

# Monotonicity of the throughput of an open queueing network in the interarrival and service times

**Citation for published version (APA):**

Adan, I. J. B. F., & Wal, van der, J. (1987). *Monotonicity of the throughput of an open queueing network in the interarrival and service times*. (Memorandum COSOR; Vol. 8705). Technische Universiteit Eindhoven.

**Document status and date:**

Published: 01/01/1987

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

EINDHOVEN UNIVERSITY OF TECHNOLOGY  
FACULTY OF MATHEMATICS AND COMPUTING SCIENCE

Memorandum COSOR 87-05

Monotonicity of the throughput of  
an open queueing network in the  
interarrival and service times

by

Ivo Adan and Jan van der Wal

Eindhoven, The Netherlands

March 1987

# Monotonicity of the throughput of an open queueing network in the interarrival and service times.

by

*Ivo Adan and Jan van der Wal*  
*University of Technology, Eindhoven*

## ABSTRACT

Using coupling and a sample path argument it is shown that the throughput of an open queueing network with general interarrival and service times is decreasing if the arrival and service times are stochastically increasing.

### 1. Introduction.

Recently the question of the behaviour of the throughput in queueing networks, particularly the monotonicity in the service rates and in the number of jobs in closed systems received a lot of attention. See e.g. Robertazzi and Lazar [1985], Suri [1985], Yao [1985], Van der Wal [1985] and Shanthikumar and Yao [1985].

Most results have been found for the case of exponential service times and product form networks. In Adan and Van der Wal [1987a] the monotonicity in the number of jobs was established for a closed network with Erlang service times, and Adan and Van der Wal [1987b] proved the monotonicity for the case of general service time distributions.

An argument for studying monotonicity can be found in the work of Van Dijk who tries to give upper and lower bounds for the throughput of nonproduct form networks, see e.g. Van Dijk and Lamond [1986].

In this paper we prove that the throughput in an open network is monotone in the interarrival and service times. Both interarrival and service times are general. The approach we use is coupling and comparison of sample paths.

The paper is organized as follows. In section 2 the model, some notations and the theorem are given. In section 3 the theorem is proved for the case that all stations are single servers. The case of multi servers is treated in section 4.

### 2. Model, notations and theorem.

Consider a queueing network with stations  $1, 2, \dots, N$  and general, though independent interarrival and service times. The interarrival times are independent of the service times. In station  $i$  there are  $L_i$  identical servers (possibly  $L_i = \infty$ ),  $i = 1, \dots, N$  and in each station the service discipline is FCFS.

We will show that the throughput of this network decreases if the interarrival times and service times stochastically increase. From now on we consider two

networks. The original one, the faster one, is denoted by F and the one with the larger interarrival and service times, the slower one, by S.

In order to establish the monotonicity we shall show that, for a specific initial state and a given realization of the sequences of interarrival and service times in the various queues of both systems and the transitions to be made, the throughput up to time  $t$  in system F is at least equal to the one in system S.

However this sample path approach only works if the interarrival and service times of both systems are related in a specific way. To achieve this we use coupling.

Let  $\underline{Z}_{ij}^C$  be the random variable denoting the  $j$ -th interarrival time from outside at station  $i$  in system  $C$ ,  $\underline{X}_{ij}^C$  be the random variable denoting the service time required by the  $j$ -th arriving job in queue  $i$  in system  $C$  and  $\underline{S}_{ij}$  be the random variable characterizing the transitions of the  $j$ -th departing job from queue  $i$ .

$i = 1, 2, \dots, N, j = 1, 2, \dots, C = S, F$ . If the  $j$ -th departing job from queue  $i$  leaves the network  $\underline{S}_{ij}$  equals zero and else  $\underline{S}_{ij}$  denotes the station where the job will jump to. For the time being we may think of Markovian routing. Note that in the multi server case the  $j$ -th departing job need not be the  $j$ -th arriving job.

In both the S and F system we assume the system to be empty at  $t = 0$ . This assumption is convenient but not very relevant.

Since we assumed that all variables  $\underline{Z}_{ij}^C$  and  $\underline{X}_{kl}^C$  are independent with

$$\underline{Z}_{ij}^S \geq_{st} \underline{Z}_{ij}^F \quad \text{and} \quad \underline{X}_{ij}^S \geq_{st} \underline{X}_{ij}^F \quad \text{for all } i \text{ and } j,$$

it follows from the coupling method that we can generate random variables  $\tilde{\underline{Z}}_{ij}^C$  and  $\tilde{\underline{X}}_{ij}^C$  such that  $\tilde{\underline{Z}}_{ij}^C$  and  $\underline{Z}_{ij}^C$ , and  $\tilde{\underline{X}}_{ij}^C$  and  $\underline{X}_{ij}^C$  are pairwise identically distributed,  $\tilde{\underline{Z}}_{ij}^C$  and  $\tilde{\underline{X}}_{kl}^C$  are independent, and

$$(1) \quad P(\tilde{\underline{Z}}_{ij}^S \geq \tilde{\underline{Z}}_{ij}^F) = 1 \quad \text{and} \quad P(\tilde{\underline{X}}_{ij}^S \geq \tilde{\underline{X}}_{ij}^F) = 1 \quad \text{for all } i \text{ and } j.$$

Let  $Z_{ij}^C, X_{ij}^C$  and  $S_{ij}, i = 1, \dots, N, j = 1, 2, \dots, C = S, F$ , be any given realization of interarrival and service times and transitions of the coupled processes. Then it follows from (1) that

$$(2) \quad Z_{ij}^S \geq Z_{ij}^F \quad \text{and} \quad X_{ij}^S \geq X_{ij}^F \quad \text{for all } i \text{ and } j$$

It will be clear that in order to show that the throughput of the faster system is at least equal to the one of the slower system, where the throughput can be defined as the expected number of service completions before time  $t$  or per unit of time, it is sufficient to consider the coupled realizations of the two systems.

From here on we follow the approach of Adan and Van der Wal [1987b].

We need some further notations.

$A_{ij}$      the time of the  $j$ -th arrival in queue  $i$ .

$D_{ij}$      the departure time of the job that arrived as  $j$ -th job in queue  $i$ . In the multi server case this need not be the time of the  $j$ -th departure.

- $A_i(t)$  total number of arrivals from in- and outside up to and including time  $t$  in queue  $i$ .
- $D_i(t)$  total number of departures up to and including time  $t$  in queue  $i$ .
- $Z_i(t)$  total number of arrivals from outside only up to and including time  $t$  in queue  $i$ .

In the sequel these variables will have a superscript S or F to indicate whether they correspond to the S or the F network.

We will show that for any realization of  $Z_{ij}^C$ 's,  $X_{ij}^C$ 's and  $S_{ij}$ 's of the coupled processes we have for all  $i$  and  $t$

$$D_i^F(t) \geq D_i^S(t).$$

I.e. for any coupled realization system F is more rewarding than system S.

Finally let  $e_1^C < e_2^C < \dots$  be the time instants in system C upon which one or more services are completed or one or more jobs arrive from outside. Then define the sequence  $t_0, t_1, \dots$  by

$$t_0 := 0$$

$$t_n := \min \{ \min \{ e_i^S \mid e_i^S > t_{n-1} \}, \min \{ e_i^F \mid e_i^F > t_{n-1} \} \}, \quad n \geq 1.$$

So  $\{t_n\}$  is the sequence of time instants upon which an event occurs in at least one of the two systems.

We make the following assumption

**Assumption**

- (i)  $\sum_{j=1}^n Z_{ij}^C \rightarrow \infty \quad (n \rightarrow \infty)$  for all  $i$  and C
- (ii)  $\sum_{j=1}^n X_{ij}^C \rightarrow \infty \quad (n \rightarrow \infty)$  for all  $i$  and C
- (iii)  $X_{ij}^C > 0$  for all  $i, j$  and C

The condition  $X_{ij}^C > 0$  guarantees that a job can complete only one service at a time and hence make at most one transition at a time. The first condition guarantees that in every interval  $[0, t]$  only a finite number of jobs arrive from outside at station  $i, i = 1, \dots, N$ . The first and second condition guarantee  $t_n \rightarrow \infty$  for  $n \rightarrow \infty$  (see Appendix).

Now we can state our main result that for all  $t$  and for each station the throughput in the F system is at least equal to the one in the S system. Recall that both systems are empty at  $t = 0$  and that  $X_{ij}^C > 0$ , so  $D_i^C(0) = 0$  for all  $i$  and  $C = S, F$ .

**Theorem.**

For all  $t \geq 0$  and all  $i = 1, 2, \dots, N$

$$(3) \quad D_i^F(t) \geq D_i^S(t)$$

Since the functions  $D_i^C$  are constant on the intervals  $[t_m, t_{m+1})$  and  $t_n \rightarrow \infty$  ( $n \rightarrow \infty$ ) it suffices to prove the theorem for the instants  $t_0, t_1, \dots$

**3. The single server case.**

The proof will be based on the following rather trivial but vital lemma stating that if arrivals come earlier then so do departures.

**Lemma 1. (Single server)**

If station  $i$  is a single server and

$$A_{ij}^F \leq A_{ij}^S \text{ for } j = 1, 2, \dots, n$$

then

$$D_{ij}^F \leq D_{ij}^S \text{ for } j = 1, 2, \dots, n$$

**Proof.**

By induction.

$n = 1$ .

$$D_{i1}^F = A_{i1}^F + X_{i1}^F \leq A_{i1}^S + X_{i1}^S = D_{i1}^S$$

Assume that the lemma holds for  $n = m$ . Then

$$(4) \quad \begin{aligned} D_{im+1}^F &= \max \{ D_{im}^F, A_{im+1}^F \} + X_{im+1}^F \\ &\leq \max \{ D_{im}^S, A_{im+1}^S \} + X_{im+1}^S = D_{im+1}^S \end{aligned}$$

which proves the lemma for  $n = m+1$ . □

Before proving the theorem we also need the following result which immediately follows from (2).

**Lemma 2.**

$$Z_i^F(t) \geq Z_i^S(t) \text{ for all } i \text{ and } t$$

As argued before it suffices to prove (3) for the sequence  $t_0, t_1, \dots$ . This will be done by induction.

For  $t_0$  inequality (3) holds by definition (all  $D_i^C(0) = 0$ ).

Assume (3) holds for  $t_0, t_1, \dots, t_m$ . Then, with  $\delta(i, j) = 1$  if  $i = j$  and 0 otherwise, for  $k = 0, 1, \dots, m$

$$\begin{aligned} A_i^F(t_k) &= \sum_{l=1}^N \sum_{j=1}^{D_l^F(t_k)} \delta(S_{lj}, i) + Z_i^F(t_k) \\ &\geq \sum_{l=1}^N \sum_{j=1}^{D_l^S(t_k)} \delta(S_{lj}, i) + Z_i^S(t_k) = A_i^S(t_k). \end{aligned}$$

Since  $A_i^C(t)$  is constant on  $[t_{k-1}, t_k)$

$$A_i^F(t) \geq A_i^S(t) \text{ for } t \leq t_m$$

thus

$$A_{ij}^F \leq A_{ij}^S \text{ for } j = 1, \dots, A_i^S(t_m),$$

and by lemma 1

$$D_{ij}^F \leq D_{ij}^S \text{ for } j = 1, \dots, A_i^S(t_m).$$

Further, for all  $j$  for which

$$D_{ij}^S \leq t_{m+1}$$

it follows from  $X_{ij}^S > 0$  that

$$A_{ij}^S \leq t_m.$$

thus  $j \leq A_i^S(t_m)$  and

$$D_{ij}^F \leq D_{ij}^S.$$

Hence

$$D_i^F(t_{m+1}) \geq D_i^S(t_{m+1}),$$

and (3) holds for  $t_{m+1}$ .

This completes the proof of the theorem for the single server case.  $\square$

#### 4. The multi server case.

The proof for the multi server case follows exactly the same lines as the single server one. Once the multi server equivalent of lemma 1 is established the rest of the argument for the proof of the theorem is identical.

The problem is to prove lemma 1 for multi servers. For that we need more notation. Define for  $C = S, F$

$$Y_{ijl}^C = \begin{cases} 1 & \text{if the } j\text{-th arriving job in station } i \\ & \text{is served by server } l \\ 0 & \text{otherwise} \end{cases}$$

$T_{ijl}^C$  the time server  $l$  in station  $i$  becomes available for the  $j$ -th arriving job, so

$$T_{ijl}^C = \max \{ D_{ik}^C Y_{ikl}^C, k = 1, 2, \dots, j-1 \}$$

and

$T_{ij}^C$  the vector of  $T_{ijl}^C$ 's :

$$T_{ij}^C = ( T_{ij1}^C, T_{ij2}^C, \dots )$$

Now we have

$$(5) \quad D_{ij}^C = \max \left\{ \min_l T_{ijl}^C, A_{ij}^C \right\} + X_{ij}^C$$

This relation is clearly more complicated than the one for  $D_{ij}^C$  in (4). Therefore we have to study the vectors  $T_{ij}^C$ 's.

Note that the case  $L_i = \infty$  is simple since then (5) reduces to

$$(6) \quad D_{ij}^C = A_{ij}^C + X_{ij}^C,$$

so  $A_{ij}^F \leq A_{ij}^S$  and  $X_{ij}^F \leq X_{ij}^S$  imply that  $D_{ij}^F \leq D_{ij}^S$ .

Now let us assume  $L_i < \infty$ .

$$(7) \quad T_{ij+1l}^C = \begin{cases} \max \{ T_{ijl}^C, A_{ij}^C \} + X_{ij}^C \\ \quad \text{if } T_{ijl}^C = \min_k T_{ijk}^C \\ \quad \text{and (the tiebreak rule) } T_{ijk}^C > T_{ijl}^C, k < l \\ T_{ijl}^C \quad \text{otherwise} \end{cases}$$

So  $T_{ij+1}^C$  is obtained from  $T_{ij}^C$  by replacing the minimal component with the smallest index. To get hold of the effect of this we need the following.

Let  $a = (a_1, \dots, a_n)$  be a vector.

Then  $Ua$  denotes the nondecreasing reordering of  $a$ :

$$Ua = (a_{i_1}, a_{i_2}, \dots, a_{i_n})$$

with  $a_{i_1} \leq a_{i_2} \leq \dots \leq a_{i_n}$ , and  $(i_1, i_2, \dots, i_n)$  a permutation of  $(1, 2, \dots, n)$ .

For two vectors  $a = (a_1, a_2, \dots, a_n)$  and  $b = (b_1, b_2, \dots, b_n)$  we write

$$a \leq b \text{ if } (Ua)_i \leq (Ub)_i, i = 1, \dots, n.$$

One may easily verify the following result.

**Lemma 3.**

Let

$$(a_1, a_2, \dots, a_n) \leq (b_1, b_2, \dots, b_n), \alpha \leq \beta, a_i = \min_l a_l, b_j = \min_l b_l,$$

then

$$(a_1, \dots, a_{i-1}, \alpha, a_{i+1}, \dots, a_n) \leq (b_1, \dots, b_{j-1}, \beta, b_{j+1}, \dots, b_n).$$

So replacing a minimal component in  $a$  by  $\alpha$  and in  $b$  by  $\beta$  with  $\alpha \leq \beta$  does not affect the ordering  $a \leq b$ .

This gives us



**Corollary.**

If

$$T_{ij}^F \leq T_{ij}^S \text{ and } A_{ij}^F \leq A_{ij}^S$$

then

$$T_{ij+1}^F \leq T_{ij+1}^S.$$

**Proof.**

Immediate from (2) and (7) and lemma 3. □

Now we can state and prove the multi server version of lemma 1.

**Lemma 4. (Multi server)**

If station  $i$  is a multi server and

$$A_{ij}^F \leq A_{ij}^S \text{ for } j = 1, 2, \dots, n$$

then

$$D_{ij}^F \leq D_{ij}^S \text{ for } j = 1, 2, \dots, n$$

**Proof.**

As observed before we only need to consider the case  $L_i < \infty$  as the lemma is trivially obtained from (6) for  $L_i = \infty$ .

Since  $T_{i1}^C = (0, \dots, 0)$ , so  $T_{i1}^F = T_{i1}^S$  (guaranteeing the initial condition of the corollary) the corollary yields us

$$T_{ij}^F \leq T_{ij}^S$$

and thus

$$\min_l T_{ijl}^F \leq \min_l T_{ijl}^S \text{ for } j = 1, 2, \dots, n.$$

By (5) this implies

$$D_{ij}^F \leq D_{ij}^S \text{ for } j = 1, 2, \dots, n.$$

which completes the proof of the lemma. □

As said before, this gives the proof of the theorem for multi servers.

**5. Conclusions and remarks.**

In the preceding we studied the open queueing network with in each station independent interarrival times and nonzero, independent and identically distributed service times and Markovian routing. It has been shown that the throughput up to time  $t$ , defined as the expected number of service completions in a station up to and including time  $t$  is decreasing if the interarrival and service times are stochastically increasing for all  $t$ . So also the average number of service completions per unit time is decreasing.

In the remainder of this section we discuss and relax most of the above conditions, such as independent interarrival times and i.i.d. service times.

#### **Independent interarrival times.**

- In fact the only assumption we have to make about the arrival process is that the arrival times are independent of service times and that the arrival processes in system F and S can be coupled such that all arrivals in system F are not later than the ones in system S with probability 1. If jobs arrive according to the same arrival process in both systems, the latter condition is trivially fulfilled, so in this case the monotonicity holds for any arbitrary arrival process as long as it is independent of the service times.

- As a consequence of the previous remark the arrival process may also be a batch arrival process. For instance the monotonicity holds if we assume that the batch sizes are independent and also independent of the interarrival and service times and stochastically smaller in system S than in F.

#### **I.i.d. service times.**

- The service times in a station may be dependent as long as the service times in system S and F can be coupled such that  $P(X_{ij}^S \geq X_{ij}^F) = 1$  for all  $i$  and  $j$ . For example, consider a network consisting of one single server station, a machine say, where jobs arrive according to an arbitrary arrival process, independent of the service times. In fact there are three different types of jobs. The probability that an arriving job is of type  $i$  is  $p_i$ ,  $i = 1, 2, 3$ . The service time distribution for type  $i$  jobs is  $G_i^C$ . Servicing is in order of arrival. If after the completion of a type  $i$  job the machine will next service a type  $j$  job, then there is a setup time distributed according to  $G_{ij}^C$ . So every job receives a generalized service time consisting of the setup time and the real service time. All service times, arrival times and setup times are independent. In this case the generalized service times may be dependent. The coupling still works and as a result the throughput is decreasing if the service times, arrival times and setup times are stochastically increasing.

The restrictions to identical service times and Markovian routing can also be relaxed. For a discussion, see Adan and Van der Wal [1987b].

The sample path approach can be used for many other network problems. For example the throughput is monotone in the number of servers. One can show that the throughput decreases if there are less servers available, say server  $l$  in station  $i$  is not available. In terms of  $T_{ij}$  this means that the component  $T_{ijl}$  is set to infinity for all  $j$ . For the proof of the monotonicity one can use the same arguments as in the multi server case.

Another example is a network with a finite capacity. If the system is full, an arriving job leaves without receiving service, and else the station where the arriving job will enter the system, is drawn from a discrete distribution. Then the throughput is monotone in the capacity of the network.

The problem monotonicity of the throughput of a queueing network with breakdowns can be solved by the sample path technique. Etc.

**Appendix.**

If

$$(i) \quad \sum_{j=1}^n Z_{ij}^C \rightarrow \infty \text{ for } n \rightarrow \infty$$

and

$$(ii) \quad \sum_{j=1}^n X_{ij}^C \rightarrow \infty \text{ for } n \rightarrow \infty$$

for  $C = S, F$ , then

$$t_n \rightarrow \infty$$

Suppose to the contrary  $t_n \rightarrow t < \infty$  for  $n \rightarrow \infty$ , then  $e_n^C \rightarrow t$  ( $n \rightarrow \infty$ ) for  $C = S$  or  $F$ , say  $C = S$ . Then we have in at least one station,  $i$  say,  $A_i^S(x) \rightarrow \infty$  ( $x \rightarrow t$ ).

From (i) it follows that only a finite number of jobs arrived from outside the network in  $[0, t)$ , say  $K$  jobs. Let us mark the  $K$  jobs in the system  $1, 2, \dots, K$ , and let  $J_1$  be the subset of  $\{1, 2, \dots, K\}$  of jobs which arrive infinitely often in station  $i$  in  $[0, t)$ . So  $J_1$  must be nonempty. Let  $J_2$  be the complement:

$$J_2 = \{1, \dots, K\} \setminus J_1.$$

Define for each job  $k \in J_2$

$t_k^A$  the last arrival in station  $i$  before  $t$

$t_k^D$  the last departure station  $i$  before  $t$

and define

$$t_0 = \max_{k \in J_2} \max(t_k^A, t_k^D)$$

Then each job  $k \in J_2$  is either trying to complete one task  $X_{ij}^S$  during the whole interval  $(t_0, t)$ , if  $t_k^A > t_k^D$ , or it is not in station  $i$  for all  $x$  in  $(t_0, t)$ , if  $t_k^D > t_k^A$ .

So there is a finite, possibly empty set  $I$  of indices of tasks performed by the  $J_2$  jobs between  $t_0$  and  $t$ . Further, since each job in  $J_1$  performs infinitely many tasks in  $[0, t)$ , all tasks given to a  $J_1$  job are completed before  $t$ , so all  $X_{ij}^S, j \in I$ , are completed before  $t$ . Since  $I$  is finite it follows with (ii) that the total service time received by the  $K$  jobs up to time  $t$  is at least equal to

$$\sum_{\substack{j=1 \\ j \in I}}^{\infty} X_{ij}^S = \infty$$

On the other hand each job can get at most  $t$  time units of service between  $0$  and  $t$ , so the  $K$  jobs can receive together at most  $Kt$  time units of service.

Contradiction, hence  $t_n \rightarrow \infty$  for  $n \rightarrow \infty$ .

**References.**

- Adan, I.J.B.F. and J. van der Wal (1987a), Monotonicity of the throughput of a closed Erlang queueing network in the number of jobs, Eindhoven University of Technology, Department of Mathematics and Computing Science, Memorandum COSOR 87-01.
- Adan, I.J.B.F. and J. van der Wal (1987b), Monotonicity of the throughput of a closed queueing network in the number of jobs, Eindhoven University of Technology, Department of Mathematics and Computing Science, Memorandum COSOR 87-03.
- Dijk, N.M. van and B.F. Lamond (1986), Bounds for the call congestion of finite single-server exponential tandem queues, to appear in Operations Research.
- Robertazzi, T.G. and A.A. Lazar (1985), On the modeling and optimal flow control of the Jacksonian network, Performance Evaluation 5, 29-43.
- Shanthikumar, J.G. and D.D. Yao (1985), Stochastic monotonicity of the queue lengths in closed queueing networks, to appear in Operations Research.
- Suri, R. (1985), A concept of monotonicity and its characterization for closed queueing networks, Operations Research 33, 606-624.
- Wal, J. van der (1985), Monotonicity of the throughput of a closed exponential queueing network in the number of jobs, Eindhoven University of Technology, Department of Mathematics and Computing Science, Memorandum COSOR 85-21.
- Yao, D.D. (1985), Some properties of the throughput function of closed networks of queues, Operations Research Letters 3, 313-317.