

Stability and bounds for single server queues in random environment

François Baccelli, A. Makowski

► **To cite this version:**

François Baccelli, A. Makowski. Stability and bounds for single server queues in random environment. RR-0536, INRIA. 1986. <inria-00076018>

HAL Id: inria-00076018

<https://hal.inria.fr/inria-00076018>

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.

IRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

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

Rapports de Recherche

N° 536

STABILITY AND BOUNDS FOR SINGLE SERVER QUEUES IN RANDOM ENVIRONMENT

François BACCELLI
Armand A. MAKOWSKI

Juin 1986

FILES D'ATTENTE EN ENVIRONNEMENT ALEATOIRE :

CONDITION DE STABILITE ET BORNES

F. BACCELLI
INRIA
Domaine de Voluceau
Rocquencourt - BP 105
78153 Le Chesnay Cédex
France

Armand A. MAKOWSKI
Electrical Eng. Dept.
University of Maryland
College Park
Maryland 20742
USA

RESUME

On considère une file d'attente à un serveur dont la vitesse de traitement est fonction d'un environnement aléatoire. La seule hypothèse sur le processus des arrivées (dates d'arrivées et demandes de service) et sur le processus des vitesses sont la stationnarité et l'ergodicité. On obtient la condition générale de stabilité. Dans le cas où ces deux processus sont indépendants, on montre que la charge de ce système est minorée (pour l'ordre convexe) par la charge dans une file d'attente avec le même processus d'arrivées et vitesse constante égale à la moyenne du processus des vitesses.

STABILITY AND BOUNDS

FOR

SINGLE SERVER QUEUES IN

RANDOM ENVIRONMENT

by

François Baccelli¹
INRIA
Domaine de Voluceau
Rocquencourt - BP 105
78153 Le Chesnay Cedex
FRANCE

and

Armand M. Makowski²
Electrical Engineering Department
University of Maryland
College Park
Maryland 20742
U.S.A.

ABSTRACT

We consider a single server queue where the speed of the server is a function of a random environment. The only assumptions concerning the arrival process (arrival dates and service requirements) and the speed process are stationarity and ergodicity. A general expression is given for the stability condition. In the case where the two processes are independent, we show that the workload in this queue is larger for convex ordering than the workload in a queue with the same arrival process but fixed deterministic speed taken as the average of the speed process.

¹On leave from INRIA : the work of this author was supported through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland at College Park, U.S.A.

²The work of this author was supported partially through N.S.F. grant; ECS-83-51386 and partially through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland at College Park, U.S.A.

1 - INTRODUCTION

Two classes of queues in random environment may be defined depending on whether the environment modulates the input stream or the speed of the server (see [PUR 74], [NEU 78.1], [NEU 78.2]). We shall consider here a single server queue in which the server works at a speed which depends upon a randomly changing environment. However, it is not hard to see that the model we propose covers actually both cases and that the two main theorems can be rewritten for the case of a modulated input just by interpreting differently the constituting processes. Intensity conservation laws have been derived for this type of queuing systems by Miyazawa ([MIY 85]). Bounds on the first moment of the stationary workload have also been derived by Rolski ([ROL 81]) in the case of modulated Poisson inputs and i.i.d. service times. The aim of the present paper is to give a general expression for the stability condition of such systems and to prove that under certain independence conditions, both stationary and transient workloads are larger for the stochastic convex ordering (see [STO 83]) than the corresponding workloads in a queue where the speed of the server is constant and equal to the mean value of the speed process.

The paper is organized as follows : the appropriate mathematical formalism to be used in the paper is outlined in a preliminary section. The model is defined in section 3. This section also contains the derivation of the stability condition. The main stochastic ordering result is established in section 4.

2 - PRELIMINARIES

We shall denote by M the space of integer-valued Radon measures on \mathbb{R} endowed with the Borel σ -field generated by the vague topology. If $N \in M$, then there exists a sequence $((t_n), n \in \mathbb{Z})$ such that $N = \sum_{n \in \mathbb{Z}} \varepsilon_{t_n}$, where ε_a is the Dirac measure at $a \in \mathbb{R}$ and the t_n 's are ordered in such a way that $t_n \leq t_{n+1}$ and $t_0 \leq 0 < t_1$.

In the following, all the random variables used are defined on a probability space $(\Omega, \mathcal{F}, P, (\theta_t, t \in \mathbb{R}))$ where θ_t is a family of measurable automorphisms of Ω leaving P invariant. P will also be assumed to be θ_t ergodic.

The arrival process in the queue will be represented by a stationary point process, that is a measurable function K from Ω to M

$$(2.1) \quad \omega \rightarrow K(\omega, dt) = \sum_{n \in \mathbb{Z}} \epsilon_{t_n},$$

such that (stationarity)

$$(2.2) \quad K(\theta_s(\omega), dt) = \sum_{n \in \mathbb{Z}} \epsilon_{t_n - s}.$$

Here the t_n 's represent the arrival dates.

This process will be assumed to be simple in the sense that a.s. $t_n < t_{n+1}$ and to have a finite intensity μ . With this point process, one associates its Palm probability (see [NEV 76]) denoted as \hat{P}_k . \hat{P}_k is a probability measure on $(\hat{\Omega}, \hat{F}_k, \theta_{t_1})$, the so called Palm space of K where

$$(2.3) \quad \left\{ \begin{array}{l} \hat{\Omega}_k = \Omega \cap (t_0 = 0), \\ \hat{F}_k = F \cap (t_0 = 0). \end{array} \right.$$

All the service variables are defined on $\hat{\Omega}_k$. We shall denote as $\beta : \hat{\Omega}_k \rightarrow \mathbb{R}^+$, the service required by the customer arriving at time t_0 (so that the customer arriving at time t_n requires $\beta \circ \theta_{t_n}$). It will be assumed that $E_{\hat{P}_k} [\beta] < \infty$.

Similarly, the speed process will be represented by a stochastic process A on (Ω, F, P, θ_t) with values in \mathbb{R}^+ :

$$(2.4) \quad \omega \rightarrow A(\omega, t), \quad \forall t \in \mathbb{R}.$$

This process will be assumed to be right continuous, with left limits and to be such that (stationarity)

$$(2.5) \quad A(\omega, t) = A(\Theta_t(\omega), 0), \quad \forall t \in \mathbb{R}$$

and (integrability)

$$(2.6) \quad E_p[A(\omega, t)] < \infty.$$

The particular case of a piecewise constant speed, which corresponds to the initial models of Purdue and Neuts can be obtained as follows : The set of dates where the speed changes is represented by a stationary point process J with intensity $\lambda < \infty$:

$$(2.7) \quad \omega \rightarrow J(\omega, dt) = \sum_{n \in \mathbb{Z}} \epsilon_{s_n}$$

and with associated Palm space $(\hat{\Omega}_j, \hat{F}_j, \hat{P}_j, \theta_{s_1})$. The speed of the server between s_0 and s_1 is then defined by a nonnegative random variable $\alpha : \hat{\Omega}_j \rightarrow \mathbb{R}^+$. In this case, the speed process is defined by the relations :

$$(2.8) \quad \left\{ \begin{array}{l} A(\omega, 0) = \alpha \theta_{s_0}(\omega), \\ A(\omega, t) = A(\Theta_t(\omega), 0). \end{array} \right.$$

3 - THE QUEUING MODEL

The basic probability space to be used in this section is $(\hat{\Omega}_k, \hat{F}_k, \hat{P}_k)$. We shall use the following notations

$$(3.1) \quad \tau = t_1 - t_0$$

and

$$(3.2) \quad \theta = \theta_\tau.$$

Let $T : \hat{\Omega}_k \rightarrow \mathbb{R}^+$ be the maximum amount of work which can be provided by the server between $t_0 = 0$ and t_1 . Clearly

$$(3.3) \quad T = \int_{t_0}^{t_1} A(s) ds = \int_0^\tau A(s) ds.$$

Hence, the workload process in the queue with initial workload zero is defined by the following recursion on $(\hat{\Omega}_k, \hat{F}_k, \hat{P}_k)$:

$$(3.4) \quad \left\{ \begin{array}{l} W_{n+1} = [W_n + \beta\theta^n - T\theta^n]^+, \quad \forall n \geq 0, \\ W_0 = 0, \end{array} \right.$$

where θ^n denotes the n -th iterate of θ , $n \in \mathbb{N}$. Consider, for later use, the Loynes sequence $(V_n, n \in \mathbb{N})$ associated with the sequence $(W_n, n \in \mathbb{N})$. That is the sequence defined on $(\hat{\Omega}_k, \hat{F}_k, \hat{P}_k)$ by the equations

$$(3.5) \quad \left\{ \begin{array}{l} V_{n+1} \circ \theta = [V_n + \beta - T]^+, \quad \forall n \geq 0, \\ V_0 = 0. \end{array} \right.$$

Due to the stationary assumption,

$$(3.6) \quad V_n = W_n \circ \theta^{-n} \sim W_n, \quad \forall n \in \mathbb{N},$$

where \sim denotes the equivalence in law.

The stability condition generalizing the one obtained by Purdue and Neuts for certain particular cases is now derived in the following theorem :

Theorem 1

Under the foregoing assumptions, if the condition

$$(3.7) \quad \mu E_{\hat{P}_k} [\beta] < E_p [A(0)]$$

is satisfied, the queue is stable in the sense that W_n converges in law to a finite random variable W_∞ .

Proof

It is well known that under the stationarity and ergodicity assumptions of section 2, if $E_{\hat{p}_k} [T] < \infty$ and

$$(3.8) \quad E_{\hat{p}_k} [\beta] < E_{\hat{p}_k} [T],$$

then, the increasing sequence $(V_n, n \in \mathbb{N})$ converges a.s. to a finite random variable V_∞ (see [LOY 62] or [MIY 85]). Owing to (3.6), the theorem will thus be proved for $W_\infty = V_\infty$ if we show that (3.7) and (3.8) are equivalent and that $E_{\hat{p}_k} [T] < \infty$.

In view of (2.6) and (3.3), we get

$$(3.9) \quad E_{\hat{p}_k} [T] = E_{\hat{p}_k} \left[\int_0^{t_1} A(\theta_s(\omega), 0) ds \right],$$

whereas the Palm inversion formula ([NEV 76]) yields

$$(3.10) \quad E_{\hat{p}_k} \left[\int_0^{t_1} A(\theta_s(\omega), 0) ds \right] = \frac{1}{\mu} E_p [A(\omega, 0)].$$

The proof of the theorem is thus concluded. \square

Remark 1 It can be shown that if $\mu E_{\hat{p}_k} [\beta] > E_p [A(\omega, 0)]$, the queue is unstable in the sense that $W_n \uparrow \infty$ a.s.. The proof is the same as the one for classical G/G/1 queues.

Remark 2 In the particular case of a piecewise constant speed process $E_p [A(\omega, 0)]$ can be rewritten as follows when using the inversion formula :

$$\begin{aligned} E_p [A(0)] &= E_p [\alpha \circ \theta_{s_0}] = \lambda E_{\hat{p}_j} \left[\int_0^{s_1} \alpha \circ \theta_{s_0} \circ \theta_t dt \right] \\ &= \lambda E_{\hat{p}_j} \left[\int_0^{s_1} \alpha \circ \theta_{s_0} dt \right] \\ &= \lambda E_{\hat{p}_j} [\alpha \cdot \sigma], \end{aligned}$$

where $\sigma = s_1 - s_0$. Thus, in this case, the stability condition also reads

$$(3.11) \quad \mu E_{\hat{P}_k} [\beta] < \lambda E_{\hat{P}_j} [\alpha \cdot \sigma].$$

4 - A LOWER BOUND

Even in the case with i.i.d. service and inter-arrival sequences and say Markovian speed process, the computation of the moments of the variable W_∞ may be quite hard. In this section, we show that in the case where the processes $(A(t), t \in \mathbb{R})$ and $((t_i, \beta_0 \theta^i) \mid i \in \mathbb{Z})$ are independent, easily computable lower bounds can be derived for the moments of W_∞ from a stochastic comparison with a queue with fixed environment.

The probability space throughout this section will be the Palm space of $K : (\hat{\Omega}_k, \hat{F}_k, \hat{P}_k)$ and it will be assumed that the condition (3.7) of theorem 1 is fulfilled.

The workload equation for the queue with fixed speed $a \equiv E_p[A(0)]$ reads :

$$(4.1) \quad \left\{ \begin{array}{l} \tilde{W}_{n+1} = [\tilde{W}_n + \beta_0 \theta^n - \tau_0 \theta^n \cdot a]^+, \quad \forall n \geq 0, \\ \tilde{W}_0 = 0. \end{array} \right.$$

Since the stability condition for this queue coincides with (3.7), if this condition holds, the sequence \tilde{W}_n also converges weakly to a finite random variable \tilde{W}_∞ which can be interpreted as the stationary workload in a classical G/G/1 queue. We shall also need later on the Loynes sequence associated with $(\tilde{W}_n, n \in \mathbb{N})$, which will be denoted as $(\tilde{V}_n, n \in \mathbb{N})$:

$$(4.2) \quad \tilde{V}_n = \tilde{W}_n \circ \theta^{-n}, \quad \forall n \in \mathbb{N}.$$

Let H be the sub σ -algebra of \hat{F}_k generated by the variables $(\tau_0 \theta^i, \beta_0 \theta^i, i \in \mathbb{Z})$ and G be the one generated by $(A(t), t \in \mathbb{R})$.

Lemma 1

If G is independent of H with respect to \hat{P}_k , the following relation holds :

$$(4.3) \quad E_{\hat{P}_k} [T_0 \theta^n | H] = t_0 \theta^n \cdot a, \quad \forall n \in \mathbb{N} \quad \hat{P}_k \text{ a.s.}$$

Proof

We have

$$(4.4) \quad \begin{aligned} E_{\hat{P}_k} [T | H] &= E_{\hat{P}_k} \left[\int_{t_0}^{t_1} A(\omega, s) ds | H \right] \\ &= \int_{-\infty}^{+\infty} E_{\hat{P}_k} [1_{(t_0, t_1)}(s) A(\omega, s) | H] ds \\ &= \int_{-\infty}^{+\infty} 1_{(t_0, t_1)}(s) E_{\hat{P}_k} [A(\omega, s) | H] ds, \end{aligned}$$

where we have used the H-mesurability of t_0 and t_1 to get the last expression. But in view of the independence assumption

$$(4.5) \quad E_{\hat{P}_k} [A(\omega, s) | H] = E_p [A(\omega, s)] = a,$$

hence the result for $n = 0$. The general formula is then obtain from lemma 2 (Appendix).

Theorem 2

Under the assumptions of lemma 1, for all $n \in \mathbb{N}$, W_n and V_n are integrable and

$$(4.6) \quad E_{\hat{P}_k} [W_n | H] \geq \tilde{W}_n,$$

$$(4.7) \quad E_{\hat{P}_k} [V_n | H] \geq \tilde{V}_n.$$

Proof

The integrability is obtained by induction from the integrability of β and T .

Owing to (3.6) and (4.2) and to lemma 2 (Appendix), equations (4.6) and (4.7) are equivalent. Let us prove (4.6) by induction. This relation obviously holds for $n = 0$. Assume it holds for n . Applying Jensen's inequality for conditional expectation in (3.4) (the function $x \rightarrow [x]^+$ is convex), we get

$$(4.8) \quad E_{\hat{p}_k} [W_{n+1} | H] \geq [E_{\hat{p}_k} [W_n | H] + E_{\hat{p}_k} [\beta_0 \theta^n | H] - E_{\hat{p}_k} [T_0 \theta^n | H]]^+.$$

From the very definition of H

$$(4.9) \quad E_{\hat{p}_k} [\beta_0 \theta^n | H] = \beta_0 \theta^n,$$

whereas from lemma 1

$$(4.10) \quad E_{\hat{p}_k} [T_0 \theta^n | H] = \tau_0 \theta^n .a.$$

We have thus

$$(4.11) \quad E_{\hat{p}_k} [W_{n+1} | H] \geq [E_{\hat{p}_k} [W_n | H] + \beta_0 \theta^n - \tau_0 \theta^n .a]^+ \geq [\tilde{W}_n + \beta_0 \theta^n - \tau_0 \theta^n .a]^+,$$

where we have used the induction assumption to get the last inequality. Thus in view of (4.1)

$$(4.12) \quad E_{\hat{p}_k} [W_{n+1} | H] \geq \tilde{W}_{n+1},$$

which concludes the proof. \square

Corollary 1

Let $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ be any increasing and convex function, then

$$(4.13) \quad E_{\hat{p}_k} [f(W_n)] \geq E_{\hat{p}_k} [f(\tilde{W}_n)], \quad \forall n \in \mathbb{N},$$

provided these expectations exist. \square

Proof

Using Jensen's inequality again, we get from (4.6)

$$(4.14) \quad E_{\hat{P}_k} [f(W_n) | H] \geq f(\bar{W}_n), \quad \forall n \in \mathbb{N},$$

so that taking expectation with respect to \hat{P}_k , we get

$$(4.15) \quad E_{\hat{P}_k} [f(W_n)] \geq E_{\hat{P}_k} [f(\bar{W}_n)], \quad \forall n \in \mathbb{N},$$

provided the expectations exist. \square

Corollary 2

Let f be as in corollary 1. If (3.7) holds, then

$$(4.16) \quad E_{\hat{P}_k} [f(W_\infty) | H] \geq f(\bar{W}_\infty)$$

and

$$(4.17) \quad E_{\hat{P}_k} [f(W_\infty)] \geq E_{\hat{P}_k} [f(\bar{W}_\infty)]$$

provided these expectations exist. \square

Proof

The existence of $E_{\hat{P}_k} [f(W_\infty)]$ entails the existence of $E_{\hat{P}_k} [f(V_n)]$ for all $n \in \mathbb{N}$, since f is increasing and $V_n \leq W_\infty$. (A similar result holds for \bar{V}_n). In view of this, we get from (4.7) and Jensen's inequality that

$$(4.18) \quad E_{\hat{P}_k} [f(V_n) | H] \geq f(\bar{V}_n), \quad \forall n \in \mathbb{N}.$$

Since $V_n \uparrow W_\infty$ and $\bar{V}_n \uparrow \bar{W}_\infty$, we get (4.16) from the monotone convergence theorem used in (4.18), whereas (4.17) is obtained when taking expectations in (4.16).

Remark 1

Corollaries 1 and 2 can be rephrased in terms of the convex ordering relation \leq_c of [STO 83] as (4.19) and (4.20) respectively

$$(4.19) \quad W_n \leq_c \tilde{W}_n, \quad \forall n \in \mathbb{N}$$

and

$$(4.20) \quad W_\infty \leq_c \tilde{W}_\infty.$$

Remark 2

The basic methodology used to derive the preceding results can be used to derive the following general comparison theorem between queuing systems in random environment.

Let $(A(t), t \in \mathbb{R})$ and $(A'(t), t \in \mathbb{R})$ be two stationary speed processes on $(\Omega, \mathcal{F}, \mathbb{P}, \theta_t)$ being both independent of the input sequence (i.e. independent of H with the notations of lemma 1). If A' is "smoother" than A in the sense that there exists a sub σ -field L of $\hat{\mathcal{F}}_k$ independent of H such that

$$(4.21) \quad E_{\hat{\mathbb{P}}_k} [A(\omega, t) \mid L] = A'(\omega, t), \quad \forall t \in \mathbb{R}, \quad \hat{\mathbb{P}}_k \text{ a.s.},$$

then, the workload processes $(W_n, n \in \mathbb{N})$ and $(W'_n, n \in \mathbb{N})$ obtained from the respective random environments A and A' satisfy the relations

$$(4.22) \quad W_n \geq_c W'_n, \quad \forall n \in \mathbb{N}$$

and

$$(4.23) \quad W_\infty \geq_c W'_\infty.$$

provided condition (3.7) is fulfilled.

Remark 3

It can also be shown that this bounding methodology can be continued to certain types of feedforward queuing networks in random environments including the

class of series networks : the total delays through such networks are always larger for convex ordering than the corresponding delays in the same network with fixed environment equal to the average environment. Here again, these bounds hold both for steady state and transient behavior (see [BAC 86]).

Acknowledgements

The authors would like to thank both P. Brémaud and the referee who suggested several improvements of the initial version of this paper.

APPENDIX

Lemma 2

Let V be an integrable positive random variable defined on the probability space $(\hat{\Omega}_k, \hat{\mathcal{F}}_k, \hat{\mathbb{P}}_k)$ and H be a θ -invariant sub σ -algebra of $\hat{\mathcal{F}}_k$. Then

$$(A.1) \quad E_{\hat{\mathbb{P}}_k} [V \circ \theta | H] = E_{\hat{\mathbb{P}}_k} [V | H] \circ \theta \quad \hat{\mathbb{P}}_k \text{ a.s.} \quad \square$$

Proof

Consider any set $A \in H$ and write :

$$\begin{aligned} E_{\hat{\mathbb{P}}_k} [V \circ \theta 1_A] &= E_{\hat{\mathbb{P}}_k} [V 1_{A \circ \theta^{-1}}] \quad (\hat{\mathbb{P}}_k \text{ is } \theta\text{-invariant}) \\ &= E_{\hat{\mathbb{P}}_k} [E_{\hat{\mathbb{P}}_k} [V | H] 1_{A \circ \theta^{-1}}] \quad (H \text{ is } \theta\text{-invariant}) \\ &= E_{\hat{\mathbb{P}}_k} [E_{\hat{\mathbb{P}}_k} [V | H] \circ \theta \cdot 1_A]. \quad \square \end{aligned}$$

BIBLIOGRAPHY

- [BAC 86] BACCELLI F. and MASSEY W.A.
"Series Parallel, Fork Join Queueing networks and their Stochastic ordering" Technical Memorandum, ATT Bell Laboratories, 1986.
- [FRA 81] FRANKEN P., KOENIG D., ARNDT V., SCHMIDT V.
"Queues and Point Processes", J. Wiley, New York, 1981.
- [LOY 62] LOYNES R.M.
"The stability of a queue with non independent inter-arrival and service times". Proc. Cambridge Philos. Soc. 58, 1962.
- [MIY 85] MIYAZAWA J.
"The intensity conservation law for queues with randomly changed service rates". J.A.P. 22, 408-418, 1985.
- [NEU 78] NEUTS M.F.
"The M/M/1 queue with randomly varying arrival and service rates". Opsearch, Volume 15, 1978, pp. 139-157.
- [NEU 78] NEUTS M.F.
"Further results on the M/M/1 queue with randomly varying rates". Opsearch, Volume 15, 1978, pp. 158-168.
- [NEV 76] NEVEU J.
"Cours sur les processus ponctuels" Ecole d'Eté de Probabilité de St Flour, Lecture Notes in Mathematics # 598, Springer Verlag, New York, 1976.
- [PUR 74] PURDUE P.R.
"The M/M/1 queue in a random environment". Operations Research, Volume 22, 1974, pp 562-569.
- [STO 83] STOYAN D.
"Comparison methods for queues and other stochastic models". John Wiley and Sons, 1983.

