Admission Control in Stochastic Event Graphs

Eitan Altman, Bruno Gaujal, Arie Hordijk

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Admission Control in Stochastic Event Graphs

Eitan ALTMAN Bruno GAUJAL and Arie HORDIJK

N° 3179

Juin, 1997

THÈME 1

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Admission Control in Stochastic Event Graphs

Eitan ALTMAN [*] Bruno GAUJAL [**] and Arie HORDIJK [***]

Thème 1 — Réseaux et systèmes
Projet Mistral, Sloop

Rapport de recherche n ° 3179 — Juin, 1997 — 33 pages

Abstract: In this paper, we show that the expected workload and the expected waiting time in (max,+) linear system under a single input sequence is multimodular. We use this result to construct the optimal deterministic admission control in the (max,+) system under rate constraints.

Key-words: Multimodularity, (max,+) algebra, event graphs, optimal control

(Résumé : tsvp)

[*] INRIA, BP 93, 2004 Route des Lucioles, 06902 Sophia Antipolis Cedex, France. E-mail: altman@sophia.inria.fr. URL:http://www.inria.fr:80/mistral/personnel/Eitan.Altman/me.html

[**] INRIA, BP 93, 2004 Route des Lucioles, 06902 Sophia Antipolis Cedex, France. E-mail: gaujal@sophia.inria.fr. Bruno Gaujal is a member of a common project between CNRS, UNSA and INRIA.

[***] Dept. of Mathematics and Computer Science, Leiden University, P.O.Box 9512, 2300RA Leiden, The Netherlands. E-mail: hordijk@wi.leidenuniv.nl. The research of Arie Hordijk was done while he was on sabbatical leave at INRIA, Sophia-Antipolis; it has been partially supported by the Ministère Français de l'Éducation Nationale et de l'Enseignement Supérieur et de la Recherche.

Contrôle d'admission dans les graphes d'événements stochastiques

Résumé : Dans cet article, nous montrons que la charge moyenne ainsi que le temps d'attente moyen dans un système (max,+) linéaire avec une seule entrée est multimodulaire. Nous utilisons ensuite ce résultat pour construire un contrôle d'admission déterministe optimal sous des contraintes de taux d'admission, dans un système (max,+) quelconque.

Mots-clé : Multimodularité, algèbre (max,+), graphe d'événement, contrôle optimal

1 Introduction

In this paper, we address the following problem. Customers arrive at a discrete event system. A fraction of at least p of these customers must be admitted. The admission is governed by a deterministic binary sequence a_k (the k-th customer is admitted if and only if a_k is equal to one; otherwise it is rejected). The questions we address are to select the deterministic sequence that minimizes (i) the expected average of any convex increasing function of the waiting time of the customers in the system, (i) the expected average of any convex increasing function of the workload in the system, at some particular times.

In [9], Hajek proved multimodularity with respect to the admission sequence for the average number of customers in a single GI/M/1 queue. Here, we consider a network of queues which forms a stochastic event graph under a general arrival process and with stationary sequence of service times for all servers of the system (i.i.d. assumptions are not required). The central point of the paper is obtain some multimodularity properties. We then rely on the theoretical basis established in [3], to obtain the optimal control.

The rest of the paper is structured as follows. In section 2, we present an overview of the model that will be studied, that is (max,+) linear systems and we give the basic evolution equation of such a system. Section 3 introduces the definitions and section 4 studies the particular case of a FIFO queue, to give an idea of the proof in the general case. The main result of section 5 is to show that the workload in a (max,+) linear system is multimodular with respect to the arrival sequence. In section 6, we give the proof of multimodularity for the traveling time of a customer in the system. Finally, in section 7, we show that a balanced admission policy for is optimal among all deterministic policies with a given admission rate by applying the results established in [3].

2 Discrete event systems and the (Max,+) algebra.

In this section, we will describe the class of networks called *stochastic event graphs*, often called (max,+) linear systems since their dynamics follows a vectorial evolution equation in the so-called (max,+) algebra. *event graphs* form a subset of the Petri nets [8] in which each place has at most one input transition and one output transition.

Before we give a more detailed introduction to (max,+) systems in section 2.1, we first give as an example, the most studied queueing model, the G/G/1 queue. If a G/G/1 queue is represented as a Petri net, then one gets the net given in Figure 1. This Petri net is an event graph (each place has one input transition and one output transition). Therefore, its dynamic can be written as a (max,+) linear system. An extensive study on (max,+) systems can be found in [4]. Here, we give a brief introduction, that will serve our means in the following.

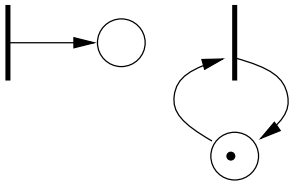

Figure 1: Event graph model of a G/G/1 queue

2.1 Dynamics of (max,+) linear systems.

More formally, an event graph is a Petri net $\mathcal{G} = (\mathcal{P}, \mathcal{Q}, \mathcal{M}_0)$, where $\mathcal{P} = \{p_1, \cdots, p_P\}$ is the set of places, $\mathcal{Q} = \{q_1, \cdots, q_Q\}$ is the set of transitions (also called *nodes* for event graphs), and $\mathcal{M}_0 : \mathcal{P} \to \{0, 1, 2, \cdots M\}$ is the initial number of tokens in each place.

An event graph is a Petri net where each place has one input transition and one output transition. In a stochastic event graph, the n-th *firing* time of transition q_i in $\mathcal{Q}$ is assumed to be a random variable $\sigma_i(n)$ with finite expectation.

A more detailed presentation of (stochastic) event graphs and Petri nets is available in numerous references, like for example [8, 4].

To any stochastic event graph, we can associate a set of matrices, $A_0(n), \cdots, A_M(n)$, all of size $Q \times Q$, where the entry (i, j) in matrix $A_k(n)$ is $\sigma_j(n-k)$ if there exists a place between transitions q_j and q_i with k initial tokens, and $-\infty$ otherwise.

If an event graph contains transitions with no input places (also called an *input* transition), we define the matrices B which gives the connections of the regular transitions with the input. More precisely, the entry i, j in B is 0 if there is a place between the input transition q_j and transition q_i, and is $-\infty$ otherwise.

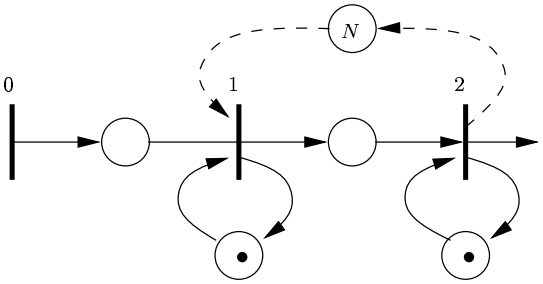

Figure 2: Event graph model of two queues in tandem

The example of two queues in tandem is given in Figure 2. If the dashed arcs are taken into account, the system corresponds to two queues in tandem with a second queue with finite buffer of size N, and blocking before service in the first queue when the second queue is full.

In this example (with the finite capacity $N = 1$), let us call $\sigma_i(n)$ the n-th firing time of transition i and $U(n)$ the sequence of firing times of the single input transition 0, which represents the input.

Then we get

$$A_0(n) = \begin{pmatrix} -\infty & -\infty \\ \sigma_1(n) & -\infty \end{pmatrix}, A_1(n) = \begin{pmatrix} \sigma_1(n-1) & \sigma_2(n-1) \\ -\infty & \sigma_2(n-1) \end{pmatrix}.$$

$$B = \begin{pmatrix} 0 \\ -\infty \end{pmatrix},$$

Now, in the general case, if we consider the state variables, $X_i(n)$, which denote the time when transition i initiates its n-th firing, then the vector $X(n) = (X_1(n), \cdots, X_Q(n))$ satisfies the following linear equation in the (max,+) notation, [4].

$$X(n) = A_0(n) \otimes X(n) \oplus A_1(n) \otimes X(n-1) \otimes \cdots \otimes A_M(n) \otimes X(n-M) \oplus B \otimes U(n). \tag{1}$$

In the case of tandem queues, given in Figure 2 the vector $U(n) = (\tau(n))$ which is the sequence of the inter-arrival times, and the general equation reduces to

$$\begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix} = A_0(n) \otimes \begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix} \oplus A_1(n) \otimes \begin{pmatrix} X_1(n-1) \\ X_2(n-1) \end{pmatrix} \oplus B \otimes (\tau(n)). \tag{2}$$

After simplifications and state expansion (see [4]), the implicit equation (1) can be written under the standard form:

$$X(n) = A(n) \otimes X(n-1) \oplus B \otimes U(n). \tag{3}$$

As for the example, the inductive solution of Equation (2) is

$$\begin{pmatrix} X_1(n) \\ X_2(n) \end{pmatrix} = A_0(n) \otimes A_1(n) \otimes \begin{pmatrix} X_1(n-1) \\ X_2(n-1) \end{pmatrix} \oplus B \otimes (\tau(n)). \tag{4}$$

.

We give the value of $A_0(n) \otimes A_1(n)$:

$$\begin{pmatrix} \sigma_1(n-1) & \sigma_2(n-1) \\ \sigma_1(n-1) + \sigma_1(n) & \sigma_2(n-1) + \sigma_1(n) \end{pmatrix}.$$

2.2 Queueing network models

The aim of this section is to give a few practical examples of queueing systems that fall in the class of (max,+) linear systems. Roughly speaking, if one considers the places in an event graph as buffers and the transitions as servers, one gets a queueing network which is (max,+) linear.

On a practical point of view, to check at first sight if a network of queues is (max,+) linear, one has to verify the following heuristic rules. If the following conditions are satisfied, then a queueing network is certainly (max,+) linear.

1 All queues in the network are FIFO with arbitrary service times.

2 The network is composed of buffer, server, fork and join nodes only.

4 No routing of customers in the network.

5 The network contains a single class of customers.

Note that in such a system, buffers may have arbitrary capacity (finite or infinite). The topology is also general and the network may contain cycles.

Typical examples are a G/G/1 queue, a queue with general blocking, (finite) queues in tandem, kanban systems [5], flexible manufacturing systems [10], fork-join queues or any parallel and/or series composition made with these elements.

The set of queueing networks satisfying these conditions does not cover the entire set of (max,+) linear systems, however, those systems may be difficult to model as queueing networks.

On the other hand, typical examples of a network which are not (max,+) linear are systems where

- the route of the customer does not depend solely on its origin. (this case covers probabilistic routing, routing depending on customer class or on the state of the system). Note that the loss of customers can be seen as a routing which depends on the state of the system;
- the service is not FIFO;
- the networks contains asynchronous superposition of several incoming streams in one buffer.

Therefore, several classical queueing networks such as Jackson networks or circuit switched networks are not (max,+) linear.

3 Multimodularity and admission control

3.1 Definition of multimodularity

For more details on multimodularity and its properties, please refer to [9, 3]. Here, we only give the basic definitions.

Let $e_i \in \mathbb{N}^m$, $i = 1, ..., m$ denote the vector having all entries zero except for a "1" in its ith place. Define $d_i = e_i - e_{i+1}$, $i = 1, ..., m-1$ (for an integer i taking values between 0 and m, we understand throughout $i - 1 = m$ for $i = 0$). Let $\{d_0 = -e_1, d_1, ..., d_{m-1}, d_m = -e_m\}$ be the base of vector for multimodularity. We have:

$$\begin{array}{rcccccccc} d_0 & = & (& -1 & 0 & 0 & \cdots & 0 & 0 &) \\ d_1 & = & (& 1 & -1 & 0 & \cdots & 0 & 0 &) \\ d_2 & = & (& 0 & 1 & -1 & \cdots & 0 & 0 &) \\ & & & & & \vdots & & & & \\ d_{m-1} & = & (& 0 & 0 & 0 & \cdots & 1 & -1 &) \\ d_m & = & (& 0 & 0 & 0 & \cdots & 0 & 1 &) \end{array} \tag{5}$$

Following Hajek [9], we say that a function f defined on $\mathbb{Z}^m$ is *multimodular* if $f(a + d_i) + f(a + d_j) \geqslant f(a) + f(a + d_i + d_j)$ for all $i \neq j$.

Here, we will only use functions defined on the positive orthant, $\mathbb{N}^m$, which is a convex subset on $\mathbb{Z}^m$ and we only require that

For all d_i and d_j with $i \neq j$ such that $a - d_i, a - dj, a - d_i - d_j$ are all in $\mathbb{N}^N$, we have $f(a - d_i) + f(a - d_j) \geqslant f(a) + f(a - d_i - d_j)$.

These changes will not alter the main properties of multimodular functions which are used in the following (see [3] for more in this issue).

3.2 Admission policy: the time slot approach

In the rest of the paper, we will use the following notations:

Let $\{T_i\}_{i=1,\cdots,N}$ be a sequence of arrival times, with the convention that $T_1 = 0$. Since all the rest of the notations are based on the original sequence, this can be considered as a time driven approach.

We denote by δ_i the ith *interval* length , that is: $\delta_i = T_{i+1} - T_i$. We assume that $\delta_0 = 0$. From now, the T sequence and the δ sequence are fixed.

As for the admission control, it is defined through an *arrival sequence*. The arrival sequence is a sequence of N integer numbers, $a = (a_1, a_2, \cdots, a_N)$, where a_i gives the number of customers admitted to the queue at time T_i. Furthermore, each customer carries a load. The load of the $j-$th customer is denoted σ_j. For convenience, we introduce the value $a_0 = 0$.

All functions will depend on a, sometimes implicitly, as for the function $\kappa(i)$ or the function $\nu(i)$, for example, which are defined in the following way.

The *counting function* $\kappa(i)$ is the number of arrivals by time T_i:

$$\kappa(i) \stackrel{\text{def}}{=} \sum_{j=1}^{i} a_j. \tag{6}$$

$$\nu(i) \stackrel{\text{def}}{=} \min\{m : \sum_{j=1}^{m} a_j \geqslant i\}. \tag{7}$$

Basically, $\nu(i)$ is the *number of intervals* elapsed at the ith arrival. We assume that $\nu(0) = 0$.

The *inter-arrival function* $\tau_i(a)$ is the i-th inter-arrival time. We have:

$$\tau_i(a) \stackrel{\text{def}}{=} \sum_{j=\nu(i)}^{\nu(i+1)-1} \delta_j. \tag{8}$$

The last inter-arrival time concerns the potential customer arriving at time T_N. Therefore its definition is different from the others: we set $\tau_{\kappa(N)}(a) = T_N - T_{\nu(\kappa(N))}$.

Figure 3 illustrates all these preliminaries.

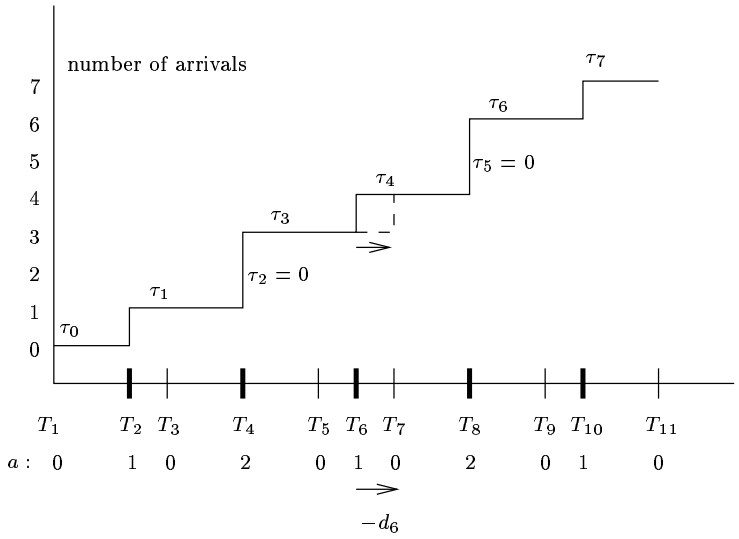

Figure 3: an example

4 The FIFO queue

In this section, we assume that the arrival stream enter a single FIFO queue. This is a simple case of a (max,+) system but the proofs presented here are typical of what happens in a more general framework.

The admission sequence a controls the arrival as illustrated in Figure 4.

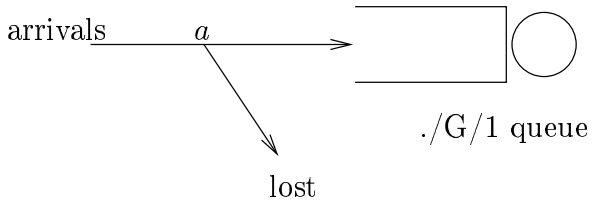

Figure 4: Admission control in one queue.

The system is a G/G/1 queue with batch arrivals. Note that this system is (max,+) linear.

In this part, we show that the function $\mathbf{E}W$, the expected workload, is multimodular on $\mathbb{N}^N$.

The workload at this time T_N under the admission control (a) is denoted by $W(a)$ and is given by the following expanded Lindley's equation:

$$W(a) = \max\left(0, (\sigma_{\kappa(N)} - \tau_{\kappa(N)}), (\sigma_{\kappa(N)} - \tau_{\kappa(N)}) + (\sigma_{\kappa(N)-1} - \tau_{\kappa(N)-1}), \cdots\right). \tag{9}$$

For convenience, we denote

$$w_j \stackrel{\text{def}}{=} \sum_{h=j}^{\kappa(N)} (\sigma_h - \tau_h).$$

Using this definition, we have:

$$W(a) = \max\left(0, w_{\kappa(N)}, \cdots, w_1\right).$$

4.1 Coupling of the service times with the customers

The rest of the section is devoted to proving multimodularity of the workload. Unfortunately, this does not hold on sample paths (as illustrated by the following example). We have to use a coupling of the service times with the customers and then, by assuming the service times are stationary, we will prove multimodularity of the expected workload.

Now, we will illustrate the difficulty of attaching the service times with customers while insuring that multimodularity holds through an example.

An example We consider a single queue with a sequence of service times $\sigma(n) = (4, 1, 1, 1, 1, 1, \cdots)$ and with possible arrival epochs are all the integer points.

We focus on the workload at time $T = 7$ under the arrival streams:

$$\begin{aligned} a &= (0,0,1,0,0,1,1,0) && (10)\\ a - d_0 &= (1,0,1,0,0,1,1,0) && (11)\\ a - d_6 &= (0,0,1,0,0,0,2,0) && (12)\\ a - d_6 - d_0 &= (1,0,1,0,0,0,2,0) && (13)\\ & && (14) \end{aligned}$$

The workload satisfies: $W(a) = 1, W(a - d_0) = 0, W(a - d_0 - d_6) = 0, W(a - d_0) = 0$, as shown in Figure 5.

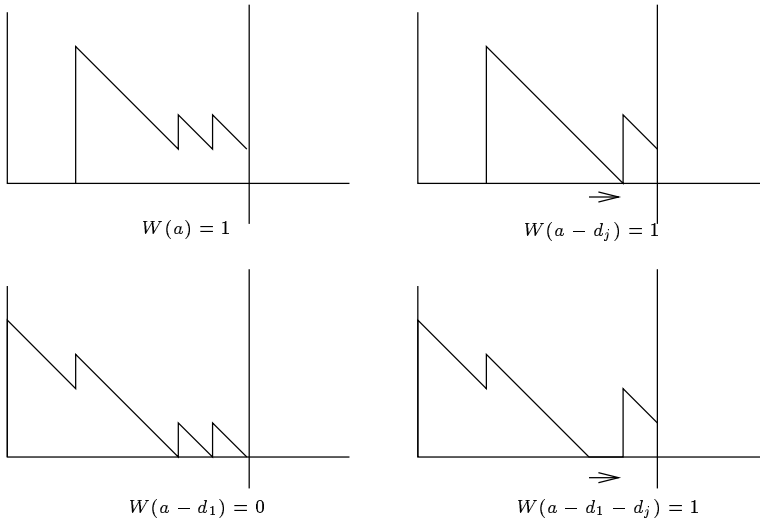

Figure 5: The workload is not multimodular.

This shows that the function W is not multimodular, since $W(a) + W(a - d_1 - d_7) > W(a - d_1) + W(a - d_7)$.

However, under a proper coupling of the service times with the customers, then W can be made multimodular.

We couple the service times with the customers in the following way:

$$\begin{array}{rcccccccccc}
a & = & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
\sigma & = & & & \sigma_1 & & & \sigma_2 & \sigma_3 & \\
 & & & & & & & & & \\
a-d_0 & = & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\
\sigma & = & \sigma_0 & & \sigma_1 & & & \sigma_2 & \sigma_3 & \\
 & & & & & & & & & \\
a-d_6 & = & (0 & 0 & 1 & 0 & 0 & 0 & 2 & 0) \\
\sigma & = & & & \sigma_1 & & & & \sigma_2,\sigma_3 & \\
 & & & & & & & & & \\
a-d_0-d_6 & = & (1 & 0 & 1 & 0 & 0 & 0 & 2 & 0) \\
\sigma & = & \sigma_0 & & \sigma_1 & & & & \sigma_2,\sigma_3 &
\end{array}$$

Under this coupling c, the workloads become as in Figure (6). We see that $W^c(a) = 0, W^c(a-d_0) = 0, W^c(a-d_0-d_6) = 1, W^c(a-d_6) = 1$, and W^c satisfies the multimodular inequality (which is an equality here),

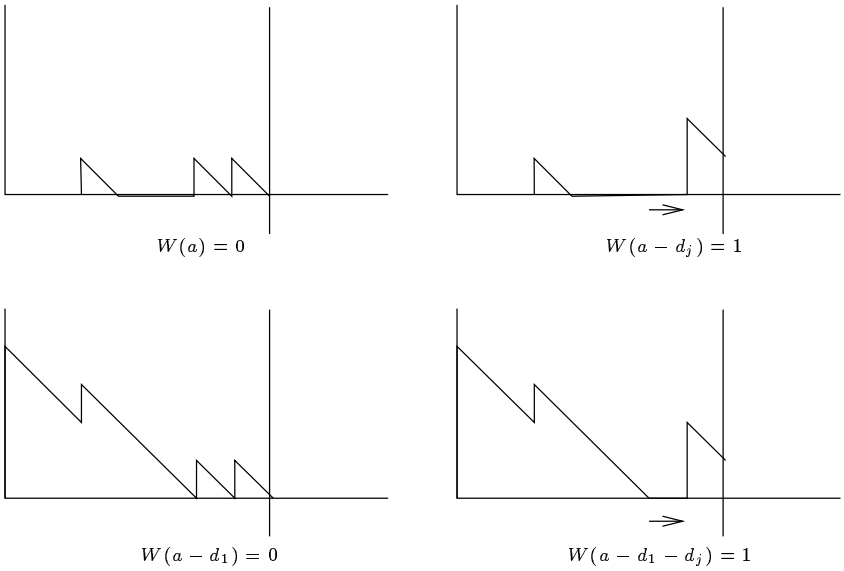

Figure 6: Multimodularity of the workload under proper coupling.

General coupling In general, the coupling of the service times with the arrival stream is done in the following way:

- Let a be an arbitrary arrival sequence in $\mathbb{N}^N$: $a = (a_1, a_2, \cdots, a_N)$. The service times are coupled with the customers entering the queue in the following fashion: With arriving batch a_i, we attach the service times:

$$\sigma_{\kappa(i-1)+1}, \sigma_{\kappa(i-1)+2}, \cdots, \sigma_{\kappa(i)}.$$

- With arrival sequence, $a - d_1$, the service time sequence for batch $a_i, i \neq 1$ is not modified and the batch attached with a_1 becomes:

$$\sigma_0, \sigma_1, \sigma_2, \cdots, \sigma_{\kappa(1)},$$

  where σ_0 is a new service time.

- If $a - d_j \in \mathbb{N}^N$, $1 < j < N$, the service time sequence for batch $a_i, i \neq j, j-1$ is not modified and the batch attached with a_{j-1} (which is not empty) becomes:

$$\sigma_{\kappa(j-2)+1}, \sigma_{\kappa(j-1)+2}, \cdots, \sigma_{\kappa(j-1)-1}$$

  and the batch attached with a_j becomes:

$$\sigma_{\kappa(j-1)}, \sigma_{\kappa(j-1)+1}, \cdots, \sigma_{\kappa(j)}.$$

  In other words, the service time $\sigma_{\kappa(j-1)}$ is moved from the $j-1$-th batch to the j-th batch.

- If $a - d_{N+1} \in \mathbb{N}^N$, the service time sequence for batch $a_i, i \neq N$ is not modified and the batch attached with a_N (which was not empty) becomes:

$$\sigma_{\kappa(N-1)+1}, \sigma_{\kappa(N-1)+2}, \cdots, \sigma_{\kappa(N)-1}.$$

  The last service time $\sigma_{\kappa(N)}$ has been removed.

- If $a - d_j - d_i \in \mathbb{N}^N$, the coupling of the arrival times is obtained by composing the modification of the coupling associated with a induced by d_i and d_j. This construction is commutative.

Note that this coupling strongly depends on the initial choice of a. If one changes the starting point a, then another coupling will be chosen. Also, the coupling is not defined on points which do not lay in the positive orthant, $\mathbb{N}^N$.

4.2 Multimodularity

We choose a point a in $\mathbb{N}^N$ and we construct the associated coupling c. We will denote by $W^c(.)$ the workload at time T_N under this coupling.

Now, using the notations given in the preliminaries, we have the following results.

Property 4.1. *Let $1 < i < N$. If $a_{i-1} > 0$, $\tau_{\kappa(i)-1}(a - d_i) = \tau_{\kappa(i)-1}(a) + \delta_i$, $\tau_{\kappa(i)}(a - d_i) = \tau_{\kappa(i)}(a) - \delta_i$. All others τ_j are unchanged.*

Proof. $-d_i$ corresponds to delaying the last acceptance at time T_i to time T_{i+1}. $\tau_{\kappa(i)}$ is the time interval between T_i and the arrival instant of the next customer that arrives after time T_i. Hence, delaying the arrival at T_i to time T_{i+1} results in increasing $\tau_{\kappa(i)-1}$ by $T_{i+1} - T_i = \delta_i$. $\tau\kappa(i)$ decreases by that value. See Figure 3 for an illustration of this proof. □

Property 4.2. *Let $0 < i < N$. If $a_i > 0$, $w_{\kappa(i)}(a - d_i) = w_{\kappa(i)} + \delta_i$. All others w_j are left unchanged.*

Proof. Follows from the Property 4.1 and the definition of w_j □

Under the previous coupling, we have the following results.

Lemma 4.3. *Let $h : \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function. The $h \circ W^c(a) + h \circ W^c(a - d_i - d_j) \leqslant h \circ W^c(a - d_i) + h \circ W^c(a - d_j)$, for all i, j and all a such that $a - d_i - d_j$, $a - d_i$ and $a - d_j$ are in $\mathbb{N}^N$.*

Proof.

- Let us first consider the case $0 < i, j < N$ and $a_i > 0, a_j > 0$. In this case, using the previous property 4.2, we have:

$$\begin{aligned} W^c(a - d_i) &= \max(W^c(a), w_{\kappa(i)} + \delta_i) && (15)\\ W^c(a - d_j) &= \max(W^c(a), w_{\kappa(j)} + \delta_j) && (16)\\ W^c(a - d_i - d_j) &= \max(W^c(a), w_{\kappa(i)} + \delta_i, w_{\kappa(j)} + \delta_j). && (17) \end{aligned}$$

  Therefore, by monotonicity of h, we have $h \circ W^c(a - d_i) + h \circ W^c(a - d_j) \geqslant h \circ W^c(a) + h \circ W^c(a - d_i - d_j)$.

- Let us now assume that $i = 0$. This case corresponds to the arrival of an extra customer at time $T_1 = 0$. This customer brings a load that we denote σ_0. In this case, we have:

$$W^c(a - d_0) = \max(W^c(a), w_1 + (\sigma_0 - \tau_0)). \quad (18)$$

  This case is treated similarly as the general case since $W^c(a - d_0)$ is of the form $\max(W^c(a), X)$.

- If $i = N$ and $a_N > 0$, then we have:

$$W^c(a - d_N) = (W^c(a) - \sigma_{\kappa(N)})^+. \quad (19)$$

  Here, for some X (see the first two cases)

$$W^c(a - d_j) = \max(W^c(a), X), \quad (20)$$

  and

$$W^c(a - d_j - d_N) = (\max(W^c(a), X) - \sigma_{\kappa(N)})^+, \quad (21)$$

  By a case analysis, we see that: If $W^c(a - d_j - d_N) = 0$ then, $W^c(a - d_N) = 0$ and by monotonicity of h, $h \circ W^c(a - d_N) + h \circ W^c(a - d_j) \geqslant h \circ W^c(a) + h \circ W^c(a - d_j - d_N)$. If $W^c(a - d_j - d_N) > 0$, then, $W^c(a - d_j) - W^c(a - d_j - d_N) = \sigma_{\kappa(N)}$. This yields,

$$W^c(a) - W^c(a - d_N) \leqslant W^c(a - d_j) - W^c(a - d_j - d_N).$$

Since h is convex and increasing,

$$\begin{aligned}
&h \circ W^c(a - d_j) - h \circ W^c(a - d_j - d_N) \\
&= h(W^c(a - d_j - d_N) + \sigma_{\kappa(N)}) - h \circ W^c(a - d_j - d_N) \\
&\geqslant h(W^c(a - d_N) + \sigma_{\kappa(N)}) - h \circ W^c(a - d_N) \\
&\geqslant h \circ W^c(a) - h \circ W^c(a - d_N).
\end{aligned}$$

This concludes the proof. □

Theorem 4.4. *If the service times are stationary, the function $E_\sigma h \circ W$ is multimodular, where E_σ denotes the expectation w.r.t. the firing time sequence σ_n.*

Proof. Let a be an arbitrary point in $\mathbb{N}^N$. Construct the associated coupling c of the services times. Under this coupling and for all i, j, Lemma 4.3 shows that $h \circ W^c(a - d_i) + h \circ W^c(a - d_j) - h \circ W^c(a) - h \circ W^c(a - d_i - d_j) \geqslant 0$. Therefore, $E_\sigma \Big(h \circ W^c(a - d_i) + h \circ W^c(a - d_j) - h \circ W^c(a) - h \circ W^c(a - d_i - d_j) \Big) \geqslant 0$. By the stationary assumption on the service times and using the fact that under coupling c, the service times involved in $W^c(.)$ are always consecutive, $E_\sigma h \circ W^c(.) = E_\sigma h \circ W(.)$, for the points $a - d_i$, $a - d_i$, $a - d_i - d_j$ and a (the expectation is invariant by shift). Finally we get

$$E_\sigma h \circ W(a - d_i) + E_\sigma h \circ W(a - d_j) - E_\sigma h \circ W(a) - E_\sigma h \circ W(a - d_i - d_j) \Big) \geqslant 0.$$ □

4.3 Number of customers in the queue with batch arrivals

Many other quantities of interest can be shown to be multimodular in different queueing models. We illustrate this fact by showing multimodularity of the number of customers in the queue with batch arrivals.

More precisely, we prove that the expected number of customers in a GI/M/1 queue with batch arrivals of general stationary size is multimodular in expectation. This is an extension of the original model studied in [9] which did not allow for batches of random size. The proof is very similar to the proof for the workload given previously.

We make the following assumption on the system:

- The customers arrive in batches of random size S_n, which form a stationary sequence.
- Each customer holds the server for a duration exponentially distributed.

The admission policy at arrival epochs consists in accepting the whole batch or rejecting the whole batch.

In general, using the fact that the service times are exponentially distributed, we construct X_k, as a sequence of independent random variables such that each X_k as the same distribution as the number of potential service completions during one inter-arrival time.

Therefore, the number of customers in the queue on the k-th arrival is:

$$N_n = \max(0, N_{n-1} + a_n S_n - X_n)$$

First, note that this recurrence equation is similar to the equation for the workload where S_k replaces σ_k and X_k replaces τ_k.

Note also that the example in Figure 5 shows that sample path multimodularity does not hold for the number of customers either. Indeed, with batches all of size one and admission policy,

$$\begin{aligned} a &= (0,0,1,0,0,1,1,0) \\ a - d_0 &= (1,0,1,0,0,1,1,0) \\ a - d_6 &= (0,0,1,0,0,0,2,0) \\ a - d_6 - d_0 &= (1,0,1,0,0,0,2,0), \end{aligned}$$

the number of customers in the queue at time T satisfies: $N(a) = 1, N(a - d_1) = 0, N(a - d_1 - d_7) = 0, N(a - d_7) = 0$.

We construct a coupling of the batch sizes which is similar to the coupling of services in the treatment of workload.

- Let a be an arbitrary arrival sequence in $\mathbb{N}^N$: $a = (a_1, a_2, \cdots, a_N)$. The batch sizes are coupled with the admission sequence in the following fashion.

  With arriving batch attached to a_i, we attach the batch size $s_{\kappa(i)}$.

- With arrival sequence, $a - d_0$, the service time sequence for batch $a_i, i \neq 1$ is not modified and the batch attached with a_1 becomes $s_0 + s_1$ where s_0 is a new service time.

- If $a - d_j \in \mathbb{N}^N$, then with arrival sequence, $a - d_j$, $0 < j < N$ the batch for $a_i, i \neq j, j-1$ is not modified and the batch attached to a_{j-1} (which was not empty) becomes 0 and the batch attached to a_j becomes: $s_{\kappa(j-1)} + s_{\kappa(j)}$. In other words, the batch $s_{\kappa(j-1)}$ is moved from the $j-1$-th slot to the j-th slot.

- If $a - d_{N+1} \in \mathbb{N}^N$, then with arrival sequence, $a - d_{N+1}$, the batch size for $a_i, i \neq N$ is not modified and the batch size attached to a_N (which was not empty) becomes 0. The batch of size $s_{\kappa(N)}$ has been removed.

- If $a - d_j - d_i \in \mathbb{N}^N$, then with arrival sequence, $a - d_j - d_i$, the coupling of the batch sizes is obtained by composing the modification of the coupling associated with a induced by d_i and d_j. This construction is also commutative.

Under this coupling, the number of customers in the queue satisfies:

$$N_n = \max(0, N_{n-1} + S_{\kappa(n)} - X_n)$$

We denote by $\mathbf{E}_s N_k$ the expectation with respect with the batch size of the number of customers.

Theorem 4.5. *The expected number of customers* $\mathbf{E}_s N_\kappa(a)$ *is multimodular.*

Proof. To prove this result, it suffices to prove that for a fixed sequence x_k and under the coupling c, the variable

$$N_n^c(a) \stackrel{\text{def}}{=} \max(0, N_{n-1}^c(a) + S_{\kappa(n)}^c - x_n),$$

satisfies the multimodular condition, $N_n^c(a)+N_n^c(a-d_j-d_i) \leqslant N_n^c(a-d_i)+N_n^c(a-d_j)$. The proof of this result is similar to the proof of the multimodularity of the function $W^c(a)$, using the analogy mentioned previously in the equations between the batch size and the service times on one hand and between the number of services and the inter-arrival times on the other hand. □

5 The general case: (max,+) systems with one input

The G/G/1 queue is a special case of a (max,+) linear system. This section will generalize the multimodularity properties to the case of an arbitrary network which is (max,+) linear and has one input.

5.1 Description of the system

In this section, we will consider a general stochastic event graph with a single input transition. As mentioned in §2, the dymanical behavior of this system is linear in the max,+) algebra.

Let $\mathcal{G} = (\mathcal{P}, \mathcal{Q}, \mathcal{M}_0)$ be an event graph with Q nodes (or transitions) and only one input (which can be seen as a source of exogenous costumers). This single input node, is represented by a special transition q_0 in the marked graph. We further assume with no loss of generality that for all transitions directly connected to q_0, the initial marking in the interconnection place is equal to zero.

For each transition q (or node) in $\mathcal{G}$, we consider all the paths π from q_0 to q. This set is denoted y $\mathcal{P}(q)$. We also denote by $M(\pi) \stackrel{\text{def}}{=} \sum_{s\in\pi} \mathcal{M}_0(s)$, the sum of the initial tokens on the path π. Now, we define

$$L(q) = \min_{\pi\in\mathcal{P}(q)} M(\pi).$$

Lemma 5.1. *When an event graph complies with the foregoing assumptions, the* $n+L(q)$*-th firing of transition* q *of* $\mathcal{G}$ *involves a token produced by the* n*-th firing of transition* q_0*.*

Proof. Let S_q be the shortest path from q_0 to q with $L(q)$ tokens. The length of S_q is called the distance from q_0 to q. The proof holds by induction on the length of S_q. If $S_q = 0$, then $q = q_0$ and the result is true. Suppose that the result is true for all transitions at "distance" $k-1$ from q_0. Choose q at distance k, then the transition q' preceding q on the path S_q

is at distance $k-1$ from q_0 and induction applies to q'. Now the place (q', q) contains m tokens. By definition of q', $L(q') = L(q) - m$, and by induction, the $n + L(q')$-th firing of transition q' uses the token number n and since the buffer place between q' and q is FIFO, the $n + L(q)$-th firing of q will use that token. □

5.2 Lindley's equation for (max,+) systems

We recall that $X_q(n)$ is the epoch when transition q fires for the n-th time with the convention that $X_q(n) = 0$ if $n \leqslant 0$. Note that because of Lemma 5.1, $X_q(n)$ is also the epoch when the $n - L(q)$-th token generated by the input is fired by transition q.

Now, let W_n be a vector , with all its elements equal to: $W_n^q \stackrel{\text{def}}{=} X_q(n + L(q)) - X_0(n)$. Using Lemma 5.1, W_n^q can be seen as the *traveling time* for customer n between its entrance in the system and its passage in transition q.

In the so-called (max,+) algebra, we recall that we have the following vectorial equation that describes the evolution of $\mathcal{G}$ (see [6]):

$$X(n) = A(n) \otimes X(n-1) \oplus B \otimes U_n.$$

If we consider only the element $X_i(n)$ and we subtract U_n on each side of this equation, we get:

$$W_n^i = A_i(n + L(i)) \otimes X(n + L(i) - 1) \oplus B_i \otimes T_{n+L(i)} - T_n.$$

We distinguish two cases. If i directly depends on the input (*i.e.* there is a place between q_0 and i), then by assumption, $L(i) = 0$. In this case, we get:

$$\begin{aligned} W_n^i &= A_i(n) \otimes X(n-1) - U_n \oplus B_i \\ &= (A_i(n) - U_n) \otimes X(n-1) \oplus B_i \end{aligned}$$

If i and q_0 are not directly connected, then $B_i = -\infty$ and we have:

$$\begin{aligned} W_n^i &= A_i(n + L(i)) \otimes X(n + L(i) - 1) - U_n \\ &= A_i(n + L(i)) \otimes X(n + L(i) - 1) - U_n \oplus B_i \\ &= (A_i(n + L(i)) - \tau_n) \otimes X(n + L(i) - 1) - U_{n-1} \oplus B_i. \end{aligned}$$

Now, we introduced the matrix $H(n)$ defined row by row:

$$H_i(n) \stackrel{\text{def}}{=} A_i(n + L(i)).$$

If we write this last equality for $W(n)$ in vectorial form, we get

$$W_n = H(n) \otimes D(-\tau_n) \otimes W_{n-1} \oplus B,$$

where

- $D(h)$ is the diagonal matrix with h on the diagonal and $-\infty$ everywhere else.
- B is the matrix which describes the input connection.

This recursion is a generalization of the Lindley equation in the case of a network.

By using elementary matrix operations in the (max,+) algebra, the equation can also be developed into:

$$W_{n+1} = B \oplus \bigoplus_{i=1}^{n} C_i,$$

with

$$C_i = \bigotimes_{j=i}^{n} \big(H(j) \otimes D(-\tau_j)\big) \otimes B.$$

In the following, we will also often use the following transformation for notation convenience. If X is a vector of size Q, then $\mathbf{X}$ is a diagonal matrix of size $Q \times Q$, with the vector X on the diagonal and $-\infty$ elsewhere.

5.3 Multimodularity

The main result established in this section is that the expectation of $W_n(q)$ is multimodular for all q. The proof is very similar to the case of the single queue and is made surprisingly even easier by using the vectorial form of the Lindley equation in the (max,+) algebra.

The multidimensional coupling of the service times in each transition with the arriving customers is done similarly as in the one queue case.

However, we first assume that the service times in transition q are indexed from $1-L(q)$ rather than from 1. The n-th firing in transition q lasts $\sigma^q_{n-L(q)}$ units of time. We also construct a coupling for transition q which is independent of the coupling for any other transition.

- Let a be an arbitrary arrival sequence in $\mathbb{N}^N$: $a = (a_1, a_2, \cdots, a_N)$. The service times are coupled with the customers entering the system in the following fashion:

  With arriving batch a_i, we attach the service times: $\sigma^q_{\kappa(i-1)+1}, \sigma^q_{\kappa(i-1)+2}, \cdots, \sigma^q_{\kappa(i)}$.
- With arrival sequence, $a-d_1$, the service time sequence for batch $a_i, i \neq 1$ is not modified and the batch attached to a_1 now becomes under this coupling: $\sigma^q_0, \sigma^q_1, \sigma^q_2, \cdots, \sigma^q_{\kappa(1)}$ where σ^q_0 is a new service time.
- If $a - d_j \in \mathbb{N}^N$, then with arrival sequence, $a - d_j$, $1 < j < N$ the service time sequence for batch $a_i, i \neq j, j-1$ is not modified and the batch attached to a_{j-1} becomes: $\sigma^q_{\kappa(j-2)+1}, \sigma^q_{\kappa(j-2)+2}, \cdots, \sigma^q_{\kappa(j-1)-1}$ and the batch attached to a_j becomes: $\sigma^q_{\kappa(j-1)}, \sigma^q_{\kappa(j-1)+1}, \cdots, \sigma^q_{\kappa(j)}$. In other words, the service time $\sigma^q_{\kappa(j-1)}$ is moved from the $j-1$-th batch to the j-th batch.

- If $a - d_{N+1} \in \mathbb{N}^N$, then with arrival sequence, $a - d_{N+1}$, the service time sequence for batch $a_i, i \neq N$ is not modified and the batch attached to a_N (which is not empty) becomes: $\sigma^q_{\kappa(N-1)+1}, \sigma^q_{\kappa(N-1)+2}, \cdots, \sigma^q_{\kappa(N)-1}$. The last service time $\sigma^q_{\kappa(N)}$ has been removed.

- If $a - d_j - d_i \in \mathbb{N}^N$, then with arrival sequence, $a - d_j - d_i$, the coupling of the arrival times is obtained by composing the modification of the coupling associated with a induced by d_i and d_j. We note again that the construction is commutative.

Now, we consider the traveling time matrix of a potential customer entering the system at time T_N under arriving stream a and its associated coupling. We denote this matrix $W(a)$. This matrix is defined by the following equation:

$$W(a) = B_{\kappa(N)} \oplus \bigoplus_{i=1}^{\kappa(N)} C_i(a), \tag{22}$$

with

$$C_i(a) = \bigotimes_{j=i}^{\kappa(N)} \big(H(j) \otimes D(-\tau_j(a))\big) \otimes B.$$

We have similar lemmas as in the one queue case, (see Properties 4.1 - 4.2).

Lemma 5.2. *Let $0 < i < N$. If $a_i > 0$, then $C_{\kappa(i)}(a - d_i) = D(\delta_i) \otimes C_{\kappa(i)}(a)$. All other C_j are left unchanged.*

Proof. Using property 4.1, we have: $D(-\tau_{\kappa(i)-1}(a - d_i)) = D(\delta_i) \otimes D(\tau_{\kappa(i)-1}(a))$ and $D(-\tau_{\kappa(i)+1}(a - d_i)) = D(-\delta_i) \otimes D(\tau_{\kappa(i)}(a))$ with all others τ_j left unchanged.

- Now, for every $j > \kappa(i)$, then $C_j(a - d_i)$ does not involved $D(-\tau_{\kappa(i)}(a - d_i))$ or $D(-\tau_{\kappa(i)-1}(a - d_i))$, and therefore is left unchanged.

- If $j < \kappa(i)$, then $C_j(a - d_i)$ involves $D(-\tau_{\kappa(i)}(a - d_i))$ and $D(-\tau_{\kappa(i)-1}(a - d_i))$. Since the matrices $D(x)$ commute with everything, and since $D(\delta_i) \otimes D(-\delta_i) = E$, the identity matrix, then, $C_j(a - d_i)$ is left unchanged.

- Finally, if $j = \kappa(i)$, then $C_j(a - d_i)$ involves $D(-\tau_{\kappa(i)}(a - d_i))$ but not $D(-\tau_{\kappa(i)-1}(a - d_i))$. Using the fact that $D(\delta_i)$ commutes with all the other matrices, we have $C_{\kappa(i)}(a - d_i) = D(\delta_i) \otimes C_{\kappa(i)}(a)$.

□

In the following we will use to simplify the equations the variables: $Z_i \stackrel{\text{def}}{=} D(\delta_i) \otimes C_{\kappa(i)}(a)$.

Lemma 5.3. *Let $h : \mathbb{R}^Q \to \mathbb{R}^Q$ be a component-wise increasing convex function. Component-wise, we have*

$$\mathbf{h} \circ \mathbf{W}(a - d_i) \otimes \mathbf{h} \circ \mathbf{W}(a - d_j) \geqslant \mathbf{h} \circ \mathbf{W}(a) \otimes \mathbf{h} \circ \mathbf{W}(a - d_j - d_i). \tag{23}$$

Proof. First note as a general remark that for any matrix M and any positive number x, $D(x) \otimes M \oplus M = D(x) \otimes M$. Now, as in the case of a single queue, we have to distinguish three cases.

- The case where $0 < i, j < N$ and $a_i, a_j > 0$. by commutativity of the $\oplus$ operator, we have

$$\begin{aligned} W(a - d_i) &= W(a) \oplus Z_i & (24)\\ W(a - d_j) &= W(a) \oplus Z_j & (25)\\ W(a - d_j - d_i) &= W(a) \oplus Z_j \oplus Z_i. & (26) \end{aligned}$$

  Now using the distributivity of $\otimes$ w.r.t. $\oplus$, we have $\mathbf{W}(a - d_i) \otimes \mathbf{W}(a - d_j) = \mathbf{W}(a) \otimes \mathbf{W}(a - d_j - d_i) \oplus \mathbf{Z}_j \otimes \mathbf{Z}_i$. This last equation interpreted in the classical algebra says that for each server q, the traveling time at time T_N satisfies:

$$W_N^q(a - d_i) + W_N^q(a - d_j) \geqslant W_N^q(a) + W_N^q(a - d_j - d_i). \tag{27}$$

  The fact that h is increasing and using the Equation (27) implies that $h \circ W_N^q(a - d_i) + h \circ W_N^q(a - d_j) \geqslant h \circ W_N^q(a) + h \circ W_N^q(a - d_j - d_i)$.

- Now we examine the case where $i = 0$. As in the single queue case, this corresponds under our coupling to the arrival of an extra customer at time $T_1 = 0$ that has a service time that we denote σ_0^q in queue q. In this case, we have:

$$W(a - d_0) = W(a) \oplus \bigotimes_{j=0}^{n} \left(A_j \otimes D(-\tau_j)\right) \otimes B_i. \tag{28}$$

  This case is treated as the general case, since $W(a - d_i)$ is of the form $W(a) \oplus Z$, for some positive vector Z.

- If $i = N$ and $a_N > 0$. This case corresponds to the removal of the last customer in the arrival batch, which happens to have arrived at time T_N. In this case, we have:

$$W(a - d_N) = S \otimes W(a) \oplus O,$$

where O is a vector composed of 0 and S is a diagonal matrix with $S_{[q,q]} = -\sigma^q_{\kappa(N)}$. Since we also have $W(a-d_j) = W(a) \oplus Z$ and $W(a-d_j-d_N) = S \otimes (W(a) \oplus Z) \oplus O = S \otimes W(a) \oplus S \otimes Z \oplus O$, then we get:

$$\begin{aligned}
&\mathbf{W}(a-d_N) \otimes \mathbf{W}(a-d_j) \\
&= (S \otimes \mathbf{W}(a) \oplus \mathbf{O}) \otimes (\mathbf{W}(a) \oplus \mathbf{Z}) \\
&= S \otimes \mathbf{W}(a) \otimes \mathbf{W}(a) \oplus \mathbf{O} \otimes \mathbf{W}(a) \oplus S \otimes \mathbf{W}(a) \otimes \mathbf{Z} \oplus \mathbf{O} \otimes \mathbf{Z},
\end{aligned}$$

and on the other hand,

$$\mathbf{W}(a) \otimes \mathbf{W}(a-d_j-d_N) \;=\; \mathbf{W}(a) \otimes S \otimes \mathbf{W}(a) \oplus \mathbf{W}(a) \otimes S \otimes \mathbf{Z} \oplus \mathbf{W}(a) \otimes \mathbf{O}.$$

Since every matrix involved in these equations is diagonal, they commute and we have:

$$\mathbf{W}(a-d_N) \otimes \mathbf{W}(a-d_j) = \mathbf{W}(a) \otimes \mathbf{W}(a-d_j-d_N) \oplus \mathbf{O} \otimes \mathbf{Z}.$$

Since $\mathbf{O} \otimes \mathbf{Z}$ is a positive-diagonal matrix, the result is established by rewriting this equation in the conventional algebra.

As with the function h, the proof is similar to the case of a single queue. □

Theorem 5.4. *If the service times in every node are mutually independent stationary sequences, the function* $\mathbf{E}_\sigma h(W_q(a))$ *is multimodular, where* $\mathbf{E}_\sigma$ *denote the expectation w.r.t. the service times in all the nodes of the system.*

Proof. The proof is similarly to the one queue case. The coupling c is compatible with stationary service times in all nodes of the system if they are stationary sequences. Therefore, the inequality given in Equation (23) implies the multimodularity of the expected traveling time w.r.t. all service times. □

Note that this theorem proves that the traveling time for a customer arriving at time slot N, that is, the time between its entrance in the system and its service in queue q, is multimodular, for all q and N.

Also note that the case of event graphs with multiple or no entries as well as the case where the marked graph is not empty initially are intractable by this means. Indeed, the coupling of the service times to the customers is not feasible *a priori* in those cases. It will depend on the sequence of arrivals in the case of multiple entries and all the service times in the case of a closed system.

6 A dual policy: counting variable and waiting time

In the previous section, we were interested in the study of the workload which is a criterion related with the server. Here we will focus on the waiting time of the customers entering

the network. Previously, all quantities were indexed by n, the number of time slots. In the section, all quantities will rather be indexed by $\kappa(n)$, the number of arrivals. First, assume that the sequence a has all its values in $\{0,1\}$. This means that simultaneous arrivals are not allowed. In this respect, the counting sequence b, will be given in the following way: $b_n = \nu(n) - \nu(n-1) - 1$. For example if $a = (0,0,1,0,0,1,0,1,1,0,0,1)$, then $b = (2,2,1,0,2)$. Note that as long as $a \in \{0,1\}^N$, then a and b represent the same information. a_n gives the number of arrivals at time slot n (time driven), and the dual variable b_k gives the number of time slots elapsed between the $(k-1)th$ arrival and the kth arrival (event driven).

6.1 Waiting time

Let S be a (max,+) linear system with a single input satisfying the assumptions given in §5.

Then, the traveling time of the k-th customer to node i is denoted by $\mathcal{W}_k^i(b)$. The vector $\mathcal{W}_k(b)$ satisfies the vectorial Lindley equation, using the function τ defined in equation (8):

$$\mathcal{W}_{k+1}(b) = H(k) \otimes D(-\tau_k) \otimes \mathcal{W}_k(b) \oplus B,$$

This can be written:

$$\mathcal{W}_{k+1} = B \oplus \bigoplus_{i=1}^{n} \mathcal{C}_i,$$

with

$$\mathcal{C}_i = \bigotimes_{j=i}^{k} \big(H(j) \otimes D(-\tau_j)\big) \otimes B.$$

This equation has essentially the same form as the equation (22).

6.2 Coupling

The coupling adapted in this case is essentially the dual of the coupling on the service times used previously. Here we rather couple the inter-arrival times, δ_i. This coupling uses the function ν instead of the function κ.

We build the coupling d in the following way.

- Let b be an arbitrary arrival sequence in $\mathbb{N}^N$: $b = (b_1, b_2, \cdots, b_N)$. The intervals are coupled with the customers entering the queue in the following fashion: With interval length b_i, we attach the intervals: $\delta_{\nu(i-1)+1}, \delta_{\nu(i-1)+2}, \cdots, \delta_{\nu(i)}$.

- With $b + d_0$, the interval sequence for length $b_i, i \neq 1$ is not modified and the length attached with b_1 (which was not empty) becomes: $\delta_1, \delta_2, \cdots, \delta_{\nu(1)}$. where δ_0 has been removed.
- If $b + d_j \in \mathbb{N}^N$, then with arrival sequence, $b + d_j$, $1 < j < N$ the interval sequence for length $b_i, i \neq j, j-1$ is not modified and the length attached with b_{j-1} becomes: $\delta_{\nu(j-2)+1}, \delta_{\nu(j-1)+2}, \cdots, \delta_{\nu(j-1)}, \delta_{\nu(j-1)+1}$ and the length attached with b_j (which is not empty) becomes: $\delta_{\nu(j-1)+2}, \cdots, \delta_{\nu(j)}$. In other words, the interval $\delta_{\nu(j-1)+1}$ is moved from the j-th length to the $(j-1)$-st length.
- If $b + d_N \in \mathbb{N}^N$, then with arrival sequence, $b + d_{N+1}$, the interval sequence for length $b_i, i \neq N$ is not modified and the length attached with b_N becomes $\delta_{\nu(N-1)+1}, \delta_{\nu(N-1)+2}, \cdots, \delta_{\nu(N)-1}, \delta_{\nu(N)+1}$. The last interval $\delta_{\nu(N)+1}$, is a new interval.
- If $b + d_j + d_i \in \mathbb{N}^N$, then with arrival sequence, $b + d_j + d_i$, the coupling of the arrival times is obtained by composing the modification of the coupling associated with b induced by d_i and d_j. This construction is commutative.

6.3 Multimodularity

In this case, we will use the direct base $\{d_0, \cdots, d_N\}$ rather than $\{-d_0, \cdots, -d_N\}$ used in the previous sections.

Property 6.1. *Let $0 < i < N$. If $a_i > 0$, then $\tau_i(b + d_i) = \tau_i(b) + \delta_{\nu(i)+1}$. $\tau_{i+1}(b + d_i) = \tau_{i+1}(b) - \delta_{\nu(i)+1}$. All others τ_j are unchanged.*

Proof. By definition. See Figure 3 □

Lemma 6.2. *Let $0 < i < N$. If $b_i > 0$, then $\mathcal{C}_{i+1}(b + d_i) = D(\delta_{\nu(i)+1}) \otimes C_{i+1}(b)$. All other $\mathcal{C}_j$ are left unchanged.*

Proof. The proof is similar to the proof of Lemma 5.2 and follows from the Property 6.1 and the definition of $\mathcal{C}_j$ □

In the following we will simplify the equations using the variables: $\mathcal{Z}_i \stackrel{\text{def}}{=} D(\delta_{\nu(i)+1}) \otimes \mathcal{C}_{i+1}(a)$.

Under the previous coupling, we have the following results:

Lemma 6.3. *Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a increasing convex function. Then, $h \circ \mathcal{W}^d(b) + h \circ \mathcal{W}^d(b + d_i + d_j) \leqslant h \circ \mathcal{W}^d(b + d_i) + h \circ \mathcal{W}^d(b - d_j)$, for all i, j and all b such that $b + d_i + d_j$, $b + d_i$ and $b + d_j$ are in $\mathbb{N}^N$.*

Proof. The proof is essentially similar to the proof of lemma 5.3. The proof follows from the following equalities by using the same technique as in the previous case.

- If $0 < i, j < N$,

$$\begin{aligned} \mathcal{W}(a + d_i) &= \mathcal{W}(a) \oplus \mathcal{Z}_i & (29) \\ \mathcal{W}(a + d_j) &= \mathcal{W}(a) \oplus \mathcal{Z}_j & (30) \\ \mathcal{W}(a + d_j + d_i) &= \mathcal{W}(a) \oplus \mathcal{Z}_j \oplus \mathcal{Z}_i. & (31) \end{aligned}$$

If $i = 0$,

$$\mathcal{W}(a + d_0) = \mathcal{W}(a) \oplus \mathcal{Z}_0. \quad (32)$$

If $i = N$,

$$\mathcal{W}(a + d_N) = \mathcal{S} \otimes \mathcal{W}(a) \oplus \mathcal{O}, \quad (33)$$

where $\mathcal{O}$ is a vector composed of 0 and $\mathcal{S}$ is a diagonal matrix with $\mathcal{S}_{[q,q]} = -\delta^q_{\nu(N)+1}$. □

Theorem 6.4. *If the intervals δ_i form a stationary sequence, the function $\mathbf{E}_\delta(h \circ \mathcal{W})$ is multimodular.*

The proof is essentially similar to the proof of Theorem 5.4.

7 Optimal admission sequence

In this section we use the multimodularity property to derive optimality of the most regular admission sequence.

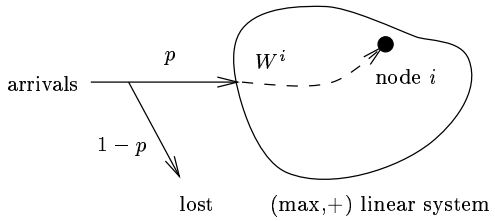

Figure 7: Sketch of the control problem

We address now the extension of the admission control problem to the case of a network. We want to admit customers in a (max,+) linear system S, under the constraint that in the long run, the fraction of customers admitted be at least p.

In the rest of the paper, we will also assume that the system S is initially empty, *i.e.* for every node q in S, there exists a path from q_0 to q that contains no tokens. More formally, this can be written,

$$L(q) = 0, \forall q \in S.$$

This assumption will be necessary to prove Lemmas 7.4 and 7.6.

Now the problem is to find which admission policy minimizes the traveling time to node i of an customer admitted in the system, as illustrated in the figure 7. As in the previous sections, the admission is governed by a sequence a_n.

7.1 Balanced Sequences

We first present the notion of balanced sequence that will represent the optimal admission sequence. More details on these notions are presented in [2] and references therein.

Definition 7.1. *The sequence* $a \stackrel{\text{def}}{=} \{a_n, n \in \mathbb{N}\}$ *is a balanced sequence if there exists two real numbers* θ *and* p *such that* $a_n = \lfloor np+\theta \rfloor - \lfloor (n-1)p+\theta \rfloor$, *where* $\lfloor x \rfloor$ *denotes the largest integer smaller or equal to* x. *The quantity* θ *is the initial phase and* p *is the rate of* a.

Note that θ can always be chosen in $[0, 1)$ and $\forall k$,

$$\lim_{d\to\infty} \frac{\sum_{i=k}^{k+d} a_i}{d} = p.$$

This justifies the fact that p is called the rate of a. Also note that $a_n \in \{\lfloor p \rfloor, \lceil p \rceil\}$.

The following theorem is proved in [7] using methods from combinatorics.

Theorem 7.2 ([7]). *The sequence* a *is balanced if and only if for any* d, k, l,

$$\left| \sum_{i=k}^{k+d} a_i - \sum_{i=l}^{l+d} a_i \right| \leqslant 1$$

This theorem gives a characterization of balanced admission sequences which says that the number of admissions among any d consecutive decisions differs by at most 1, for all d. This gives the intuition why such a sequence is "balanced".

The following lemma gives the relation between a balanced admission sequence a and its corresponding counting sequence b.

Lemma 7.3. *Let* a *be any sequence in* $\{0, 1\}$ *and* b *be the associated counting sequence* ($b_n = \nu(n) - \nu(n-1) - 1$). *Then,* a *is balanced with rate* p *if and only if* b *is balanced with rate* $(1-p)/p$.

Proof. First, let us prove that if a is balanced, then b is balanced. Note that b is composed of only two integers n and $n+1$, otherwise, a would contain the intervals 10^n01, $10^{n+j}1$, with $j \geqslant 2$. This contradicts the balance condition for a given in Theorem 7.2. Now, let us suppose that b is not balanced. This means that there exists two intervals of b of the same length k that contain i(resp. $i+j$) times the integer $(n+1)$, with $j \geqslant 2$. For the sequence a, this yields a first interval of length $k(n_1)+i+1$ with $k+1$ times the integer 1 and a second sequence of length $k(n+1)+i+j-1$ that contains $k-1$ times the integer 1. Since $j \geqslant 2$, the length of the second interval is larger than the length of the first one.

We can extract a sub-interval of a with at most $k-1$ times the integer 1. This contradicts the fact that a is balanced.

Now let us suppose that b is balanced. By definition of a balanced sequence, b contains only two integers $(n, n+1)$, where $n = \lfloor p \rfloor$. Let us assume that a is not balanced. Using Theorem 7.2, this means that a contains two intervals of the same length l, with respectively k and $k+j$ times the integer 1 and $j \geqslant 2$. The interval with the smaller number 1 is extended on both sides so that it contains exactly $k+j$ ones, starting with a 1 and ending with a 1. The second interval is shrunk on both side so that it starts and ends with a 1. Their respective length are now $l_1 \geqslant l+2 \geqslant l_2+2$. For b this yields two intervals of the same length:$k+j-1$. They contain b_1 (resp. b_2) times the integer $n+1$. This means that for the intervals in a, their length are respectively: $l_1 = 1+(k+j-1)(n+1)+b_1$ and $l_2 = 1+(k+j-1)(n+1)+b_1$. The fact that $l_1 \geqslant l+2 \geqslant l_2+2$ implies that $b_1 \geqslant b_1+2$. This contradicts the fact that b is balanced.

At this point we have established that a is balanced if and only if b is also balanced. As for the rates, assume that a has rate p, Then if we consider an interval in a with k times 1, its length is $l = k/p + o(k)$. For b, the associated interval is of length $k+o(k)$ and sums to $l-k+o(k)$. The rate is $(l-k)/k = (1-p)/p$. □

7.2 The time slot approach

Now, we are ready to consider the problem to find the optimal admission policy with a given rate. We denote by $\mathbf{E}_{\sigma,\delta}(.)$ the expectation w.r.t. the inter-arrival times and the service times in all nodes in the system. Following the technical conditions given in [3], we have to make sure that the expected traveling times to node i under the admission sequence a, satisfy the properties given in Lemma 7.4. Let $h : \mathbb{R}^Q \to \mathbb{R}^Q$ be an arbitrary convex function component-wise non-decreasing. For simplicity we denote in the following $W_n^i(a_1, \cdots, a_n)$ instead of $h \circ W_n(i)(a_1, \cdots, a_n)$.

Lemma 7.4. *Assume that the inter-arrival times and the service times are stationary sequences, independent of each other. The following properties are true:*
i- $\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n)$ *is non-decreasing.*
ii- $\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n) = \mathbf{E}_{\sigma,\delta} W_m^i(0, \cdots, 0, a_1, \cdots, a_n)$, $n < m$.
iii- $\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n)$ *is multimodular.*

Proof. We prove the three properties in sequence.

i- Using the extended Lindley formula, the expected traveling time $\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n)$ is also increasing.

ii- Let us fix n and m with $n < m$. Since the inter-arrival times are stationary, we can couple the intervals such that $\forall 1 \leqslant j \leqslant n$, $\delta_j^{(n)} = \delta_{j+m-n}^{(m)}$. (this is merely a shift of the sequence of intervals for W_m^i). Under this coupling and because the system is initially empty, we have:

$$W_n(i)(a_1, \cdots, a_n) = W(i)_m(0, \cdots, 0, a_1, \cdots, a_n),$$

and,

$$\mathbf{E}_\sigma W_n^i(a_1, \cdots, a_n) = \mathbf{E}_\sigma W_m^i(0, \cdots, 0, a_1, \cdots, a_n).$$

Therefore, if $n < m$,

$$\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n) = \mathbf{E}_{\sigma,\delta} W_m^i(0, \cdots, 0, a_1, \cdots, a_n).$$

iii- Theorem 5.4 shows that $\mathbf{E}_\sigma W_n^i(a_1, \cdots, a_n)$ is multimodular, thus $\mathbf{E}_{\sigma,\delta} W_n^i(a_1, \cdots, a_n)$ is multimodular as well. □

Now, we choose the cost function $g^i(a)$ of a policy $a = (a_1, a_2, \cdots)$ to be the Cezaro sum of the expected traveling time to a given node i, of a potential customer admitted the queue at time T_n, composed by an arbitrary increasing convex function, that is:

$$g^i(a) \stackrel{\text{def}}{=} \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\sigma,\delta} W_n^i(a_1, ..., a_n).$$

Note that $g^1(a)$ could be called the average traveling time at epochs $\{T_n\}$. In the case the sequence $\{T_n\}$ is a Poisson process, $g^i(a)$ is the time-average of the traveling times.

To find which admission policy minimizes the cost $g^1(.)$, we need to use the notion of *balanced sequences*.

We denote the balanced sequence with phase θ and rate p by

$$a_k^p(\theta) \stackrel{\text{def}}{=} \lfloor kp + \theta \rfloor - \lfloor (k-1)p + \theta \rfloor. \tag{34}$$

We can restate the general theorems given in [3] in the particular case studied here.

Theorem 7.5. *For all sequence a such that*

$$\underline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} a_n \geqslant p,$$

then

$$g^i(a) \geqslant g^i(a^p(\theta)),$$

for any $\theta \in [0, 1]$.

Proof. First note that the balanced sequence $a_k^p(\theta)$ has an asymptotic rate equal to p. Then, Corollary 4.1 in [3] can be applied with since all necessary assumptions on W are satisfied. □

7.3 The counting approach, the bounded case

In this section, we use the counting sequence b rather than the admission sequence a and therefore, we rather focus on the function $h \circ \mathcal{W}(i)_n(b_1, \cdots, b_n)$, which is denoted $\mathcal{W}^i_n(b_1, \cdots, b_n)$ for the sake of notation simplicity.

Lemma 7.6. *Assume that the service time and the inter-arrival are stationary sequences, independent of each other and that they are bounded from above and from below by S and D respectively. i-* $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n)$ *is decreasing.*
ii- $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n) \geqslant \mathbf{E}_{\delta,\sigma}\mathcal{W}^i_m(b_{n-m+1}, \cdots, b_n)$, $n > m$.
iii- $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n) = \mathbf{E}_{\delta,\sigma}\mathcal{W}^i_m(\mathrm{b}, \cdots, \mathrm{b}, b_{n-m+1}, \cdots, b_n)$, $n < m$, *where* $\mathrm{b} = \lceil QS/D \rceil$.
iv- $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n)$ *is multimodular.*

Proof. The proof holds by using stationarity of the inter-arrival time and the the service time.

i- Using the extended Lindley's Formula, it is clear that $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n)$ is decreasing.

ii- Let us fix n and m with $n > m$. Since the service times are stationary, we can couple the service times such that $\forall 1 \leqslant j \leqslant m,\ \sigma^{(n)}_{j+n-m} = \sigma^{(m)}_j$. (this is merely a shift of the sequence of inter-arrival times for $\mathcal{W}^i_m$). We also couple the intervals in the following way:

$$\forall 1 \leqslant j \leqslant \nu(m),\ \delta^{(n)}_{j+\nu(n-m)} = \delta^{(m)}_j.$$

Under this coupling and because the system is initially empty, we have:

$$\mathcal{W}^i_n(b_1, \cdots, b_n) \quad \geqslant \quad \mathcal{W}^i_m(b_{n-m+1}, \cdots, b_n).$$

Therefore, the inequality holds for the expected values as well: if $n > m$,

$$\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n) \geqslant \mathbf{E}_{\delta,\sigma}\mathcal{W}^i_m(b_{n-m+1}, \cdots, b_n).$$

*iii-*Using the same coupling as for *ii* and using the fact that if, $b_1 \geqslant \mathrm{b}$, then the network is in its original state after the first arrival, we get

$$\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n) = \mathbf{E}_{\delta,\sigma}\mathcal{W}^i_{n+1}(\mathrm{b}, b_{n-m+1}, \cdots, b_n).$$

An easy induction gives the result.

iv- Theorem 6.4 shows that $\mathbf{E}_{\delta}\mathcal{W}^i_n(b_1, \cdots, b_n)$ is multimodular, thus $\mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, \cdots, b_n)$ is multimodular as well. □

Similarly to the previous case, we choose the cost function of any policy $b = (b_1, b_2, \cdots)$ to be the Cezaro sum of the expected traveling time of all customers admitted the queue to a given node i:

$$\gamma^i(b) = \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma}\mathcal{W}^i_n(b_1, ..., b_n).$$

We recall that the balanced policy with phase θ and rate p is

$$b_k^r(\theta) \stackrel{\text{def}}{=} \lfloor kr + \theta \rfloor - \lfloor (k-1)r + \theta \rfloor. \tag{35}$$

We can restate the general theorems given in [3] in the particular case studied here.

Theorem 7.7. *Under the foregoing assumptions, For all sequence b such that*

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} b_n \leqslant r,$$

then $\gamma^i(b) \geqslant \gamma^i(b^r(\theta))$, for any $\theta \in [0,1]$.

Proof. Lemma 7.3 shows that the counting sequence associated with $b_k^r(\theta)$ is a balanced sequence with rate r.

Then, the optimization theory developed in [3] can be applied since all necessary assumptions on $\mathbf{E}_{\sigma,\delta}\mathcal{W}$ are satisfied. □

Note that Lemma 7.3 says that the average workload and the average waiting time are both optimized by the same admission sequence $a^p(\theta)$.

7.4 The unbounded case

If the service times and the inter-arrival times are not bounded in the original event graph $\mathcal{G}$, then, we use a fixed quantity Z and we introduce a new system $\mathcal{G}^Z$ where all services times, σ_n^q are replaced by the sequence $\sigma_n^{Z,q} \stackrel{\text{def}}{=} \min(Z, \sigma_n^q)$. The inter-arrival times δ_k are also replaced by $\delta^Z \stackrel{\text{def}}{=} \max(1/Z, \delta_k)$. In the new system $\mathcal{G}^Z$, the service times are bounded from above by Z and the inter-arrival times are bounded from below by $1/Z$.

In $\mathcal{G}^Z$, $\mathcal{W}_n^{Z,i}$ is the traveling time of the n-th customer to the i-th node.

If Θ is a random variable with uniform distribution on $[0,1)$,

$$b_k^r(\Theta) \stackrel{\text{def}}{=} \lfloor kr + \Theta \rfloor - \lfloor (k-1)r + \Theta \rfloor$$

is called the randomized balanced policy with rate p.

Following the results in [3], the time average of $\mathbf{E}_{\delta,\sigma}\mathcal{W}^Z$, is minimized by the randomized balanced policy, as well by the one with θ fixed (at an arbitrary value). We have for all policy b with rate less than p,

$$\begin{aligned}
\gamma^{Z,i}(b) &= \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma} \mathcal{W}_n^{Z,i}(b_1, ..., b_n) \\
&\geqslant \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma,\Theta} \mathcal{W}_n^{Z,i}(b_1^p(\Theta), ..., b_n^p(\Theta)) \\
&\stackrel{\text{def}}{=} \gamma^{Z,i}(b^p).
\end{aligned}$$

To prove the optimality of the randomized balanced sequence for the original system $\mathcal{G}$ where the service times and the inter-arrival times are unbounded, we need to let Z go to infinity in the previous inequality. For that, we need several lemmas.

Lemma 7.8. *The random sequence $b^r(\Theta)$ with Θ uniform on $[0,1)$ is stationary and $\mathbf{P}(b_k^r(\Theta)=1)=r$.*

Proof. By definition, $b_k^r(\Theta)=\lfloor kq+\Theta\rfloor-\lfloor(k-1)r+\Theta\rfloor$. The fact that $\int_0^1\lfloor x+\Theta\rfloor d\Theta=x$ for any x shows that

$$\forall k,\quad \int_0^1 b_k^q(\Theta)d\Theta=r.$$

□

Lemma 7.9. *If $\{\sigma_n\}$ is a stationary sequence, then $\{\min(\sigma_n,Z)\}$ and $\{\max(\sigma_n,Z)\}$ are also stationary.*

Proof. The proof follows by definition of stationary sequences □

We define the variable $\beta_k\stackrel{\text{def}}{=}\sum_{i=1}^k b_i$. The inter-arrival times satisfy

$$\tau_k=\sum_{i=B_k}^{\beta_{k+1}-1}\delta_i. \tag{36}$$

One sees that τ_k is a sum of a random number of random variables.

Lemma 7.10. *Assume that the process $\{b_k\}$ is stationary , and the process $\{\delta_i\}$ is stationary and independent of $\{b_k\}$. Then, the process $\{\tau_k\}$ is stationary.*

Proof. We compute the distribution of the joint process $\tau_2,\cdots,\tau_k$. It is determined by all the probabilities:

$$\begin{aligned}P(\delta,b)&\stackrel{\text{def}}{=}\mathbf{P}(b_2=n_2,\cdots,b_k=n_k,\delta_{b_1+1}\leqslant d_1,\cdots,\delta_{b_1+n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k})\\&=\sum_i\mathbf{P}(b_1=i,b_2=n_2,\cdots,b_k=n_k,\delta_{i+1}\leqslant d_1,\cdots,\delta_{i+n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k}).\end{aligned}$$

Now,

$$\begin{aligned}P(\delta,b)&=\sum_i\mathbf{P}(b_i=i,b_2=n_2,\cdots,b_k=n_k)\mathbf{P}(\delta_{i+1}\leqslant d_1,\cdots,\delta_{i+n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k})\\&=\sum_i\mathbf{P}(b_1=i,b_2=n_2,\cdots,b_k=n_k)\mathbf{P}(\delta_1\leqslant d_1,\cdots,\delta_{n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k})\\&=\sum_i\mathbf{P}(b_0=i,b_1=n_2,\cdots,b_{k-1}=n_k)\mathbf{P}(\delta_1\leqslant d_1,\cdots,\delta_{n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k})\\&=\mathbf{P}(b_1=n_2,\cdots,b_{k-1}=n_k,\delta_1\leqslant d_1,\cdots,\delta_{n_1+\cdots+n_k}\leqslant d_{n_1+\cdots+n_k}),\end{aligned}$$

where the first equality follows from the independence of b and δ, the second follows from the stationarity of δ and the third from the stationarity of b.

This last expression gives the distribution of the joint process $\tau_1, \cdots \tau_{k-1}$. Since all of this holds for all k, the process $\{\tau_k\}$ is stationary. □

Lemma 7.11. *Under the randomized balanced policy, if $\{\sigma_k^i\}$ and $\{\delta_k\}$ are stationary sequences, then $\{\sigma_k^{Z,i}\}$ and $\{\tau_k^Z\}$ are also stationary.*

Proof. The proof comes from a direct combination of the three previous lemmas. □

Lemma 7.12. *If $\{\sigma_k^i\}$ and $\{\delta_k\}$ are stationary sequences and if the system $\mathcal{G}^Z$ is empty originally, then for all n,*

$$\mathcal{W}_n^{Z,i}(b^p(\Theta)) \leqslant_s \mathcal{W}_{n+1}^{Z,i}(b^p(\Theta)),$$

where $\leqslant_s$ denotes the stochastic order.

Proof. When the service time and the inter-arrival sequence are stationary (as it is the case here), this result is well known (see for example [1]). We can set $\mathcal{W}_0^{Z,i}(b^p(\Theta)) = 0$ and we have $\mathcal{W}_0^{Z,q}(b^p(\Theta)) \leqslant_s \mathcal{W}_1^{Z,i}(b^p(\Theta))$. Now, the Lindley formula for (max,+) systems can be used in this case.

$$\mathcal{W}_k^Z(b^p(\Theta)) = H(k-1) \otimes D(-\tau_{k-1}) \otimes \mathcal{W}_{k-1}^Z(b^p(\Theta)) \oplus B,$$

and

$$\mathcal{W'}_{k+1}^Z(b^p(\Theta)) = H'(k) \otimes D(-\tau'_k) \otimes \mathcal{W'}_k^Z(b^p(\Theta)) \oplus B,$$

where the prime denote another sample path. By induction, we can assume that component-wise, $\mathcal{W}_k^Z(b^p(\Theta)) \geqslant_s \mathcal{W}_{k-1}^Z(b^p(\Theta))$. Now, stationarity of the service times in all queues and of inter-arrival times comes from Lemma 7.11. We can couple the service times and the inter-arrival times such that for all i, $\sigma_k^{i'} = \sigma_{k-1}^i$ and $\tau'_k = \tau_{k-1}$. Under this coupling, $H(k-1) = H'(k)$ and $D(-\tau'_k) = D(-\tau_{k-1})$. Hence

$$\mathcal{W'}_{k+1}^Z(b^p(\Theta)) \geqslant \mathcal{W'}_k^Z(b^p(\Theta))$$

component-wise, under this coupling. □

The optimality of the randomized balanced policy is established by the following theorem.

Theorem 7.13. *Consider the system $\mathcal{G}$ with stationary inter-arrival times and stationary service times in all queues. For all policy b such that*

$$\overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} b_n \leqslant p,$$

then $\gamma^i(b) \geqslant \gamma^i(b^p)$.

Proof. Remark that the quantity $\mathcal{W}_k^{Z,i}(b)$ is increasing in Z and

$$lim_{Z\to\infty}\mathcal{W}_k^{Z,i}(b) = \sup_Z \mathcal{W}_k^{Z,i}(b) = \mathcal{W}_k^i(b).$$

The proof follows from this series of inequalities.

$$\begin{aligned}
\gamma^i(b) &= \overline{\lim_{N\to\infty}} \sup_Z \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma} \mathcal{W}_n^{Z,i}(b_1, ..., b_n) \right) \\
&\geqslant \sup_Z \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma} \mathcal{W}_n^{Z,i}(b_1, ..., b_n) \\
&\geqslant \sup_Z \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma,\Theta} \mathcal{W}_n^{Z,i}(b_1^p(\Theta), ..., b_n^p(\Theta)).
\end{aligned}$$

By Lemma 7.12, we know that $\mathbf{E}_{\delta,\sigma,\Theta}\mathcal{W}_n^{Z,i}(b_1^p(\Theta), ..., b_n^p(\Theta))$ is increasing in n. Therefore, the Cezaro limit equals the supremum on all n. We continue the previous inequalities:

$$\begin{aligned}
\gamma^i(b) &\geqslant \sup_Z \sup_n \mathbf{E}_{\delta,\sigma,\Theta} \mathcal{W}_n^{Z,i}(b_1^p(\Theta), ..., b_n^p(\Theta)) \\
&= \sup_n \sup_Z \mathbf{E}_{\delta,\sigma,\Theta} \mathcal{W}_n^{Z,i}(b_1^p(\Theta), ..., b_n^p(\Theta)) \\
&= \overline{\lim_{N\to\infty}} \frac{1}{N} \sum_{n=1}^{N} \mathbf{E}_{\delta,\sigma,\Theta} \mathcal{W}_n^{i}(b_1^p(\Theta), ..., b_n^p(\Theta)) \\
&= \gamma^i(b^p).
\end{aligned}$$

□

8 Conclusion

We investigated in this paper the admission control problem in a general setting. We were motivated by the pioneering work of Hajek [9], who solved the admission control problem into a single queue in a Markovian framework. We study the admission into a general network having a (max,+) linear structure, under a general probabilistic framework: only stationarity and ergodicity assumptions are used instead of a Markovian setting.

Structural properties of the waiting time of any (max,+) linear system are established. In particular, we show that the waiting time seen as a function of the admission sequence is multimodular. Multimodularity can be seen as the "integer" counterpart of convexity. From that point on, we can use general theorems that we developed in [3] to describe the optimal admission policy among all admission sequence with a given long range acceptance range.

In a forthcoming paper [2], we use again the framework established in [3] to another class of optimal control problems: the routing in systems composed of several (max,+) linear systems.

Acknowledgment

The authors would like to thank François Baccelli for his help in Section 7.4 concerning stationarity issues.

References

[1] F. Baccelli abd Z. Liu. On a class of stochastic recursive equations arising in queueing theory. *Annals of Probability*, 21(1):350–374, 1992.

[2] Eitan Altman, Bruno Gaujal, and Arie Hordijk. Balanced sequences and optimal routing. Technical report, INRIA, May 1997.

[3] Eitan Altman, Bruno Gaujal, and Arie Hordijk. Multimodularity, convexity and optimization properties. Technical report, INRIA, May 1997.

[4] F. Baccelli, G. Gohen, G.J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity.* Wiley, 1992.

[5] François Baccelli. Ergodic theory of stochastic Petri networks. *Annals of Probability*, 20(1):375–396, 1992.

[6] François Baccelli and Volker Schmidt. Taylor series expansions for poisson driven (max,+) linear systems. *The annals of Applied Probability*, 6(1):138–185, 1996.

[7] E.M. Coven and G.A. Hedlund. Sequences with minimal block growth. *Math. System Theory*, 7(2), 1973.

[8] Jörg Desel and Javier Esparza. *Free Choice Petri Nets.* Cambridge Tracts in Theorical Computer Science, 1995.

[9] Bruce Hajek. Extremal splittings of point processes. *Mathematics of Operation research*, 10(4), November 1985.

[10] H. Hillion and J.-M. Proth. Performance evaluation of job shop systems using timed event graphs. *IEEE Transaction on Automatic Control*, 34(1):3–9, 1989.

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

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399