



Stability of non markovian polling systems

Laurent Massoulié

► To cite this version:

Laurent Massoulié. Stability of non markovian polling systems. RR-2148, INRIA. 1993. inria-00074524

HAL Id: inria-00074524

<https://inria.hal.science/inria-00074524>

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

Stability of non Markovian Polling Systems

Laurent MASSOULIÉ

N° 2148

Novembre 1993

PROGRAMME 1

Architectures parallèles,
bases de données,
réseaux et systèmes distribués

*Rapport
de recherche*

1993

Stability of non Markovian Polling Systems

Laurent Massoulié *

November 29, 1993

Abstract

In this article we consider polling systems with markovian server routings, and where each station is attended according to a specific policy. A stationary regime for this system is constructed under general statistical assumptions (stationarity, ergodicity) on the input processes to the stations (in particular it is not required that these processes be mutually independent).

The method of construction is as follows: one constructs recursively a sequence of stationary regimes for fictive systems that approximate in some sense the original polling system; the stationary regime is then identified as the limit of this sequence of stationary processes. The main tools for these results are Palm calculus and Birkhoff's ergodic theorem.

It is shown by a coupling argument that this stationary regime is minimal in the stochastic ordering sense. The assumptions on the service policies allow to consider the purely gated policy, the α -limited policy, the binomial-gated policy and others. As a by-product sufficient conditions for the stationary regime of a $G/G/1/\infty$ queue with multiple server vacations are obtained.

Keywords: polling systems, stationary regime, stability, ergodicity, coupling, point processes, Palm calculus, stochastic ordering, queues with server vacations.

*Laboratoire des Signaux et Systèmes, CNRS-ESE, Plateau de Moulon, 91192 Gif-sur-Yvette, France.

Stabilité de systèmes de polling non markoviens*

Laurent Massoulié †

29 Novembre 1993

Résumé

On considère dans cet article un réseau de polling avec routages markoviens du serveur et où chaque station est traitée selon une politique de service spécifique. Un régime stationnaire pour ce système est construit sous des hypothèses statistiques générales (stationnaires ergodiques) sur les processus d'arrivée dans les stations, qu'on ne suppose pas mutuellement indépendants.

La technique de construction est la suivante: une suite de régimes stationnaires pour des systèmes fictifs approximant le système de polling est construite récursivement; le régime stationnaire est alors identifié comme la limite de cette séquence de processus stationnaires. Les principaux outils utilisés sont le calcul de Palm et le théorème ergodique de Birkhoff.

Une technique de couplage permet de montrer que ce régime stationnaire est minimal au sens de l'ordre stochastique. Les hypothèses faites sur les politiques de service permettent de considérer entre autres les politiques "purely gated", " α -limited" et "binomial-gated". On obtient accessoirement des conditions suffisantes pour qu'une file $G/G/1/\infty$ avec vacances multiples du serveur soit ergodique.

Mots Clés: systèmes de polling, régime stationnaire, stabilité, ergodicité, couplage, processus ponctuels, calcul de Palm, ordre stochastique, files d'attente avec vacances du serveur.

*Ce rapport est publié dans le cadre de la collaboration avec P. Brémaud.

†Laboratoire des Signaux et Systèmes, CNRS-ESE, Plateau de Moulon, 91192 Gif-sur-Yvette, France.

1 Introduction

In the study of polling systems, the issue of stability is of especial importance. Heuristic stability conditions for polling systems have been known for a long time (see Takagi [12] p. 11 and the references therein), but proofs of the existence of a stationary regime under these conditions lacked until the recent publication of several articles that settled the question for a wide scope of markovian polling systems (that is to say, polling systems in which the arrivals to each queue are mutually independent M/G processes). For instance, Borovkov and Schassberger [4] prove that a polling ring with markovian server routings (after attending queue i , server switches to queue j with probability $p_{i,j}$, a switch-over time $W_{i,j}$ with mean $w_{i,j}$ occurring between departure from i and arrival to j) and with limited service policy at each queue is stable if and only if:

$$\rho + \frac{\lambda_i}{\pi_i F_i} w < 1 \text{ for all } i \in \{1, \dots, K\} \quad (1)$$

where λ_i is the arrival rate at queue i , ρ is the total traffic intensity $\sum_{i=1}^K \lambda_i \sigma_i$, $w = \sum_{i,j=1}^K \pi_i p_{i,j} w_{i,j}$ is the mean switch-over time between departure and arrival of the server to consecutively visited queues (here, $(\pi_i)_{1 \leq i \leq K}$ denotes the stationary distribution of the transition matrix $(p_{i,j})_{1 \leq i,j \leq K}$ and F_i is the mean of the maximal number of customers that may be served in a single visit to queue i).

In [7], Fricker and Jaibi consider polling systems where the server attends the queues in periodic (but not necessarily cyclic) order, and uses different service policies at different stages in a cycle; they prove that such systems are stable if and only if

$$\rho + \frac{\lambda_i}{F_i} w < 1 \text{ for all } i \in \{1, \dots, K\} \quad (2)$$

where w is the mean of the total switch-over time in a cycle, and F_i is the mean of the maximal number of customers that may be served per cycle in queue i (in the earlier paper of Georgiadis and Szpankowski [8], this result is derived for cyclic polling systems with limited service policy at each queue; see also Altman et al. [1], where Foster's criteria are used to derive sufficient conditions of stability for cyclic polling systems with mixed service policies).

In the case of polling systems fed by non-Poissonian arrival processes, the only stability result known to us is the existence of a stationary regime for a $G/G/1/\infty$ queue with multiple server vacations (this system can be viewed as a polling system with a single station attended according to the exhaustive policy), which is a byproduct of the decomposition result established in Doshi [6]. However the case of General Ergodic arrival processes is of great interest for applications, since the Poisson assumption is generally unrealistic (for instance, if the arrival processes into a polling system are the outputs from another queueing system, they usually fail to be Poisson and also fail to be mutually independent).

In this article, we consider polling systems where the arrivals at each station are stationary ergodic (G/G), not necessarily independent processes, the server switches from queue to queue according to markovian routings as in Borovkov and Schassberger's model, and each queue is attended according to a specific policy. A stationary regime is constructed under the assumption that condition

$$\rho + \left(\sum_{j=1}^K \frac{\lambda_j}{\pi_j F_j} \right) w < 1 \quad (3)$$

holds (here, w , the (π_j) and F_j are as in (1)). The policies that we consider are of gated type, and satisfy a monotonicity condition that is inspired from the one imposed on the policies considered in [7]; we at

first consider deterministic policies only, but, as is explained in Section 7, the reasoning can be extended to include random policies such as the binomial gated policy introduced by Levy (see [9]).

A stationary regime for the polling system under consideration is constructed in a recursive way: the procedure is initiated by considering the stationary workload M^0 of the system where the server attends the customers of each queue indifferently, and takes vacations at each time it empties the queue or returns from a vacation to an empty queue, where the stationary sequence of vacations S_n is the sequence of switch-over times occurring in the real polling model. The existence of this initial workload process follows from [6]. However, the question whether this stationary regime is ergodic or not is not answered in [6]. Since ergodicity plays a crucial role in our proof we establish in the appendix sufficient conditions for the stationary regime of a $G/G/1/\infty$ queue with multiple server vacations to be ergodic. The monotonicity properties of the service policies enable to construct a sequence of stationary workloads M^m , $m \geq 0$ that is non decreasing, so that a limiting process M^∞ exists. From an ergodic argument it follows that this limiting process is infinite with probability zero or one; it is then shown that under condition (3) the process M^∞ is almost surely finite, and that it is a stationary workload for the original polling system. It is also shown by a coupling construction that M^∞ is the minimal stationary workload.

In order to establish the existence of a stationary regime for a system when the only assumption made on the input process into the system (here, the arrival processes at the polling stations) is its ergodicity, the only general method available consists in using some monotonicity property of the system to construct a monotonic sequence of processes that converges eventually to the desired stationary regime (see in the book of Brandt, Franken and Lisek [5] e.g. the proof of Loynes' theorem and Brandt's construction of stationary regimes for the $G/G/m/\infty$ queue). In this regard, the scheme of proof presented above is not surprising. The main difficulty in order to apply such a scheme of proof resides in the choice of the right monotonicity properties of the system, and in the choice of an adequate recursive construction. One can note also that in the proof of Loynes' theorem, a random variable $M(0)$ is constructed recursively, and if the workload of the $G/G/1/\infty$ queue at time 0 is $M(0)$, then the workload process on \mathbb{R}^+ is stationary; the recursive construction involved in this proof is "local", involving the construction of the workload at time 0 rather than the construction of the whole workload process (in this sense, Brandt's construction of stationary regimes for the $G/G/m/\infty$ queue is also "local"). On the other hand, the complexity of polling systems makes it impossible to prove the existence of stationary regimes with such a local construction: indeed, in order to construct the workload in the system on \mathbb{R}^+ , knowledge of the workload at time 0 is not sufficient; one also needs to know the number of customers present in each station, which queue is currently attended or towards which queue the server is currently switching, the number of customers still to be served in the queue currently attended before the next switch-over time or the remaining switch-over time if the server is currently switching... Therefore the recursive construction carried over in this paper is a "global" one, involving at each step whole processes $\{M^m(t)\}_{t \in \mathbb{R}}$ rather than random variables $M^m(0)$.

This paper is organized as follows: section 2 is devoted to the presentation of some classical results of Palm Calculus, a convenient tool when dealing with ergodic point processes; the polling model considered here is described precisely in Section 3; the recursive construction of the stationary workload and the proof that condition (3) is sufficient for the stability of the model constitute Section 4 and 5 respectively. In Section 6 it is shown that the stationary workload constructed is stochastically dominated by any transient workload in the asymptotic $t \rightarrow \infty$; this implies that, if there are more than one stationary regime for the model, the one constructed here is the smallest in the stochastic ordering sense. In Section 7 it is explained how the reasonings of Sections 4 and 5 can be extended to random policies, and a

necessary condition for stability of the model is given; as a concluding remark we point out a possible way to prove that this necessary condition is also sufficient for stability of the model.

2 Background on Palm Calculus

For the proofs of the results stated in this Section and for further details and references, the reader is addressed to the book of Baccelli and Brémaud [3], chapter 1.

Let a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ be given, on which a measurable semi-group of automorphisms $\{\theta_t\}$, $t \in \mathbf{R}$ is defined. Suppose that \mathbf{P} is stationary in the sense that for all $t \in \mathbf{R}$, $\mathbf{P} \circ \theta_t^{-1} = \mathbf{P}$. A locally finite point process N is given defined on $(\Omega, \mathcal{F}, \mathbf{P})$ such that for all $C \in \mathcal{B}(\mathbf{R})$, $\omega \in \Omega$, $t \in \mathbf{R}$

$$N(\theta_t \omega, C) = N(\omega, C + t) \quad (4)$$

Suppose that its average intensity is finite and non-null:

$$0 < \lambda(N) = \mathbf{E}[N(0, 1)] < \infty$$

The Palm probability \mathbf{P}_N^0 associated with (\mathbf{P}, N) is defined by

$$\mathbf{P}_N^0(A) = \frac{1}{(b-a)\lambda(N)} \mathbf{E}\left[\int_{(a,b)} (\mathbf{1}_A \circ \theta_s) N(ds)\right] \quad (5)$$

It does not depend on $(a, b] \subset \mathbf{R}$ and verifies

$$\begin{cases} \mathbf{P}_N^0[N(\{0\}) = 1] = 1 \\ \mathbf{P}_N^0 \circ \theta_{T_n} = \mathbf{P}_N^0 \\ \mathbf{E}_N^0[T_1] = \frac{1}{\lambda(N)} \end{cases} \quad (6)$$

where $\{T_n\}$, $n \in \mathbf{Z}$, is the sequence of points of N with the convention that $T_0 \leq 0 < T_1$ (implying in particular $\mathbf{P}_N^0[T_0 = 0] = 1$). The original, θ_t -invariant probability \mathbf{P} can be recovered from the Palm probability \mathbf{P}_N^0 by Ryll-Nardzewski and Slivnyak's inversion formula

$$\mathbf{P}(A) = \lambda(N) \mathbf{E}_N^0\left[\int_0^{T_1} \mathbf{1}_A \circ \theta_s ds\right] \quad (7)$$

Let \mathbf{P}^0 be a probability on (Ω, \mathcal{F}) such that

$$\mathbf{P}^0(\Omega^0) = 1 \quad (8)$$

where $\Omega^0 = \{T_0 = 0\}$. Suppose that \mathbf{P}^0 is θ_{T_n} -invariant:

$$\mathbf{P}^0(\theta_{T_n} \in \cdot) = \mathbf{P}^0(\cdot), \quad n \in \mathbf{Z} \quad (9)$$

Moreover assume that

$$0 < \mathbf{E}^0[T_1] < \infty$$

The probability \mathbf{Q} defined by

$$\mathbf{Q}(A) = \frac{1}{\mathbf{E}^0[T_1]} \mathbf{E}^0\left[\int_0^{T_1} \mathbf{1}_A \circ \theta_s ds\right] \quad (10)$$

is θ_t -invariant, and \mathbf{P}^0 is then the Palm probability \mathbf{Q}_N^0 associated with the stationary point process $(N, \theta_t, \mathbf{Q})$ (this is Slivnyak's inverse construction).

Let N' be another locally finite point process defined on $(\Omega, \mathcal{F}, \mathbf{P})$, for which (4) is also satisfied (one says that the processes N and N' are jointly θ_t -stationary). Suppose that the average intensity of N' is non-null and finite

$$0 < \lambda(N') = \mathbf{E}[N'(0, 1)] < \infty$$

The two Palm probabilities \mathbf{P}_N^0 and $\mathbf{P}_{N'}^0$, are connected by Neveu's exchange formula

$$\text{for all } f \geq 0, \lambda(N) \mathbf{E}_N^0[f] = \lambda(N') \mathbf{E}_{N'}^0\left[\int_{(0, T'_1]} (f \circ \theta_s) N(ds)\right] \quad (11)$$

where $\{T'_n\}$, $n \in \mathbf{Z}$ is the sequence of points of N' with the convention that $T'_0 \leq 0 < T'_1$.

3 Model Description

A polling system with K infinite buffer queues, indexed by $i \in \{1, \dots, K\}$ is served by a single server. A cycle is the time interval between successive arrivals of the server at queues. At the end of service at queue i , a K -faced die is tossed independently of every other randomness involved in the model to choose which queue will be served next; the die will choose queue j with probability $p_{i,j}$ (markovian routing of the server). The markovian matrix $(p_{i,j})_{1 \leq i, j \leq K}$ is assumed to be irreducible, and thus admits a unique stationary probability distribution $(\pi_i)_{1 \leq i \leq K}$.

Assume that queue i is selected for service in the n^{th} cycle; upon completion of service at queue i , the switch of the server to queue j (selected by the die roll) requires a switch-over time $W_{i,j}(n)$. The sequences $(W_{i,j}(n))_{n \in \mathbf{Z}}$, $1 \leq i, j \leq K$ are mutually independent, each one is i.i.d. and independent of the server routings, of the customers arrival processes and of the customers service times. Note

$$w = \sum_{i,j=1}^K \pi_i p_{i,j} w_{i,j}$$

the mean of the switch-over time in one cycle.

A deterministic function $f_i : \mathbf{N} \rightarrow \mathbf{N}$ determines the number of customers served during a cycle where queue i is selected for service; f_i is called the service policy at queue i . If the server finds X_i customers waiting at queue i upon arrival at that queue, it will serve $f_i(X_i)$ customers of that queue (in *FIFO* order) before switching to another queue and beginning the next cycle. For every $i \in \{1, \dots, K\}$, the service policy f_i is assumed to satisfy the following conditions:

- (i) for all $x \in \mathbf{N}$, $f_i(x) \leq x$
- (ii) for all $x, y \in \mathbf{N}$, $x \leq y \Rightarrow f_i(x) \leq f_i(y)$
- (iii) for all $x, y \in \mathbf{N}$, $x \leq y \Rightarrow x - f_i(x) \leq y - f_i(y)$

Hypothesis (i) says that the policy at each queue is gated: if a gate is closed behind the last customer in waiting room i upon arrival of the server to that queue, then customers arriving after closure of the gate will not be served in the considered cycle. The monotonicity assumptions (ii) and (iii) will play a crucial role in Section 3, and are to be compared with the monotonicity assumption P4 satisfied by the service policies considered in [7]. We note $F_i = \sup_{x \in \mathbf{N}} f_i(x)$; if $F_i < +\infty$, one says that policy f_i is limited

with bound F_i ; if $F_i = +\infty$, f_i is called an unlimited policy. We give some simple examples of policies satisfying (i) – (iii):

$$\begin{cases} f(x) = x & \text{(purely gated policy)} \\ f(x) = x \wedge a & \text{for some integer } a \text{ (} a\text{-limited policy)} \\ f(x) = [(x - a)^+] \wedge b & \text{for some integers } a, b \\ f(x) = \text{int}(\alpha x) & \text{for some } \alpha \in (0, 1), \text{ where } \text{int} \text{ is the integer part function} \end{cases}$$

Finally, the marked point processes $A_i = \{T_n(i), \sigma_n(i)\}_{n \in \mathbb{Z}}$, $1 \leq i \leq K$, describing the arrival times of the customers and their service times, are only assumed to be jointly stationary, ergodic processes independent of the switch-over times and of the routings of the server.

4 The Recursive Construction Scheme

4.1 Informal description

The construction procedure is initiated by considering the single server queue with aggregated arrival process $A = \sum A_i$ and server's vacations occurring when there is no work in the system, the sequence of vacations $\{S_n\}$ being exactly the sequence of switch-over times that occur in the original polling system. The stationary workload process M^0 is constructed from the point process $\Delta^0 = \{\delta_n^0\}$ counting vacations beginnings (a vacation of length S_n begins at time δ_n^0). When the server in this system (system 0) returns from a vacation to a queue i (i.e., at some time $\delta_n^0 + S_n$ where S_n is of type $W_{i,j}$, $j \in \{1, \dots, K\}$) it first serves an appropriate number of type i customers according to the service policy f_i (this is achieved at time $\delta_{n+1}^1 \in [\delta_n^0 + S_n, \delta_{n+1}^0]$). The corresponding congestion process is $X^0 = (X_1^0, \dots, X_K^0)$.

One considers then the (virtual) vacation system where vacation S_n occurs at δ_n^1 rather than at δ_n^0 , the corresponding workload process being denoted M^1 . In system 1, the server returning from a vacation to a queue i (i.e., at some time $\delta_n^1 + S_n$ with S_n of type $W_{j,i}$, $j \in \{1, \dots, K\}$) first attends an appropriate number of type i customers according to service policy f_i ; the corresponding congestion process is X^1 . The construction of system 1's stationary regime (Δ^1, M^1, X^1) is possible (i.e., it can be done on a single cycle of the M^0 process) due to the monotonicity properties (ii) and (iii) of the service policies f_i .

By applying the procedure to M^1 and so forth, a sequence of stationary regimes (Δ^m, M^m, X^m) , $m \geq 0$ is obtained. Each virtual vacation system m approximates the original polling system in the following sense: in system m ($m \geq 0$) the server returning from a vacation to a queue i first attends type i customers according to the service policy f_i . However, when this is achieved, service may be provided to other customers before the next vacation occurs.

The essential feature of this approximation scheme (apart from its stationarity) is non decreasingness of the processes M^m ($M^m \leq M^{m+1}$, $m \geq 0$).

4.2 First Step

Stationary regime of a $G/G/1/\infty$ queue with multiple server vacations: Let $\{Y_n\}$ be a stationary Markov chain on $\{1, \dots, K\}$ with transition probabilities $(p_{i,j})$. Use the sequences of switch-over times $\{W_{i,j}(n)\}$ (that are i.i.d., mutually independent and independent of $\{Y_n\}$) to construct the marked

point process $\Delta \doteq \{\delta_n, s_n\}$ as follows:

$$\begin{cases} \delta_0 = 0 \\ \delta_{n+1} - \delta_n = W_{Y_{n-1}, Y_n}(n); s_n = (W_{Y_{n-1}, Y_n}(n), Y_n) \text{ for all } n \in \mathbb{Z} \end{cases}$$

Let $(\Omega', \mathcal{F}', \mathbf{P}^0)$ be the canonic representation of the process Δ (i.e., a point $\omega' \in \Omega'$ consists in a sample path of the process Δ , and \mathbf{P}^0 is the probability law of Δ). The operators θ'_t , $t \in \mathbb{R}$ are defined on $(\Omega', \mathcal{F}', \mathbf{P}^0)$ as follows: for any $\omega' \in \Omega'$, $t \in \mathbb{R}$, $\theta'_t \omega'$ is the sample path ω' shifted t time units to the left. It is easily seen that \mathbf{P}^0 is θ'_{δ_n} -invariant and that $\mathbf{P}^0[\delta_0 = 0] = 1$. Also, $\mathbf{E}^0[\delta_1] = w$ is non null and finite. Slivnyak's inverse construction can then be applied, yielding a θ'_t -invariant probability \mathbf{P}' on (Ω', \mathcal{F}') . Consider the $G/G/1/\infty$ queue with aggregated input process $A = \sum_{i=1}^K A_i$; under condition (3) its traffic intensity $\rho = \sum_{i=1}^K \rho_i$ is strictly less than 1. Loynes' theorem guarantees then existence of a stationary and ergodic workload process $\{M(t)\}$ for this queue. Let $(\Omega'', \mathcal{F}'', \mathbf{P}'')$ be the canonic representation of the two processes (A, M) ; for any $\omega'' \in \Omega''$, $t \in \mathbb{R}$, $\theta''_t \omega''$ is the sample path ω'' shifted t time units to the left. Stationarity of the processes (A, M) is exactly θ''_t -invariance of \mathbf{P}'' . Endow the product space

$$(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbf{P}}) = (\Omega' \times \Omega'', \mathcal{F}' \times \mathcal{F}'', \mathbf{P}' \times \mathbf{P}'')$$

with the shift $\tilde{\theta}_t$

$$\tilde{\theta}_t \tilde{\omega} = (\theta'_t \omega', \theta''_t \omega''), \text{ for all } t \in \mathbb{R}, \tilde{\omega} = (\omega', \omega'') \in \tilde{\Omega}$$

It is readily seen that $\tilde{\mathbf{P}}$ is $\tilde{\theta}_t$ -invariant.

Let $B = \{b_n\}$ (respectively $I = \{i_n\}$) denote the sequence of the beginnings of the server's busy (respectively idle) periods when the workload is M . Let us place ourselves under the Palm probability $\tilde{\mathbf{P}}_I^0$ of I . We assume that the points of I and B are numbered in the following way: $\forall n \in \mathbb{Z}, i_n \leq b_n < i_{n+1}$.

Remark 1 . $\tilde{\mathbf{P}}$ -independence between (A, M) and Δ implies that (A, M) and Δ are also $\tilde{\mathbf{P}}_I^0$ -independent, and the law of Δ under $\tilde{\mathbf{P}}_I^0$ is the same as under $\tilde{\mathbf{P}}$.

Construct now the marked point process $\Delta^0 = \{\delta_n^0, (S_n, U_n)\}$ as follows:

$$\begin{cases} \forall n \in \mathbb{Z}, \text{ no point } \delta_n^0 \text{ falls in } (b_n, i_{n+1}] \\ \forall n \geq 0, \text{ the restriction of } \Delta^0 \text{ to } (i_n, b_n] \text{ is the restriction of } \Delta \text{ to } (\sum_{j=0}^{n-1} b_j - i_j, \sum_{j=0}^n b_j - i_j] \\ \forall n < 0, \text{ the restriction of } \Delta^0 \text{ to } (i_n, b_n] \text{ is the restriction of } \Delta \text{ to } (-\sum_{j=n}^{-1} b_j - i_j, -\sum_{j=n+1}^{-1} b_j - i_j] \end{cases}$$

One might say that Δ^0 is constructed by filling the idle periods $(i_n, b_n]$ with Δ ; we note ϕ this operation:

$$\Delta^0 = \phi(M, \Delta)$$

Finally, let $(\Omega, \mathcal{F}, \mathbf{P}_I^0)$ be the canonic representation of the three processes (A, M, Δ^0) when the joint law of (A, M, Δ) is $\tilde{\mathbf{P}}_I^0$, and let θ_t be the operator on (Ω, \mathcal{F}) that shifts jointly the three processes (A, M, Δ^0) t time units to the left.

Proposition 1 . Under \mathbf{P}_I^0 , the processes (A, M, Δ^0) are jointly θ_{i_1} -stationary.

Proof: we note that:

$$\forall n \geq 1, \theta_{i_n} \Delta^0 = \phi(\theta_{i_n} M, \tilde{\theta}_{\sum_{j=0}^{n-1} b_j - i_j} \Delta) \quad (12)$$

A conditioning on M and Remark 1 show that:

$$\mathbf{P}_I^0(\theta_{i_1}(A, M, \Delta^0) \in \cdot) = \tilde{\mathbf{E}}_I^0 \int \mathbf{1}_{\{[\tilde{\theta}_{i_1} A, \tilde{\theta}_{i_1} M, \phi(\theta_{i_1} M, \tilde{\theta}_{i_1} \Delta)] \in \cdot\}} \tilde{\mathbf{P}}(\Delta \in d\delta)$$

Use of $\tilde{\theta}_i$ -stationarity of (A, M) under $\tilde{\mathbf{P}}_I^0$ then yields the conclusion.

Use of Slivnyak's inverse construction provides then a θ_t -invariant probability \mathbf{P} on (Ω, \mathcal{F}) . Note that, since for all $t \in \mathbb{R}$, $\theta_t(A, M) = \tilde{\theta}_t(A, M)$, the joint law of (A, M) under \mathbf{P} is the same as under $\tilde{\mathbf{P}}$. The points δ_n^0 of Δ^0 are the instants of departure in vacation of the server, and the first coordinate S_n of their mark (S_n, U_n) is the vacation's length, the vacation S_n being of type W_{U_{n-1}, U_n} . The amount of work present in the system at time t when vacations are interpreted as additional work is denoted $M^0(t)$ (see figure 1) and is computed according to the formula

$$M^0(t) = S_\nu + \left\{ \sum_{i=1}^K \sum_{n \in \mathbb{Z}} \mathbf{1}_{(\delta_n^0, t]}(T_n(i)) \cdot \sigma_n(i) \right\} - (t - \delta_\nu^0)$$

where $\nu = \sup\{n \in \mathbb{Z} \mid S_n \leq t\}$. The construction of M^0 is compatible with the shift θ_t , i.e.

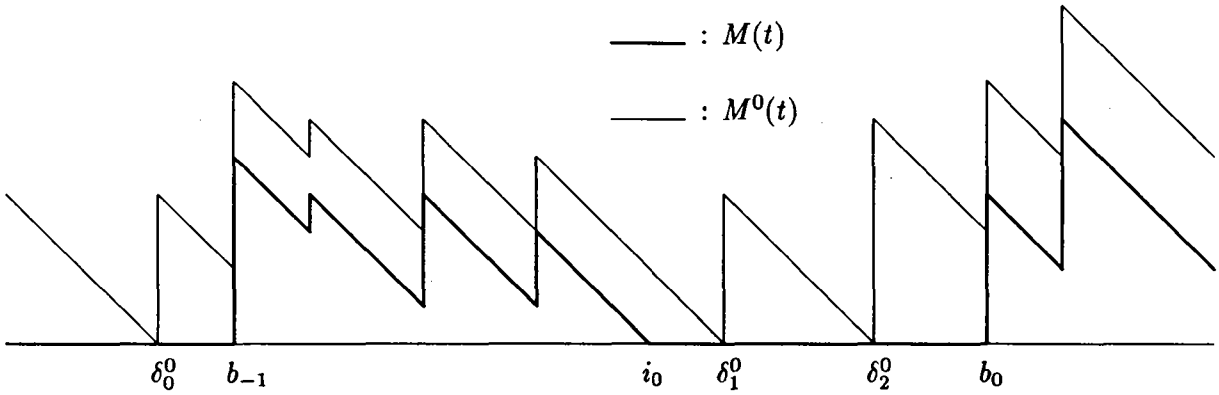


Figure 1: Construction of M^0

$$\theta_t M^0(s) = M^0(t + s), \quad t, s \in \mathbb{R}$$

Let Δ_i^0 count the points δ_m^0 such that $U_{m-1} = i$.

Lemma 1 . For all $i \in \{1, \dots, K\}$, the process Δ_i^0 has its intensity equal to $\frac{\pi_i(1-\rho)}{w}$; under the associated Palm probability $\mathbf{P}_{\Delta_i^0}^0$, the sequence $\{S_n, U_n\}_{n \geq 0}$ is independent of the processes (A, M) and of $(\delta_n^0, S_n, U_n)_{n < 0}$; $\{U_n\}_{n \geq -1}$ is a Markov chain on $\{1, \dots, K\}$ with transition matrix $(p_{k,l})$ started from i ; conditionally on $\{U_n\}_{n \geq 0}$ the sequence $\{S_n\}_{n \geq 0}$ is independent, S_n being distributed as W_{U_{n-1}, U_n} for all $n \geq 0$.

Proof: Use Neveu's exchange formula between I and Δ_i^0 to obtain, for bounded functionals f and g :

$$\begin{aligned} \lambda(\Delta_i^0) \cdot \mathbf{E}_{\Delta_i^0}^0 \{ f[A, M, (\delta_n^0, S_n, U_n)_{n < 0}] \cdot g[(S_n, U_n)_{n \geq 0}] \} = \\ \lambda(I) \mathbf{E}_I^0 \sum_{n \in \mathbb{Z}} \mathbf{1}_{(0, i_1]}(\delta_n^0) \mathbf{1}_{\{U_{n-1} = i\}} f[\theta_{\delta_n^0} A, \theta_{\delta_n^0} M, (\delta_m^0, S_m, U_m)_{m < n}] \cdot g[(S_m, U_m)_{m \geq n}] \end{aligned} \quad (13)$$

where $\lambda(N)$ denotes the intensity of the point process N . Recalling the construction of Δ^0 from Δ , the expression of the law of $\{S_n, U_n\}_{n \geq 0}$ under $\mathbf{P}_{\Delta_i^0}^0$ and its independence properties follow from this formula.

To check that $\lambda(\Delta_i^0) = \frac{\pi_i(1-\rho)}{w}$, take $f = g = 1$ in (13); this reads:

$$\lambda(\Delta_i^0) = \frac{1}{\mathbf{E}_I^0(i_1)} \mathbf{E}_I^0 \sum_{n \in \mathbf{Z}} 1_{(0, i_1]}(\delta_n^0) \cdot 1_{\{U_{n-1}=i\}}$$

A conditioning on b_0, i_1 yields:

$$\lambda(\Delta_i^0) = \frac{\pi_i}{w} \cdot \frac{\mathbf{E}_I^0(b_0 - i_0)}{\mathbf{E}_I^0(i_1)}$$

The proof is concluded by noting that $\frac{\mathbf{E}_I^0(b_0 - i_0)}{\mathbf{E}_I^0(i_1)} = 1 - \rho$, which is a classical queueing result (see [3] chapter 2).

The stationary congestion process $\{X^0(t)\} = \{X_1^0(t), \dots, X_K^0(t)\}$ of the number of customers present in the system is constructed as follows. At each point δ_n^0 , the system is empty; the number of customers present at any instant t of the ensuing vacation period $(\delta_n^0, \delta_n^0 + S_n]$ is then:

$$X^0(t) = A(\delta_n^0, t] \text{ for all } t \in (\delta_n^0, \delta_n^0 + S_n]$$

The congestion at queue i at $\delta_n^0 + S_n$ is thus

$$X_i^0(\delta_n^0 + S_n) = A_i(\delta_n^0, \delta_n^0 + S_n]$$

Let $U_n = i$; a server entering queue i at time $\delta_n^0 + S_n$ and finding $X_i^0(\delta_n^0 + S_n)$ customers waiting there, if it behaved as in the model described in Section 2, would serve the first $f_i[X_i^0(\delta_n^0 + S_n)]$ customers in the waiting line and then switch to another queue. We thus impose that the first $f_i[X_i^0(\delta_n^0 + S_n)]$ type i -customers present at $\delta_n^0 + S_n$ be served first; this is achieved at some time $\delta_{n+1}^1 \in [\delta_n^0 + S_n, \delta_{n+1}^0]$ (recall that the policies f_i are gated), and the number of customers present in the system at any time $t \in [\delta_n^0 + S_n, \delta_{n+1}^1]$ is well defined. The charge $X^0(t)$ in the remaining interval $(\delta_{n+1}^1, \delta_{n+1}^0]$ is defined as follows: after δ_{n+1}^1 , the server attends the customers in FIFO order, with no distinction between customers of different classes. Note that the resulting stationary point process $\{\delta_n^1\}$ might fail to satisfy the usual notational convention for point processes $T_0 \leq 0 < T_1$.

We describe informally the first step of the recursive construction before defining the rigorous framework. As was said, if the server served exactly according to the polling policy, the congestion process would be unchanged on $[\delta_n^0 + S_n, \delta_{n+1}^1]$ but the next vacation S_{n+1} would begin at δ_{n+1}^1 rather than at δ_{n+1}^0 : it would be advanced. In system 1, the vacation S_n is taken at δ_n^1 rather than at δ_n^0 , but system 0 and system 1 are in the same state at every δ_n^1 . The corresponding workload process M^1 has the following expression (see figure 2):

$$\begin{cases} \text{for all } n \in \mathbf{Z}, t \in [\delta_n^0, \delta_{n+1}^1), & M^1(t) = M^0(t) \\ \text{for all } n \in \mathbf{Z}, t \in [\delta_n^1, \delta_n^0), & M^1(t) = M^0(t) + S_n \end{cases}$$

The process Δ^1 counting vacations beginnings in system 1 is

$$\Delta^1 = \{\delta_n^1, (S_n, U_n)\}$$

It remains to build the congestion process X^1 . Set

$$X^1(\delta_n^1) = X^0(\delta_n^1) \text{ for all } n \in \mathbf{Z}$$

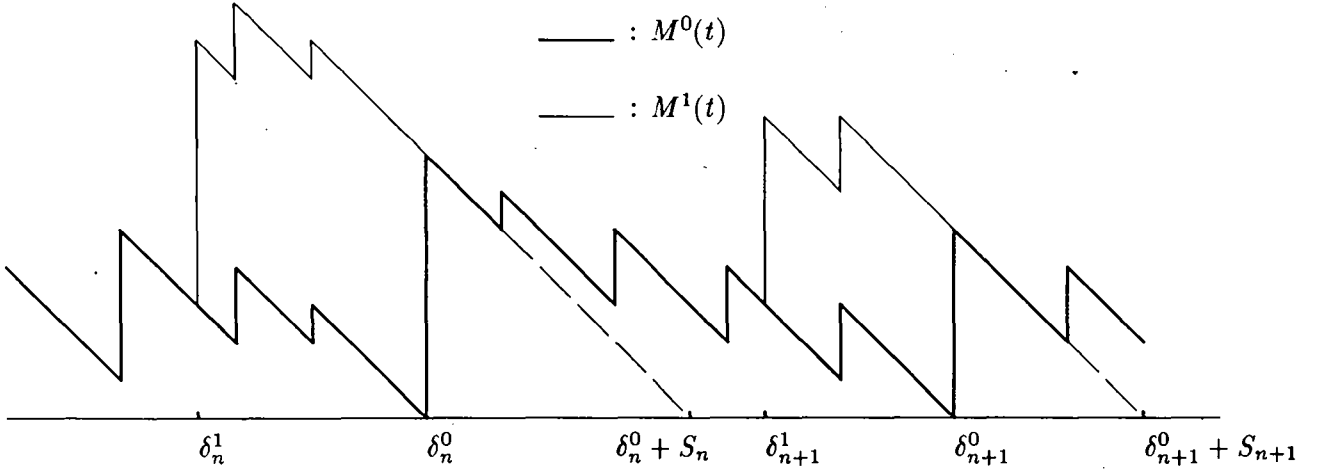


Figure 2: Construction of M^1

Since in system 1, the server starts vacation S_n at δ_n^1 , for any $t \in [\delta_n^1, \delta_n^1 + S_n]$ the congestion vector $X^1(t)$ equals $X^0(\delta_n^1) + A(\delta_n^1, t]$. Starting at δ_n^1 with $X^1(\delta_n^1) = X^0(\delta_n^1)$, in order that $X^1(\delta_{n+1}^1) = X^0(\delta_{n+1}^1)$, it is necessary that the same customers be served in system 0 and in system 1 during the cycle $(\delta_n^1, \delta_{n+1}^1]$. One also wants that system 1 approximates the original polling system as system 0 does, i.e. one wants the server in system 1, upon arrival at queue i , to attend first type i customers according to the service policy f_i .

Let $U_n = i$; if in system 0 the number of type i -customers attended in the period $(\delta_n^1, \delta_{n+1}^1]$ is greater than the image by f_i of the number of type i -customers present in system 1 at $\delta_n^1 + S_n$, i.e. is greater than

$$f_i\{X_i^0(\delta_n^1) + A_i(\delta_n^1, \delta_n^1 + S_n)\}$$

then the announced construction of X^1 on $(\delta_n^1 + S_n, \delta_{n+1}^1]$ is feasible: the first $f_i[X_i^1(\delta_n^1 + S_n)]$ type i -customers present at $\delta_n^1 + S_n$ in system 1 are attended first; this is achieved at $\delta_{n+1}^2 \in [\delta_n^1 + S_n, \delta_{n+1}^1]$. During $(\delta_{n+1}^2, \delta_{n+1}^1]$ the server attends those customers served in system 0 between $(\delta_n^1, \delta_{n+1}^1]$ that are not those $f_i[X_i^1(\delta_n^1 + S_n)]$ customers served during $[\delta_n^1 + S_n, \delta_{n+1}^2]$. Service in system 1 is non preemptive, FIFO in each class of customers separately, but with intricate priority rules between classes.

4.3 Implementation of the construction

system m : Let M^m be the stationary charge of the $G/G/1/\infty$ queue with arrival process A , and with vacations S_n occurring at points δ_n^m ; the vacation S_n is of type W_{U_{n-1}, U_n} . The marked point process $\Delta^m = \{\delta_n^m, (S_n, U_n)\}$ is stationary. A vacation occurs only after the end of the previous one, i.e. $\delta_{n+1}^m \geq \delta_n^m + S_n$. The congestion process X^m behaves as follows: at the return of a vacation S_n , the number of customers in the system is $X^m(\delta_n^m + S_n)$. If $U_n = i$, exactly $f_i[X_i^m(\delta_n^m + S_n)]$ type i -customers are served

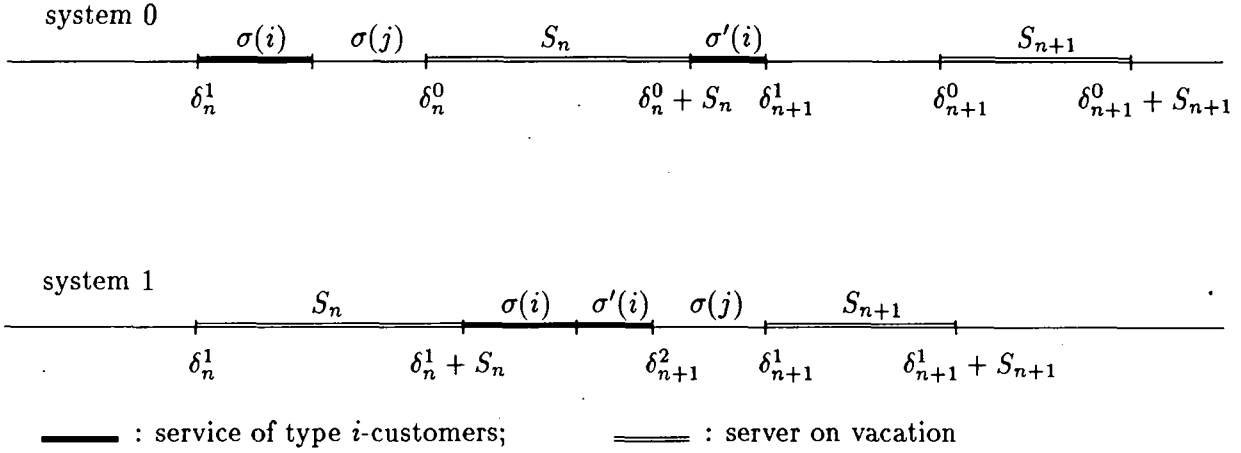


Figure 3: Construction of X^1

in FIFO order; this is achieved at $\delta_{n+1}^{m+1} \in [\delta_n^m + S_n, \delta_{n+1}^m]$. During $(\delta_{n+1}^{m+1}, \delta_{n+1}^m]$, non-preemptive service is given in FIFO order in each customer class, but with priority rules among classes. Also, vacations begin exactly upon completion of a customer's service (no preemption occurring there either).

Let D_i^m denote the point process that counts departures from the service booth of class i -customers in system m .

Lemma 2 . *If $U_n = i$, the number of class i -customers $D_i^m(\delta_n^{m+1}, \delta_{n+1}^{m+1}]$ served in system m during $(\delta_n^{m+1}, \delta_{n+1}^{m+1}]$ is greater than or equal to:*

$$f_i\{X_i^m(\delta_n^{m+1}) + A_i(\delta_n^{m+1}, \delta_n^{m+1} + S_n)\}$$

Proof: Note first the straightforward identity for all $t, s \in \mathbf{R}$

$$t \leq s \Rightarrow X_i^m(t) + A_i(t, s] = X_i^m(s) + D_i^m(t, s] \quad (14)$$

Taking $t = \delta_n^{m+1}$, $s = \delta_n^m + S_n$ one obtains

$$X_i^m(\delta_n^{m+1}) + A_i(\delta_n^{m+1}, \delta_n^m + S_n] = X_i^m(\delta_n^m + S_n) + D_i^m(\delta_n^{m+1}, \delta_n^m + S_n] \quad (15)$$

Since $\delta_n^{m+1} \leq \delta_n^m$, (15) implies

$$X_i^m(\delta_n^{m+1}) + A_i(\delta_n^{m+1}, \delta_n^{m+1} + S_n] \leq X_i^m(\delta_n^m + S_n) + D_i^m(\delta_n^{m+1}, \delta_n^m + S_n] \quad (16)$$

The policy f_i is non decreasing; applying it to both sides of (16) preserves the inequality there:

$$f_i\{X_i^m(\delta_n^{m+1}) + A_i(\delta_n^{m+1}, \delta_n^{m+1} + S_n]\} \leq f_i\{X_i^m(\delta_n^m + S_n) + D_i^m(\delta_n^{m+1}, \delta_n^m + S_n]\}$$

It follows from assumption (iii) made on f_i that for any two non negative integers x, y

$$f_i(x + y) \leq f_i(x) + y$$

This implies

$$f_i\{X_i^m(\delta_n^{m+1}) + A_i(\delta_n^{m+1}, \delta_n^{m+1} + S_n]\} \leq f_i\{X_i^m(\delta_n^m + S_n)\} + D_i^m(\delta_n^{m+1}, \delta_n^m + S_n]$$

But the congestion process X^m is such that, if $U_n = i$, the number $D_i^m(\delta_n^m + S_n, \delta_{n+1}^{m+1})$ is exactly equal to $f_i\{X_i^m(\delta_n^m + S_n)\}$ and the conclusion of the Lemma follows.

Define the workload process M^{m+1} by letting, for every $n \in \mathbf{Z}$:

$$\begin{cases} \text{for all } t \in [\delta_{n-1}^m, \delta_n^{m+1}), & M^{m+1}(t) = M^m(t) \\ \text{for all } t \in [\delta_n^{m+1}, \delta_n^m), & M^{m+1}(t) = M^m(t) + S_n \end{cases} \quad (17)$$

The marked point process Δ^{m+1} is:

$$\Delta^{m+1} = \{\delta_n^{m+1}, (S_n, U_n)\} \quad (18)$$

Construct the congestion process X^{m+1} as follows: for every $n \in \mathbf{Z}$, set $X^{m+1}(\delta_n^{m+1}) = X^m(\delta_n^{m+1})$. For $t \in (\delta_n^{m+1}, \delta_n^{m+1} + S_n]$ (the n^{th} vacation period in system $m+1$) let:

$$X^{m+1}(t) = X^{m+1}(\delta_n^{m+1}) + A(\delta_n^{m+1}, t]$$

The value of the congestion process at $t = \delta_n^{m+1} + S_n$ is then:

$$X^{m+1}(\delta_n^{m+1} + S_n) = X^m(\delta_n^{m+1}) + A(\delta_n^{m+1}, \delta_n^{m+1} + S_n]$$

According to Lemma 2, in system m , the number of type i -customers served during $(\delta_n^{m+1}, \delta_{n+1}^{m+1}]$ is greater than $f_i\{X_i^{m+1}(\delta_n^{m+1} + S_n)\}$; we impose then that after $\delta_n^{m+1} + S_n$, the server in system $m+1$ attends the first $f_i\{X_i^{m+1}(\delta_n^{m+1} + S_n)\}$ type i -customers that were present at $\delta_n^{m+1} + S_n$, and this will be achieved at a time $\delta_{n+1}^{m+2} \in [\delta_n^{m+1} + S_n, \delta_{n+1}^{m+1}]$. Finally, to define X^{m+1} on the remaining interval $[\delta_{n+1}^{m+2}, \delta_{n+1}^{m+1}]$, we let the server attend those customers served in system m during $(\delta_n^{m+1}, \delta_{n+1}^{m+1}]$ that are not the $f_i\{X_i^{m+1}(\delta_n^{m+1} + S_n)\}$ type i -customers it already took care of, and this is done in the same order as in system m . One easily checks that the induction hypotheses made on system m are satisfied by system $m+1$.

5 Existence of a stationary regime for the polling model

5.1 Sufficiency of condition (3) for the convergence of the scheme

In the sequel, it is assumed that the probability \mathbf{P} is ergodic for the shift $\{\theta_t\}$. A sufficient condition for this property to hold is provided by the following lemma, the proof of which is postponed to the appendix.

Lemma 3 . *Assume that for some i, j in $\{1, \dots, K\}$, the routing probability $p_{i,j}$ is non zero and the distribution of the corresponding switch-over times $W_{i,j}$ is spread-out. Then the stationary processes $\{A, M, \Delta^0\}$ are ergodic.*

Equation (17) implies that for every $t \in \mathbf{R}$, the sequence $M^m(t)$ increases with m ; the limiting process $M^\infty = \lim_{m \rightarrow \infty} M^m$ is thus well defined. In every system m , the server works at unit speed, so that:

$$M^m(0) \leq M^m(t) + t \text{ for all } m \in \mathbf{N}, t \in \mathbf{R}^+ \quad (19)$$

It thus follows for all $t \in \mathbf{R}^+$ that

$$\lim_{m \rightarrow \infty} M^m(0) = +\infty \Rightarrow \lim_{m \rightarrow \infty} M^m(t) = +\infty$$

That is to say, the event $\{\lim_{m \rightarrow \infty} M^m(0) = +\infty\}$ is θ_t -invariant; θ_t -ergodicity of \mathbf{P} implies then that $\mathbf{P}(\lim_{m \rightarrow \infty} M^m(0) = +\infty)$ equals 0 or 1.

Note R_i^m the process that counts, in system m , the returns of the server from vacations of type $W_{j,i}$ for some j :

$$R_i^m = \{r_n^m(i)\}_{n \in \mathbb{Z}} = \{\delta_n^m + S_n; U_n = i\}_{n \in \mathbb{Z}}$$

Lemma 4 . *for every $m \in \mathbb{N}$, every $i \in \{1, \dots, K\}$, R_i^m has its intensity equal to $\frac{\pi_i(1-\rho)}{w}$. Under the associated Palm probability the sequence $\{S_n, U_n\}$ of vacations and of their types that the server, in system m , shall experience on \mathbb{R}^+ , is independent of $\{M^m(t), X^m(t)\}_{t \leq 0}$ and its distribution does not depend on m ($\{U_n\}$ is a Markov chain with transition matrix $(p_{k,l})$ started from i , and conditionally on $\{U_n\}$ the sequence $\{S_n\}$ is independent with each S_n being distributed as W_{U_{n-1}, U_n}).*

The proof is done by induction on m , the passage from m to $m+1$ being made by Neveu's exchange formula (11) between the point processes R_i^m and R_i^{m+1} , Lemma 1 providing the seed $m=0$. Th detailed arguments will be omitted here.

We introduce the notation:

$$W^m(t) = \sum_{n \in \mathbb{Z}} \mathbf{1}_{\{\delta_n^m < t\}} (\delta_n^m + S_n - t)^+ \text{ for all } m \in \mathbb{N}, t \in \mathbb{R} \quad (20)$$

When the server is at work in system m , W^m equals 0, otherwise it represents the time it will take to finish its current vacation.

Corollary 1 . *The law of $W^m(0)$ under \mathbf{P} does not depend on m . This implies in particular that*

$$\begin{aligned} (i) \quad & \text{The family of random variables } \{W^m(0)\}_{m \in \mathbb{N}} \text{ is tight under } \mathbf{P} \\ & \text{and} \\ (ii) \quad & \mathbf{P}(W^m(0) > 0) = 1 - \rho \text{ for all } m \in \mathbb{N} \end{aligned} \quad (21)$$

Proof: The inversion formula (7) reads

$$\mathbf{P}(W^m(0) > s) = \frac{\pi_i(1-\rho)}{w} \mathbf{E}_{R_i^m}^0 \int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > s\}} dt \text{ for all } s \in \mathbb{R}^+, m \in \mathbb{N}$$

By definition of $r_1^m(i)$ and of $W^m(t)$, the right hand side of this expression depends only on the vacations and of their types $\{S_n, U_n\}$ that the server in system m will experience on \mathbb{R}^+ , the law of which, under $\mathbf{P}_{R_i^m}^0$, is independent of m . This shows that the law of $W^m(0)$ does not depend on m . Tightness of $\{W^m(0)\}_{m \in \mathbb{N}}$ follows, and this shows also that $\mathbf{P}(W^m(0) > 0)$ does not depend on m . Since $\mathbf{P}(W^0(0) > 0) = 1 - \rho$, (21) (ii) follows.

Assume now that $\mathbf{P}(\lim_{m \rightarrow \infty} M^m(0) = +\infty) = 1$; then $M^m(0)$ tends to $+\infty$ in probability:

$$\lim_{m \rightarrow \infty} \mathbf{P}(M^m(0) > s) = 1 \text{ for all } s > 0$$

Part (i) of Corollary 1 and the majorization

$$M^m(0) \leq \sum_{i=1}^K \sum_{n=0}^{X_i^m(0)} \sigma_{-n}(i) + W^m(0)$$

imply that $\sum_{i=1}^K X_i^m(0)$ tends also to $+\infty$ in probability, that is

$$\lim_{m \rightarrow \infty} \mathbf{P}(\sum_{i=1}^K X_i^m(0) > s) = 1 \text{ for all } s > 0 \quad (22)$$

This equation and part (ii) of (21) yield:

$$\lim_{m \rightarrow \infty} \mathbf{P}(\sum_{i=1}^K X_i^m(0) > s; W^m(0) > 0) = 1 - \rho \text{ for all } s > 0 \quad (23)$$

Use the elementary inequality $\mathbf{1}_{\{X+Y>x+y\}} \leq \mathbf{1}_{\{X>x\}} + \mathbf{1}_{\{Y>y\}}$ to deduce

$$\liminf_{m \rightarrow \infty} [\sum_{i=1}^K \mathbf{P}(X_i^m(0) > s; W^m(0) > 0)] \geq 1 - \rho \text{ for all } s > 0 \quad (24)$$

The following Lemma will ensure that the above inequality (24) cannot hold when condition (3) is in force.

Lemma 5 . For every $m \geq 0$, every $i \in \{1, \dots, K\}$, one has

$$\mathbf{P}_{R_i^m}^0(X_i^m(0) \geq x) \leq \frac{\lambda_i w}{\pi_i(1-\rho)f_i(x)} \text{ for all } x \in \mathbb{N} \quad (25)$$

and

$$\mathbf{P}(X_i^m(0) > x + y; W^m(0) > 0) \leq \frac{\lambda_i w}{\pi_i f_i(x)} + \epsilon_i(y) \text{ for all } x, y \in \mathbb{N} \quad (26)$$

where $\epsilon_i(y)$ does not depend on m , and tends to 0 as y tends to $+\infty$.

Proof: By construction, the number $D_i^m(r_0^m(i), r_1^m(i))$ of type i -customers served in system m during $(r_0^m(i), r_1^m(i))$ is greater than or equal to $f_i[X_i^m(r_0^m(i))]$; this and non-decreasingness of f_i imply:

$$\mathbf{P}_{R_i^m}^0(X_i^m(0) \geq x) \leq \frac{1}{f_i(x)} \mathbf{E}_{R_i^m}^0\{D_i^m(r_0^m(i), r_1^m(i))\}$$

But, since D_i^m has intensity λ_i , the term $\mathbf{E}_{R_i^m}^0\{D_i^m(r_0^m(i), r_1^m(i))\}$ is equal to $\frac{\lambda_i w}{\pi_i(1-\rho)}$, and (17) holds. Use the inversion (7) formula to obtain:

$$\mathbf{P}(X_i^m(0) \leq x + y; W^m(0) > 0) = \frac{\pi_i(1-\rho)}{w} \mathbf{E}_{R_i^m}^0 \int_0^{r_1^m(i)} \mathbf{1}_{\{X_i^m(t) \leq x+y\}} \mathbf{1}_{\{W^m(t) > 0\}} dt$$

Note that, for any $t \in (0, r_1^m(i))$,

$$X_i^m(t) \leq X_i^m(0) + A_i(0, r_1^m(i))$$

which yields

$$\mathbf{P}(X_i^m(0) \leq x + y; W^m(0) > 0) \geq \frac{\pi_i(1-\rho)}{w} \mathbf{E}_{R_i^m}^0 \{[1 - \mathbf{1}_{\{X_i^m(0) > x\}} - \mathbf{1}_{\{A_i(0, r_1^m(i)) > y\}}] \int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > 0\}} dt\} \quad (27)$$

According to Lemma 4, $\int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > 0\}} dt$ is $\mathbf{P}_{R_i^m}^0$ -independent of $X_i^m(0)$, and has mean value $\frac{w}{\pi_i}$. This and (25) imply:

$$\mathbf{P}(X_i^m(0) \leq x + y; W^m(0) > 0) \geq 1 - \rho - \frac{\lambda_i w}{\pi_i f_i(x)} - \frac{\pi_i(1-\rho)}{w} \mathbf{E}_{R_i^m}^0 \{ \mathbf{1}_{\{A_i(0, r_1^m(i)) > y\}} \int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > 0\}} dt \} \quad (28)$$

Denote by G_i the cumulative distribution function of $\int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > 0\}} dt$ under $\mathbf{P}_{R_i^m}^0$ (according to Lemma 4, G_i does not depend on m). Note also that, since $\mathbf{E}_{R_i^m}^0\{A_i(0, r_1^m(i))\} = \frac{\lambda_i w}{\pi_i(1-\rho)}$, the probability $\mathbf{P}_{R_i^m}^0(A_i(0, r_1^m(i)) > y)$ is less than $\frac{\lambda_i w}{\pi_i(1-\rho)y}$. It follows then from the extremal correlation property of two random variables with given marginals (see Whitt [13]) that:

$$\mathbf{E}_{R_i^m}^0\{\mathbf{1}_{\{A_i(0, r_1^m(i)) > y\}} \int_0^{r_1^m(i)} \mathbf{1}_{\{W^m(t) > 0\}} dt\} \leq \int_{1 - \frac{\lambda_i w}{\pi_i(1-\rho)y}}^1 G_i^{-1}(u) du$$

Finally, set

$$\epsilon_i(y) = \frac{\pi_i(1-\rho)}{w} \int_{1 - \frac{\lambda_i w}{\pi_i(1-\rho)y}}^1 G_i^{-1}(u) du$$

and check from (28) that (26) holds.

Theorem 1 . Under condition (3), that is if

$$\rho + \left(\sum_{i=1}^K \frac{\lambda_i}{\pi_i F_i}\right) w < 1$$

the limiting workload process $M^\infty = \lim_{m \rightarrow \infty} M^m$ is \mathbf{P} -almost surely finite.

Proof: If it were not so, $M^\infty(0)$ would be almost surely infinite, and (24) would hold. But according to Lemma 5, for any $\epsilon > 0$ and any $i \in \{1, \dots, K\}$, one can select an s_i sufficiently large to ensure:

$$\mathbf{P}(X_i^m(0) > s_i; W^m(0) > 0) \leq \frac{\lambda_i w}{\pi_i F_i} + \epsilon \text{ for all } m \in \mathbf{N}$$

If ϵ is chosen sufficiently small, (16) cannot hold with $s = \sup_i s_i$, and $M^\infty(0)$ must therefore be almost surely finite if condition (3) holds; θ_t -invariance of the event $\{M^\infty(0) = +\infty\}$ implies then that the whole process M^∞ is almost surely finite.

5.2 Identification of the stationary regime

Define for every $t \in \mathbf{R}$, every $m \in \mathbf{N}$:

$$\nu^m(t) = \max\{n \in \mathbf{Z}; \delta_n^m \leq t\}$$

Equation (17) implies the following equality:

$$M^m(t) = M^0(t) + \sum_{i=\nu^0(t)+1}^{\nu^m(t)} S_i \text{ for all } t \in \mathbf{R}, m \in \mathbf{N} \quad (29)$$

Since for all $t \in \mathbf{R}$, $M^m(t)$ increases to the almost surely finite limit $M^\infty(t)$ as $m \rightarrow \infty$, $\nu^m(t)$ increases also to an almost surely finite limit $\nu^\infty(t)$ as $m \rightarrow \infty$. Note that $\nu^\infty(t)$ is integer-valued, and increases with t ; the point δ_n^∞ such that $\nu^\infty[(\delta_n^\infty)^-] = n - 1$, and $\nu^\infty(\delta_n^\infty) = n$ is thus well defined, and for every $n \in \mathbf{Z}$, δ_n^m decreases to δ_n^∞ as $m \rightarrow \infty$. The sequence $\{\Delta^m\}$ converges then as $m \rightarrow \infty$ to a limiting process

$$\Delta^\infty = \{\delta_n^\infty, (S_n, U_n)\}$$

We now show that the process X^m also admits a proper limit as $m \rightarrow \infty$.

Let n be some fixed integer. Note that, on $[\delta_n^m, \delta_n^m + S_n)$, the processes X_i^n are greater than the X_i^{m-1} . For m large enough, it holds that:

$$\delta_n^m < \delta_n^\infty + S_n \leq \delta_n^m + S_n$$

Thus, for large m , the sequences $X_i^m(\delta_n^\infty + S_n)$ increase. Since $M^\infty(\delta_n^\infty + S_n)$ is finite, the $X_i^m(\delta_n^\infty + S_n)$ reach a finite limit $X_i^\infty(\delta_n^\infty + S_n)$ in finite (random) time m_0 . This implies in fact that, after m_0 , the sequence δ_n^m is constant and equal to δ_n^∞ : indeed, if after m_0 there was an m such that $\delta_n^{m+1} < \delta_n^m$, the customers attended in system m during $[\delta_n^{m+1}, \delta_n^m)$ would not have been served in system $m+1$ at time $\delta_n^\infty + S_n$, which would lead to a strictly positive difference between $X_i^{m+1}(\delta_n^\infty + S_n)$ and $X_i^m(\delta_n^\infty + S_n)$ for some i , a contradiction with m_0 's definition. Therefore, by construction, for any integer k , the processes $\{M^m, X^m, \Delta^m\}$ remain constant on the interval $[\delta_n^\infty + S_n, \delta_{n+k}^\infty + S_{n+k})$ for m larger than $m_0 + k$. The processes $\{M^m, X^m, \Delta^m\}$ converge thus to limiting processes $\{M^\infty, X^\infty, \Delta^\infty\}$, and these limits are reached in finite time on any bounded subset of \mathbf{R} .

Let us check now that the processes $(M^\infty, \Delta^\infty, X^\infty)$ describe a stationary regime for the polling system presented in Section 2. Since the processes (M^m, Δ^m, X^m) reach their limits in finite time on any bounded subset of \mathbf{R} , it is easily seen that the three limiting processes describe indeed the regime of a vacation system with arrival processes $\{A_i\}$ and vacations S_n occurring at times δ_n^∞ , where non-preemptive service is provided to every customer, the service being provided in FIFO order in each customer class separately. In order that the server in this limiting system follows the expected polling policy, it is enough that between arrival to a queue i and the next switch-over the server attends type i customers only with the number of type i customers served chosen according to policy f_i . Let $D_i^\infty, i \in \{1, \dots, K\}$ count departures of type i customers in the limiting system. It is then enough to check that

$$\forall i \in \{1, \dots, K\}, \forall n \in \mathbf{Z}, D_i^\infty(\delta_n^\infty + S_n, \delta_{n+1}^\infty) = \mathbf{1}_{\{U_n=i\}} \cdot f_i(X_i^\infty(\delta_n^\infty + S_n)) \quad (30)$$

This is the case indeed, as can be seen from the following computations for arbitrary $n \in \mathbf{Z}$

$$\begin{aligned} D_i^\infty(\delta_n^\infty + S_n, \delta_{n+1}^\infty) &= \lim_{m \rightarrow \infty} D_i^m(\delta_n^\infty + S_n, \delta_{n+1}^\infty) \\ &= \lim_{m \rightarrow \infty} D_i^m(\delta_n^m + S_n, \delta_{n+1}^{m+1}) \\ &= \lim_{m \rightarrow \infty} [f_i(X_i^m(\delta_n^m + S_n))] \mathbf{1}_{\{U_n=i\}} \\ &= \mathbf{1}_{\{U_n=i\}} \cdot f_i(X_i^\infty(\delta_n^\infty + S_n)) \end{aligned}$$

In the above, the third equality is a consequence of the service policy in system m , $m \geq 0$, while the limits can be taken in the first, second and fourth equalities because the processes X^m , D^m and Δ^m reach their limits in finite time on bounded sets.

In whole rigour, one should also check that the sequence $\{S_n, U_n\}$ of switch-over times and of routings in the limiting system satisfy the expected statistical hypotheses, and in particular that the server's switch-over times and routings after an arrival of the server to some queue i are independent of the past behaviour of the system. Calling R_i^∞ the point process that counts arrivals of the server to queue i in the limiting system, use of Lemma 4 and of Neveu's exchange formula between R_i^n and R_i^∞ provide an easy way to check that this is the case indeed; we omit the detailed arguments here.

6 Minimality of the stationary workload M^∞

Let us assume here that the conditions of Lemma 3 hold, i.e. there exists a couple (i, j) such that $p_{i,j}$ is non-zero and the distribution of the corresponding switch-over time $W_{i,j}$ is spread-out. In this section it will be shown that for any initial configuration a of the polling system, the corresponding transient

workload $M^{\infty,a}(t)$ is, in the asymptotic $t \rightarrow \infty$, stochastically larger than the stationary workload M^∞ in the sense that for any integer k and any t_1, \dots, t_k , any s_1, \dots, s_k one has:

$$\liminf_{t \rightarrow +\infty} \mathbf{P}(M^{\infty,a}(t+t_1) > s_1, \dots, M^{\infty,a}(t+t_k) > s_k) \geq \mathbf{P}(M^\infty(t_1) > s_1, \dots, M^\infty(t_k) > s_k) \quad (31)$$

The initial configuration a consists of the state of each queue (number of customers, corresponding remaining service times and order of arrival), the station where the server is or where it goes if it is currently switching, the number of customers it will serve before switching or the remaining switch-over time if it is currently switching.

Given this initial configuration, compute the corresponding initial workload $M^{0,a}(0)$, and build the workload $M^{0,a}(t)$ of the queueing system with initial workload $M^{0,a}(0)$, aggregated arrival process $A = \sum_i A_i$ and with multiple server's vacations $S_n = W_{U_{n-1}, U_n}(n)$ begining when the server empties the system or when it returns from a vacation to an empty queue (the value U_0 is the queue that the server attends or where to it is switching at time 0). Let $\Delta^{0,a} = \{\delta_n^{0,a}, (S_n, U_n)\}$ count the vacations beginings in this system.

Lemma 6 . *Under the assumptions of Lemma 3, the processes $\{M^{0,a}, \Delta^{0,a}\}$ couple with the stationary processes $\{M^0, \Delta^0\}$.*

Proof: Construct the workload M^a of the system with initial charge $M^{0,a}(0)$, with aggregated arrival process $A = \sum A_i$, and without server's vacations. Let T^a denote the smallest real t such that $M^a(t)$ is zero, and let T denote the smallest real t such that the stationary workload $M(t)$ of this system is zero. Since $\rho < 1$, both quantities T^a and T are almost surely finite, and it is easily seen that $T^a \vee T$ is a coupling time for the processes M^a and M . Construct now the processes $(M^{0,a}, \Delta^{0,a})$ and (M^0, Δ^0) by "filling" the idle periods of M^a and M with processes Δ^a and Δ respectively (cf. 4.2). Note:

$$\begin{cases} i(t) = \int_0^t \mathbf{1}_{\{M(s)=0\}} ds \\ i^a(t) = \int_0^t \mathbf{1}_{\{M^a(s)=0\}} ds \end{cases}$$

$i^a(t)$ (respectively $i(t)$) is the time that the server spent idling on $[0, t)$ when the workload is M^a (respectively M). Let $\theta_t^+ Z$ denote the restriction to \mathbf{R}^+ of some process Z shifted t time units to the left. We have, for $t \geq T \vee T^a$:

$$\begin{cases} \theta_t^+ \Delta^{0,a} = \phi(\theta_t^+ M, \theta_{i^a(t)}^+ \Delta^a) \\ \theta_t^+ \Delta^0 = \phi(\theta_t^+ M, \theta_{i(t)}^+ \Delta) \end{cases} \quad (32)$$

The processes $\theta_{i^a(T \vee T^a)} \Delta^a$ and $\theta_{i(T \vee T^a)} \Delta$ are both transient versions of the stationary process Δ , which is regenerative with respect to those points δ_n with $Y_n = i$, noted $\Delta_i = \{\delta_n(i)\}$, and, under the hypotheses of Lemma 3, the distribution of the length of the regeneration cycles $\mathbf{P}_{\Delta_i}^0(\delta_1(i) \in \cdot)$ is spread-out (see Corollary 4 in the Appendix). It follows then from Corollary 1.4, page 141 of Asmussen [2] that one can construct both processes $\theta_{i^a(T \vee T^a)} \Delta^a$ and $\theta_{i(T \vee T^a)} \Delta$ so that they couple in finite time T' . In that case, (32) and the definitions of $i(t)$, $i^a(t)$ imply that the processes $\Delta^{0,a}$ and Δ^0 are coupled after any t large enough to ensure

$$i(t) - i(T \vee T^a) \geq T'$$

and there exists such finite t with probability one since $i(t)$ tends to $+\infty$ as $t \rightarrow +\infty$, and T' is almost surely finite.

Corollary 2 . *For any initial configuration a of the polling system, the corresponding transient workload process $M^{\infty,a}$ is asymptotically stochastically greater than M^∞ , i.e. (31) holds.*

Proof: Starting from the processes $(M^{0,a}, \Delta^{0,a})$, one can apply the recursive construction of Section 4 and build a sequence of processes $(M^{m,a}, \Delta^{m,a}, X^{m,a})$ with the same initial configuration a in each system m . The processes $M^{m,a}$ are increasing, and a closer look at the construction shows that the workloads $M^{m,a}$ and $M^{\infty,a}$ coincide on $[0, \delta_{m+1}^{m+1,a})$; this implies that $M^{\infty,a}$ dominates every process $M^{m,a}$, and thus, for any integer k , any t_1, \dots, t_k , any s_1, \dots, s_k :

$$\mathbf{P}(M^{\infty,a}(t_1) > s_1, \dots, M^{\infty,a}(t_k) > s_k) \geq \mathbf{P}(M^{m,a}(t_1) > s_1, \dots, M^{m,a}(t_k) > s_k)$$

Let now N be some (random) integer such that $(M^{0,a}, \Delta^{0,a})$ and (M^0, Δ^0) are already coupled at δ_N^0 ; by construction, the processes $(M^{m,a}, \Delta^{m,a}, X^{m,a})$ and (M^m, Δ^m, X^m) will be coupled at time δ_{m+N}^m . Then, for all $m \geq 0$,

$$\liminf_{t \rightarrow \infty} \mathbf{P}(M^{\infty,a}(t + t_1) > s_1, \dots, M^{\infty,a}(t + t_k) > s_k) \geq \mathbf{P}(M^m(t_1) > s_1, \dots, M^m(t_k) > s_k) \quad (33)$$

The proof is completed by letting m tend to $+\infty$ in (33).

7 Concluding Remarks

7.1 Extension of the result to random policies

Consider the following service policy: for every $i \in \{1, \dots, K\}$, upon arrival at queue i , the server picks up a random number ξ that is uniformly distributed on $(0, 1)$ and independent of everything else. If X_i denotes the number of customers waiting at queue i at that time, the server will attend the first $f_i(X_i, \xi)$ type i -customers before switching to another queue, where f_i is some deterministic function from $\mathbf{N} \times (0, 1)$ into \mathbf{N} . Assume that f_i satisfies the following conditions for μ -almost every $u \in (0, 1)$

- (i') for all $x \in \mathbf{N}$, $f_i(x, u) \leq x$
- (ii') for all $x, y \in \mathbf{N}$, $x \leq y \Rightarrow f_i(x, u) \leq f_i(y, u)$
- (iii') for all $x, y \in \mathbf{N}$, $x \leq y \Rightarrow x - f_i(x, u) \leq y - f_i(y, u)$

where μ is the Lebesgue measure on $(0, 1)$. For every $i \in \{1, \dots, K\}$, let

$$F_i = \lim_{x \rightarrow \infty} \int_0^1 f_i(x, u) du$$

The reasonings of Sections 4 and 5 are easily adapted to check that a polling system with such random policies is stable under condition (3): add to the marks (S_n, U_n) of the points $\{\delta_n^0\}$ a third coordinate ξ_n such that the $\{\xi_n\}$ are i.i.d., uniform on $(0, 1)$ and independent of everything else. In system m , the point $\delta_{n+1}^{m+1} \in (\delta_n^m, \delta_{n+1}^m]$ is the time at which the first $f_i[X_i^m(\delta_n^m + S_n), \xi_n]$ type i -customers present at $\delta_n^m + S_n$ are served ($U_n = i$), and the mark of δ_n^m is (S_n, U_n, ξ_n) . The proofs in Section 5 translate to that framework without difficulty. A typical example of a random policy satisfying (i') – (iii') is the binomial-gated policy introduced by Levy (see [9]): upon arrival at queue i , if there are X customers waiting there, the number of customers that will be served during this cycle is distributed according to the binomial distribution with parameters p and X , where $p \in (0, 1]$ is a given parameter. This policy fits the framework just defined: indeed, let:

$$f_i(x, \xi) = y \mid \sum_{j=0}^{y-1} C_x^j p^j (1-p)^{x-j} \leq \xi < \sum_{j=0}^y C_x^j p^j (1-p)^{x-j}$$

and classical properties of the binomial distribution imply that f_i satisfies (ii'), (iii').

7.2 A necessary condition for stability

Assume that the polling system described in Section 2 admits a stationary regime. Let $\Delta = \{\delta_n, (S_n, U_n)\}$ count the server's departures from queues: at δ_n , the server leaves queue U_{n-1} and a switch-over time S_n occurs before queue U_n is reached. Let D_i count the departures of type i -customers from the ticket booth. Stability of the system implies that D_i has intensity λ_i . Let $\sigma_i(i)$ denote the service time of the customer leaving the system at some point t of D_i ; use the identity

$$\delta_1 = S_0 + \sum_{i=1}^K \int_{(0, \delta_1]} \sigma_i(i) D_i(dt)$$

to check that the intensity of Δ equals $\frac{1-\rho}{w}$. The process Δ_i :

$$\Delta_i = \{\delta_n(i)\} = \{\delta_n; U_n = i\}$$

is of intensity $\frac{\pi_i(1-\rho)}{w}$; this implies that:

$$\mathbf{E}_{\Delta_i}^0 D_i(0, \delta_1(i)) = \frac{\lambda_i w}{\pi_i(1-\rho)}$$

But the model is so that:

$$D_i(0, \delta_1(i)) = f_i(X_i(S_0)) \leq F_i$$

And thus, if the system is stable, necessarily:

$$\forall i \in \{1, \dots, K\}, \rho + \frac{\lambda_i w}{\pi_i F_i} \leq 1 \quad (34)$$

In view of the necessary and sufficient conditions (1) and (2), one is induced to believe that, for the model considered, a necessary and sufficient condition for stability would be:

$$\forall i \in \{1, \dots, K\}, \rho + \frac{\lambda_i w}{\pi_i F_i} < 1 \quad (35)$$

(we carefully avoid to mention the critical case $\rho + \frac{\lambda_i w}{\pi_i F_i} = 1$). There might be a way to prove that (35) is indeed sufficient for stability, using the results of Sections 4 and 5: remark first that, if only one queue (say, queue 1) is attended according to a limited policy (i.e., $F_i = +\infty$ for $i \neq 1$), conditions (3) and (35) coincide. Remark now that, to start the recursive construction in 4.2, it is only needed that, between $\delta_n^0 + S_n$ and δ_{n+1}^0 , if $U_n = i$, the number of type i -customers attended is greater than $f_i(X_i(\delta_n^0 + S_n))$. One can thus try to construct a stationary regime under condition (35) by induction on the number of queues with limited service policy, starting the recursion with the stationary regime of the polling system where queue 1 is attended according to policy f_1 , and every other queue is served according to the purely gated policy, then replacing the purely gated policy at queue 2 by policy f_2 , and so on.

8 Appendix

This part is devoted to the proof of Lemma 3; the result is established in a slightly more general framework, which will in fact allow an improved clarity in the arguments.

Let $\{A, N, Y\}$ be some jointly stationary processes defined on $(\Omega, \mathcal{F}, \mathbf{P})$, that are compatible with the

shift $\{\tilde{\theta}_t\}$ defined on (Ω, \mathcal{F}) . It is assumed that \mathbf{P} is ergodic for the shift $\{\tilde{\theta}_t\}$. The marked point process $N = \{T_n, \sigma_n\}$ has finite, non null intensity $\lambda(N)$. The marks σ_n belong to \mathbb{R}^+ and satisfy for all $n \in \mathbb{Z}$

$$\mathbf{P}(T_n + \sigma_n \leq T_{n+1}) = 1$$

It is also assumed that $\mathbf{E}_N^0(\sigma_0)$ is strictly positive. The process Y is \mathbf{P} -independent of $\{A, N\}$, and is regenerative with respect to some stationary point process $R = \{r_n\}$, that has a finite, non-zero intensity $\lambda(R)$. Define under \mathbf{P}_N^0 the process Y^0 by letting:

$$\begin{cases} \forall n \in \mathbb{Z}, Y^0 \equiv 0 \text{ on } [T_n + \sigma_n, T_{n+1}) \\ \forall n \geq 0, \text{ the restriction of } Y^0 \text{ to } [T_n, T_n + \sigma_n) \text{ is the restriction of } Y \text{ to } [\sum_{j=0}^{n-1} \sigma_j, \sum_{j=0}^n \sigma_j) \\ \forall n < 0, \text{ the restriction of } Y^0 \text{ to } [T_n, T_n + \sigma_n) \text{ is the restriction of } Y \text{ to } [-\sum_{j=n}^{-1} \sigma_j, -\sum_{j=n+1}^{-1} \sigma_j) \end{cases}$$

The process Y^0 is constructed by "filling" the intervals $[T_n, T_n + \sigma_n)$ with the process Y , which is noted:

$$Y^0 = \phi(N, Y)$$

On the canonic representation of the three processes (A, N, Y^0) when the underlying probability is \mathbf{P}_N^0 , define the operators θ_t , $t \in \mathbb{R}$ that shift jointly the three processes (A, N, Y^0) t time units to the left (by a slight abuse of notation, we shall still note $(\Omega, \mathcal{F}, \mathbf{P}_N^0)$ this canonic representation). One can note that $\tilde{\theta}_t(A, N) = \theta_t(A, N)$ for all $t \in \mathbb{R}$. Remark that:

$$\theta_{T_n} Y^0 = \phi(\theta_{T_n} N, \tilde{\theta}_{\sum_{j=0}^{n-1} \sigma_j} Y) \text{ for all } n \geq 1 \quad (36)$$

The same arguments as in the proof of Proposition 1 apply to show that the processes $\{A, N, Y^0\}$ are θ_{T_1} -stationary under \mathbf{P}_N^0 .

Theorem 2 . *If the distribution $\mathbf{P}_R^0(r_1 \in \cdot)$ of the length of Y 's regeneration cycles is spread-out, then the probability \mathbf{P}_N^0 is ergodic for the shift θ_{T_1} when restricted to the σ -field generated by $\{A, N, Y^0\}$.*

Proof: Let $\{(A, N, Y^0) \in C\}$ be some θ_{T_1} -invariant event. That is to say:

$$\{\omega : (A, N, Y^0) \in C\} = \{\omega : \theta_{T_n}(A, N, Y^0) \in C\} \text{ for all } n \in \mathbb{N}$$

This implies

$$\mathbf{P}_N^0[(A, N, Y^0) \in C] = \mathbf{P}_N^0\left[\bigcup_{m \geq 0} \bigcap_{n \geq m} \{\theta_{T_n}(A, N, Y^0) \in C\}\right] \quad (37)$$

and:

$$\forall k, l, m, k \leq l, \mathbf{P}_N^0[(A, N, Y^0) \in C] = \mathbf{P}_N^0\left[\bigcap_{k \leq j \leq l} \{\theta_{T_j}(A, N, Y^0) \in C\} \bigcap \bigcap_{n \geq m} \{\theta_{T_n}(A, N, Y^0) \in C\}\right] \quad (38)$$

Define on (Ω, \mathcal{F}) the filtration $\{\mathcal{F}_t\}$ by:

$$\mathcal{F}_t = \sigma\{(A, N, Y^0) |_{[-t, +t]}\}$$

The families (g_t, h_t) defined by:

$$\begin{cases} g_t = \mathbf{E}_N^0[1_{\{\bigcap_{k \leq j \leq t} \theta_{T_j}(A, N, Y^0) \in C\}} | \mathcal{F}_t] \\ h_t = \mathbf{E}_N^0[1_{\{\bigcap_{n \geq 0} \theta_{T_n}(A, N, Y^0) \in C\}} | \mathcal{F}_t] \end{cases}$$

are uniformly integrable $\{\mathcal{F}_t\}$ -martingales, and thus converge in L^1 -norm to their respective limits $1_{\{\bigcap_{k \leq j \leq l} \theta_{T_j}(A, N, Y^0) \in C\}}$ and $1_{\{\bigcap_{n \geq 0} \theta_{T_n}(A, N, Y^0) \in C\}}$ (this is a standard result from Martingale Theory; see e.g. Neveu [10]). Let $\epsilon > 0$ be fixed. We can then choose a $K > 0$ such that:

$$\begin{cases} \mathbf{E}_N^0 | g_K - 1_{\{\bigcap_{k \leq j \leq l} \theta_{T_j}(A, N, Y^0) \in C\}} | \leq \epsilon \\ \mathbf{E}_N^0 | h_K - 1_{\{\bigcap_{n \geq 0} \theta_{T_n}(A, N, Y^0) \in C\}} | \leq \epsilon \end{cases} \quad (39)$$

and g_K, h_K depend on the behaviour of (A, N, Y^0) only on the interval $[-K, K]$. Noting that the random variables g_K, h_K take their values in $[0, 1]$, it is easily seen that the right-hand-side of (38) equals, within an error e_1 of magnitude less than 2ϵ :

$$\mathbf{E}_N^0[g_K(A, N, Y^0)\theta_{T_m}h_K(A, N, Y^0)]$$

that we note, making use of (36):

$$\mathbf{E}_N^0[g'(A, N, Y)h'(\theta_{T_m}A, \tilde{\theta}_{T_m}N, \theta_{\sum_{j=0}^{m-1} \sigma_j}Y)] \quad (40)$$

By construction of Y^0 , the functionals g' and h' depend on the behaviour of their last argument only on $[-K, +K]$. Let q be some fixed integer. The term (40) is equal to

$$\mathbf{E}_N^0[g'(A, N, Y)h'(\theta_{T_m}A, \theta_{T_m}N, \theta_{\sum_{j=0}^{m-1} \sigma_j}Y)1_{\{K < r_q\}}1_{\{\sum_{j=0}^{m-1} \sigma_j \geq K+r_q\}}] \quad (41)$$

within an error e_2 of magnitude less than

$$1 - \mathbf{P}_N^0(\{K < r_q\} \cap \{\sum_{j=0}^{m-1} \sigma_j \geq K + r_q\})$$

To evaluate the term (41), condition on A, N first; then, since Y is time stationary and independent of A, N under \mathbf{P}_N^0 , apply the inversion formula (7) for the point process $R = \{r_n\}$, and use the fact that the r_n are regeneration points for Y to check that (41) equals:

$$\begin{aligned} \mathbf{E}_N^0[\lambda(R)\mathbf{E}_R^0 \{ \int_{r_q}^{r_q+1} g'(A, N, \tilde{\theta}_s Y) 1_{\{K < -s\}} 1_{\{\sum_{j=0}^{m-1} \sigma_j \geq K-s\}} \\ \times \mathbf{E}_R^0(h'(\theta_{T_m}A, \theta_{T_m}N, \tilde{\theta}_{\sum_{j=0}^{m-1} \sigma_j + s}Y | A, N).ds | A, N) \}] \end{aligned} \quad (42)$$

But, by hypothesis, the distribution $\mathbf{P}_R^0(r_1 \in \cdot)$ of the length of Y 's regeneration cycles is spread-out. It follows then from Corollary 1.4, page 141 of Asmussen [2] that the law of $\tilde{\theta}_t Y$ under \mathbf{P}_R^0 converges in variation to the law of Y under \mathbf{P}_N^0 as $t \rightarrow \infty$. The term (42) is thus equal to

$$\mathbf{E}_N^0[\mathbf{E}_N^0(g'(A, N, Y)1_{\{K < r_q\}}1_{\{\sum_{j=0}^{m-1} \sigma_j \geq K+r_q\}} | A, N). \mathbf{E}_N^0(h'(\theta_{T_m}A, \theta_{T_m}N, Y) | A, N)] \quad (43)$$

within an error e_3 that tends to 0 as $m \rightarrow \infty$ (the ergodic theorem implies that $\sum_{j=0}^{m-1} \sigma_j$ tends to infinity as $m \rightarrow \infty$). Use time stationarity of Y and its independence from A, N under \mathbf{P}_N^0 to replace Y , in the argument of h' , by $\tilde{\theta}_{\sum_{j=0}^{m-1} \sigma_j} Y$; thus (43) equals:

$$\mathbf{E}_N^0[\mathbf{E}_N^0(g_K(A, N, Y^0)1_{\{K < r_q\}}1_{\{\sum_{j=0}^{m-1} \sigma_j \geq K+r_q\}} | A, N). \mathbf{E}_N^0(\theta_{T_m}h_K(A, N, Y^0) | A, N)] \quad (44)$$

It follows from (39) that (44) equals

$$\mathbf{E}_N^0[\mathbf{P}_N^0(\bigcap_{k \leq j \leq l} \theta_{T_j}(A, N, Y^0) \in C \mid A, N) \mathbf{P}_N^0(\bigcap_{n \geq m} \theta_{T_n}(A, N, Y^0) \in C \mid A, N)] \quad (45)$$

within an error of magnitude less than

$$2\epsilon + 1 - \mathbf{P}_N^0(\{K < r_q\} \cap \{\sum_{j=0}^m \sigma_j \geq K + r_q\})$$

Letting successively m , then q tend to $+\infty$, the errors e_3 , then e_2 vanish; since ϵ is arbitrary, one obtains:

$$\mathbf{P}_N^0[(A, N, Y^0) \in C] = \mathbf{E}_N^0[\mathbf{P}_N^0(\bigcap_{k \leq j \leq l} \{\theta_{T_j}(A, N, Y^0) \in C\} \mid A, N) \cdot \mathbf{P}_N^0(\bigcup_{m \geq 0} \bigcap_{n \geq m} \{\theta_{T_n}(A, N, Y^0) \in C\} \mid A, N)] \quad (46)$$

Let now successively l , then k go to infinity, to establish:

$$\mathbf{P}_N^0[(A, N, Y^0) \in C] = \mathbf{E}_N^0[(\mathbf{P}_N^0(\bigcup_{m \geq 0} \bigcap_{n \geq m} \{\theta_{T_n}(A, N, Y^0) \in C\} \mid A, N))^2]$$

This equality and (37) show that the $\sigma(A, N)$ -measurable random variable

$$\mathbf{P}_N^0(\bigcup_{m \geq 0} \bigcap_{n \geq m} \{\theta_{T_n}(A, N, Y^0) \in C\} \mid A, N)$$

is almost surely $\{0, 1\}$ -valued; it is also θ_{T_1} -invariant and is therefore almost surely constant because \mathbf{P}_N^0 is ergodic for the shift θ_{T_1} when restricted to $\sigma(A, N)$, so that:

$$\mathbf{P}_N^0[(A, N, Y^0) \in C] \in \{0, 1\}$$

Ergodicity of (A, N, Y^0) is proven.

Corollary 3 . *the result of Lemma 3 holds.*

Proof: In order to apply Theorem 2 to the processes (A, I, Δ^0) , we only have to check that Δ is regenerative, with spread-out cycle length distribution. Take as regeneration points those δ_n with $Y_n = i$, noted $\{\delta_n(i)\} = \Delta_i$. The distribution $\mathbf{P}_{\Delta_i}^0(\delta_1(i) \in \cdot)$ of the regeneration cycles satisfies:

$$\mathbf{P}_{\Delta_i}^0(\delta_1(i) \in dt) \geq p_{i,j} \cdot \mathbf{P}_{\Delta_i}^0(\delta_1(i) \in dt \mid Y_1 = j) \quad (47)$$

The distribution appearing in the right-hand-side of (47) is the convolution of $W_{i,j}$'s distribution and of some other probability distribution, and is thus spread-out when $W_{i,j}$'s distribution is so; the distribution of the cycle length dominates a spread-out distribution, and is also spread-out.

References

- [1] Altman E., Konstantopoulos P., Liu Z., "Stability, Monotonicity and Invariant Quantities in General Polling Systems", Report INRIA Centre Sophia Antipolis, France (1991).
- [2] Asmussen S., "Applied Probability and Queues", John Wiley & Sons (1987).
- [3] Baccelli F., Brémaud P., "Elements of Queueing Theory", Springer-Verlag (1993).

- [4] Borovkov A. A., Schassberger R., "Ergodicity of Queueing Systems", Oral Communication, Conference on Applied Probability in Engineering, Computer and Communication Sciences, Paris (1993).
- [5] Brandt A., Franken P., Lisek B., "Stationary Stochastic Models", John Wiley & Sons (1992).
- [6] Doshi B., "Generalizations of the Stochastic Decomposition Results for Single Server Queues with Vacations", Commun. Statist. -Stochastic Models, 6 (2), 307-333 (1990).
- [7] Fricker C., Jaibi M. R., "Monotonicity and Stability of Periodic Polling Models", Research Memorandum FEW 559 Tilburg University, The Netherlands (1992).
- [8] Georgiadis L., Szpankowski W., "Stability of Token Passing Rings", Queueing Systems 11, 7-33 (1992).
- [9] Levy H., "Analysis of Cyclic Polling Systems with Binomial-Gated Service", Performance of Distributed and Parallel Systems, Hasewaga T., Takagi H., Takahashi Y. (Eds), Elsevier (1989).
- [10] Neveu J., "Mathematical Foundations of the Calculus of Probability", Holden-Day, San Francisco, London, Amsterdam (1965).
- [11] Neveu J., "Sur les mesures de Palm de deux processus ponctuels stationnaires", Zeitschrift fur Wahrsch., 34, 199-203 (1976).
- [12] Takagi H., "Queueing Analysis of Polling Models: an Update", Stochastic Analysis of Computer and Communications Systems, Hideaki Takagi (Editor), Elsevier Science Pub., 267-318 (1990).
- [13] Whitt W., "Bivariate Distributions with Given Marginals", Ann. Stat. 4, 1280-1289 (1976).



Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399



★ R R - 2 1 4 8 ★