PERFORMANCE '90
P.J.B. King, I. Mitrani and R.J. Pooley

Exact Solution for Networks of Parallel Queues with Finite Buffers

I. F. Akyildiz and N. M. van Dijk***

* School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332
U. S. A.

** Vrije University
Amsterdam
Netherlands

ABSTRACT

Networks of parallel queues with interdependent service capacities, finite buffers and accessibility constraints are studied. An invariance condition is provided in terms of the service and blocking protocols. It is shown that this condition guarantees a product form for the equilibrium state probabilities. The solution unifies various existing product form results such as for reversible networks with blocking, networks without blocking but station interdependent servicing and networks with exponential (e.g., FCFS) and nonexponential (e.g., Processor Sharing)) queues. It is shown that the equilibrium state probability distribution has the insensitivity property, i.e., it depends upon the service requirements only throughout their means.

Key Words: Product Form, Insensitivity, Parallel Queues, Finite Buffers,Blocking

1. Introduction

Product form results for queueing networks and their relationships with notions of partial balance have been extensively studied over the last two decades [5,6,7,8,12,16,18]. For exponential networks with a fixed routing, no blocking and station independent servicing the product form is a common feature. For networks with blocking or load dependent servicing the results are much more restrictive. The product form is generally restricted to exponential networks with finite queue size constraints provided the routing is completely reversible and the servicing is load-independent [1,2,11,16,19]. Although blocking results with nonreversible routing have been reported [11,12], various situations with capacity constraints remain open such as that with routing which is only partially reversible. Conversely, for exponential networks with fixed routing and no blocking, product form results have also been reported with station interdependent servicing provided the service rates at a particular queue are defined by a special functional form [7,16].

These product form results for exponential networks, remain valid for networks with non-exponential queues (insensitivity phenomenon) provided that at these queues a detailed notion of partial balance is satisfied per position or per job (local or job-local balance). These conditions are satisfied when the service discipline is symmetric [4,7,8,16] or satisfies a more general service invariance condition [11].

Akyildiz's work was supported in part by School of Information and Computer Science, ICS, of Georgia Tech and by the Air Force Office of the Scientific Research (AFOSR) under Grant AFOSR-88-0028.

The network model considered in this paper has not been covered by the above references because

i) it includes blocking due to common constraints of collections of parallel queues,
ii) it allows service interdependencies within such collections and
iii) it does not require reversibility of the routing all over but only where blocking can occur.

The literature on systems with parallel queues is rather extensive due to their practical interest but has been restricted to commonly shared pools with assumptions of Poisson input and exponential service [9,14,17,21]. Under these assumptions product form results have been established. In practice, however, exponential service is rather restrictive and input streams in closed systems are generally non-Poisson.

This paper aims to extend the above results to non-exponential service, non-Poisson input requirements and interdependencies of parallel queues both with respect to rejecting and servicing jobs. The main results of this paper are:

i) An insensitive product form expression
ii) A concrete blocking and service invariance condition
iii) A number of new product form examples with the novel aspects of:
   a) A general interdependent blocking of parallel queues
   b) A general interdependent servicing of parallel queues
   c) A reversible routing only where blocking can arise

The proof of the product form result is straightforward and self-contained and is based upon verifying particular balance conditions. The insensitivity is established using the notion of balance per job and an intermediate step with mixtures of Erlang distributions.

Although insensitivity results are well-known [20,23], combined results such as in this paper have not been reported explicitly before. Additionally, we are considering features like blocking and service interdependencies in the network model which was also not addressed in combined form before. Moreover, this paper provides a concrete condition in terms of systems protocol in contrast to abstract condition of reversibility requirements [16,19] or Poisson input assumption for blocking to take place [21] or separability assumptions [14,17]. This concrete condition unifies and extends product form examples [9,15,17,21,22] and provides product form blocking structures [1,2], in queueing networks with multiclass and parallel queue interdependencies. Such a condition has not been reported before.

The presentation is restricted to closed queueing networks. However, the extension to open queueing networks is straightforward.

The organization of this paper is as follows: Section 2 describes the various model protocols. The essential invariance condition for the blocking and servicing protocols is presented in section 3. The product form result is derived in section 4.

2. Network Description

The system consists of N stations having multiple servers and M fixed number of jobs. There are T possible job types. Each station s has $Q(s)$ parallel queues, for $s = 1,...,N$. A job entering station s requires service at queue $q(s,t)$ depending on its present type t. After completing service a job of type t at queue $q(s,t)$ of station s goes with probability $p_{s,s'}^{t,t'}$ to the queue $q(s',t')$ of the station s' and changes its type to t'. The job can be rejected by the destination station s' based upon the present job configuration at that station. This blocking and its protocol will be described in section 2.1. Various queues at a station provide service at interdependent service rates as will be described in section 2.2. The service allocation to the jobs at a queue is governed by a queueing discipline which will be described in section 2.3. In section 2.4 we specify the service distributions.

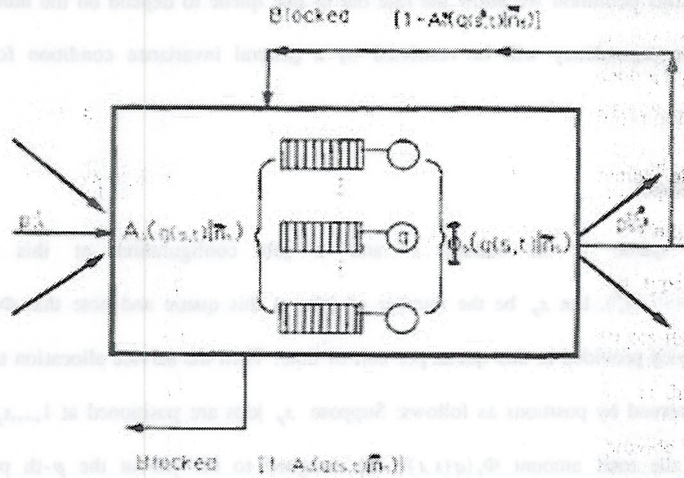

Figure 1. Structure of Station s

2.1. Blocking Protocol

Let $\bar{n}_s = (n_s^1, n_s^2, ..., n_s^T)$ denote the state that n_s^t jobs of type t are present at station s for $t = 1,2,...,T$ and $s = 1,...,N$. Suppose that a job of a type t' completes service at station s' and requests service at station s with its type number changed into t, while the state is $\bar{n}_s$. This request is accepted with probability $A_s(q(s,t) \mid \bar{n}_s)$ and the job is allocated to queue q at station s. If the request is rejected, the job retains its type number t' and has to restart a new service at station s' as a new arriving job. Note that the rejected job is always accepted by the source station. This blocking protocol is known as the rejection [1, 2] or Type III [3] or communication blocking [16,22] or also as repetitive service blocking [3].

One may observe that the function $A_s(q(s,t) \mid \bar{n}_s)$ allows the blocking probability of a type t job to depend not only on the total number of type t jobs (such as due to a capacity constraint at the corresponding queue $q(s,t)$) but also on the number of jobs of other types (such as due to a common in-or output channel). This multi-type dependent blocking will be restricted by a general invariance condition in section 3.

2.2. Service Rates

If station s is in state $\bar{n}_s = (n_s^1, n_s^2, \ldots, n_s^T)$ denoting that n_s^t jobs of type t are present at one of its queues, then the number of jobs at each individual queue is given, since each type-t job has a corresponding queue number $q(s,t)$.

The rate at which queue q provides service is given by:

$$\Phi_s(q(s,t) \mid \bar{n}_s)$$

where we assume that this function has either the value 0 if there are no jobs at queue q or has positive value otherwise.

Note that by this definition we allow the rate out of one queue to depend on the number of jobs at other queues. The server dependency will be restricted by a general invariance condition for blocking protocol defined in section 3.

2.3. Service Disciplines

Consider a queue q at station s and a job configuration at this station given by $\bar{n}_s = (n_s^1, n_s^2, \ldots, n_s^T)$. Let x_p be the number of jobs at this queue and note that $\Phi_s(q(s,t) \mid \bar{n}_s)$ is the total amount of service provided at this queue per unit of time. Then the service allocation to the individual jobs at this queue is governed by positions as follows: Suppose x_p jobs are positioned at $1,\ldots,x_p$. Then $\Gamma_{s,q}(p \mid \bar{n}_s)$ is the fraction of the total amount $\Phi_s(q(s,t) \mid \bar{n}_s)$ assigned to the job at the p-th position, $p = 1,\ldots,x_p$. $\delta_{s,q}(p \mid \bar{n}_s)$ is the probability that the last entered job at queue q from the jobs present has been assigned position p, $p = 1,\ldots,x_p$.

When a job at position p completes its service the jobs at positions $p+1,\ldots,x_p$ shift to positions $p,\ldots,x_p-1$. When a job is assigned position p, the jobs at previous positions $p,\ldots,x_p$ are shifted to positions $p+1,\ldots,x_p+1$. We assume that

$$\sum_p \Gamma_{s,q}(p \mid \bar{n}_s) = \sum_p \delta_{s,q}(p \mid \bar{n}_s) = 1 \tag{1}$$

We will distinguish two types of service disciplines. A discipline is said to be *non-symmetric* when it adopts the above description without further conditions. A discipline is said to be *symmetric* when in addition

$$\Gamma_{s,q}(p \mid \bar{n}_s) = \delta_{s,q}(p \mid \bar{n}_s) \qquad p = 1,\ldots,x_p \qquad \text{for all } \bar{n}_s \tag{2}$$

Let S be the set of all symmetric service disciplines [16]. Chandy and Martin [7] call them *station balancing*. Various disciplines can be parametrized in the above manner [16]. Most notably are the standard BCMP disciplines [5]:

- FCFS: First Come First Served ($\notin$ S)
- PS-1: Processor Sharing single server ($\in$ S)
- IS: Infinite Servers ($\in$ S)
- LCFS-PR: Last Come First Served Pre-emptive Resume ($\in$ S)

2.4. Distribution Functions

The service distribution of a job of class t at station s depends upon the service discipline of its queue $q(s,t)$ and has a distribution function of the form.

$$G_s^t = \begin{cases} E(1, \mu_{s,q}) & \text{for} \quad q(s,t) \notin S \\ \sum_{k=1}^{\infty} a_s^t(k) \; E(k, v_s^t) & \text{for} \quad q(s,t) \in S \end{cases} \tag{3}$$

where $E(k,\alpha)$ denotes an Erlang k-distribution with mean k/α and $a_s^t(k)$ denotes the probability that the distribution consists of k successive exponential phases with parameter v_s^t assuming $\sum_k a_s^t(k) = 1$. Hence,

$$\tau_s^t = \begin{cases} \frac{1}{\mu_{s,q}} & \text{for} \quad q(s,t) \notin S \\ \sum_{k=1}^{\infty} a_s^t(k) \; E(k, v_s^t) & \text{for} \quad q(s,t) \in S \end{cases} \tag{4}$$

is the mean service requirement of a type t-job at station s while

$$R_s^t(r) = \frac{[\sum_{k=r} a_s^t(k)]}{[v_s^t \; \tau_s^t]} \tag{5}$$

is known from the renewal theory as the stationary excess probability of "r" residual exponential phases up to a next renewal in a renewal process with renewal function G_s^t for $q(s,t) \in S$. Informally, the function in (3) requires all jobs at a non-symmetric queue to have an exponential service with one and the same parameter regardless of job type, while a job at a symmetric queue may have a general mixture of Erlang service distributions depending on its job type. The restriction to these mixtures will be used in section 4 to justify a Markovian analysis. The proof of our results will thus be established for these mixtures only. It is well-known, however, that any nonnegative probability distribution can be arbitrarily closely approximated by these mixtures (in the sense of weak convergence, [10]). Based upon standard weak convergence limit theorems for the probability measures of the sample paths on appropriate so-called D-spaces [4,13], the insensitivity result can therefore be extended to arbitrary service distributions.

3. Conditions

First it is to be noticed that the routing probabilities $p_{..}^{..}$, the possible changes of job types and the blocking functions $A(.|.)$ together with the blocking protocol will exclude certain state configurations which are described below. Let R be a set of all reachable configurations $\vec{N} = (\vec{n}_1, \ldots, \vec{n}_N)$ with a given starting configuration $\vec{N}^* = (\vec{n}_1^*, \ldots, \vec{n}_N^*)$ and exponential service times with unit mean at any queue for any job. We assume R to be irreducible. Throughout this paper, we will restrict our attention to configurations within R. We define a station configuration $\vec{n}_s$ admissible if there exists a configuration within R with $\vec{n}_s$ restricted to station s. We are now ready to present our conditions.

3.1. Partial Reversible Routing

The routing probabilities from one station to another are subject to a partial reversibility condition. Informally, it requires the routing to be reversible wherever jobs can be rejected, but it allows arbitrary fixed routing probabilities where jobs cannot be rejected. More precisely, without loss of generality, assume that there exists a unique probability distribution satisfying the following traffic equations:

$$\lambda_s^t = \sum_{s't'} \lambda_{s'}^{t'} \, p_{s's}^{t't} \qquad (s = 1,\ldots,N \; ; \; t \in T) \tag{6}$$

Then additionally we require the following *Partial Reversibility Condition:*

For any station s and type t such that for some admissible configuration $\vec{n}_s$ and queue $q = q(s,t)$:

$$A_s(q \mid \vec{n}_s) < 1 \tag{7}$$

we have

$$\lambda_s^t \, p_{ss'}^{tt'} = \lambda_{s'}^{t'} \, p_{s's}^{t't} \qquad \textit{for all } s't' \quad \textit{with} \quad p_{ss'}^{tt'} > 0 \quad \textit{or} \quad p_{s's}^{t't} > 0 \tag{8}$$

Note that the standard reversibility condition [16] requires equation (8) to hold for any (s,s',t,t'). While in our case it is required only for a subset satisfying equation (7). Note also that our partial reversibility condition, equation (8) has merely to do with routing in contrast to the quasi-reversibility [16].

The reason for including this partial rather than global reversibility condition is twofold:

i) It allows us to simultaneously analyze systems with as well as without blocking. Even without blocking the results of this paper are new as they involve a service interdependence of parallel queues.

ii) New examples with blocking but a global non-reversible structure can be covered. Blocking results in the literature either require reversibility all over the network [1,2,13,16,19] or provide a general non-reversibility routing condition but exclude for example First Come First Served station types [11].

3.2. Blocking and Service Invariance Condition

In order to present a general condition upon the interdependent blocking and servicing at a particular station we need to introduce some notation. We will focus on a fixed station s. For a vector $\bar{n} = (n^1, n^2, \ldots, n^T)$ with $n^1 + n^2 + \cdots + n^T = n$ denoting that n jobs are present at station s of which n^t are of type t, $t = 1,2,\ldots,T$. Let $\bar{T}(\bar{n})$ be the corresponding vector with the first n^{t_1} components equal to t_1, the first type number t in increasing order with $n^t > 0$, the next n^{t_2} equal to t_2, the second t with $n^t > 0$, etc. Conversely, for any given vector of job types $(j_1, j_2, \ldots, j_k)$ let $\bar{n}(j_1, \ldots, j_k)$ be the vector of corresponding numbers n^t of jobs of type $t = 1,2,\ldots$ Furthermore throughout for a given number $\bar{n}_s$ at station s, let $\bar{m}_s = (m_s^1, \cdots, m_s^{Q(s)})$ denote the corresponding queue sizes m_s^q at queue $q = 1,\ldots,Q(s)$ and let $m_s = m_s^1 + \cdots + m_s^{Q(s)} = n_s$ be the total number at station s.

Invariance Condition:

For any station s, any admissible vector $\bar{n}_s$ (at station s) and with $\bar{T}(\bar{n}_s)$ the corresponding queue size vector of size n, the product

$$\prod_{k=1}^{n} \frac{A_s(q(s,j_k) \mid \bar{n}(j_1,\ldots,j_{k-1}))}{\Phi_s(q(s,j_k) \mid \bar{n}(j_1,\ldots,j_{k-1},j_k))} \tag{9}$$

is invariant for all permutations

$$(j_1, \ldots, j_n) \in \bar{T}(\bar{n}_s) \tag{10}$$

This invariance product is denoted by $P_s(\bar{n}_s)$ for $n > 0$ while we introduce

$P_s(\bar{n}) = 1$ for $n = 0$

Informally, this condition requires that it does not matter in which order the jobs of the various types arrive if we consider $A(.|.) / \Phi(.|.)$ as state dependent arrival rate.

REMARKS.

i) Note that as we have required $\Phi_s(q(s,t)|\bar{n}) > 0$ for any queue q and $\bar{n}$, the functions $A_s(.|.)$ must necessarily be positive for reaching any admissible vector $\bar{n}_s$. We will have, however, $A_s(.|.) = 0$ when the acceptance of a next job would lead to a non-admissible state vector.

ii) (Decoupling). Clearly, the invariance condition is guaranteed by separately satisfying the invariance of the products

$$\prod_{k=1}^{n} A_s\left[q(s,j_k) \mid \bar{n}(j_1, \ldots, j_{k-1})\right] \tag{11}$$

and

$$\prod_{k=1}^{\hat{}} \Phi_s\left[q(s,j_k) \mid \bar{n}(j_1,\ldots,j_{k-1},j_k)\right] \tag{12}$$

for all permutations $(j_1,\ldots,j_k) \subseteq \bar{T}(\bar{n}_s)$. Although examples can be found for which (9) holds while (11) and (12) fail, the conditions (11) and (12) seem much more realistic as they decouple blocking and servicing. We will therefore restrict our examples given in section 5 to (11) and (12).

iii) (Convex Blocking). An important subclass of blocking satisfying the invariance condition (11) is obtained by assuming that $A_s(.|\bar{n}_s)$ depends upon $\bar{n}_s$ only by $\bar{m}_s$: the vector of queue sizes $m_1, m_2, \ldots$ by

$$A_s(q(s,t) \mid \bar{n}_s) = \begin{cases} 1 & \text{if} \quad \bar{m}_s + e_s^q \in B_s \\ 0 & otherwise \end{cases} \tag{13}$$

where $\bar{m}_s \pm e_s^q$ denotes the vector equal to $\bar{m}_s$ with one job more (+ sign) or less (− sign) at queue q and where B_s is a set such that

$$\bar{m}_s = (m_1,\ldots,m_{Q(s)}) \in B_s \longrightarrow \bar{m}_s - e_s^q = B_s \qquad (\text{for } all \ q) \tag{14}$$

that is, blocking arises only to prohibit departures from B_s where B_s satisfies (14). Due to [9,15] we call such blocking *coordinate convex*. In [9,15] this blocking type is considered for networks which consist of only one station. Additionally these references allow nonexponential service times and strictly Poisson arrival times. Here we allow queueing networks having more than one station with this type of blocking and with non-exponential service times and non-Poisson arrival times.

The verification of (11) is immediate as accessible states $\bar{n}_s$ are necessarily restricted to B_s so that the product (11) is equal to 1 for any accessible state $\bar{n}_s$ regardless of the chosen permutation.

4. Product Form Solution

In this section we will derive the insensitive product form results. At different stations, queues, positions, job-types and residual service amounts need to be specified, some notational complexity is unavoidable. Let

$$\bar{V} = [\, \bar{V}_1, \bar{V}_2, \cdots, \bar{V}_N \,]$$

with

$$\bar{V}_s = [\, \bar{Q}_1^s, \bar{Q}_2^s, \ldots, \bar{Q}_{Q(s)}^s \,] \qquad \text{for } (s = 1,\ldots,N)$$

where

$$\bar{Q}_s^q = [(t^1_{s,q}, r^1_{s,q}), \cdots, (t^{m_{s,q}}_{s,q}, r^{m_{s,q}}_{s,q})\,] \qquad \text{for} \quad (q = 1,\ldots,Q(s))$$

to denote for each station s and each queue q at this station that $m_{s,q}$ jobs are present at this queue of which the job at position p has a job-type number $t^p_{s,q}$ and a residual number of exponential service phases $r^p_{s,q}$.

Further, for a given state $\bar{V}$ let

$$\bar{V} + [s',q'] \; (t',p',r') - [s,q] \; (t,p,r)$$

be the state that differs from $\bar{V}$ in that the job at position p in queue q at station s with $t^p_{s,q} = t$ and $r^p_{s,q} = r$ has moved to position p' in queue q' at station s' with $t^{p'}_{s',q'} = t'$ and $r'_{s',q'} = r'$. Here it is to be noted that for a job at a non-symmetric queue the number of residual exponential service phases is necessarily equal to 1. Similarly, for a given vector $\bar{n}_s = (n_s^1 , n_s^2 , \cdots , n_s^T)$. Let

$$\bar{n}_s \pm e_s^t$$

denote the vector that differs from $\bar{n}_s$ in one job more (+ sign) or less (- sign) from type t. Finally, recall that $q(s,t)$ is the queue number for a job type t at station s and that S denotes the set of all symmetric queues.

Now we are ready to present the two main theorems of this paper. The first theorem is the key theorem which contains more detailed information than needed. The second theorem is a more practical consequence of theorem 1 and shows the insensitivity property. Now let us assume that there exists a unique stationary distribution $\Pi(.)$ for the $\bar{V}$ process restricted to an irreducible set $\bar{R}$.

Theorem 1. With C as a normalization constant and $P_s(.)$ defined by (9) we have

$$\Pi(\bar{V}) = C \prod_{s=1}^{N} P_s(\bar{n}_s) \left\{ \prod_{t=1}^{T} (\lambda_s^t)^{n_s^t} \right\} \left\{ \prod_{q \notin S} \left(\frac{1}{\mu_{s,q}}\right)^{m_{s,q}} \right\} \left\{ \prod_{q \in S} \prod_{p=1}^{m_{s,q}} \tau_s^{t^p_{s,q}} R_s^{t^p_{s,q}} (r^p_{s,q}) \right\} \quad (15)$$

Before presenting the proof, let us give a direct consequence of this theorem. Now let

$$\bar{N} = (\bar{n}_1 , \bar{n}_2 , ... \bar{n}_N)$$

where

$$\bar{n}_s = (n_s^1 , n_s^2 , ... n_s^T)$$

to denote that n_s^t jobs of type t are present at station s for all possible t and s. By standard calculus or from renewal theory notice that for any s, t:

$$\sum_{r=1}^{\infty} R_s^t(r) = 1 \quad (16)$$

Therefore by summing over all possible numbers $r = r^p_{s,q}$ of residual exponential phases for any p,s,q and by disregarding the specification of positions p for the jobs individually, we can conclude the following main result from theorem 1. It shows that the steady state joint queue size vector has a product form and is insensitive for symmetric queues.

Theorem 2. With C as a normalization constant, the steady state distribution for admissible states is given by:

$$\Pi(\bar{N}) = C \prod_{s=1}^{N} P_s(\bar{n}_s) \left\{ \prod_{t=1}^{T} (\lambda_s^t)^{n_s^t} \right\} \left\{ \prod_{q \notin S} \left(\frac{1}{\mu_{s,q}}\right)^{m_{s,q}} \right\} \left\{ \prod_{[t \,|\, q(s,t) \in S]} (\tau_s^t)^{n_s^t} \right\} \quad (17)$$

In the following we present the proof of theorem 1.

Proof.

By virtue of the Markovian structure of the $\bar{V}$ process it is sufficient to verify the global balance (or equilibrium) equations. These require that the total rate (or probability flow) out of any state due to a change at any of the queues $q = 1,\dots,Q(s)$ for $s = 1,\dots,N$, is equal to the rate into that state due to a change at any of these queues. However, this in turn is guaranteed if for each queue $q \in \{1,\dots,Q(s)\}$ and for $s = 1,\dots,N$ individually, we can establish:

The rate out of any state due to a change at queue q $=$ (18)
The rate into that state due to a change at queue q

In the following we will verify (18) for non-symmetric and symmetric queues, respectively. In particular, for symmetric queues we will even establish (18) by

The rate out of any state due to a change at position p at queue q $=$ (19)
The rate into that state due to a change at position p at queue q

Now consider a fixed state $\bar{V}$ and station $s \in \{1,\dots,N\}$

i) *Nonsymmetric Queue.* Consider a fixed queue $q \in \{1,\dots,Q(s)\}$ and for convenience let

$$t(p) = t_{s,q}^{p} \qquad \text{for} \qquad p = 1,\dots,n,$$

The rate out of state $\bar{V}$ due to a change at queue q is given by

$$\Pi(\bar{V}) \;\; \Phi_s(q(s,t)|\bar{n}_s) \;\; \mu_{s,q} \;\; [\; \sum_{p} \Gamma_{s,q}(p|\bar{n}_s)] \tag{20}$$

The rate into this state due to a change at q is equal to

$$\sum_{p} \Bigg\{ \sum_{s',t'} \; \sum_{p'} \; \Pi\Big[\bar{V} + [s',q(s',t')]\;(t',p',1) - [s,q]\;(t(p),p,1)\Big] \cdot \tag{21}$$

$$\cdot \Bigg\{ \Phi_{s'}\Big[q(s',t') \mid \bar{n}_{s'} + e_{s'}^{t'}\Big] \cdot \Gamma_{s',q(s',t')}(p'|\bar{n}_{s'}' + e_{s'}^{t'}) \cdot v_{s'}^{t'} \cdot p_{t',s}^{t',t(p)} \Bigg\} \cdot \delta_{s,q}(p|\bar{n}_s) \cdot A_s(q(s,t)|\bar{n}_s - e_s^{t(p)}) \Bigg\}$$

$$+ \sum_{p} \Bigg\{ \sum_{p'} \Pi\Big[\bar{V} + [s,q](t(p),p',1) - [s,q]\,(t(p),p,1)\Big]\,\mu_{s,q} \cdot$$

$$\cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \Gamma_{s,q}(p'|\bar{n}_s) \cdot \Bigg[\sum_{s',t'} p_{t,s}^{t(p),s'} \;[\, 1 - A_{s'}\Big[q(s',t')|\bar{n}_{s'}\Big]]\Bigg] \; \delta_{s,q}(p|\bar{n}_s) \Bigg\}$$

where

$$v_{s'}^{t'} = \mu_{s',q(s',t')} \qquad \text{for} \;\; q(s',t') \notin S$$

Now first recall that

$$A_s(q(s,t)|\bar{n}_s - e_s^{t(p)}) \; > \; 0$$

for any admissible $\bar{n}_s$ and q as we have assumed $\quad \Phi_s(q(s,t)|\bar{n}_s) > 0$

provided $n_s > 0$.

Similarly

$$\Phi_{s'}(q(s',t')|\bar{n}_{s'} + e_{s'}^{t'}) > 0$$

for any admissible $\bar{n}_{s'} + e_{s'}^{t'}$.

As a result, by substituting (15) and using the invariance of the product (17) for expression (18) we conclude that for any admissible state $\bar{V}$ and admissable state $\left[\bar{V} + [s',q(s',t')]\ (t',p',1) - [s,q](t(p),p,1)\right]$ with $s' \neq s$.

$$\Pi\left[\bar{V} + [s',q(s',t')](t',p',1) - [s,q]\,(t(p),p,1)\right] = \tag{22}$$

$$\Pi(\bar{V}) \cdot \left[\frac{A_{s'}(q(s',t') \mid \bar{n}_{s'})}{A_s(q(s,t) \mid \bar{n}_s - e_s^{t(p)})}\right] \cdot \left[\frac{\Phi_s(q(s,t)|\bar{n}_s)}{\Phi_{s'}(q(s',t')|\bar{n}_{s'} + e_{s'}^{t'})}\right] \cdot$$

$$\cdot \left[\frac{\lambda_{s'}^{t'}}{\lambda_s^{t(p)}}\right] \left\{1_s\left[q(s',t')\right] \cdot \mu_{s,q} \cdot \tau_{s'}^{t'} \cdot R_{s'}^{t'}(1) + \left[1 - 1_s\left[q(s',t')\right]\right] \frac{\mu_{s,q}}{\mu_{s',q(s',t')}}\right\}$$

where

$$1_s(q) = \begin{cases} 1 & \text{for} \quad q \in S \\ 0 & \text{for} \quad q \notin S \end{cases} \tag{23}$$

Furthermore also by (15) it is valid that

$$\Pi\left[\bar{V} + [s,q](t(p),p',1) - [s,q](t(p),p,1)\right] = \Pi(\bar{V}) \tag{24}$$

for all p, p'.

By substituting (22) and (24) into (21) and using

$$\sum_{p'} \Gamma_{s',q}(p',1) = \sum_{p'} \Gamma_{s,q}(p',1) = 1$$

as in (1) and recalling

$$\nu_{s'}^{t'} = \mu_{s',q(s',t')} \qquad \text{for} \quad q(s',t') \in S$$

we can then rewrite (21) as

$$\Pi(\bar{V}) \cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \mu_{s,q} \cdot \sum_p \delta_{s,q}(p|\bar{n}_s) \cdot \tag{25}$$

$$\left\{\frac{1}{\lambda_s^{t(p)}} \cdot \left[\sum_{s' \in S, t'} p_{s't'}^{t(p)} \cdot \lambda_{s'}^{t'} \cdot A_{s'}\left[q(s',t')|\bar{n}_{s'}\right] \cdot\right.\right.$$

$$\sum_{s' \in S, t'} p_{s't'}^{t(p)} \cdot \lambda_{s'}^{t'} \cdot A_{s'}\left[q(s',t')|\bar{n}_s\right] \cdot \nu_{s'}^{t'} \cdot \tau_{s'}^{t'}\, R_{s'}^{t'}(1)$$

$$\left.\left. + \sum_{s',t'} p_{s't'}^{t(p)s} \cdot \left\{1 - A_{s'}\left[q(s',t')|\bar{n}_{s'}\right]\right\}\right]\right\}$$

Noting that
$$R_{s'}^{t'}(1) = \frac{1}{v_{s'}^{t'}\,\tau_{s'}^{t'}} \tag{26}$$

and by virtue of (4) we thus obtain for (21)

$$\Pi(\bar{V}) \quad \Phi_s(q\,|\bar{n}_s)\,\mu_{s,q} \sum_{p} \delta_{s,q}(p\,|\bar{n}_s) \cdot \tag{27}$$
$$\frac{1}{\lambda_s^{t(\varphi)}} \cdot \left\{ \sum_{s't'} \lambda_{s'}^{t'}\, p_{s't'}^{t(\varphi)} \cdot A_{s'}\left[q(s't')|\bar{n}_{s'}\right] + \sum_{s't'} \cdot \lambda_s^{t(\varphi)}\, p_{s't'}^{t(\varphi)}\, \left[1 - A_{s'}(q(s't')|\bar{n}_{s'})\right] \right\}$$

Now the partial reversibility conditions (7 and 8) need to be taken into account. When

$$A_{s'}(q(s't')\,|\,\bar{n}_{s'}) = 1 \qquad for\ all \quad s't'$$

with $p_{s't'}^{t(\varphi)} > 0$ or $p_{s't'}^{t(\varphi)} > 0$, the second term within the bracket of (27) is equal to 0 and the equality of (20) and (27) and thus (21) directly follow from the traffic equations (6) and (1).

When however

$$A_{s'}(q(s't')\,|\,\bar{n}_{s'} + e_{s'}^{t'}) < 1 \qquad for \quad some \quad s't'$$

with $p_{s't'}^{t'} > 0$ or $p_{s't'}^{t'} > 0$ and assuming $\bar{n}_{s'} + e_{s'}^{t'}$ to be admissible, the partial reversibility conditions (7 and 8) reduce equation (27) to

$$\Pi(\bar{V}) \cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \mu_{s,q} \cdot \sum_{p} \delta_{s,q}(p\,|\bar{n}_s) \cdot \tag{28}$$
$$\frac{1}{\lambda_s^{t(\varphi)}} \left\{ \sum_{s't'} \lambda_s^{t(\varphi)}\, p_{s't'}^{t(\varphi)} \left[A_{s'}\left[q(s't')|\bar{n}_{s'}\right] + \left\{1 - A_{s'}\left[q(s't')|\bar{n}_{s'}\right]\right\}\right]\right\}$$

so that also in this case equality of (20 and 21) is proven by the virtue of $\sum_{s't'} p_{s't'}^{t'} = 1$ and equation (1).

As the state $\bar{V}$, station s and queue q of station s were arbitrarily chosen, we have hereby verified (18) for any non-symmetric queue q.

ii) *Symmetric Queue.* Now consider a fixed queue $q \in \{1,...,Q(s)\}$ as well as a fixed position $p \in \{1,...,n_s\}$ at this queue. For convenience let $t = t_{s,q}^p$ and $r = r_{s,q}^p$. In the following we verify (18) by means of (19).

The rate out of state $\bar{V}$ due to a change at position p of queue q is given by

$$\Pi(\bar{V}) \;\; \Phi_s(q(s,t)|\bar{n}_s) \;\; \Gamma_{s,q}(p\,|\bar{n}_s) \;\; v_s^t \tag{29}$$

The rate into state $\bar{V}$ due to a change at position p of queue q equals

$$\Pi\left[\bar{V} + [s,q]\,(t,p,r+1) - [s,q](t,p,r)\right] \cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \Gamma_{s,q}(p\,|\bar{n}_s) \cdot v_s^t + \tag{30}$$
$$\sum_{s't'} \sum_{p'} \Pi\left[\bar{V} + [s',q(s',t')]\,(t',p',1) - [s,q]\,(t,p,r)\right] \cdot$$
$$\left\{\Phi_{s'}\left[q(s't')\,|\,\bar{n}_{s'} + e_{s'}^{t'}\right] \cdot \Gamma_{s',q(s',t')}(p'|\bar{n}_{s'} + e_{s'}^{t'}) \cdot p_{s't}^{t's} \cdot A_s\left[q(s,t)|\bar{n}_s - e_s^t\right] \cdot \delta_{s,q}(p\,|\bar{n}_s)\right\} \cdot a_s^t(r) +$$

$$\sum_{r'} \Pi\left[\bar{V} + [s,q](t,p',1) - [s,q](t,p,r)\right]$$

$$\left\{ \Phi_s(q(s,t)|\bar{n}_s) \cdot \Gamma_{s,q}\left[p'|\bar{n}_s\right] \cdot v_s^t \cdot \sum_{s',t'} p_{s,t}^{s',t'} \left[1 - A_{s'}\left[q(s',t') \mid \bar{n}_{s'}\right]\right] \right\} \cdot \delta_{s,q}(p|\bar{n}_s) \cdot a_s^t(r)$$

where as before $v_{s'}^{t'} = \mu_{s',q(s',t')}$ for $q(s',t') \notin S$.

By taking the remark made after equation (21) and also equation (23) into account we conclude similarly to (22) and (24) that:

$$\Pi\left[\bar{V} + [s,q]\ (t,p,r+1) - [s,q]\ (t,p,r)\right] = \Pi(\bar{V}) \left[\frac{R_s^t(r+1)}{R_s^t(r)}\right] \tag{31}$$

$$\Pi\left[\bar{V} + [s,q]\ (t,p',1) - [s,q]\ (t,p,r)\right] = \Pi(\bar{V}) \left[\frac{R_s^t(1)}{R_s^t(r)}\right] \tag{32}$$

$$\Pi\left[\bar{V} + [s',q(s',t')]\ (t',p',1) - [s,q](t,p,r)\right] = \Pi(\bar{V}) \left[\frac{\lambda_{s'}^{t'}}{\lambda_s^t} \right] \cdot \tag{33}$$

$$\left[\frac{A_{s'}\left[q(s',t') \mid \bar{n}_{s'}\right]}{A_s\left[q(s,t) \mid \bar{n}_s - e_s^t\right]}\right] \cdot \left[\frac{\Phi_s(q(s,t)|\bar{n}_s)}{\Phi_{s'}\left[q(s',t')|\bar{n}_{s'} + e_{s'}^{t'}\right]}\right] \cdot \left\{\frac{1_s(q(s',t')) \cdot \tau_{s'}^{t'} \cdot R_{s'}^{t'}(1)}{\tau_s^t \cdot R_s^t(r)} + \frac{[1 - 1_s(q(s',t'))]}{[\mu_{s',q(s',t')} \cdot \tau_s^t \cdot R_s^t(r)]}\right\}$$

By substituting (31), (32) and (33) into (30) and using $\sum_{p'} \Gamma_{s,q}(p',1) = \sum_{p'} \Gamma_{s,q}(p',1) = 1$ as by (1) and recalling

$$v_{s'}^{t'} = \mu_{s',q(s',t')} \qquad \text{for} \quad q(s',t') \in S$$

Similarly to (24) we can rewrite (30) as:

$$\Pi(\bar{V}) \cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \Gamma_{s,q}(p|\bar{n}_s) \cdot \frac{v_s^t}{[R_s^t(r)]} \cdot \tag{34}$$

$$\left\{ R_s^t(r+1) + \left[\frac{\delta_{s,q}(p|\bar{n}_s)}{\Gamma_{s,q}(p|\bar{n}_s)}\right] \cdot \frac{a_s^t(r)}{[v_s^t\ \tau_s^t]} \cdot \left[\sum_{s' \in S,t'} p_{s',t'}^{s,t} \cdot \lambda_{s'}^{t'} \cdot A_{s'}\left[q(s',t') \mid \bar{n}_{s'}\right] \cdot \right.\right.$$

$$\cdot \sum_{s' \in S,t'} p_{s',t'}^{s,t} \cdot \lambda_{s'}^{t'} \cdot A_{s'}\left[q(s',t')|\bar{n}_{s'}\right] \cdot v_{s'}^{t'} \cdot \tau_{s'}^{t'} \cdot R_{s'}^{t'}(1) +$$

$$\left.\left. \sum_{s',t'} \lambda_s^t \cdot p_{s,t}^{s',t'} \cdot \frac{\left[1 - A_{s'}\left[q(s',t')|\bar{n}_{s'}\right]\ v_s^t\ \tau_s^t\ R_s^t(1)\right]}{\lambda_s^t} \right] \right\}$$

The symmetry condition (2) and expression (26) reduce this expression (34) to:

$$\Pi(\bar{V}) \cdot \Phi_s(q(s,t)|\bar{n}_s) \cdot \Gamma_{s,q}(p|\bar{n}_s) \cdot \frac{v_s^t}{R_s^t(r)} \cdot \left\{ R_s^t(r+1) + \frac{a_s^t(r)}{[v_s^t \cdot \tau_s^t]} \cdot \frac{1}{\lambda_s^t} \cdot \right. \tag{35}$$

$$\left.\left[\sum_{s',t'} \lambda_{s'}^{t'} \cdot p_{s',t'}^{s,t} \cdot A_{s'}\left[q(s',t')|\bar{n}_{s'}\right] \sum_{s',t'} \cdot \lambda_s^t \cdot p_{s,t}^{s',t'} \cdot \left\{1 - A_{s'}\left[q(s',t')|\bar{n}_{s'}\right]\right\}\right]\right\}$$

with

$$R_s^t(r) = R_s^t(r+1) + \frac{a_s^t(r)}{[v_s^t \ \ \tau_s^t]} \tag{36}$$

According to (5), equality of (29) and (35) (and thus (30)) now follows similarly to that of (20) and (27) as based upon the traffic equation (6) and the partial reversibility condition (7 and 8). We have thus verified (19) for any p so that (18) is also secured. With $\bar{V}$ station s and queue q arbitrarily chosen, (18) is thus guaranteed also for any symmetric queue which completes the proof of theorem 1.

REMARKS.

i) A similar product form expression can be given that only concerns the total queue length at each station by averaging (17) over the various job types.

ii) The results are directly applicable to open networks. To this end, one only needs to adjust the traffic equations (6) to include exterior arrivals and departures. The details are standard and therefore omitted.

5. Conclusion

A product form expression is established for queueing networks of parallel queues with interdependent servicing and blocking. The expression unifies various existing product form results such as for reversible networks with blocking, networks without blocking but station interdependent servicing and networks with exponential (e.g., FCFS) and nonexponential (e.g., Processor Sharing) queues, and also allows examples with non-reversible routing and blocking. A sufficient condition for this product form to hold is given in concrete terms of servicing and of blocking functions. The product form is insensitive to service distributional forms at symmetric queues. The proof is notationally complex but conceptually straightforward and self-contained as based upon different partial balance notions. Variations such as zero service times and modified blocking protocols can easily be built in. In the future it would be interesting to derive an efficient computational algorithm for the normalization constant C and the formulas for performance measures as in [2].

6. References

1. Akyildiz, I.F. and von Brand H., "Exact Solutions for Open, Closed and Mixed Queueing Networks with Rejection Blocking", *Theoretical Computer Science* Journal, North-Holland, May 1989, pp. 203-219.

2. Akyildiz, I.F. and von Brand H., "Computational Algorithms for Networks of Queues with Rejection Blocking", *ACTA Informatica*, Springer Verlag, Vol. 26, July 1989, pp. 559-576.

3. Akyildiz, I. F. and Perros, H. G., "Queueing Networks with Finite Buffers", *Special Issue in Performance Evaluation Journal*, Vol. 10, No. 4, pp. 149-151, December 1989.

4. Barbour, A., "Networks of Queues and the Method of Stages", *Adv. Appl. Prob.*8, pp. 584-591, 1976.

5. Baskett, F., Chandy, M., Muntz, R. and Palacios, J., "Open, Closed and Mixed Networks of Queues with Different Classes of Customers", *J.A.C.M.*, Vol. 22, April 1975, pp. 248-260.

6. Chandy, K.M., Howard, J.H. and Towsley, D.F., "Product Form and Local Balance in Queueing Networks", *J.A.C.M.* 24, pp. 250-263, 1977.

7. Chandy, K.M. and Martin, A.J., "A Characterization of Product-form Queueing Networks", *J.A.C.M.*, Vol. 30, pp. 286-299, 1983.

8. Cohen, J.W., "The Multiple Phase Service Network with Generalized Processor Sharing", *Acta Informatica*, 12, pp. 245-284, 1979.

9. Foschini, G.J. and Gopinath, B., "Sharing Memory Optimally", *IEEE Trans. Comm.*, 31, pp. 352-359, 1983.

10. Hordijk, A. and Schassberger, R., "Weak Convergence of Generalized Semi-Markov Processes", *Stochastic Process. Appl.*, 12, pp. 271-291 1982.

11. Hordijk, A. and van Dijk, N.M., "Networks of Queues with Blocking", Performance '81 (ed. K.J. Kylstra). *North Holland*, pp. 51-65, 1981.

12. Hordijk, A. and van Dijk, N.M., "Adjoint Processes, Job-Local-Balance and Insensitivity of Stochastic Networks", *Bull. 44th Session Int. Stat. Inst.*, vol. 50, pp. 776-788, 1982.

13. Hordijk, A. and van Dijk, N.M., "Networks of Queues. Part I: Job-Local-Balance and the Adjoint Process. Part II: General Routing and Service Characteristics", Lecture Notes in Control and Informational Sciences, *Springer Verlag*, Vol. 60, 158-205, 1983.

14. Kamoun, F. and Kleinrock, L., "Analysis of Finite Storage in a Computer Network Node Environment Under General Traffic Conditions", *IEEE Trans. Comm.* 28, 992-1003, 1980.

15. Kaufman, J., "Blocking in a Shared Resource Environment", *IEEE Trans. Comm.* 29, 1474-1481, 1981.

16. Kelly, F.P., "Reversibility and Stochastic Networks", *Wiley*, 1979.

17. Lam, S.S. and Reiser, M., "Congestion Control of Store and Forward Networks by Input Buffer Limits", *IEEE Trans Comm.* 27, pp. 127-133, 1979.

18. Noetzel, A.S., "A Generalized Queueing Discipline for Product Form Network Solutions. *Journal of the ACM* 26(4), pp. 779-793, October 1979.

19. Pittel, B., "Closed Exponential Networks of Queues with Saturation. The Jackson-Type Stationary Distribution and its Asymptotic Analysis", *Math. Oper. Res.*, pp. 357-378, 1979.

20. Schassberger, R. , "The Insensitivity of Stationary Probabilities in Networks of Queues", *Adv. Appl. Probl.* 10, 906-912, 1978.

21. Swiderski, J., "Unified Analysis of Local Flow Control Mechanisms in Message-Switched Networks", *IEEE Trans. Comm.* 32, 1286-1293, 1984.

22. Van Dijk, N.M. and Tijms, H.C., "Insensitivity in Two-Node Blocking Models with Applications", Teletraffic Analysis and Computer Performance Evaluation, *North Holland*, 329-340, 1986.

23. Whittle, P., Partial Balance and Insensitivity", *J. Appl. Prob.* 22, 168-176, 1985.