard University, Cambridge MA, pp. 318-325 (1976).

SOLBERG, JAMES J., "A Mathematical Model of Computerized Manufacturing Systems," in Proceedings, 4th International Conference on Production Research, Tokyo, Japan (August 1977).

STECKE, KATHRYN E., "Experimental Investigation of a Computerized Manufacturing System," Master's Thesis, School of Industrial Engineering, Purdue University, W. Lafayette IN 47907 (December 1977).

STECKE, KATHRYN E., "A Hierarchical Approach to Production Planning in Flexible Manufacturing Systems," in Proceedings, Twentieth Annual Allerton Conference on Communication, Control, and Computing, Monticello IL (October 6-8, 1982).

STECKE, KATHRYN E., "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems," Management Science, Vol. 29, No. 3, pp. 273-288 (March 1983).

STECKE, KATHRYN E. and MORIN, THOMAS L., "Optimality of Balanced Workloads in Flexible Manufacturing Systems," Working Paper No. 289, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI 48109 (January 1982).

STECKE, KATHRYN E. AND SOLBERG, JAMES J., "The CMS Loading Problem," Report No. 20, NSF Grant No. APR 74 15256, School of Industrial Engineering, Purdue University, W. Lafayette IN 47907 (February 1981a).

STECKE, KATHRYN E. and SOLBERG, JAMES J., "Loading and Control Policies for a Flexible Manufacturing System," International Journal of Production Research, Vol. 19, No. 5, pp. 481-490 (September-October, 1981b).

STECKE, KATHRYN E. and SOLBERG, JAMES J., "The Optimality of Unbalanced Workloads and Machine Group Sizes for Flexible Manufacturing Systems," Working Paper No. 290, Division of Research, Graduate School of Business Administration, The University of Michigan, Ann Arbor MI 48109 (January 1982).