



Series-parallel fork-join queueing networks and their stochastic ordering

François Baccelli, W.A. Massey

► To cite this version:

François Baccelli, W.A. Massey. Series-parallel fork-join queueing networks and their stochastic ordering. RR-0534, INRIA. 1986. inria-00076020

HAL Id: inria-00076020

<https://inria.hal.science/inria-00076020v1>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
BP 105
78153 Le Chesnay Cedex
France
tel (1) 39 63 55 11

Rapports de Recherche

N° 534

**SERIES - PARALLEL,
FORK - JOIN QUEUEING
NETWORKS AND
THEIR STOCHASTIC
ORDERING**

François BACCELLI
William A. MASSEY

Juin 1986

**RESEAUX SERIE PARALLELE DE FILES D'ATTENTE AVEC BRANCHEMENT
ET JONCTION ET LEUR ORDONNANCEMENT STOCHASTIQUE**

François Baccelli
INRIA
Domaine de Voluceau
Rocquencourt - BP 105
78153 Les Chesnay Cédex
FRANCE

William A. Massey
AT&T Bell Laboratories
Murray Hill, New Jersey
07974
USA

RESUME

On définit une nouvelle classe de réseaux de files d'attente obtenue en combinant des réseaux en série et des réseaux avec branchement et jonction de tâches. Des bornes sont obtenues sur les statistiques des temps de traversée de ces réseaux grâce à l'extension du schéma de Loynes et du théorème d'ordonnancement convexe. L'étude de tels réseaux est motivée par les mécanismes de synchronisation rencontrés en algorithmique parallèle et en productique.

ABSTRACT

We define a new class of queueing systems that are formed by combining series and Fork-Join networks. For such networks, we derive bounds on the moments of the total delay to traverse such networks, for both transient and stationary case. We accomplish this by generalizing both Loynes increasing schema and the convex ordering theorem. These results have applications to flexible manufacturing and parallel processing.

Series-Parallel, Fork-Join Queueing Networks and Their Stochastic Ordering

*François Baccelli**
William A. Massey

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

1. INTRODUCTION

The class of queueing networks that we consider in the present paper is, roughly speaking, the "product" of two simpler classes. The first is the class of series networks obtained when concatenating any number, say K of FIFO queues linearly [5]. These networks are fed by an external input process. Each external arrival epoch creates an arrival in the first queue and each departure from the n -th queue for $n \leq K - 1$, creates an arrival to the $n + 1$ -th queue. The network has a single output stream which is the output process of the K -th queue.

The second class consists of the "Fork-Join networks" (see [6] and [7]). They are obtained by arranging K FIFO queues in parallel. These networks are also fed by a single external arrival stream and produce a single output stream. The "fork" occurs when every arrival epoch of the external stream triggers K simultaneous arrivals of customers in the K parallel queues. Every individual queue processes customers entering its buffer on a FIFO basis. Once serviced, a customer leaves this queue and enters the *join buffer*, a common buffer which is given in addition to the K parallel queues. As soon as all the K customers that were triggered by an earlier external arrival epoch reach this join buffer, they all vanish simultaneously. This is the "join" event or an epoch for the external departure process.

The queueing models we define here arise in many areas:

- Flexible Manufacturing

A set of machines can concurrently process items arriving on a conveyor. Here the simultaneous work demands are generated by the arrival of the items. The departure of an item (from this set of machines to another set) takes place when all the work

* INRIA, Domaine de Voluceau Rocquencourt, B.P. 105, 78153 Le Chesnay Cedex, France

demands it created are complete.

- Parallel Processing

A set of processors has to execute, independently, parallel subroutines created simultaneously by a "co-begin" primitive and all the return arguments of these subroutines are needed to continue the execution of the parallel program (the "co-end" primitive).

Apart from the subclass of Jackson series networks, the type of queueing networks we consider here remain basically unsolved. It can be shown that the "synchronizations" destroy all nice properties like insensitivity or product form, so that every problem becomes particular and computationally hard. For instance, in the particular case of fork-join networks, only the two-dimensional problem has received attention up to now (see [6] and [8]).

Several classes of stochastic ordering principles have been considered in the queueing literature. It was shown for instance, that an increased input or decreased output intensity leads to higher moments of the waiting or response times for wide classes of queueing systems (see [1]). Another type of ordering comes from the idea that an increased variability of either the input or the service statistics should also lead to higher waiting or response times. This has been discussed by several authors in the context of isolated queues (see [2], [3], and [4]). The latter ordering principle was used in [7] to compare the moments of the delays experienced by customers traversing fork-join networks to the related moments of product form networks. Both upper and lower bounds were derived from this comparison, which holds for all dimensions.

The aim of this paper is to extend the scope of this ordering and bounding technique to the larger class of series-parallel, fork-join networks which are rigorously defined in Section 2. The stochastic ordering theorem is proven in Section 3, while Section 4 is devoted to the derivation of bounds of practical interest for particular networks.

2. DEFINING THE NETWORKS

Consider the service times of the successive customers entering a FIFO single server queue to be given $\{\sigma_n\}_{n \geq 0}$, $\sigma_n \in R^+$. For any external arrival pattern $a_0 \leq a_1 \leq \dots \leq a_n \leq \dots$, $a_n \in R^+$, the queueing mechanism fully determines an external output pattern $b_0 \leq b_1 \leq \dots \leq b_n \leq \dots$, $d_n \in R^+$ given by the formula

$$(2.1) \quad b_n = a_n + W_n + \sigma_n, \quad n \geq 0,$$

where W_n is obtained by the induction

$$(2.2) \quad \begin{cases} W_{n+1} = \max(0, W_n + \sigma_n + a_n - a_{n+1}) \\ W_0 = 0. \end{cases}$$

Let us denote this transformation mechanism by the symbolic formula:

$$(2.3) \quad \{b_n\} = f(\{a_n\}).$$

Consider now a class C of networks with FIFO queues and the following characteristics:

- There is a unique external arrival stream of customers and a unique external departure stream.
- The queues are indexed on a partially ordered set (depending upon the network) say B for network β .
- When considering the service times of the successive customers entering queue j , say $\{\sigma_n^j\}_{n \geq 0}$, as given for all j in B , the knowledge of an arrival pattern $\{a_n\}_{n \geq 0}$ fully determines the departure pattern by a formula

$$(2.4) \quad \{d_n\} = \phi(\{a_n\}).$$

The class C is now defined recursively as follows:

- It contains all the networks formed of one FIFO queue.
- If γ and γ' are two networks of C both the concatenation of γ and γ' and the parallelization of γ and γ' belong to C .

The *concatenation* of γ and γ' , defined by the symbolic formulas

$$(2.5) \quad \{d_n\} = \phi(\{a_n\}), \quad \{d_n\} = \phi'(\{a_n\}),$$

is a new network of queues containing the queues of γ and γ' ordered in such a way that the partial orders of γ and γ' are respected and that the queues of γ' are all "larger" than those of γ . The new external output process is the external output process of γ' obtained when feeding γ' with the output process of γ , so that the symbolic formula for the new network is

$$(2.6) \quad \{d_n\} = \phi'(\phi(\{a_n\})).$$

The *parallelization* of γ and γ' is a new network containing the queues of γ and γ' in a new order compatible with the initial orders and such that for any pair of queues $j \in \gamma$ and $j' \in \gamma'$, j and j' are not comparable. The external output process of the new network is the one obtained when taking the external input processes of γ and γ' both equal to the external input of the new network and when triggering the n -th departure date of the new network by the later of the n -th departure dates of γ and γ' . The symbolic formula of the new network is hence given by

$$(2.7) \quad \{d_n\} = \max(\phi(\{a_n\}), \phi'(\{a_n\})).$$

where the max is taken coordinatewise. An example of such a network is given in figure 1:

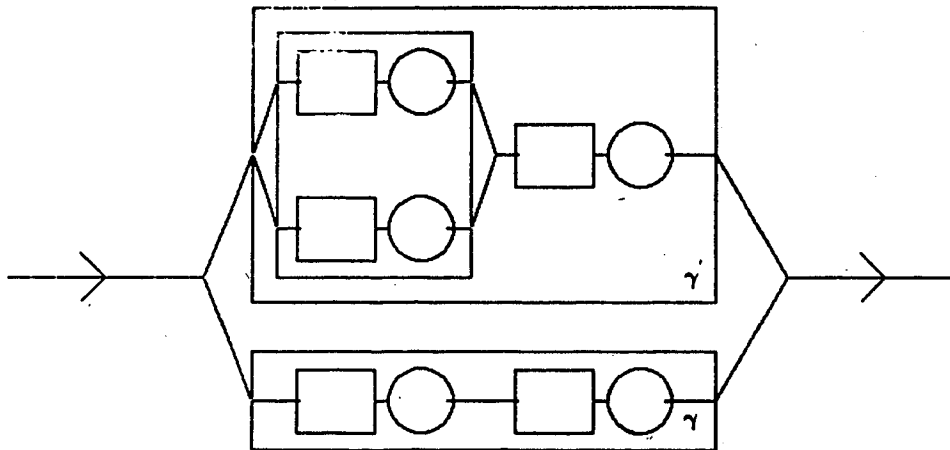


Figure 1

The following two lemmas will be needed in Section 3.

Lemma 1. *Let β be a network in C . The queues indexed with minimal elements of B all have their n -th arrival date triggered by the n -th external arrival date of β . The n -th external departure date is triggered by the latest of the n -th departure dates from the queues indexed with maximal elements of B .*

Proof: A network formed with one queue satisfies both properties. So do the concatenation and the parallelization of γ and γ' if both γ and γ' satisfy these properties. ■

For $\beta \in C$ and $j \in B$, let $\pi(j)$ be the set of immediate predecessors of j in B :

$$(2.8) \quad \pi(j) = \{i \in B \mid i < j \text{ and there exists no } k \text{ such that } i < k < j\}$$

Lemma 2. *If $\pi(j) = \emptyset$, the n -th arrival to queue j coincides with the n -th external arrival date. If $\pi(j) \neq \emptyset$, the n -th arrival to queue j coincides with the latest of the n -th departure epochs of the queues indexed in $\pi(j)$.*

Proof: The first property is just a rephrasing of part of Lemma 1. Assume now that every queue in γ and γ' satisfies the second property. Owing to Lemma 1, one checks directly that every queue in the concatenation or parallelization of γ and γ' will satisfy the property too. ■

3. STOCHASTIC ORDERING

Let $\beta \in C$ and $j \in B$. For $n \geq 0$, let σ_n^j be the service requirement of the n -th customer to be served by queue j and d_n^j be the total delay experienced between the n -th external arrival date and the beginning of the n -th service in queue j .

Lemma 3. *For $n \geq 0$*

$$(3.1) \quad d_{n+1}^j = \max\left(\max_{i \in \pi(j)} (d_{n+1}^i + \sigma_{n+1}^i), d_n^j + \sigma_n^j + a_n - a_{n+1}\right),$$

where the maximum over an empty set is zero by definition and

$$(3.2) \quad d_0^j = \max_{i \in \pi(j)} (d_0^i + \sigma_0^i).$$

Proof: The boundary condition of equation (3.2) follows directly from Lemma 2.

If $\pi(j) = \emptyset$, the first assertion of Lemma 2 shows that we should have

$$(3.3) \quad d_{n+1}^j = \max(0, d_n^j + \sigma_n^j + a_n - a_{n+1}), n \geq 1$$

which is exactly equation (1).

If $\pi(j) \neq \emptyset$, the second assertion of Lemma 2 shows that the $n + 1$ -th arrival date to queue j takes place at time

$$(3.4) \quad a_{n+1} + \max_{i \in \pi(j)} (d_{n+1}^i + \sigma_{n+1}^i).$$

Since the server of queue j becomes available for serving the $n + 1$ -th customer at time

$$(3.5) \quad a_n + d_n^j + \sigma_n^j,$$

It follows that d_{n+1}^j is equal to the expression in the r.h.s of equation (3.1). ■

We are now in position to prove the stochastic ordering result. Consider a network β in C and assume that all the random variables $\{a_n\}_{n \geq 0}$ and $\{\sigma_n^j\}_{n \geq 0}$, $j \in B$ are defined on the probability space (Ω, F, P) .

Let now $(\hat{a}_n)_{n \geq 0}$ and $(\hat{\sigma}_n^j)_{n \geq 0}$, $j \in B$ on $(\Omega, \mathcal{F}, P)$ be a set of "smoother" arrival and service processes in the sense that there exists a sub σ algebra of $\mathcal{F}$ say $\mathcal{G}$ such that for all $n \geq 0$,

$$(3.6) \quad \hat{a}_{n+1} - \hat{a}_n \leq E[a_{n+1} - a_n | \mathcal{G}] \text{ a.s.}$$

and for all j in B ,

$$(3.7) \quad \hat{\sigma}_n^j \leq E[\sigma_n^j | \mathcal{G}] \text{ a.s.}$$

Let $\hat{d}_n^j$ be the delay variable obtained with this new arrival and service pattern in β . The main result of the paper is the following theorem:

Theorem 4. For $n \geq 0$ and $j \in B$,

$$(3.8) \quad \hat{d}_n^j \leq E[d_n^j | \mathcal{G}] \text{ a.s.}$$

Proof: The property is first established for $n = 0$. If $\pi(j) = \emptyset$, equation (3.2) shows that

$$(3.9) \quad d_0^j = \hat{d}_0^j = 0,$$

so that (3.8) is trivially satisfied. If $\pi(j) \neq \emptyset$, applying Jensen's inequality for conditional

expectations to (3.2) yields:

$$(3.10) \quad E[d_i | \mathcal{G}] \geq \max_{i \in \pi(j)} (E[d_i | \mathcal{G}] + \hat{\sigma}_i'),$$

so that if the predecessors of j satisfy property (3.8), so does queue j since (3.10) implies then:

$$(3.11) \quad E[d_i | \mathcal{G}] \geq \max_{i \in \pi(j)} (\hat{d}_i' + \hat{\sigma}_i') = \hat{d}_j'.$$

Assume now that the property was established for all queues up to rank n . If $\pi(j) \neq \emptyset$ (3.1) shows that

$$(3.12) \quad d_{n+1}^j = \max \left(d_n^j + \sigma_n^j + a_n - a_{n+1}, 0 \right),$$

so that Jensen's inequality together with (3.6) and (3.7) imply that

$$(3.13) \quad E[d_{n+1}^j | \mathcal{G}] \geq \max \left(E[d_n^j | \mathcal{G}] + \hat{\sigma}_n' + \hat{a}_n - \hat{a}_{n+1}, 0 \right).$$

Hence, since (3.8) is satisfied for rank n , we get from (3.13) that

$$(3.14) \quad E[d_{n+1}^j | \mathcal{G}] \geq \max(\hat{d}_n^j + \hat{\sigma}_n' + \hat{a}_n - \hat{a}_{n+1}, 0) = \hat{d}_{n+1}^j \quad \text{a.s.},$$

so that the property is true for rank $n+1$ too. If $\pi(j) \neq \emptyset$, Jensen's inequality applied to (3.1) and equations (3.6) and (3.7) imply that

$$(3.15) \quad E[d_{n+1}^j | \mathcal{G}] \geq \max \left(\max_{i \in \pi(j)} (E[d_{n+1}^i | \mathcal{G}] + \hat{\sigma}_{n+1}^i), E[d_n^j | \mathcal{G}] + \hat{\sigma}_n' + \hat{a}_n - \hat{a}_{n+1} \right).$$

Using now the ordering property for rank n , we get

$$(3.16) \quad E[d_{n+1}^j | \mathcal{G}] \geq \max \left(\max_{i \in \pi(j)} (E[d_{n+1}^i | \mathcal{G}] + \hat{\sigma}_{n+1}^i), \hat{d}_n^j + \hat{\sigma}_n' + \hat{a}_n - \hat{a}_{n+1} \right) \quad \text{a.s.}$$

Assume that the property is satisfied for the predecessors of j , we get that it is then satisfied by queue j too since (3.16) entails that

$$(3.17) \quad E[d_{n+1}^j | \mathcal{G}] \geq \max \left(\max_{i \in \pi(j)} (\hat{d}_{n+1}^i + \hat{\sigma}_{n+1}^i), \hat{d}_n^j + \hat{\sigma}_n' + \hat{a}_n - \hat{a}_{n+1} \right) = \hat{d}_{n+1}^j \quad \text{a.s.}$$

and this finishes the proof. ■

Corollary 5. *Let f be any positive convex increasing function on R^+ . Then, for all $n \geq 0$ and $j \in B$, provided $f(d_n^j)$ and $f(\hat{d}_n^j)$ are both integrable*

$$(3.18) \quad E[f(d_n^j)] \geq E[f(\hat{d}_n^j)].$$

Proof: Due to Jensen's inequality

$$(3.19) \quad E[f(d_n^j) | \mathcal{G}] \geq f(E[d_n^j | \mathcal{G}]),$$

so that using equation (3.18) and the increasingness of f ,

$$(3.20) \quad E[f(d_n^j) | \mathcal{G}] \geq f(\hat{d}_n^j).$$

Equation (3.18) follows now directly from (3.20). ■

Remark

Consider a two queue series network and denote as W_n^j , $n \geq 1$, $j = 1, 2$ the waiting time of the n -th customer to enter queue j . We have the following inductions for the W_n^j 's, initialized by the conditions $W_0^j = 0$:

$$(3.21) \quad W_{n+1}^1 = \max(W_n^1 + \sigma_n^1 + a_n - a_{n+1}, 0)$$

$$(3.22) \quad W_{n+1}^2 = \max(W_n^2 + \sigma_n^2 + a_n^1 - a_{n+1}^1, 0)$$

where the a_n^1 's are the departure epochs from queue 1:

$$(3.23) \quad a_{n+1}^1 - a_n^1 = \sigma_{n+1}^1 + \max(a_{n+1} - a_n - \sigma_n^1 - W_n^1, 0).$$

Notice that due to the decreasingness of the r.h.s of (3.23) considered as a function of W_n^1 , we cannot derive from this any simple comparison result between $(a_{n+1}^1 - a_n^1)$ and $(\hat{a}_{n+1}^1 - \hat{a}_n^1)$ when using Jensen's inequality as before. This suggests that the W_n^j 's do not satisfy the same type of ordering property as the d_n^j 's in general.

The bound derived in Corollary 5 is basically transient. In order to continue this bound to the steady state case, let us make the following assumptions:

- The interarrival times $\{a_{n+1} - a_n\}_{n \geq 0}$ form a stationary and ergodic sequence of integrable random variables on $(\Omega, \mathcal{F}, P)$.

- The service requirements for queue j , $\{\sigma_n^j\}_{n \geq 0}$, form a stationary and ergodic sequence of integrable random variables on $(\Omega, \mathcal{F}, P)$, for all $j \in B$.

The following theorem generalizes Loynes' result for $G/G/1$ queues (see [9]).

Theorem 6. *Let $j_0 \in B$. Under the above conditions, if for each j in $\pi(j_0) \cup \{j_0\}$ we have*

$$(3.24) \quad E\{\sigma_n^j\} < E[a_{n+1} - a_n],$$

then the distribution functions of the random variables $\{d_n^{j_0}\}_{n \geq 0}$ converge weakly to a proper distribution function on R^+ when n goes to infinity.

Proof: See the Appendix. ■

In the sequel, we shall denote by $d_\infty^{j_0}$ any random variable on $(\Omega, \mathcal{F}, P)$ with this limiting distribution (see the appendix also for the construction of such a random variable).

Corollary 7. *Assume that the conditions of Lemma 6 hold for both $\{a_{n+1} - a_n\}_{n \geq 0}$ and $\{\sigma_n^j\}_{n \geq 0}$ as well as $\{\hat{a}_{n+1} - \hat{a}_n\}_{n \geq 0}$ and $\{\hat{\sigma}_n^j\}_{n \geq 0}$. Then for any positive, increasing and convex function $f: R_+ \rightarrow R_+$,*

$$(3.25) \quad E[f(d_\infty^{j_0})] \geq E[f(\hat{d}_\infty^{j_0})],$$

provided both quantities are finite.

Proof: Inequality (3.25) is obtained from (3.18) by letting n go to infinity, and using the weak convergence result of Lemma 6. ■

4. APPLICATIONS

Unless specified otherwise, β is any network in C .

A. Determinism minimizes response times.

The property that deterministic interarrival times (respectively service times) minimizes the response times among all $G/G/1$ FIFO queues with the same mean interarrival time (respectively service time), as shown in [2] or [3], can be extended directly to our network β , using Theorem 4.

Consider the case where the sequences $\{a_{n+1} - a_n\}_{n \geq 0}$ and $\{\sigma_n^j\}_{n \geq 0}$ with $j \in B$, are mutually independent.

Let $\hat{d}_n^j$ be the response time for the following respective characteristics:

$$(4.1) \quad \begin{aligned} \hat{a}_{n+1} - \hat{a}_n &= E[a_{n+1} - a_n], \\ \hat{\sigma}_n^j &= \sigma_n^j, \end{aligned}$$

and $\tilde{d}_n^j$ is the response time corresponding to:

$$(4.2) \quad \begin{aligned} \tilde{a}_{n+1} - \tilde{a}_n &= a_{n+1} - a_n, \\ \tilde{\sigma}_n^{j_0} &= E[\sigma_n^{j_0}] && \text{for some } j_0, \\ \tilde{\sigma}_n^j &= \sigma_n^j && \text{for all } j \neq j_0. \end{aligned}$$

Notice that β has the same stability condition for both timings.

Corollary 8. *There exist two sub σ -fields, $\hat{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ such that for any positive, increasing, and convex function f we have for all $n \geq 0$ and $j \in B$,*

$$(4.3) \quad f(\hat{d}_n^j) \leq E[f(d_n^j) | \hat{\mathcal{G}}]$$

$$(4.4) \quad f(\tilde{d}_n^j) \leq E[f(d_n^j) | \tilde{\mathcal{G}}].$$

From these two inequalities follow

$$(4.5) \quad \begin{aligned} E[f(\hat{d}_n^j)] &\leq E[f(d_n^j)] \\ E[f(\tilde{d}_n^j)] &\leq E[f(d_n^j)], \end{aligned}$$

provided integrability holds everywhere. Moreover, these last two bounds extend to steady state under additional stationarity, ergodicity, and stability conditions.

Proof: This follows from Theorem 4 and its first corollary, when taking for $\hat{\mathcal{G}}$ the σ -field generated by $\{\sigma_n^j\}_{n \geq 0, j \in B}$, for $\tilde{\mathcal{G}}$ the one generated by $\{a_{n+1} - a_n, \sigma_n^j\}_{n \geq 0, j \in B}$, and then using the mutual independence assumptions. ■

B. Lower bounds for the delays in parallel networks.

Theorem 4 also provides computable lower and upper bounds for the following problem, a special case of which was considered in [7]. Let β be a network of C obtained by the parallelization of K subnetworks of C , say $\alpha_1, \dots, \alpha_K$ (i.e. $\beta = \beta_K$ where $\beta_1 = \alpha_1$ and β_l is the parallelization of β_{l-1} and α_l for $l > 1$). Denote as $M(B)$ the set of maximal elements in B , which coincides with the union of the maximal elements of A_l (the index set of α_l). Consider now the problem of evaluating the moments of the successive total delays through β ;

$$(4.7) \quad T_n = \max_{j \in M(B)} (d_n^j + \sigma_n^j), \quad n \geq 0.$$

Let T_n^l (resp. $\hat{T}_n^l$) denote the total delay through α_l when timing this network with the interarrival times $a_{n+1} - a_n$ (resp. $\hat{a}_{n+1} - \hat{a}_n = E(a_{n+1} - a_n)$) and the service times σ_n^j (resp. $\hat{\sigma}_n^j = \sigma_n^j$), $j \in A_l$, $n \geq 0$.

Corollary 9. *Let f and $\hat{\mathcal{G}}$ be defined as in Corollary 8. For any $n \geq 0$,*

$$(4.8) \quad E[f(T_n) | \hat{\mathcal{G}}] \geq \max_{1 \leq l \leq K} f(\hat{T}_n^l)$$

and

$$(4.9) \quad E[f(T_n)] \geq E[\max_{1 \leq l \leq K} f(\hat{T}_n^l)],$$

provided these quantities are integrable. Under the additional assumptions mentioned in Corollary 8, this last bound extends to steady state statistics.

Proof: Using Jensen's inequality and independence, we get

$$(4.10) \quad E[f(T_n) | \hat{\mathcal{G}}] \geq \max_{j \in M(B)} f(E[d_n^j | \hat{\mathcal{G}}] + \sigma_n^j),$$

so that (4.3) implies that

$$(4.11) \quad E[f(T_n) | \hat{\mathcal{G}}] \geq \max_{j \in M(B)} f(\hat{d}_n^j + \sigma_n^j).$$

Since $\hat{T}_n^l = \max_{j \in M(A_l)} (\hat{d}_n^j + \sigma_n^j)$, this last equation entails (4.8) and then (4.9). ■

Remark.

Notice that due to our mutual independence assumption on the sequences $\{\sigma_n^l\}_{n \geq 0}$, the random variables T_n^l are mutually independent for any fixed n . Hence the lower bound of equation (4.10) reduces to the determination of the expected value of the maximum of K independent random variables related to the subnetworks α_l , $l = 1, \dots, K$.

C. Upper bounds for delays in parallel networks with divisible interarrival times.

Consider the case where the arrival process of the preceding parallel network is divisible in the sense that there exist K mutually independent sequences of positive random variables $\{\hat{a}_{n+1}^l - \hat{a}_n^l\}_{n \geq 0, l=1, \dots, K}$ which satisfy the following mean condition:

$$(4.12) \quad a_{n+1} - a_n = \frac{1}{K} \sum_{l=1}^K \hat{a}_{n+1}^l - \hat{a}_n^l,$$

for all $n \geq 0$. Let $\{\hat{T}_n^l\}_{n \geq 0}$ represent the sequence of total delays through α_l when timed with the interarrival times $\{\hat{a}_{n+1}^l - \hat{a}_n^l\}_{n \geq 0}$ and the service requirements $\hat{\sigma}_n^l = \sigma_n^l$, $n \geq 0$ and $j \in A_l$. Notice that here again the network α_l has the same stability condition for both the original and the new timings.

Corollary 10. *Let f be as in Corollary 8. There exists a sub σ -field $\mathcal{G}$ of $\mathcal{F}$ such that for all $n \geq 0$,*

$$(4.13) \quad E[\max_{1 \leq l \leq K} f(\hat{T}_n^l) | \mathcal{G}] \geq f(T_n),$$

and

$$(4.14) \quad E[f(T_n)] \leq E[\max_{1 \leq l \leq K} f(\hat{T}_n^l)],$$

provided integrability holds everywhere. As before, this last bound extends to steady state statistics under additional assumptions.

Proof: Take for $\mathcal{G}$ the σ -field generated by the variables $\{a_{n+1} - a_n, \sigma_n^l\}_{j \in B, n \geq 0}$. For any $n \geq 0$, one gets from the various independence assumptions and from the exchangeability of the $\hat{a}_{n+1}^l - \hat{a}_n^l$, $l = 1, \dots, K$, that

$$(4.15) \quad E[\hat{a}_{n+1}^l - \hat{a}_n^l | \mathcal{G}] = a_{n+1} - a_n.$$

Using Jensen's inequality, we get

$$(4.16) \quad \begin{aligned} E[f(\max_{1 \leq l \leq K} \hat{T}_n^l) | \mathcal{G}] &= E[f(\max_{1 \leq l \leq K} \max_{j \in M(A_l)} (d_n^l + \sigma_n^j)) | \mathcal{G}] \\ &\geq f(\max_{1 \leq l \leq K} \max_{j \in M(A_l)} (E[\hat{d}_n^l | \mathcal{G}] + \sigma_n^j)). \end{aligned}$$

This together with Corollary 5 entail

$$(4.17) \quad E[f(\max_{1 \leq l \leq K} \hat{T}_n^l) | \mathcal{G}] \geq f(\max_{1 \leq l \leq K} \max_{j \in M(A_l)} d_n^l + \sigma_n^j) = f(T_n).$$

This completes the proof of (4.13) and (4.14). ■

Notice that for this upper bound, the random variables $\hat{T}_n^l$ are also mutually independent and can be obtained by considering the subnetworks $\{\alpha_l\}_{l=1, \dots, K}$ in isolation.

D. Networks in a random environment.

The problem of determining the statistics of isolated queues with time varying interarrival times (service times) was considered, for the Markovian case in [10]. For the general $G/G/1$ FIFO queue, bounds are also available, when the variations depend upon an independent and stationary "environment" random process. It was shown in [4] that the waiting time statistics in such a queueing system are bounded from below by those of the same queue with the environment process kept fixed to its mean value. Theorem 4 allows us to extend this result (also based on convex ordering) to any network β of C . We shall limit the discussion to the case where the environment process modulates the interarrival times only. As in [4], the environment process is assumed to be a random process $\{V(t)\}_{t \in \mathbb{R}^+}$ on $(\Omega, \mathcal{F}, P)$ being t -stationary and with piecewise constant values in $\mathbb{R}^+$. The network itself is characterized by a sequence of positive random variables $\{a_{n+1} - a_n\}_{n \geq 0}$ (to be "modulated" by V) and by the service times sequences $\{\sigma_n^j\}_{n \geq 0, j \in B}$ all mutually independent. All these random variables are supposed to be integrable with $E(V(t)) = 1$ holding in particular. $\{V(t)\}_{t \geq 0}$ is also assumed to be independent of the timing sequences. The modulation of the variables $a_{n+1} - a_n$ is obtained, like in the case of modulated Poisson process, by accelerating

time proportionally to V , so that the effective interarrival times in the network come from the sequence $(\hat{a}_{n+1} - \hat{a}_n)_{n \geq 0}$ defined by $\hat{a}_0 = 0$, and

$$(4.18) \quad \hat{a}_{n+1} - \hat{a}_n = \int_{a_n}^{a_{n+1}} V(s) ds$$

Let $\{\hat{d}_n^j\}$ be the sequence of response times in this network and $\{d_n^j\}$ be the sequence obtained when fixing $V(t)$ equal to 1, its mean value.

Corollary 11. *Let f be as in Corollary 8. There exists a sub σ -field $\mathcal{G}$ of $\mathcal{F}$ such that for all $n \geq 0$ and j in B ,*

$$(4.19) \quad E[f(\hat{d}_n^j) | \mathcal{G}] \geq f(d_n^j) \text{ a.s.}$$

and

$$(4.20) \quad E[f(\hat{d}_n^j)] \geq E[f(d_n^j)],$$

provided integrability is granted. If both the sequences $\{a_{n+1} - a_n\}_{n \geq 0}$ and $\{\sigma_n^j\}_{n \geq 0, j \in B}$ are stationary and ergodic sequences and $V(t)$ is a stationary and ergodic process, β has the same stability condition for both timings, and the bound (4.20) extends to steady state when this stability condition is satisfied.

Proof. Let $\mathcal{G}$ be the σ -field generated by the variables $\{a_{n+1} - a_n, \sigma_n^j\}_{j \in B, n \geq 0}$. The following basic property is established in [4]. Under the enforced assumptions, for any $n \geq 0$,

$$(4.21) \quad E[\hat{a}_{n+1} - \hat{a}_n | \mathcal{G}] = a_{n+1} - a_n,$$

so that in particular

$$(4.22) \quad E[\hat{a}_{n+1} - \hat{a}_n] = E[a_{n+1} - a_n].$$

Equation (4.19) follows from (4.21) and Theorem 4 in the usual way. The last assertion follows from the fact that the sequence $\hat{a}_{n+1} - \hat{a}_n$ is also stationary and ergodic (see [4]) and from (4.22).

■

5. APPENDIX

The basic idea for proving Lemma 6 consists of generalizing the schema of Loynes for the response time of a G/G/1 queue [9], to the response times d_n^j of our network. Let us first consider the sequences $\{a_{n+1}-a_n\}_{n \geq 0}$ and $\{\sigma_n^j\}_{n \geq 0}$ for all $j \in B$ as the right half of certain bi-infinite sequences $\{a_{n+1}-a_n\}_{n \in \mathbb{Z}}$ and $\{\sigma_n^j\}_{n \in \mathbb{Z}}$ on $(\Omega, \mathcal{F}, P)$. We shall assume that $(\Omega, \mathcal{F}, P)$ is the canonical space for these sequences. Let θ denote the discrete left shift on this canonical space. Hence P will be assumed to be θ -invariant (stationarity) and θ -ergodic. Let us denote by τ the difference $a_1 - a_0$, and by σ^j the variable σ_0^j . Consider now the schema $\{\delta_n^j\}_{n \geq 0}$ defined by $\delta_0^j = 0$, and for $n \geq 0$:

$$(5.1) \quad \delta_{n+1}^j \circ \theta = \max \left(\max_{i \in \pi(j)} (\delta_{n+1}^i + \sigma^i \circ \theta), \delta_n^j + \sigma^j - \tau \right).$$

Lemma 1. *For any $j \in B$, the sequence $\{\delta_n^j\}_{n \geq 0}$ is increasing.*

Proof: Let us first prove this for the minimal elements. It is clear that $\delta_1^j \geq 0 = \delta_0^j$. Assume now that $\delta_n^j \geq \delta_{n-1}^j$ for some $n \geq 1$. From (5.1), we get:

$$(5.2) \quad \delta_{n+1}^j \circ \theta = \max(\delta_n^j + \sigma^j - \tau, 0) \geq \max(\delta_{n-1}^j + \sigma^j - \tau, 0) = \delta_n^j \circ \theta.$$

By induction, the δ_n^j 's are thus increasing. Now consider j such that $\pi(j) \neq \emptyset$. By induction hypothesis, we can assume that the δ_n^i 's are increasing for $i \in \pi(j)$. It is clear again that $\delta_1^j \geq 0 = \delta_0^j$. Assuming now that $\delta_n^j \geq \delta_{n-1}^j$, by (5.1) we get

$$(5.2) \quad \delta_{n+1}^j \circ \theta \geq \max \left(\max_{i \in \pi(j)} (\delta_{n+1}^i + \sigma^i \circ \theta), \delta_n^j + \sigma^j - \tau \right).$$

Since the δ_n^i are increasing for $i \in \pi(j)$ we get

$$(5.3) \quad \delta_{n+1}^j \circ \theta \geq \max \left(\max_{i \in \pi(j)} (\delta_n^i + \sigma^i \circ \theta), \delta_n^j + \sigma^j - \tau \right)$$

and so for this case δ_n^j increases in n too. ■

Lemma 2. *Let δ_∞^j be the limiting value of the increasing sequence δ_n^j when n goes to infinity. We then have that δ_∞^j is finite a.s. if the condition (3.24) is satisfied for all $i \in \pi(j) \cup \{j\}$. If there exists an $i \in \pi(j) \cup \{j\}$ such that $E(\sigma_n^i) > E(a_{n+1} - a_n)$, then $\delta_\infty^j = \infty$ a.s.*

Proof: The limiting variables δ_∞^j satisfy the pathwise equation:

$$(5.4) \quad \delta_\infty^j \circ \theta = \max \left\{ \max_{i \in \pi(j)} (\delta_\infty^i + \sigma^j \circ \theta), \delta_\infty^j + \sigma^j - \tau \right\}.$$

Now consider the minimal elements of B for which (5.4) reduces to

$$(5.5) \quad \delta_\infty^j \circ \theta = \max(0, \delta_\infty^j + \sigma^j - \tau).$$

Equation(5.5) shows that the event $\{\delta_\infty^j = \infty\}$ is θ -invariant. Therefore, this event is either of probability 0 or 1. Assume that it is of probability 1. By the increasingness property, we have

$$(5.6) \quad E[\max(\delta_n^j + \sigma^j - \tau, 0) - \delta_n^j] = E[\delta_{n+1}^j \circ \theta - \delta_n^j] = E[\delta_{n+1}^j - \delta_n^j] \geq 0.$$

From this we get

$$(5.7) \quad \lim_{n \rightarrow \infty} E[\max(\delta_n^j + \sigma^j - \tau, 0) - \delta_n^j] \geq 0.$$

Using Fatou's lemma, this inequality is preserved with limit taken inside the expectation. If we assume that $\delta_n^j \uparrow \infty$, then we get

$$(5.8) \quad E[\sigma^j] \geq E[\tau].$$

Now taking the contrapositive of this argument, we see that

$$(5.9) \quad E[\sigma^j] < E[\tau]$$

is sufficient to have δ_∞^j finite a.e. This completes the proof of the first part of the lemma for the minimal elements.

Assume now that the property holds for all $i \in \pi(j)$ and that all of the δ_∞^i are finite a.e. The proof for the sufficiency of condition (5.9) to have δ_∞^j finite a.e. proceeds as follows. The event $\{\delta_\infty^j = \infty\}$ is shown to be θ -invariant from (5.5). The inequality

$$(5.10) \quad E \left[\lim_{n \rightarrow \infty} \left[\max \left\{ \max_{i \in \pi(j)} (\delta_n^i + \sigma^j \circ \theta), \delta_n^j + \sigma^j - \tau \right\} - \delta_n^j \right] \right] \geq 0$$

is then established using increasingness as in (5.6). Under the assumption $\delta_\infty^i < \infty$ a.e. for all $i \in \pi(j)$, the hypothesis $\delta_\infty^j = \infty$ used exactly as in (5.10) implies that queue j satisfies condition

(5.8). The rest of the proof follows exactly as before. ■

Proof of Lemma 6: Notice that $d_n^j = \delta_n^j \circ \theta^{-n}$ (by induction). Hence d_n^j and δ_n^j have the same distribution due to the θ -invariance of P . The weak convergence of the law of d_n^j to a proper distribution is now a direct consequence of the increasing convergence a.e. of δ_n^j to the finite random variable δ_∞^j .

References

- [1] W. Whitt. "Comparing and Counting processes and queues" *Adv. Appl. Prob.* 13, 207-220 1981.
- [2] D. Stoyan. "Comparison methods for queues and other stochastic models". English translation (D. J. Daley editor), J. Wiley and sons, New York, 1984.
- [3] W. Whitt. "Minimizing delays in the GI/G/1 queue". *Opns. Res.* Vol 32, pp. 41-51, 1984.
- [4] F. Baccelli, A. Makowski. "Stability and bounds for single server queues in random environment". Submitted to *stochastic models*, 1985.
- [5] W. A. Massey. Calculating Exit Times for Series Jackson Networks I. To appear in *Journal of Applied Probability*.
- [6] L. Flatto, S. Hahn. "Two parallel queues created by arrivals with two demands, I". *SIAM J. Appl. Math.* Vol. 44, pp. 1041-1053, 1984.
- [7] F. Baccelli, A. Makowski. "Simple computable bounds for the Fork-Join queue" *Proceedings of the Johns Hopkins conference on Information Sciences and Systems - Johns Hopkins University*, 1985.
- [8] F. Baccelli. "Two parallel queues created by arrival with two demands: The M/G/2 symmetrical case". Report INRIA No. 426, July 1985.
- [9] R. M. Loynes. *The Stability of a Queue with Non Independent Interarrival and Service Times*. *Proceedings of the Cambridge Philosophical Society*, Vol. 5, pp. 497-520, 1962.
- [10] W. A. Massey. *Asymptotic Analysis of the Time Dependent M/M/1 Queue*. *Mathematics of Operations Research*. Vol. 10, No. 2, May 1985, pp. 305-327.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

