Stationary Distribution Analysis of a Queueing Model with Local Choice

Plinio S. Dester[1]
INRIA Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France
plinio.santini-dester@polytechnique.edu

Christine Fricker
Département d'informatique de l'ENS, École normale supérieure, CNRS, PSL Research University, 75005 Paris, France,
INRIA Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France
Christine.Fricker@inria.fr

Hanene Mohamed[2]
Université Paris Nanterre, Modal'X, UPL, 92000 Nanterre, France
hanene.mrad@u-parisnanterre.fr

Abstract

The paper deals with load balancing between one-server queues on a circle by a local choice policy. Each one-server queue has a Poissonian arrival of customers. When a customer arrives at a queue, he joins the least loaded queue between this queue and the next one, ties solved at random. Service times have exponential distribution. The system is stable if the arrival-to-service rate ratio called load is less than one. When the load tends to zero, we derive the first terms of the expansion in this parameter for the stationary probabilities that a queue has 0 to 3 customers. We investigate the error, comparing these expansion results to numerical values obtained by simulations. Then we provide the asymptotics, as the load tends to zero, for the stationary probabilities of the queue length, for a fixed number of queues. It quantifies the difference between policies with this local choice, no choice and the choice between two queues chosen at random.

2012 ACM Subject Classification Mathematics of computing → Stochastic processes

Keywords and phrases queueing model, local choice, stationary analysis, balance equations, power series expansion

Digital Object Identifier 10.4230/LIPIcs.AofA.2018.22

Acknowledgements We would like to thank Nathanael Enriquez for helpful discussions and the introduction to reference [4].

1 Introduction

1.1 A load balancing policy

The paper deals with the impact of choice between two neighbors in a large set of queues. Load balancing is present in a wide literature and includes various policies as choice, offloading, redundancy or work stealing ([6], [17], [8] and others) for example. The two-choice policy is a well-known distributed way to improve load balancing. See [14] and [12] for one-server

[1] PGMO
[2] ANR11-LBX-0023

29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018).
Editors: James Allen Fill and Mark Daniel Ward; Article No. 22; pp. 22:1–22:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

queues. For this policy, the arriving customers choose two queues at random and join the shortest one, ties being solved at random. The paper focuses on the case where only local choice can be processed. This case occurs in many applications with geographical constraints, like vehicle-sharing systems or cloud computing.

1.2 The model

The model we present is called *local choice model*. It consists in a set of N one-server queues with infinite capacity where customers arrive at each queue according to independent Poisson processes with rate λ, which means that inter-arrival times are independent with exponential distribution with parameter λ. When a customer arrives at queue i, $1 \leq i \leq N$, he chooses between queues i and $i+1$ the least loaded one and joins it. By convention, queue $N+1$ is queue 1. If queues i and $i+1$ have the same number of customers, he joins one of these queues with probability $1/2$. The service times are iid with exponential distribution with parameter μ. When the customer is served, he leaves the system. All inter-arrival and service times are independent. The load ρ is by definition λ/μ.

1.3 The problem

The main issue addressed in the paper concerns the marginal distribution of the number of customers in one queue at equilibrium for the *local choice model*. We investigate the asymptotics of the stationary probabilities for one queue as the load tends to zero. The number N of queues is fixed throughout the paper. We compare them with the same quantities for the *random choice model*, where an arriving customer chooses two queues at random and joins the least loaded one and *the no choice model*, where a customer who arrives at queue i is served at this queue.

The *no choice model* is simply N independent $M/M/1$ queues. The *random choice model* is classical, see [14] and [12]. For $\rho < 1$, the limiting stationary tail probability, i.e. the limit as N gets large of the stationary probability that a queue has more than k customers, is doubly exponentially decreasing, more precisely is ρ^{2^k-1}, $k \geq 0$. This doubly exponential decrease is known in the literature as *the power of choice*. Indeed it is much smaller than the tail probability ρ^k, $k \geq 0$ in the *no choice model* as in the $M/M/1$ queue, the queue length stationary distribution is geometric with parameter ρ. What is this tail probability for the *local choice model*?

1.4 The results

They concern the *local choice model* previously described. In the paper, N is fixed and $\rho < 1$ to ensure the ergodicity of the queue length process. We consider the stationary probabilities as analytical functions of parameter ρ. Based on some crucial arguments (see Lemmas 2 and 3), an induction procedure provides all the terms of the power series expansion. We apply this procedure to find the first terms explicitly. Then, in the study of the marginal distribution for one queue, it gives the first terms (at order 6) of the expansion in ρ of the stationary probability that a queue has m customers, for small m ($m \leq 3$). This expansion is an approximation for the stationary probability at light traffic, which is compared to simulations.

The main result of the paper gives the asymptotics as ρ tends to 0 of the stationary probability that a queue has m customers, for any m. It is claimed in Proposition 8 that these asymptotics are $2\rho^{2m-1}$ for $N=2$ and $12(\rho/2)^{2m-1}$ for $N \geq 3$. It gives the rate of decay in parameter ρ of the stationary queue length at light traffic, which is ρ^2 for $N = 2$ and $(\rho/2)^2$ for $N \geq 3$. In other words, compared to the N independent $M/M/1$ queues, the *local choice model* does not lead an improvement as large as in the *random choice model* which is doubly exponential.

1.5 Related work

The choice between two queues at random among N one-server queues is well understood for N large via mean-field method for the late 90's with [14] and [12] and knows a great interest in literature. Nevertheless, local choice is a quite challenging open problem in queueing theory. As far as we know, very few papers investigate the problem. For this model, where the underlying graph is linear, more precisely a circle, and more generally for a graph $G = (V, E)$, [7] gives an approximation of the steady-state queue length distribution which seems numerically accurate compared to simulations. This approximation, called pair-approximation, is obtained from the empirical measure on pairs of neighbors. It is a mean field limit as the graph gets large. But this limit, solution of an ODE, is hard to study analytically. In [7], the expression of the ODE is explicitly given, but its equilibrium point is investigated by numerical simulations.

The series expansion of the stationary probabilities in parameter ρ is the key tool in [2] for the study of the *JSQ model*. It is the classical model of N queues, where arriving customers join the shortest queue among all the queues. The paper gives asymptotics in light traffic for the mean and the variance of the total number of customers at equilibrium. Nevertheless the method to obtain them is quite different.

1.6 Related models

Some papers deal with such models, but without departure. They are called urn models in computer science literature, and deposition models or crystal growth models in statistical physics. The problems addressed in both cases are quite different.

THE URN MODEL. Urns are put at vertices of a finite graph $G = (V, E)$ with $|V| = N$. Arrival of balls are associated to edges. For each ball, an edge is chosen at random and the ball is put in the least loaded of the two end-points of the edge. The problem of the maximum number of balls per urn for N balls in N urns is investigated. The conclusion is that the *power of choice* does not hold for d-regular graphs, d constant, as this maximum is not in $\log \log N$ (see [10], also [3] and references therein). But the main difference with our study is that we deal with the stationary regime. The poor load balancing result in the urn problem might come from the fact that with N balls in N urns, the equilibrium is not reached.

THE CRYSTAL GROWTH MODEL. In this model, consider N sites $1, \dots, N$. There is also no departure. Particles arrive at each site, say i, at rate λ. If the two (respectively just one or none) neighboring sites $i-1$ and $i+1$ have more particles than site i, the arrival rate at the site i is β_2, β_1 and β_0, respectively. [9, 1, 5] give ergodic conditions for the shape process, which is Markov. Our arrival process is a variant of this model in the special case where $\beta_0 = 0$ and $\beta_2 = 2\beta_1$ (see Section 2 for details). Note that if we extend the *local choice model*, to the case where the customers, arriving at queue i, choose between the two neighboring queues i and $i+1$ with some probability α and do not choose otherwise, it will still fit in this framework as a variant, but with $\beta_0 \neq 0$.

1.7 Outline

The paper is organized as follows. Section 2 gives the model description and the notations. Section 3 leads to the induction procedure to obtain the power series expansion of the stationary state probabilities of the model. Section 4 deals with the results for the marginal distribution for one queue.

2 Model description and notations

Consider a system of N queues with infinite capacities, each of them served by one server at rate μ. In all the following, queue $N+1$ means queue 1. The arrival rate at each queue is λ but the arriving customer at queue i joins the least loaded queue between queues i and $i+1$, ties being solved at random. All inter-arrival and service times are independent with exponential distribution. The i.i.d. Bernoulli variables with parameter 1/2 introduced to solve the ties are independent of the previous random variables. By definition, $\rho = \lambda/\mu$.

2.1 The state process

For $1 \leq i \leq N$, let $X_i(t)$ be the number of customers at queue i at time t and $X(t) = (X_i(t))_{1\leq i\leq N}$. The queue length process $(X(t))_{t\geq 0}$ is a Markov process on state space $\mathbb{N}^N$ with Q-matrix Q, given for $n = (n_1, \ldots, n_N)$ here and in all the following, by its non-negative components, for $1 \leq i \leq N$,

$$Q(n, n + e_i) = \lambda c_i(n)$$
$$Q(n, n - e_i) = \mu \mathbb{1}_{n_i > 0}$$

where $c : \mathbb{N} \times \mathbb{N}^N \longrightarrow \mathbb{R}_+$, called the *contribution function*, quantifies the amount of arrivals at the different queues and $(e_i)_{1\leq i\leq N}$ is the canonical basis of $\mathbb{R}^N$.

For our *local choice model*, this contribution function is called *local choice function* and is denoted by c^{lc}. Function c^{lc} at queue i, depends only on the state of this queue and the two neighboring queues $i-1$ and $i+1$ and is defined by

$$c_i^{lc}(n) = d(n_i, n_{i+1}) + d(n_i, n_{i-1}) \text{ where } d(k,l) = \frac{1}{2}\mathbb{1}_{\{k=l\}} + \mathbb{1}_{\{k<l\}} \tag{1}$$

with, by convention, $n_0 = n_N$ and $n_{N+1} = n_1$. Dispatching function d is the basis of our *local choice model* since it implements the load balancing policy: join the least loaded among two neighboring queues.

▶ Remark. The local choice function c^{lc} can also be defined by

$$c_i^{lc}(n) = \omega(\Delta_{i-1}n, -\Delta_i n),$$

in terms of the shape function Δ defined by $n \mapsto \Delta n = (\Delta_1 n, \ldots, \Delta_N n)$ where $\Delta_j n = n_j - n_{j+1}$, $1 \leq j \leq N$ and the so-called deposition function ω given by

$$\omega(a,b) = \frac{1}{2}\left(\mathbb{1}_{\{a=0\}} + \mathbb{1}_{\{b=0\}}\right) + \mathbb{1}_{\{a>0\}} + \mathbb{1}_{\{b>0\}},\ a, b \in \mathbb{Z}. \tag{2}$$

Note that the Gates-Wescott process studied in [5] is the shape process $(\Delta X(t))$ for the model without departure associated to the following deposition function

$$\omega(a,b) = \beta_{\mathbb{1}_{\{a>0\}}+\mathbb{1}_{\{b>0\}}},\ a, b \in \mathbb{Z}, \tag{3}$$

with β_0, β_1 and $\beta_2 > 0$.

3 An algorithm to compute the stationary distribution

In this section we study, for N fixed, the queue length process $(X(t))$ at stationarity. We prove first that $(X(t))$ is ergodic for $\rho < 1$ if $c = c^{lc}$. See Proposition 1. For a general c, if $(X(t))$ is ergodic, it has a unique invariant measure $y = (y_n,\ n \in \mathbb{N}^N)$ on $\mathbb{N}^N$, solution of the global balance equations

$$\sum_{n' \in \mathbb{N}^N} y(n')Q(n', n) = 0, \quad n \in \mathbb{N}^N. \tag{4}$$

Our aim is not to solve these equations but rather look for an analytical solution for y of the form

$$y_n(\rho) = \sum_{k \geq 0} \alpha_k(n)\rho^k,\ n \in \mathbb{N}^N.$$

Assuming the existence of $\varepsilon > 0$ such that the solution of the global balance equations $y_n(\rho)$ has a serie expansion for $0 < \rho < \varepsilon$, we prove that each α_k, $k \geq 0$, has a finite support. See Lemma 4. Then we explain the algorithm to obtain by induction on $k \geq 0$ the explicit expressions of α_k and compute explicitely the first terms.

3.1 Ergodicity for c^{lc}

For local choice, contribution function c^{lc} is given by equation (1). The following result gives us the necessary and sufficient condition for ergodicity of the Markov state process $(X(t))$ in this case.

▶ **Proposition 1** (Ergodicity). *For $c = c^{lc}$, the Markov process $(X(t))_{t\geq 0}$ is ergodic if $\rho < 1$ and transient if $\rho > 1$.*

The proof based on Foster's criterion is postponed in Appendix.

3.2 Power series expansion in ρ of the stationary probabilities

For ρ such that $(X(t))$ is ergodic, let $y(\rho) = (y_n(\rho), n \in \mathbb{N}^N)$ be its invariant measure, the unique solution of the global balance equations

$$\left(\sum_{i=1}^N \mathbb{1}_{\{n_i>0\}} + \rho \sum_{i=1}^N c_i(n)\right) y_n(\rho) = \sum_{i=1}^N y_{n+e_i}(\rho) + \rho \sum_{i=1}^N \mathbb{1}_{\{n_i>0\}} c_i(n-e_i) y_{n-e_i}(\rho), \quad n \in \mathbb{N}^N \tag{5}$$

obtained by plugging the expression of Q in equation (4).

We look for an invariant measure $(y_n, n \in \mathbb{N}^N)$ satisfying the following condition.

$(\mathbf{H_0})$ There exists $\varepsilon > 0$, such that, for $\rho \in [0, \varepsilon[$ and $n \in \mathbb{N}^N$,
$y_n(\rho)$ can be written as a series expansion of the form

$$y_n(\rho) = \sum_{k \geq 0} \alpha_k(n)\rho^k. \tag{6}$$

▶ Remark. In all the following, (H_0) will be assumed. This question of analyticity of stationary probabilities of a family of Markov chains depending on one parameter is the major issue addressed by [16, Chapter IV] (see also [15, Chapter 7]). The main tool for proving such analyticity is the Lyapunov function in Foster's criterion for ergodicity. We get a quadratic Lyapunov function to prove the ergodicity in Proposition 1 (see the proof in Appendix A). But the dynamics of our model do not allow to apply the results of [16, Chapter IV] or [15, Chapter 7], due to the contribution function part. This question is the object of future work.

Moreover the following technical assumption

$$\sum_{k\geq 0}\sum_{n\in\mathbb{N}^N}\alpha_k(n)\rho^k=\sum_{n\in\mathbb{N}^N}\sum_{k\geq 0}\alpha_k(n)\rho^k,\quad \rho<\varepsilon \tag{7}$$

is used.

▶ Remark. According to Proposition 1, for $c=c^{lc}$, as analyticity requires the existence of the stationary measure, thus implicitly the ergodicity of process $(X(t))$, it holds that $\varepsilon\leq 1$. Note that assumption (H_0) could have been written with 1 instead of ε. We introduce ε in (H_0) of this form because some results in the following apply for more general c than c^{lc}, where the ergodicity condition can be written $\rho<\varepsilon$.

Under assumption (H_0), for each $n\in\mathbb{N}^N$, $\rho\mapsto y_n(\rho)$ is C^∞ on $[0,\varepsilon[$ and $\alpha_k(n)=y_n^{(k)}(0)/k!$. Taking the k-th derivative in the global balance equations (5) with respect to ρ and evaluating it at $\rho=0$, it holds that, for any $n\in\mathbb{N}^N$ and $k\in\mathbb{N}^*$,

$$\left(\sum_{i=1}^N \mathbb{1}_{\{n_i>0\}}\right)\alpha_k(n)=\\ \sum_{i=1}^N\alpha_k(n+e_i)+\sum_{i=1}^N \mathbb{1}_{\{n_i>0\}}c_i(n-e_i)\alpha_{k-1}(n-e_i)-\left(\sum_{i=1}^N c_i(n)\right)\alpha_{k-1}(n). \tag{8}$$

3.3 Some crucial lemmas

Equation (8) allows us to prove that, for k fixed, α_k has a finite support. It is the purpose of Lemma 4. For that, we need to prove the two following technical lemmas. Lemma 2, formulated with α for sake of simplicity, will be applied for each $\alpha_k, k\geq 1$. Before that, let us introduce the following set

$$\mathcal{A}_k\overset{def}{=}\{n\in\mathbb{N}^N, n_1+n_2+\ldots+n_N=k\},\ k\in\mathbb{N}. \tag{9}$$

▶ **Lemma 2.** *Let $\alpha:\mathbb{N}^N\longrightarrow\mathbb{R}$ and $k_0\in\mathbb{N}^*$ be such that, for $n=(n_1,\ldots,n_N)$ with $|n|=n_1+\cdots+n_N>k_0$,*

(i) $\alpha(n)\geq 0$,

(ii) *the following recurrence equation holds,*

$$\left(\sum_{i=1}^N \mathbb{1}_{\{n_i>0\}}\right)\alpha(n)=\sum_{i=1}^N\alpha(n+e_i), \tag{10}$$

(iii) $\sum_{n,|n|>k_0}\alpha(n)<\infty$

then, for all n such that $|n|>k_0$, $\alpha(n)=0$.

Proof. Let $k_0 \in \mathbb{N}^*$ be fixed. First, we claim that, for any $k > k_0$,

$$\sum_{n \in \mathcal{A}_k} \sum_{i=1}^{N} \alpha(n + e_i) = \sum_{n \in \mathcal{A}_{k+1}} \left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \right) \alpha(n). \tag{11}$$

Indeed, for $n \in \mathcal{A}_{k+1}$, for i such that $n_i \neq 0$, $\alpha(n)$ can be written as $\alpha(\hat{n} + e_i)$, for a unique $\hat{n} \in \mathcal{A}_k$. The number of elements in $\mathcal{A}_k$ that can generate n when we add them to e_i is exactly equal to the number of non-zero coordinates n_i of n, $1 \leq i \leq N$. Therefore, equation (11) holds.

Then we replace $\sum_{i=1}^{N} \alpha(n + e_i)$ in the left-hand side of (11) by the left-hand side of equation (10). It yields, for any $k > k_0$,

$$\sum_{n \in \mathcal{A}_k} \left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \right) \alpha(n) = \sum_{n \in \mathcal{A}_{k+1}} \left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \right) \alpha(n).$$

Thus, for any $k > k_0$,

$$\sum_{n \in \mathcal{A}_k} \left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \right) \alpha(n) = C \tag{12}$$

where C is non-negative due to (i) and independent of k. As $\sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \leq N$,

$$\sum_{n \in \mathcal{A}_k} N\alpha(n) \geq C.$$

If $C > 0$, $\sum_{k > k_0} \sum_{n \in \mathcal{A}_k} \alpha(n)$ diverges, since $\sum_{n \in \mathcal{A}_k} \alpha(n) \geq \dfrac{C}{N} > 0$. But, this contradicts the fact that $\sum_{n, |n| > k_0} \alpha(n) < \infty$. Thus $C = 0$. Using the fact that $\alpha(n) \geq 0$ in equation (12), for all n such that $|n| > k_0$, $\alpha(n) = 0$. ◀

The following lemma is a key argument for both computing the $\alpha_k(n)$ (see Section 3.4) and in the proof of Lemma 4.

▶ **Lemma 3.** *The following property holds:*

$$\sum_{n \in \mathbb{N}^N} \alpha_k(n) = 0, \quad k > 0. \tag{13}$$

Proof. Permuting the sums, by equation (7), it holds that, for $\rho < \varepsilon$,

$$\sum_{k \geq 0} \left(\sum_{n \in \mathbb{N}^N} \alpha_k(n) \right) \rho^k = \sum_{n \in \mathbb{N}^N} \left(\sum_{k \geq 0} \alpha_k(n) \rho^k \right) = \sum_{n \in \mathbb{N}^N} y_n(\rho) = 1.$$

as $y(\rho) = (y_n(\rho), n \in \mathbb{N}^N)$ is a probability measure. The left-hand side of this equation is a power series whose all the terms except the first one are null. It ends the proof. ◀

We can now prove the following result.

▶ **Lemma 4.** *Let $k \in \mathbb{N}$. For all n, $|n| > k$, $\alpha_k(n) = 0$.*

Proof. We prove this assertion by induction on k. Take $k = 0$. From equation (6), $y_n(0) = \alpha_0(n)$. As the invariant measure for $\rho = 0$ (no arrival) is $y_n(0) = \delta_{0_N}$, the Dirac mass at $(0, \dots, 0) \in \mathbb{N}^N$ denoted by 0_N, the assertion is true for $k = 0$. Let $k \in \mathbb{N}$ be fixed. If we suppose that the assertion holds for $k' \leq k$, then Lemma 2, applied to $\alpha = \alpha_{k+1}$ and $k_0 = k + 1$, guarantees that the assertion is true for $k' = k + 1$. Indeed let us check assertions (i), (ii) and (iii) for all n with $|n| > k + 1$. Let such a n be fixed. In equation (8), $\alpha_{(k+1)-1}(n) = \alpha_{(k+1)-1}(n - e_i) = 0$ since $|n| > k$ and $|n - e_i| > k$ and induction assumption. Therefore equation (8) is rewritten as equation (10), giving (ii). Moreover, by induction assumption, in equation (6), $\alpha_{k+1}(n)$ represents the first possible non-zero coefficient for $y_n(\rho)$. This coefficient $\alpha_{k+1}(n) \geq 0$, because otherwise, as

$$y_n(\rho) \sim_{\rho \to 0} \alpha_{k+1}(n)\rho^{k+1},$$

it would exist ρ such that $y_n(\rho) < 0$, which is false as $y(\rho)$ is a probability measure. It gives (i). Eventually, by equation (13), $\sum_{n=(n_1,\dots,n_N)} \alpha_{k+1}(n) = 0$ and, as $\sum_{|n| \leq k+1} \alpha_{k+1}(n)$ is finite, then $\sum_{|n|>k+1} \alpha_{k+1}(n)$ is finite too, which is (iii). ◀

3.4 Induction procedure

The algorithm to obtain all the coefficients $\alpha_k(n)$ is an induction procedure on $k \geq 0$. We use that $\alpha_0 = \delta_{0_N}$ and key equation (8). For $k \geq 1$, assume that we know the coefficients $\alpha_{k-1}(n)$, for all $n \in \mathbb{N}^N$ and find the coefficients $\alpha_k(n)$, $n \in \mathbb{N}^N$. First, by Lemma 4, $\alpha_k(n) = 0$ for $|n| > k$. Second we derive each coefficient $\alpha_k(n)$ for $n \in \mathcal{A}_k$, defined by equation (9), as the left-hand side of equation (8). Indeed, in the right-hand side of the same equation, the first term is null due to Lemma 4. The other terms are known as coefficients for $k - 1$. By the same procedure, we compute the $\alpha_k(n)$ for $n \in \mathcal{A}_{k-1}$: Since $n + e_i \in \mathcal{A}_k$, we still know also the first term of the right-hand side of equation (8). Then we determine the coefficients for $n \in \mathcal{A}_{k-2}$, $n \in \mathcal{A}_{k-3}$ and so on, until $n \in \mathcal{A}_1$. It remains to compute the last coefficient $\alpha_k(0_N)$. It is given by the additional equation (13) in Lemma 3.

▶ Remark. For Lemma 4 and the previous induction procedure, we do not use the specific expression (1) of contribution function c. We just choose ρ in the domain of analyticity of the $y_n, n \in \mathbb{N}^N$. What follows remains valid for a general contribution function c satisfying the following additional assumptions

($\mathbf{H_1}$) For $n \in \mathbb{N}^N,\ c_1(n) + \cdots + c_N(n) = N$.

($\mathbf{H_2}$) c is invariant by cyclic permutation or reflection (reverse order).

More precisely, the second assumption means that, for such a permutation σ on $\{1, 2, ..., N\}$, for $n \in \mathbb{N}^N$ and $1 \leq i \leq N$, $c_{\sigma(i)}(\sigma(n)) = c_i(n)$. These assumptions are obviously true for the local choice function $c = c^{lc}$ defined by equation (1).

3.5 Deriving the first terms

Let us derive the coefficients until order 3 under (H_0), (H_1) and (H_2). It is given by the following proposition.

▶ **Proposition 5.** *For $k = 0$,*

$$\alpha_0(n) = \mathbb{1}_{\{n=0_N\}}. \tag{14}$$

For $k=1$,

$$\begin{cases} \alpha_1(0_N) & = -N, \\ \alpha_1(e_i) & = 1,\ 1 \leq i \leq N \\ \alpha_1(n) & = 0 \ \textit{otherwise.} \end{cases} \tag{15}$$

For $k=2$, for $i,j \in \{1,2,...,N\}$,

$$\begin{cases} \alpha_2(0_N) & = \dfrac{1}{2}(N^2 - Nc_1(e_1)), \\ \alpha_2(e_i) & = -N, \\ \alpha_2(e_i+e_j) & = c_i(e_j), \\ \alpha_2(n) & = 0 \ \textit{otherwise.} \end{cases} \tag{16}$$

For $k=3$, for all $i,j,l \in \{1,2,...,N\}$, $i \neq j$, $j \neq l$ and $l \neq i$,

$$\begin{cases} \alpha_3(0_N) & = -\sum_{n\neq 0_N} \alpha_3(n) \\ \alpha_3(e_i) & = \frac{1}{2}(N^2 - Nc_1(e_1)) \\ \alpha_3(e_i+e_j) & = \frac{1}{2}\left(\sum_{v=1}^N \alpha_3(e_i+e_j+e_v) - 3Nc_i(e_j)\right), \\ \alpha_3(2e_i) & = \frac{1}{2}\left(\sum_{v=1}^N c_i(e_v)c_i(e_i+e_v) - 3Nc_i(e_i)\right), \\ \alpha_3(e_i+e_j+e_l) & = \frac{1}{3}(c_i(e_j)c_l(e_i+e_j) + c_j(e_l)c_i(e_j+e_l) \\ & \quad +c_l(e_i)c_j(e_l+e_i)), \\ \alpha_3(2e_i+e_j) & = \frac{1}{2}\left(c_i(e_j)c_i(e_i+e_j) + c_i(e_i)c_j(2e_i)\right), \\ \alpha_3(3e_i) & = c_1(e_1)c_1(2e_1) \\ \alpha_3(n) & = 0 \ \textit{otherwise.} \end{cases} \tag{17}$$

Proof. For $\rho = 0$, the solution is $y_n(0) = \mathbb{1}_{\{n=0_N\}}$, which gives the coefficients for $k=0$. For $k=1,2$ and 3, we use the method previously described and assumptions (H_1) and (H_2). ◀

It is interesting to notice that, for $k=0$ and 1, the coefficients $\alpha_k(n)$ do not depend on the choice function c. It means that, for ρ sufficiently small, the choice policy does not influence the system. For $k \geq 4$, the expressions become huge, which is not a problem if performed numerically.

4 Marginal distribution for one queue

Our objective is to study the expansion of the stationary probability that queue i, $1 \leq i \leq N$, has $m \in \mathbb{N}$ customers assuming an analytical solution for y. We give the series expansion at order 6, for small m ($m \leq 3$), for the local choice contribution function. Moreover we investigate the accuracy of this expansion, compared to numerical values obtained by simulations. Then the main result of the section provides the first term of the expansion for every $m \geq 1$.

As our system is invariant by cyclic permutation, by assumption (H_2), for $m \in \mathbb{N}$ and $i \in \{1,\dots,N\}$, the probability that queue i has m customers does not depend on i. This probability, denoted by $\pi_m(\rho)$, is given by

$$\pi_m(\rho) = \sum_{n=(n_1,\dots,n_N),n_1=m} y_n(\rho). \tag{18}$$

Under assumption (H_0) that $y_n(\rho)$ is analytical on $[0,\varepsilon[$, $\pi_m(\rho)$ has a series expansion, that can be written as

$$\pi_m(\rho) = \sum_{k\geq 0} \phi_k(m)\rho^k, \quad 0 \leq \rho < \varepsilon \tag{19}$$

where $\phi_k(m) = \pi_m^{(k)}(0)/k!$ is given from equation (6) by

$$\phi_k(m) = \mathbb{1}_{\{m\leq k\}} \sum_{n_2+n_3+\ldots+n_N \leq k-m} \alpha_k(m, n_2, n_3, \ldots, n_N). \tag{20}$$

4.1 Expansion for a general contribution function

Note that, in equation (20), $\phi_m(m)$ is the first possibly non-null coefficient of the expansion of $\pi_m(\rho)$. This follows directly from Lemma 4. Moreover this coefficient is derived in the following proposition, which also gives the third order expansion of the π_m's.

▶ **Proposition 6.** *If the choice function c satisfies (H_0), (H_1) and (H_2), then*

$$\begin{cases} \pi_0(\rho) = 1-\rho, \\ \pi_1(\rho) = \rho - c_1(e_1)\rho^2 + \left(Nc_1(e_1) - \sum_{j=1}^N c_1(e_1+e_j)c_1(e_j)\right)\rho^3 + \mathcal{O}(\rho^4), \\ \pi_2(\rho) = c_1(e_1)\rho^2 - \left(Nc_1(e_1) - \sum_{j=2}^N c_1(e_1+e_j)c_1(e_j)\right)\rho^3 + \mathcal{O}(\rho^4), \\ \pi_m(\rho) = \left(\prod_{j=1}^{m-1} c_1(je_1)\right)\rho^m + \mathcal{O}(\rho^{m+1}), \quad m\geq 3 \end{cases} \tag{21}$$

where ρ tends to 0.

Proof. Equation (21) comes straightforwardly from equation (19) and two intermediate results, Lemma 9 and Lemma 10, postponed in Appendix. Note that, as at equilibrium the rates of incoming and outgoing customers are the same, i.e., $N\lambda = N\mu(1-\pi_0)$, it gives another way to obtain that $\pi_0(\rho) = 1-\rho$. ◀

4.2 Expansions for the local choice contribution function

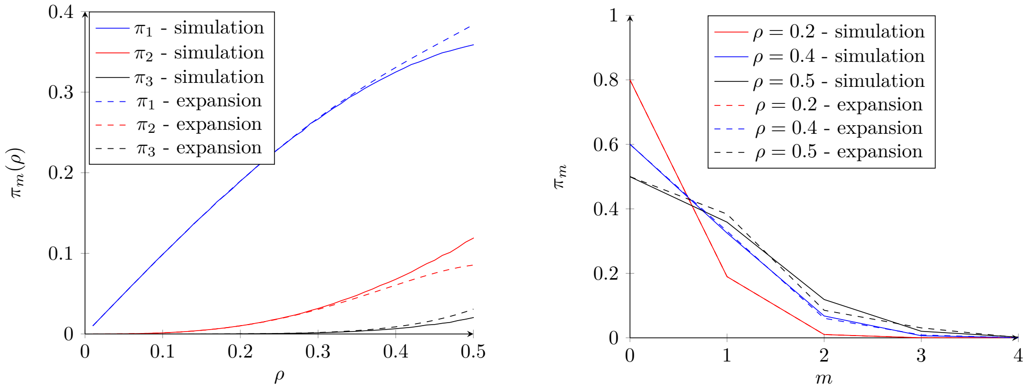
Equation (21) can be rewritten in the case of the local choice function $c = c^{lc}$ defined by equation (1). It gives the following result.

▶ **Corollary 7.** *For the local choice function c^{lc}, for $N \geq 3$*

$$\pi_0(\rho) = 1-\rho, \qquad \pi_1(\rho) = \rho - \frac{3}{2}\rho^3 + \mathcal{O}(\rho^4),$$
$$\pi_2(\rho) = \frac{3}{2}\rho^3 + \mathcal{O}(\rho^4), \qquad \pi_m(\rho) = \mathcal{O}(\rho^{m+1}), \quad m > 2.$$

For $N = 2$, the coefficient $3/2$ of ρ^3 is replaced by 2.

The main point is that $\phi_m(m)$ is null in this case. The aim will be to find the first non vanishing term of the expansion of $\pi_m(\rho)$ for every $m \geq 1$. It is the purpose of Section 4.6. Let us begin by giving more terms in the previous series expansion.



(a) For $m = 1$, 2 and 3, as a function of ρ. **(b)** For $\rho = 0.2,\ 0.4$ and 0.5.

Figure 1 Invariant distribution $\pi = (\pi_m, m \in \mathbb{N})$ of the number of customers in one queue for a set of N queues with local choice.

4.3 Further expansions for the local choice function

As the amount of cases to analyze grows exponentially with k, it is rather difficult to obtain further series expansions. The following expansions are obtained with help of mathematical software. For that, we observe the following property, which remains to be proved, that for each $k \in \mathbb{N}$, there exists $N_0(k)$ such that if $N > N_0(k)$, for each $m \geq 1$, $\phi_k(m)$ does not depend on N. For small values of k, it is easy to see that this property holds, given the recurrence equation and the local choice function. Using this, from global balance equations (5) for some N sufficiently large, the following result holds. For $\rho < 1$ tending to 0,

$$\begin{aligned}
\pi_0(\rho) &= 1 - \rho \\
\pi_1(\rho) &= \rho - \frac{3}{2}\rho^3 + \frac{11}{8}\rho^4 - \frac{7}{3}\rho^5 + \frac{10727}{2880}\rho^6 + \mathcal{O}(\rho^7) \\
\pi_2(\rho) &= \frac{3}{2}\rho^3 - \frac{11}{8}\rho^4 + \frac{47}{24}\rho^5 - \frac{1583}{320}\rho^6 + \mathcal{O}(\rho^7) \\
\pi_3(\rho) &= \frac{3}{8}\rho^5 + \frac{11}{9}\rho^6 + \mathcal{O}(\rho^7) \\
\pi_i(\rho) &= \mathcal{O}(\rho^7), \quad i > 3.
\end{aligned}$$

4.4 Validation by simulation

In Figure 1, we investigate numerically the accuracy of the previous expansion. Recall that π is the stationary queue length distribution of any queue in this symmetric system of N queues. In figure 1a, we plot π_m for $m = 0$, 1, 2 and 3 as a function of ρ given by simulation and by the series expansion at order 6. The conclusion is that the previous series expansion gives a quite good approximation for small values of ρ ($\rho \leq 0.3$), reasonable for $\rho \leq 0.4$. Figure 1b gives the distribution for different small values of ρ. It indicates that, as ρ increases, the distribution deviates from a geometric distribution. Moreover, the series expansion gives a quite good approximation for $\rho \leq 0.4$.

4.5 No choice policy: the case of independent queues

For the case where each queue receives independently customers at rate λ and serves them at rate μ, the contribution function becomes $c_i(n) = 1$, $n \in \mathbb{N}^N$ and $i \in \mathbb{N}$. We can easily verify that

$$\alpha_k(n) = (-1)^{k-|n|}\binom{N}{k-|n|}\mathbb{1}_{|n|\le k}$$

satisfies equation (8), where $|n| = n_1 + \ldots + n_N$. Using equation (20), we have for any $r \in \mathbb{N}$, $0 \le r \le k$,

$$\begin{aligned}\phi_k(k-r) &= (-1)^r \sum_{i=0}^{r} (-1)^i \binom{N-2+i}{i}\binom{N}{r-i} \\ &= (-1)^r \mathbb{1}_{\{r\le 1\}}. \end{aligned} \tag{22}$$

The term $\binom{N-2+i}{i}$ comes from the fact that we need to distribute the remaining i customers in the remaining $N-1$ queues. The last equality, of the form $a_r = b_r$ for all $r \in \mathbb{N}$, is obtained proving that the generating functions $\sum_{r\ge 0} a_r z_r$ and $\sum_{r\ge 0} b_r z_r$ are equal by developing the product

$$1 + z = (1+z)^N \frac{1}{(1+z)^{N-1}}.$$

With straightforward algebra, plugging equation (22) in equation (19), we retrieve that the stationary distribution $\pi(\rho) = (\pi_m(\rho), m \in \mathbb{N})$ for one queue is the geometric distribution with parameter ρ, as each queue is a $M/M/1$ queue with arrival-to-service-rate ratio $\rho = \lambda/\mu$.

4.6 Main result: Asymptotics for the stationary queue length distribution in light traffic

Let us then present the main result.

▶ **Proposition 8.** *For the local choice function c^{lc} defined by equation (1) and under assumption (H_0), for $m \ge 2$, the stationary probability $\pi_m(\rho)$ that a queue has m customers verifies*

$$\pi_m(\rho) = \begin{cases} 12\left(\dfrac{\rho}{2}\right)^{2m-1} + \mathcal{O}(\rho^{2m}) & \text{if } N \ge 3 \\ 2\rho^{2m-1} + \mathcal{O}(\rho^{2m}) & \text{if } N = 2 \end{cases}$$

when ρ tends to zero.

Proposition 8 guarantees that, for ρ sufficiently small, the probability of having m customers in the queue follows a geometric decay of parameter $\rho^2/4$ as m grows. The following table illustrates where the local choice is situated.

Table 1 Comparison of asymptotics for the stationary probability that a queue has more than k customers in light traffic (as parameter ρ tends to 0) for different allocation policies, $N \ge 3$.

Allocation policy	$u_k = \sum_{k\ge k} \pi_m$
No-choice	$\sim \rho^m$
Local choice	$\sim (\rho/2)^{2k-1}$
Random choice	$\sim \rho^{2^k-1}$

As expected, the performance of *local choice* policy is between the other two policies. However, for light traffic, its behavior is closer to *no choice* than to *random choice*. Indeed, the two first asymptotics are exponential while the third one is double exponential in ρ.

The light-traffic asymptotics obtained in this paper are for the limit when t tends to $+\infty$ first and then N tends to $+\infty$, since the asymptotic result is independent of N for $N \geq 3$, while from mean-field approximation for the *random choice model* the limit is when N first and then t tends to $+\infty$. The comparison we made is rigorous and justified by the interchange of the order of these two limits, see [14].

References

1. E. Andjel, M. Menshikov, and V. Sisko. Positive recurrence of processes associated to crystal growth models. *Annals of Applied Probability*, pages 1059–1085, 2006.
2. Johannes Pieter Cornelis Blanc. The power-series algorithm applied to the shortest-queue model. *Operations Research*, 40(1):157–167, 1992.
3. P. Bogdan, T. Sauerwald, A. Stauffer, and H. Sun. Balls into bins via local search. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 16–34. SIAM, 2013.
4. F. Ezanno. *Systèmes de particules en interaction et modèles de déposition aléatoire*. PhD thesis, Aix-Marseille Université, 2012.
5. F. Ezanno. Some results about ergodicity in shape for a crystal growth model. *Electronic Journal of Probability*, 18(33):1–20, 2013.
6. Ayalvadi Ganesh, Sarah Lilienthal, D Manjunath, Alexandre Proutiere, and Florian Simatos. Load balancing via random local search in closed and open systems. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 287–298. ACM, 2010.
7. N. Gast. The power of two choices on graphs: the pair-approximation is accurate? *ACM SIGMETRICS Performance Evaluation Review*, 43(2):69–71, 2015.
8. Nicolas Gast and Gaujal Bruno. A mean field model of work stealing in large-scale systems. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 13–24. ACM, 2010.
9. D. Gates and M. Westcott. Markov models of steady crystal growth. *Annals of Applied Probability*, pages 339–355, 1993.
10. K. Kenthapadi and R. Panigrahy. Balanced allocation on graphs. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 434–443. Society for Industrial and Applied Mathematics, 2006.
11. John Lamperti. Criteria for the recurrence or transience of stochastic process. i. *Journal of Mathematical Analysis and applications*, 1(3-4):314–330, 1960.
12. M. Mitzenmacher. *The Power of Two Choices in Randomized Load Balancing*. PhD thesis, Berkeley, 1996.
13. P. Robert. *Stochastic networks and queues*, volume 52. Springer, 2013.
14. N. Vvedenskaya, R. Dobrushin, and F. Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii*, 32(1):20–34, 1996.
15. G. Fayolle, V. Malyshev and M. Menshikov. *Topics in the constructive theory of countable Markov chains*, Cambridge university press, 1995.
16. V. A. Malyshev and M. Menshikov. Ergodicity, continuity and analyticity of countable Markov chains. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 39, 3-48, 1979.
17. K. Gardner, M. Harchol-Balter, A. Scheller-Wolf, M. Velednitsky and S. Zbarsky. Redundancy-d: The power of d choices for redundancy. *Operations Research*, 65(4): 1078–1094, 2017.

A Proof of Proposition 1

Proof. Assume that $\rho < 1$. We prove ergodicity by Foster's criterion for Markov processes based on a Lyapunov function (see for example [13, Proposition 8.14]). Here the Lyapunov function f is quadratic, given by $f(n) = n_1^2 + \cdots + n_N^2$, $n = (n_1, \ldots, n_N)$.

Let us denote $|n| = \sum_{i=1}^N n_i$. The infinitesimal generator is given by

$$\begin{aligned} Lf(n) &= \sum_{n' \in \mathbb{N}^N} Q(n, n')(f(n') - f(n)) \\ &= \sum_{i=1}^N \lambda c_i(n)(f(n + e_i) - f(n)) + \mathbb{1}_{n_i>0}\mu(f(n - e_i) - f(n)), \end{aligned} \tag{23}$$

for $f : \mathbb{N}^N \to \mathbb{R}$ with finite support. With straightforward algebra, using equation (1), it holds that

$$n_1 c_1^{lc}(n) + \cdots + n_N c_N^{lc}(n) \leq |n| \text{ and } c_1^{lc}(n) + \cdots + c_N^{lc}(n) = N. \tag{24}$$

This gives

$$\begin{aligned} L(f)(n) &= \lambda \sum_{i=1}^N c_i^{lc}(n)((n_i + 1)^2 - n_i^2) + \mu \sum_{i=1}^N \mathbb{1}_{n_i>0}((n_i - 1)^2 - n_i^2) \\ &= 2\lambda \sum_{i=1}^N c_i^{lc}(n) n_i + \lambda \sum_{i=1}^N c_i^{lc}(n) - 2\mu \sum_{i=1}^N n_i + \mu \sum_{i=1}^N \mathbb{1}_{n_i>0} \\ &\leq (\lambda + \mu)N - 2(\mu - \lambda)|n|. \end{aligned} \tag{25}$$

By the equivalence of norms in $\mathbb{R}^N$, there is a constant $C > 0$ such that, for all n, $\sqrt{f(n)} \leq C^{-1}|n|$ where $|n| = n_1 + \cdots + n_N$. Thus, if $f(n) > K$ then $|n| \geq C\sqrt{K}$. As $\rho = \lambda/\mu < 1$, K can be chosen large enough to get $\gamma = -(\lambda + \mu)N + 2(\mu - \lambda)C\sqrt{K} > 0$.

Thus, by equation (25), if $f(n) > K$ then $L(f)(n) \leq -\gamma$. Moreover the set $F = \{n \in \mathbb{N}^N, f(n) \leq K\}$ is finite and the random variables $\sup_{0 \leq s \leq 1} f(X(s))$ and $\int_0^1 L(f)(X(s))ds$ are integrable. Indeed,

$$\sup_{0 \leq s \leq 1} f(X(s)) \leq C^{-2} \sup_{0 \leq s \leq 1} |X(s)|^2 \leq C^{-2}(\mathcal{N}_{\lambda N}([0, 1[)^2$$

where the arrival process in the system, denoted by $\mathcal{N}_{\lambda N}$, is a Poisson process with intensity $\lambda N ds$, as the sum of the N independent Poisson processes with parameter λ of arrivals at the N queues. Using again equation (25),

$$\int_0^1 L(f)(X(s))ds \leq (\lambda + \mu)N.$$

Thus, the Markov process $(X(t))_{t \geq 0}$ is ergodic if $\rho < 1$.

If $\rho > 1$, we apply [13, Theorem 8.10], a simplified version of a Lamperti's result, to prove the transience of the embedded Markov chain (M_n) at jump times of $(X(t))$. It gives the transience of $(X(t))$. Let g be defined by $g(n) = n_1 + \cdots + n_N$. Using that $c_1^{lc}(n) + \ldots + c_N^{lc}(n) = N$, see equation (24), for all $n \in \mathbb{N}^N$,

$$\mathbb{E}_n(g(M_1) - g(n)) = Lg(n) = \lambda \sum_{i=1}^N c_i(n) - \mu \sum_{i=1}^N \mathbb{1}_{\{n_i>0\}} \geq (\lambda - \mu)N > 0.$$

Moreover, for all $n \in \mathbb{N}^N$,

$$\mathbb{E}_n(|g(M_1) - g(n)|^2) = \sum_{n' \in \mathbb{N}^N} Q(n, n')|g(n') - g(n)|^2$$
$$= \lambda \sum_{i=1}^{N} c_i(n) + \mu \sum_{i=1}^{N} \mathbb{1}_{\{n_i > 0\}} \leq (\lambda + \mu)N,$$

thus $\sup_{n \in \mathbb{N}^N} \mathbb{E}_n(|g(M_1) - g(n)|^2) < \infty$. The sufficient conditions for applying [13, Theorem 8.10] hold. It ends the proof. ◀

B Two lemmas

▶ **Lemma 9.** *For integer k, $0 \leq k \leq 3$, the coefficients $\phi_k(m)$, $m \in \mathbb{N}$, are given by*

$$\phi_0(0) = 1, \text{ and } \phi_0(m) = 0, \; m > 0,$$
$$\phi_1(0) = -1, \; \phi_1(1) = 1 \text{ and } \phi_1(m) = 0, \; m > 1,$$
$$\phi_2(0) = 0, \; \phi_2(1) = -c_1(e_1), \; \phi_2(2) = c_1(e_1) \text{ and } \phi_2(m) = 0, \; m > 2,$$
$$\phi_3(0) = 0, \; \phi_3(1) = -\phi_3(2) = Nc_1(e_1) - \sum_{j=1}^{N} c_1(e_1 + e_j)c_1(e_j),$$
$$\phi_3(3) = c_1(e_1)c_1(2e_1) \text{ and } \phi_3(m) = 0, \; m > 3.$$

Proof. We use, for $k \leq 3$, the expressions of α_k given by Proposition 5 to compute ϕ_k. ◀

▶ **Lemma 10.** *For $k \geq 1$, $\phi_k(k) = \alpha_k(ke_1) = \prod_{j=1}^{k-1} c_1(je_1)$.*

Proof. For $k \in \mathbb{N}^*$, by equation (20), $\phi_k(k) = \alpha_k(ke_1)$. Taking $n = ke_1$ in equation (8),

$$\alpha_k(ke_1) = \sum_{i=1}^{N} \alpha_k(ke_1 + e_i) + c_1((k-1)e_1)\alpha_{k-1}((k-1)e_1) - N\alpha_{k-1}(ke_1).$$

By Lemma 4, for any i, $1 \leq i \leq N$, $\alpha_k(ke_1 + e_i) = 0$ and $\alpha_{k-1}(ke_1) = 0$. It gives that

$$\phi_k(k) = c_1((k-1)e_1)\phi_{k-1}(k-1).$$

This recurrence equation in $\phi_k(k)$ leads to the desired result, since $\phi_1(1) = 1$. ◀

C Proof of Proposition 8

Proof. In the proof, the following definition will be used.

▶ **Definition 11.** The state $n = (n_1, \dots, n_N)$ exists at order k if and only if, in equation (6), $\alpha_k(n) \neq 0$.

First step. To prove Proposition 8, the first step is to obtain that, for a state $n = (n_1, \dots, n_N)$ existing at order k, the maximum possible queue length is $\lceil k/2 \rceil$. Indeed, by Lemma 4, n exists at order k only if $|n| \leq k$. Moreover, the following result holds.

▶ **Lemma 12.** *Let $k \in \mathbb{N}$ and $n = (n_1, \dots, n_N) \in \mathbb{N}^N$. If $|n| \leq k$ and $n_1 > \lceil k/2 \rceil$ then $\alpha_k(n) = 0$.*

Proof. The following assertion is proved by induction on $p \geq 0$.

($\mathcal{B}_p$) For $k \in \mathbb{N}$ and $n = (n_i)_{1 \le i \le N}$, if $|n| = k - p$ and $n_1 > \lceil k/2 \rceil$ then $\alpha_k(n) = 0$.

Let us prove ($\mathcal{B}_0$). Let $k \in \mathbb{N}$ and n such that $|n| = k$ and $n_1 > \lceil k/2 \rceil$. As $|n| = k$, by Lemma 4, for each i, $1 \le i \le N$, $\alpha_k(n+e_i) = \alpha_{k-1}(n) = 0$. Thus equation (8) is rewritten as

$$\left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i>0\}}\right) \alpha_k(n) = \sum_{i=1}^{N} \mathbb{1}_{\{n_i>0\}} c_i^{lc}(n-e_i)\alpha_{k-1}(n-e_i). \tag{26}$$

As $|n| \le k$ and $n_1 > \lceil k/2 \rceil$, $n_2 + n_N \le k - n_1 < k - \lceil k/2 \rceil < n_1$. Thus $n_2 + n_N \le k - \lceil k/2 \rceil - 1 < n_1 - 1$.

It means that each neighboring queue of queue 1 has strictly less than $n_1 - 1$ customers. Thus the contribution on queue 1 for our local choice function c^{lc} defined by equation (1) gives $c_1^{lc}(n-e_1) = 0$ and equation (26) can be rewritten

$$\left(\sum_{i=1}^{N} \mathbb{1}_{\{n_i>0\}}\right) \alpha_k(n) = \sum_{i=2}^{N} \mathbb{1}_{\{n_i>0\}} c_i^{lc}(n-e_i)\alpha_{k-1}(n-e_i). \tag{27}$$

Therefore, state n exists at order k only if there is $i_1 \ne 1$ such that $\alpha_{k-1}(n-e_{i_1}) \ne 0$. But $|n-e_{i_1}| = k-1$ and we can repeat the previous arguments for $k-1$ instead of k and $n-e_{i_1}$ instead of n, with $(n-e_{i_1})_1 > \lceil k/2 \rceil \ge \lceil (k-1)/2 \rceil$, and so on until we obtain $n_1 e_1$. In conclusion, n exists at order k only if $\alpha_{n_1}(n_1 e_1) \ne 0$. It contradicts Lemma 10. Therefore $\alpha_k(n) = 0$.

Assume now, for $p \ge 1$, that ($\mathcal{B}_{p-1}$) is true, and prove ($\mathcal{B}_p$). For that, let $k \in \mathbb{N}$ and n be such that $|n| = k - p$ and $n_1 > \lceil k/2 \rceil$. By induction assumption ($\mathcal{B}_{p-1}$), applied to k and $n + e_i$ as $|n + e_i| = k - (p-1)$, then to $k-1$ and n as $|n| = k - 1 - (p-1)$, it holds that $\alpha_k(n+e_i) = \alpha_{k-1}(n) = 0$. Then the arguments used for ($\mathcal{B}_0$) give that $\alpha_k(n) = 0$. It ends the proof. ◀

One can then deduce easily the following result.

▶ **Lemma 13.** *Let m be in $\mathbb{N}^*$. The first possibly non vanishing term of the expansion when ρ tends to zero of the stationary probability $\pi_m(\rho)$ that a queue has m customers is $\phi_{2m-1}(m)\rho^{2m-1}$.*

Proof. For $m \in \mathbb{N}$, by definition, see (19), $\pi_m(\rho) = \sum_{k \ge 0} \phi_k(m)\rho^k$ with

$$\phi_k(m) = \sum_{\substack{n=(m,n_2,\dots,n_N) \\ |n| \le k}} \alpha_k(m, n_2, \dots, n_N).$$

If $k < 2m - 1$ then, for $n = (m, n_2, \dots, n_N)$ such that $|n| \le k$, $n_1 = m > \lceil k/2 \rceil$. Thus, by Lemma 12, all the $\alpha_k(m, n_2, \dots, n_N)$ in the right-hand side of the previous equation are null for $k < 2m - 1$. It ends the proof. ◀

Second step. Moreover the states which exist at order $k = 2m - 1$ with one queue with the maximum value m correspond just to two neighboring queues with m and $j < m$. It is given by the following lemma.

▶ **Lemma 14.** *If $|n| \le k = 2m - 1$ (k odd), $n_1 = m$ and there exists two distinct j and l, different from 1, such that $n_j > 0$ and $n_l > 0$ then $\alpha_k(n) = 0$.*

Proof. The following assertion is proved by induction on $p \ge 0$

($\mathcal{B}_p$) For $k = 2m-1$, $m \in \mathbb{N}$, for n such that $|n| = k - p$, $n_1 = m$, $n_j > 0$ and $n_l > 0$ with j and l distinct, different from 1, then $\alpha_k(n) = 0$.

Let us prove ($\mathcal{B}_0$). Let $k = 2m-1$ and n chosen as indicated. As $|n| = k$, by Lemma 4, for each i, $1 \le i \le N$, one gets $\alpha_k(n + e_i) = \alpha_{k-1}(n) = 0$. As before, using Lemma 2, equation (26) holds. By assumption, as in the proof of Lemma 12, it holds that each neighboring queue of queue 1 has strictly less than $n_1 - 1$ customers, which yields $c_1(n - e_1) = 0$. Thus equation (26) can be rewritten equation (27). We conclude as in the proof of Lemma 12. ◀

Step 3. We distinguish two cases:
Case 1: $N \ge 3$. From equation (20) and applying Lemma 14,

$$\phi_{2m-1}(m) = \sum_{i=1}^{m-1} \alpha_{2m-1}(m, i, 0, ...0) + \sum_{i=1}^{m-1} \alpha_{2m-1}(m, 0, ..., 0, i) + \alpha_{2m-1}(m, 0, ..., 0)$$

then by symmetry,

$$\phi_{2m-1}(m) = 2 \sum_{i=1}^{m-1} \alpha_{2m-1}(m, i, 0, ...0) + \alpha_{2m-1}(m, 0, 0, ...0). \tag{28}$$

This means that only these terms are non null. The rest of the proof consists in deriving them.

Let n_1 and n_2 be chosen as follows: $n_1 = (k+1)/2$ and $n_2 = (k-1)/2$. Using the same arguments as in Lemma 12, equation (8) gives, for $k = 2m-1$ with m integer and $m \ge 2$,

$$2\alpha_{2m-1}(m, m-1, 0, ..., 0) = \frac{1}{2}\alpha_{2(m-1)}(m-1, m-1, 0, ..., 0). \tag{29}$$

Let $k = 2m$, and $n = (m, m, 0, ..., 0)$. For $m \in \mathbb{N}$, $m \ge 2$, as $c_2^{lc}(m, m-1, 0, ..., 0) = 1$, equation (8) gives

$$2\alpha_{2m}(m, m, 0, ..., 0) = 2\alpha_{2m-1}(m, m-1, 0, ..., 0). \tag{30}$$

Combining equations (29) and (30), for $m \ge 3$,

$$\alpha_{2m-1}(m, m-1, 0, ..., 0) = \frac{1}{2^2}\alpha_{2m-3}(m-1, m-2, 0, ..., 0)$$

and then, using equation (17) to show that $\alpha_3(2, 1, 0, ..., 0) = 3/8$, for $m \ge 3$,

$$\alpha_{2m-1}(m, m-1, 0, ..., 0) = \frac{1}{2^{2(m-2)}}\alpha_3(2, 1, 0, ..., 0) = \frac{3}{2^{2m-1}}. \tag{31}$$

Then, for $n = (m, i, 0, ..., 0)$, for $0 < i < m-1$, from equation (8),

$$2\alpha_{2m-1}(m, i, 0, ..., 0) = \alpha_{2m-1}(m, i+1, 0, ..., 0).$$

By induction and using equation (31), for $0 < i < m-1$,

$$\alpha_{2m-1}(m, i, 0, ..., 0) = \frac{1}{2^{m-1-i}}\alpha_{2m-1}(m, m-1, 0, ..., 0) = \frac{3}{2^{2m-1}}\frac{1}{2^{m-1-i}}. \tag{32}$$

With similar arguments and then using equation (32) for $i = 1$,

$$\begin{aligned}\alpha_{2m-1}(m, 0, 0, ..., 0) &= \alpha_{2m-1}(m, 1, 0, ..., 0) + \alpha_{2m-1}(m, 0, ..., 0, 1) \\ &= \frac{6}{2^{2m-1}}\frac{1}{2^{m-2}}.\end{aligned} \tag{33}$$

Plugging equations (31), (32) and (33) in equation (28),

$$\begin{aligned}\phi_{2m-1}(m) &= 2\frac{3}{2^{2m-1}} + 2\sum_{i=1}^{m-2} \frac{3}{2^{2m-1}} \frac{1}{2^{m-1-i}} + \frac{6}{2^{2m-1}} \frac{1}{2^{m-2}} \\ &= \frac{6}{2^{2m-1}} \left(\sum_{i=1}^{m-1} \frac{1}{2^{m-1-i}} + \frac{1}{2^{m-2}} \right) = \frac{12}{2^{2m-1}}\end{aligned}$$

Using it in Lemma 13 gives the result.

Case 2: $N = 2$. With similar arguments, equation (28) is rewritten in this case

$$\phi_{2m-1}(m) = \sum_{i=1}^{m-1} \alpha_{2m-1}(m, i) + \alpha_{2m-1}(m, 0).$$

while equations (29) and (30) become $2\alpha_{2m-1}(m, m-1) = \alpha_{2(m-1)}(m-1, m-1)$ and $2\alpha_{2m}(m, m) = 2\alpha_{2m-1}(m, m-1)$. Following exactly the same lines as in case 1, one gets

$$\phi_{2m-1}(m) = \sum_{i=1}^{m-1} \frac{1}{2^{m-1-i}} + \frac{1}{2^{m-2}} = 2.$$

It ends the proof. ◀