



On the Rates of Convergence of Erlang's Model

Christine Fricker, Philippe Robert, Danielle Tibi

► **To cite this version:**

Christine Fricker, Philippe Robert, Danielle Tibi. On the Rates of Convergence of Erlang's Model. [Research Report] RR-3368, INRIA. 1998. <inria-00073321>

HAL Id: inria-00073321

<https://hal.inria.fr/inria-00073321>

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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the rates of convergence of Erlang's model

Christine Fricker, Philippe Robert, Danielle Tibi

No 3368

Février 1998

THÈME 2



*R*apport
de recherche



On the rates of convergence of Erlang's model

Christine Fricker*, Philippe Robert*, Danielle Tibi*

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n° 3368 — Février 1998 — 23 pages

Abstract: The convergence to equilibrium of the renormalized $M/M/N/N$ queue is analyzed. Some upper bounds on the distance to equilibrium are given and we prove the cutoff property for two regimes of this queue. We use simple probabilistic methods, such as coupling technics and martingales, to derive our results.

Key-words: The $M/M/N/N$ queue. Convergence to equilibrium. Cutoff property. Hitting times of the $M/M/\infty$ queue. Ornstein-Uhlenbeck processes. Gamma functions.

(Résumé : tsvp)

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

* Université de Paris 7, URA 1321, 2 Place Jussieu, 75251 Paris Cedex 05, France

Sur les vitesses de convergence des modèles d'Erlang

Résumé : On étudie la convergence à l'équilibre de la file d'attente $M/M/N/N$ renormalisée. Des bornes sur la distance à l'équilibre sont obtenues et nous montrons la propriété de cutoff pour deux régimes de cette file. Nos méthodes sont essentiellement probabilistes.

Mots-clé : La file $M/M/N/N$. Convergence à l'équilibre. La propriété de cutoff. Les temps d'atteinte de la file $M/M/\infty$. Les processus d'Ornstein-Uhlenbeck. Les fonctions gamma

1 Introduction

We consider here the classical M/M/N/N queue with input rate λ and release rate 1. As it stands, this model is very well known and basic. Its invariant distribution π_N is given by the truncated Poisson distribution with parameter λ . An important part of the work on the Erlang loss systems has concerned the loss probability in a transient mode $P(X_N(t) = N)$, if $X_N(t)$ is the number of customers at time t . Most of the results have been proved via clever analytical methods; various analytic expansions have been proposed for this model and its generalizations, see for example [20], [18], [12] or [19] and the references therein. Our purpose in this note is to give some insight on the time to reach equilibrium. We try as far as possible to give, using probabilistic methods, explicit and computable bounds on the rates of convergence of the Markov chains associated to these queueing systems. Our approach is mainly probabilistic using simple tools such as martingales and coupling technics. As a by product we derive some results for the M/M/ ∞ queue using martingale methods.

Because the process $(X_N(t))$ is reversible, it is possible, using a diagonalization procedure, to express all the related quantities concerning the approach to equilibrium. Nevertheless, the explicit expression of the eigenvalues and the eigenvectors is not easy to handle to provide useful expressions. We shall study the renormalized process, i.e. the case where λ is replaced by λN . The normalizing parameter N "simplifies" the analysis of the model by giving the major mode of the convergence as $N \rightarrow +\infty$. Despite the underlying phenomena driving the convergence to equilibrium are very simple and already apparent in the analysis of the convergence of the renormalized processes, or at the logarithmic level in the large deviations analysis (see [22]), it turns out that getting simple, accurate and explicit bounds on the rate of convergence is not easy; in this note we give a partial answer to this problem. We also prove that the associated Markov chain has the cutoff property: before some time, which depends on N , the chain is very far from equilibrium for some initial states, but after this time the chain is at equilibrium.

2 The renormalized model

We recall here some elementary results concerning the renormalization of this queue. If $X_N(t)$ (we take the right continuous version of $t \rightarrow X_N(t)$) denotes the number of customers at time t , then the following differential equation is satisfied

$$dX_N(t) = 1_{\{X_N(t-) < N\}} \mathcal{N}_{\lambda N}(dt) - \sum_1^{X_N(t-)} \mathcal{N}_1^i(dt), t \geq 0, \quad (1)$$

where \mathcal{N}_x denotes a Poisson process with parameter x and we add an upper index if we need many of them. All the Poisson processes are assumed to be independent. The generator of this Markov process is given by

$$\Omega_N(f) = \lambda N(f(x+1) - f(x))1_{\{x < N\}} + x(f(x-1) - f(x)), \quad (2)$$

for any function f on the integers.

The equation (1) can be rewritten as

$$dX_N(t) = dM_N(t) + \lambda N \mathbf{1}_{\{X_N(t-) < N\}} dt - X_N(t-) dt, \quad t \geq 0, \quad (3)$$

where $M_N(t)$ is a martingale defined by

$$M_N(t) = \int_0^t \mathbf{1}_{\{X_N(s-) < N\}} (\mathcal{N}_{\lambda N}(ds) - \lambda N ds) - \sum_{i=1}^{+\infty} \int_0^t \mathbf{1}_{\{X_N(s-) \geq i\}} (\mathcal{N}_1^i(ds) - ds), \quad (4)$$

whose quadratic variation is given by

$$\lambda N \int_0^t \mathbf{1}_{\{X_N(s-) < N\}} ds + \int_0^t X_N(s-) ds.$$

Because $X_N(t) \leq \mathcal{N}_{\lambda N}([0, t])$, we get that

$$\lim_{N \rightarrow +\infty} E \left(\left(\frac{M_N(t)}{N} \right)^2 \right) = 0.$$

Using Doob's inequality, it is straightforward to prove that if $\lim_{N \rightarrow +\infty} \frac{X_N(0)}{N} = \bar{X}(0)$, almost surely

$$\frac{1}{N} \left(X_N(t) - X_N(0) - \lambda N \int_0^t \mathbf{1}_{\{X_N(s-) < N\}} ds + \int_0^t X_N(s-) ds \right)$$

converges to 0 uniformly on compact sets. Hence almost surely $\bar{X}_N(t) = \frac{X_N(t)}{N}$ converges uniformly on compact sets to $\bar{X}(t)$ such that

$$d\bar{X}(t) = \begin{cases} \lambda dt - \bar{X}(t) dt & \text{if } \bar{X}(t) < 1 \\ (\lambda \wedge 1 - \bar{X}(t)) dt & \text{if } \bar{X}(t) = 1 \end{cases}$$

with $a \wedge b = \min(a, b)$; equivalently

$$\bar{X}(t) = (\lambda + (\bar{X}(0) - \lambda)e^{-t}) \wedge 1.$$

Three different regimes. As it is obvious from the previous equation, the model has three regimes. For each of them, we describe its behavior on a normal space scale; these results are well known and easily proved using the theorems on the convergence of a sequence of Markov chains and their invariant measures, see [8].

1. $\lambda > 1$. At the scale of the renormalized process, the queue becomes full after a finite time. The variable $Y_N(t) = N - X_N(\frac{t}{N})$ is a Markov process whose generator is given by

$$\lambda(f(y-1) - f(y))\mathbf{1}_{\{y>0\}} + \left(1 - \frac{y}{N}\right)(f(y+1) - f(y)), \quad y \in \mathbb{N},$$

for some function f . As $N \rightarrow +\infty$, this generator converges to the generator of a birth and death process whose jumps of $+1$ [resp. -1] have intensity 1 [resp. λ]. The invariant measure of this asymptotic generator is the geometric distribution with parameter $\frac{1}{\lambda}$ and it is the limit, in distribution, of the invariant measures of $(Y_N(t))$. Asymptotically the queue has a geometric number of idle servers.

2. $\lambda < 1$. The queue is never full. In this case consider the process $Y_N(t) = \frac{X_N(t) - \lambda N}{\sqrt{N}}$, its generator is given by

$$\lambda N \left(f\left(y + \frac{1}{\sqrt{N}}\right) + f\left(y - \frac{1}{\sqrt{N}}\right) - 2f(y) \right) + y\sqrt{N} \left(f\left(y - \frac{1}{\sqrt{N}}\right) - f(y) \right),$$

if $y < (1 - \lambda)\sqrt{N}$, for some measurable function f on \mathbb{R} . If f is twice differentiable on \mathbb{R} , then the generator converges to

$$\lambda f''(y) - yf'(y), \quad y \in \mathbb{R},$$

which is the generator of an Ornstein-Ühlenbeck process with parameter λ . The invariant measure of $(Y_N(t))$ converges to a normal distribution with variance λ .

3. $\lambda = 1$. The queue becomes full at infinity. The process $Y_N(t) = \frac{N - X_N(t)}{\sqrt{N}}$, its generator is given by

$$N \left(f\left(y + \frac{1}{\sqrt{N}}\right) + f\left(y - \frac{1}{\sqrt{N}}\right) - 2f(y) \right) - y\sqrt{N} \left(f\left(y + \frac{1}{\sqrt{N}}\right) - f(y) \right),$$

for $y > 0$ and by

$$N \left(f\left(\frac{1}{\sqrt{N}}\right) - f(0) \right),$$

for $y = 0$. Asymptotically for f twice differentiable on \mathbb{R}_+ such that $f'(0) = 0$, the generator converges to

$$\begin{cases} f''(y) - yf'(y) & y > 0 \\ \frac{1}{2}f''(0) & y = 0 \end{cases},$$

which is the generator of the reflected Ornstein-Ühlenbeck process on \mathbb{R}_+ . The invariant measure of $(Y_N(t))$ converges in distribution to the absolute value of a standard normal random variable.

Remark 1. The renormalizing properties of Erlang loss networks are presented in [15] and results on the convergence to Ornstein-Ühlenbeck processes can be found in [24] and [14].

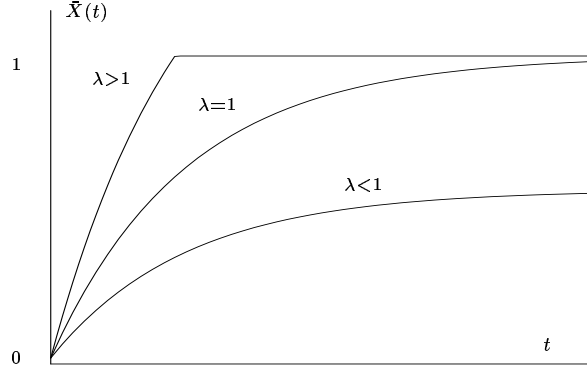


Figure 1: The three different regimes

3 The $M/M/\infty$ process

In the following it will be useful to consider the process $(Y_N(t))$ without any upper boundary, i.e. that satisfies the following differential equation

$$dY_N(t) = \mathcal{N}_{\lambda N}(dt) - \sum_1^{Y_N(t-)} \mathcal{N}_1^i(dt), t \geq 0. \quad (5)$$

Our main result in this section is the following proposition.

Proposition 1. For $c \geq 0$,

$$(\mathcal{E}_c(t)) = \left((1 + ce^t)^{Y_N(t)} e^{-\lambda N c e^t} \right) \quad (6)$$

is a martingale. Consequently, if $Y_N(0) = x \in \mathbb{N}$, for each $\alpha > 0$,

$$\left(e^{-\alpha t} \int_0^{+\infty} (1 + c)^{Y_N(t)} c^{\alpha-1} e^{-\lambda N c} dc \right),$$

is also a martingale.

Remark 2. The martingale $(\mathcal{E}_c(t))$ is central for this model. It plays the role of the exponential martingale (or Wald's martingale) for the random walks. As far as we know it is not present in the literature, although many of the results for this model can be obtained with this martingale. If its expression is complicated, it is quite natural in the following sense. It

is well known, see Chihara [6], that for $w \geq 0$,

$$(1+w)^x e^{-\lambda N w} = \sum_{n=0}^{+\infty} C_n^{\lambda N}(x) \frac{w^n}{n!}, \quad (7)$$

where $C_n^{\lambda N}(x)$ is the n -th Poisson-Charlier polynomial. The Poisson-Charlier polynomials are orthogonal polynomials with respect to the Poisson distribution with parameter λN ; they are associated with the Poisson process. From an analytical point of view, they are related to the differential equations of the birth and death process $(Y_N(t))$, see [16]. In this setting, the above proposition can be rephrased as follows : the function $\phi : (t, x) \rightarrow (1 + ce^t)^x e^{-\lambda N ce^t}$ is space-time harmonic with respect to the generator Λ_N of $(Y_N(t))$, i.e.

$$\frac{\partial \phi}{\partial t}(t, x) = -\Lambda_N(\phi)(t, x) = -\lambda N (\phi(t, x+1) - \phi(t, x)) - x(\phi(t, x-1) - \phi(t, x)).$$

Replacing w by ce^t in (7) and identifying the coefficients of the expansion with respect to c , we deduce the following corollary.

Corollary 2. For any $n \geq 0$, if $C_n^a(x)$ denotes the n -th Poisson-Charlier polynomial with parameter a ,

$$(e^{nt} C_n^{\lambda N}(Y_N(t))),$$

is a martingale, in particular

$$(e^t (Y_N(t) - \lambda N)), \quad (e^{2t} (Y_N(t)^2 - (1 + 2\lambda N)Y_N(t) + \lambda^2 N^2)),$$

are martingales.

Proof of Proposition 1. For $c \geq 0$, $\mathcal{E}_c(t)$ is integrable since $Y_N(t) \leq \mathcal{N}_{\lambda N}[0, t]$. Using the equation (5), we get that

$$\begin{aligned} \frac{d\mathcal{E}_c(t)}{\mathcal{E}_c(t-)} &= Y_N(t-) \frac{ce^t}{1+ce^t} dt - \lambda N ce^t dt + ce^t \mathcal{N}_{\lambda N}(dt) - \frac{ce^t}{1+ce^t} \sum_1^{Y_N(t-)} \mathcal{N}_1^i(dt), \\ &= ce^t (\mathcal{N}_{\lambda N}(dt) - \lambda N dt) - \frac{ce^t}{1+ce^t} \sum_1^{Y_N(t-)} (\mathcal{N}_1^i(dt) - dt), \end{aligned}$$

hence $(\mathcal{E}_c(t))$ is a martingale. If h is some borelian function on \mathbb{R}_+ such that

$$\int_0^{+\infty} |h(c)| E \left((1+c)^{Y_N(0)} \right) e^{-\lambda N c} dc < +\infty,$$

then

$$\left(\int_0^{+\infty} \mathcal{E}_c(t) h(c) dc \right)$$

is also a martingale. We conclude our proposition by taking $h(c) = c^{\alpha-1}$. □

For $x \geq 0$, we denote by T_x the hitting time of x ,

$$T_x = \inf\{s \geq 0 / Y_N(s) = x\}.$$

This variable will, in some cases, measure accurately the time to reach equilibrium for our truncated process. The next proposition presents some results concerning hitting times for the process $(Y_N(t))$. We begin with an elementary lemma.

Lemma 3. *If $(Z(t))$ is a birth and death process on \mathbb{N} , denoting by $\gamma_x > 0$ [resp. δ_x] the birth [resp. death] rate starting from $x \in \mathbb{N}$, then if $a < b$, the probability of never returning to a starting from b is given by*

$$\frac{\sum_{a+1}^b \prod_{a+1}^{k-1} \frac{\delta_i}{\gamma_i}}{\sum_{a+1}^{+\infty} \prod_{a+1}^{k-1} \frac{\delta_i}{\gamma_i}}.$$

Proof. Denote by H_b this probability, then clearly $H_a = 0$ and for $b > a$,

$$(\gamma_b + \delta_b)H_b = \gamma_b H_{b+1} + \delta_b H_{b-1},$$

thus $H_{b+1} - H_b = \frac{\delta_b}{\gamma_b}(H_b - H_{b-1})$. It is then easy to conclude. \square

Proposition 4. *For $\alpha > 0$ and $0 \leq a \leq b$,*

$$E_a(e^{-\alpha T_b}) = \frac{\Gamma_a(\alpha)}{\Gamma_b(\alpha)}, \quad (8)$$

$$E_b(e^{-\alpha T_a}) = \frac{\Psi_b(\alpha)}{\Psi_a(\alpha)}, \quad (9)$$

where for $y \geq 0$,

$$\Gamma_y(\alpha) = \int_0^{+\infty} (1+c)^y c^{\alpha-1} e^{-\lambda N c} dc,$$

and

$$\Psi_y(\alpha) = \Gamma_y(\alpha) \sum_{k=y}^{+\infty} \frac{\Gamma_0(k+1)}{\Gamma_k(\alpha)\Gamma_{k+1}(\alpha)},$$

$\Gamma_0(\alpha)(\lambda N)^\alpha$ being the usual Gamma function.

Remark 3. The relation (8) is an integral representation of a classical result, see Takacs [23]. The Laplace transform (9) has been obtained under another form in [12], the authors used operator technics and complex analysis to derive it.

Proof. If $Y_N(0) = a \leq b$, the stopped martingale

$$Z(t) = e^{-\alpha(t \wedge T_b)} \int_0^{+\infty} (1+c)^{Y_N(t \wedge T_b)} c^{\alpha-1} e^{-\lambda N c} dc,$$

is bounded, hence it is converging in L_1 , so

$$E_a(Z(T_b)) = E_a(Z(0)),$$

$$E_a \left(e^{-\alpha T_b} \int_0^{+\infty} (1+c)^{Y_N(T_b)} c^{\alpha-1} e^{-\lambda N c} dc \right) = E_a \left(e^{-\alpha T_b} \right) \Gamma_b(\alpha) = \Gamma_a(\alpha),$$

which proves the first part of the proposition.

We assume that $Y_N(0) = b \geq a$, if we take $x > b$, using again the martingale of Proposition 1 and $T = T_a \wedge T_x$, we get

$$1 = \frac{\Gamma_b(\alpha)}{E_b \left(e^{-\alpha T_a} 1_{\{T_a < T_x\}} \right) \Gamma_a(\alpha) + E_b \left(e^{-\alpha T_x} 1_{\{T_x \leq T_a\}} \right) \Gamma_x(\alpha)}. \quad (10)$$

At this point, it is natural to let $x \rightarrow +\infty$, the first term of the denominator of the above equation converges to $E_b \left(e^{-\alpha T_a} \right) \Gamma_a(\alpha)$ which is the quantity of interest. To get the behavior of the second term

$$J_x = E_b \left(e^{-\alpha T_x} 1_{\{T_x \leq T_a\}} \right) \Gamma_x(\alpha),$$

we shall use a technic of change of measure. Let us introduce some notations; define Q_b the probability distribution on the sample paths of $(Y_N(t))$ starting from b by

$$Q_b(f) = \frac{1}{\Gamma_b(\alpha)} E_b \left(f \cdot e^{-\alpha t} \Gamma_{Y_N(t)}(\alpha) \right),$$

for any measurable functional f on the sample paths of $(Y_N(t))$, depending only of the coordinates $\leq t$. Because $\left(\frac{e^{-\alpha t} \Gamma_{Y_N(t)}(\alpha)}{\Gamma_b(\alpha)} \right)$ is a positive martingale whose expected value is 1, the definition of Q_b is consistent, see [21] for example. Under Q_b , the process $(Y_N(t))$ is also a birth and death process whose transitions are given by

$$Q_b(Y_N(t+h) = k+1 / Y_N(t) = k) = \frac{P_b(Y_N(t+h) = k+1, Y_N(t) = k) e^{-\alpha(t+h)} \Gamma_{k+1}(\alpha)}{P_b(Y_N(t) = k) e^{-\alpha t} \Gamma_k(\alpha)},$$

hence starting from k , the birth rate is $\lambda N \frac{\Gamma_{k+1}(\alpha)}{\Gamma_k(\alpha)}$ and the death rate $k \frac{\Gamma_{k-1}(\alpha)}{\Gamma_k(\alpha)}$. The stopped martingale

$$\left(e^{-\alpha T_x \wedge t} \Gamma_{Y_N(T_x \wedge t)}(\alpha) \right)$$

is uniformly integrable, depends only of the coordinates $\leq T_x$, and its limit is $e^{-\alpha T_x} \Gamma_x(\alpha)$, hence for $t \in \mathbb{R}$,

$$\lim_{t \rightarrow +\infty} E \left(e^{-\alpha(T_x \wedge t)} \Gamma_{Y_N(T_x \wedge t)} \mathbf{1}_{\{T_x \wedge t \leq T_a\}} \right) = J_x.$$

According to the definition of Q_b ,

$$E \left(e^{-\alpha(T_x \wedge t)} \Gamma_{Y_N(T_x \wedge t)} \mathbf{1}_{\{T_x \wedge t \leq T_a\}} \right) = Q_b(T_x \wedge t \leq T_a),$$

consequently we get that

$$J_x = \Gamma_b(\alpha) Q_b(T_x \leq T_a),$$

hence

$$\lim_{x \rightarrow +\infty} J_x = \Gamma_b(\alpha) Q_b(T_a = +\infty),$$

and according to (10),

$$E_b(e^{-\alpha T_a}) = \frac{\Gamma_b(\alpha) Q_b(T_a < +\infty)}{\Gamma_a(\alpha)}.$$

Now we use our lemma to get that

$$Q_b(T_a < +\infty) = \frac{\sum_b^{+\infty} \frac{k!}{(\lambda N)^k \Gamma_k(\alpha) \Gamma_{k+1}(\alpha)}}{\sum_a^{+\infty} \frac{k!}{(\lambda N)^k \Gamma_k(\alpha) \Gamma_{k+1}(\alpha)}},$$

our proposition is proved. \square

Remark 4. If for $\alpha > 0$ and $n \in \mathbb{N}$ we denote by $\phi_n(\alpha) = E_n(e^{-\alpha T_{n-1}})$, then it is easily seen that

$$\phi_n(\alpha) = \frac{n}{n + \lambda N + \alpha - \lambda N \phi_{n+1}(\alpha)},$$

if we replace temporarily λN by λ , we get that $\lambda \phi_n(\alpha - \lambda)$ has a continued fraction representation

$$\frac{\lambda n}{n + \alpha - \frac{\lambda(n+1)}{n+1 + \alpha - \frac{\lambda(n+2)}{n+2 + \alpha - \frac{\lambda(n+3)}{n+3 + \alpha - \dots}}}}$$

This kind of representation occurs in the problems of enumeration of paths. In this framework, families of orthogonal polynomials play also a major role. Despite the combinatorial results, see [9] or chapter 5 of [11], have some similarity with the results obtained from a probabilistic point of view, it is not clear how one can get a clear picture of these phenomena.

Remark 5 (Related results for Ornstein-Ühlenbeck processes). As a stochastic process, the Ornstein-Ühlenbeck process (U_t) verifies the stochastic differential equation

$$dU_t = \sqrt{2\lambda}dB_t - U_t dt,$$

where (B_t) is a standard Brownian motion. Because $\left(\frac{Y_N(t) - \lambda N}{\sqrt{N}}\right)$ converges in distribution to (U_t) , the martingale we have in the discrete case

$$\left(\mathcal{E}_{\frac{c}{\sqrt{N}}}(t)\right) = \left(\left(1 + \frac{c}{\sqrt{N}}e^t\right)^{Y_N(t)} e^{-\lambda N \frac{c}{\sqrt{N}} e^t}\right)$$

converges in distribution to

$$\left(e^{ce^t U_t - \frac{\lambda}{2} c^2 e^{2t}}\right),$$

which is therefore also a martingale; this can also be checked directly using the exponential martingale of the standard Brownian motion. For $\alpha > 0$, as in the discrete case, if we integrate with respect to $c^{\alpha-1}dc$, we get that

$$\left(e^{-\alpha t} \int_0^{+\infty} e^{cU_t - \frac{\lambda}{2} c^2} c^{\alpha-1} dc\right),$$

is a martingale. The classical formulas for the Laplace transforms of the hitting times of the Ornstein-Ühlenbeck process are then easy to derive, see [13] for this kind of approach. If τ_x is the hitting time of x by U_t , we get using the stopped martingale, for $\alpha > 0$ and $a \leq b$, that

$$E_a(e^{-\alpha\tau_b}) = \frac{\int_0^{+\infty} e^{ca - \frac{\lambda}{2} c^2} c^{\alpha-1} dc}{\int_0^{+\infty} e^{cb - \frac{\lambda}{2} c^2} c^{\alpha-1} dc},$$

and by symmetry (or by replacing c by $-c$ in the martingale $\mathcal{E}_{\frac{c}{\sqrt{N}}}$), the Laplace transform of the hitting time has a similar expression,

$$E_b(e^{-\alpha\tau_a}) = \frac{\int_0^{+\infty} e^{-cb - \frac{\lambda}{2} c^2} c^{\alpha-1} dc}{\int_0^{+\infty} e^{-ca - \frac{\lambda}{2} c^2} c^{\alpha-1} dc}.$$

In the discrete case, as we have seen, the distribution of the hitting time of a starting from b is more difficult to express.

4 The time to reach equilibrium

In this section we study the time at which the process is stationary. The quantity we take to measure this is a kind of worst case of the total variation distance, more specifically.

Definition 5. *The distance to equilibrium at time t is defined as*

$$d_N(t) = \max_{0 \leq x \leq N} \|P_x(t) - \pi_N\|_{tv},$$

where P_x is the probability distribution for the process $(X_N(t))$ which starts at x and the norm $\|\cdot\|_{tv}$ is defined as

$$\|P - Q\|_{tv} = \sup_{A \subset \mathbb{N}} |P(A) - Q(A)| = \frac{1}{2} \sum_{i=0}^{+\infty} |P(\{i\}) - Q(\{i\})|,$$

for P and Q probability measures on \mathbb{N} . As we will see, there is a phase transition for all the models we consider. Roughly speaking, before some time t_N , the process is very far from equilibrium, i.e. for $t < t_N$, $d_N(t) \sim 1$ and after that time it is at equilibrium, i.e. $d_N(t) \sim 0$. We prove this result as well as some estimates for $d_N(t)$. This kind of phase transition is often encountered in the context of random walks where a sharp transition occurs, the cutoff property, see [1], [2] and [7]. In a queueing context, the results of this kind obtained without the explicit expressions of the transient behavior of the Markov process are quite rare, see [3].

Throughout this section we shall denote by $(X_N^x(t))$, $x = 0, \dots, N$, the Markov processes starting from x with generator defined by (2). We couple them so that if for some $t \geq 0$, $X_N^x(t) = X_N^y(t)$, then $X_N^x(s) = X_N^y(s)$ for all $s \geq t$. We shall consider only monotonic couplings; for these couplings, when X_N^0 meets X_N^N , then all the processes are on the same trajectory. If Z_N is some independent random variable whose distribution is the truncated Poisson distribution with parameter λN , then $(X_N^{Z_N}(t))$ is a stationary version of the process $(X_N(t))$. The number $d_N(t)$ measures how fast all these Markov processes stick to $(X_N^{Z_N}(t))$.

4.1 The super-critical case, $\lambda > 1$

Here the situation is quite simple, the queue has to be full very quickly and will remain nearly saturated forever. When the process $(X_N^0(t))$ reaches N , all the processes are stuck to the sample path of $(X_N^0(t))$, in particular all of them will be on the stationary sample path. Recall that $T_N = \inf\{s \geq 0 / X_N^0(s) = N\}$, and according to the definition of $d_N(t)$,

$$d_N(t) = \max_{0 \leq x \leq N} \sup_{A \subset \mathbb{N}} |E \left(1_{\{X_N^x(t) \in A\}} - 1_{\{X_N^{Z_N}(t) \in A\}} \right)| \leq P(T_N > t). \quad (11)$$

The following proposition gives an upper bound for $d_N(t)$. As we shall see it is not optimal as a function of N , however it is quite accurate for the dependence in the time variable.

Proposition 6. *The hitting time T_N has an exponential moment of order $N(\sqrt{\lambda} - 1)^2$. Furthermore,*

$$E \left(e^{N(\sqrt{\lambda} - 1)^2 T_N} \right) \leq \lambda^{\frac{N}{2}}, \quad (12)$$

and consequently

$$d_N(t) \leq \lambda^{N/2} e^{-N(\sqrt{\lambda}-1)^2 t}, \quad t \geq 0. \quad (13)$$

Proof. Clearly T_N has the same distribution as the hitting time of N by the process $(Y_N(t))$ starting from 0. We shall use a standard martingale argument. For $\xi \in \mathbb{R}$, denote by $f(x) = e^{\xi x}$, using (5), if $Y_N(0) = 0$, we get

$$df(Y_N(t)) = dM_N(t) + f(Y_N(t-)) (\lambda N(e^\xi - 1) + (e^{-\xi} - 1)Y_N(t-)) dt,$$

for some martingale $(M_N(t))$. For $\xi \leq 0$, the stopped process satisfies,

$$\begin{aligned} df(Y_N(t \wedge T_N)) &\leq dM_N(t \wedge T_N) \\ &\quad + \mathbf{1}_{\{t < T_N\}} (f(Y_N(t-)) (\lambda N(e^\xi - 1) + N(e^{-\xi} - 1))) dt. \end{aligned}$$

If we minimize $(\lambda N(e^\xi - 1) + N(e^{-\xi} - 1))$ for $\xi \leq 0$, we find that $\xi_0 = -\frac{1}{2} \log(\lambda)$ achieves the minimum. For this value of ξ , we get that for $s \leq t$

$$\begin{aligned} f(Y_N(t \wedge T_N)) - f(Y_N(s \wedge T_N)) &\leq M_N(t \wedge T_N) \\ &\quad - M_N(s \wedge T_N) - N(\sqrt{\lambda} - 1)^2 \int_{T_N \wedge s}^{T_N \wedge t} f(Y_N(s-)) ds. \end{aligned}$$

We deduce that $\left(e^{N(\sqrt{\lambda}-1)^2 t \wedge T_N} (\sqrt{\lambda})^{-Y_N(t \wedge T_N)} \right)$ is a supermartingale, in particular

$$E \left(e^{N(\sqrt{\lambda}-1)^2 t \wedge T_N} \right) \leq \sqrt{\lambda}^N.$$

Letting t converges to $+\infty$ gives (12), then using Chebyshev's inequality and (11), we get our upper bound (13). \square

The process is already at equilibrium when $(X_N^0(t))$ hits N , for the renormalized process it corresponds to the time T such that

$$\bar{X}(T) = \lambda(1 - e^{-T}) = 1,$$

that is $T = \log \frac{\lambda}{\lambda-1}$.

Using (13), we find that $d_N(t) < e^{-1}$, if

$$t \geq \frac{1}{2} \frac{\log \lambda}{(\sqrt{\lambda} - 1)^2} + o(1) > \log \frac{\lambda}{\lambda - 1},$$

our bound (13) is thus not really tight as we shall see in the next proposition; we prove that $\log \frac{\lambda}{\lambda-1}$ is indeed the time to reach equilibrium for this model. However, one can show,

see [10], that the second largest eigenvalue $\lambda_{2,N} < 0$ of the generator Ω_N , which gives the exponential rate of convergence to equilibrium, satisfies

$$\lim_{N \rightarrow +\infty} \frac{\lambda_{2,N}}{N} = -(\sqrt{\lambda} - 1)^2.$$

The exponential rate of convergence to 0 in t is thus accurate in (13).

Proposition 7.

$$\lim_{N \rightarrow +\infty} d_N(t) = \begin{cases} 1 & \text{if } t < \log \frac{\lambda}{\lambda-1} \\ 0 & \text{if } t > \log \frac{\lambda}{\lambda-1}. \end{cases}$$

Proof. According to Corollary 2, for the process $(Y_N(t))$ starting from 0, the process $(e^t(Y_N(t) - \lambda N))$ is a martingale. Notice that T_N is also the hitting time of N for $(X_N^0(t))$. Thus

$$E(e^{t \wedge T_N} (X_N^0(t \wedge T_N) - \lambda N)) = -\lambda N, \quad t \geq 0.$$

Using (12), for N sufficiently large, it is easily seen that $E(e^{T_N}) < +\infty$ and Lebesgue's convergence theorem gives

$$E(e^{T_N}) = \frac{\lambda}{\lambda-1}.$$

With the same argument for the martingale

$$(e^{2t} (Y_N(t)^2 - (1 + 2\lambda N)Y_N(t) + \lambda^2 N^2))$$

of Corollary 2, we get, for N such that $N(\sqrt{\lambda} - 1)^2 > 2$, that

$$E(e^{2T_N}) = \frac{\lambda^2}{(\lambda-1)^2 - \frac{1}{N}}.$$

Using Chebyshev's inequality we get that for $\varepsilon > 0$,

$$P\left(T_N - \log \frac{\lambda}{\lambda-1} > \varepsilon\right) \leq \frac{1}{N(e^\varepsilon - 1)^2((\lambda-1)^2 - \frac{1}{N})}.$$

For $t \geq \log \frac{\lambda}{\lambda-1}$, we get that

$$P(T_N > t) \leq \frac{1}{N\left(\frac{\lambda-1}{\lambda}e^t - 1\right)^2((\lambda-1)^2 - \frac{1}{N})},$$

so $\lim_{N \rightarrow +\infty} d_N(t) = 0$.

Now we fix some $t < \log \frac{\lambda}{\lambda-1}$ and $\varepsilon > 0$. We know that the invariant distribution of $(N - X_N(t))$ converges in distribution to a geometric distribution with parameter $\frac{1}{\lambda}$. In

particular, there exists some $K > 0$ such that for all $N \geq 1$, $\tilde{\pi}_N(\{N, \dots, N - K\}) \geq 1 - \varepsilon$. If we denote by T_N^K , the hitting time of $N - K$ by $(X_N^0(t))$ with the same method as before we get

$$E\left(e^{T_N^K}\right) = \frac{\lambda}{\lambda - 1 + \frac{K}{N}}, \quad \text{and} \quad E\left(\left(e^{T_N^K} - \frac{\lambda}{\lambda - 1}\right)^2\right) \leq \frac{C_K}{N},$$

for some constant C_K .

For $t < \log \frac{\lambda}{\lambda - 1}$, we have

$$P(T_N^K \leq t) \leq \frac{C_K(\lambda - 1)^2}{\lambda^2 N \left(\frac{\lambda - 1}{\lambda} e^t - 1\right)^2},$$

hence there exists some N_0 such that $P(T_N^K \leq t) \leq \varepsilon$ for $N \geq N_0$. Now

$$\begin{aligned} \|P(X_N^0(t) \in \cdot) - \pi_N\|_{tv} &= \sup_{A \subset \mathbb{N}} |P(X_N^0(t) \in A) - \pi_N(A)| \\ &\geq |P(X_N^0(t) \geq N - K) - \pi_N(\{N - K, \dots, N\})|, \end{aligned}$$

using that for $N \geq N_0$, $P(X_N^0(t) \geq N - K) \leq P(T_N^K \leq t) \leq \varepsilon$, we get

$$\|P(X_N^0(t) \in \cdot) - \pi_N\|_{tv} \geq 1 - 2\varepsilon.$$

As ε is arbitrary, our proposition is proved. \square

Remark 6.

The picture 2 shows a smooth convergence to equilibrium, the probability distributions drift continuously to the equilibrium measure. In particular, the probability of the region, which is not negligible for the distribution at time t and for the equilibrium distribution, increases smoothly from 0 to 1. The curve of the middle depicts this phenomenon. Here, at our scale, because of the sharp transition, this curve does not exist.

It is easily seen, using (8), that T_N converges in distribution to $\log \frac{\lambda}{\lambda - 1}$. The first part of our proof could have used this property, we shall proceed like that in the next section for the case $\lambda = 1$. Here, we have chosen to use the martingales which show a much stronger convergence, i.e. the convergence of the exponential moment of order 2 of T_N .

4.2 The critical case, $\lambda = 1$

Proposition 8. *If T_N denotes the hitting time of N starting from 0, then $T_N - \frac{\log N}{2}$ converges in distribution to a variable with density on \mathbb{R}*

$$\psi(x) = \sqrt{\frac{2}{\pi}} e^{-x} e^{-\frac{1}{2} e^{-2x}}.$$

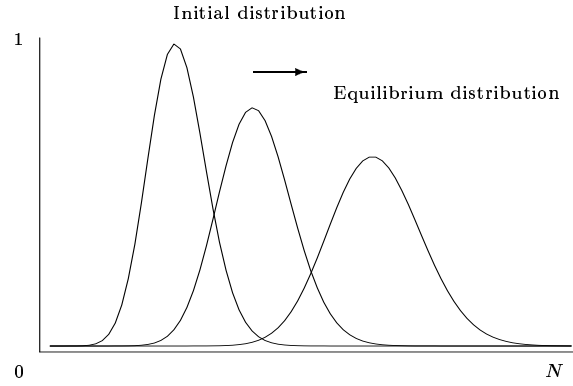


Figure 2: Convergence of the probability distributions to equilibrium

Proof. We shall use the representation (8). For $\alpha > 0$, the function $\Gamma_N(\alpha)$ can be written as

$$N^\alpha \Gamma_N(\alpha) = \sqrt{N}^\alpha 2^{\frac{\alpha}{2}-1} \int_0^{+\infty} \left(1 + \sqrt{\frac{2x}{N}}\right)^N x^{\frac{\alpha}{2}-1} e^{-\sqrt{2Nx}} dx.$$

The expression under the integral in the right hand side of the above equation is easily seen as a non increasing function of N . Using Lebesgue's theorem, we get that

$$\lim_{N \rightarrow +\infty} \frac{\sqrt{N}^\alpha \Gamma_N(\alpha)}{2^{\frac{\alpha}{2}-1}} = \int_0^{+\infty} e^{-x} x^{\frac{\alpha}{2}-1} dx = \Gamma\left(\frac{\alpha}{2}\right),$$

hence

$$\lim_{N \rightarrow +\infty} E\left(e^{-\alpha(T_N - \frac{\log N}{2})}\right) = \frac{\Gamma(\alpha)}{2^{\frac{\alpha}{2}-1} \Gamma\left(\frac{\alpha}{2}\right)},$$

Legendre's formula for the Gamma function, see [25], gives

$$\lim_{N \rightarrow +\infty} E\left(e^{-\alpha(T_N - \frac{\log N}{2})}\right) = \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{\alpha+1}{2}\right) 2^{\frac{\alpha}{2}}.$$

The last expression is the Laplace transform of the probability density on \mathbb{R}

$$\sqrt{\frac{2}{\pi}} e^{-x} e^{-\frac{1}{2}e^{-2x}},$$

which is the distribution of $-\log|B_1|$, where B_1 is some standard normal random variable. Our proposition is proved. \square

Remark 7. The above result shows that the distribution of T_N is sharply concentrated around $\frac{\log N}{2}$, this is the main reason for the phase transition that will be proved in the next proposition.

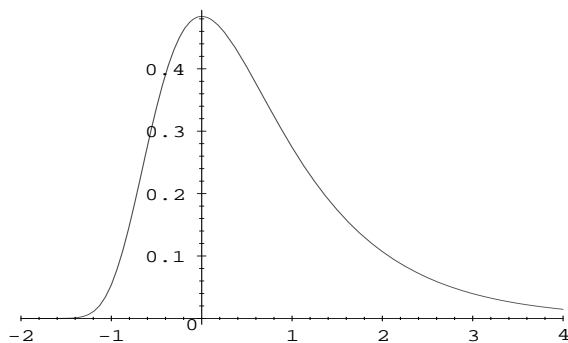


Figure 3: The distribution of $T_N - \frac{\log N}{2}$

Proposition 9. *In the case $\lambda = 1$, then for any sequence $(\phi(N))$ converging to $+\infty$, then*

$$\lim_{N \rightarrow +\infty} d_N \left(\frac{\log N}{2} + a\phi(N) \right) = \begin{cases} 1 & \text{if } a < 0 \\ 0 & \text{if } a > 0 \end{cases} ;$$

moreover for any $a \in \mathbb{R}$,

$$\limsup_{N \rightarrow +\infty} d_N \left(\frac{\log N}{2} + a \right) \leq \sqrt{\frac{2}{\pi}} \int_0^{e^{-a}} e^{-\frac{x^2}{2}} dx. \tag{14}$$

Proof. If we use (11) we get that

$$d_N \left(\frac{\log N}{2} + a\phi(N) \right) \leq P \left(T_N \geq \frac{\log N}{2} + a\phi(N) \right),$$

we have seen before that $T_N - \log N/2$ converges in distribution. The right hand side of the inequality must converge to 0 for $a > 0$. The inequality (14) is also a direct consequence of Proposition 8.

We know that the invariant distribution of $\frac{N - X_N(t)}{\sqrt{N}}$ converges in distribution to the distribution of the absolute value of a normal random variable. Hence, for $\varepsilon > 0$, there exists some $A > 0$ such that, for N sufficiently large,

$$\pi_N \{0, \dots, [N - A\sqrt{N}]\} \leq \varepsilon.$$

Using again (8), we get that for $\alpha > 0$,

$$N^\alpha E_0 \left(e^{-\alpha T_{\lfloor N - A\sqrt{N} \rfloor}} \right) = \frac{\Gamma(\alpha)}{\int_0^{+\infty} \left(1 + \frac{x}{N}\right)^{\lfloor N - A\sqrt{N} \rfloor} x^{\alpha-1} e^{-x} dx},$$

in a similar way as before we get that

$$\lim_{N \rightarrow +\infty} E_0 \left(e^{-\alpha \left(T_{\lfloor N - A\sqrt{N} \rfloor} - \frac{\log N}{2} \right)} \right) = \frac{2\Gamma(\alpha)}{2^{\frac{\alpha}{2}} \int_0^{+\infty} e^{-x - A\sqrt{2x}} x^{\alpha/2-1} dx}. \quad (15)$$

in particular $T_{\lfloor N - A\sqrt{N} \rfloor} - \frac{\log N}{2}$ converges in distribution (notice that the right hand side of the above identity tends to 1 as $\alpha \rightarrow 0$). Because

$$P_0 \left(X_N \left(\frac{\log N}{2} + a\phi(N) \right) \geq \lfloor N - A\sqrt{N} \rfloor \right) \leq P_0 \left(T_{\lfloor N - A\sqrt{N} \rfloor} - \frac{\log N}{2} \leq a\phi(N) \right),$$

using the convergence in distribution of $T_{\lfloor N - A\sqrt{N} \rfloor} - \frac{\log N}{2}$, for $a < 0$, the right hand side tends to 0 as $N \rightarrow +\infty$. We conclude the proof as in the super-critical case. Our proposition is proved. \square

In this case the stationary process lives in a region below N whose width is of the order of \sqrt{N} . The way the process converges to equilibrium can be pictured as follows: First reach the region $\lfloor N - A\sqrt{N} \rfloor$ and then within a bounded time stick to the stationary sample path. If we use the renormalized process, the time T to reach $\lfloor N - A\sqrt{N} \rfloor$ is $\frac{1}{2} \log N + o(1)$ as N gets large. This is indeed the right value to reach equilibrium. The last part of the travel i.e. from $\lfloor N - A\sqrt{N} \rfloor$ to equilibrium is basically the time to reach equilibrium for a reflected Ornstein-Uhlenbeck process starting from A .

Remark 8. Using Diaconis' terminology, see [7], $(X_N(t))$ has a $[\frac{\log N}{2}, \phi(N)]$ cutoff for any sequence $(\phi(N))$ converging to infinity.

4.3 The sub-critical case, $\lambda < 1$

We begin with an upper bound on $d_N(t)$.

Proposition 10.

$$d_N(t) \leq (N+1)e^{-t}, \quad t \geq 0. \quad (16)$$

Proof. We use a coupling argument: We imagine that the jobs are balls in an urn with N places. We consider two processes, process $(X_N^1(t))$ starting with N red balls and process $(X_N^2(t))$ beginning with an empty urn. We couple these two processes as follows: the balls

that arrive are green and the arrivals are the same for both processes. If a green ball is not rejected for the processes 1 and 2, then its residence time in the urn will be given by the same value exponentially distributed for both processes. If a green ball is rejected in 1 but not in 2, then one of the red balls in 1 is painted in green and its residual residence time has the same value as for the new green ball in 2. Notice that with this coupling we have $X_N^2(t) \leq X_N^1(t)$ and $X_N^2(t) < X_N^1(t)$ if and only if there are still some red balls in 1. Furthermore when $X_N^2(t) = X_N^1(t)$, then the two processes are the same after t . When the urn does not contain any red ball, the process is already at equilibrium; remember that the stationary process is between these two processes. The time S_N after which no red ball is present is certainly smaller than the maximum of N independent exponentially distributed random variables (Z_i) with parameter 1, hence

$$P(S_N > t) \leq P\left(\sum_{i=1}^N 1_{\{Z_i \geq t\}} > 0\right).$$

If Q_N is the distribution of $\sum_{i=1}^N 1_{\{Z_i \geq t\}}$ and $Pois(x)$ denotes the Poisson distribution with parameter x , Le Cam's inequality, see [4], gives

$$\|Q_N - Pois(Ne^{-t})\|_{tv} \leq e^{-t}, \quad t \geq 0, \quad (17)$$

in particular,

$$\left|P\left(\sum_{i=1}^N 1_{\{Z_i \geq t\}} > 0\right) - (1 - e^{-Ne^{-t}})\right| \leq e^{-t},$$

hence,

$$P(S_N > t) \leq P\left(\sum_{i=1}^N 1_{\{Z_i \geq t\}} > 0\right) \leq (N + 1)e^{-t}.$$

Of course, the last inequality can be proved with elementary means, the term in the middle is simply the probability that the maximum of N exponential random variables exceeds t . The inequality (17) will be used in the sequel. Our proposition is proved. \square

Remark 9. The second largest eigenvalue converges to -1 when $N \rightarrow +\infty$, see [10], so as in the super-critical case our bound (16) is correct as a function of t . Notice however that this bound gives $\log N$ as an upper bound for the time to reach equilibrium. We shall see in the next proposition that $\frac{\log N}{2}$ is the correct order of magnitude.

Proposition 11. *In the case $\lambda < 1$, then for any sequence $(\phi(N))$ such that $\lim_{N \rightarrow +\infty} \phi(N) = +\infty$, then*

$$\lim_{N \rightarrow +\infty} d_N \left(\frac{\log N}{2} + a\phi(N) \right) = \begin{cases} 1 & \text{if } a < 0 \\ 0 & \text{if } a > 0. \end{cases}$$

Proof. From the case $\lambda = 1$, we know that the first part of our proposition is true ; the process which starts from 0 needs at least the time $\frac{\log N}{2}$ to reach the good region around λN for equilibrium, otherwise the process is at distance 1. All we have to prove is that this time is indeed sufficient to reach equilibrium. The difference with the case $\lambda = 1$ is that the process can start above λN , at N for example. The hitting time of λN starting from N is less easy to handle than the hitting time from 0 to λN . We shall use the following procedure : construct a coupling of two processes $(X^1(t)), (X^2(t))$, the first one starting from N and the second from 0. If T_m is the meeting time of these two processes, then clearly

$$d_N(t) \leq P(T_m > t). \quad (18)$$

If we assume that the two processes have no upper boundary, then it certainly increases their meeting time ; the inequality (18) is thus still true with this meeting time.

The coupling We fix $\varepsilon > 0$ and $A > 0$. We consider the same coupling as before between X^1 and X^2 , the arrivals and the services of the customers are the same, until the time

$$S_A = \inf \left\{ s > 0 / X_N^1(s) - X_N^2(s) \leq A\sqrt{N} \right\}.$$

After this stopping time S_A , we suppose that the processes behave independently until they meet.

According to the proof of Proposition 10,

$$P(S_A > t) \leq P \left(\sum_{i=1}^N 1_{\{Z_i \geq t\}} > A\sqrt{N} \right),$$

where the (Z_i) are i.i.d. exponential variables with parameter 1. And inequality (17) gives

$$\left| P \left(\sum_{i=1}^N 1_{\{Z_i \geq t\}} > A\sqrt{N} \right) - P \left(\mathcal{N}_1[0, Ne^{-t}] \geq A\sqrt{N} \right) \right| \leq e^{-t}. \quad (19)$$

After time S_A , the two processes behave as processes Y_N^1 and Y_N^2 , such that $Y_N^1(0) - Y_N^2(0) = \lfloor A\sqrt{N} \rfloor$, and

$$dY_N^i(t) = \mathcal{N}_{\lambda N}^i(dt) - \sum_{k=1}^{Y_N^i(t-)} \mathcal{N}_1^{i,k}(dt), \quad t \geq 0, \quad i = 1, 2.$$

It is easily seen that $\left(\frac{Y_N^1(t) - Y_N^2(t)}{\sqrt{N}} \right)$ converges in distribution to $(U(t))$ the Ornstein-Ühlenbeck with parameter 2λ starting from A . For $\delta > 0$, because of the ergodicity of Ornstein-Ühlenbeck processes, the hitting time $\tau_\delta = \inf \{ s > 0 / U(s) \leq -\delta \}$ is finite. There exists

some $T > 0$ such that $P(\tau_\delta > T) \leq \varepsilon$, and some N_0 , such that for $N \geq N_0$,

$$\begin{aligned} P_A \left(\inf_{s \leq T} \frac{Y_N^1(s) - Y_N^2(s)}{\sqrt{N}} > -\delta \right) &\leq P_A \left(\inf_{s \leq T} U(s) > -\delta \right) + \varepsilon \\ &= P(\tau_\delta > T) + \varepsilon \leq 2\varepsilon. \end{aligned} \quad (20)$$

The meeting time T_m of these two processes is certainly smaller than S_A plus the time to reach $-\delta$ starting from $[A]$ for the process $\left(\frac{Y_N^1(t) - Y_N^2(t)}{\sqrt{N}}\right)$. For $a > 0$, and $N \geq N_0$,

$$\begin{aligned} P \left(T_m > \frac{\log N}{2} + a\phi(N) \right) &\leq \\ &P \left(S_A > \frac{\log N}{2} + a\phi(N) - T \right) + P_A \left(\inf_{s \leq T} \frac{Y_N^1(s) - Y_N^2(s)}{\sqrt{N}} > -\delta \right), \end{aligned}$$

using (19) and (20), we get

$$\begin{aligned} P \left(T_m > \frac{\log N}{2} + a\phi(N) \right) &\leq \\ &\frac{1}{\sqrt{N}} e^{-a\phi(N)+T} + P \left(\mathcal{N}_1[0, \sqrt{N}e^{-a\phi(N)+T}] \geq A\sqrt{N} \right) + 2\varepsilon, \end{aligned}$$

the law of large numbers for Poisson processes, i.e. for $\eta > 0$, $\lim_{x \rightarrow +\infty} \frac{\mathcal{N}_1[0, x\eta]}{x} = \eta$ a.s. shows that the last term converges to 0 as $N \rightarrow +\infty$. Our proposition is proved. \square

Remark 10. The meeting time T_m of the two processes of the above proof satisfies probably a proposition similar to Proposition 8. But due to the complex representation of the hitting time of λN starting from N , it is very unlikely that a simple argument will give such a result, i.e. without jumping into the analytic intricacies of the asymptotic analysis of equation (9).

On the cutoff property In the critical case the second largest eigenvalue $\lambda_{2,N}$, is 2 and asymptotically 1 in the case $\lambda < 1$, see [10], but the order of magnitude of the time for the cutoff does not change. The main explanation which is given for the cutoff property is that an accumulation of eigenvalues occurs below $\lambda_{2,N}$ as $N \rightarrow +\infty$, and this phenomenon balances the influence of the variable $\lambda_{2,N}$. Here, this is certainly not the reason since it is well known, see [5], that there is a gap of size at least 1 between all the eigenvalues.

Acknowledgments

This work has benefited from discussions with Philippe Flajolet.

References

- [1] D. ALDOUS AND P. DIACONIS, *Strong uniform times and finite random walks*, Advances in Applied Mathematics, 8 (1987), pp. 69–97.
- [2] D. J. ALDOUS, *Random walks on finite groups and rapidly mixing Markov chains*, in Séminaire de Probabilités de Strasbourg XVII, L. notes in Math. 986, ed., Berlin, 1983, Springer Verlag, pp. 243–297.
- [3] V. ANANTHARAM, *The settling time of a closed Jackson network*, 1988. Unpublished.
- [4] A. BARBOUR, L. HOLST, AND S. JANSON, *Poisson approximation*, Oxford science publications, 1992.
- [5] V. E. BENEŠ, *The covariance function of a simple truck group, with applications to traffic measurement*, Bell System Technical Journal, 40 (1961), pp. 117–148.
- [6] T. S. CHIHARA, *An introduction to orthogonal polynomials*, Gordon and Breach Science Publishers, New York, 1978. Mathematics and its Applications, Vol. 13.
- [7] P. DIACONIS, *The cutoff phenomenon in finite Markov chains*, Proceedings of the National Academy of Sciences of the United States of America, 93 (1996), pp. 1659–1664.
- [8] S. ETHIER AND T. KURTZ, *Markov Processes, characterization and convergence*, Probability and mathematical statistics, John Wiley & Sons Ltd, 1986.
- [9] P. FLAJOLET, *Combinatorial aspects of continued fractions*, Discrete Mathematics, 32 (1980), pp. 125–161.
- [10] C. FRICKER, P. ROBERT, AND D. TIBI, *On the spectral gap of the M/M/N/N queue*, 1997. In preparation.
- [11] I. GOULDEN AND D. JACKSON, *Combinatorial enumeration*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Inc., New York, 1983.
- [12] F. GUILLEMIN AND A. SIMONIAN, *Transient characteristics of an M/M/∞ system*, Advances in Applied Probability, 27 (1995), pp. 862–888.
- [13] D. I. HADJIEV, *The first passage problem for generalized Ornstein-Ühlenbeck processes with nonpositive jumps*, in Séminaire de probabilités, XIX, 1983/84, vol. 1123 of Lecture Notes in Math., Springer, Berlin, 1985, pp. 80–90.
- [14] P. HUNT, *Loss networks under diverse routing: the symmetric star network*, Advances in Applied Probability, 25 (1995), pp. 255–272.
- [15] P. HUNT AND T. KURTZ, *Large loss networks*, Stochastic Processes and their Applications, 53 (1994), pp. 363–378.

-
- [16] S. KARLIN AND J. MCGREGOR, *The classification of birth and death processes*, Transactions of the AMS, 86 (1957), pp. 366–400.
- [17] F. KELLY, *Loss networks*, Annals of Applied Probability, 1 (1991), pp. 319–378.
- [18] C. KNESSL, *On the transient behavior of the $M/M/m/m$ loss model*, Communications in Statistics. Stochastic Models, 6 (1990), pp. 749–776.
- [19] D. MITRA AND A. WEISS, *The transient behavior in Erlang's model for large trunk groups and various traffic conditions*, tech. rep., AT&T Bell Laboratories, 1988. in the Proceedings of the Twelfth International Teletraffic Congress, Torino - Italy.
- [20] G. F. NEWELL, *The $M/M/\infty$ service system with ranked servers in heavy traffic*, vol. 231 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1984. With a preface by Franz Ferschl.
- [21] L. ROGERS AND D. WILLIAMS, *Diffusions, Markov processes and Martingales*, vol. Volume 2: Itô calculus, John Wiley & Sons Ltd, 1987.
- [22] A. SHWARTZ AND A. WEISS, *Large deviations for performance analysis*, Stochastic Modeling Series, Chapman & Hall, London, 1995.
- [23] L. TAKACS, *Introduction to the theory of queues*, Oxford University Press, New York, 1962.
- [24] W. WHITT, *On the heavy-traffic limit theorem for $GI/G/\infty$ queues.*, Advances in Applied Probability, 14 (1982), pp. 171–190.
- [25] E. T. WHITTAKER AND G. N. WATSON, *A course of modern analysis*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1996. Reprint of the fourth (1927) edition.



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399