Evènements rares dans les réseaux.

Marc Lelarge

Evénements rares dans les réseaux

THÈSE

présentée et soutenue publiquement le 7 février 2005

pour l'obtention du

Doctorat de l'Ecole Polytechnique

(spécialité Mathématiques Appliquées)

par

M. Marc Lelarge

Composition du jury

Rapporteurs :	Prof. Venkat Anantharam	University of California Berkeley
	Prof. Sem Borst	CWI & Bell Labs
	Prof. Neil O'Connell	University College Cork
Examinateurs :	Prof. François Baccelli	INRIA-ENS (directeur de thèse)
	Prof. Philippe Bougerol	Université Paris 6
	Prof. Serguei Foss	Heriot-Watt University
	Dr. Laurent Massoulié	Microsoft
	Prof. David McDonald	Université d'Ottawa

Mis en page avec la classe thloria.

Je tiens tout d’abord à remercier François Baccelli pour son soutien constant et ses conseils toujours judicieux. Ce fut pour moi un grand honneur et un réel plaisir de découvrir le monde de la recherche au sein de son groupe.

I wish to thank Serguei Foss for his contagious enthusiasm. He never refused me his help and advice. This thesis owes a great debt to these inspiring researchers.

I wish to thank Venkat Anantharam, Sem Borst and Neil O’Connell for accepting to read a first version of this thesis and writing a report.

Je tiens à remercier Philippe Bougerol pour avoir accepté de présider mon jury ainsi que Laurent Massoulié et David McDonald pour avoir accepté de faire partie de ce jury.

Je remercie tous ceux qui m’ont supporté et aidé pendant ce travail et en particulier un grand merci à Dohy Hong pour la joyeuse et studieuse ambiance qu’il a su maintenir dans notre bureau S1. Un grand merci aussi à Emmanuel, Charles, Augustin, Bartek, Pierre, Ki-Baek pour les bons moments passés à l’ENS...

Table des matières

Chapitre 1

Introduction

Nous nous intéressons dans cette thèse à l'étude d'événements rares dans des réseaux de communication. Dans un premier temps, nous introduisons la classe des réseaux monotones séparables qui nous permettra une analyse systématique de réseaux de grande dimension. Parmi ceux-ci, nous appliquerons notre théorie en détail aux réseaux (max,plus)-linéaires et aux réseaux de Jackson généralisés.

La première étape de notre étude consiste à comprendre la dynamique de ces réseaux. Nous décrivons leur comportement fluide, ce qui permet d'écrire les conditions de stabilité du réseau et de construire les variables d'état (telles que temps d'attente aux différentes stations, tailles des files d'attente...) dans leur régime stationnaire.

L'étude trajectorielle du réseau nous permet ensuite de comprendre le comportement aléatoire du réseau. Nous calculons les asymptotiques des probabilités d'événements rares (dont la probabilité tend vers 0) et décrivons "comment" ces événements se produisent. Nous montrons que le "comportement" du réseau est radicalement différent selon les hypothèses probabilistes faites sur les temps de service.

Dans le cas de distributions sous-exponentielles, l'événement rare est dû à un unique grand service qui bloque une station du réseau tandis que dans le cas de distributions à queue exponentielle, l'événement rare est dû à une conjonction de nombreux temps de services anormalement longs. Ces heuristiques sont rendues précises par les calculs des probabilités considérées. Dans un dernier temps, nous étudions l'impact d'une structure de dépendance entre les différents temps de service grâce au mouvement Brownien fractionnaire.

Cadre général

Dans ce chapitre, nous introduisons le cadre monotone séparable qui a été développé par François Baccelli et Serguei Foss [13]. Les principales propriétés de ces réseaux sont rappelées, en particulier la condition de stabilité d'un réseau monotone séparable est connue sous des hypothèses probabilistes générales (stationarité et ergodicité des processus d'entrée).

Nous donnons des conditions naturelles sous lesquelles un réseau (max,plus)-linéaire appartient à cette classe. Nous montrons que les graphes d'événements utilisés pour modéliser des réseaux de communication ainsi que les mécanismes de synchronisation de certains protocoles font partie de cette classe.

Concernant les réseaux de Jackson généralisés, nous écrivons les équations d'évolution sous la forme d'une équation de point fixe dans un espace fonctionnel. Cette écriture permet d'avoir un cadre unifié permettant de décrire un système ayant une dynamique discrète ou une dynamique fluide. Dans le premier cas, les fonctions considérées sont des fonctions de comptage tandis que dans le second cas, les fonctions sont simplement supposées croissantes.

Modèles fluides

Dans ce chapitre, nous nous intéressons au comportement limite du réseau lorsque le temps est accéléré par un facteur n tandis que la variable d'espace (i.e. la taille des files d'attente ou la charge des stations) est divisée par ce même facteur n.

Dans le cas des réseaux de Jackson généralisés, nous utilisons des propriétés de monotonie et de convexité des operateurs définissant l'équation de point fixe pour caractériser la limite fluide. Dans un cadre aléatoire stationnaire ergodique, ces limites fluides correspondent à des lois fortes des grands nombres. Nous faisons le lien entre ce calcul de limite fluide et la condition de stabilité du réseau.

Nous étudions ensuite le cas de files de type GPS (Generalized Processor Sharing). Différentes files d'attente se partagent un serveur de manière égalitaire. Si toutes les files sont pleines, chacune reçoit une proportion de la capacité du serveur. Si l'une des files est vide, la capacité normalement allouée à cette file est redistribuée parmi les files occupées. Nous calculons la limite fluide du système dans le cas où le système global est instable. Nous montrons que certaines files peuvent cependant rester stables et nous caractérisons l'ensemble de ces files.

Asymptotiques sous-exponentielles

Dans ce chapitre, nous étudions le comportement du réseau dans le cas où la distribution des temps de service dans chaque station est sous-exponentielle. Sous cette hypothèse, une méthodologie générale a éte développée par François Baccelli et Serguei Foss [14] pour étudier les asymptotiques de réseaux monotones séparables. La forme générale de ces asymptotiques est donnée mais les constantes doivent être calculées au cas par cas.

L'idée générale est que l'événement rare se produit selon un événement typique : si la charge du réseau est grande au temps 0, ceci est dû à un grand service qui a bloqué une station à un moment dans le passé. Hormis ce temps de service exceptionnellement long, le réseau se comporte ”normalement”, en particulier il est bien approximé par sa limite fluide.

Dans le cas de réseaux (max,plus)-linéaires, cette limite fluide est bien connue et donnée par les exposants de Lyapunov. Ceci nous permet d'exprimer les constantes des asymptotiques sous-exponentielles en fonction de ces exposants. Dans le cas des réseaux de Jackson généralisés et des systèmes de type GPS, les calculs du chapitre précédent permettent de conclure.

Grandes déviations

Dans ce chapitre, nous traitons le cas où les distributions des temps de service ont des queues exponentielles. Ce cas est complémentaire du chapitre précédent et les techniques probabilistes sont différentes.

La première étape consiste à développer une méthodologie générale pour ces hypothèses probabilistes. Un processus sous-additif est naturellement associé à un réseau monotone séparable (de la même manière qu'une marche aléatoire est naturellement associée à une file d'attente avec un serveur). Nous avons d'abord montré que le principe de grande déviation correspondant au cas du maximum d'une marche aléatoire s'étend au cas du maximum d'un processus sous-additif. Ceci nous permet d'obtenir un résultat général pour tout réseau monotone séparable en fonction d'une transformée de Laplace asymptotique.

Dans le cas des réseaux (max,plus)-linéaires, les propriétés d'idempotence de l'algèbre (max,plus) nous permettent d'exprimer cette transformée de Laplace en fonction des différentes composantes du réseau. Dans le cas des réseaux de Jackson, nous donnons le principe de grandes déviations trajectorielles du processus de la longueur des files d'attente à chaque station. Ce résultat est original et étend le seul cas connu correspondant à des temps de service exponentiels. En particulier, la preuve utilise une extension du principe de contraction qui peut avoir des applications à d'autres systèmes ou réseaux.

Asymptotiques pour des réseaux (max,plus)-linéaires browniens fractionnaires

Ce dernier chapitre traite le cas où les temps de service ont des queues exponentielles mais contrairement aux deux chapitres précédents la suite des temps de service à chaque station a une structure de dépendance. En particulier, nous calculons l'effet d'une dépendance à long terme (observée empiriquement dans le trafic internet) sur les performances générales d'un réseau (max,plus)-linéaire.

Overview

The goal of this thesis is the study of rare events in stochastic networks. What we call rare events are events with very small probability. The one dimensional example of such an event is the tail of the stationary workload W of a stable single server queue,

$$\mathbb{P}(W > x) \quad \text{as } x \to \infty.$$

This is exactly the kind of asymptotics we want to study in a network setting. Dealing with networks instead of single server queue means that we have now a multidimensional object to understand. This naturally raises intricate mathematical problems and a problem of methodology too. The range of interesting networks one can build from very simple bricks is now exploding. One has to find proofs that are sufficiently systematic to cover a whole set of networks. If one finds a very suitable technique for a very specific brick, there is little hope that his technique will extend to a non-negligible subset of the possible networks !

To avoid this kind of annoyance, we chose another approach. We first study the general properties of a set of networks, namely the set of monotone separable networks. This class has enough structure to enable us to derive general properties for various networks. Knowing if this class of networks is negligible is then more a matter of philosophy... anyway it covers several classical networks !

In Chapter 2, we present the general framework of monotone separable networks and three subclasses : (max,plus)-linear systems, generalized Jackson networks and generalized processor sharing (GPS) queues. The class of monotone separable networks was first introduced by François

Baccelli and Serguei Foss to study stability of such networks [13]. In particular they constructed the stationary version of generalized Jackson networks in [12]. With the first section of Chapter 3, the dynamic of such networks is quite well understood. In the second section of this chapter, we construct the stationary regime of a GPS system. The new part consisting in the study of the overloaded system, we show that even if the whole system is unstable, there exist subsystems that are stable. In Chapter 4, we address the core of the problem, and thanks to results of previous chapters, we are able to derive subexponential asymptotics for these three classes of networks. The technique used has been proposed by Baccelli and Foss in [14]. Thanks to the three cases we explore, we show both the power of the method and its limits. We end the chapter with some thoughts to generalize it.

Chapters 5 and 6 are quite independent. We study the same objects, namely monotone separable networks, but under different stochastic assumptions. The techniques used in these chapters are completely different from previous chapter. Chapter 5 deals with standard large deviations theory and show that the monotone separable framework may be well suited for such large deviations studies. In Chapter 6, we study the impact of correlation between successive service times at a same station. The study of fractional Brownian motion enable us to get some results in this direction.

Chapitre 2

General Framework

2.1 Monotone Separable Networks

2.1.1 Framework

The framework described in this section has been developed by François Baccelli and Serguei Foss and results of this section can be found in [13], [10] and [14].

Consider a stochastic network described by the following framework :

– The network has a single input point process N, with points $\{T_n\}_{-\infty<n<\infty}$; for all $m \leq n \in \mathbb{Z}$, let $N_{[m,n]}$ be the $[m, n]$ restriction of N, namely the point process with points $\{T_\ell\}_{m\leq\ell\leq n}$.
– The network has a.s. finite activity for all finite restrictions of N : for all $m \leq n \in \mathbb{Z}$, let $X_{[m,n]}(N)$ be the time of last activity in the network, when this one starts empty and is fed by $N_{[m,n]}$. We assume that for all finite m and n as above, $X_{[m,n]}$ is finite.

We assume that there exists a set of functions $\{f_\ell\}$, $f_\ell : \mathbb{R}^\ell \times K^\ell \to \mathbb{R}$, such that :

$$X_{[m,n]}(N) = f_{n-m+1}\{(T_\ell, \zeta_\ell),\ m \leq \ell \leq n\}, \tag{2.1}$$

for all n, m and N, where the sequence $\{\zeta_n\}$ is that describing service times and routing decisions.

We say that a network described as above is monotone-separable if the functions f_n are such that the following properties hold for all N :

1. **Causality :** for all $m \leq n$,
$$X_{[m,n]}(N) \geq T_n;$$
2. **External monotonicity :** for all $m \leq n$,
$$X_{[m,n]}(N') \geq X_{[m,n]}(N),$$
whenever $N' := \{T'_n\}$ is such that $T'_n \geq T_n$ for all n, a property which we will write $N' \geq N$ for short ;
3. **Homogeneity :** for all $c \in \mathbb{R}$ and for all $m \leq n$
$$X_{[m,n]}(N + c) = X_{[m,n]}(N) + c;$$
4. **Separability :** for all $m \leq \ell < n$, if $X_{[m,\ell]}(N) \leq T_{\ell+1}$, then
$$X_{[m,n]}(N) = X_{[\ell+1,n]}(N).$$

2.1.2 Maximal Dater

By definition, for $m \leq n$, the $[m,n]$ maximal dater is

$$Z_{[m,n]}(N) := X_{[m,n]}(N) - T_n = X_{[m,n]}(N - T_n).$$

Note that $Z_{[m,n]}(N)$ is a function of $\{\zeta_\ell\}_{m \leq \ell \leq n}$ and $\{\tau_\ell\}_{m \leq \ell \leq n}$ only, where $\tau_n = T_{n+1} - T_n$. In particular, $Z_n := Z_{[n,n]}(N)$ is a function of ζ_n only and does not depend on $\{\tau_\ell\}_{-\infty < \ell < \infty}$.

Lemma 1. Internal monotonicity of X and Z
Under the above conditions, the variables $X_{[m,n]}$ and $Z_{[m,n]}$ satisfy the internal monotonicity property : for all N, $m \leq n$,

$$\begin{aligned} X_{[m-1,n]}(N) &\geq X_{[m,n]}(N), \\ Z_{[m-1,n]}(N) &\geq Z_{[m,n]}(N). \end{aligned}$$

In particular, the sequence $\{Z_{[-n,0]}(N)\}$ is non-decreasing in n. Put

$$Z := Z_{(-\infty,0]}(N) = \lim_{n \to \infty} Z_{[-n,0]}(N) \leq \infty.$$

Lemma 2. Sub-additive property of Z
Under the above conditions, $\{Z_{[m,n]}\}$ satisfies the following sub-additive property : for all $m \leq \ell < n$, for all N,

$$Z_{[m,n]}(N) \leq Z_{[m,\ell]}(N) + Z_{[\ell+1,n]}(N).$$

2.1.3 Stationary Ergodic Setting and Main Stability Results

Assume the variables $\{\tau_n, \zeta_n\}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P}, \theta)$, where θ is an ergodic, measure-preserving shift transformation, such that $(\tau_n, \zeta_n) \circ \theta = (\tau_{n+1}, \zeta_{n+1})$. The following integrability assumptions are also assumed to hold :

$$\mathbb{E}[\tau_n] := \lambda^{-1} := a < \infty, \quad \mathbb{E}[Z_n] < \infty.$$

Lemma 3. *Under the foregoing ergodic assumptions, either $Z = \infty$ a.s. or $Z < \infty$ a.s.*

The network is stable if $Z < \infty$ a.s. and unstable otherwise.

Denote by Q the degenerate input process with all its points equal to 0 : $T_n(Q) = 0$ for all n. In view of Lemma 33, the Kingman's sub-additive ergodic theorem gives :

Lemma 4. *Under the foregoing ergodic assumption, there exists a non-negative constant $\gamma(0)$ such that*

$$\lim_{n \to \infty} \frac{Z_{[-n,-1]}(Q)}{n} = \lim_{n \to \infty} \frac{\mathbb{E}\left[Z_{[-n,-1]}(Q)\right]}{n} = \gamma(0) \; a.s.$$

The main result on the stability region will be proved in the next section :

Theorem 1. *(a) If $\lambda\gamma(0) < 1$, then $Z < \infty$ a.s.*

(b) If $Z < \infty$ a.s., then $\lambda\gamma(0) \leq 1$.

2.1.4 Upper $G/G/1/\infty$ Queue and Lower Bound for the maximal Dater

We first derive a lower bound that will give us part (b) of Theorem 12 and will be useful in the large deviation analysis of monotone separable networks.

Proposition 1. *We have the following lower bound*

$$Z \geq \sup_{n\geq 0} \left(Z_{[-n,0]}(Q) + T_{-n} - T_0 \right).$$

Proof.
For n fixed, let N^n be the point process with point $T_j^n = T_{-n} - T_0$, for all j. Then

$$\begin{aligned} Z_{[-n,0]} &= X_{[-n,0]}(N) - T_0 \geq X_{[-n,0]}(N^n) \\ &= X_{[-n,0]}(Q) + T_{-n} - T_0 = Z_{[-n,0]}(Q) + T_{-n} - T_0, \end{aligned}$$

where we used external monotonicity in the first inequality and homogeneity between the first and second line. □

Proof of Theorem 12 part (b).
Suppose that $\lambda\gamma(0) > 1$, then we have

$$\liminf_{n\to\infty} \frac{Z_{[-n,0]}(N)}{n} \geq \gamma(0) - a > 0,$$

which concludes the proof of part (b). □

We assume now that $\gamma(0) < a$. We pick an integer $L \geq 1$ such that

$$\mathbb{E}\left[Z_{[-L,-1]}(Q) \right] < La, \tag{2.2}$$

which is possible in view of Lemma 34. Without loss of generality, we assume that $T_0 = 0$. Part (a) of Theorem 12 will follow from the following proposition :

Proposition 2. *The stationary maximal dater Z is bounded from above by the stationary response time $\hat{R}$ in the $G/G/1/\infty$ queue with service times*

$$\hat{s}_n := Z_{[L(n-1)+1,Ln]}(Q)$$

and inter-arrival times $\hat{\tau}_n := T_{Ln} - T_{L(n-1)}$, where L is the integer defined in (5.17). Since $\mathbb{E}[\hat{s}_1] < \mathbb{E}[\hat{\tau}_1] = La$, this queue is stable. With the convention $\sum_0^{-1} = 0$, we have,

$$Z \leq \hat{s}_0 + \sup_{k\geq 0} \sum_{i=-k}^{-1} (\hat{s}_i - \hat{\tau}_{i+1}) =: \hat{R}.$$

Proof.

To an input process N, we associate the following upper bound process, $N^+ = \{T_n^+\} \geq N$, where $T_n^+ = T_{kL}$ if $n = (k-1)L+1, \dots, kL$. Then for all n, since we assumed $T_0 = 0$, we have thanks to the external monotonicity,

$$X_{[-n,0]}(N) = Z_{[-n,0]}(N) \leq X_{[-n,0]}(N^+) = Z_{[-n,0]}(N^+). \tag{2.3}$$

We show that for all $k \geq 1$,

$$Z_{[-kL+1,0]}(N^+) \leq \hat{s}_0 + \sup_{-k+1\leq i\leq 0} \sum_{j=-i}^{-1} (\hat{s}_j - \hat{\tau}_{j+1}). \tag{2.4}$$

This inequality will follow from the two next lemmas

Lemma 5. *Assume $T_0 = 0$. For any $m < n \leq 0$,*

$$Z_{[m,0]}(N) \leq Z_{[n,0]}(N) + (Z_{[m,n-1]}(N) - \tau_{n-1})^+.$$

Proof of Lemma 35.

Assume first that $Z_{[m,n-1]}(N) - \tau_{n-1} \leq 0$, which is exactly $X_{[m,n-1]}(N) \leq T_n$. Then by the separability property, we have

$$Z_{[m,0]}(N) = X_{[m,0]}(N) = X_{[n,0]}(N) = Z_{[n,0]}(N).$$

Assume now that $Z_{[m,n-1]}(N) - \tau_{n-1} > 0$. Let $N' = \{T'_j\}$ be the input process defined as follows

$$\begin{aligned} \forall j \leq n-1, \quad T'_j &= T_j, \\ \forall j \geq n, \quad T'_j &= T_j + Z_{[m,n-1]}(N) - \tau_{n-1}. \end{aligned}$$

Then we have $N' \geq N$ and $X_{[m,n-1]}(N') \leq T'_n$, hence by the external monotonicity, the separability and the homogeneity properties, we have

$$\begin{aligned} Z_{[m,0]}(N) &= X_{[m,0]}(N) \leq X_{[m,0]}(N') \\ &= X_{[n,0]}(N') = X_{[n,0]}(N) + Z_{[m,n-1]}(N) - \tau_{n-1} = Z_{[n,0]}(N) + Z_{[m,n-1]}(N) - \tau_{n-1}. \end{aligned}$$

□

From this lemma we derive directly

Lemma 6. *Assume $T_0 = 0$. For any $n < 0$,*

$$Z_{[n,0]}(N) \leq \sup_{n \leq k \leq 0} \left(\sum_{i=k}^{-1} (Z_i - \tau_{i+1}) \right) + Z_0,$$

with the convention $\sum_0^{-1} = 0$

Applying Lemma 36 to $Z_{[-kL+1,0]}(N^+)$ gives (5.19). We now return to the proof of Proposition 22. We have

$$\begin{aligned} Z &= \lim_{k \to \infty} Z_{[-kL+1,0]} \\ &= \sup_{k \geq 0} Z_{[-kL+1,0]}(N) \\ &\leq \sup_{k \geq 0} Z_{[-kL+1,0]}(N^+) \quad \text{thanks to (5.18)} \\ &\leq \sup_{k \geq 0} \left(\hat{s}_0 + \sup_{-k+1 \leq i \leq 0} \sum_{j=-i}^{-1} (\hat{s}_j - \hat{\tau}_{j+1}) \right) = \hat{R}, \quad \text{thanks to (5.19).} \end{aligned}$$

from Lemma 36. □

2.2 (max,plus)-Linear Systems and Event Graphs

2.2.1 (max,plus)-Linear Systems

Most of the material of this section is taken from the book *Synchronization and Linearity* [11]. Some notations are taken from [16].

Definition 1. *The (max,plus) semi-ring* $\mathbb{R}_{\max}$ *is the set* $\mathbb{R} \cup \{-\infty\}$*, equipped with* $\max$*, written additively (i.e.* $a \oplus b = \max(a, b)$*) and the usual sum, written multiplicatively (i.e.* $a \otimes b = a + b$*). The zero element is denoted* $\epsilon = -\infty$*.*

For matrices of appropriate sizes, we define $(A \oplus B)^{(i,j)} = A^{(i,j)} \oplus B^{(i,j)} = \max(A^{(i,j)}, B^{(i,j)})$, $(A \otimes B)^{(i,j)} = \bigoplus_k A^{(i,k)} \otimes B^{(k,j)} = \max_k(A^{(i,k)} + B^{(k,j)})$.

Let s be an arbitrary fixed natural number. Assume the following to be given :
– $\{T_n, n \in \mathbb{N}\}$, where $T_n \in \mathbb{R}$, the arrival time sequence ;
– $\{A_n, n \in \mathbb{N}\}$, where A_n is a $s \times s$ matrix ;
– $\{B_n, n \in \mathbb{N}\}$, where B_n, is a $s-$dimensional vector.
The associated (max,plus)-linear recurrence is that with state variable sequence $\{X_n, n \in \mathbb{N}\}$, where X_n is a $s-$dimensional vector, which satisfies the evolution equation :

$$X_{n+1} = A_{n+1} \otimes X_n \oplus B_{n+1} \otimes T_{n+1}. \tag{2.5}$$

We assume w.l.o.g. that A_n has no null column (= $(\epsilon \ldots \epsilon)'$) and that if the i-th line of A_n is null, then $B_n^{(i)} \geq 0$.

To each (max,plus)-linear recurrence, one associates a network in the sense of the last section, with $\zeta_n = (A_n, B_n)$ and

$$X_{[m,n]}(N) = \bigoplus_{1 \leq i \leq s} \bigoplus_{m \leq k \leq n} (D_{[k+1,n]} \otimes B_k \otimes T_k)^{(i)},$$

where for $k < n$, $D_{[k+1,n]} = \bigotimes_{j=n}^{k+1} A_j = A_n \otimes \cdots \otimes A_{k+1}$ and $D_{[n+1,n]} = E$, the identity matrix (the matrix with all its diagonal entries equal to 0 and all its non-diagonal ones equal to ϵ). If one defines

$$Y_{[m,n]} = \bigoplus_{m \leq k \leq n} D_{[k+1,n]} \otimes B_k \otimes T_k,$$

it is easy to check that $Y_{[m,m]} = B_m \otimes T_m$, that for all $n \geq m$,

$$Y_{[m,n+1]} = A_{n+1} \otimes Y_{[m,n]} \oplus B_{n+1} \otimes T_{n+1}$$

and that $X_{[m,n]}(N) = \max_i(Y_{[m,n]})^{(i)}$.

We denote by $\mathbf{0}$ the vector with all its entries equal to 0.

Lemma 7. *The network associated with a (max,plus)-linear recurrence is monotone-separable provided* $A_n \otimes \mathbf{0} \leq B_n \oplus \mathbf{0}$ *for all* n*.*

Proof.

The first three properties are immediate. Let us prove that separability holds under the last assumption. If $X_{[m,l]}(N) \leq T_{l+1}$, then $Y_{[m,l]} \leq \mathbf{0} \otimes T_{l+1}$.

So by monotonicity,

$$\begin{aligned} A_{l+1} \otimes Y_{[m,l]} &\le A_{l+1} \otimes \mathbf{0} \otimes T_{l+1} \\ &\le B_{l+1} \otimes T_{l+1} \oplus \mathbf{0} \otimes T_{l+1}. \end{aligned}$$

Hence we have

$$\begin{aligned} A_{l+1} \otimes Y_{[m,l]} \oplus B_{l+1} \otimes T_{l+1} &\le B_{l+1} \otimes T_{l+1} \oplus \mathbf{0} \otimes T_{l+1} \\ Y_{[m,l+1]} &\le Y_{[l+1,l+1]} \oplus \mathbf{0} \otimes T_{l+1}. \end{aligned} \tag{2.6}$$

But $\max_i B_{l+1}^{(i)} \ge 0$, hence we have $\max_i Y_{[l+1,l+1]}^{(i)} \ge T_{l+1}$. And then

$$X_{[m,l+1]}(N) = \max_i Y_{[m,l+1]}^{(i)} \le \max_i Y_{[l+1,l+1]}^{(i)} = X_{[l+1,l+1]}(N).$$

We show by induction that for all $n \ge l+1$,

$$Y_{[m,n]} \le Y_{[l+1,n]} \otimes \mathbf{0} \otimes T_{l+1}. \tag{2.7}$$

In view of (5.24), it is true for $n = l+1$. Suppose it is true for n, then we have by monotonicity,

$$\begin{aligned} A_{n+1} \otimes Y_{[m,n]} &\le A_{n+1} \otimes Y_{[l+1,n]} \oplus B_{n+1} \otimes T_{l+1} \oplus \mathbf{0} \otimes T_{l+1} \\ Y_{[m,n+1]} &\le Y_{[l+1,n+1]} \oplus \mathbf{0} \otimes T_{l+1}, \quad \text{since } T_{n+1} \ge T_{l+1}. \end{aligned}$$

Now taking the maximum over the indices in (5.25) gives $X_{[m,n]}(N) \le X_{[l+1,n]}(N)$, but the converse inequality is clearly true in view of the definition of the mapping $X(.)$. Hence we have finally

$$X_{[m,n]}(N) = X_{[l+1,n]}(N).$$

□

In this case we can define the maximal dater

$$Z = \bigoplus_{1 \le i \le s} \bigoplus_{k \le 0} \left(D_{[k+1,0]} \otimes B_k \otimes T_k\right)^{(i)} - T_0$$

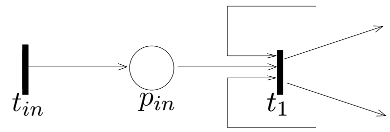
2.2.2 Event Graphs

In this section, we first describe what we define as an event graph and then show that these objects belong to the class of (max,plus)-linear systems and under some additional assumptions to the class of monotone separable networks.

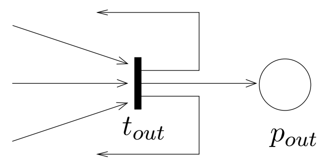
Consider a bipartite oriented graph $\mathcal{G}$ with two types of nodes : transitions (denoted by bars) and places (denoted by circles), and with an integer marking of each place. We will only consider the class of *event graphs*, which is the class of such bipartite graphs where each place has exactly one upstream and one downstream transition. An example of such a graph is provided below where the integer marking of a place (here 0 or 1) is depicted by tokens. We will also assume that the event graph is live, namely that there is no circuit with only places of zero marking.

A transition without predecessor is called a source ; similarly a transition with no successor is called a sink ; we will consider networks that have exactly one source and one sink and we will adopt the following notation :

– For the source :



– For the sink :



Consider an event graph, together with ($\mathcal{T}$ denote the set of transitions) :
– a sequence of non-negative, real variables σ_n^t, $t \in \mathcal{T}$, $n \geq 0$;
– an increasing sequence of real variables T_n, $n \geq 0$.
We show below that to such a triple, one can associate a (max,plus)-linear recurrence of type (2.5).

For this, take $K = |\mathcal{T}|$, and we identify $\mathcal{T}$ with $\{1, \ldots, K\}$. We adopt a numbering of coordinates such that coordinate 1 is the source and K the sink. For all $m = 0, \ldots, L$, where L is the maximal value of the initial marking, define $a_m(n)$ to be the $K \times K$ matrix with entries

$$(a_m(n))^{(i,j)} = \begin{cases} \sigma_n^i & \text{if there is two hop path from } j \text{ to } i \text{ with a place with marking } m \\ \varepsilon & \text{otherwise.} \end{cases} \tag{2.8}$$

Let b the K-dimensional vector with all its entries equal to ε, but the first, which is equal to 0. Let then x_n be the sequence of K-dimensional vectors defined by the recurrence relation

$$x_n = a_0(n) \otimes x_n \oplus \cdots \oplus a_L(n) \otimes x_{n-L} \oplus b \otimes T_n. \tag{2.9}$$

The reduction to a (max,plus)-recurrence is then obtained as follows : the matrix a_0 can be assumed to be strictly triangular w.l.o.g. thanks to the liveness assumption (see [11]). Therefore the matrix

$$a_0(n)^* = E \oplus a_0(n) \oplus a_0(n)^2 \oplus \ldots$$

is well defined and when defining $\overline{a}_i(n) = a_0(n)^* \otimes a_i(n)$ and $\overline{b}(n) = a_0(n)^* \otimes b$, we obtain

$$x_n = \overline{a}_1(n) \otimes x_{n-1} \oplus \cdots \oplus \overline{a}_L(n) \otimes x_{n-L} \oplus \overline{b}(n) \otimes T_n. \tag{2.10}$$

Then, with the following notation

$$X_n = \begin{pmatrix} x_{n-L+1} \\ \vdots \\ x_n \end{pmatrix},$$

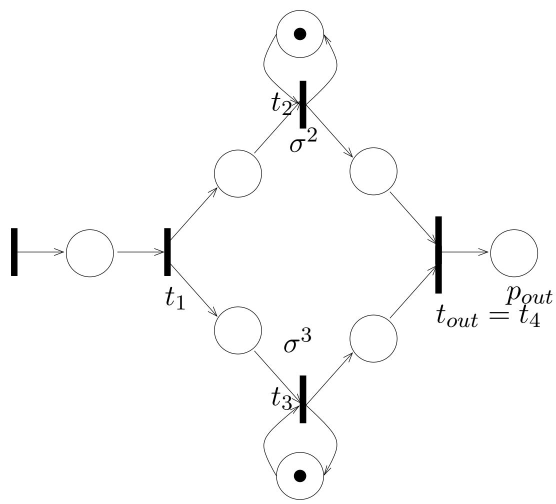
we get the desired equation, namely $X_n = A_n X_{n-1} \oplus B_n T_n$, when taking

$$A_n = \begin{pmatrix} \epsilon & E & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \epsilon & & & & E \\ \overline{a}_L(n) & \overline{a}_{L-1}(n) & \ldots & \ldots & \overline{a}_1(n) \end{pmatrix}, \qquad B_n = \begin{pmatrix} \epsilon \\ \vdots \\ \vdots \\ \epsilon \\ \overline{b}(n) \end{pmatrix}. \tag{2.11}$$

So to each event graph, one can associate a (max,plus)-linear recurrence and therefore a network.

Remark 1. One can drop coordinate i if column i has only ϵ entries (indeed, in this case coordinate i is never used in the recursion). We can drop coordinates successively. We will not do this for the last column, which is associated to the last activity.

Here is an example. Consider the following graph :



and take all sigma's equal to 0 but for transitions 2 and 3 for which we take some sequences σ_n^2 and σ_n^3 respectively. Here, $K = 4$, $L = 1$ and the matrices are

$$a_0(n) = \begin{pmatrix} \epsilon & \epsilon & \epsilon & \epsilon \\ \sigma_n^2 & \epsilon & \epsilon & \epsilon \\ \sigma_n^3 & \epsilon & \epsilon & \epsilon \\ \epsilon & 0 & 0 & \epsilon \end{pmatrix}, \qquad a_1(n) = \begin{pmatrix} \epsilon & \epsilon & \epsilon & \epsilon \\ \epsilon & \sigma_n^2 & \epsilon & \epsilon \\ \epsilon & \epsilon & \sigma_n^3 & \epsilon \\ \epsilon & \epsilon & \epsilon & \epsilon \end{pmatrix}.$$

The evolution equations are :

$$\begin{aligned} x_n^{(1)} &= T_n, \\ x_n^{(2)} &= [x_n^{(1)} \oplus x_{n-1}^{(2)}] \otimes \sigma_n^2, \\ x_n^{(3)} &= [x_n^{(1)} \oplus x_{n-1}^{(3)}] \otimes \sigma_n^3, \\ x_n^{(4)} &= x_n^{(2)} \oplus x_n^{(3)}. \end{aligned}$$

Denoting $\sigma_n^{i \vee j} = \max(\sigma_n^i; \sigma_n^j)$, we get :

$$\begin{aligned} x_n^{(1)} &= T_n, \\ x_n^{(2)} &= x_{n-1}^{(2)} \otimes \sigma_n^2 \oplus T_n \otimes \sigma_n^2, \\ x_n^{(3)} &= x_{n-1}^{(3)} \otimes \sigma_n^3 \oplus T_n \otimes \sigma_n^3, \\ x_n^{(4)} &= x_{n-1}^{(2)} \otimes \sigma_n^2 \oplus x_{n-1}^{(3)} \otimes \sigma_n^3 \oplus T_n \otimes [\sigma_n^{2 \vee 3}]. \end{aligned}$$

So we have a (max,plus)-linear recurrence with

$$A_n = \begin{pmatrix} \epsilon & \epsilon & \epsilon & \epsilon \\ \epsilon & \sigma_n^2 & \epsilon & \epsilon \\ \epsilon & \epsilon & \sigma_n^3 & \epsilon \\ \epsilon & \sigma_n^2 & \sigma_n^3 & \epsilon \end{pmatrix} \qquad B_n = \begin{pmatrix} 0 \\ \sigma_n^2 \\ \sigma_n^3 \\ \sigma_n^{2\vee 3} \end{pmatrix}.$$

We drop coordinate 1 but we keep coordinate 4 in our recursion for the reasons explained above. It is easy to check that what was just done is in fact equivalent to the generic way of transforming (2.9) into (2.5) which was presented above.

Here it is easy to check that $x_n^{(2)} \geq x_{n-1}^{(2)}$ and $x_n^{(3)} \geq x_{n-1}^{(3)}$ for all $n \geq 1$, and that this in turn implies that $x_n(4) \geq x_{n-1}(4)$. Hence we can take (note that the size of the matrix is $s = 3 < KL = 4$) :

$$A_n = \begin{pmatrix} \sigma_n^2 & \epsilon & \epsilon \\ \epsilon & \sigma_n^3 & \epsilon \\ \sigma_n^2 & \sigma_n^3 & 0 \end{pmatrix} \qquad B_n = \begin{pmatrix} \sigma_n^2 \\ \sigma_n^3 \\ \sigma_n^{2\vee 3} \end{pmatrix}.$$

Similar modifications can be made in FIFO networks where for all i, $x_n^{(i)} \geq x_{n-1}^{(i)}$.

Remark 2. Although we will not need this in what follows, we find it useful to stress that one can also associate to all event graphs some token dynamics (see [11] p. 69 and following). If one sees Equation (2.9) as an extension of Lindley's equation (initially for the G/G/1 queue) to event graphs, the token dynamics of event graphs can then be seen as a generalization of that of customers in such a queue, see the Section 2.2.4 with examples.

We just showed how one can associate to such an event graph a (max,plus)-linear system. In particular note that the matrices A_n and the vector B_n produced have a fixed structure : for each n and each i, j, $A_n^{(i,j)}$ and $B_n^{(i)}$ are either almost surely finite, or else almost surely equal to $-\infty$. Indeed, we even showed more. Let m be the number of timed transitions in the event graph. The set of timed transitions is denoted $\mathcal{T}_{timed} = \{t(1), \ldots, t(m)\} \subset \mathcal{T} = \{1, \ldots, K\}$ and we take $\zeta_n = (\sigma_n^{t(1)}, \ldots, \sigma_n^{t(m)})$. We showed that the matrices and vectors $\{A_n, B_n\}$ that are used in the recursion are obtained via two applications $\mathcal{A}$ and $\mathcal{B}$ such that :

$$\begin{array}{rccc} \mathcal{A} : & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,s)}\left(\mathbb{R}_{\max}\right) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{A}(\sigma), \end{array}$$

$$\begin{array}{rccc} \mathcal{B} : & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,1)}\left(\mathbb{R}_{\max}\right) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{B}(\sigma), \end{array}$$

via the formula

$$\begin{array}{rcl} \mathcal{A}(\zeta_n) & = & A_n, \\ \mathcal{B}(\zeta_n) & = & B_n. \end{array}$$

Note that our notation are consistent and to an event graph, one associates a network in the sense of section 2.1.1 with

$$X_{[m,n]}(N) = \bigoplus_{1 \leq i \leq s} \bigoplus_{m \leq k \leq n} \left(\left\{ \bigotimes_{j=n}^{k+1} \mathcal{A}(\zeta_j) \right\} \otimes \mathcal{B}(\zeta_k) \otimes T_k \right)^{(i)},$$

with the convention $\bigotimes_{j=n}^{n+1} = E$, the identity matrix.

We show now that under some additional conditions the class of the event graphs is a subclass of the monotone separable class. In what follows, we will always assume that these conditions are satisfied.

Proposition 3. *Consider an event graph such that*

A1 For all $i \in \mathcal{T}$, there exists a tokenless path in the oriented graph $\mathcal{G}$, going from t_1 to t_{out} through i ;

A2 Each transition i is either untimed (with $\sigma_n^i \equiv 0$) or recycled, namely such that there exists a place p with marking 1 such that p is both a predecessor and a successor of i (a natural way of making the event graph FIFO).

Then the network associated with this event graph is monotone-separable.

The fact that one can associate a (max,plus)-linear recurrence as described in Section 2.2.1 with $A_n \otimes \mathbf{0} = B_n \oplus \mathbf{0}$ follows from the next proposition. Hence thanks to Lemma 39, we know that this system belongs to the monotone separable framework.

Proposition 4. *To an event graph verifying conditions of previous proposition, one can always associate two applications $\mathcal{A}$ and $\mathcal{B}$ such that*

1. *there exists I, J such that $I \cap J = \emptyset$, $I \cup J = [1, s]$ and (we omit the ζ since the following properties are true for all $\zeta \in \mathbb{R}_+^m$)*

$$\forall i \in J,\ \mathcal{B}^{(i)} = \epsilon,\ \max_j \mathcal{A}^{(i,j)} = 0;$$
$$\forall i \in I,\ \max_j \mathcal{A}^{(i,j)} = \max_{j \in I} \mathcal{A}^{(i,j)} = \mathcal{B}^{(i)}.$$

2. *for all i, $A^{(i,i)} \geq 0$ and for all $k \in [1, m]$, there exists j such that $A^{(j,j)}(\zeta) = \sigma^{(k)}$.*

Proof.

We will show that the matrices given by (2.11) satisfy the desired conditions. In order for matrix A_n in (2.11) to satisfy point 2, we must add zeros on the diagonal. Note that due to the FIFO assumption, this is always possible. Now the second part of point 2 follows from the fact that under Assumption A2 of Proposition 3, each $\sigma^{(k)}$ is on the diagonal of a_1 which is the diagonal of $\overline{a}_1$ too since a_0 is strictly triangular thanks to Assumption A1.

The first point follows (with $J = [1, (L-1)K]$ and $I = [(L-1)K+1, LK]$) from the Lemma 9 proved in section 2.2.5.

□

Remark 3. It is clear that previous conditions are symetric. More precisely if π is a permutation of $[1, s]$ and $\mathcal{A}_\pi^{(i,j)} = \mathcal{A}^{(\pi(i),\pi(j))}$, $\mathcal{B}_\pi^{(i)} = \mathcal{B}^{(\pi(i))}$ then it is equivalent to chec the conditions on the couple $(\mathcal{A}, \mathcal{B})$ or $(\mathcal{A}_\pi, \mathcal{B}_\pi)$. This fact will be used in the next section to get a generic form for the applications.

2.2.3 Reducible and Irreducible Event Graphs

Two transitions of an event graph will be said to belong to the same communication class if there is a directed path in $\mathcal{G}$ from the first to the second and another one from the second to the first. We denote by $\mathcal{C}_1, \ldots, \mathcal{C}_d$ these communication classes, which form a partition of the set of transitions. By construction, these communication classes can be arranged according to a partial order denoted $\prec$. The numbering is assumed to be compatible with this partial order : $\mathcal{C}_i \prec \mathcal{C}_j \Rightarrow i \leq j$. By definition there is always a place between 2 transitions, hence we can consider the graph (still denoted $\mathcal{G}$) where we delete the places. The set of vertices is $\mathcal{T}$ and there is an edge (i, j) if there is a two hop path from i to j.

There is a natural way (see Section 2.3 of [11]) to associate to the applications $\mathcal{A}$ and $\mathcal{B}$ some graphs $\mathcal{G}_{\mathcal{A}} = (\mathcal{V}, \mathcal{E}_{\mathcal{A}})$ and $\mathcal{G}_{\mathcal{B}} = (\{0\} \cup \mathcal{V}, \mathcal{E}_{\mathcal{B}})$. Let $\mathcal{V} = \{1, \ldots, s\}$. If $\mathcal{A}^{(j,i)}(\sigma) > \epsilon$, then the edge (i, j) belongs to $\mathcal{E}_{\mathcal{A}}$ and has weight $\mathcal{A}(\sigma)^{(j,i)}$. If $\mathcal{B}^{(i)}(\sigma) > \epsilon$, then the edge $(0, i)$ belongs to $\mathcal{E}_{\mathcal{B}}$ and has weight $\mathcal{B}^{(i)}(\sigma)$. We denote $\mathcal{G}_{\mathcal{A}\cup\mathcal{B}} = \mathcal{G}_{\mathcal{A}} \cup \mathcal{G}_{\mathcal{B}}$.

We will denote by Ξ (resp. Ξ_n) the set of paths in $\mathcal{G}_{\mathcal{A}\cup\mathcal{B}}$ from node 0 to node s (resp. with length n, where the length is the number of edges of the path).

Remark 4. The simplification on the matrices A_n and B_n correspond to the following operations on the graph $\mathcal{G}_{\mathcal{A}\cup\mathcal{B}}$: if there exists no edge starting from vertices i, then we delete this node i and all the edges that link to this vertices. We operate recursively. The final graph corresponds exactly to the simplified matrices A_n and B_n. Indeed the simplifications will not affect any result on the underlying event graph and we can deal with the matrices of (2.11) for the proofs.

We refer to [11] (page 42) for the interpretation of product of matrices in term of paths in graph. Let $\mathcal{C}'_1, \ldots, \mathcal{C}'_{d'}$ be the communication classes of $\mathcal{G}_{\mathcal{A}}$ and $\lessdot$ the associated partial order. We assume that $\mathcal{C}'_i \lessdot \mathcal{C}'_j \Rightarrow i \leq j$.

Lemma 8. *The sets $\{\mathcal{C}_1, \ldots, \mathcal{C}_d, \prec\}$ and $\{\mathcal{C}'_1, \ldots, \mathcal{C}'_{d'}, \lessdot\}$ are isomorphic. In particular $d = d'$.*

Proof.

We consider the matrix before simplification and we omit the subscript n :

$$A = \begin{pmatrix} \epsilon & E & & & \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \epsilon & & & & E \\ \overline{a}_L & \overline{a}_{L-1} & \ldots & \ldots & \overline{a}_1 \end{pmatrix},$$

where $\bar{a}_i = a_0^* \otimes a_i$. We take the following notation $i^k = (L-k)K + i$ for $1 \leq i \leq K$ and $1 \leq k \leq L$. Then the upper part of the matrix $\begin{pmatrix} \epsilon & E & & \\ \vdots & & \ddots & \\ \vdots & & & \ddots \\ \epsilon & & & E \end{pmatrix}$ correspond in $\mathcal{G}_{\mathcal{A}}$ to edges between vertices i^k and i^{k+1} with null weight. For the lower part of the matrix, the coefficients of the submatrix $\bar{a}_k$ give the weight of edges between i^k and j^1 (if $\bar{a}_k^{(j,i)} > \epsilon$). Now if $i \to j$ is an edge of $\mathcal{G}$, then by construction there exists an indice k such that $a_k^{(j,i)} > \epsilon$. Hence there exists a path in $\mathcal{G}_{\mathcal{A}}$ from i^1 to j^1, namely : $i^1 \to i^2 \to \cdots \to i^k \to j^1$. Now if $i^1 \to j^1$ is an edge of $\mathcal{G}_{\mathcal{A}}$, then $i \to j$ is an edge of $\mathcal{G}$. And the lemma follows. □

The set $\{\mathcal{C}_1, \ldots, \mathcal{C}_d, \prec\}$ is by definition an acyclic graph. Hence by permuting the indices, we obtain for the matrix in the evolution equation of the event graph the following block structure (after simplification) :

$$A_n = \left(\begin{array}{ccccccc} A_n(1,1) & | & \epsilon & | & \epsilon & | & \epsilon \\ - & - & - & - & - & - & - \\ A_n(2,1) & | & A_n(2,2) & | & \epsilon & | & \epsilon \\ - & - & - & - & - & - & - \\ & \vdots & & \vdots & & \vdots & \\ - & - & - & - & - & - & - \\ A_n(d,1) & | & A_n(d,2) & | & & | & A_n(d,d) \end{array} \right), B_n = \left(\begin{array}{c} B_n(1) \\ - \\ B_n(2) \\ - \\ \vdots \\ - \\ B_n(d) \end{array} \right),$$

where each $A_n(i,i)$ is an irreducible matrix of size s_i (corresponding to communication class $\mathcal{C}_i$).

As the output transition is necessarily in the last communication class ("last" refers here to the partial order $\prec$), this choice of numbering can be made compatible with our earlier assumption that the last coordinate is that of the output transition.

2.2.4 Event Graphs : Examples

Queues in tandem

We consider two queues in tandem, as illustrated in Figure 2.1. Let σ_n^i be the n-th service time

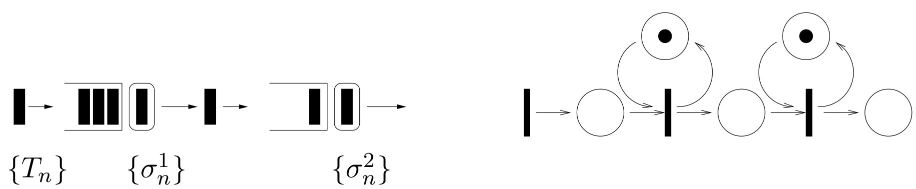


FIG. 2.1 – Queues in Tandem

at the i-th server. We denote by x_n^1 (resp. x_n^2) the end of the n-th service in queue 1 (resp. 2). We have then

$$x_n^1 = (T_n \oplus x_{n-1}^1) \otimes \sigma_n^1, \quad (2.12)$$
$$x_n^2 = (x_n^1 \oplus x_{n-1}^2) \otimes \sigma_n^2. \quad (2.13)$$

Putting (2.12) in (2.13) gives

$$x_n^2 = (x_{n-1}^1 \oplus T_n) \otimes (\sigma_n^1 \otimes \sigma_n^2) \oplus x_{n-1}^2 \otimes \sigma_n^2,$$

with

$$X_n = \begin{pmatrix} x_n^1 \\ x_n^2 \end{pmatrix},$$
$$A_n = \begin{pmatrix} \sigma_n^1 & \epsilon \\ \sigma_n^1 + \sigma_n^2 & \sigma_n^2 \end{pmatrix},$$
$$B_n = \begin{pmatrix} \sigma_n^1 \\ \sigma_n^1 + \sigma_n^2 \end{pmatrix},$$

so that

$$X_n = A_n \otimes X_{n-1} \oplus B_n \otimes T_n.$$

Tree Queueing Network

Consider a tree with nodes numbered by $1, 2, \ldots, m$ such that j is a successor of i (we write $j \in Suc(i)$) implies $i \leq j$. In particular node 1 is the root. We associate to this tree the network with m queues and in which departure process of queue i is the input process of queues $j \in Suc(i)$. Queues in tandem is a special case of tree network. Note that in the literature, such tree networks are also referred to as disassembly networks.

We take the example of a tree with three queues, as illustrated in Figure 2.2. The end-to-end delay here is defined as the delay for a customer to traverse all the queues, which is taken care of by the dummy node (a max operator) in the end.

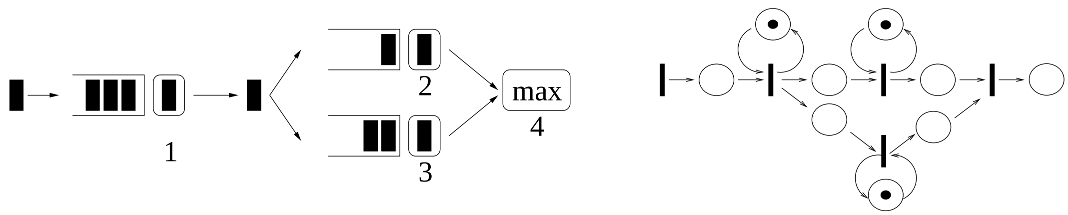


FIG. 2.2 – Tree Network

Let σ_n^i be the n-th service time at the i-th server. We denote by x_n^i the end of the n-th service in server i. We have then

$$x_n^1 = (T_n \oplus x_{n-1}^1) \otimes \sigma_n^1, \tag{2.14}$$
$$x_n^2 = (x_n^1 \oplus x_{n-1}^2) \otimes \sigma_n^2, \tag{2.15}$$
$$x_n^3 = (x_n^1 \oplus x_{n-1}^3) \otimes \sigma_n^3, \tag{2.16}$$
$$x_n^4 = x_n^2 \oplus x_n^3. \tag{2.17}$$

Putting equation (2.14) in (2.15), (2.16), (2.17) we obtain the desired recursion equation with

$$X_n = \begin{pmatrix} x_n^1 \\ x_n^2 \\ x_n^3 \\ x_n^4 \end{pmatrix}, \quad B_n = \begin{pmatrix} \sigma_n^1 \\ \sigma_n^1 + \sigma_n^2 \\ \sigma_n^1 + \sigma_n^3 \\ \sigma_n^1 + \max(\sigma_n^2, \sigma_n^3) \end{pmatrix},$$

$$A_n = \begin{pmatrix} \sigma_n^1 & \epsilon & \epsilon & \epsilon \\ \sigma_n^1 + \sigma_n^2 & \sigma_n^2 & \epsilon & \epsilon \\ \sigma_n^1 + \sigma_n^3 & \epsilon & \sigma_n^3 & \epsilon \\ \sigma_n^1 + \max(\sigma_n^2, \sigma_n^3) & \sigma_n^2 & \sigma_n^3 & 0 \end{pmatrix}.$$

Notice that both precedent examples are feed-forward networks.

Queueing network with fixed window control

We consider now m queues in tandem with a window-based control which does not allow more than L customers in the system. In other words, the n th customer can enter the first queue only after the $n - L$ th customer leaves the last queue in the tandem queueing network. We denote by x_n^i the end of the n-th service in queue i.

For the network of two queues in Figure 2.3, we have then

$$x_n^1 = (T_n \oplus x_{n-1}^1 \oplus x_{n-L}^2) \otimes \sigma_n^1, \tag{2.18}$$
$$x_n^2 = (x_n^1 \oplus x_{n-1}^2) \otimes \sigma_n^2. \tag{2.19}$$

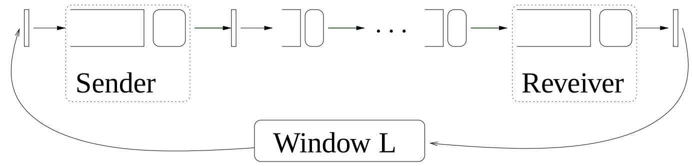


FIG. 2.3 – Tandem Queueing Network with Fixed Window Control

From these equations, if we put (2.18) in (2.19), we obtain the desired recursion equation with

$$X_n = \begin{pmatrix} x^2_{n-L+1} \\ x^2_{n-L+2} \\ \vdots \\ \vdots \\ x^2_{n-1} \\ x^1_n \\ x^2_n \end{pmatrix}, \quad B_n = \begin{pmatrix} \epsilon \\ \epsilon \\ \vdots \\ \vdots \\ \epsilon \\ \sigma^1_n \\ \sigma^1_n + \sigma^2_n \end{pmatrix},$$

$$A_n = \begin{pmatrix} \epsilon & 0 & \epsilon & \dots & \dots & \epsilon \\ \epsilon & \epsilon & 0 & \epsilon & \dots & \epsilon \\ \vdots & & & \ddots & & \\ \vdots & & & & \ddots & \\ \epsilon & & & & & 0 \\ \sigma^1_n & \epsilon & \dots & \epsilon & \sigma^1_n & \epsilon \\ \sigma^1_n + \sigma^2_n & \epsilon & \dots & \epsilon & \sigma^1_n + \sigma^2_n & \sigma^2_n \end{pmatrix}.$$

In the tree network case, the window control with size L is implemented in such a way that the n th customer can enter the first queue (root of the tree) only after the all the $n-L$ th customers quit the leave queues in the tree queueing network.

2.2.5 Event Graphs : proofs

Lemma 9. *For all* $k \in [1, L]$ *and all* $i \in [1, K]$*, we have* $\max_j \overline{a}_k^{(i,j)} \leq \max_j \overline{a}_1^{(i,j)} = \overline{b}^{(i)}$

Proof. Thanks to Assumption A1, we have by construction for all $k \geq 1$,

$$a_k\mathbf{0} \oplus \mathbf{0} \leq a_0\mathbf{0}.$$

Hence by monotonicity, we have

$$a_0^*(a_k\mathbf{0} \oplus \mathbf{0}) \leq a_0^* a_0 \mathbf{0} = a_0^*\mathbf{0}, \tag{2.20}$$

since clearly $\mathbf{0} \leq a_o\mathbf{0}$. From (2.20), we derive that $a_0^* a_k \mathbf{0} \leq a_0^*\mathbf{0}$. Since $(a_0^*\mathbf{0})^{(i)} = \max_j (a_0^*)^{(i,j)} = (a_0^*)^{(i,1)} = \overline{b}^{(i)}$, we showed that

$$\max_j \overline{a}_k^{(i,j)} \leq \overline{b}^{(i)}.$$

Now for a_1, we have thanks to Assumption A2 that ensures that the event graph is FIFO :

$$a_1\mathbf{0} \geq \mathbf{0}.$$

Hence we have $a_0^* a_1 \mathbf{0} \geq a_0^* \mathbf{0}$ and finally $\overline{a}_1 \mathbf{0} = \overline{b}$ which concludes the proof. □

2.3 Generalized Jackson Networks

In this section, we introduce the (single class) generalized Jackson network. Such networks have been considered among others by Jackson [61] or Gordon and Newell [52]. The framework that we use here is that of Baccelli and Foss in [12].

We will take the following notation

1. $\mathbb{A}_1$ (resp. $\mathbb{A}_1^*$) is the set of non-negative sequences : $u = \{u_i\}_{1\leq i\leq n}$, such that $n \leq +\infty$, and $u_i \geq 0$, (resp. $u_i > 0$) for all $i \leq n$;
2. $\mathbb{A}_2$ (resp. $\mathbb{A}_2^*$) is the set of non-decreasing sequences : $U = \{U_i\}_{1\leq i\leq n}$, such that $n \leq +\infty$, and $0 \leq U_i \leq U_{i+1}$ (resp. $0 < U_i < U_{i+1}$) for all $i \leq n-1$;

We will denote by $\mathbb{A}$ (resp. $\mathbb{A}^*$) the set of discrete measure on $\mathbb{R}_+$ such that there exists $U \in \mathbb{A}_2$ (resp. $U \in \mathbb{A}_2^*$) with $d\mathscr{U} = \sum_{1\leq i\leq n} \delta_{U_i}$. To such a measure we can associate a sequence $u \in \mathbb{A}_1$ (resp. $u \in \mathbb{A}_1^*$) in the following manner $u_i = U_i - U_{i-1}$, for $i \geq 1$ and with the convention $U_0 = 0$. $\mathbb{A}_3$ (resp. $\mathbb{A}_3^*$) will denote the set of counting functions : $\mathcal{U} : \mathbb{R}_+ \to \mathbb{N}$ such that $\mathcal{U}(t) = \sum_{1\leq i\leq n} \mathbb{1}_{\{U_i \leq t\}} = \int_0^t d\mathscr{U}$ with $d\mathscr{U} \in \mathbb{A}$ (resp. $d\mathscr{U} \in \mathbb{A}^*$). Clearly the spaces $\mathbb{A}, \mathbb{A}_1, \mathbb{A}_2$ and $\mathbb{A}_3$ are isomorphic and the same holds with $\mathbb{A}^*, \mathbb{A}_1^*, \mathbb{A}_2^*$ and $\mathbb{A}_3^*$.

2.3.1 Single Server Queue

A single server queue will be defined by $\mathbf{Q} = \left(\tau^A, \sigma\right)$, where $\tau^A = \{\tau_i^A\}_{1\leq i\leq n}$ and $\sigma = \{\sigma_i\}_{1\leq i\leq n}$ belong to $\mathbb{A}_2$ and $\mathbb{A}_1$ respectively. The interpretations are the following : customer i arrives in the queue at time τ_i^A and its service time is σ_i.
Associated to a queue $\mathbf{Q}$, we define the departure process $\{\tau_i^D\}_{1\leq i\leq n} \in \mathbb{A}_2$ by

$$\begin{cases} \tau_1^D = \tau_1^A + \sigma_1, \\ \tau_i^D = \max[\tau_i^A, \tau_{i-1}^D] + \sigma_i, \quad 2 \leq i \leq n. \end{cases} \tag{2.21}$$

τ_i^D is the departure time of customer i. Expanding this recursion yields

$$\tau_i^D = \max_{j=1\ldots i} (\tau_j^A + \sigma(j,i)), \quad \text{for} \quad 1 \leq i \leq n, \tag{2.22}$$

with the notation $\sigma(j,i) = \sigma_j + \cdots + \sigma_i$. Hence we defined a mapping $\Phi : \mathbb{A} \times \mathbb{A} \mapsto \mathbb{A}$ such that :

$$\tau^D = \{\tau_i^D\}_{1\leq i\leq n} = \Phi(\mathbf{Q}). \tag{2.23}$$

We will use the following notation for the different counting functions :
- $A(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i^A \leq t\}}$;
- $\Sigma(t) = \sum_{n=1}^{\infty} \mathbb{1}_{\{\sigma(1,n) \leq t\}}$;
- $D(t) = \sum_{i=1}^{\infty} \mathbb{1}_{\{\tau_i^D \leq t\}}$.

For any non-decreasing function F, we denote by $F^{\leftarrow}(x) = \inf\{t,\ F(t) \geq x\}$ the pseudo-inverse of F (which is left-continuous). We have $F^{\leftarrow}(x) \leq u \Leftrightarrow x \leq F(u)$. Moreover, we use the notation $\wedge$ for $\min$ and $\vee$ for $\max$. The following lemma gives a new description of the departure process in term of counting functions.

Lemma 10. *Given a queue $\mathbf{Q} \in \mathbb{A}^* \times \mathbb{A}$, let $D = \Phi(\mathbf{Q})$ where Φ is the mapping defined by Equations (2.22) and (2.23). In term of counting functions, we have :*

$$D(t) = A(t) \wedge \inf_{0 \leq s \leq t} \Sigma[t - s + \Sigma^{\leftarrow}(A(s))]. \tag{2.24}$$

Proof.

For $1 \leq j$, we define the point process Γ_j as follows :

$$\begin{aligned} \tau_n^{\Gamma_j} &= 0 \quad \text{for } 1 \leq n \leq j-1, \\ \tau_n^{\Gamma_j} &= \tau_j^A + \sigma(j,n) \quad \text{for } n \geq j. \end{aligned}$$

The construction of Γ_j is depicted in Figure 2.4 and we have for $j \geq 1$ and with the convention $\sigma(1,0) = 0$,

$$\begin{aligned} \Gamma_j(t) &= j-1 \quad \text{for } t < \tau_j^A, \\ \Gamma_j(t) &= \Sigma(t - \tau_j^A + \sigma(1, j-1)) \quad \text{for } t \geq \tau_j^A. \end{aligned}$$

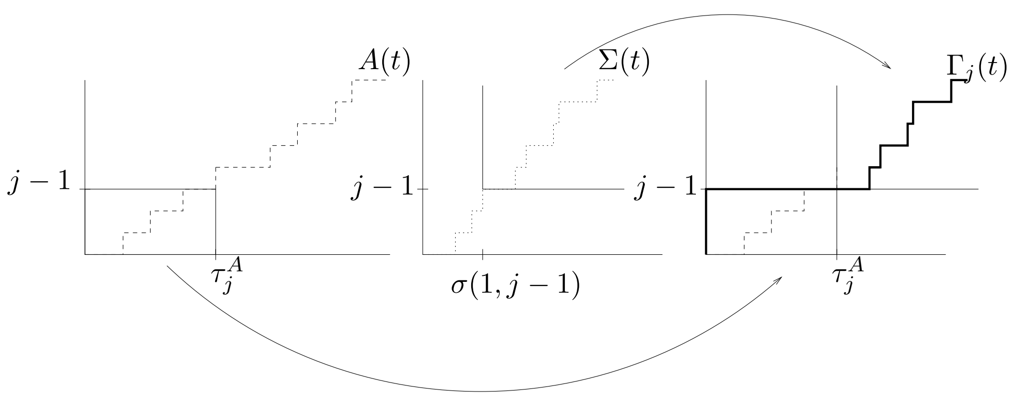


FIG. 2.4 – Construction of Γ_j

Thanks to (2.22), we have $t \geq \tau_n^D \Leftrightarrow \forall j \leq n,\ t \geq \tau_n^{\Gamma_j}$, hence we have

$$D(t) \geq n \quad \Leftrightarrow \quad \inf_{j \leq n} \Gamma_j(t) \geq n,$$

but we have for all $j \geq n+1$, $\Gamma_j(t) \geq n$, for all t, hence $D(t) = \inf_{j \geq 1} \Gamma_j(t)$. We have

$$\inf_{j \geq 1} \Gamma_j(t) = \inf_{\{j \geq 1,\, \tau_j^A \leq t\}} \Sigma[t - \tau_j^A + \sigma(1, j-1)] \wedge A(t).$$

We now show that $\inf_{j\geq 1}\Gamma_j(t) = A(t)\wedge \inf_{0\leq s\leq t}\Sigma[t-s+\Sigma^{\leftarrow}(A(s))]$. Since $\tau_j^A \in \mathbb{A}_2^*$, on each interval $[\tau_{j-1}^A,\tau_j^A)$ (we use the convention $\tau_0^A=0$), we have $A(s)=j-1$ and the function $s\mapsto \Sigma[t-s+\Sigma^{\leftarrow}(j-1)]$ is non-increasing, hence we have

$$\inf_{s\in[\tau_{j-1}^A,\tau_j^A)} \Sigma[t-s+\Sigma^{\leftarrow}(A(s))] \quad = \quad \Sigma[t-\tau_j^A+\sigma(1,j-1)].$$

Moreover, we have for $\tau_k^A\leq t<\tau_{k+1}^A$,

$$\inf_{s\in[\tau_k^A,t)} \Sigma[t-s+\Sigma^{\leftarrow}(A(s))] = \Sigma(\Sigma^{\leftarrow}(k)) \geq k = A(t).$$

Finally we have

$$\begin{aligned} A(t)\wedge \inf_{0\leq s\leq t}\Sigma[t-s+\Sigma^{\leftarrow}(A(s))] &= A(t)\wedge \inf_{\{j\geq 1,\,\tau_j^A\leq t\}}\Sigma[t-\tau_j^A+\sigma(1,j-1)] \\ &= \inf_{j\geq 1}\Gamma_j(t). \end{aligned}$$

□

Remark 5. Equations (2.22) and (2.24) give two equivalent definitions of the mapping $\Phi:\mathbb{A}^*\times\mathbb{A}\to\mathbb{A}$. But for $\tau^A\in\mathbb{A}$, only Equation (2.22) gives the right definition of Φ. In particular notice that we always have $\tau_i^D\geq\sigma(1,i)\vee\tau_i^A$, from which we derive $D(t)\leq\Sigma(t)\wedge A(t)$.

2.3.2 Generalized Jackson Networks

We recall here the notation introduced in [12], to describe a generalized Jackson network with K nodes.

The networks we consider are characterized by the fact that service times and routing decisions are associated with stations and not with customers. This means that the j-th service on station k takes $\sigma_j^{(k)}$ units of time, where $\{\sigma_j^{(k)}\}_{j\geq 1}$ is a predefined sequence. In the same way, when this service is completed, the leaving customer is sent to station $\nu_j^{(k)}$ (or leaves the network if $\nu_j^{(k)}=K+1$) and is put at the end of the queue on this station, where $\{\nu_j^{(k)}\}_{j\geq 1}$ is also a predefined sequence, called the routing sequence. The sequences $\{\sigma_j^{(k)}\}_{j\geq 1}$ and $\{\nu_j^{(k)}\}_{j\geq 1}$, where k ranges over the set of stations, are called the driving sequences of the net. A generalized Jackson network will be defined by

$$\mathbf{JN} = \left\{\{\sigma_j^{(k)}\}_{j\geq 1},\ \{\nu_j^{(k)}\}_{j\geq 1},\ n^{(k)},\ 0\leq k\leq K\right\}.$$

where $(n^{(0)},n^{(1)},\ldots,n^{(K)})$ describes the initial condition. The interpretation is as follows : for $k\neq 0$, at time $t=0$, in node k, there are $n^{(k)}$ customers with service times $\sigma_1^{(k)},\ldots,\sigma_{n^{(k)}}^{(k)}$ (if appropriate, $\sigma_1^{(k)}$ may be interpreted as a residual service time).

Node 0 models the external arrival of customers in the network. Hence,

- if $n^{(0)}=0$, there is no external arrival.
- if $\infty>n^{(0)}\geq 1$, then for all $1\leq j\leq n^{(0)}$, the arrival time of the j-th customer in the network takes place at $\sigma_1^{(0)}+\cdots+\sigma_j^{(0)}$ and it joins the end of the queue of station $\nu_j^{(0)}$. Hence $\sigma_j^{(0)}$ is the j-th inter-arrival time. Note that in this case, there may be a finite number of customers passing through a given station so that the network is actually well defined once a finite sequence of routing decisions and service times are given on this station.

- if $n^{(0)} = \infty$, then when taking for instance the sequence $\{\sigma_j^{(0)}\}_{j\geq 1}$ i.i.d., the arrival process is a renewal process etc.

To each node of a generalized Jackson network, we can associate the following counting functions in $\mathbb{A}$:

1. $K+1$ functions associated to the service times $\sigma^{(k)}$ (as in the single server queue) ;
2. $K(K+1)$ functions that counts the number of customers routed from a node $\{0,\ldots,K\}$ to a node $\{1,\ldots,K\}$;
3. $K+1$ functions associated to $n^{(k)}$.

Hence a generalized Jackson network with K nodes is an object in $\mathbb{A}^{(K+1)(K+2)} = \mathbb{A}^{\mathbf{JN}}$.
We will use the following notation for each of these counting functions :

- $N = (n^{(0)},\ldots,n^{(K)})$, with $n^{(i)} \geq 0$;
- $\sigma^{(k)} = \{\sigma_j^{(k)}\}_{j\geq 1}$ and $\sigma^{(k)}(1,n) = \sum_{j=1}^{n} \sigma_j^{(k)}$, for $0 \leq k \leq K$;
- $\Sigma^{(i)}(t) = \sum_n \mathbb{1}_{\{\sigma^{(i)}(1,n)\leq t\}}$, for $0 \leq i \leq K$;
- $P_{i,j}(n) = \sum_{l\leq n} \mathbb{1}_{\{\nu_l^{(i)}=j\}}$, for $0 \leq i \leq K, \quad 1 \leq j \leq K+1$.

We denote the arrival and departure processes of each queue k of the networks by $A^{(k)}$ and $D^{(k)}$ respectively, with the following notation $\mathbf{A} = (A^{(1)},\ldots,A^{(K)})$ and $\mathbf{D} = (D^{(1)},\ldots,D^{(K)})$. We give a procedure that constructs the processes $\mathbf{A}$ and $\mathbf{D}$:

$\mathbf{Procedure\ 1(JN)}$:

$-1-$ $\quad t := 0;$
$\quad \mathbf{for} \quad i \geq 0 \quad \mathbf{do}$
$\quad\quad R^{(i)}(t) := \sigma_1^{(i)}; \quad A^{(i)}(t) := n^{(i)}; \quad D^{(i)}(t) := 0;$
$\quad \mathbf{od}$

$-2-$ $\quad V := \min_{\{i,\, A^{(i)}(t)-D^{(i)}(t)\geq 1\}} R^{(i)}(t); \quad \gamma := \arg \min_{\{i,\, A^{(i)}(t)-D^{(i)}(t)\geq 1\}} R^{(i)}(t);$

$-3-$ $\quad \mathbf{if} \quad V = \infty \quad \mathbf{then} \quad END;$
$\quad \mathbf{fi}$

$-4-$ $\quad D^{(\gamma)}(t+V) := D^{(\gamma)}(t) + 1; \quad A^{(\gamma)}(t+V) := A^{(\gamma)}(t);$
$\quad \mathbf{if} \quad A^{(\gamma)}(t+V) - D^{(\gamma)}(t+V) \geq 1 \quad \mathbf{then} \quad R^{(\gamma)}(t+V) := \sigma^{(\gamma)}_{D^{(\gamma)}(t+V)+1}; \quad \mathbf{fi}$
$\quad j := \nu^{(\gamma)}_{D^{(\gamma)}(t+V)};$
$\quad \mathbf{if} \quad j \neq K+1 \quad \mathbf{then} \quad A^{(j)}(t+V) := A^{(j)}(t) + 1; \quad D^{(j)}(t+V) := D^{(j)}(t);$
$\quad\quad\quad \mathbf{if} \quad A^{(j)}(t) - D^{(j)}(t) = 0 \quad \mathbf{then} \quad R^{(j)}(t+V) := \sigma^{(j)}_{A^{(j)}(t+V)}; \quad \mathbf{fi}$
$\quad \mathbf{fi}$
$\quad \mathbf{for} \quad i \notin \{\gamma, j\} \quad \mathbf{do}$
$\quad\quad R^{(i)}(t+V) := R^{(i)}(t) - V; \quad A^{(i)}(t+V) := A^{(i)}(t); \quad D^{(i)}(t+V) := D^{(i)}(t);$
$\quad \mathbf{od}$
$\quad t := t + V;$

$-5-$ $\quad \mathbf{goto} \quad 2;$

Remark 6. Since each sequence $\{\sigma_j^{(k)}\}_{j\geq 1}$ or $\{\nu_j^{(k)}\}_{j\geq 1}$ is infinite, the variables $\nu^{(\gamma)}_{D^{(\gamma)}(t+V)}$, $\sigma^{(\gamma)}_{D^{(\gamma)}(t+V)}$ and $\sigma^{(j)}_{A^{(j)}(t+V)}$ in step 4 are always available :

- if $\sum_{i=0}^K n^{(i)} < +\infty$ then the procedure ends in step 3 ;
- if $\sum_{i=0}^K n^{(i)} = +\infty$, the procedure never ends, this corresponds to a network with infinite number of customers. In this case there exists $T \leq \infty$ such that $\lim_{t\to T} A(t) = \lim_{t\to T} D(t) = \infty$.

We take the following notation : given a departure process for queue 0 : $\Sigma^{(0)}$, and departure processes for the queues $i \in [1, K]$: $\mathbf{X} = \{X^{(i)}\}_{1\leq i\leq K}$, and an initial number of customers in each queue $n^{(i)}$, we construct the following arrival processes $\mathbf{Y} = \{Y^{(i)}\}_{1\leq i\leq K}$:

$$Y^{(i)}(t) = n^{(i)} + P_{0,i}(\Sigma^{(0)}(t) \wedge n^{(0)}) + \sum_{j=1}^K P_{j,i}(X^{(j)}(t)). \tag{2.25}$$

We denote this by $\mathbf{Y} = \Gamma(\mathbf{X}, \mathbf{JN})$.
Given an arrival process for each queue : $\mathbf{Y}$, we define the corresponding departure process $\mathbf{X}$ and denote it by $\mathbf{X} = \Phi(\mathbf{Y}, \mathbf{JN})$. Hence, we have $X^{(i)} = \Phi(Y^{(i)}, \Sigma^{(i)})$, where Φ was defined for the single server queue in (2.22).

Proposition 5. $\mathbf{A}$ *and* $\mathbf{D}$*, the arrival and departure processes of the generalized Jackson network are the unique solution of the fixed point equation*

$$\begin{cases} \mathbf{A} = \Gamma(\mathbf{D}, \mathbf{JN}), \\ \mathbf{D} = \Phi(\mathbf{A}, \mathbf{JN}). \end{cases} \tag{2.26}$$

We will denote by Ψ *the mapping from* $\mathbb{A}^{\mathbf{JN}}$ *to* $\mathbb{A}^2$ *that to any Jackson network* $\mathbf{JN}$ *associates the corresponding couple* $(\mathbf{A}, \mathbf{D})$*.*

Proof.

If we define $J^{(k)} = \sup\{j, \sum_{i=1}^j \sigma_i^{(k)} = 0\}$, the generalized Jackson network is equivalent to the following

$$\left\{ \{\sigma_j^{(k)}\}_{j\geq J^{(k)}+1}, \{\nu_j^{(k)}\}_{j\geq J^{(k)}+1}, n^{(k)} + \sum_{i=0}^K P_{i,k}(J^{(i)}) \right\}.$$

Hence, we can assume that $J^{(k)} = 0$, for all k and we have for time $t = 0$, $A^{(i)}(0) = n_i$, $D^{(i)}(0) = 0$. For $t \geq 0$ let

$$\begin{aligned} \tilde{D}^{(i)}(t) &= A^{(i)}(0) \wedge \inf_{0\leq s\leq t} \Sigma^{(i)}[t - s + \Sigma^{(i)\leftarrow}(A^{(i)}(0))] \\ \tilde{A}^{(i)}(t) &= n^{(i)} + P_{0,i}(\Sigma^{(0)}(t) \wedge n^{(0)}) + \sum_{j=1}^K P_{j,i}(\tilde{D}^{(j)}(t)). \end{aligned}$$

Now consider $t_1 = \inf\{t \geq 0,\ \exists i,\ \tilde{D}^{(i)}(t-) \neq \tilde{D}^{(i)}(t), \text{or } \Sigma^{(0)}(t-) \neq \Sigma^{(0)}(t)\}$ the first time of jump for processes $\tilde{D}$ and $\tilde{A}$. Thus

$$\begin{aligned} A^{(i)}(t) &= \tilde{A}^{(i)}(t) \quad \text{for } 0 \leq t \leq t_1, \\ D^{(i)}(t) &= \tilde{D}^{(i)}(t) \quad \text{for } 0 \leq t \leq t_1, \end{aligned}$$

provide a solution pair to (5.33) over $t \in [0, t_1]$, moreover this solution is exactly the one constructed by the previous procedure. Now suppose a solution pair $(\mathbf{A}, \mathbf{D})$ has been constructed on $[0, t_n]$, where t_n is a jump point for one of the $A^{(i)}$, $D^{(i)}$. As above let $X(s) = \mathbf{A}(s)$ for $s \leq t_n$, $X(s) = \mathbf{A}(t_n)$ for $s > t_n$, and for $t \geq t_n$ define,

$$\begin{aligned}\tilde{D}^{(i)}(t) &= X^{(i)}(t) \wedge \inf_{0 \leq s \leq t} \Sigma^{(i)}[t - s + \Sigma^{(i)\leftarrow}(X^{(i)}(s))], \\ \tilde{A}^{(i)}(t) &= n^{(i)} + P_{0,i}(\Sigma^{(0)}(t) \wedge n^{(0)}) + \sum_{j=1}^{K} P_{j,i}(\tilde{D}^{(j)}(t)).\end{aligned}$$

Letting $t_{n+1} = \inf\{t \geq t_n,\ \exists i,\ \tilde{D}^{(i)}(t-) \neq \tilde{D}^{(i)}(t), \text{or } \Sigma^{(0)}(t-) \neq \Sigma^{(0)}(t)\}$, one concludes as above. The uniqueness of $(\mathbf{A}, \mathbf{D})$ and the fact that $(\mathbf{A}, \mathbf{D})$ are constructed by **Procedure 1** are a consequence of this construction procedure. □

Remark 7.

1. This construction is very similar to the construction of the reflection mapping made in the proof of Theorem 2.1 of [26] ;
2. This property gives the connection between two possible descriptions of a generalized Jackson network. One of these descriptions has been given in words at the beginning of this section and is depicted with more rigor in the **Procedure 1**. The other description is in term of fixed point Equation (5.33) which has already been introduced by Majewski in [71]. These two descriptions are equivalent in the special case of discrete inputs and an empty network at time $t = 0-$.

2.4 Generalized Processor Sharing Queues

Consider a processor which offers service to inputs arriving from a variety of sources. If one wishes to offer different levels of service to different types of sources, then separate customer classes are needed and a service policy must be established. Generalized Processor Sharing is a policy that has been proposed for use in high-speed data networks.

Consider the following model of N coupled $G/G/FIFO$ queues. Each queue is served in accordance with the Generalized Processor Sharing (GPS) discipline, which operates as follows. Queue j is assigned a weight ϕ^j, with $\sum_{j=1}^{N} \phi^j = 1$. If all queues are backlogged, then queue j is served at speed ϕ^j. If some of the queues are empty, then the excess capacity is redistributed among the backlogged queues in proportion to their respective weights. All customers within each queue are served in a FIFO order.

More formally we can construct the workload of each queues as follows. Let $\{T_n^A, \sigma_n, c_n\}$ be a simple marked process, with $\sigma_n > 0$ and $c_n \in \{1, \ldots, N\}$. The interpretations are the following : customer n arrives in the queue c_n at time T_n^A and its service time is σ_n. We will say that this customer is of class $c_n \in \{1, \ldots, N\}$ and we denote by $\tau_n = T_{n+1}^A - T_n^A$ the inter-arrival times. We denote by $W_Y^i[n] := W_Y^i(T_n^A-)$ the workload of queue i at time T_n^A- with initial condition $W_Y^i[0] = Y^i$. The sequence $\{W_Y[n] = (W_Y^1[n], \ldots, W_Y^N[n])\}$ is generated by the recurrence

$$W_Y[n+1] = h(W_Y[n], \sigma_n, c_n, \tau_n), \quad n = 0, 1 \ldots$$

where the function h is defined by the following equations :

$$\begin{aligned} W^j(T_k^A) &= W^j(T_k^A-) + \sigma_k \mathbb{1}_{\{c_k=j\}}, & (2.27)\\ \frac{dW^j}{dt}(t) &= -r^j(t) \quad \text{for } T_k^A \le t < T_{k+1}^A, & (2.28)\\ r^j(t) &= \begin{cases} \frac{\phi^j}{\sum_{\ell \notin I(t)} \phi^\ell} & j \notin I(t), \\ 0 & j \in I(t); \end{cases} & (2.29)\\ I(t) &= \{i,\ W^i(t) = 0\}. & (2.30) \end{aligned}$$

Equations (2.27), (2.28), (2.29) and (2.30) show how to construct the workload process of each queue for $t \ge T_0^A$.

Note that we have

$$\sum_i W_Y^i[n+1] = \left(\sum_i W_Y^i[n] + \sigma_n - \tau_n \right)^+ ,$$

the recurrence for the sum of the component of $W_Y[n]$ reduces to the Lindley's equation.

The stability of the GPS queues follows directly from the stability of the single server queue with input process $\{T_n^A, \sigma_n\}_{n\in\mathbb{Z}}$, since the sum of the workload of each queue is exactly the workload of this single server queue.

Chapitre 3

Fluid Models

3.1 Fluid Limit of Generalized Jackson Networks

In this section, we consider a (single class) generalized Jackson network and its fluid limit as introduced in Section 2.3.

In [26], Chen and Mandelbaum derive the fluid approximation for generalized Jackson networks. The queue-length, busy-time and workload processes are obtained from the input processes through the oblique reflexion mapping due to Skorokhod [81] in a one-dimensional setting and to Harrison and Reiman [56] in the context of open networks. Using this fluid approach and assuming that service times and inter-arrival times are independent and identically distributed (i.i.d.), Dai shows in [29] that generalized Jackson networks are stable (i.e. positive Harris recurrent) when the nominal load is less than one at each station. The first stability result for generalized Jackson networks under ergodic assumptions can be found in the paper of Foss [43]. In [71], Majewski derives an unified formalism which allows discrete and fluid customers. The input for the model are the cumulative service times, the cumulative external arrivals and the cumulative routing decisions of the queues. A path space fixed point equation characterizes the corresponding behavior of the network.

The framework that we use here is that of Baccelli and Foss in [12], where only stationarity and ergodicity on the data are assumed. Denote by X_0^n the time to empty the system when n customers arrive at the same time from the outside world in the network. Thanks to a subadditive argument, the following limit is shown to hold in [12]

$$\lim_{n\to\infty} \frac{X_0^n}{n} = \gamma(0) \quad \text{a.s.} \tag{3.1}$$

The constant $\gamma(0)$ corresponds to the maximal throughput capacity of the network. In fact the saturation rule [13] makes this intuition rigorous and ensures that if $\rho = \lambda\gamma(0) < 1$ then the network is stable. In this chapter, we provide a new proof of (3.1) using fluid approximations which gives an explicit formula for the constant $\gamma(0)$. One contribution of this paper is to provide a connection between the fluid approximation of a generalized Jackson network and the stability condition for this network under stationary and ergodic assumptions on the data. In particular, no i.i.d. assumptions are needed (on inter-arrival times or service times) and we can consider more general routing mechanism than Bernouilli routing.

The other application of this section will be linked to the calculation of tails in generalized Jackson networks with subexponential service distributions in next chapter. We are able to give here the behavior (in the fluid scale) of the network on a "rare" event. We refer to the next chapter for an exact notion of what we mean by rare event.

Results of [26] or [71] will be of minor help for us since a lot of work would be required to obtain our explicit result form theirs. For these reasons, we took a different approach. For each time t, we are able to give an explicit formulation of the fluid limit. The simplicity of the result is due to the concavity of the processes in the fluid scale ; a property which had not been proved yet to the best of our knowledge. In other words, given some drifts for the input processes, when a queue becomes empty, it remains empty forever. It seems that this basic fact has not been exploited yet. It allows us to reduce the computation of the fluid limits (which are solution of a fixed-point network equation in a functional space as described in [71]) to the computation of some traffic intensity for a simplified network that evolves in time. Hence for a fixed time, we only have to compute a fixed point solution of some traffic equations (see Section 3.1.2). Proposition 8 gives the fluid approximation of generalized Jackson network. To obtain the time to empty the system, we just notice that if the network is processing fluid, then one of the queues has work since the initial time. This gives us a very compact way of obtaining the constant $\gamma(0)$ (Theorem 2 of Section 3.1.3). Proposition 9 is a slight extension of the main Theorem 2 and will be needed in the computation of the fluid picture of a generalized Jackson network in the specific case of one big jump see next chapter.

This section is based on the paper [67].

3.1.1 The case of Single Server Queue

For any sequence of functions $\{f^n\}$, we define the corresponding scaled sequence $\{\hat{f}^n\}$ as follows : $\hat{f}^n(t) = \frac{f^n(nt)}{n}$. We say that $f^n \to f$ uniformly on compact sets, or simply $f^n \to f$ u.o.c. if for each $t > 0$,

$$\sup_{0 \le u \le t} |f^n(u) - f(u)| \to 0 \quad \text{as } n \to \infty.$$

We first recall the following lemma known as Dini's Theorem :

Lemma 11. *Let $\{f^n\}$ be a sequence of nondecreasing functions on $\mathbb{R}_+$ and let f be a continuous function on $\mathbb{R}$. Assume that $f^n(t) \to f(t)$ for all t (weak convergence is denoted by $f^n \to f$). Then $f^n \to f$ u.o.c.*

The following Lemma can be found in Billingsley [19] page 287 :

Lemma 12. *If f_n are nondecreasing functions and $f_n \to f$, then $f_n^{\leftarrow} \to f^{\leftarrow}$.*

Proposition 6. *Consider a sequence of single server queues $\{\mathbf{Q}^n\} = \{\tau^{A,n}, \sigma^n\} \in (\mathbb{A} \times \mathbb{A})^{\mathbb{N}}$ with associated arrival process $\tau^{A,n}$ such that $\hat{A}^n(t) \to \hat{A}(t)$ for all $t > 0$, with $\hat{A}$ concave on $\mathbb{R}_+$, and associated service time process σ^n such that $\hat{\Sigma}^n(t) \to \mu t$ for all $t \ge 0$, with $\mu \ge 0$, then $\hat{D}^n \to \hat{D}$ u.o.c, with $\hat{D}(t) = \mu t \wedge \hat{A}(t)$.*

Proof :
First observe that thanks to Remark 5, we have $D^n(t) \le \Sigma^n(t) \wedge A^n(t)$, hence making the fluid scaling and taking the limit in n, we have $\hat{D}(t) \le \mu t \wedge \hat{A}(t)$. Proposition 6 follows in the case $\mu = 0$ by Lemma 11. We consider now the case $\mu > 0$ and first assume that : $\mathbf{Q}^n \in \mathbb{A}^* \times \mathbb{A}$ for all n and $\hat{A}(0) = 0$.

Since $\hat{A}(0) = 0$, $\hat{A}$ is continuous on $\mathbb{R}_+$ and Lemma 11 gives $\hat{A}^n \to \hat{A}$ u.o.c. Moreover thanks to Lemma 12, the sequences $\hat{\Sigma}^n$ and $\hat{\Sigma}^{n\leftarrow}$ converge u.o.c. to the respective functions $t \mapsto \mu t$ and $t \mapsto \frac{t}{\mu}$.
For fixed $t \geq 0$, we have thanks to uniformity on compact sets,

$$\begin{aligned}
\lim_{n\to\infty} \frac{D^n(nt)}{n} &= \lim_{n\to\infty} \inf_{0\leq u\leq t} \frac{1}{n}\Sigma^n[n(t-u) + (\Sigma^n)^{\leftarrow}(A^n(nu))] \wedge \frac{A^n(nt)}{n} \\
&= \inf_{0\leq u\leq t} \lim_{n\to\infty} \frac{1}{n}\Sigma^n[n(t-u) + (\Sigma^n)^{\leftarrow}(A^n(nu))] \wedge \lim_{n\to\infty} \frac{A^n(nt)}{n} \\
&= \inf_{0\leq u\leq t} [\mu(t-u) + \hat{A}(u)] \wedge \hat{A}(t) \\
&= \mu t \wedge \hat{A}(t),
\end{aligned}$$

where the last equality follows from concavity of $\hat{A}$. Now using Lemma 11, the result follows in this case.
To extend the result to the case : $\mathbf{Q}^n \in \mathbb{A} \times \mathbb{A}$, we consider the sequence $\tau_i^{B,n} = \tau_i^{A,n} + 1/i$ which belongs to $\mathbb{A}^*$. For any $\epsilon > 0$, we have for $n \geq 1/\epsilon$, $A^n(n(t-\epsilon)) \leq B^n(nt) \leq A^n(nt)$. Hence $\hat{A}(t-\epsilon) \leq \hat{B}(t) \leq \hat{A}(t)$ and since $\hat{A}$ is continuous, we have $\hat{B} = \hat{A}$. Moreover, since $\tau_i^{B,n} \geq \tau_i^{A,n}$, we have $D_B^n = \Phi(B^n, \Sigma^n) \leq \Phi(A^n, \Sigma^n)$, and we can apply the first part of the proof to $\hat{B}$, hence $D_B^n(t) \to \hat{A}(t) \wedge \mu t$ and the result follows in this case.
The case $\hat{A}(0) \neq 0$ can be dealt with the same monotonicity argument. For any $\epsilon > 0$, consider the sequence $\tau_i^{C,n} = \tau_i^{B,n} \vee i\epsilon$. We have $\hat{C}(t) = \frac{t}{\epsilon} \wedge \hat{A}(t)$ and $\tau_i^{C,n} \geq \tau_i^{A,n}$. We can apply the first part of the proof to $\hat{C}$, hence $D_C^n(t) \to \hat{C}(t) \wedge \mu t$. For $\epsilon \leq \mu^{-1}$, we get $\hat{D}(t) \geq \mu t \wedge \hat{A}(t)$. □

3.1.2 Fluid Limit and Bottleneck Analysis

Bottleneck Analysis

We first define the **N**on **C**apture condition as follows :
Condition (NC) : we say that the $K \times K$ matrix $P = (p_{i,j})_{1\leq i,j\leq K}$ satisfies **(NC)**, if P is a substochastic matrix such that the following stochastic matrix

$$R = \begin{pmatrix} p_{1,1} & \dots & p_{1,K} & 1 - \sum_i p_{1,i} \\ & p_{i,j} & & \vdots \\ p_{K,1} & \dots & p_{K,K} & 1 - \sum_i p_{K,i} \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

has only $K+1$ as absorbing state, i.e if (X_n) is a Markov chain with transition matrix R, almost surely (X_n) is equal to $K+1$ eventually.

Lemma 13. *Let P be a $K \times K$ substochastic matrix. The following properties are equivalent :*

1. *P satisfies* **(NC)** *;*
2. *the Perron Frobenius eigenvalue of P is $r < 1$;*
3. *$(I - P')$ is invertible.*

Proof :
1 $\Rightarrow$ 2 $\Rightarrow$ 3 by Corollary 1 page 8 and Corollary 2 page 31 of Seneta [79]. To see that 3 $\Rightarrow$ 1, just write the equations for the expected number of visits for the Markov chain (X_n) with transition matrix R, to state $i \neq K+1$, $V_i = \mathbb{E}\left[\sum_n \mathbb{1}_{\{X_n=i\}}\right]$:

$$V_i = \mathbb{P}\left[X_0 = i\right] + \sum_{j=1}^{K} p_{j,i} V_j \quad \text{for all } i \in [1, K]. \tag{3.2}$$

Since $(I - P')$ is invertible, (3.2) has a finite solution. Hence the only absorbing state of (X_n) is $K+1$. $\square$
For $\mathbf{x}$ and $\mathbf{y}$ two vectors of $\mathbb{R}^K$, we will write $\mathbf{x} \geq \mathbf{y}$ if $x_i \geq y_i$, for all i.
For any matrix P, any vector $\alpha \in \mathbb{R}_+^K$ and any $\mathbf{y} \in \mathbb{R}_+^K$, we define $\mathbf{F}_\alpha : \mathbb{R}_+^K \to \mathbb{R}_+^K$ and $\mathbf{G}_\mathbf{y} : \mathbb{R}_+^K \to \mathbb{R}_+^K$ with

$$\begin{aligned} (\mathbf{F}_\alpha)_i(x_1, \ldots, x_K) &= \alpha_i + \sum_{j=1}^{K} p_{j,i} x_j, \\ (\mathbf{G}_\mathbf{y})_i(x_1, \ldots, x_K) &= x_i \wedge y_i. \end{aligned}$$

Proposition 7. *If the matrix P satisfies* **(NC)**, *the fixed point equation $\mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{x}) = \mathbf{x}$ has an unique solution $\mathbf{x}(\alpha, \mathbf{y})$. Moreover, $(\alpha, \mathbf{y}) \mapsto \mathbf{x}(\alpha, \mathbf{y})$ is a continuous, non-decreasing function.*

Remark 8. These relations already appeared in Massey [72] and Chen and Mandelbaum [26] see section 3.1. In fact as pointed out in [26], we can use Tarski's fixed point theorem (Tarski [82]) to get the existence of this fixed point (called inflow in [26]). But we give here a self-contained proof that shows continuity and monotonicity properties of the solution.

Proof :
Existence of a solution to the fixed point equation is an easy consequence of monotonicity. Since $\mathbf{F}_\alpha$ and $\mathbf{G}_\mathbf{y}$ are non-decreasing functions and $\mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(0) \geq 0$, we see that $(\mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y})^n(0) \nearrow \mathbf{b}$. We have $\mathbf{b} \leq \mathbf{F}_\alpha(\mathbf{y})$ and $\mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{b}) = \mathbf{b}$.
For a given subset Δ of $[1, K]$ and $\mathbf{y} \in \mathbb{R}_+^K$ we define $\mathbf{F}_{\alpha,\mathbf{y}}^\Delta : \mathbb{R}_+^K \to \mathbb{R}_+^K$ by

$$(\mathbf{F}_{\alpha,\mathbf{y}}^\Delta)_i(x_1, \ldots, x_K) = \alpha_i + \sum_{j \in \Delta} p_{j,i} y_j + \sum_{j \in \Delta^c} p_{j,i} x_j.$$

$F_\bullet^\Delta(\bullet)$ depend only on $\{x_i,\ i \in \Delta^c\}$ and $\mathbf{F}_\alpha = \mathbf{F}_{\alpha,\mathbf{y}}^\emptyset$.
We fix $\mathbf{y} \in \mathbb{R}_+^K$ and first study the case $\mathbf{F}_{\alpha,\mathbf{y}}^\Delta(\mathbf{x}) = \mathbf{x}$.
This equation is

$$\left\{ \begin{array}{rcl} x_1 &=& \alpha_1 + \sum_{j \in \Delta} p_{j,1} y_j + \sum_{j \in \Delta^c} p_{j,1} x_j, \\ &\vdots& \\ x_K &=& \alpha_K + \sum_{j \in \Delta} p_{j,K} y_j + \sum_{j \in \Delta^c} p_{j,K} x_j. \end{array} \right.$$

In fact, we only have to calculate $\{x_i,\ i \in \Delta^c\}$ and then, we obtain $\{x_i,\ i \in \Delta\}$. Renumbering the indexes of $\mathbf{x}$, and taking into account only those in Δ^c, we have

$$\left\{ \begin{array}{rcl} x_1 &=& \lambda_1(\alpha, \mathbf{y}) + \sum_{j=1}^n p_{j,1}^\Delta x_j, \\ &\vdots& \\ x_n &=& \lambda_n(\alpha, \mathbf{y}) + \sum_{j=1}^n p_{j,n}^\Delta x_j. \end{array} \right. \tag{3.3}$$

$P^\Delta = (p^\Delta_{i,j}; i,j = 1,\dots,n)$ is a substochastic matrix and $I - P^\Delta$ is invertible (even for $\Delta = \emptyset$ see Lemma 13). Hence, if $\lambda(\alpha,\mathbf{y}) = (\lambda_1(\alpha,\mathbf{y}),\dots,\lambda_n(\alpha,\mathbf{y}))$, Equation (3.3) has only one solution given by :

$$\tilde{\mathbf{x}}^\Delta = \lambda(\alpha,\mathbf{y}) + \tilde{\mathbf{x}}^\Delta P^\Delta \quad \Leftrightarrow \quad \tilde{\mathbf{x}}^\Delta = \lambda(\alpha,\mathbf{y})(I - P^\Delta)^{-1}.$$

We now return to our fix point problem $\mathbf{x} = \mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{x})$. To show uniqueness of the solution, take any solution $\mathbf{z} = \mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{z})$. We have $\mathbf{z} \geq 0$ hence $\mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{z}) \geq \mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(0)$ and then $\mathbf{z} \geq \mathbf{b}$. Let $A = \{i,\ z_i > y_i\}$ and $B = \{i,\ b_i > y_i\}$. Of course, we have $B \subset A$ and $\mathbf{b} = \tilde{\mathbf{x}}^B$ since $\mathbf{F}^B_{\alpha,\mathbf{y}}(\mathbf{b}) = \mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{b}) = \mathbf{b}$. Moreover, we have

$$\begin{aligned} z_i &= \alpha_i + \sum_{j\in B} p_{j,i} y_j + \sum_{j\in A\setminus B} p_{j,i} y_j + \sum_{j\notin A} p_{j,i} z_j, \\ (\mathbf{F}^B_{\alpha,\mathbf{y}})_i(\mathbf{z}) &= \alpha_i + \sum_{j\in B} r_{j,i} y_j + \sum_{j\in A\setminus B} p_{j,i} z_j + \sum_{j\notin A} p_{j,i} z_j, \end{aligned}$$

hence, we have $\mathbf{F}^B_{\alpha,\mathbf{y}}(\mathbf{z}) \geq \mathbf{z}$. But since $(\mathbf{F}^B_{\alpha,\mathbf{y}})^n(\mathbf{z}) \nearrow \tilde{\mathbf{x}}^B = \mathbf{b}$, we have $\mathbf{b} \geq \mathbf{z}$. Finally $\mathbf{z} = \mathbf{b}$.
For any Δ, $(\alpha,\mathbf{y}) \mapsto \tilde{\mathbf{x}}^\Delta(\alpha,\mathbf{y}) = \lambda(\alpha,\mathbf{y})(I - P^\Delta)^{-1}$ is a continuous non-decreasing function. Fix any $(\alpha,\mathbf{y})$, and define $A = \{i,\ x_i(\alpha,\mathbf{y}) \geq y_i\}$, $B = \{i,\ x_i(\alpha,\mathbf{y}) > y_i\}$. We have of course $\mathbf{x}(\alpha,\mathbf{y}) = \tilde{\mathbf{x}}^A(\alpha,\mathbf{y}) = \tilde{\mathbf{x}}^B(\alpha,\mathbf{y})$ and for $(\beta,\mathbf{z})$ in a neighborhood of $(\alpha,\mathbf{y})$, we have $\mathbf{x}(\beta,\mathbf{z}) \in \{\tilde{\mathbf{x}}^A(\beta,\mathbf{z}),\ \tilde{\mathbf{x}}^B(\beta,\mathbf{z})\}$, and the continuity of $(\alpha,\mathbf{y}) \mapsto \mathbf{x}(\alpha,\mathbf{y})$ follows from that of $(\alpha,\mathbf{y}) \mapsto \tilde{\mathbf{x}}^\Delta(\alpha,\mathbf{y})$.
Now to see that this function is non decreasing, take $(\beta,\mathbf{z}) \geq (\alpha,\mathbf{y})$, we have

$$\mathbf{F}_\beta \circ \mathbf{G}_\mathbf{z}(\mathbf{x}(\alpha,\mathbf{y})) \geq \mathbf{F}_\alpha \circ \mathbf{G}_\mathbf{y}(\mathbf{x}(\alpha,\mathbf{y})) = \mathbf{x}(\alpha,\mathbf{y})$$

and the sequence $\{(\mathbf{F}_\beta \circ \mathbf{G}_\mathbf{z})^n(\mathbf{x}(\alpha,\mathbf{y}))\}_{n\geq 0}$ increases to $\mathbf{x}(\beta,\mathbf{z})$. □

Fluid Limit for Generalized Jackson Networks

We consider the following sequence of Jackson networks :

$$\begin{aligned} \mathbf{JN}^n &= \{\sigma^n, \nu^n, N^n\}, \quad \text{with}, \\ \lim_{n\to\infty} \frac{N^n}{n} &= (n^{(0)}, n^{(1)}, \dots, n^{(K)}), \quad n^{(0)} \leq +\infty, \quad n^{(i)} < \infty,\ i \neq 0. \end{aligned}$$

Thanks to **Procedure 1** given in appendix, we can construct the corresponding input and output processes $\mathbf{A}^n$ and $\mathbf{D}^n$. We assume that the driving sequences satisfy

$$\begin{aligned} & \hat{\Sigma}^{(0),n}(t) \to \Sigma^{(0)}(t), \quad \text{where } t \mapsto \Sigma^{(0)}(t) \wedge n^{(0)} \text{ is a concave function,} \\ \forall k \geq 1,\quad & \hat{\Sigma}^{(k),n}(t) \to \mu^{(k)} t, \quad \forall t \geq 0 \quad (\mu^{(k)} \geq 0), \\ & \hat{P}^n_{i,j}(t) \to p_{i,j} t \quad \forall t \geq 0. \end{aligned}$$

We suppose that the routing matrix $P = (p_{i,j})_{1\leq i,j\leq K}$ satisfies **(NC)**.

Proposition 8. *The processes $\mathbf{A}^n$ and $\mathbf{D}^n$ converge uniformly on compact sets to a fluid limit defined by*

$$\hat{A}^{(i)}(t) = n^{(i)} + p_{0,i}(\Sigma^{(0)}(t) \wedge n^{(0)}) + \sum_{j=1}^K p_{j,i} \hat{D}^{(j)}(t), \tag{3.4}$$

$$\hat{D}^{(i)}(t) = \hat{A}^{(i)}(t) \wedge \mu^{(i)} t. \tag{3.5}$$

Remark 9. 1. Existence and uniqueness of solutions to Equations (3.4) and (3.5) follow directly from Proposition 7 as shown in the proof. Moreover, it easily follows from the proof that each component of $\mathbf{A}$ and $\mathbf{D}$ is concave and if $\Sigma^{(0)}$ is piece-wise linear then so are the processes $\mathbf{A}$ and $\mathbf{D}$.

2. Theorem 7.1 of [26] gives the fluid approximation of a generalized Jackson network, if we take a linear function for $\Sigma^{(0)}$, then from $(\hat{\mathbf{A}}, \hat{\mathbf{D}})$, we can calculate explicitly the solution of the equations of this Theorem.

Proof :
For any fixed $n \geq 1$, we define the sequences of processes $\{\mathbf{A}_t^n(k), \mathbf{D}_t^n(k)\}_{k\geq 0}$ and $\{\mathbf{A}_b^n(k), \mathbf{D}_b^n(k)\}_{k\geq 0}$ with the same recurrence equation :

$$\begin{cases} \mathbf{A}^n(k+1) = \Gamma(\mathbf{D}^n(k), \mathbf{JN}^n), \\ \mathbf{D}^n(k+1) = \Phi(\mathbf{A}^n(k+1), \mathbf{JN}^n), \end{cases}$$

but with different initial conditions $\mathbf{D}_t^n(0) = (\Sigma^{(1),n}, \ldots, \Sigma^{(K),n})$ and $\mathbf{D}_b^n(0) = (0, \ldots, 0)$.
We recall the notation :

$$\begin{aligned} \Gamma_i(\mathbf{X}, \mathbf{JN}^n)(t) &= n^{(i),n} + P^n_{0,i}(\Sigma^{(0),n}(t) \wedge n^{(0),n}) + \sum_{j=1}^K P^n_{j,i}(X_j(t)), \\ \Phi_i(\mathbf{X}, \mathbf{JN}^n)(t) &= \Phi(X_i, \sigma^{(i),n})(t), \end{aligned}$$

and we will use the scaled sequences $\hat{\mathbf{A}}^n(k)(t) = \frac{\mathbf{A}^n(k)(nt)}{n}$ and $\hat{\mathbf{D}}^n(k)(t) = \frac{\mathbf{D}^n(k)(nt)}{n}$. We introduce the mappings $\Gamma^s : \mathbb{C}^K \to \mathbb{C}^K$ and $\Phi^s : \mathbb{C}^K \to \mathbb{C}^K$ that appear in Equations (3.4) and (3.5) (where $\mathbb{C}$ is the set of continuous functions on $\mathbb{R}_+$) :

$$\begin{aligned} \Gamma_i^s(x_1, \ldots, x_K)(t) &= n^{(i)} + p_{0,i}(\Sigma^{(0)}(t) \wedge n^{(0)}) + \sum_{j=1}^K p_{j,i} x_j(t), \\ \Phi_i^s(x_1, \ldots, x_K)(t) &= x_i(t) \wedge \mu^{(i)} t. \end{aligned}$$

The following lemma holds for both top and bottom sequences, hence we omit the $\cdot_t$ or $\cdot_b$.

Lemma 14. *Assume that for a fixed k, $\hat{\mathbf{D}}^n(k) \to \hat{\mathbf{D}}(k)$ u.o.c. and that each component of $\hat{\mathbf{D}}(k)$ is a concave function. Then we have*

$$\hat{\mathbf{A}}^n(k+1) \xrightarrow{n\to\infty} \Gamma^s(\hat{\mathbf{D}}(k)) = \hat{\mathbf{A}}(k+1) \quad \textit{u.o.c. and} \quad \hat{\mathbf{D}}^n(k+1) \xrightarrow{n\to\infty} \Phi^s(\hat{\mathbf{A}}(k+1)) = \hat{\mathbf{D}}(k+1) \quad \textit{u.o.c.}$$

and components of $\hat{\mathbf{A}}(k+1)$ and $\hat{\mathbf{D}}(k+1)$ are concave functions.

Proof of the lemma :
For any fixed t, we have

$$\frac{A^{(i),n}(k+1)(nt)}{n} = \frac{n^{(i),n}}{n} + \frac{P^n_{0,i}(\Sigma^{(0),n}(nt) \wedge n^{(0),n})}{n} + \sum_{j=1}^K \frac{P^n_{i,j}(D^{(j),n}(k)(nt))}{n}.$$

Hence thanks to Lemma 11, we have $\hat{\mathbf{A}}^n(k+1) \xrightarrow{n\to\infty} \Gamma^s(\hat{\mathbf{D}}(k))$ u.o.c. and each component of $\hat{\mathbf{A}}(k+1) = \Gamma^s(\hat{\mathbf{D}}(k))$ is clearly a concave function. Now thanks to Proposition 6 the result follows. □

We now return to the proof of Proposition 8.

We have $\hat{\mathbf{A}}(k+1) = \Gamma^s \circ \Phi^s(\hat{\mathbf{A}}(k))$. This equation gives the relation between 2 functions of a real parameter t. But we can fix this parameter and then we obtain for any fixed t an equation between real numbers that we write $\hat{\mathbf{A}}(k+1)(t) = \Gamma^s \circ \Phi^s(\hat{\mathbf{A}}(k)(t))$ (even if $\Gamma^s \circ \Phi^s$ is supposed to act on functions). Moreover as a consequence of Proposition 7, we know that the fixed point equation $\Gamma^s \circ \Phi^s(\zeta(t)) = \zeta(t)$ has an unique solution, namely $\zeta(t) = \mathbf{x}(\alpha, \mu^{(1)}t, \ldots, \mu^{(K)}t)$, with $\alpha = (n^{(1)}+p_{0,1}(\Sigma^{(0)}(t)\wedge n^{(0)}), \ldots, n^{(i)}+p_{0,i}(\Sigma^{(0)}(t)\wedge n^{(0)}), \ldots, n^{(K)}+p_{0,K}(\Sigma^{(0)}(t)\wedge n^{(0)}))$. For any t, the sequence $\{\hat{\mathbf{A}}_b(k)(t)\}_{k\geq 1}$ (resp. $\{\hat{\mathbf{A}}_t(k)(t)\}_{k\geq 1}$) is non decreasing (resp. non increasing). We have $\hat{\mathbf{A}}_b(k)(t) \xrightarrow{k\to\infty} \zeta(t)$ and $\hat{\mathbf{A}}_t(k)(t) \xrightarrow{k\to\infty} \zeta(t)$ and $\hat{\mathbf{D}}_b(k)(t) \xrightarrow{k\to\infty} \Phi^s(\zeta(t))$ and $\hat{\mathbf{D}}_t(k)(t) \xrightarrow{k\to\infty} \Phi^s(\zeta(t))$.

Moreover, fix any $n \geq 1$, the mappings $. \mapsto \Gamma(., \mathbf{JN}^n)$ and $. \mapsto \Phi(., \mathbf{JN}^n)$ are non decreasing and :

$$\begin{cases} \mathbf{A}^n = \Gamma(\mathbf{D}^n, \mathbf{JN}^n), \\ \mathbf{D}^n = \Phi(\mathbf{A}^n, \mathbf{JN}^n). \end{cases}$$

Hence, for all $k \geq 0$, we have :

$$\begin{array}{ccc} \mathbf{A}_b^n(k) \leq & \mathbf{A}^n & \leq \mathbf{A}_t^n(k), \\ \mathbf{D}_b^n(k) \leq & \mathbf{D}^n & \leq \mathbf{D}_t^n(k). \end{array}$$

We have :

$$\begin{array}{ccccc} \frac{\mathbf{A}_b^n(k)(nt)}{n} & \leq & \frac{\mathbf{A}^n(nt)}{n} & \leq & \frac{\mathbf{A}_t^n(k)(nt)}{n}, \\ \hat{\mathbf{A}}_b(k)(t) & \leq & \liminf_n \frac{\mathbf{A}^n(nt)}{n} \leq \limsup_n \frac{\mathbf{A}^n(nt)}{n} & \leq & \hat{\mathbf{A}}_t(k)(t), \end{array}$$

hence, we have

$$\forall t, \quad \lim_n \frac{\mathbf{A}^n(nt)}{n} = \zeta(t).$$

The result follows from Lemma 11. □

3.1.3 Maximal Dater Asymptotic

Motivation

We first recall the definition of simple Euler network from Section 4.1 of [12]. Consider a route $p = (p_1, \ldots, p_L)$ with $1 \leq p_i \leq K$ for $i = 2, \ldots, L-1$. Such a route is successful if $p_1 = 0$ and $p_L = K+1$. We can associate to such a route a routing sequence ν and a vector ϕ as follows

($\oplus$ means concatenation) :

$$
\begin{array}{ll}
\textbf{Procedure 2}(\mathbf{p}): & \\
-1- & \textbf{for} \quad k = 0 \dots K \quad \textbf{do} \\
 & \quad \nu^{(k)} := \emptyset; \\
 & \quad \phi^{(k)} := 0; \\
 & \textbf{od} \\
-2- & \textbf{for} \quad i = 1 \dots L-1 \quad \textbf{do} \\
 & \quad \nu^{(p_i)} := \nu^{(p_i)} \oplus p_{i+1}; \\
 & \quad \phi^{(p_i)} := \phi^{(p_i)} + 1; \\
 & \textbf{od}
\end{array}
$$

Note that $\phi^{(j)}$ is the number of visits to node j in such a route.
A simple Euler network is a generalized Jackson network $E = \{\sigma, \nu, N\}$, with $N = (1, 0, \dots, 0)$. The routing sequence $\nu = \{\nu_i^{(k)}\}_{i=1}^{\phi^{(k)}}$ is generated by a successful route and $\sigma = \{\sigma_i^{(k)}\}_{i=1}^{\phi^{(k)}}$ is a sequence of real-valued non-negative numbers, representing service times.
Consider now a sequence of simple Euler networks, say $\{E(l)\}_{l=1}^{+\infty}$ where $E(l) = \{\sigma(l), \nu(l), 1\}$. We define σ and ν to be the infinite concatenation of the $\{\sigma(l)\}_{l=1}^{+\infty}$ and $\{\nu(l)\}_{l=1}^{+\infty}$. Denote by σ_c the sequence obtained from σ in the following manner

$$\sigma_c = (c\sigma^{(0)}, \sigma^{(1)}, \dots, \sigma^{(K)}).$$

We consider the corresponding sequence of Jackson networks $\mathbf{JN}_c^n = \{\sigma_c, \nu, N^n\}$, with $N^n = (n, 0, \dots, 0)$. The Jackson network $\mathbf{JN}_c^n$ corresponds to an empty network with n customers in node 0 at time $t = 0$. We will denote by X_c^n the time to empty the system $\mathbf{JN}_c^n$, called maximal dater of the network. Thanks to the Euler property of $\{E(i)\}_{i \geq 1}$, we know that for all n, $X_c^n < +\infty$ (see [12]). We suppose that

$$\lim_{n\to\infty} \frac{\sigma_c^{(0)}(1,n)}{n} = \frac{c}{\lambda}, \tag{3.6}$$

$$\lim_{n\to\infty} \frac{\sigma^{(k)}(1,n)}{n} = \frac{1}{\mu^{(k)}}, \quad 1 \leq k \leq K, \quad (\mu^{(k)} > 0) \tag{3.7}$$

$$\lim_{n\to\infty} \frac{P_{i,j}(n)}{n} = p_{i,j}, \quad 0 \leq i \leq K, \quad 1 \leq j \leq K+1. \tag{3.8}$$

We assume that $P = (p_{i,j})_{1 \leq i,j \leq K}$ satisfies **(NC)**. We denote by π_i the solution of the following system

$$\forall i \in [1, K], \quad \pi_i = p_{0,i} + \sum_{j=1}^{K} p_{j,i}\pi_j. \tag{3.9}$$

The constant π_i is the expected number of visits to site i for the Markov chain with transition matrix P and with initial distribution $p_{0,i}$ (see proof of Lemma 13). We will prove the following theorem :

Theorem 2. *Under the previous conditions, we have for all* $c \geq 0$

$$\lim_{n\to\infty} \frac{X_c^n}{n} = \max_{1 \leq i \leq K} \frac{\pi_i}{\mu^{(i)}} \vee \frac{c}{\lambda}.$$

Proof of Theorem 2

Given a routing matrix $P = (p_{i,j}; i,j = 0,\dots,K)$ that satisfies **(NC)** and a vector $\alpha = (\alpha_1,\dots,\alpha_K) \in \mathbb{R}_+^K$, we denote by π_i^α the solution of the following system (see Lemma 13)

$$\forall i \in [1,K], \quad \pi_i^\alpha = \alpha_i + \sum_{j=1}^K p_{j,i}\pi_j^\alpha.$$

Proposition 9. *Consider a sequence of Jackson networks as in Proposition 8 such that $\mu^{(k)} > 0$ for all k, $\Sigma^{(0)}(t) = \lambda t/c$ with $\lambda > 0$ and $c \geq 0$ (with the convention : $*/0 = +\infty$) and $X^n < +\infty$ for all n, we denote $\alpha = (n^{(1)} + n^{(0)}p_{0,1},\dots,n^{(K)} + n^{(0)}p_{0,K})$, we have*

$$\lim_{n\to\infty} \frac{X_c^n}{n} = \max_{1\leq i\leq K} \frac{\pi_i^\alpha}{\mu^{(i)}} \vee \frac{cn^{(0)}}{\lambda}.$$

Proof :
Lower bound :
Consider the auxiliary Jackson network $\tilde{\mathbf{JN}}^n = \{0, \nu^n, N^n\}$, and the associated vector $\mathbf{Y}(n)$, where $Y^{(i)}(n)$ is the total number of customers that go through node i in this network. We have

$$Y^{(i)}(n) = n^{(i),n} + P_{0,i}^n(n^{(0),n}) + \sum_{j=1}^K P_{j,i}^n(Y^{(j)}(n)).$$

Hence $\lim_n \frac{Y^{(i)}(n)}{n} = \pi_i^\alpha$ thanks to assumption **(NC)** on P.
Now consider the original network $\mathbf{JN}_c^n$. The number of customers that go through node i is still $Y^{(i)}(n)$. Hence we have the following inequality for the maximal dater of node $i \geq 1$, $X^{(i),n} \geq \sigma^{(i),n}(1, Y^{(i)}(n))$. And for node 0, $X_c^{(0),n} \geq \sigma_c^{(0),n}(1, n^{(0),n})$. Hence, we have

$$\liminf_{n\to\infty} \frac{X^{(i),n}}{n} \geq \lim_{n\to\infty} \frac{\sigma^{(i),n}(1,Y^{(i)}(n))}{n} = \frac{\pi_i^\alpha}{\mu^{(i)}},$$
$$\liminf_{n\to\infty} \frac{X_c^{(0),n}}{n} \geq \lim_{n\to\infty} \frac{\sigma_c^{(0),n}(1,n^{(0),n})}{n} = \frac{cn^{(0)}}{\lambda}.$$

Since $X_c^n = \max_{1\leq i\leq K} X^{(i),n} \vee X_c^{(0),n}$, the lower bound follows.
Upper bound :
We consider the original Jackson network. Thanks to Proposition 8, we know that the corresponding input and output processes $\mathbf{A}^n$ and $\mathbf{D}^n$ converge to a fluid limit $\hat{A}$ and $\hat{D}$ respectively. Let $T^{(i)} = \inf\{t > 0,\ \hat{A}^{(i)}(t) = \hat{D}^{(i)}(t)\}$, $T = \max_{i\in[1,K]} T^{(i)}$ and $M = T \vee cn^{(0)}/\lambda$. We have

$$\forall t \geq M, \quad \hat{A}^{(i)}(t) = n^{(i)} + p_{0,i}n^{(0)} + \sum_{j=1}^K p_{j,i}\hat{A}^{(j)}(t),$$

hence, we have

$$\forall t \geq M, \quad \hat{A}^{(i)}(t) = \hat{D}^{(i)}(t) = \pi_i^\alpha. \tag{3.10}$$

We denote $i_0 = \arg\max\{T^{(i)}\}$ and we have $\hat{A}^{(i_0)}(T) = \hat{D}^{(i_0)}(T) = \mu^{(i_0)}T$ by concavity of $\hat{A}^{(i_0)}$, hence $T = \frac{\pi^\alpha_{i_0}}{\mu^{(i_0)}}$. Moreover, Equation (3.10) implies :

$$\forall t \geq M, \quad \frac{Y^{(i)}(n) - D^{(i),n}(nt)}{n} \xrightarrow{n\to\infty} 0, \quad \text{where } Y^{(i)}(n) \text{ is the total number of customers that go through node } i.$$

Since $X^n_c < +\infty$, we know that for any t,

$$X^n_c \quad \leq \quad nt + \sum_{i=1}^{K} \sigma^{(i),n}(D^{(i),n}(nt), Y^{(i)}(n)) + \sigma_c^{(0),n}(\Sigma^{(0),n}(nt), n^{(0),n}),$$

taking $t = M$, we have $\limsup_{n\to\infty} \frac{X^n_c}{n} \leq M = T \vee \frac{cn^{(0)}}{\lambda} = \frac{\pi^\alpha_{i_0}}{\mu^{(i_0)}} \vee \frac{cn^{(0)}}{\lambda}$, and the result follows.
$\square$

proof of Theorem 2 :
It is easy to see that assumptions of Proposition 8 hold for the Jackson networks $\mathbf{JN}^n_c = \{\sigma_c, \nu, N^n\}$, with $n^{(i)} = 0$, except the $n^{(0)} = 1$.$\square$

Stability of Generalized Jackson Networks

We now give the connection between this fluid limit and the stability region of generalized Jackson networks under stationary ergodic assumptions following [12].
Assume that we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, endowed with an ergodic measure-preserving shift θ. Consider a sequence of simple Euler networks, say $\{E(n)\}_{n=-\infty}^{\infty}$ where $E(n) = \{\sigma(n), \nu(n), 1\}$.
Let $\xi(n) = \{\{\sigma(n)\}, \{\nu(n)\}\}$. The stochastic assumptions of Section 4.1 of [12] are as follows :

- the variables $\{\sigma(n)\}$, $\{\nu(n)\}$ are random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$;
- the random variable $\xi(n)$ satisfy the relation $\xi(n) = \xi(0) \circ \theta^n$ for all n, which implies that $\{\xi(n)\}_n$ is stationary and ergodic ;
- all the expectations $\mathbb{E}\left[\phi^{(k)}(0)\right]$ and $\mathbb{E}\left[\sum_{i=1}^{\phi^{(k)}(0)} \sigma_i^{(k)}(0)\right]$ are finite ($\phi^{(j)}(n)$ is obtained by **Procedure 2** on $E(n)$).

In such a setting, we can find Ω_0, such that on Ω_0 conditions (3.7), (3.8) and **(NC)** hold and $\mathbb{P}(\Omega_0) = 1$. Thanks to the strong law of large numbers, we have almost surely :

$$\begin{aligned}
\frac{\phi^{(j)}(1) + \cdots + \phi^{(j)}(n)}{n} \quad &\to \quad \mathbb{E}\left[\phi^{(j)}(0)\right] < +\infty, \\
\frac{\sum_{i=1}^{\phi^{(j)}(1)} \sigma_i^{(j)}(1) + \cdots + \sum_{i=1}^{\phi^{(j)}(n)} \sigma_i^{(j)}(n)}{n} \quad &\to \quad \mathbb{E}\left[\sum_{i=1}^{\phi^{(j)}(0)} \sigma_i^{(j)}(0)\right] < +\infty.
\end{aligned}$$

From these equations, we derive condition (3.7) :

$$\lim_{n\to\infty} \frac{\sigma^{(j)}(1,n)}{n} = \frac{\mathbb{E}\left[\sum_{i=1}^{\phi^{(j)}(0)} \sigma_i^{(j)}(0)\right]}{\mathbb{E}\left[\phi^{(j)}(0)\right]} := \frac{1}{\mu^{(j)}} \quad \text{a.s.}$$

With the same kind of arguments, we show that limit (3.8) holds almost surely. To show that P satisfies **(NC)**, we denote $V^{(j)} = \mathbb{E}\phi^{(j)}(0)$ and $V^{(j)}(n) = \phi^{(j)}(1) + \cdots + \phi^{(j)}(n)$ and thanks to the Euler's property of the graphs, we have $V^{(i)}(n) = P_{0,i}(n) + \sum_{j=1}^{K} P_{j,i}(V^{(j)}(n))$, hence $V^{(i)} = p_{0,i} + \sum_{j=1}^{K} p_{j,i} V^{(j)}$. Equation (3.9) has a finite solution, hence P satisfies **(NC)** and $V^{(i)} = \pi_i$ (see Lemma 13). Now we can define Ω_0 as follows :

$$\Omega_0 = \left\{ \frac{\sigma^{(k)}(1,n)}{n} \to \frac{1}{\mu^{(k)}},\ \frac{P_{i,j}(n)}{n} \to p_{i,j},\ \frac{V^{(j)}(n)}{n} \to \pi_j \right\}.$$

We will take the conventional notation : $\mu^{(0)} = \lambda$ for the intensity of the external arrival.
The limit calculated in Theorem 2 is exactly the constant $\delta(c)$ defined in Equation (85) of [12]. On the event Ω_0, Theorem 2 applies and gives a new proof of Theorem 15 of [12] which says that $\delta(0) = \gamma(0) = \max_i \pi_i/\mu^{(i)}$. Moreover, the lower bound of Lemma 6 (in [12]) is shown to be in fact the exact value of $\delta(c)$. Theorems 13 and 14 of [12] give the stability condition of a Jackson-type queueing networks in an ergodic setting. To be more precise : for $m \le n \le 0$, we define $\sigma_{[m,n]}$ and $\nu_{[m,n]}$ to be the concatenation of the $\{\sigma(k)\}_{m\le k\le n}$ and $\{\nu(k)\}_{m\le k\le n}$ and then define the corresponding generalized Jackson networks :

$$\mathbf{JN}_{[m,n]} = \{\sigma_{[m,n]}, \nu_{[m,n]}, N_{[m,n]}\}, \quad \text{with} \quad N_{[m,n]} = (n-m+1, 0, \dots, 0).$$

We define $X_{[m,n]}$ to be the time to empty the generalized Jackson network $\mathbf{JN}_{[m,n]}$ and $Z_{[m,n]} = X_{[m,n]} - \sum_{i=1}^{n-m+1} \sigma_{[m,n],i}^{(0)}$ the associated maximal dater. Note that notation is consistent with [12]. The sequence $Z_{[-n,0]}$ is an increasing sequence. So there exists a limit $Z = \lim_{n\to\infty} Z_{[-n,0]}$ (which may be either finite or infinite). We call this limit Z the maximal dater of the generalized Jackson network $\mathbf{JN} = \{\sigma, \nu, N\}$ where σ and ν are the infinite concatenation of the $\{\sigma(k)\}_{k\le 0}$ and $\{\nu(k)\}_{k\le 0}$ and $N = (+\infty, 0, \dots, 0)$. Let A be the event

$$A = \{Z = \lim_{n\to\infty} Z_{[-n,0]} = \infty\}. \tag{3.11}$$

This event is of crucial interest since a finite stationary construction of the state of the network can only be made on the complementary part of A. In other words, $Z < \infty$ iff the network is stable. The following Theorem follows from Theorem 13 and 14 of [12]

Theorem 3. *Let* $\rho = \lambda \max_{1\le i\le K} \frac{\pi_i}{\mu^{(i)}}$.
If $\rho < 1$, *then* $\mathbb{P}(A) = 0$.
If $\rho > 1$, *then* $\mathbb{P}(A) = 1$.

Remark 10. There exists a parallel stream of work which uses sample path methods (quite different from the one described in this paper) to prove a weaker form of stability called pathwise stability or rate stability. Rate stability means that the long-run average departures must equal the long-run average arrivals at each station with probability 1. In Chen [25], it is proved that for a multiclass queueing network under work-conserving service disciplines, the weak stability of the fluid model implies the rate stability of the stochastic network. We refer to the paper of Chen [25] for a detailed definition of fluid model and weak stability ; the main result of [25] is that under the usual traffic conditions, a generalized Jackson network is rate stable. Further in Dai [30], under weak strong law of large numbers assumptions, it is proved that if $\rho > 1$ (with our notation) then the number of customers in the network diverges to infinity with probability 1 as time $t \to \infty$ (see Proposition 5.1 of [30]). This result corresponds to the second part of our Theorem. Anyway to prove that

$\rho < 1$ ensures stability of the generalized Jackson network, Dai in [29] needs i.i.d. assumptions and additional conditions on the inter-arrival times that are unbounded and spread out. In this section, we use fluid limits to derive the same result under stationary and ergodic conditions only.

3.1.4 Rare Events in Generalized Jackson Networks

The aim of this section is to give a picture of one kind of rare event when the maximal dater of a generalized Jackson network becomes very big. Under some stochastic assumptions, one can prove that large maximal daters occur when a single large service time has taken place in one of the stations, and all other service times are close to their mean see next chapter. We now give the corresponding fluid picture.

The One Big Jump Framework

We consider a sequence of simple Euler networks, say $\{E(n)\}_{n=-\infty}^{+\infty}$ where $E(n) = \{\sigma(n), \nu(n), 1\}$. Considering the corresponding $\mathbf{JN}_{[-n,+\infty]}$ network, we assume that

$$\hat{\Sigma}^{(0),n}(t) \rightarrow t/a, \quad \forall t, \tag{3.12}$$

$$\forall k \geq 1, \quad \hat{\Sigma}^{(k),n}(t) \rightarrow \mu^{(k)} t, \quad \forall t, \tag{3.13}$$

$$\forall i,j, \quad \hat{P}^n_{i,j}(t) \rightarrow p_{i,j} t, \quad \forall t. \tag{3.14}$$

We assume that $P = (p_{i,j})_{1 \leq i,j \leq K}$ satisfies **(NC)** and we take the following notation :

$$\forall i \in [1,K], \quad \pi_i = p_{0,i} + \sum_{k=1}^{K} p_{k,i} \pi_k, \tag{3.15}$$

$$\forall i \in [1,K], \quad x_i = p_{0,i} + \sum_{k \neq j} p_{k,i} x_k \Rightarrow x_j = p_j, \tag{3.16}$$

$$\forall i \in [1,K], \quad \pi_{j,i} = \delta_{j,i} + \sum_{k=1}^{K} p_{k,i} \pi_{j,k}. \tag{3.17}$$

Equation (3.15) is the traditional traffic equation of the network in term of number of customers. In Equation (3.16), p_j corresponds to the amount of traffic coming in queue j if this one is blocked (its departure process is null). Note that in this case $x_i \leq \pi_i$. Equation (3.17) corresponds to the traffic equation in the network when there is no input from the outside world and only queue j is active. We introduce the corresponding loads :

$$b_j = \frac{\pi_j}{\mu^{(j)}}, \quad b = \max_j b_j \quad \text{and,} \quad b_{j,i} = \frac{\pi_{j,i}}{\mu^{(i)}}, \quad B_j = \max_i b_{j,i}.$$

We assume that the stability condition $b < a$ holds. We suppose that the big jump occurs in the simple Euler network $-n$, hence we replace $E(-n)$ by an extra E which is not "typical" in the following sense : a big service time σ takes place on station j and within the set of service times of the simple Euler network E. Let us look at the corresponding maximal dater $Z_{[-n,0]}(E)$ in the fluid scale suggested by the limit of Proposition 9 :

– if $\sigma > na$, then the number of customers blocked in station j at time σ is of the order of np_j, whereas the number of customers in the other stations is small. So, according to Proposition9, the time to empty the network from time σ on should be of the order np_jB_j ; hence, in this case, the maximal dater in question should be of the order of $\sigma - na + np_jB_j$;

– if $\sigma < na$, then at time σ, the number of customers blocked in station j is of the order of $p_j \frac{\sigma}{a}$, and the other stations have few customers ; from time σ to the time of the last arrival (which is of the order of na), station k has to serve approximately the load $p_j \frac{\sigma}{a} b_{j,k}$ generated by these blocked customers plus the load $(na - \sigma)\frac{b_k}{a}$ generated by the external arrivals on the time interval from σ to the last arrival. On this time interval, the service capacity is of the order of $(na - \sigma)$. Hence the maximal dater should be of the order of $\max_k \left\{ p_j b_{j,k} \frac{\sigma}{a} + (na - \sigma)\frac{b_k}{a} - (na - \sigma) \right\}^+$.

It is now natural to introduce the following function :

$$f^j(\sigma, n) = \mathbb{1}_{\{\sigma > na\}} \{\sigma - na + np_j B_j\} + \mathbb{1}_{\{\sigma \le na\}} \max_k \left\{ p_j b_{j,k} \frac{\sigma}{a} + \left(\frac{b_k}{a} - 1\right)(na - \sigma)\right\}^+ \quad (3.18)$$

We now return to rigor and consider the network $\mathbf{JN}^n(E)$ with input $\{\tilde{E}(k)\}_{k=-n}^{\infty}$, where $\tilde{E}(k) = E(k)$ for all $k > -n$ and $\tilde{E}(-n) = E$. That is, if we denote by $\sigma^{(k),n}$ and $\nu^{(k),n}$ the concatenations $(\{\sigma^{(k)}(E)\}, \{\sigma^{(k)}(-n+1)\}, \ldots, \{\sigma^{(k)}(0)\}, \ldots)$ and $(\{\nu^{(k)}(E)\}, \{\nu^{(k)}(-n+1)\}, \ldots, \{\nu^{(k)}(0)\}, \ldots)$ respectively, then

$$\mathbf{JN}^n(E) \quad = \quad \{\sigma^n(E), \nu^n(E), N^n\}, \quad \text{with} \quad N^n = (n, 0, \ldots, 0).$$

The maximal dater of order $[-n, 0]$ in this network will be denoted by $\tilde{Z}^n(E)$. Of course $\tilde{Z}^n(E(n)) = Z_{[-n,0]}$. For all simple Euler networks $E = (\sigma, \nu, 1)$, let $Y^{(j)}(E) = \sum_{u=1}^{\phi^{(j)}} \sigma_u^{(j)}$.

Let z_n be some sequence of positive real numbers, we define :

$$\begin{aligned}
\mathbf{U}^j(n) &= \{E \text{ is a simple Euler network such that } Y^{(k)}(E) \le z_n \ \forall k \ne j\}, \\
\mathbf{V}^j(n) &= \{E \in \mathbf{U}^j(n),\ Y^{(j)}(E) \ge n(a-b),\ \phi^{(j)} \le L\},
\end{aligned}$$

Proposition 10. *Under the previous assumptions, there exists a sequence $z_n \to \infty$ with $\frac{z_n}{n} \to 0$, such that we have*

$$\sup_{E \in \mathbf{V}^j(n)} \left| \frac{\tilde{Z}^n(E) - f^j(Y^{(j)}(E), n)}{n} \right| \xrightarrow{n \to \infty} 0. \quad (3.19)$$

3.1.5 Computation of the Fluid Limit

We take a sequence of simple Euler networks $F_n \in \mathbf{V}^j(n)$ and denote by $\mathbf{JN}^n = \mathbf{JN}^n(F_n)$. Since z_n/n tends to 0, we have

$$\begin{aligned}
& \hat{\Sigma}^{(0),n}(t) \ \to\ t/a, \quad \forall t, \quad \text{a.s.} \\
\forall k \ne j \ge 1, \quad & \hat{\Sigma}^{(k),n}(t) \ \to\ \mu^{(k)} t, \quad \forall t, \quad \text{a.s.} \\
\forall i, j, \quad & \hat{P}_{i,j}^n(t) \ \to\ p_{i,j} t, \quad \forall t, \quad \text{a.s.}
\end{aligned}$$

We denote by $\zeta_n = Y^{(j)}(F_n) \in [n(a-b), +\infty)$ and by T_n the time for station j to complete its $\phi^{(j)}(F_n)$ first services in the network $\mathbf{JN}^n$. From monotonicity, we get $\zeta_n \le T_n \le \zeta_n + \sum_{k \ne j} Y^{(k)}(F_n)$. Hence, we have $\lim_{n\to\infty} \frac{T_n}{\zeta_n} = 1$, since $z_n/\zeta_n \le z_n/n(a-b) \to 0$. We first suppose that $\zeta_n/n \to \zeta < +\infty$. Then $\mathbf{JN}^n$ is such that $\Sigma^{(j),n}(t) \le L$ for $t \le T_n$. Hence

$\frac{\Sigma^{(j),n}(nt)}{n} \leq \frac{L}{n}$ for $nt \leq T_n$, so that $\hat{\Sigma}^{(j),n}(t) \to 0$, for all $t \leq \zeta$. We see that this last fluid limit does not hold on the whole positive real line. Nevertheless consider the Jackson networks with the same driving sequences as $\mathbf{JN}^n$ except for station j where we take the concatenation of $(\{\sigma^{(j)}(F_n)\}, \infty, \dots)$. For this new network, the fluid limit for station j holds on $\mathbb{R}_+$ and we can directly apply Proposition 8. But it is easy to see that for $t \leq T_n$, this network and the original Jackson network $\mathbf{JN}^n$ have exactly the same dynamic. Hence, Proposition 8 applies for $t \leq \zeta$, so that for each k, the sequence $\{\hat{A}^{(k),n}\}$ converges u.o.c. to a limit $\hat{A}^{(k)}$ when n tends to ∞, with a similar result and notation for the departure process. We have with $\lambda = a^{-1}$,

$$\begin{aligned} \hat{A}^{(i)}(t) &= p_{0,i}\lambda(t \wedge a) + \sum_{k=1}^{K} p_{k,i}\hat{D}^{(k)}(t), \\ \hat{D}^{(i)}(t) &= \hat{A}^{(i)}(t) \wedge \tilde{\mu}^{(i)} t \quad \text{with } \tilde{\mu}^{(i)} = \mu^{(i)} \text{ for } i \neq j \text{ and } \tilde{\mu}^{(j)} = 0. \end{aligned}$$

We can rewrite the first expression :

$$\hat{A}^{(i)}(t) = \lambda p_{0,i}(t \wedge a) + \sum_{k \neq j} p_{k,i}\hat{D}^{(k)}(t).$$

Hence with the notation introduced in previous section, we have

$$\begin{aligned} \hat{A}^{(i)}(t) &= \hat{D}^{(i)}(t) = \lambda x_i(t \wedge a) \leq \lambda \pi_i(t \wedge a) \quad \text{for} \quad t \leq \zeta \quad \text{and} \quad i \neq j, \\ \hat{A}^{(j)}(t) &= \lambda p_j(t \wedge a) \quad \text{for} \quad t \leq \zeta. \end{aligned}$$

In what follows, we will consider the new Jackson network obtained by taking as initial condition the state of the initial network at time T_n and as routing and service sequences the routing decisions and (residual) service still unused at this time. This network will be denoted by $\bar{\mathbf{JN}}^n = \{\bar{\sigma}^n, \bar{\nu}^n, \bar{N}^n\}$, with

$$\begin{aligned} \bar{\sigma}^{(0),n} &= \left\{\Sigma^{(0),n\leftarrow}(\Sigma^{(0),n}(T_n)+1) - T_n, \sigma^{(0),n}_{\Sigma^{(0),n}(T_n)+2}, \dots\right\}, \\ \bar{\nu}^{(0),n} &= \left\{\nu^{(0),n}_{\Sigma^{(0),n}(T_n)+1}, \nu^{(0),n}_{\Sigma^{(0),n}(T_n)+2}, \dots\right\}, \\ \bar{N}^{(0),n} &= n - \Sigma^{(0),n}(T_n), \end{aligned}$$

and for $i \neq 0$,

$$\begin{aligned} \bar{\sigma}^{(i),n} &= \left\{r^{(i),n}, \sigma^{(i),n}_{D^{(i),n}(T_n)+2}, \dots\right\}, \\ \bar{\nu}^{(i),n} &= \left\{\nu^{(i),n}_{D^{(i),n}(T_n)+1}, \nu^{(i),n}_{D^{(i),n}(T_n)+2}, \dots\right\}, \\ \bar{N}^{(i),n} &= A^{(i),n}(T_n) - D^{(i),n}(T_n), \\ r^{(i),n} &= \begin{cases} \sigma^{(i),n}_{D^{(i),n}(T_n)+1} & \text{if } A^{(i),n}(T_n) = D^{(i),n}(T_n), \\ D^{(i),n\leftarrow}\big(D^{(i),n}(T_n)+1\big) - T_n & \text{else.} \end{cases} \end{aligned}$$

We have :

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} A^{(i),n}(T_n) &= \lim_{n\to\infty} \frac{1}{n} D^{(i),n}(T_n) = \hat{A}^{(i)}(\zeta) = \hat{D}^{(i)}(\zeta) \quad \text{for } i \neq j, \\ \lim_{n\to\infty} \frac{1}{n} A^{(j),n}(T_n) &= \lambda p_j(\zeta \wedge a) \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{n} D^{(j),n}(T_n) = 0. \end{aligned}$$

Hence, we have

$$\begin{array}{rcl} \dfrac{\bar{N}^n}{n} & \to & (\lambda(a-\zeta)^+, 0, \dots, \lambda p_j(\zeta \wedge a), \dots, 0) \\ \hat{\hat{\Sigma}}^{(0),n}(t) & \to & \lambda t, \quad \forall t, \\ \forall i \geq 1, \quad \hat{\hat{\Sigma}}^{(i),n}(t) & \to & \mu^{(i)} t, \quad \forall t, \\ \forall i,j, \quad \hat{\hat{P}}^n_{i,j}(t) & \to & p_{i,j} t, \quad \forall t. \end{array}$$

We can apply Proposition 9 with a parameter α that depends on the quantity $a - \zeta$:

– if $\zeta \geq a$, then we have $\alpha = p_j e_j$, with $e_j = (0, \dots, 1, \dots, 0)$ with the one is in j-th position and

$$\pi_i^\alpha = p_j \pi_{j,i}.$$

Proposition 9 gives

$$\frac{\tilde{Z}_n(F_n) + na - T_n}{n} \quad \to \quad p_j \max_i \frac{\pi_{j,i}}{\mu^{(i)}}. \tag{3.20}$$

Hence, we have

$$\tilde{Z}_n(F_n) \;=\; (T_n - na + np_j B_j)(1 + o(n)) = f^j(T_n, n)(1 + o(n)). \tag{3.21}$$

– if $\zeta < a$, we have $\alpha = \lambda(a-\zeta)P_0 + \lambda p_j \zeta e_j$, where $P_0 = (p_{0,1}, \dots, p_{0,K})$ and

$$\pi_i^\alpha = \lambda \left[(a - \zeta)\pi_i + p_j \pi_{j,i} \zeta \right].$$

Proposition 9 gives

$$\frac{\tilde{Z}_n(F_n) + na - T_n}{n} \to (a-\zeta) \vee \lambda \max_i \left[\frac{(a-\zeta)\pi_i + p_j \pi_{j,i}\zeta}{\mu^{(i)}} \right].$$

Hence, we have

$$\begin{array}{rcl} \tilde{Z}_n(F_n) & = & (1 + o(n)) \max_i \left[p_j b_{j,i} \dfrac{T_n}{a} + (na - T_n)(\dfrac{b_i}{a} - 1) \right]^+ \\ & = & f^j(T_n, n)(1 + o(n)). \end{array}$$

The case $\zeta_n/n \to \infty$ corresponds to $\zeta = \infty$. Results until Equation (3.20) hold true in this context, hence Equation (3.21) holds true.

Finally we proved that for any sequence $F_n \in \mathbf{V}^j(n)$ with $Y^{(j)}(F_n) = \zeta_n \in [n(a-b), +\infty)$ such that $\zeta_n/n \to \zeta \leq +\infty$,

$$\left| \frac{\tilde{Z}_n(F_n) - f^j(\zeta_n, n)}{n} \right| \xrightarrow{n \to \infty} 0. \tag{3.22}$$

But the result holds for any sequence $F_n \in \mathbf{V}^j(n)$. Consider any sequence $F_n \in \mathbf{V}^j(n)$ and suppose that

$$\limsup_{n \to \infty} \left| \frac{\tilde{Z}_n(F_n) - f^j(Y^{(j)}(F_n), n)}{n} \right| = l > 0.$$

By extracting a subsequence of $\{F_n\}$, we can replace lim sup by lim. Moreover by doing once more an extraction, we may suppose that $Y^{(j)}(F_n)/n \to \zeta \leq +\infty$ and for this subsequence, limit (3.22) is violated. Hence for any sequence F_n, we have

$$\left|\frac{\tilde{Z}_n(F_n) - f^j(Y^{(j)}(F_n), n)}{n}\right| \xrightarrow{n\to\infty} 0. \tag{3.23}$$

We consider now a sequence $F_n \in \mathbf{V}^j(n)$ such that

$$\left|\frac{\tilde{Z}_n(F_n) - f^j(Y^{(j)}(F_n), n)}{n}\right| \geq \sup_{\{E\in\mathbf{V}^j(n)\}} \left|\frac{\tilde{Z}_n(E) - f^j(Y^{(j)}(E), n)}{n}\right| - \epsilon_n,$$

with $\epsilon_n \to 0$. Thanks to (3.23), we see that (15) holds. □

Remark 11. In the stochastic framework of section 6.2, we see that assumptions on the limits (3.12), (3.13) and (3.14) are fulfilled. In particular, if the sequence of simple Euler networks $\{E(n)\}_{n=-\infty}^{+\infty}$ is i.i.d, then we deduce from previous proposition that :

$$\sup_{\{E\in\mathbf{V}^j(n)\}} \left|\frac{\tilde{Z}^n(E) - f^j(Y^{(j)}(E), n)}{n}\right| \xrightarrow{n\to\infty} 0 \quad \text{a.s.}$$

3.2 Fluid Limit for GPS Queues

The purpose of this section is to construct the stationary workload at each queues of a GPS system under fairly general stochastic assumptions, namely stationarity and ergodicity. This construction is quite simple in the case $\rho < 1$ where ρ is the total load of the system. In the case $\rho \geq 1$, we show that there are still some queues that can be stable in the following sense : for any initial condition the workload process of these queues couples in finite time with an unique stationary workload process. For the unstable queues, we show the existence of a mean service rate and give its expression in a closed form formula. To the best of our knowledge there is no such result available in the literature. With this work, the stability of GPS systems is fully understood.

The other application of this section will be linked to the calculation of tails in GPS systems with subexponential service distributions in next chapter. We are able to give here the behavior (in the fluid scale) of the system on a “rare” event. We refer to the next chapter for an exact notion of what we mean by rare event. Note that the work of Dupuis and Ramanan [38] [39] allows to construct the transient fluid limit of a GPS system.

3.2.1 Construction of the Stationary Regime

Since we will make an extensive use of notation introduced in Section 2.4, we repeat it here.

Consider the following model of N coupled $G/G/FIFO$ queues. Each queue is served in accordance with the Generalized Processor Sharing (GPS) discipline, which operates as follows. Queue j is assigned a weight ϕ^j, with $\sum_{j=1}^N \phi^j = 1$. If all queues are backlogged, then queue j is served at speed ϕ^j. If some of the queues are empty, then the excess capacity is redistributed among the backlogged queues in proportion to their respective weights. All customers within each queue are served in a FIFO order.

More formally we can construct the workload of each queues as follows. Let $\{T_n^A, \sigma_n, c_n\}$ be a simple marked process, with $\sigma_n > 0$ and $c_n \in \{1, \ldots, N\}$. The interpretations are the following : customer n arrives in the queue c_n at time T_n^A and its service time is σ_n. We will say that this customer is of class $c_n \in \{1, \ldots, N\}$ and denote by $\tau_n = T_{n+1}^A - T_n^A$ the inter-arrival times. We denote by $W_Y^i[n] := W_Y^i(T_n^A-)$ the workload of queue i at time T_n^A- with initial condition $W_Y^i[0] = Y^i$. The sequence $\{W_Y[n] = (W_Y^1[n], \ldots, W_Y^N[n])\}$ is generated by the recurrence

$$W_Y[n+1] = h(W_Y[n], \sigma_n, c_n, \tau_n), \quad n = 0, 1 \ldots$$

where the function h is defined by the following equations :

$$\begin{aligned} W^j(T_k^A) &= W^j(T_k^A-) + \sigma_k \mathbf{1}_{\{c_k = j\}}, & (3.24)\\ \frac{dW^j}{dt}(t) &= -r^j(t) \quad \text{for } T_k^A \le t < T_{k+1}^A, & (3.25)\\ r^j(t) &= \begin{cases} \frac{\phi^j}{\sum_{\ell \notin I(t)} \phi^\ell} & j \notin I(t), \\ 0 & j \in I(t); \end{cases} & (3.26)\\ I(t) &= \{i,\ W^i(t) = 0\}. & (3.27) \end{aligned}$$

Equations (3.24), (3.25), (3.26) and (3.27) show how to construct the workload process of each queue for $t \ge T_0^A$.

Note that we have

$$\sum_i W_Y^i[n+1] = \left(\sum_i W_Y^i[n] + \sigma_n - \tau_n \right)^+,$$

the recurrence for the sum of the component of $W_Y[n]$ reduces to the Lindley's equation.

The stability of the GPS queues follows directly from the stability of the single server queue with input process $\{T_n^A, \sigma_n\}_{n\in\mathbb{Z}}$, since the sum of the workload of each queue is exactly the workload of this single server queue.

But if the single server queue is unstable, there are still some stable queues in the GPS system. In this section we show this result by constructing the corresponding stationary workload of these queues.

We first recall some basic results on the single server queue and refer to Chapter 2 of [10] for more details on the next result. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with measurable flow θ_t, $t \in \mathbb{R}$, such that $(\mathbb{P}, \theta_t)$ is ergodic. Let T^A be a point process defined on $(\Omega, \mathcal{F})$. Assume T^A is simple and compatible with $\{\theta_t\}$. We assume that this arrival process has finite intensity λ and let the sequence (σ, c) be a sequence of marks of the arrival process that describes the amount of required service and the class of customer n. Let $\rho = \lambda \mathbb{E}_{T^A}^0[\sigma_0]$ be the traffic intensity. The process (T^A, σ, c) can be obtained by the superposition of independent point processes of finite intensity (see Section 1.4.2 of [10]).

For the $G/G/1/\infty$ queue, the evolution of the workload process $W(t)$ between two successive arrivals is described by Lyndley's equation :

$$W(t) = (W(T_n^A-) + \sigma_n - (t - T_n^A))^+, \quad t \in [T_n^A, T_{n+1}^A), \tag{3.28}$$

where $a^+ = \max(a, 0)$.

Theorem 4. *Under the stability condition*

$$\rho < 1,$$

there exists a unique finite workload process $\{W(t)\}$, $t \in \mathbb{R}$, *compatible with the flow* $\{\theta_t\}$, *and satisfying equation (3.28) for all* $t \in \mathbb{R}$. *This process is such that*

$$W(0) = \sup_{n \le 0} \left(T_n + \sum_{i=0}^{n} \sigma_i \right)^+ .$$

Moreover, there are an infinite number of negative indices n *and an infinite number of positive indices* n *such that*

$$W(T_n^A -) = 0.$$

If the traffic intensity ρ *is strictly larger than 1, there exists no finite* $\mathbb{P}-$*stationary workload process* $\{W(t)\}$, $t \in \mathbb{R}$.

Returning to our GPS model, if $\rho < 1$, it is easy to construct the workload process of each queue (compatible with θ_t). Let $\{W(t)\}$, $t \in \mathbb{R}$, be the unique $\mathbb{P}-$ stationary workload associated with $\{T_n^A, \sigma_n\}_{n\in\mathbb{Z}}$. The point process E defined by

$$E(B) = \sum_{n\in\mathbb{Z}} \mathbf{1}_{\{T_n^A \in B\}} \mathbf{1}_{\{W(T_n^A-)=0\}},$$

counts the points T_n^A at which an arriving customer finds an empty system. Clearly E is compatible with $\{\theta_t\}$. Let $\{U_n\}$, $n \in \mathbb{Z}$, be the sequence of points of E, with the usual convention $U_0 \le 0 < U_1$. Then we can construct the unique $\mathbb{P}$-stationary workload process $\{(W^1(t), \dots, W^N(t))\}$ of the GPS queues using the mapping h we defined above on each cycle $[U_n, U_{n+1})$ with initial condition 0.

In the case $\rho > 1$, we can still construct a workload process for the "stable" queues of the GPS system.

We are looking for a random variable $Z \ge 0$ which verifies the functional equation :

$$Z \circ \theta = h(Z, \sigma, c, \tau), \quad \mathbb{P}^0_{T^A}\text{-a.s.}, \tag{3.29}$$

where $\theta = \theta_{T_1}$, $\sigma = \sigma_0$, $c = c_0$ and $\tau = \tau_0$. Recall that the mapping h was defined above section and that Z is a vector of size N. If there exists such a random variable with at least one finite component, we define the workload sequence $\{W[n]\}$ by

$$W[n] = Z \circ \theta^n, \quad n \in \mathbb{Z},$$

and the associated workload process $\{W(t)\}$ by equations (3.24), (3.25), (3.26), (3.27), and with

$$W(T_n^A -) = W[n].$$

The fact that $W(T_n^A-)$ is indeed the limit as $t \to T_n^A$, $t < T_n^A$ of $W(t)$ follows from (3.29). We refer to Section 2.2.1 of [10] for more details.

Indeed with the stochastic assumptions we made, the sequence $\{W_Y[n]\}$, $n \ge 0$ is generated by a stochastic recurrence as defined in Section 2.5 of [10] :

$$W_Y[n+1] = h(W_Y[n], \sigma_n, c_n, \tau_n), \quad n = 0, 1 \dots \tag{3.30}$$

Moreover we are in the framework of Section 2.5.2 of [10]. The state space is $\mathbb{R}_+^N$ and $\le$ is the coordinate-wise partial ordering and $0 = (0, \dots, 0)$. The mapping h is such that

– $0 \leq h(W, \sigma, c, \tau)$, for all $W \in \mathbb{R}_+^N$ and (σ, c, τ) ;
– $W \leq W'$ implies $h(W, \sigma, c, \tau) \leq h(W', \sigma, c, \tau)$, for all (σ, c, τ) ;
– for all (σ, c, τ), $W \mapsto h(W, \sigma, c, \tau)$ is a continuous mapping from $\mathbb{R}_+^N$ to itself.

Hence we can use a Loynes ’ sequence (as in the single server case) to find a minimal stationary solution to (3.30). However the Loynes variable associated with the stochastic recurrence will not belong to $\mathbb{R}_+^N$ when $\rho > 1$. Some of its components will be infinite but we will show in the next theorem that some components are finite almost surely. To make it precise, we need some additional notation.

Let λ^ℓ be the intensity of the point process $T^{A,\ell}$ that counts the points of $\{T_n^A\}$ with mark $c_n = \ell$. Let $\rho^\ell = \lambda^\ell \mathbb{E}^0_{T^{A,\ell}}[\sigma_0]$. We have $\rho = \sum_{\ell=1}^N \rho^\ell$, see section 1.4.3 of [10]. We assume without loss of generality that

$$\frac{\rho^1}{\phi^1} \leq \cdots \leq \frac{\rho^N}{\phi^N}. \tag{3.31}$$

Define

$$\begin{aligned}
R_k &= \frac{1 - \sum_{j=1}^{k-1} \rho^j}{\sum_{j=k}^N \phi^j}, \\
K &= \max_{k=1,\dots,N} \left\{ \frac{\rho^k}{\phi^k} < R_k \right\}, \\
S &= \{1, \dots, K\}, \\
R &= \frac{1}{\sum_{j \notin S} \phi^j} \left(1 - \sum_{j \in S} \rho^j \right).
\end{aligned}$$

We will show that S is indeed the set of “stable” queues. This set is empty iff $\rho^1/\phi^1 \geq 1$, in this case we take $K = 0$ and $R = 1$.

For any k, we will consider the GPS system where queues $i > k$ are always backlogged, i.e. for all $j \leq k$,

$$\frac{d\bar{W}^{j,[k]}}{dt}(t) = -r^{j,[k]}(t) \quad \text{for } T_n^A \leq t < T_{n+1}^A \tag{3.32}$$

$$\bar{W}^{j,[k]}(T_n^A) = \bar{W}^{j,[k]}(T_n^A-) + \sigma_n \mathbb{1}_{\{c_n = j\}} \tag{3.33}$$

$$r^{j,[k]}(t) = \begin{cases} \frac{\phi^j}{\sum_{\ell \notin I^{[k]}(t)} \phi^\ell} & j \notin I^{[k]}(t), \\ 0 & j \in I^{[k]}(t) \end{cases} \tag{3.34}$$

$$I^{[k]}(t) = \left\{ i \leq k,\ \bar{W}^{i,[k]}(t) = 0 \right\}. \tag{3.35}$$

Note that for all $\ell > k$, we have $\ell \notin I^{[k]}(t)$ for all t. The interpretation for it is that queues with indice larger than $k+1$ are always backlogged.

Theorem 5. *Under previous condition on the input process (T^A, σ, c), we have the following properties :*

– *there exists a unique finite workload process $\{(\bar{W}^{1,[K]}(t), \dots, \bar{W}^{K,[K]}(t))\}$, $t \in \mathbb{R}$, compatible with the flow $\{\theta_t\}$, and satisfying equations (3.32), (3.33), (3.34) and (3.35) for all*

$t \in \mathbb{R}$ with $k = K$. Moreover, there are an infinite number of negative indices n and an infinite number of positive indices n such that

$$\sum_{i=1}^{K} \bar{W}^{i,[K]}(T_n^A-) = 0.$$

– *if $K+1 \leq N$, under the additional condition $\frac{\rho^{K+1}}{\phi^{K+1}} > R$, there exists no finite $\mathbb{P}-$stationary workload process $\{W^i(t)\}$, $t \in \mathbb{R}$ for any $i \geq K+1$. For any finite initial condition $Y \in \mathbb{R}_+^N$, we can define the workload of each queue for $t \geq 0$, following equations (3.24), (3.25), (3.26) and (3.27), and we have for $i \geq K+1$*

$$W_Y^i(t) \sim (\rho^i - \phi^i R)t \quad \textit{as } t \to \infty.$$

Proof.

If $\rho < 1$, then $K = N$ and the result follows from previous construction on the cycles.

We assume now that $1 \leq K \leq N-1$. The proof will proceed as follows : for each $k \leq K$, we will show that there exists an unique finite workload process $(\bar{W}^{1,[k]}(t), \dots, \bar{W}^{k,[k]}(t))$ compatible with the flow $\{\theta_t\}$ and that corresponds to a GPS system where queues $k+1, \dots, N$ are always backlogged. Moreover there are an infinite number of negative indices n and an infinite number of positive indices n such that $\sum_{i \leq k} \bar{W}^{i,[k]}(T_n^A-) = 0$.

For simplicity we will note $\bar{W}^{i,[k]}(t) = \bar{W}^{i,[k]}(0) \circ \theta_t := \bar{W}^{i,[k]} \circ \theta_t$.

For $t \geq 0$, we will denote by $\bar{W}_Y^{[k]}(t) = (\bar{W}_Y^{1,[k]}(t), \dots \bar{W}_Y^{k,[k]}(t))$ the process that satisfies equations (3.32), (3.33), (3.34) and (3.35) for $t \geq 0$ and with initial condition $\bar{W}_Y^{i,[k]}(0) = Y^i$.

The first step is easy. We have $1 \in S$, hence $\rho^1 < \phi^1$. Thanks to Theorem 4, there exists a unique workload process $\tilde{W}^1 \circ \theta_t$ that satisfies

$$\tilde{W}^1(t) = \left(\tilde{W}^1(T_n^{A,1}-) + \sigma_n^1 - \phi^1(t - T_n^{A,1})\right)^+, \quad t \in [T_n^{A,1}, T_{n+1}^{A,1}).$$

We have clearly $\bar{W}^{1,[1]} = \tilde{W}^1$ and we have $\bar{W}^{1,[1]}(T_n^{A,1}-) = 0$ for infinitely many positive or negative n.

For the second step, let define the following random variable :

$$\tilde{r}^2 := \mathbb{1}_{\{\bar{W}^{1,[1]}>0\}} + \frac{1}{\sum_{j \neq 1} \phi^j} \mathbb{1}_{\{\bar{W}^{1,[1]}=0\}}.$$

We have

$$\begin{aligned}
\mathbb{E}[\tilde{r}^2] &= \mathbb{P}(\bar{W}^{1,[1]} > 0) + \frac{\mathbb{P}(\bar{W}^{1,[1]} = 0)}{\sum_{j \neq 1} \phi^j} \\
&= \frac{\rho^1}{\phi^1} + \frac{\phi^1 - \rho^1}{\phi^1(\sum_{j \neq 1} \phi^j)} \\
&= \frac{1 - \rho^1}{\sum_{j \neq 1} \phi^j} = R_2.
\end{aligned}$$

Now consider the following recursion

$$\tilde{W}^2(t) = \left(\tilde{W}^2(T_n^{A,2}-) + \sigma_n^2 - \phi^2 \int_{[T_n^{A,2},t)} \tilde{r}^2 \circ \theta_u du\right)^+, \quad t \in [T_n^{A,2}, T_{n+1}^{A,2}).$$

We have

$$\mathbb{E}^0_{T^{A,2}}\left[\int_{[0,T_1^{A,2})}\tilde{r}^2\circ\theta_u du\right]=\frac{R_2}{\lambda_2},$$

hence we have $2\in S$ implies that $\tilde{W}^2$ is stable in the sense of Theorem 4. We denote $\tilde{W}^2\circ\theta_t$ the unique workload process compatible with the shift.

We return now to our GPS system where queues $3,4,\dots,N$ are always backlogged. Observe that the following functions

$$\bar{W}_0^{1,[2]}(t)\circ\theta_{-t}\quad\text{and}\quad\bar{W}_0^{2,[2]}(t)\circ\theta_{-t}\quad\text{are increasing in } t.$$

Moreover we have (just by looking at the service rates)

$$\bar{W}_0^{1,[2]}(t)\circ\theta_{-t}\le\bar{W}_0^{1,[1]}(t)\circ\theta_{-t}\xrightarrow{t\to\infty}\bar{W}^{1,[1]}.$$

And since $\bar{W}_0^{1,[2]}(u)\le\bar{W}_0^{1,[1]}(u)\le\bar{W}^{1,[1]}(u)$, we have by looking at the service rate again,

$$\bar{W}_0^{2,[2]}(t)\circ\theta_{-t}\le\tilde{W}_0^{2}(t)\circ\theta_{-t}\xrightarrow{t\to\infty}\tilde{W}^{2}.$$

Hence we proved that

$$\begin{array}{rcl}\bar{W}_0^{1,[2]}(t)\circ\theta_{-t} & \xrightarrow{t\to\infty} & \bar{W}^{1,[2]}\le\bar{W}^{1,[1]},\\ \bar{W}_0^{2,[2]}(t)\circ\theta_{-t} & \xrightarrow{t\to\infty} & \bar{W}^{2,[2]}\le\tilde{W}^{2},\end{array}$$

where $(\bar{W}^{1,[2]},\bar{W}^{2,[2]})\circ\theta_t$ corresponds to a GPS system where queues $3,\dots,N$ are always backlogged.

We have that if $\mathbb{P}(\bar{W}^{1,[1]}+\tilde{W}^2=0)=0$ then $\rho^2\ge\phi^2R_2$. This follows from the fact that if $\bar{W}^{1,[1]}+\tilde{W}^2>0$ $\mathbb{P}$-a.s., then we have

$$\begin{array}{rcl}\rho^1+\rho^2 & \ge & \lim\limits_{t\to\infty}\dfrac{1}{t}\displaystyle\int_0^t\left(\phi^1\mathbf{1}_{\{\bar{W}^{1,[1]}(u)>0\}}+\phi^2\tilde{r}^2(u)\right)du\\ & = & \rho^1+\phi^2R_2.\end{array}$$

In our case, we have $\rho^2<\phi^2R_2$, hence we have $(\bar{W}^{1,[1]}+\tilde{W}^2)(T_n^A-)=0$ infinitely often, and the same result holds for $(\bar{W}^{1,[2]}+\bar{W}^{2,[2]})(T_n^A-)$.

We now show uniqueness of this solution : consider any finite solution $(Z_1,Z_2)\circ\theta_t$, then we have

$$\left(\bar{W}_0^{1,[2]}+\bar{W}_0^{2,[2]}\right)(t)\le\left(\bar{W}_{Z_1}^{1,[2]}+\bar{W}_{Z_1}^{2,[2]}\right)(t)\le\left(\bar{W}_{Z_1}^{1,[1]}+\tilde{W}_{Z_2}^{2}\right)(t).$$

Let

$$\nu=\inf\left\{t\ge0,\ \left(\bar{W}_{Z_1}^{1,[1]}+\tilde{W}_{Z_2}^{2}\right)(t)=0\right\}.$$

With the same kind of arguments as above, we have $\left(\bar{W}_{Z_1}^{1,[1]}+\tilde{W}_{Z_2}^{2}\right)(t)>0$ for all $t\ge0$, implies that $\rho^1+\rho^2\ge\rho^1+\phi^2R_2$. Hence we have $\nu<\infty$ $\mathbb{P}$-a.s. Thus for any finite initial condition (Z_1,Z_2), there exists a finite time ν such that

$$\forall t\ge\nu,\ \bar{W}_0^{1,[2]}(t)=\bar{W}_{Z_1}^{1,[2]}\quad\text{and,}\quad\bar{W}_0^{2,[2]}(t)=\bar{W}_{Z_1}^{2,[2]}(t).$$

In particular taking $Z_1 = \bar{W}^{1,[2]}$ and $Z_2 = \bar{W}^{2,[2]}$, we have that $\left(\bar{W}_0^{1,[2]}(t), \bar{W}_0^{2,[2]}(t)\right)$ and $\left(\bar{W}^{1,[2]}(t), \bar{W}^{2,[2]}(t)\right)$ couple. This in turn implies that $\left(\bar{W}_{Z_1}^{1,[2]}(t), \bar{W}_{Z_1}^{2,[2]}(t)\right)$ and $\left(\bar{W}^{1,[2]}(t), \bar{W}^{2,[2]}(t)\right)$ couple. And we have for sufficiently large t,

$$\begin{aligned} Z_1 &= \bar{W}_{Z_1}^{1,[2]}(t) \circ \theta_{-t} = \bar{W}^{1,[2]}(t) \circ \theta_{-t} = \bar{W}^{1,[2]}, \\ Z_2 &= \bar{W}_{Z_2}^{2,[2]}(t) \circ \theta_{-t} = \bar{W}^{2,[2]}(t) \circ \theta_{-t} = \bar{W}^{2,[2]}. \end{aligned}$$

This finishes step 2.

For $k \leq N$, we assume that $\bar{W}^{1,[k-1]}, \ldots, \bar{W}^{k,[k-1]}$ are given. We construct the random variable

$$\tilde{r}^k = \frac{1}{\sum_{j=1}^N \phi^j \mathbb{1}_{\{j \notin I^{[k-1]}\}}}.$$

We have by construction

$$\sum_{j=1}^{k-1} \phi^j \tilde{r}^k \mathbb{1}_{\{\bar{W}^{j,[k-1]} > 0\}} + \tilde{r}^k \sum_{j=k}^{N} \phi^j = 1,$$

and since $\phi^j \tilde{r}^k \circ \theta_t = r^{j,[k-1]}(t)$ is exactly the service rate of queue $\bar{W}^{j,[k-1]}$,

$$\mathbb{E}\left[\phi^j \tilde{r}^k \mathbb{1}_{\{\bar{W}^{j,[k-1]} > 0\}}\right] = \rho^j.$$

This implies that

$$\mathbb{E}\left[\tilde{r}^k\right] = R_k. \tag{3.36}$$

We consider the following recursion

$$\tilde{W}^k(t) = \left(\tilde{W}^k(T_n^{A,k}-) + \sigma_n^k - \phi^k \int_{[T_n^{A,k}, t)} \tilde{r}^k \circ \theta_u du\right)^+, \quad t \in [T_n^{A,k}, T_{n+1}^{A,k}).$$

We have

$$\mathbb{E}^0_{T^{A,k}}\left[\int_{[0, T_1^{A,k})} \tilde{r}^k \circ \theta_u du\right] = \frac{R_k}{\lambda_k},$$

hence we have $k \in S$ implies that $\tilde{W}^k$ is stable in the sense of Theorem 4. We denote $\tilde{W}^k \circ \theta_t$ the unique workload process compatible with the shift. Now the proof is similar to step 2. We have that $\bar{W}_0^{i,[k]} \circ \theta_{-t}$ are increasing in t and that

$$\begin{aligned} \bar{W}_0^{i,[k]}(t) \circ \theta_{-t} &\leq \bar{W}_0^{i,[k-1]}(t) \circ \theta_{-t}, \quad \forall i \leq k-1, \\ \bar{W}_0^{k,[k]}(t) \circ \theta_{-t} &\leq \tilde{W}_0^{k}(t) \circ \theta_{-t}. \end{aligned}$$

Hence we can show existence and then uniqueness by a coupling argument.

It remains to show that queues that are not in S are unstable under the additional condition $\frac{\rho^{K+1}}{\phi^{K+1}} > R$.

First assume that $K = N-1$. In this case $\frac{\rho^N}{\phi^N} > R$ implies that $\rho > 1$. Hence thanks to Theorem 4, we know that there exists no finite $\mathbb{P}-$stationary workload process $\{\sum_{i=1}^N W^i(t)\}$, $t \in \mathbb{R}$. Now for any finite workload process $(\mathcal{W}^1(t), \dots, \mathcal{W}^N(t))$ of queues $1, \dots, N$ for $t \geq 0$, we have for all $i \leq N-1$,

$$\mathcal{W}^i(t) \leq \bar{W}^{i,[N-1]}_{\mathcal{W}^i(0)}(t), \quad \forall t \geq 0.$$

This shows that $\mathcal{W}^N(t) \to \infty$ as $t \to \infty$. Indeed we have $\mathcal{W}^N(t) \sim (\rho - 1)t = \left(\rho^N - \phi^N R\right) t$. In this case, the proposition follows.

We assume now that $\frac{\rho^{K+1}}{\phi^{K+1}} > R$ (with the possible value 0 for K, in which case $R = 1$). This ensures that $\rho > 1$. Thanks to the ordering of the indices, we have

$$\frac{\sum_{i=K+1}^N \rho^i}{\sum_{i=K+1}^N \phi^i} > R.$$

If we replace the classes $K+1, \dots N$ by an unique class with weight $\sum_{i=K+1}^N \phi^i$, the workload obtain for this virtual class is clearly a lower bound for the sum of the workloads of classes $K+1, \dots N$. The argument above applies to this virtual class which receives mean service rate $\sum_{i=K+1}^N \phi^i R$ and we have for any finite workload process $(\mathcal{W}^1(t), \dots, \mathcal{W}^N(t))$ defined on $\mathbb{R}_+$ (we denote $\mathcal{I}(t) = \{i,\ \mathcal{W}^i(t) = 0\}$),

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t \frac{\phi^{K+1} + \cdots + \phi^N}{\sum_{j \in \mathcal{I}(u)} \phi^j} du \leq \sum_{i=K+1}^N \phi^i R < \sum_{i=K+1}^N \rho^i.$$

Hence for $i \geq K+1$, we have

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t \frac{\phi^i}{\sum_{j \in \mathcal{I}(u)} \phi^j} du \leq \phi^i R < \rho^i.$$

Hence we have $\mathcal{W}^i(t) \to \infty$ as $t \to \infty$ for $i \geq K+1$. From which we derive that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{\phi^i}{\sum_{j \in \mathcal{I}(u)} \phi^j} du = \phi^i R < \rho^i,$$

and then

$$\mathcal{W}^i(t) \sim (\rho^i - \phi^i R)t \quad \text{as } t \to \infty.$$

□

Remark 12.

1. We imposed $\frac{\rho^{K+1}}{\phi^{K+1}} > R$ in order to avoid the critical case (corresponding to $\rho = 1$ in the single server queue).
2. The constants K and R already appeared in the work of Borst, Boxma and Jelenković [21]. But the approach of these authors is completely different. They assume the existence of the mean service rates for each flow (see their Appendix C) and then derive the equations they must solve. They use these equations to get the so-called GPS inequalities.

The study of the case $\rho > 1$ is indeed interesting in itself ! Consider a GPS system with weight $\phi^1, \dots, \phi^N$ but with greedy classes, i.e. the classes $K+1, \dots, N$ are continuously claiming their full share of the link rate. The other classes behave "normally", i.e. the input processes $(T^{A,\ell}, \sigma)$ satisfy the stationary ergodic conditions. The previous theorem gives us the following result

Proposition 11. *The subsystem consisting of queues $1, \dots, K$ is stable in the sense that there exists an unique stationary workload process if*

$$\max_{j=1}^{K} \frac{\rho^j}{\phi^j} < \frac{1 - \sum_{i=1}^K \rho^i}{1 - \sum_{i=1}^K \phi^i} := R. \tag{3.37}$$

Moreover there exists a mean service rate for the greedy queues, in the following sense : for any finite initial condition Y, let $I_Y(t) = \{i \leq K,\ W_Y^i(t) = 0\}$, then we have

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \frac{1}{\sum_{i=1}^K \phi^i \mathbf{1}_{i \notin I_Y(u)} + \sum_{i=K+1}^N \phi^i} = R.$$

Hence the mean service rate of greedy queue j is $\phi^j R$.

In the case

$$\max_j \frac{\rho^j}{\phi^j} > R,$$

there is at least one of the queues $1, \dots, K$ which is unstable.

Indeed, it is not very hard to see that the system described in Proposition 11 belongs to the monotone separable framework. Hence we computed the constant $\gamma(0)$ that corresponds to this system. We can rewrite the stability condition (3.37) as follows

$$\lambda\gamma(0) = \lambda \max_j \left\{ \mathbb{E}_{T^A}^0 \left[\sigma_0\right] + \left(1 - \sum_{i=1}^K \phi^i\right) \frac{\mathbb{E}_{T^A}^0 \left[\sigma_0 \mathbf{1}_{\{c_0 = j\}}\right]}{\phi^j} \right\} < 1,$$

which gives the explicit formulation of $\gamma(0)$.

Let $F(S_1, \dots, S_K)$ be the time to empty the system described in Proposition 11 if at time 0, the load of queue i is S_i. The mapping F is clearly continuous and satisfy the following scaling property :

$$\forall \kappa > 0, \quad F(\kappa S_1, \dots, \kappa S_K) = \kappa F(S_1, \dots, S_K).$$

Let $S_i(n) = \sum_{j=0}^n \sigma_j \mathbf{1}_{\{c_j = i\}}$, we have

$$\begin{array}{ccccc} \frac{Z_{[0,n]}(Q)}{n} & = & \frac{F(S_i(n))}{n} & = & F\left(\frac{S_i(n)}{n}\right) \\ \downarrow & & & & \downarrow \\ \gamma(0) & & = & & F\left(\mathbb{E}_{T^A}^0 \left[\sigma_0 \mathbf{1}_{\{c_0 = i\}}\right]\right) \end{array}$$

Hence we have by identification

$$F(S_1, \dots, S_K) = \max_{j=1,\dots,K} \left\{ S_1 + S_2 + \dots + S_K + \frac{1 - \sum_{i=1}^K \phi^i}{\phi^j} S_j \right\}. \tag{3.38}$$

This is certainly not the most direct way to compute the mapping F. Indeed a simple computation would have given the expression (3.38), from which we would have been able to compute $\gamma(0)$. Then the stability condition $\lambda\gamma(0) < 1$ would have ensured us that there exists a finite minimal stationary solution. Anyway with this approach we would not have been able to prove the second order ergodic results and the coupling-convergence results (in particular uniqueness).

3.2.2 Rare Events in GPS Queues

In this section we consider a stable GPS system with $\rho < 1$. W.l.o.g we assume that the ordering (3.31) holds. We are interested in the effect of a very big service time of size σ arriving in queue j at time $T_0^{A,j}$. Hence we consider workload process given by equations (3.24), (3.25), (3.26) and (3.27) for $t \geq T_0^{A,j}$, with initial condition $(W^1(T_0^{A,j}-), \dots, W^N(T_0^{A,j}-))$, i.e. in the stationary regime but we replace σ_0 by a deterministic value σ. We assume w.l.o.g that $T_0^{A,j} = 0$ and we denote $W^{\{j\}}(\sigma, t) = (W^{1,\{j\}}(\sigma, t), \dots, W^{N,\{j\}}(\sigma, t))$ the corresponding workload process.

Let $T(\sigma) > 0$ be the time for queue j to empty. On the interval $[0, T(\sigma)]$, the queue j is always backlogged. Hence we are exactly in the situation of Proposition 11 with queue j as greedy queue and if

$$\max_{i \neq j} \frac{\rho^i}{\phi^i} > \frac{1 - \sum_{i \neq j} \rho^i}{1 - \sum_{i \neq j} \phi^i},$$

then at least one queue $i \neq j$ begins to grow on this period of time. Hence the situation at time $T(\sigma)$ is that some queues are very big and will remain backlogged for a long period of time. Indeed we are still in the situation of Proposition 11 but this time with a set of greedy queues.

It is now quite natural to introduce the following notation corresponding to a GPS system in which queues $\{1, \dots, N\} \backslash D$ are greedy. Given a set $D = \{d_1, \dots, d_n\} \subset \{1, \dots, N\}$, with $d_1 \leq \dots \leq d_n$, we still have

$$\frac{\rho^{d_1}}{\phi^{d_1}} \leq \dots \leq \frac{\rho^{d_n}}{\phi^{d_n}}.$$

Hence results of previous section apply and we denote

$$\begin{aligned} K(D) &= \max_{i=1,\dots,n} \left\{ i, \ \frac{\rho^{d_i}}{\phi^{d_i}} < \frac{1 - \sum_{\ell=1}^{i-1} \rho^{d_\ell}}{\sum_{\ell=i}^{n} \phi^{d_\ell} + \sum_{j \notin D} \phi^j} \right\} \\ S(D) &= \{d_1, \dots, d_{K(D)}\}, \\ R(D) &= \frac{1}{\sum_{j \notin S(D)} \phi^j} \left(1 - \sum_{j \in S(D)} \rho^j \right), \end{aligned}$$

with the convention $\sum_{-1}^{0} = \sum_{\emptyset} = 0$.

In the case $D = \{j\}$, we will use the notation (j) instead of $(\{j\})$. From the proof of Theorem 5, it is clear that all queues $i < j$ remain stable when j is greedy. Indeed we have

$$\sum_{i=1}^{N} \rho^i < 1 \Rightarrow R(j)\phi^j > \rho^j. \tag{3.39}$$

In view of results of previous section, the interpretation is the following. Denote by $W^{d_i}(D)$ the stationary workload of queue d_i when queues that are not in D continuously claim their full share of the link rate. Then we have

$$d_i \in S(D) \Rightarrow W^{d_i}(D) < \infty. \tag{3.40}$$

Note that $W^{d_i}(D)$ is an upper bound for the stationary workload of queue d_i, W^{d_i} (which is well-defined as $\rho < 1$). Hence (3.40) provides an upper bound that is independent of what happens in queues that are not in D. In other words, the queue d_i is insensitive from the point of view of stability to the queues that are not in D. In particular if

$$\frac{\rho^i}{\phi^i} < R(j), \tag{3.41}$$

then queue i is insensitive to queue j. Note that it is always the case if $\frac{\rho^i}{\phi^i} < 1$.

We proceed now to the analyze of the effect of a very big service of type j when condition (3.41) is not satisfied. We will give a superscript $.^{\{j\}}$ to the constants that are calculated in this case. We denote

$$\begin{aligned}
N^{\{j\}} &= N - |S(j)| \\
f_1^{\{j\}} &= \frac{1}{\phi^j R(j) - \rho^j}, \\
\gamma^{k,\{j\}}(t) &= \phi^k R(j) \mathbf{1}_{\{t \le f_1^{\{j\}}\}}, \\
z_1^{k,\{j\}} &= \int_0^{f_1^{\{j\}}} \left(\rho^k - \gamma^{k,\{j\}}(u)\right)^+ du, \\
I_1 &= \{1, \dots, N\} \backslash (S(j) \cup \{j\}), \\
i_1^{\{j\}} &= j.
\end{aligned}$$

We have the following interpretation for these constants in term of fluid queues (which will be more detailed in next proposition). Queue j empties at time $f_1^{\{j\}}\sigma$ and at this time, the workload of queues $k \in I_1$ reaches level $z_1^{k,\{j\}}\sigma$, whereas other queues are empty (in the fluid approximation). Hence at time $f_1^{\{j\}}\sigma$, queues $k \in I_1$ are backlogged and will have a service rate $\phi^k R(I_1)$, whereas other queues including j are stable. Then define

$$\begin{aligned}
\left\{i_2^{\{j\}}\right\} &= \arg\min_{i \in I_1} \left\{ \frac{z_1^{i,\{j\}}}{\phi^i R(I_1) - \rho^i} \right\}, \\
f_2^{\{j\}} &= \inf_{i \in I_1} \left\{ \frac{z_1^{i,\{j\}}}{\phi^i R(I_1) - \rho^i} \right\} + f_1^{\{j\}}, \\
\gamma^{k,\{j\}}(t) &= \phi^k R(I_1) \mathbf{1}_{\{f_1^{\{j\}} < t \le f_2^{\{j\}}\}}, \\
z_2^{k,\{j\}} &= \int_0^{f_2^{\{j\}}} \left(\rho^k - \gamma^{k,\{j\}}(u)\right)^+ du \\
I_2 &= I_1 \backslash \left\{i_2^{\{j\}}\right\}.
\end{aligned}$$

The interpretation is the following : at time $f_2^{\{j\}}\sigma$, queues $\left\{i_2^{\{j\}}\right\}$ empty whereas queues in I_2

reach levels $z_2^{k,\{j\}}\sigma$. More generally we define

$$\begin{aligned}
f_{\ell+1}^{\{j\}} &= \inf_{i \in I_\ell} \left\{ \frac{z_\ell^{i,\{j\}}}{\phi^i R(I_\ell) - \rho^i} \right\} + f_\ell^{\{j\}}, \\
\left\{ i_{\ell+1}^{\{j\}} \right\} &= \arg\min_{i \in I_\ell} \left\{ \frac{z_\ell^{i,\{j\}}}{\phi^i R(I_\ell) - \rho^i} \right\}, \\
I_{\ell+1} &= I_\ell \backslash \left\{ i_\ell^{\{j\}} \right\}, \\
\gamma^{k,\{j\}}(t) &= \phi^k R(I_\ell) \mathbf{1}_{\{f_\ell^{\{j\}} < t \le f_{\ell+1}^{\{j\}}\}}, \\
z_{\ell+1}^{k,\{j\}} &= \int_0^{f_{\ell+1}^{\{j\}}} \left(\rho^k - \gamma^{k,\{j\}}(u) \right)^+ du.
\end{aligned}$$

For all $k \in \{1, \dots, N\}$, we defined a function $\gamma^{k,\{j\}}(t)$ for $t \le \sigma/(1-\rho)$ that we extend for $t > \sigma/(1-\rho)$ by $\gamma^{k,\{j\}}(t) = \rho^k$. We can now define the function

$$\begin{aligned}
w^{k,\{j\}}(\sigma, t) &= \int_0^t \left(\rho^k - \gamma^{k,\{j\}}(u/\sigma) \right)^+ du \quad \forall j \neq k, \\
w^{j,\{j\}}(\sigma, t) &= \left(\sigma + \int_0^t (\rho^j - \phi^j R(j)) du \right)^+.
\end{aligned}$$

Let $w^{\{j\}}(\sigma, t) = (w^{1,\{j\}}(\sigma, t), \dots, w^{N,\{j\}}(\sigma, t))$ be the multidimensional function.

Since the sequence of sets $\{I_\ell\}$ is decreasing, it is easy to see that $R(I_{\ell+1}) > R(I_\ell)$. Indeed Figure 3.1 shows what the function $w^{\{j\}}(\sigma, .)$ looks like for fixed σ.

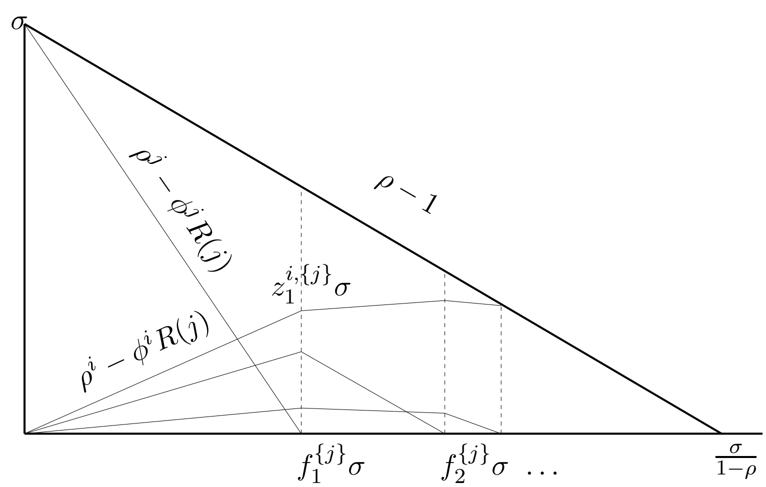


FIG. 3.1 – function $w^{\{j\}}(\sigma, .)$ for fixed σ

Proposition 12. *Under previous condition, we have for any constant $\alpha, \beta > 0$, as $n \to \infty$,*

$$\sup_{\sigma > n\alpha, t \le \beta} \left| \frac{W^{\{j\}}(\sigma, nt) - w^{\{j\}}(\sigma, nt)}{n} \right| \to 0, \quad \textit{a.s.}$$

Proof.
We consider a sequence σ^n such that $\sigma^n > n\alpha$. We suppose that

$$\frac{\sigma^n}{n} \quad \to \quad \sigma \leq +\infty.$$

We will show that

$$\sup_{0\leq t\leq\beta}\left|\frac{W^{\{j\}}(\sigma^n, nt) - w^{\{j\}}(\sigma^n, nt)}{n}\right| \to 0,$$

which is sufficient to prove the proposition. For simplicity, we denote $W_n^{\{j\}}(t) = W^{\{j\}}(\sigma^n, t)$ and $w_n^{\{j\}}(t) = w^{\{j\}}(\sigma^n, t)$.

We first assume that $\sigma < \infty$. Let T_1^n be the first positive time at which queue j becomes empty, ie queue j is backlogged on $[0, T_1^n]$. Hence we have thanks to the result on the mean service rate of Proposition 11,

$$\lim_{n\to\infty}\frac{W_n^{j,\{j\}}(T_1^n)}{n} \quad = \quad \sigma + (\rho^j - \phi^j R(j))\left(\lim_{n\to\infty}\frac{T_1^n}{n}\right) = 0,$$

from which we derive $\lim_{n\to\infty}\frac{T_1^n}{n} = \sigma/(\phi^j R(j) - \rho^j)$. Now for $0 \leq t \leq \sigma f_1^{\{j\}}$, we can apply Proposition 11 and we have

$$\begin{aligned}\frac{W_n^{\ell,\{j\}}(nt)}{n} \quad &\to \quad (\rho^\ell - \phi^\ell R(j))^+ t, \quad \forall \ell \neq j,\\ \frac{W_n^{j,\{j\}}(nt)}{n} \quad &\to \quad \sigma + (\rho^j - \phi^j R(j))t.\end{aligned}$$

We have shown in the case $\sigma < \infty$ that for all j,

$$\sup_{0\leq t\leq f_1^{\{j\}}\sigma}\left|\frac{W_n^{\{j\}}(nt) - w_n^{\{j\}}(nt)}{n}\right| \to 0.$$

Moreover, we see that at time T_1^n, the queues $k \in I_1$ are backlogged. Define T_n^2 the first time at which one of these queues become empty. Using Proposition 11 in the same manner, we obtain that $T_2^n/n \to \sigma f_2^{\{j\}}$ and that,

$$\sup_{0\leq t\leq f_2^{\{j\}}\sigma}\left|\frac{W_n^{\{j\}}(t) - w_n^{\{j\}}(nt)}{n}\right| \to 0.$$

Hence in the case $\sigma < \infty$, the proposition follows by iterating the same kind of arguments.

In the case $\sigma = +\infty$, since $T_1^n \geq \sigma^n$, we have for sufficiently large n, we have $T_1^n \geq n\beta$. Hence for all $k \neq j$, we have with the same argument as above that

$$\sup_{0\leq t\leq\beta}\left|\frac{W_n^{k,\{j\}}(nt) - w_n^{k,\{j\}}(nt)}{n}\right| \to 0,$$

and for $k = j$, we have for all $t \leq \beta$,

$$\frac{W_n^{j,\{j\}}(nt) - \sigma^n}{n} \quad \to \quad (\rho^j - \phi^j R(j))t.$$

This concludes the proof. $\square$

Chapitre 4

Subexponential Asymptotics

4.1 Introduction

Subexponential distributions are a special case of heavy-tailed distributions. The name arises from one of their properties, that their tails decrease more slowly than any exponential tail. This implies that large values can occur in a sample with non-negligible probability and makes the subexponential distributions candidates for modeling situations where some extremely large values occur in a sample compared to the mean size of the data.

4.1.1 Some Definitions and Notations

Notation
Here and later in the paper, for positive functions f and g, the equivalence $f(x) \sim dg(x)$ with $d > 0$ means $f(x)/g(x) \to d$ as $x \to \infty$. By convention, the equivalence $f(x) \sim dg(x)$ with $d = 0$ means $f(x)/g(x) \to 0$ as $x \to \infty$, this will be written $f(x) = o(g(x))$. We will also use the notation $f(x) = \mathcal{O}(g(x))$ to mean $\limsup f(x)/g(x) < \infty$ and $\liminf f(x)/g(x) > 0$.
In what follows, $\epsilon(x)$ denotes a function such that $\epsilon(x) \xrightarrow{x\to\infty} 0$. The function ϵ may vary from place to place ; for example, $\epsilon(x) + \epsilon(x) = \epsilon(x)$, $\epsilon(x)(1 + \epsilon(x)) = \epsilon(x)$, etc. Similarly, we will write $\epsilon(x, y)$ for $\epsilon(x) + \epsilon(y)$, or $\epsilon(x)\epsilon(y)$, etc.
The tail of the distribution function F is denoted $\overline{F}(x) = 1-F(x)$. We recall here some definitions

Definition 1. *A distribution function F on $\mathbb{R}_+$ is long tailed if for any $y > 0$,*

$$\overline{F}(x + y) \sim \overline{F}(x) \quad \textit{as} \quad x \to \infty.$$

We introduce a proper subset of the class of long tailed distributions, the class of subexponential distributions denoted by $\mathcal{S}$:

Definition 2. *A distribution function F on $\mathbb{R}_+$ is called subexponential if $\overline{F^{*2}}(x) \sim 2\overline{F}(x)$.*

For basic properties of subexponential distribution see [51] and [40].

Definition 3. *A positive measurable function f on $[0, +\infty)$ is called regularly varying with index $\alpha \in \mathbb{R}$ ($f \in \mathcal{R}(\alpha)$) if*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\alpha} \quad \textit{for all } t > 0.$$

Definition 4. *A positive measurable function h on $[0,+\infty)$ is called rapidly varying ($h \in \mathcal{R}(-\infty)$) if*

$$\lim_{x\to\infty} \frac{h(tx)}{h(x)} = 0 \quad \textit{for all } t > 1.$$

For example, Weibull or lognormal random variables have tail distributions that are rapidly varying.

For a distribution function F on the positive real line with finite first moment $M = \int_0^\infty \overline{F}(u)du$, the integrated tail distribution F^s of F is defined by

$$\overline{F}^s(x) := 1 - F^s(x) = \min\{1, \int_x^\infty \overline{F}(u)du\}.$$

We will need the following lemma later on

Lemma 15. *If F^s is long tailed, then there exists a non-decreasing integer valued function $N_x \to \infty$ such that for all finite non-negative real number b, we have*

$$\sum_{n=0}^{N_x} \overline{F}(x+nb) = o\left(\overline{F}^s(x)\right), \quad \textit{as } x \to \infty.$$

Proof.

For any integer n, we have (for sufficiently large x)

$$\begin{aligned}\overline{F}^s(x) = \int_x^\infty \overline{F}(u)du &\geq b\overline{F}(x+b) + b\overline{F}(x+2b) + \cdots + b\overline{F}(x+nb) + \int_{x+nb}^\infty \overline{F}(u)du \\ &= b\sum_{k=1}^n \overline{F}(x+kb) + \overline{F}^s(x+nb).\end{aligned}$$

Since F^s is long tailed, we have for fixed n, as $x \to \infty$,

$$\frac{\sum_{k=1}^n \overline{F}(x+kb)}{\overline{F}^s(x)} \to 0,$$

from which the lemma follows. □

We present now what we call Veraverbeke's theorem. Let S_n be a random walk with negative drift, namely $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables such that $\mathbb{E}[X_0] = -\mu < 0$. Let define

$$S_0 = 0, \quad S_n = \sum_{i=1}^n X_i, \quad \text{and,} \quad M = \sup_{n\geq 0} S_n.$$

The condition $\mathbb{E}[X_0] = -\mu < 0$ ensures that M is a.s. finite. Assume that there exists a distribution function F on $[0,\infty)$ such that

$$\mathbb{P}(X_1 > x) \sim d\overline{F}(x) \quad \text{with } d > 0 \text{ as } x \to \infty.$$

Theorem 6. *If F^s is subexponential then the random variable M is subexponential and we have*

$$\mathbb{P}(M > x) \sim \frac{d}{\mu}\overline{F}^s(x) \quad \text{as } x \to \infty. \tag{4.1}$$

This result was first proved in different contexts by Borovkov [20], Cohen [27], Veraverbeke in [84], Pakes in [77] and Embrechts and Veraverbeke in [41]. There are now probabilistic proofs of this result see for example the work of Asmussen [6] or Zachary [85]. Finally, we remark that a proof of the converse of Veraverbeke's Theorem - that (for $\mathbb{E}[X_0]$ finite and negative) the relation (4.1) implies the subexponentiality of F^s- is given by Korshunov in [66].

The fact that equivalence (4.1) is a product of the integrated tail and of the inverse of the drift, has a very nice queueing interpretation. Consider two independent sequences $\{X, X_i\}_{i\in\mathbb{N}}$ and $\{Y, Y_i\}_{i\in\mathbb{N}}$ of i.i.d. random variables. Define

$$M_{X+Y} = \sup_{n\geq 0}\sum_{i-1}^{n}(X_i + Y_i) \quad \text{and} \quad M_{X+\mathbb{E}[Y]} = \sup_{n\geq 0}\sum_{i-1}^{n}(X_i + \mathbb{E}[Y_i]).$$

Assume that $\mathbb{P}(X + Y > x) \sim \mathbb{P}(X > x)$ and that the tail distribution of X satisfies the condition of Veraverbeke's Theorem. Then we have by rewriting (4.1),

$$\mathbb{P}(M_{X+Y} > x) \sim \mathbb{P}(M_{X+\mathbb{E}[Y]} > x) \quad \text{as } x \to \infty.$$

This is an example of reduced load equivalence : the tail asymptotics of the workload is dominated by the heaviest input and is asymptotically the same as the one of a system fed by this heaviest input and in which we replace the rest by its mean. This kind of equivalence has first been understood by Agrawal, Makowski and Nain in [2] for a single server with fluid inputs and generalized to more general input by Jelenković, Momčilović and Zwart in [63]. For other results concerning various models of single server queue, we refer to the works of Zwart [87] and Likhanov and Mazumdar [69], [70]. In a network setting, we will show that the same kind of results hold, the heaviest tail distribution dominates the asymptotics. If different stations in the network have service times with the same kind of tail distribution, then each of them will contribute to the asymptotics. New arguments have to be found to derive the asymptotics.

4.1.2 The Single Big Event Theorem

In this section we summarize results from the work of Baccelli and Foss [14].

A corollary of Veraverbeke's theorem already proved by Anantharam [3] (in the regularly case) and by Asmussen and Klüppelberg [7] states that, in the $GI/GI/1$ queue, large workload occur on a typical event where a single large service time has taken place in the distant past, and all other service time are close to their mean. The main result of this section is to extend the notion of typical event to subexponential monotone separable networks : large maximal daters occur when a single large service time has taken place in one of the stations and all other service time are close to their mean.

We recall assumptions of [14], the notations were introduced in Section 2.1.

(IA) : the sequences $\{\zeta_n\}$ and $\{\tau_n\}$ are mutually independent and each of them consists of i.i.d. random variables.

(AA) : For all i,

$$Z_i = Z_{[i,i]} = Y_i^{(1)} + \cdots + Y_i^{(r)},$$

where the random variables $Y_i^{(j)}$ are non negative, independent of inter-arrival times and such that the sequence of random vectors $(Y_i^{(1)}, \dots Y_i^{(r)})$ is i.i.d. ; general dependences between the components of the vector $(Y_i^{(1)}, \dots Y_i^{(r)})$ are allowed. In addition

$$Z_{[0,n]}(Q) \geq \max_{j=1,\dots,r} \sum_{i=n}^{0} Y_i^{(j)} \quad \text{a.s.}$$

Consider a distribution F on $\mathbb{R}_+$ such that the following holds :

1. F is subexponential, with finite first moment $M = \int_0^\infty \overline{F}(u)du$.
2. The integrated tail distribution F^s is subexponential.
3. For all $j = 1, \dots, r$

$$\mathbb{P}\left(Y_1^{(j)} > x\right) \sim d^{(j)}\overline{F}(x),$$

with $\sum_j d^{(j)} > 0$.

We make the following assumption on the $Y_i^{(j)}$'s,

$$(\mathbf{H}): \quad \mathbb{P}\left(\sum_1^r Y_1^{(j)} > x\right) \sim \mathbb{P}\left(\max_1^r Y_1^{(j)} > x\right) \sim \sum_1^r \mathbb{P}\left(Y_1^{(j)} > x\right).$$

Theorem 7 and 8 of [14] state

Theorem 7. *Let Z be the stationary maximal dater of some monotone separable network. For any x and for $j = 1, \dots, r$, let $\{K_{n,x}^j\}$ be a sequence of events such that*

1. *for any n and j, the event $K_{n,x}^j$ and the random variable $Y_{-n}^{(j)} = \sum_{k=1}^{\phi^{(j)}(-n)} \sigma_k^{(j)}(-n)$ are independent ;*
2. *for any j, $\mathbb{P}(K_{n,x}^j) \to 1$ uniformly in $n \geq N_x$ as $x \to \infty$.*

For all sequences $\epsilon_n \to 0$, we denote $x_n = x + n(a - b + \epsilon_n)$. Then, as $x \to \infty$,

$$\mathbb{P}[Z > x] \sim \sum_{j=1}^{K} \sum_{n \geq N_x} \mathbb{P}[Z > x, Y_{-n}^{(j)} > x_n, K_{n,x}^j], \tag{4.2}$$

and

$$\mathbb{P}[Z > x] = \mathcal{O}(\overline{F}^s(x)).$$

The equivalence (4.2) will be the key relationship for the exact asymptotics in the next sections. It shows that for the monotone separable network also, whenever the maximal dater is large, at most one of the service times is large whereas all other ones are moderate.

4.2 Asymptotics of Subexponential Max Plus Networks : the Stochastic Event Graph Case

This section is focused on the derivation of the tail asymptotics of the steady state end-to-end response times in open, single input, stochastic event graphs [11], a class of networks which are known to admit a (max,plus)-linear representation.

To the best of our knowledge, within this class of networks, under subexponential statistical assumptions, exact asymptotics are only known for the following special cases :

- the case of dimension 1 ; this type of asymptotics is known as Pake's [77] or Veraverbeke's theorem [84], and most often expressed as a property of the waiting or response times in the G/G/1 queue (this can also be seen as a property of extrema of random walks) ;
- the case of irreducible event graphs [17], a first class of networks with general dimension that contains the G/G/1 queue as a special case ;
- the case of tandem queues [14], a class of reducible event graphs with a specific linear topology, which also contains the G/G/1 queue as a special case.

This section is based on the paper [16].

4.2.1 Stochastic Assumptions

Model Description and Stochastic Assumptions

For now on, we consider an event graph as described in Section 2.2.2, with $m \leq K$ timed transitions, namely $\mathcal{T}_{timed} = \{t(1), \ldots, t(m)\}$, satisfying the assumptions in Property 3, and with associated recursion :

$$X_n = A_n \otimes X_{n-1} \oplus B_n \otimes T_n$$

of dimension $s \leq KL$. This means that the matrices $\{A_n, B_n\}$ and vectors that are used in the recursion are obtained via two applications f and g such that :

$$\begin{array}{rccc} \mathcal{A} : & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,s)}\left(\mathbb{R}_{\max}\right) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{A}(\sigma), \end{array}$$

$$\begin{array}{rccc} \mathcal{B} : & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,1)}\left(\mathbb{R}_{\max}\right) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{B}(\sigma), \end{array}$$

via the formula

$$\begin{array}{rcl} \mathcal{A}(\zeta_n) & = & A_n, \\ \mathcal{B}(\zeta_n) & = & B_n. \end{array}$$

with $\zeta_n = (\sigma_n^{t(1)}, \ldots, \sigma_n^{t(m)})$.

We now assume that the following independence assumption holds :

Assumption 1. **(IA)** The sequences $\{\zeta_n\}$ and $\{\tau_n \equiv T_{n+1} - T_n\}$ are mutually independent and each of them consists of i.i.d. random variables.

The following assumptions are also assumed to hold

$$\mathbb{E}(\tau_0) \equiv \lambda^{-1} \equiv a < \infty, \quad \mathbb{E}(\sigma_0^{t(i)}) \equiv b^{t(i)} < \infty \quad \forall i = 1, \ldots, m.$$

This implies in particular $\mathbb{E}(Z_{[0,0]}) < \infty$.
Under these assumptions, considering the matrices $A_n(k,k)$, we have for all i and j :

$$\frac{(A_{-1}(k,k) \otimes A_{-2}(k,k) \otimes \cdots \otimes A_{-n}(k,k))^{(i,j)}}{n} \to \gamma_k \quad \text{both a.s. and in } L_1$$

where γ_k is a constant referred to as the top Lyapunov exponent of the sequence $\{A_n(k,k)\}$, see theorem 7.27 (p. 325) in [11].
In addition, we assume stability of the system, namely $\max_k \gamma_k = \gamma < a$ (see [13]).

We will also adopt the following notations :

- if $j \in \mathcal{C}_i$, we denote $\gamma_{(j)} = \gamma_i$;
- for all transitions i, the subset of transitions j such that there is a directed path in $\mathcal{G}$ from i to j is denoted $[\geq i]$;
- finally, we define

$$\Gamma_{(\geq i)} \quad = \quad \max_{k \in [\geq i]} \gamma_{(k)}.$$

The subexponential assumptions are now the following :

Assumption 2. **(SE)** The service times $\sigma^{t(k)}$ are independent r.v., with respective mean $b^{t(k)}$. There exists a distribution function F on $\mathbb{R}_+$ such that :

- **(SE.1)** F is subexponential, with finite first moment M.
- **(SE.2)** The integrated tail distribution F^s is subexponential.
- **(SE.3)** The following equivalence holds when x tends to ∞ :

$$\mathbb{P}(\sigma_1^{t(i)} > x) \sim c^{t(i)} \overline{F}(x),$$

  for all $i = 1, \ldots, m$ with $\sum_{i=1}^m c^{t(i)} = c > 0$.

For $i \notin \mathcal{T}_{timed}$, we will denote $\forall k,\ \sigma_k^i = 0$ and $c^i = 0$. Under **(SE.1)** and **(SE.3)**, we have (see [17] or [51]) :

Lemma 16.

$$\mathbb{P}(\sum_{k=1}^{K} \sigma_1^k > x) \sim \mathbb{P}(\max_{1 \leq k \leq K} \sigma_1^k > x) \sim \sum_{k=1}^{K} c^k \overline{F}(x).$$

Preliminary Results

Lemma 17. *For any event graph as described in Section 2.2.2, there exists some sets $\mathcal{K}_j$ such that $\bigcup_j \mathcal{K}_j = [1;s]$ and*

$$B_n^{(s)} \quad = \quad \bigoplus_j \bigotimes_{k \in \mathcal{K}_j} \sigma_n^k = \max_j \sum_{k \in \mathcal{K}_j} \sigma_n^k.$$

Moreover, $\forall j$ there exists only one integer $k(j)$ such that :

$$\begin{aligned} (A_n)^{(k(j),k(j))} &\geq \sigma_n^j, \\ (A_n)^{(s,k(j))} &\geq \sigma_n^j, \\ (B_n)^{(k(j))} &\geq \sigma_n^j. \end{aligned}$$

The following two properties hold (referred to as (AA') in what follows) :

$$Z_i = Z_{[i,i]} = \bigoplus_j \bigotimes_{k \in \mathcal{K}_j} \sigma_i^k = \max_j \sum_{k \in \mathcal{K}_j} \sigma_i^k, \qquad \text{(AA'-1)}$$

and, when denoting by Q the point process with all its points in 0

$$Z_{[n,0]}(Q) \quad \geq \quad \max_k \sum_{i=n}^{0} \sigma_i^k. \qquad \text{(AA'-2)}$$

Proof.

The first part is proved in Section 4.2.3. Thanks to Lemma **??**, we have

$$Z_i = \max_j B_i^{(j)} = B_i^{(s)},$$

and for the second part :

$$\begin{aligned} Z_{[n,0]}(Q) = \max_{n \leq k \leq 0} \left[(A_0 \otimes \cdots \otimes A_{k+1} \otimes B_k)^{(s)} \right] &\geq (A_0 \otimes \cdots \otimes A_{n+1} \otimes B_n)^{(s)} \\ &\geq (A_0)^{(s,k(j))} + \cdots + (A_{n+1})^{(k(j),k(j))} + (B_n)^{(k(j))} \\ &\geq \sigma_0^j + \cdots + \sigma_{n+1}^j + \sigma_n^j, \end{aligned}$$

for all j. □

Lemma 18. *For all positive integers L, let*

$$\hat{s}_n = Z_{[L(n-1)+1,Ln]}(Q).$$

We have

$$\max_k \sum_{i=L(n-1)+1}^{Ln} \sigma_i^k \leq \hat{s}_n \leq \sum_{k=1}^{m} \sum_{i=L(n-1)+1}^{Ln} \sigma_i^k. \qquad (4.3)$$

Proof.

The first inequality follows from (AA'-2). The second one follows from $Z_i = \max_j \sum_{k \in \mathcal{K}_j} \sigma_i^k \leq \sum_{k \in [1;s]} \sigma_i^k$, and the sub-additivity of Z. □

We will assume that assumptions **(IA)** and **(SE)** hold throughout this paper without restating it. Moreover N_x will denote a non-deceasing integer-valued function tending to infinity such that for all finite real numbers b,

$$\sum_{n=0}^{N_x} \overline{F}(x + nb) = o(\overline{F}^s(x)).$$

The existence of this function follows from the fact that F^s is long-tailed (see [14]).

Proposition 13. *Let Z be the stationary maximal dater of the event graph : $Z \equiv \lim_{n\to\infty} Z_{[-n,0]}$. For any x and for $j = 1, \ldots, r$, let $\{K^j_{n,x}\}$ be a sequence of events such that*

1. *for any n and j, the event $K^j_{n,x}$ and the random variable σ^j_{-n} are independent ;*
2. *for any j, $\mathbb{P}(K^j_{n,x}) \to 1$ uniformly in $n \geq N_x$ as $x \to \infty$.*

For all sequences η^j_n, $j = 1, \ldots, s$, tending to 0, put

$$A^j_{n,x} = K^j_{n,x} \cap \{\sigma^j_{-n} > x + n(a - \gamma + \eta^j_n)\}, \quad A^j_x = \bigcup_{n=N_x}^{\infty} A^j_{n,x} \quad \text{and} \quad A_x = \bigcup_{j=1}^{s} A^j_x.$$

Then, as $x \to \infty$,

$$\mathbb{P}[Z > x] \sim \mathbb{P}[Z > x, A_x] \sim \sum_{j=1}^{s} \sum_{n \geq N_x} \mathbb{P}[Z > x, A^j_{n,x}].$$

Proof.

The proof is omitted but uses the same arguments as the proof of Theorem 8 in [14]. The only difference lies in the fact that Condition (AA) in [14] has to be replaced by (AA'), defined in Lemma 17. But under (AA'), (7) of [14] still holds as shown in Lemma 18, which is enough to prove the desired result. □

4.2.2 Exact Tail Asymptotic

Theorem 8.

$$\mathbb{P}(Z > x) \quad \sim \quad \left(\sum_{i=1}^{s} \frac{c^i}{a - \Gamma_{(\geq i)}} \right) \overline{F}^s(x), \tag{4.4}$$

with :

$$\Gamma_{(\geq i)} \quad = \quad \max_{k \in [\geq i]} \gamma_{(k)}.$$

Proof.

For the sake of simplicity, we give a proof in the case of constant inter-arrival times only. In fact, it was shown in [14] Section 7.3., that the result extends to the stochastic framework we introduced.

Lower bound :

Thanks to Proposition 13, we have

$$\mathbb{P}(Z > x) \sim \sum_{n=N_x}^{\infty} \sum_{j=1}^{s} \mathbb{P}(Z > x, A^j_{x,n}).$$

We have to find appropriate sequences $\{K^j_{n,x}\}$ and $\{\eta^j_n\}$.
For all j, we have $(B_{-n})^{(k(j))} \geq \sigma^j_{-n}$. Hence we have

$$Z \geq \sigma^j_{-n} + (A_{-1} \otimes A_{-2} \otimes \ldots \otimes A_{-n+1})^{(s,k(j))} - na. \tag{4.5}$$

Consider the events

$$K^j_{n,x} = \left\{ (A_{-1} \otimes A_{-2} \otimes \ldots \otimes A_{-n+1})^{(s,k(j))} \geq n(\Gamma_{(\geq j)} - \eta^j_n) \right\},$$

and choose a sequence $\eta^j_n \to 0$ such that $\mathbb{P}[K^j_{n,x}] \to 1$ uniformly in $n \geq N_x$ as $x \to \infty$. Then from (4.5), we have

$$\begin{aligned} \mathbb{P}(Z > x, A^j_{x,n}) &\geq \mathbb{P}(\sigma^j_{-n} > x + n(a - \gamma + \eta^j_n),\ \sigma^j_{-n} > x + na - n(\Gamma_{(\geq j)} - \eta^j_n)) \\ &\geq (1 + o(1))\mathbb{P}(\sigma^j_{-n} > x + n\left[a + \eta^j_n - \min(\gamma, \Gamma_{(\geq j)})\right]). \end{aligned}$$

But we have for all j, $\Gamma_{(\geq j)} \leq \Gamma_{(\geq 1)}$ and $\gamma = \Gamma_{(\geq 1)}$.
Hence we get an equivalent in $\frac{c^j}{a - \Gamma_{(\geq j)}} \overline{F}^s(x)$.
Upper bound :
We have

$$\mathbb{P}(Z > x, A_x) = \sum_{j=1}^{s} \sum_{n \geq N_x} \mathbb{P}(Z > x, \sigma^j_{-n} > x + n(a - \gamma + \eta^j_n), K^j_{x,n}).$$

As

$$\mathbb{P}(Z > x, \sigma^j_{-n} > x + n(a - \gamma + \eta_n), K^j_{x,n}) \leq \mathbb{P}(\sigma^j_{-n} > x + n(a - \gamma + \eta_n)),$$

we have an upper bound in $(1 + o(1))\frac{c^j}{a - \Gamma_{(\geq 1)}} \int_x^\infty \bar{F}(y)dy$.
We consider now the case $\Gamma_{(\geq j)} < \Gamma_{(\geq 1)}$.
We then have the following decomposition :

$$\begin{aligned} Z &= \max\left\{ Z_{[-n+1;0]}; \max_{k \geq 0}[(A_{-1} \otimes \cdots \otimes A_{-n-k+1} \otimes B_{-n-k})^{(s)} - (n+k)a] \right\} \\ &\equiv \max\left\{U_n; V_n\right\}, \\ V_n &= \max\left\{ (A_{-1} \otimes \cdots \otimes A_{-n+1} \otimes B_{-n})^{(s)} - na; \right. \\ &\qquad \left. \max_{k \geq 1}[(A_{-1} \otimes \cdots \otimes A_{-n-k+1} \otimes B_{-n-k})^{(s)} - (n+k)a] \right\} \\ &\equiv \max\left\{Z^1_n; Z^2_n\right\}. \end{aligned}$$

Thanks to Lemma 19, we have $Z^2_n \leq Z^1_n + R_n$, where $R_n = Z_{[-\infty,-n-1]}$ is a random variable independent of σ^j_{-n}. Hence we have

$$\begin{aligned} V_n > x &\Rightarrow Z^1_n > x \quad \text{or} \quad Z^2_n > x \\ &\Rightarrow Z^1_n > x \quad \text{or} \quad Z^1_n + R_n > x \\ &\Rightarrow Z^1_n + R_n > x. \end{aligned}$$

Hence

$$\mathbb{P}(Z > x, A^j_{n,x}) \leq \mathbb{P}(\max\{Z^1_n + R_n, U_n\} > x, A^j_{n,x}).$$

We will denote $PA_n = A_{-1} \otimes \cdots \otimes A_{-n+1}$. We have then

$$\begin{aligned} Z_n^1 &= \max[Z_n^{(\geq j)}, Z_n^{(\geq j)^c}] \quad \text{with} \\ Z_n^{(\geq j)} &= \max_{i\in[\geq j]} [PA_n^{(s,i)} + (B_{-n})^{(i)}] - na, \\ Z_n^{(\geq j)^c} &= \max_{i\in[\geq j]^c} [PA_n^{(s,i)} + (B_{-n})^{(i)}] - na. \end{aligned}$$

Since $U_n \leq Z$ a.s., $\mathbb{P}(U_n \leq x) \to 1$ uniformly in n as $x \to \infty$. Since the distribution of $R_n = Z_{[-\infty,-n-1]}$ does not depend on n, $R_n/n \to 0$ in probability. Due to the SLLN, $\max_{i\in[\geq j]^c}[PA_n^{(s,i)} + (B_{-n})^{(i)}]/n \to c_j \leq \gamma$ and $\max_{i\in[\geq j]}[PA_n^{(s,i)}]/n \to \Gamma_{(\geq j)}$. For $i \in [\geq j]$, we have $(B_{-n})^{(i)} \leq \sigma_{-n}^j + \sum_{k\neq j} \sigma_{-n}^k$. We denote $\zeta_n^j = \sum_{k\neq j} \sigma_{-n}^k$, we have $\zeta_n^j/n \to 0$ in probability. Therefore, there exists a sequence $\epsilon_n \downarrow 0$, $n\epsilon_n \to \infty$ such that

$$\mathbb{P}\Big\{U_n \leq x, R_n \leq n\epsilon_n, \max_{i\in[\geq j]^c}[PA_n^{(s,i)} + (B_{-n})^{(i)}] \leq n(\gamma + \epsilon_n), \max_{i\in[\geq j]}[PA_n^{(s,i)}] \leq n(\Gamma_{(\geq j)} + \epsilon_n), \zeta_n^j \leq n\epsilon_n\Big\} \to 1$$

uniformly in $n \geq N_x$ as $x \to \infty$. Denote the latter event $K_{n,x}^j$. For $i \in [\geq j]^c$, the random variables $(B_{-n})^{(i)}$ and σ_{-n}^j are independent, hence $K_{n,x}^j$ is independent of σ_{-n}^j. Moreover, observe that on $K_{n,x}^j$, we have

$$\begin{aligned} \{\max\{Z_n^1 + R_n, U_n\} > x\} &= \{Z_n^1 + R_n > x\} \\ &\subset \{n(\gamma + \epsilon_n) - na + R_n > x\} \cup \{n(\Gamma_{(\geq j)} + \epsilon_n) + n\epsilon_n + \sigma_{-n}^j + n\epsilon_n - na > x\}. \end{aligned}$$

Put $\eta_n^j = -3\epsilon_n$. Then

$$\begin{aligned} \mathbb{P}(Z > x, A_{n,x}^j) &\leq \mathbb{P}(R_n > x + n(a - \gamma - \epsilon_n, K_{n,x}^j)\mathbb{P}(\sigma_{-n}^j > x + n(a - \gamma + \eta_n^j)) \\ &\quad + \mathbb{P}(\sigma_{-n}^j > x + n(a - \Gamma_{(\geq j)} - 3\epsilon_n), K_{n,x}^j) \\ &= o(1)\mathbb{P}(\sigma_{-n}^j > x + n(a - \gamma + \eta_n^j)) + (1 + o(1))\mathbb{P}(\sigma_{-n}^j > x + n(a - \Gamma_{(\geq j)} + \eta_n^j)), \end{aligned}$$

and the desired asymptotics follows. □

4.2.3 Two Technical Lemmas

Proof of Lemma 17.

The first property is a mere rewriting of the definition of $\overline{b}_n = a_0(n)^* \otimes b$. Remark **??**, which gives the relation between the matrices $\overline{a}_1$ and a_0^*, allows one to establish the last properties. Indeed, the maximum in $(v_1)^{(i)} = \max_j (a_1)^{(i,j)}$ is on the diagonal. Moreover, we have $(\overline{a}_1)^{(k,k)} = \max_i[(a_0^*)^{(k,i)} + (a_1)^{(i,k)}] \geq (a_1)^{(k,k)}$, which ensures the existence of $k(j)$ such that $(B_n)^{(k(j))} \geq \sigma_n^j$ and $(A_n)^{(k(j),k(j))} \geq \sigma_n^j$ because the diagonal terms of $\overline{a}_1$ are diagonal terms of A too.
Moreover, we have

$$\begin{aligned} (\overline{a}_1)^{(s,k)} &= \max_i[(a_0^*)^{(s,i)} + (a_1)^{(i,k)}] \\ &\geq (a_0^*)^{(s,k)} + (a_1)^{(k,k)} \\ &\geq (a_1)^{(k,k)} \end{aligned}$$

and then, we have $(A_n)^{(s,k(j))} \geq \sigma_n^j$.
□

We will denote for $a > 0$ and $u \leq v$:

$$Z_{[u,v]} = \max_{u \leq i \leq v} \left[(D_{[i+1,v]} \otimes B_i)^{(s)} - (v-i)a \right].$$

Lemma 19. *We denote for $n \geq 1$:*

$$\begin{aligned} Z_n^1 &= (A_0 \otimes \cdots \otimes A_{-n+1} \otimes B_{-n})^{(s)} - na, \\ Z_n^2 &= \max_{k \geq 0} [(A_0 \otimes \cdots \otimes A_{-n-k+1} \otimes B_{-n-k})^{(s)} - (n+k)a]. \end{aligned}$$

We have then

$$Z_n^2 \leq Z_n^1 + Z_{[-\infty,-n-1]}.$$

Proof. We have only to prove that

$$\tilde{Z}_n^2 \leq Z_n^1 + Z_{[-\infty,-n-1]},$$

with $\tilde{Z}_n^2 = \max_{k \geq 1}[(A_{-1} \otimes \cdots \otimes A_{-n-k+1} \otimes B_{-n-k})^{(s)} - (n+k)a]$. We will assume that $k \geq 1$ in what follows and we denote :

$$\begin{aligned} D_1 &= A_0 \otimes \cdots \otimes A_{-n+1}, \\ DB_k &= D_{[-n-k+1,-n-1]} \otimes B_{-n-k}, \\ Z^{k,2} &= (D_1 \otimes A_{-n} \otimes DB_k)^{(s)} - (n+k)a. \end{aligned}$$

With this notation, we have $Z_n^1 = (D_1 \otimes B_{-n})^{(s)} - na$ and $\tilde{Z}_n^2 = \max_{k \geq 1}[Z^{k,2}]$. We have then

$$\begin{aligned} Z^{k,2} &= \max_{i,j}[D_1^{(s,i)} + (A_{-n})^{(i,j)} + DB_k^{(j)}] - (n+k)a \\ &\leq \max_{i,j}[D_1^{(s,i)} + (A_{-n})^{(i,j)}] + \max_j[DB_k^{(j)}] - (n+k)a. \end{aligned}$$

First show that

$$\max_i[D_1^{(s,i)} + (A_{-n})^{(i,j)}] \leq \max_i[D_1^{(s,i)} + (B_{-n})^{(i)}].$$

Indeed, thanks to condition 2, we have $\max_{i \in I}[D_1^{(s,i)} + (A_{-n})^{(i,j)}] \leq \max_{i \in I}[D_1^{(s,i)} + (B_{-n})^{(i)}]$.
We have then only to prove that $\max_{i \in J}[D_1^{(s,i)} + (A_{-n})^{(i,j)}] \leq \max_{i \in I}[D_1^{(s,i)} + (B_{-n})^{(i)}]$.
But we have $\max_{i \in J}[D_1^{(s,i)} + (A_{-n})^{(i,j)}] = \max_{i \in J}[D_1^{(s,i)}]$ because $(A_{-n})^{(i,j)} = 0$ for $i \in J$. Moreover we have $\max_{i \in J}[D_1^{(s,i)}] \leq \max_{i \in I}[D_1^{(s,i)}] \leq \max_{i \in I}[D_1^{(s,i)} + (B_{-n})^{(i)}]$ and the equality follows.
Finally, we have

$$Z^{k,2} \leq Z_n^1 + \max_j[DB_k^{(j)}] - ka.$$

But $[DB_k^{(j)}] = (D_{[-n-k+1,-n-1]} \otimes B_{-n-k})^{(j)} \leq (D_{[-n-k+1,-n-1]} \otimes B_{-n-k})^{(s)}$, and we have then $Z^{k,2} \leq Z_n^1 + (D_{[-n-k+1,-n-1]} \otimes B_{-n-k})^{(s)} - ka$, and

$$\begin{aligned} Z_n^2 &\leq Z_n^1 + \max_{k \geq 1}[(D_{[-n-k+1,-n-1]} \otimes B_{-n-k})^{(s)} - ka] \\ &\leq Z_n^1 + Z_{[-\infty,-n-1]}. \end{aligned}$$

□

4.3 Tails in Generalized Jackson Networks with Subexponential Service Time Distributions

To the best of our knowledge, the literature on generalized Jackson networks with heavy tailed service times is limited to tandem queues. Bounds on the tail asymptotics of waiting and response times were considered in [17] and [57]. Exact asymptotics for these quantities were obtained in [14]. The present section addresses the case of generalized Jackson networks with arbitrary topology. It focuses on a key state variable, already used in the past for determining the stability region of such networks [13], [12], which is the time to empty the network when stopping the arrival process (this variable boils down to the virtual workload in an isolated queue or to the sojourn time for queues in tandem). The aim of this section is to derive an exact asymptotic for the tail of this state variable in the stationary regime. The main ingredients for the derivation of this result are

- a generalization of the so called "single big event theorem", well known for isolated queues, to such generalized Jackson networks which was established in [14] ; In the $GI/GI/1$ queue, this theorem states that in the case of subexponential service times, large workloads occur on a typical event where a single large service time has taken place in a distant past, and all other service time are close to their mean. Similarly, in generalized Jackson networks with subexponential service times, large maximal daters occur when a single large service time has taken place in one of the stations, and all other service times are close to their mean.
- the identification of the role played by fluid limits within the context of the single big event theorem for this class of networks ;
- the combination of these fluid limits and heavy tailed calculus which allows one to derive the closed forms formulas for the asymptotics.

Although this result sheds light on the way such a network experiences a deviation from its normal behavior, it is in no way final as the tail behavior of other state variables such as stationary queue size are still unknown. The derivation of the (more complex) asymptotic behavior of these other state variables was already obtained using a similar methodology in the particular case of tandem queues [14]. The extension of these queue size asymptotics to generalized Jackson networks with arbitrary topology seems to require much more effort and will not be pursued in the present section. The proposed method should however extend to other characteristics of stationary workload like for instance the sum of the residual service times of all customers present in the network at some given time.

This section is based on the paper [15].

4.3.1 Stochastic Assumptions

Service time and routing sequences

We recall here the notation used to describe a generalized Jackson network with K nodes.
The networks we consider are characterized by the fact that service times and routing decisions are associated with stations and not with customers. This means that the j-th service on station k takes $\sigma_j^{(k)}$ units of time, where $\{\sigma_j^{(k)}\}_{j\geq 1}$ is a predefined sequence. In the same way, when this service is completed, the leaving customer is sent to station $\nu_j^{(k)}$ (or leaves the network if $\nu_j^{(k)} = K+1$) and is put at the end of the queue on this station, where $\{\nu_j^{(k)}\}_{j\geq 1}$ is also a predefined sequence,

called the routing sequence. The sequences $\{\sigma_j^{(k)}\}_{j\geq 1}$ and $\{\nu_j^{(k)}\}_{j\geq 1}$, where k ranges over the set of stations, are called the driving sequences of the net. Node 0 models the external arrival of customers in the network, then the arrival time of the j-th customer in the network takes place at $\sigma_1^{(0)}+\cdots+\sigma_j^{(0)}$ and it joins the end of the queue of station $\nu_j^{(0)}$. Hence $\sigma_j^{(0)}$ is the j-th inter-arrival time.

The sample path construction we introduce here is that of [12]. The main interest of such a construction is that some monotonicity properties are preserved. These monotonicity properties as shown in [14] are crucial for our asymptotic calculation.

A generalized Jackson network will be defined by

$$\mathbf{JN} = \left\{ \{\sigma_j^{(k)}\}_{j\geq 1},\ \{\nu_j^{(k)}\}_{j\geq 0},\ n^{(k)},\ 0\leq k\leq K \right\},$$

where $N=(n^{(0)},n^{(1)},\ldots,n^{(K)})$ describes the initial condition. The interpretation is as follows : for $i\neq 0$, at time $t=0$, in node i, there are $n^{(i)}$ customers with service times $\sigma_1^{(i)},\ldots,\sigma_{n^{(i)}}^{(i)}$ (if appropriate, $\sigma_1^{(i)}$ may be interpreted as a residual service time).

The interpretation of $n^{(0)}$ is as follows :

- if $n^{(0)}=0$, there is no external arrival.
- if $\infty > n^{(0)} \geq 1$, then for all $1\leq j\leq n^{(0)}$, the arrival time of the j-th customer in the network takes place at $\sigma_1^{(0)}+\cdots+\sigma_j^{(0)}$. Note that in this case, there may be a finite number of customers passing through a given station so that the network is actually well defined once a finite sequence of routing decisions and service times is given on this station.
- if $n^{(0)}=\infty$, then when taking for instance the sequence $\{\sigma_j^{(0)}\}_{j\geq 1}$ independent and identically distributed (i.i.d.), the arrival process is a renewal process etc.

Euler route, Euler network

Consider a route $r=(r_1,\ldots,r_\phi)$ with $1\leq r_i\leq K$ for $i=2\ldots\phi-1$. Such a route is *successful* if $r_1=0$ and $r_\phi=K+1$. To such a route, we associate a routing sequence $\nu=(\nu^{(0)},\ldots,\nu^{(K)})$ as follows ($\oplus$ means here concatenation and $\emptyset$ the empty sequence) :

$$
\begin{array}{lll}
\mathbf{Procedure(r)} : & & \\
 & 1 & \mathbf{for}\quad k=0\ldots K \quad \mathbf{do} \\
 & & \quad \nu^{(k)} := \emptyset; \\
 & & \quad \phi^{(k)} := 0; \\
 & & \mathbf{od} \\
 & 2 & \mathbf{for}\quad i=1\ldots\phi-1 \quad \mathbf{do} \\
 & & \quad \nu^{(r_i)} := \nu^{(r_i)} \oplus r_{i+1}; \\
 & & \quad \phi^{(r_i)} := \phi^{(r_i)} + 1; \\
 & & \mathbf{od}
\end{array}
$$

Note that $\phi^{(j)}$ is the number of visits to node j in such a route.

A simple Euler network is a generalized Jackson network

$$E=\{\sigma,\nu,N\},$$

with $N = (1, 0, \dots, 0) = 1$, such that the routing sequence $\nu = \{\nu_i^{(k)}\}_{i=1}^{\phi^{(k)}}$ is generated by a successful route and such that $\sigma = \{\sigma_i^{(k)}\}_{i=1}^{\phi^{(k)}}$ is a sequence of real-valued non-negative numbers, representing service times.

Consider a sequence of simple Euler networks, say $\{E(n)\}_{n=-\infty}^{0}$, where $E(n) = \{\sigma(n), \nu(n), 1\}$. For $m \le n \le 0$, we define $\sigma_{[m,n]}$ and $\nu_{[m,n]}$ to be the concatenation of $\{\sigma(k)\}_{m \le k \le n}$ and $\{\nu(k)\}_{m \le k \le n}$ and then define the *composed* generalized Jackson network :

$$\mathbf{JN}_{[m,n]} = \{\sigma_{[m,n]}, \nu_{[m,n]}, N_{[m,n]}\}, \quad \text{with} \quad N_{[m,n]} = (n - m + 1, 0, \dots, 0).$$

Maximal dater

As proved in [12], for all possible values of $\nu(p)$ and $\sigma(p)$ in the simple Euler networks, for all integers $m \le n$, the composed network $\mathbf{JN}_{[m,n]}$ stays empty forever after some finite time. We denote by $X_{[m,n]}$ the time to empty $\mathbf{JN}_{[m,n]}$ forever and by $Z_{[m,n]} = X_{[m,n]} - \sum_{i=1}^{n-m+1} \sigma_{[m,n],i}^{(0)}$ the associated maximal dater. The sequence $Z_{[-n,0]}$ is an increasing sequence. We define the maximal dater of the generalized Jackson network $\mathbf{JN} = \{\sigma, \nu, N\}$ where σ and ν are the infinite concatenation of the $\{\sigma(n)\}_n$ and $\{\nu(n)\}_n$ and $N = (+\infty, 0, \dots, 0)$, by

$$Z = \lim_{n \to \infty} Z_{[-n,0]}. \tag{4.6}$$

To all generalized Jackson network $\mathbf{JN}_{[m,n]}$, we also associate the generalized Jackson network $\mathbf{JN}_{[m,n]}(Q)$ in which driving sequences are the same as in the original network except for the sequence $\{\sigma_j^{(0)}\}$ that is now $\sigma_j^{(0)} = 0$ for all j. Similarly we define $Z_{[m,n]}(Q)$ the time to empty the generalized Jackson network $\mathbf{JN}_{[m,n]}(Q)$.

Let

$$Y_i^{(k)} = \sum_{j=1}^{\phi^{(k)}(i)} \sigma_j^{(k)}(i) \tag{4.7}$$

be the total load brought by (external) customer i to station k. Note that

$$\begin{aligned} Z_i &= Z_{[i,i]} = Y_i^{(1)} + \dots + Y_i^{(K)}, \quad \forall i \\ Z_{[n,0]}(Q) &\ge \max_{j=1,\dots,K} \sum_{i=n}^{0} Y_i^{(j)}, \quad \forall n \le 0. \end{aligned}$$

Lemma 4 of [13] also implies that

$$\lim_{n \to \infty} \frac{Z_{[-n,0]}(Q)}{n} = b = \max_{1 \le k \le K} \mathbb{E}\left[Y_1^{(k)}\right] \quad \text{a.s.} . \tag{4.8}$$

Assumption 1, on the independence of routing and service times

All the sequences $\{\nu^{(k)}\}$ and $\{\sigma^{(k')}\}$ are mutually independent for k, k' ranging over the set of stations.

Assumption 2, on the independence of service times
We will assume the service times are independent for different stations and i.i.d. in each station with finite mean : $\mathbb{E}(\sigma^{(j)}) = \frac{1}{\mu^{(j)}} > 0$ for all $1 \leq j \leq K$.

Assumption 3, on routing
We assume that each of the successful routes used to build ν is obtained by a Markov chain on the state space $\{0, 1, \ldots, K, K+1\}$ with transition matrix

$$R = \begin{pmatrix} 0 & r_{0,1} & \ldots & \ldots & r_{0,K} & 0 \\ \vdots & r_{1,1} & r_{1,2} & \ldots & r_{1,K} & r_{1,K+1} \\ \vdots & r_{2,1} & r_{2,2} & \ldots & r_{2,K} & r_{2,K+1} \\ \vdots & & & & & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 & 1 \end{pmatrix}$$

This is equivalent to assuming that the routing decisions $\{\nu_j^{(k)}\}$ in station k are i.i.d. in j, independent of everything else, and such that the routing decision selects station i with probability $\mathbb{P}[\nu^{(k)} = i] = r_{k,i}$.
The fact that the routes built with this Markovian procedure are successful implies that state $K+1$ is the only absorbing state of this chain and all other states are transient ; we then have the very same Markovian routing assumptions as in (exponential) Jackson networks. More generally, when denoting by $\mathbb{E}_k$ the law of the chain with initial condition k, and V_j the number of visits of this absorbing chain in state j, we define :

$$\mathbb{E}_0[V_k] = \pi_k, \quad \mathbb{P}_0[V_k \geq 1] = p_k, \quad \mathbb{E}_k[V_j] = \pi_{k,j}. \tag{4.9}$$

We will use the following notation :

$$b_j = \frac{\pi_j}{\mu^{(j)}}, \quad b_{j,i} = \frac{\pi_{j,i}}{\mu^{(i)}}, \quad B_j = \max_i b_{j,i}.$$

With this notation, we have $b = \max_i \pi_i/\mu^{(i)} = \max_i b_i$. Let $\lambda^{-1} = \mathbb{E}[\sigma_0] = a$. Throughout this paper we will assume that the stability condition holds :

$$\lambda b < 1. \tag{4.10}$$

Theorem 13 of [12] applies so that if $\lambda b < 1$ then $Z < \infty$ a.s. ; conversely, if $\lambda b > 1$, $Z = \infty$ a.s.

Example 1. *As an example,we will consider a network with the following routing matrix*

$$R = \left\{ \begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & p & 1-p \\ 0 & q & 0 & 1-q \\ 0 & 0 & 0 & 1 \end{array} \right\}.$$

In this case, we have

$$\left\{ \begin{array}{l} \pi_1 = 1 + q\pi_2, \\ \pi_2 = p\pi_1. \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \pi_1 = \frac{1}{1-pq}, \\ \pi_2 = \frac{p}{1-pq}. \end{array} \right.$$

Similarly, we have :

$$\left\{\begin{array}{l}\pi_{1,1}=\pi_1,\\ \pi_{1,2}=\pi_2.\end{array}\right. \quad \left\{\begin{array}{l}\pi_{2,1}=\frac{q}{1-pq},\\ \pi_{2,2}=\frac{1}{1-pq}.\end{array}\right. \quad \text{and} \quad \left\{\begin{array}{l}p_1=1,\\ p_2=p.\end{array}\right.$$

Hence we have

$$b=B_1=\frac{1}{1-pq}\max\left(\frac{1}{\mu^{(1)}},\frac{p}{\mu^{(2)}}\right) \quad \text{and} \quad B_2=\frac{1}{1-pq}\max\left(\frac{q}{\mu^{(1)}},\frac{1}{\mu^{(2)}}\right).$$

In particular for $p=1$ and $q=0$, we are dealing with the case of queues in tandem and we have $b=B_1=\max(1/\mu^{(1)},1/\mu^{(2)})$ and $B_2=1/\mu^{(2)}$.

Assumption 4, on the subexponentiality of service times
The assumptions concerning service times are the following : there exists a distribution function F on $\mathbb{R}_+$ such that :

1. F is subexponential, with finite first moment M.
2. The integrated distribution F^s is subexponential.
3. The following equivalence holds when x tends to ∞ :

$$\mathbb{P}(\sigma_1^{(k)}>x)\sim c^{(k)}\overline{F}(x),$$

for all $k=1,\dots,K$ with $\sum_{k=1}^K c^{(k)}=c>0$.

4.3.2 Main Result

We first introduce some notations ; the intuitive meaning of these quantities will be given later on.
Let $f^j(\sigma,n)$ be the following piece-wise linear function of (σ,n), where σ and n are non-negative real numbers :

$$f^j(\sigma,n)=\mathbb{1}_{\{\sigma>na\}}\{\sigma-na+np_jB_j\}+\mathbb{1}_{\{\sigma\le na\}}\max_k\left\{p_jb_{j,k}\frac{\sigma}{a}+\left(\frac{b_k}{a}-1\right)(na-\sigma)\right\}^+ \tag{4.11}$$

and for all positive real numbers x, and all $j=1,\dots,K$, let $\Delta^j(x)$ be the following domain :

$$\Delta^j(x) \quad = \quad \{(\sigma,t)\in\mathbb{R}_+^2,\ f^j(\sigma,t)>x\}. \tag{4.12}$$

Remark 13. We may rewrite function f as follows

$$f^j(\sigma,n)=\left\{\sigma-na+\max_k\left(nb_k-\frac{\sigma}{a}(b_k-p_jb_{j,k});np_jb_{j,k}\right)\right\}^+.$$

This is due to the fact that $p_jb_{j,k}\le b_k$. In particular, we see that

$$f^j(\sigma,n)\ge\sigma-na-np_jB_j. \tag{4.13}$$

Example 2. *We continue with previous example and we have :*

$$\begin{aligned} f^1(\sigma,n) &= \left\{\sigma - na + \frac{n}{1-pq}\max\left(\frac{1}{\mu^{(1)}};\frac{p}{\mu^{(2)}}\right)\right\}^+, \\ f^2(\sigma,n) &= \left\{\sigma - na + \max\left(\frac{npq}{(1-pq)\mu^{(1)}},\frac{np}{(1-pq)\mu^{(2)}};\frac{n}{(1-pq)\mu^{(1)}} - \frac{\sigma}{a\mu^{(1)}}\right)\right\}^+. \end{aligned}$$

In the specific case of queues in tandem ($p = 1$ and $q = 0$), these equations reduce to

$$\begin{aligned} f^1(\sigma,n) &= \left\{\sigma - na + n\max\left(\frac{1}{\mu^{(1)}};\frac{1}{\mu^{(2)}}\right)\right\}^+, \\ f^2(\sigma,n) &= \left\{\sigma - na + \frac{n}{\mu^{(2)}}\right\}^+. \end{aligned}$$

And in this specific case, the corresponding domains are

$$\Delta^1(x) = \left\{\sigma > x + t\left(a - \max\left(\frac{1}{\mu^{(1)}};\frac{1}{\mu^{(2)}}\right)\right)\right\}, \tag{4.14}$$

$$\Delta^2(x) = \left\{\sigma > x + t\left(a - \frac{1}{\mu^{(2)}}\right)\right\}. \tag{4.15}$$

Theorem 9. *Consider a stable generalized Jackson network with subexponential service time distributions satisfying assumptions 1-4. Let Z denote its stationary maximal dater at customer arrivals. When $x \to \infty$,*

$$\mathbb{P}[Z > x] \sim \sum_{j=1}^{K} \pi_j \int\int_{\{(\sigma,t)\in\Delta^j(x)\}} \mathbb{P}\left[\sigma^{(j)} \in d\sigma\right] dt. \tag{4.16}$$

This equation may be rewritten with the constants $\{\alpha_i^j, \beta_i^j, \gamma_i^j\}_{0\le i\le l}$ that will be calculated in Lemma 21 below as follows :

$$\mathbb{P}[Z > x] \sim \sum_{j=1}^{K} \pi_j \left\{ \sum_{i=0}^{l} \sum_{\{\alpha_i^j x \le n < \alpha_{i+1}^j x\}} \mathbb{P}\left[\sigma^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j\right]\right\}, \tag{4.17}$$

or with $\delta_i^j = 1/\beta_i^j + \alpha_i^j\gamma_i^j$ and $d^{(j)} = \pi_j c^{(j)}$,

$$\mathbb{P}[Z > x] \sim \sum_{j=1}^{K} d^{(j)} \left\{ \sum_{i=0}^{l} \frac{1}{\gamma_i^j}\left[\overline{F}^s(\delta_i^j x) - \overline{F}^s(\delta_{i+1}^j x)\right]\right\}. \tag{4.18}$$

1. *If $\overline{F}^s \in \mathcal{R}(-\alpha)$, with $\alpha > 0$, we can rewrite Equation (4.18) as :*

$$\frac{\mathbb{P}[Z > x]}{\overline{F}^s(x)} \to \sum_{j=1}^{K} d^{(j)} \left\{ \sum_{i=0}^{l} \frac{1}{\gamma_i^j}\left[(\delta_i^j)^{-\alpha} - (\delta_{i+1}^j)^{-\alpha}\right]\right\}. \tag{4.19}$$

2. *If* $\overline{F}^s \in \mathcal{R}(-\infty)$, *then, we have*

$$\frac{\mathbb{P}[Z > x]}{\overline{F}^s(x)} \to \sum_{j=1}^{K} \frac{d^{(j)}}{a - p_j B_j}. \tag{4.20}$$

Remark 14. In view of Inequality (4.13), we see that $\{\sigma > x + t(a - p_j B_j)\} \subset \Delta^{(j)}(x)$, hence,

$$\mathbb{P}[Z > x] \geq \sum_{j=1}^{K} \pi_j \sum_{n \geq 0} \mathbb{P}\left[\sigma^{(j)} > x + n(a - p_j B_j)\right] \sim \sum_{j=1}^{K} \frac{d^{(j)}}{a - p_j B_j} \overline{F}^s(x).$$

Hence in general the asymptotic as described in Equation (4.20) is a lower bound. In the rapidly varying case, this lower bound is tight. In the regularly varying case, the complete shape of the domain has to be taken in consideration and indeed the part $\Delta^j(x) \setminus \{\sigma > x + t(a - p_j B_j)\}$ is not anymore negligible due to the scale property of regularly varying functions. More insights into the shape of the domain will be given in Section 4.3.6.

Example 3. *In the specific case of queues in tandem, thanks to (4.14) and (4.15), we see that Equation (4.16) reduces to*

$$\mathbb{P}[Z > x] \;\sim\; \sum_{n \geq 0} \mathbb{P}\left[\sigma^{(1)} > x + n\left(a - \max\left(\frac{1}{\mu^{(1)}}, \frac{1}{\mu^{(2)}}\right)\right)\right] + \sum_{n \geq 0} \mathbb{P}\left[\sigma^{(2)} > x + n\left(a - \frac{1}{\mu^{(2)}}\right)\right],$$

which corresponds to the exact asymptotic of Theorem 9 of [14].

4.3.3 Technical Conditions

Under Assumption 1-3, the properties **(IA)** and **(AA)** of [14], which read

- **(IA)** the sequence of simple Euler networks $\{E(n)\}_{n=0}^{-\infty}$ consists of i.i.d. random variables.
- **(AA)** the random variables $\{Y_i^{(k)}\}$ are independent of the inter-arrival times, and such that the sequence of random vectors $(Y_i^{(1)}, \ldots, Y_i^{(K)})$ is i.i.d. (general dependences between the components of the vector $(Y_i^{(1)}, \ldots, Y_i^{(K)})$ are allowed),

are both satisfied.

Under Assumption 1, the variable Z associated to $\mathbf{JN} = \{\sigma, \nu, N\}$ represents the stationary maximal dater of the generalized Jackson network, namely the time that it would take in steady state to clear the workload of all customers present in the system when stopping future arrivals. Under Assumption 4, the assumptions **(SE)** and **(H)** of [14] are satisfied :

- **(SE)** For all $k = 1, \ldots, K$

$$\mathbb{P}(Y_1^{(k)} > x) \sim \pi^k \mathbb{P}(\sigma^{(k)} > x) \sim d^{(k)} \overline{F}(x),$$

  with $d^{(k)} = c^{(k)} \pi_k$ and then $d := \sum_k d^{(k)} > 0$.
- **(H)**

$$\mathbb{P}(\sum_{k=1}^{K} Y_1^{(k)} > x) \sim \mathbb{P}(\max_{1 \leq k \leq K} Y_1^{(k)} > x) \sim \sum_{k=1}^{K} \mathbb{P}(Y_1^{(k)} > x) \sim d\overline{F}(x).$$

See Sections 4.4.2 and 7.2 of [14].
Under Assumption 4, there exists a non-decreasing integer-valued function $N_x \to \infty$ and such that, for all finite real numbers b,

$$\sum_{n=0}^{N_x} \overline{F}(x+nb) = o\left(\overline{F}^s(x)\right), \quad x \to \infty \tag{4.21}$$

(see Section 4.1.2 of [14]).

4.3.4 Single Big Event Theorem

As already mentioned, one of the tools we will use within this setting is the ”single big event theorem” for generalized Jackson networks. More precisely, Theorems 7 and 8 of [14] give the following result :

Proposition 14. *Let Z be the stationary maximal dater of the generalized Jackson network defined in (4.6). For any x and for $j = 1, \dots, r$, let $\{K^j_{n,x}\}$ be a sequence of events such that*

1. *for any n and j, the event $K^j_{n,x}$ and the random variable $Y^{(j)}_{-n} = \sum_{k=1}^{\phi^{(j)}(-n)} \sigma^{(j)}_k(-n)$ are independent ;*
2. *for any j, $\mathbb{P}(K^j_{n,x}) \to 1$ uniformly in $n \geq N_x$ as $x \to \infty$.*

For all sequences $\epsilon_n \to 0$, we denote $x_n = x + n(a - b + \epsilon_n)$. Then, as $x \to \infty$,

$$\mathbb{P}[Z > x] \sim \sum_{j=1}^{K} \sum_{n \geq N_x} \mathbb{P}[Z > x, Y^{(j)}_{-n} > x_n, K^j_{n,x}],$$

and

$$\mathbb{P}[Z > x] = \mathcal{O}(\overline{F}^s(x)).$$

This property leads to the following and more handy result :

Corollary 1. *Take any sequence of events $\{K^j_n\}$ such that for any j, K^j_n and the random variable $Y^{(j)}_{-n}$ are independent and $\mathbb{P}(K^j_n) \to 1$ as $n \to \infty$. Take $z_x \to \infty$, $z_x = o(x)$, such that $\overline{F}^s(x \pm z_x) \sim \overline{F}^s(x)$, and denote :*

$$G(x) = \sum_{j=1}^{K} \sum_{n \geq N_x} \mathbb{P}\left[Z_{[-n,0]} > x, K^j_n, Y^{(j)}_{-n} > x_n, \phi^{(j)}(-n) \leq L\right].$$

Then, we have :

$$(1 + \epsilon(x))G(x) \leq \mathbb{P}[Z > x] \leq (1 + \epsilon(x))G(x - z_x) + \epsilon(L, x)\overline{F}^s(x). \tag{4.22}$$

If G is long tailed, we have as $x \to \infty$

$$\mathbb{P}[Z > x] \sim G(x).$$

The proof is forwarded at the end of the section.

4.3.5 Fluid Limit

We recall some results from Section 3.1.4. We have to find sequences of events $\{K_n^j\}$ allowing one to calculate the sum

$$\sum_{n\geq N_x} \mathbb{P}\left[Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L\right] \tag{4.23}$$

where as above, $x_n = x + n(a - b + \epsilon_n)$.
The events in question will be based on the piece-wise linear functions $f^j(\sigma, n)$ defined in (4.11). Let us describe the intuitive reason for introducing this function. Assume the big service time is equal to σ and takes place on station j and within the set of service times of the simple Euler network $E(-n)$. Let us look at the maximal dater $Z_{[-n,0]}$ in the fluid scale suggested by the a.s. limit of (4.8) :

- if $\sigma > na$, then the number of customers blocked in station j at time σ is of the order of np_j, whereas the number of customers in the other stations is small. So, according to (4.8), the time to empty the network from time σ on should be of the order np_jB_j ; hence, in this case, the maximal dater in question should be of the order of $f^j(\sigma, n)$ indeed ;
- if $\sigma < na$, then at time σ, the number of customers blocked in station j is of the order of $p_j\frac{\sigma}{a}$, and the other stations have few customers ; from time σ to the time of the last arrival (which is of the order of na), station k has to serve approximately the load $p_j\frac{\sigma}{a}b_{j,k}$ generated by these blocked customers plus the load $(na - \sigma)\frac{b_k}{a}$ generated by the external arrivals on the time interval from σ to the last arrival. On this time interval, the service capacity is of the order of $(na - \sigma)$. Hence the maximal dater should again be of the order of $f^j(\sigma, n)$.

We now return to rigor.
Consider a generalized Jackson network built from the i.i.d. sequence of simple Euler networks $\{E(k)\}$. To all simple Euler networks E and all positive integers n, we associate the network $\mathbf{JN}^n(E)$ with input $\{\tilde{E}(k)\}_{k=-n}^{\infty}$, where $\tilde{E}(k) = E(k)$ for all $k > -n$ and $\tilde{E}(-n) = E$. That is, if we denote by $\sigma^{(k),n}$ and $\nu^{(k),n}$ the concatenations $(\{\sigma^{(k)}(E)\}, \{\sigma^{(k)}(-n+1)\}, \ldots, \{\sigma^{(k)}(0)\}, \ldots)$ and $(\{\nu^{(k)}(E)\}, \{\nu^{(k)}(-n+1)\}, \ldots, \{\nu^{(k)}(0)\}, \ldots)$ respectively, then

$$\mathbf{JN}^n(E) \quad = \quad \{\sigma^n(E), \nu^n(E), 0, N^n\}, \quad \text{with} \quad N^n = (n, 0, \ldots, 0).$$

The maximal dater of order $[-n, 0]$ in this network will be denoted by $\tilde{Z}^n(E)$. Of course $\tilde{Z}^n(E(n)) = Z_{[-n,0]}$. More generally, we will add the superscript n to any other function associated to a network to mean that of network $\mathbf{JN}^n(E)$.
For all simple Euler networks $E = (\sigma, \nu, 1)$, let $Y^{(j)}(E) = \sum_{u=1}^{\phi^{(j)}} \sigma_u^{(j)}$.
We are now in a position to state the main result pertaining to the fluid limit. Let ϵ_n, z_n be some sequences of positive real numbers ; we define :

$$\begin{aligned}
\mathbf{U}^j(n) &= \{E \text{ is a simple Euler network such that } Y^{(k)}(E) \leq z_n \ \forall k \neq j\}, \\
\mathbf{V}^j(n) &= \{E \in \mathbf{U}^j(n),\ Y^{(j)}(E) \geq n(a-b),\ \phi^{(j)} \leq L\}, \\
K_n^j &= \left\{\sup_{\{E\in\mathbf{V}^j(n)\}} \left|\frac{\tilde{Z}^n(E) - f^j(Y^{(j)}(E), n)}{n}\right| \leq \epsilon_n\right\} \cap \left\{E(-n) \in \mathbf{U}^j(n)\right\}.
\end{aligned} \tag{4.24}$$

We first recall a result that derives directly from Proposition 10 and the remark following this proposition.

Proposition 15. *Under the previous assumptions, there exists a sequence $z_n \to \infty$ with $\frac{z_n}{n} \to 0$, such that we have*

$$\sup_{\{E \in \mathbf{V}^j(n)\}} \left| \frac{\tilde{Z}^n(E) - f^j(Y^{(j)}(E), n)}{n} \right| \xrightarrow{n\to\infty} 0 \quad \textit{a.s.}$$

Lemma 20. *Let $\{K_n^j\}$ be the sequence of events defined above. K_n^j and the random variable $Y_{-n}^{(j)}$ are independent and there exist sequences $\epsilon_n \to 0$ and $z_n \to \infty$ with $\frac{z_n}{n} \to 0$, such that we have $\mathbb{P}[K_n^j] \to 1$ as $n \to \infty$.*

Proof.

The left-hand part of the definition of K_n^j depends on $\{E(k)\}_{k=-n+1}^{0}$ and the right-hand part depends only on $E(-n)$, hence we have

$$\mathbb{P}[K_n^j] = \mathbb{P}\left[\sup_{\{E \in \mathbf{V}^j(n)\}} \left| \frac{\tilde{Z}^n(E) - f^j(Y^{(j)}(E), n)}{n} \right| \le \epsilon_n \right] \mathbb{P}\left[E(-n) \in \mathbf{U}^j(n)\right].$$

The distribution of $E(-n)$ does not depend on n, hence $Y^{(i)}(E(-n))/n \to 0$ a.s. since its mean is finite. Therefore, there exists a sequence $z_n \to \infty$, $z_n/n \to 0$ such that

$$\mathbb{P}(Y^{(i)}(E(-n)) \le n z_n,\ \forall i \neq j) = \mathbb{P}\left[E(-n) \in \mathbf{U}^j(n)\right] \to 1$$

uniformly in $n \ge N_x$ as $x \to \infty$.

The first term derives directly from Proposition 15. Therefore, there exist sequences $\epsilon_n \to 0$ and $z_n \to \infty$ and $\frac{z_n}{n} \to 0$, such that we have $\mathbb{P}[K_n^j] \to 1$ uniformly in $n \ge N_x$ as $x \to \infty$. □

4.3.6 Computation of the Exact Asymptotics

Thanks to Lemma 20, it is easy to see that the sequence of events $\{K_n^j\}$ defined in (4.24) satisfies assumptions of Corollary 1. Moreover, we will see that we are now able to calculate the sum in Equation (4.23) which will give the exact asymptotic for $\mathbb{P}[Z > x]$ in Theorem 9. Before stating this result, we need to introduce some notation.

On the event $K_n^j \cap \left\{ Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \le L \right\}$, we have

$$Z_{[-n,0]} = f^j(Y_{-n}^{(j)}, n) + n\eta_n, \quad \text{with } \eta_n \text{ r.v. such that } |\eta_n| \le \epsilon_n.$$

Then $\{Z_{[-n,0]} > x\} = \{f^j(Y_{-n}^{(j)}, n) > x - n\eta_n\}$. In order to prove equivalence (4.16), we will first give an explicit form for the domains Δ^j defined in (4.12).

Lemma 21. *There exist constants $\{\alpha_i^j, \beta_i^j, \gamma_i^j\}_{0 \le i \le l}$ (given in closed form in the proof of the lemma as function of the quantities p_j and $b_{j,k}$ defined in Section 4.3.1) with $0 = \alpha_0^j \le \alpha_1^j \ldots \le \alpha_l^j$, such that :*

$$\Delta^j(z) \quad = \quad \bigcup_{i=0}^{l} \left\{ \alpha_i^j z \le t < \alpha_{i+1}^j z,\ \sigma > \frac{z}{\beta_i^j} + t\gamma_i^j \right\}, \tag{4.25}$$

with the convention $\alpha^j_{l+1} = +\infty$*. Moreover, we have*

$$\alpha^j_0 = 0, \quad \alpha^j_1 = 1/p_jB_j, \quad \beta^j_0 = 1, \quad \gamma^j_0 = a - p_jB_j,$$

for all j*. In addition,* $\beta^j_i \leq 1$ *for all* i, j *and the following inclusion holds :*

$$\{\sigma \geq z + t(a - p_jB_j)\} \subset \Delta^j(z). \tag{4.26}$$

Proof.
The domain Δ^j may be divided in two parts :

$$\begin{aligned}\Delta^j(z) &= \{(\sigma, t),\ f^j(\sigma, t) > z\}. \\ &= \{\sigma > ta,\ \sigma > z + t(a - p_jB_j)\} \cup \left\{\sigma \leq ta,\ \sigma > a \min_k \frac{z + t(a - b_k)}{a - b_k + p_jb_{j,k}}\right\}.\end{aligned}$$

For the first part, we have (see Figure 4.1) :

$$\{\sigma > ta,\ \sigma > z + t(a - p_jB_j)\} = \left\{0 \leq t < \frac{z}{p_jB_j},\ \sigma > z + t(a - p_jB_j)\right\} \cup \left\{\frac{z}{p_jB_j} \leq t,\ \sigma > ta\right\}.$$

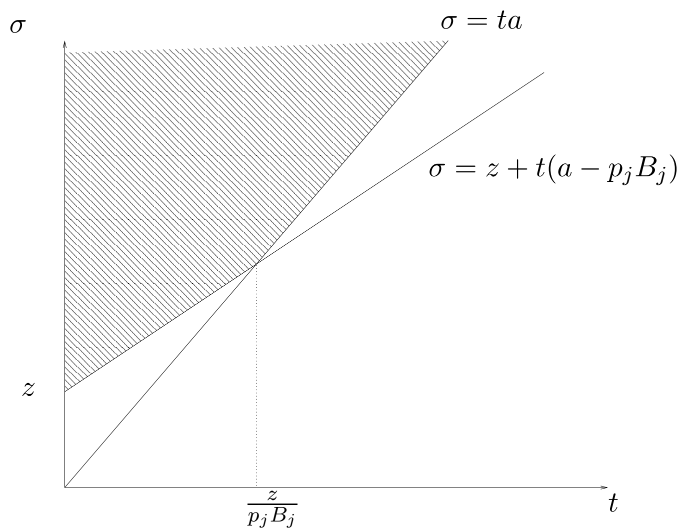


FIG. 4.1 – First part of $\Delta^j(z)$

For the second part, we have (see Figure 4.2) :

$$\left\{\sigma \leq ta,\ \sigma > a \min_k \frac{z + t(a - b_k)}{a - b_k + p_jb_{j,k}}\right\} = \bigcup_k \left\{\frac{z}{p_jb_{j,k}} \leq t,\ a\frac{z + t(a - b_k)}{a - (b_k - p_jb_{j,k})} < \sigma \leq ta\right\}.$$

Now, it is easy to see that the lemma holds (see Figure 4.3).
The inequality on the β's follows directly from the fact that $p_jb_{j,k} \leq b_k$ from which, we have

$$\frac{a}{a + p_jb_{j,k} - b_k} \geq 1.$$

□

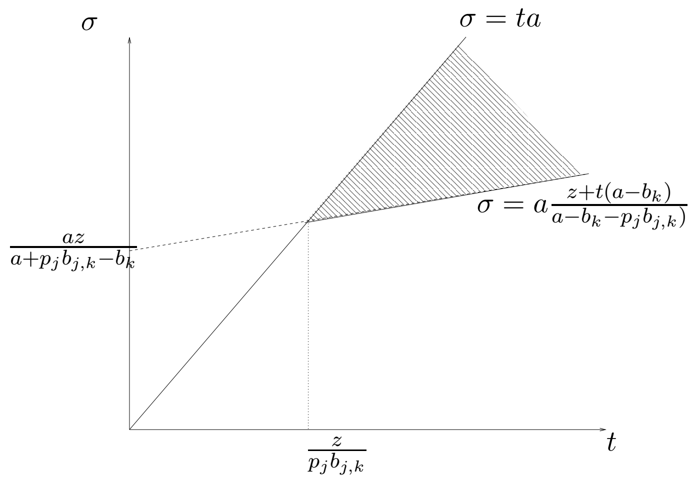


FIG. 4.2 – Construction of the second part of $\Delta^j(z)$

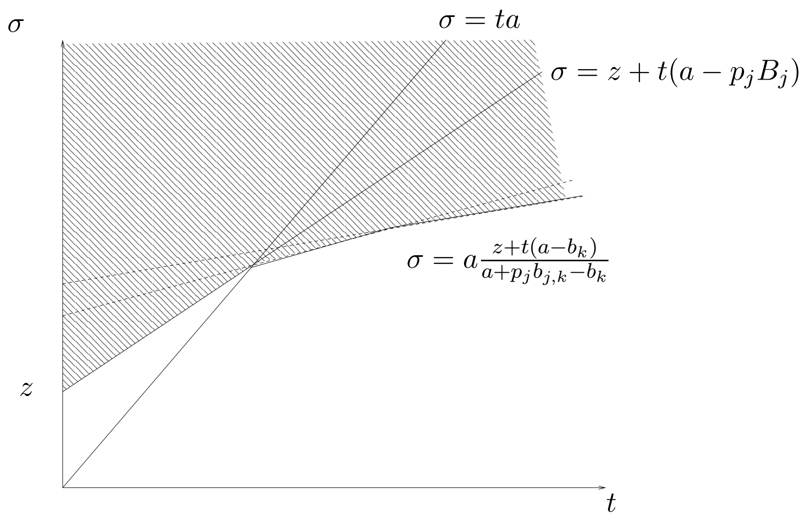


FIG. 4.3 – Domain $\Delta^j(z)$

Lemma 22. *Let X be a random variable such that $\overline{F}^s \in \mathcal{S}$, $\epsilon_n \to 0$ as $n \to \infty$ and*

$$\frac{a(x)}{x} \to a, \quad \frac{b(x)}{x} \to b, \quad \text{with } 0 < a < b \text{ as } x \to \infty.$$

If $\overline{F}(x) = \mathbb{P}[X > x]$, for $\alpha \geq 1,\ \beta > 0$, we have as $x \to \infty$

$$\sum_{a(x) \leq n < b(x)} \mathbb{P}[X > \alpha x + n(\beta + \epsilon_n)] - \sum_{ax \leq n < bx} \mathbb{P}[X > \alpha x + n\beta] = o(\overline{F}^s(x)).$$

Proof.

For the simplicity of notation, we assume that $a(x) \leq ax$ for all x. We have

$$\begin{aligned}\sum_{a(x)\leq n\leq ax} \mathbb{P}[X > \alpha x + n(\beta+\epsilon_n)] &= \frac{1+\epsilon(x)}{\beta}\int_{\alpha x + a(x)\beta}^{\alpha x + ax\beta} \overline{F}(u)du \\ &\leq \frac{1+\epsilon(x)}{\beta}\frac{ax-a(x)}{ax}\overline{F}^s(\alpha x)\end{aligned}$$

since $\overline{F}(x)$ is non-increasing. Hence, we have only to prove the lemma for $a(x) = ax$ and $b(x) = bx$. We have the following bound with $\delta_x = \sup_{n\geq ax}\epsilon_n$.

$$\begin{aligned}\sum_{ax\leq n< bx} \mathbb{P}[X > \alpha x + n(\beta+\epsilon_n)] - \mathbb{P}[X > \alpha x + n\beta] &\leq \sum_n \mathbb{P}[X \in (\alpha x + n\beta, \alpha x + n(\beta+\delta_x)] \\ &= (1+\epsilon(x))\overline{F}^s(\alpha x)\left(\frac{1}{\beta} - \frac{1}{\beta+\delta_x}\right) \\ &= o(\overline{F}^s(\alpha x)) = o(\overline{F}^s(x)).\end{aligned}$$

□

Proof of Theorem 9.

Thanks to Corollary 1, we know that the tail asymptotic of the maximal dater is linked to the quantity $S(j)$ defined by

$$S(j) = \sum_{n\geq N_x} \mathbb{P}\left[Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L\right].$$

On the event $A_{n,x}^j = K_n^j \cap \left\{Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L\right\}$, we have

$$\begin{aligned}\{Z_{[-n,0]} > x\} &= \{f^j(Y_{-n}^{(j)}, n) > x - n\eta_n\} \\ &= \{(Y_{-n}^{(j)}, n) \in \Delta^j(x - n\eta_n)\}.\end{aligned}$$

Clearly $\Delta^j(z)$ is a non-increasing function of z and we define

$$D_-^j = \Delta^j(x - n\epsilon_n) \supset \Delta^j(x - n\eta_n) \supset \Delta^j(x + n\epsilon_n) = D_+^j.$$

For simplicity of notation, we write $Y^{(j)} = Y_{-n}^{(j)}$ and $\phi^{(j)} = \phi^{(j)}(-n)$. We assume w.l.o.g. that ϵ_n is a decreasing sequence, hence for $n \geq N_x$, $\epsilon_n \leq \epsilon_{N_x} = \epsilon_x$ and we have for $n \geq N_x$

$$\begin{aligned}A_+(n) &= \mathbb{P}\left[(Y^{(j)}, n) \in D_+^j\right] \\ &= \sum_{i=0}^{l} \mathbf{1}_{\{\alpha_i^j(x+n\epsilon_n)\leq n<\alpha_{i+1}^j(x+n\epsilon_n)\}}\mathbb{P}\left[Y^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j + \frac{n\epsilon_n}{\beta_i^j}\right] \\ &\leq \sum_{i=0}^{l} \mathbf{1}_{\{\alpha_i^j x\leq n<\alpha_{i+1}^j(x+n\epsilon_x)\}}\mathbb{P}\left[Y^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j + \frac{n\epsilon_n}{\beta_i^j}\right].\end{aligned}$$

Then we have

$$\begin{aligned}\sum_{n\geq N_x} A_+(n) \leq & \sum_{\{N_x\leq n<\alpha_1^j x(1+\epsilon(x))\}} \mathbb{P}\left[Y^{(j)} > x + n(a - p_j B_j) + n\epsilon_n\right] \\ & + \sum_{i=1}^{l} \sum_{\{\alpha_i^j x\leq n<\alpha_{i+1}^j x(1+\epsilon(x))\}} \mathbb{P}\left[Y^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j + \frac{n\epsilon_n}{\beta_i^j}\right].\end{aligned}$$

Thanks to Assumption 2, we know that $Y^{(j)}$ satisfies assumption of Lemma 47 and we have

$$\begin{aligned}\sum_{n\geq N_x} A_+(n) &= \sum_{\{0\leq n<\alpha_1^j x\}} \mathbb{P}\left[Y^{(j)} > x + n(a - p_j B_j)\right] \\ &\quad + \sum_{i=1}^{l} \sum_{\{\alpha_i^j x\leq n<\alpha_{i+1}^j x\}} \mathbb{P}\left[Y^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j\right] + \epsilon(x)\overline{F}^s(x) \\ &= (1+\epsilon(x))\sum_{i=0}^{l} \sum_{\{\alpha_i^j x\leq n<\alpha_{i+1}^j x\}} \mathbb{E}[\phi^{(j)}]\mathbb{P}\left[\sigma^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j\right] + \epsilon(x)\overline{F}^s(x),\end{aligned}$$

where the last equality follows from assumption **(SE)**. But we have

$$S(j) \leq \sum_{n\geq N_x} A_+(n).$$

We now look at the lower bound. With the same arguments as above, we easily get with

$$A_-(n) = \mathbb{P}\left[(Y^{(j)}, n) \in D_-^j\right],$$

that,

$$\sum_{n\geq N_x} A_-(n) = \sum_{n\geq N_x} A_+(n) + \epsilon(x)\overline{F}^s(x).$$

We now show that

$$\sum_{n\geq N_x} A_-(n) = \sum_{n\geq N_x} \mathbb{P}\left[(Y^{(j)}, n) \in D_-^j, A_{n,x}^j\right] + \epsilon(x, L)\overline{F}^s(x).$$

Consider the difference

$$\begin{aligned}A_-(n) - \mathbb{P}\left[(Y^{(j)}, n) \in D_-^j, A_{n,x}^j\right] &\leq \mathbb{P}\left[(Y^{(j)}, n) \in D_-^j, \phi^{(j)}(-n) > L\right] \\ &\leq \mathbb{P}\left[Y^{(j)} \geq x + n(a - p_j B_j - \epsilon_n), \phi^{(j)}(-n) > L\right]\end{aligned}$$

where the last inequality follows from inclusion (4.26) of Lemma 21. With the same kind of argument as in Corollary 1, we have

$$\sum_{n\geq N_x} A_-(n) - \mathbb{P}\left[(Y^{(j)}, n) \in D_-^j, A_{n,x}^j\right] \leq \epsilon(x, L)\overline{F}^s(x).$$

Hence, we proved that when $x \to \infty$, we have

$$S(j) \sim \sum_{i=0}^{l} \sum_{\{\alpha_i^j x\leq n<\alpha_{i+1}^j x\}} \mathbb{E}[\phi^{(j)}]\mathbb{P}\left[\sigma^{(j)} > \frac{x}{\beta_i^j} + n\gamma_i^j\right].$$

Now since this quantity is long tailed, we use Corollary 1 to derive the asymptotic for $\mathbb{P}[Z > x]$.
□

4.3.7 Proof of Corollary 1

The proof is based on Proposition 14, which shows that we have

$$\mathbb{P}[Z > x] = (1 + \epsilon(x)) \left(\sum_{j=1}^{K} \sum_{n \geq N_x} \mathbb{P}[Z > x, Y_{-n}^{(j)} > x_n, K_n^j] \right).$$

Since $Z \geq Z_{[-n,0]}$, we have

$$\begin{aligned}\mathbb{P}[Z > x, Y_{-n}^{(j)} > x_n, K_n^j] &\geq \mathbb{P}[Z_{[-n,0]} > x, Y_{-n}^{(j)} > x_n, K_n^j] \\ &\geq \mathbb{P}[Z_{[-n,0]} > x, Y_{-n}^{(j)} > x_n, K_n^j, \phi^{(j)}(-n) \leq L].\end{aligned}$$

Hence we have

$$\mathbb{P}[Z > x] \quad \geq \quad (1 + \epsilon(x)) \left(\sum_{j=1}^{K} \sum_{n \geq N_x} \mathbb{P}[Z_{[-n,0]} > x, Y_{-n}^{(j)} > x_n, K_n^j, \phi^{(j)}(-n) \leq L] \right).$$

We now derive the upper bound. Take $z_x \to \infty$ such that $\overline{F}^s(x+z_x) \sim \overline{F}^s(x)$, then when $x \to \infty$, we have

$$\mathbb{P}[Z_{[-\infty,-n-1]} < z_x] = \mathbb{P}[Z < z_x] \to 1.$$

We define $\tilde{x} = x + z_x$, and $\tilde{K}_{n,x}^j = K_n^j \cap \{Z_{[-\infty,-n-1]} \leq z_x\}$. Observe that $\tilde{K}_{n,x}^j$ satisfies also assumptions of Proposition 14. By sub-additivity, we have $Z \leq Z_{[-\infty,-n-1]} + Z_{[-n,0]}$ (see [12]), hence

$$\begin{aligned}\mathbb{P}(Z > \tilde{x}, \tilde{K}_{n,x}^j, Y_{-n}^{(j)} > \tilde{x}_n) &\leq \mathbb{P}(Z_{[-\infty,-n-1]} + Z_{[-n,0]} > \tilde{x}, \tilde{K}_{n,x}^j, Y_{-n}^{(j)} > \tilde{x}_n) \\ &\leq \mathbb{P}(Z_{[-n,0]} > x, \tilde{K}_{n,x}^j, Y_{-n}^{(j)} > x_n) \\ &\leq \mathbb{P}(Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n).\end{aligned}$$

We now make the truncation of ϕ.

$$\begin{aligned}A(n) &= \mathbb{P}\left[Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n\right] \\ &\leq \mathbb{P}\left[Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L\right] + \mathbb{P}\left[Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) > L\right] \\ &= \mathbb{P}\left[Z_{[-n,0]} > x, K_n^j, Y^{(j)}(E(-n)) > x_n, \phi^{(j)}(-n) \leq L\right] + B(n).\end{aligned}$$

We will use the following result due to Kesten (for a proof see Athreya and Ney [9]) :

Lemma 23. *Let $X \in \mathcal{S}$ and let S_n be the sum of n independent copies of X. Then for every $\epsilon > 0$, there exists $K(\epsilon) > 0$ such that*

$$\sup_{x \geq 0} \frac{\mathbb{P}[S_n > x]}{\mathbb{P}[X > x]} \leq K(\epsilon)(1+\epsilon)^n, \quad n = 1, 2, \ldots$$

Recall that $\mathbb{P}(\phi^{(j)}(0) = l) = \delta^l(1-\delta)$ for some $0 < \delta < 1$, hence take ϵ such that $(1+\epsilon)\delta < 1$, and we have

$$\begin{aligned} B(n) &= \sum_{l \geq L+1} \mathbb{P}[\phi^{(j)}(-n) = l]\mathbb{P}\left[\sum_{k=1}^{l} \sigma_k^{(j)}(-n) > x_n\right] \\ &\leq \sum_{l \geq L+1} \delta^l(1-\delta)K(\epsilon)(1+\epsilon)^l \mathbb{P}[\sigma^{(j)} > x_n] \\ &\leq (1-\delta)K(\epsilon)\mathbb{P}[\sigma^{(j)} > x_n]\frac{((1+\epsilon)\delta)^{L+1}}{1-(1+\epsilon)\delta}. \end{aligned}$$

Then, we have

$$\sum_{n \geq N_x} A(n) \leq \sum_{n \geq N_x} \mathbb{P}\left[Z > x, K_n^j, Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L\right] + \epsilon(x, L)\overline{F}^s(x).$$

Since $\tilde{K}_{n,x}^j$ satisfies assumptions of Proposition 14, we have

$$\begin{aligned} \mathbb{P}(Z > \tilde{x}) &= (1+\epsilon(\tilde{x}))\sum_{j=1}^{K}\sum_{n \geq N_{\tilde{x}}} \mathbb{P}(Z > \tilde{x}, \tilde{K}_{n,x}^j, Y_{-n}^{(j)} > \tilde{x}_n) \\ &\leq (1+\epsilon(x))\sum_{j=1}^{K}\sum_{n \geq N_x} \mathbb{P}(Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n) \\ &\leq (1+\epsilon(x))\sum_{j=1}^{K}\sum_{n \geq N_x} \mathbb{P}(Z_{[-n,0]} > x, K_n^j, Y_{-n}^{(j)} > x_n, \phi^{(j)}(-n) \leq L) \\ &\quad + \epsilon(x, L)\overline{F}^s(x). \end{aligned}$$

Hence, we have showed with the notation of the lemma

$$\begin{aligned} (1+\epsilon(x))G(x) &\leq \mathbb{P}(Z > x) \\ \mathbb{P}(Z > x + z_x) &\leq (1+\epsilon(x))G(x) + \epsilon(x, L)\overline{F}^s(x). \end{aligned}$$

From these inequalities, we directly derive inequality (4.22). If G is long tailed, we can choose $z_x \to \infty$ such that $G(x+z_x) \sim G(x)$ and $\overline{F}^s(x+z_x) \sim \overline{F}^s(x)$, and the last statement of the corollary follows. □

4.4 Tails in GPS Queues with Subexponential Service Time Distributions

In this section we look at the impact of priority and scheduling mechanisms on long-tailed traffic phenomena. The importance of scheduling in the presence of heavy tails was first recognized by Anantharam in [4]. The present section specifically examines the effectiveness of Generalized Processor Sharing. The framework is quit similar to the work of Borst, Boxma and Jelenković [21]. We will see how our results complete theirs.

The GPS system does not fit exactly in the framework of [14]. If we consider the global workload (which is a $G/G/1$ server queue), we have a monotone separable network, but the service times are not i.i.d. and we are unable to apply directly the results of [14]. On the other hand, if we consider only one queue in isolation, this is not anymore a monotone separable network (the condition of homogeneity fails and the upper bound of section 5.3.2 is not anymore available). Hence we chose to adapt the argument to our framework. For this specific case, it will allow us to remove the assumption on the subexponentiality of F and to extend Veraverbeke's theorem to a more general setting.

4.4.1 Stochastic Assumptions

The framework is the same as in Sections 2.4 and 3.2 but we assume in addition that each arrival process is a renewal process. We recall briefly the notations. We consider a GPS system constituted of N GI/GI inputs denoted $\{T_n^{A,j}, \sigma_n^j\}_{n\in\mathbb{Z}}$ for $j \in \{1,\dots,N\}$. The weight of input j is ϕ^j and the process $\{T_n^A, \sigma_n, c_n\}$ is the superposition of the processes $\{T_n^{A,j}, \sigma_n^j, j\}$. Let $\mathbb{E}[T_1^A - T_0^A] = \lambda^{-1}$ and $\rho = \lambda \mathbb{E}^0_{T^A}[\sigma_0]$ be the traffic intensity. For each renewal process, we denote

$$\mathbb{E}[\tau_0^j] = \mathbb{E}[T_1^{A,j} - T_0^{A,j}] = \frac{1}{\lambda^j} < \infty,$$

and $\rho^j = \lambda^j \mathbb{E}^0_{T^{A,j}}[\sigma_0]$. We have $\rho = \sum_{j=1}^N \rho^j$. We assume moreover that for any $i \neq j$, we have

$$\rho^i \neq R(j)\phi^i,$$

we recall that this is always true in the case $i = j$ since $\rho < 1$ see (3.39).

The assumptions concerning service times are the following : there exists a distribution function F on $\mathbb{R}_+$ such that :

1. F has finite first moment M.
2. The integrated distribution F^s is subexponential.
3. The following equivalence holds when x tends to ∞ :

   $$\mathbb{P}(\sigma_0^j > x) \sim d^j \overline{F}(x),$$

   for all $j = 1,\dots,N$ with $\sum_{j=1}^N d^j > 0$.

The notation must be understood as follows in the Palm setting,

$$\mathbb{P}(\sigma_0^j > x) = \mathbb{P}^0_{T^{A,j}}(\sigma_0 > x) = \mathbb{P}^0_{T^A}(\sigma_0 > x | c_0 = j).$$

Remark 15. We did NOT assume that F is subexponential.

We take the notation of Section 3.2.2 to define the following domains :

$$\Delta^{i,\{j\}}(x) \quad = \quad \left\{ (\sigma,t) \in \mathbb{R}_+^2,\ w^{i,\{j\}}(\sigma,t) > x \right\}.$$

We are now able to state the main result

Theorem 10. *Consider a stable GPS system of N queues satisfying previous conditions. Let W^i be the stationary workload of queue i. When $x \to \infty$,*

$$\mathbb{P}(W^i > x) = \sum_{j=1}^{N} \lambda^j \int\int_{\{(\sigma,t)\in\Delta^{i,\{j\}}(x)\}} \mathbb{P}\left(\sigma^j \in d\sigma\right) dt + o\left(\overline{F}^s(x)\right). \tag{4.27}$$

In what follows we give an explicit computation of the integral on the right-hand side of (4.27). We give here some explicit cases

1. *if $d^i > 0$, then we can replace the equality by an equivalence and delete the $o\left(\overline{F}^s(x)\right)$ term in the right-hand term of (4.27) ; moreover, if $\overline{F}^s \in \mathcal{R}(-\infty)$, then we have*

$$\frac{\mathbb{P}(W^i > x)}{\overline{F}^s(x)} \to \frac{\lambda^i d^i}{\lambda(\phi^i R(i) - \rho^i)}.$$

2. *if $d^i = 0$ and $\overline{F}^s \in \mathcal{R}(-\infty)$ or $\frac{\rho^i}{\phi^i} < min_{j\neq i} R(j)$, then we have $\mathbb{P}(W^i > x) = o\left(\overline{F}^s(x)\right)$.*

Remark 16. Note that in case 3, we do not have the exact asymptotics. We will come back to this and discuss relations with existing results in the literature in Section 4.4.4.

4.4.2 Big Event Theorem

We first construct an upper bound for W. We consider N virtual $GI/GI/1$ queues with respective input process $\{T_n^{A,j}, \sigma_n^j\}_{n\in\mathbb{Z}}$ and with server capacity $\tilde{r}^j = \rho^j + \frac{1-\rho}{N}$. We denote by $\tilde{W}^j$ the workload at time 0 of these single server queues and $\tilde{W} = \tilde{W}^1 + \cdots + \tilde{W}^N$. More formally, we define

$$\xi_n^j = \sigma_n^j - \tilde{r}^j \tau_n^j, \quad S_{-n}^j = \sum_{i=-n}^{0} \xi_i^j, \quad M^j = \sup_{n\geq 0} S_{-n}^j.$$

With these definitions, we have

$$\tilde{W}^j = \left(M^j + \tilde{r}^j T_0^{A,j}\right)^+.$$

Thanks to Veraverbeke's theorem, we have

$$\mathbb{P}(\tilde{W}^j > x) \sim \frac{N\lambda^j d^j}{1-\rho} \overline{F}^s(x).$$

Moreover the random variables $\tilde{W}^j$ are independent of each other, hence we have

$$\mathbb{P}(\tilde{W} > x) \sim \sum_{j=1}^{N} \mathbb{P}(\tilde{W}^j > x) \sim \sum_{j=1}^{N} \frac{N\lambda^j d^j}{1-\rho} \overline{F}^s(x). \tag{4.28}$$

The following corollary follows the line of Corollary 5 of [14].

Corollary 2. *For any x and $j = 1, \ldots, N$, let $\{K_{n,x}^j\}$ be a sequence of events such that*

1. *for any n, the event $K_{n,x}^j$ and the random variables (σ_{-n}, c_{-n}) are independent ;*

2. $\inf_{n \geq N_x} \mathbb{P}\left(K^j_{n,x}\right) \to 1$ *as* $x \to \infty$.

For any sequence $\eta_n \to 0$, *let*

$$\begin{aligned} A^j_{n,x} &= K^j_{n,x} \cap \left\{ \sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\ c_{-n} = j \right\} \\ A_x &= \bigcup_{j=1}^{N} \bigcup_{n \geq N_x} A^j_{n,x}. \end{aligned}$$

Then as $x \to \infty$,

$$\mathbb{P}(\tilde{W} > x) \sim \mathbb{P}(\tilde{W} > x, A_x) \sim \mathbb{P}(A_x) \sim \sum_{j=1}^{N} \sum_{n \geq N_x} \mathbb{P}\left(A^j_{n,x}\right). \qquad (4.29)$$

Proof.
The proof follows the one of Corollary 5 of [14]. First note that

$$\begin{aligned} \sum_{j=1}^{N} \sum_{n \geq N_x} \mathbb{P}\left(A^j_{n,x}\right) &= \sum_{j=1}^{N} \sum_{n \geq N_x} \mathbb{P}\left(K^j_{n,x}\right) \mathbb{P}\left(\sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\ c_{-n} = j\right) \\ &\sim \sum_{j=1}^{N} \sum_{n \geq N_x} \frac{\lambda^j}{\lambda} \mathbb{P}\left(\sigma^j_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right)\right) \\ &\sim \sum_{j=1}^{N} \frac{\lambda^j}{\lambda} \frac{N\lambda}{1-\rho} d^j \overline{F}^s(x) = \sum_{j=1}^{N} \frac{N\lambda^j d^j}{1-\rho} \overline{F}^s(x). \end{aligned}$$

Thus, if the sequences $\{K_{n,x}\}$ and $\{\eta_n\}$ are such that, for all sufficiently large x,

1. the events $A^j_{n,x}$ are disjoint for all $n \geq N_x$;
2. $A^j_{n,x} \subset \{\tilde{W} > x\}$ for all $n \geq N_x$;

then

$$\begin{aligned} \mathbb{P}(\tilde{W} > x) &\geq \mathbb{P}(\tilde{W} > x, A_x) = \mathbb{P}(A_x) \\ &= \sum_{j=1}^{N} \sum_{n \geq N_x} \mathbb{P}\left(A^j_{n,x}\right) \sim \sum_{j=1}^{N} \frac{N\lambda^j d^j}{1-\rho} \overline{F}^s(x). \end{aligned}$$

Combining with (4.28), we get the equivalence (4.29).

We now construct two specific sequences $\{K^j_{n,x}\}$ and $\{\eta_n\}$ satisfying points 1 and 2 above and the conditions of the corollary.

We define the following function

$$C^j(n) = \sum_{k=-n}^{0} \mathbb{1}_{\{c_k = j\}} - 1.$$

On the event $\{c_{-n} = j\}$, we have $T^A_{-n} = T^{A,j}_{-C^j(n)}$, $\sigma_{-n} = \sigma^j_{-C^j(n)}$. We can find a non-increasing sequence $\epsilon_n \to 0$ such that $n\epsilon_n \to \infty$ and such that the probabilities of the following events tend to 1 as $n \to \infty$,

$$\begin{aligned} L_{n,x} &= \left\{ \left| \frac{S^j_{-k}}{k} - \frac{\rho-1}{N\lambda^j} \right| \le \epsilon_k,\ N_x \le k \le C^j(n-1),\ 1 \le j \le N \right\}, \\ M^j_n &= \left\{ \left| \frac{C^j(n-1)}{n} - \frac{\lambda^j}{\lambda} \right| \le \epsilon_n \right\}, \\ N^j_n &= \left\{ \left| T_0^{A,j} \right| \le \frac{n\epsilon_n}{\tilde{r}^j} \right\}. \end{aligned}$$

Hence the event $K^j_{n,x} = L_{n,x} \cap M^j_n \cap N^j_n$ satisfy the conditions of the corollary. Moreover on the event $\{c_{-n} = j\}$, we have $C^j(n) = C^j(n-1) + 1$ and,

$$S^j_{-C^j(n)} = \sigma_{-n} + S^j_{-C^j(n-1)}.$$

Now if we take $\eta_n = \sqrt{\epsilon_n}$, we have

$$\begin{aligned} \tilde{W} &\ge S^j_{-C^j(n)} - n\epsilon_n \\ &> x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right) + n\left(\frac{\lambda^j}{\lambda} - \epsilon_n\right)\left(\frac{\rho-1}{N\lambda^j} - \epsilon_{(\lambda^j/\lambda)n-1}\right) - n\epsilon_n, \end{aligned}$$

and we see that for sufficiently large n, we have $\tilde{W} > x$. The fact that the event $A^j_{n,x}$ are disjoint follows from the fact that for sufficiently large x, we have $\epsilon_{N_x} \le (1-\rho)/(N\lambda^j)$. Indeed on the event $A^j_{n,x}$, we have $S^j_{-C^j(n)} > x$ and $S^j_{-C^j(n)+1} \le (C^j(n)-1)((\rho-1)/(N\lambda^j) + \epsilon_{N_x}) \le 0$. The event $\{S^j_n > x\} \cup \{S^j_{n-1} \le 0\}$ are clearly disjoint in n. With the same kind of argument, we see that the events $A^j_{n,x}$ are disjoint in j. The end of the proof, i.e. showing that the corollary is true for any sequence $K^j_{n,x}$ is exactly the same as in the proof of Corollary 5 of [14] and is skipped. □

From this corollary we derive the following proposition

Proposition 16. *For any x and $j = 1, \dots, N$, let $\{K^j_{n,x}\}$ be a sequence of events such that*

1. *for any n, the event $K^j_{n,x}$ and and the random variables (σ_{-n}, c_{-n}) are independent ;*
2. $\inf_{n \ge N_x} \mathbb{P}\left(K^j_{n,x}\right) \to 1$ *as* $x \to \infty$.

For any sequence $\eta_n \to 0$, let

$$\begin{aligned} A^j_{n,x} &= K^j_{n,x} \cap \left\{ \sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\ c_{-n} = j \right\} \\ A_x &= \bigcup_{j=1}^{N} \bigcup_{n \ge N_x} A^j_{n,x}. \end{aligned}$$

Then for any random variable $W \le \tilde{W}$, we have as $x \to \infty$,

$$\mathbb{P}(W > x) = \mathbb{P}(W > x, A_x) + o\left(\overline{F}^s(x)\right) \tag{4.30}$$

$$= \sum_{j=1}^{N} \sum_{n \ge N_x} \mathbb{P}\left(W > x, A^j_{n,x}\right) + o\left(\overline{F}^s(x)\right). \tag{4.31}$$

Proof.
We have

$$\begin{aligned}\mathbb{P}(W > x) &= \mathbb{P}(W > x, A_x) + \mathbb{P}(W > x, A_x^c) \\ &\le \mathbb{P}(W > x, A_x) + \mathbb{P}(\tilde{W} > x, A_x^c),\end{aligned}$$

but thanks to previous corollary we have that $\mathbb{P}(\tilde{W} > x, A_x^c) = o\left(\overline{F}^s(x)\right)$. Hence we have

$$\mathbb{P}(W > x, A_x) \le \mathbb{P}(W > x) \le \mathbb{P}(W > x, A_x) + o\left(\overline{F}^s(x)\right),$$

which gives (4.30). The end of the proof is a repetition of the proof of last corollary and is skipped.
□

At this stage we are able to prove the following proposition which extends Veraverbeke's theorem to a more general setting.

Proposition 17. *Let W be the stationary workload of a single server queue fed by the superposition of N independent GI/GI processes. Assume moreover that*

$$\mathbb{P}^0_{T^A}(\sigma_0 > x) = \overline{F}(x),$$

and that F^s is subexponential. Then we have

$$\mathbb{P}(W > x) \sim \frac{\lambda}{1-\rho}\overline{F}^s(x).$$

Proof.
First note that $\tilde{W} \ge W$. Hence we can apply previous proposition, with

$$K^j_{n,x} = \left\{ \left| \frac{S_{-k}}{k} - \frac{\rho - 1}{\lambda} \right| \le \epsilon_k,\ N_x \le k \le n-1, \left|T_0^A\right| \le n\epsilon_n \right\},$$

where $S_{-k} = \sum_{i=-k}^{0} \sigma_i - \tau_i$. On the event $A^j_{n,x}$, we have $W = \sigma_{-n} + S_{-n+1} + T_0^A$, hence we have

$$\begin{aligned}\sum_{j=1}^{N}\sum_{n\ge N_x} \mathbb{P}(W > x, A^j_{n,x}) &\sim \sum_{j=1}^{N}\sum_{n\ge N_x} \mathbb{P}\left(\sigma_{-n} > x + n\left(\frac{1-\rho}{\lambda} + 2\epsilon_n\right), c_{-n} = j\right) \\ &\sim \frac{\lambda}{1-\rho}\overline{F}^s(x).\end{aligned}$$

□

Remark 17. This result extends Theorem 4.1 of Asmussen, Schmidli and Schmidt [8], in which the arrival process is the superposition of renewal processes but the service times are supposed to be i.i.d.

4.4.3 Computation of the Exact Asymptotics

We have to find a sequence of events $\{K^j_{n,x}\}$ in order to compute the following sum

$$S^{i,\{j\}} = \sum_{n\ge N_x} \mathbb{P}\left(W^i > x, K^j_{n,x}, \sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\ c_{-n} = j\right)$$

A first case is easy : when queue i remains stable even if queue j is continuously backlogged.

Lemma 24. *Assume that*

$$\frac{\rho^i}{\phi^i} < R(j). \tag{4.32}$$

Then we have

$$S^{i,\{j\}} = o\left(\overline{F}^s(x)\right).$$

Proof.

Under condition (4.32), we know thanks to Proposition 11, we know that the stationary workload of queue i exists when queue j is continuously backlogged. We denote $W^i(j)$ this workload. We have

$$W^i \leq W^i(j) < \infty,$$

and $W^i(j)$ is clearly independent of $(T_n^{A,j}, \sigma_n^j)$. Hence we have

$$\begin{aligned} S^{i,\{j\}} &= \sum_{n\geq N_x} \mathbb{P}\left(W^i > x, K^j_{n,x}, \sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\, c_{-n} = j\right) \\ &\leq \mathbb{P}(W^i(j) > x) \sum_{n\geq N_x} \mathbb{P}\left(\sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda} + \eta_n\right),\, c_{-n} = j\right) \\ &= o\left(\overline{F}^s(x)\right). \end{aligned}$$

□

We consider now the case

$$\frac{\rho^i}{\phi^i} > R(j).$$

In this case when queue j experiences a long backlog (due to a very big service time), queue i is no longer stable and the fluid limit corresponding to this queue is no longer 0. The remaining steps of the proof of Theorem 10 are similar to those of section 4.3.6.

Let ϵ_n be some sequence of positive real numbers, we define

$$K_n^j = \left\{ \sup_{\substack{\sigma > n\frac{1-\rho}{N\lambda} \\ t\leq 2a}} \left| \frac{W^{\{j\}}(\sigma, nt) - w^{\{j\}}(\sigma, nt)}{n} \right| \leq \epsilon_n,\ \left| \frac{T_n^A}{n} - a \right| \leq \epsilon_n \right\} \circ \theta_{T^A_{-n}}.$$

Thanks to the results of Section 3.2.2, we have the following lemma

Lemma 25. *Let $\{K_n^j\}$ be the sequence of events defined above. K_n^j and the random variables σ_{-n} and c_{-n} are independent. There exists a sequence $\epsilon_n \to 0$ such that we have $\mathbb{P}(K_n^j) \to 1$ as $n \to \infty$.*

On the event $K_n^j \cap \left\{\sigma_{-n} > x + n\left(\frac{1-\rho}{N\lambda}\right),\, c_{-n} = j\right\}$, we have (thanks to the continuity of $w^{i,\{j\}}$),

$$W^i = w^{i,\{j\}}(\sigma_{-n}, na) + n\eta_n, \quad \text{with } \eta_n \text{ a r.v. such that } |\eta_n| \leq \epsilon_n.$$

We will need the following lemma on the shape of the domain $\Delta^{i,\{j\}}(x)$.

Lemma 26. *There exist constants* $\{\alpha_k^{i,\{j\}}, \beta_k^{i,\{j\}}, \gamma_k^{i,\{j\}}\}_{0\le k\le \ell}$ *with* $\alpha_0^{i,\{j\}} < \alpha_1^{i,\{j\}} < \cdots < \alpha_\ell^{i,\{j\}}$, $\beta_k^{i,\{j\}} \le 1$*, such that*

$$\Delta^{i,\{j\}}(x) = \bigcup_{k=0}^{\ell} \left\{ \alpha_k^{i,\{j\}} x \le t < \alpha_{k+1}^{i,\{j\}},\ \sigma > \frac{x}{\beta_k^{i,\{j\}}} + t\gamma_k^{i,\{j\}} \right\},$$

with $\alpha_{\ell+1}^{i,\{j\}} = +\infty$*. Moreover, we have*

$$\Delta^{i,\{i\}}(x) = \left\{\sigma > x + \left(\phi^i R(i) - \rho^i\right)\lambda t\right\}.$$

This lemma follows directly from the definition of the function $w^{i,\{j\}}$. We will give some example, to show how to compute the constants of the lemma, which in turn will lead to the computation of the integral in Theorem 10.

Example 4. *1.* **Case** $N = 2$

We suppose that

$$\frac{\rho^1}{\phi^1} < 1 < \frac{\rho^2}{\phi^2}.$$

The first inequality is imposed by the stability condition $\rho < 1$ *and we suppose the second one in order to be in the following interesting case : a big service in queue* 1 *induces an instability of queue* 2*. The corresponding fluid limit* $w^{\{1\}}$ *is depicted on Figure 4.4.*

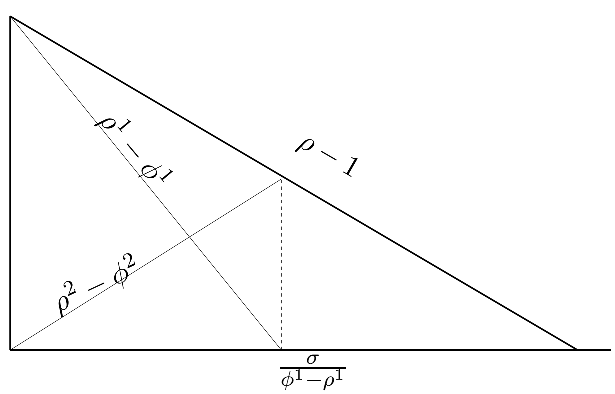


FIG. 4.4 – GPS with two classes : big service in class 1

The corresponding domains are easy to compute and given in Figure 4.5,

$$\begin{aligned}
\Delta^{1,\{1\}}(x) &= \left\{(\sigma,t),\ \sigma > x + (\phi^1 - \rho^1)t\right\},\\
\Delta^{2,\{1\}}(x) &= \left\{(\sigma,t),\ t > \frac{x}{\rho^2-\phi^2},\ \sigma > x + (1-\rho)t\right\}.
\end{aligned}$$

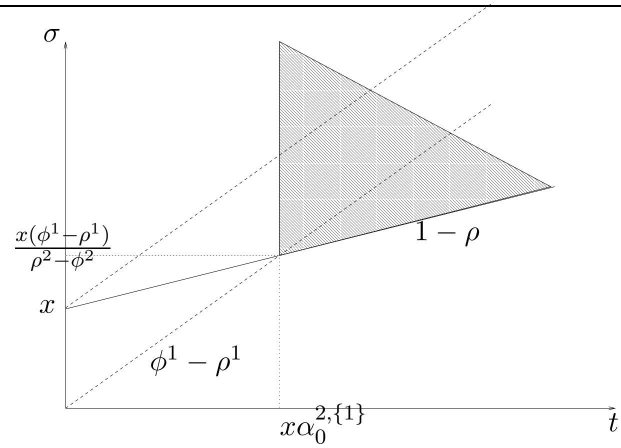


FIG. 4.5 – GPS with two classes : the domain $\Delta^{2,\{1\}}(x)$

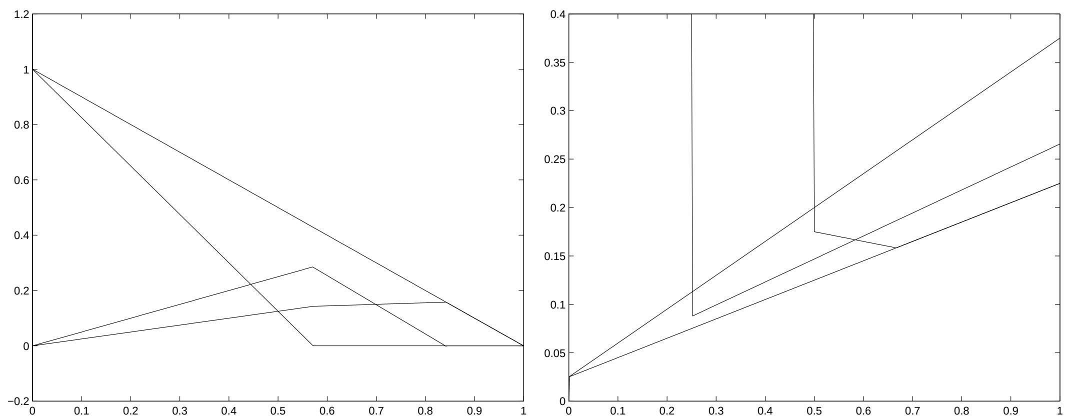


FIG. 4.6 – The case $N = 3$

2. **case** $N = 3$

   In previous case, once we fixed the traffic intensities ρ^1 and ρ^2, the behavior of the system depends only on one parameter, the ratio ϕ^1/ϕ^2. In the case $N = 3$, there are many different possible cases. We draw with matlab the fluid limits (in the case of a big service in queue 1) and the corresponding domains for parameters such that

$$\frac{\rho^1}{\phi^1} < 1 < \frac{\rho^2}{\phi^2} < \frac{\rho^3}{\phi^3},$$

   namely, the parameters are the following :

$$\rho^1 = 0.2,\ \phi^1 = 0.55, \quad \rho^2 = 0.5,\ \phi^2 = 0.4, \quad \rho^3 = 0.1,\ \phi^3 = 0.05.$$

We return now to the proof of Theorem 10. Following exactly the steps of the demonstration of Theorem 9, we can show that

$$
\begin{aligned}
S^{i,\{j\}}(x) &\sim \sum_{k=0}^{\ell} \sum_{\alpha_k^{i,\{j\}} x \leq na < \alpha_{k+1}^{i,\{j\}} x} \mathbb{P}\left(\sigma_{-n} > \frac{x}{\beta_k^{i\{j\}}} + na\gamma_k^{i,\{j\}},\ c_{-n} = j\right) \\
&= \frac{\lambda^j}{\lambda a} \sum_{k=0}^{\ell} \sum_{\alpha_k^{i,\{j\}} x \leq n < \alpha_{k+1}^{i,\{j\}} x} \mathbb{P}\left(\sigma^j > \frac{x}{\beta_k^{i\{j\}}} + n\gamma_k^{i,\{j\}}\right) \\
&= \lambda^j \sum_{k=0}^{\ell} \sum_{\alpha_k^{i,\{j\}} x \leq n < \alpha_{k+1}^{i,\{j\}} x} \mathbb{P}\left(\sigma^j > \frac{x}{\beta_k^{i,\{j\}}} + n\gamma_k^{i,\{j\}}\right).
\end{aligned}
$$

This term is of order $d^j\overline{F}^s(x/\beta_0^{i,\{j\}})$ and hence $o(\overline{F}^s(x))$ as soon as F^s is rapidly varying. Summing over j, we obtain the equality (4.27) of the Theorem, which concludes the proof.

4.4.4 Some Extensions

In the case $d^i = 0$, there are some cases in which Theorem 10 does not give the exact asymptotic. The following case is not covered by previous Theorem and follows easily from the same kind of argument,

Proposition 18. *Suppose that $\rho^i < \phi^i$ and that $\mathbb{P}(\sigma_1^i > x) = F_i(x)$ is such that F_i^s is subexponential, then we have as $x \to \infty$,*

$$
\begin{aligned}
\mathbb{P}(W^i > x) &\sim \sum_{j=1}^{N} \lambda^j \int\int_{\{(\sigma,t)\in\Delta^{i,\{j\}}(x)\}} \mathbb{P}\left(\sigma^j \in d\sigma\right) dt \\
&\sim \frac{\lambda^i}{\lambda(\phi^i R(i) - \rho^i)} \overline{F}_i^s(x).
\end{aligned}
$$

Proof.

The stationary workload of the $GI/GI/1$ queue with input process $\{T_n^{A,i}, \sigma_n^i\}_{n\in\mathbb{Z}}$ and service rate ϕ^i is clearly a stable upper-bound for W^i. The proposition follows from exactly the same arguments as above. □

In general we are unable to give the exact asymptotics of queue i if the heaviest class say k does not contribute to it, i.e. if queue i remains stable even if the heaviest class is backlogged. In this case, our upper bound is quite rough and the workload of queue i when k is backlogged (namely $W^i(k)$ with our notation) is clearly a better upper bound. Moreover this upper bound belongs to the monotone separable framework but (except if the arrival processes are Poisson point processes) the $G/G/1$ upper bound used in the proof of Theorem 7 is not a $GI/GI/1$ single server queue.

One can consider a stable feed-forward network of flows where each node has a GPS discipline. The same kind of techniques as we did for the single server queue apply. Indeed the only non-trivial thing to find is an upper-bound. To each flow we associate a system of queues in tandem (and we chose the rate such that each system is stable as for $\tilde{W}$). Hence we associate to the original network, a network of queues in tandem with a fork at the beginning and a join at the end. The maximal dater of this virtual network will be an upper-bound for the maximal dater of the original one (because the virtual network is not work-conserving). Moreover this network belongs to the (max,plus)-class studied in the second section of this chapter, hence we have the exact asymptotics of its maximal dater and the rest will follow like in the single server case. This is a work in progress.

We compare now our results with the results of Borst, Boxma and Jelenković in [21].

Our theorem deals with instantaneous input but using the same kind of arguments as Jelenković and Lazar in [62], these results should extend to fluid input with on-period that are regularly varying but with some conditions on the rate during on-periods.

In [21], the authors deal with a possibly unstable GPS system and derive the tail asymptotic of the stable queues. Our approach does not cover this case.

Theorem 3.1 of [21] deals with the case $\rho^i < \phi^i$ which correspond exactly to previous proposition. If we have $d^i > 0$, Theorem 4.1 and 5.1 are somehow extended by our results. Indeed in [21], the authors impose some restrictions on the parameters of the system so that the sum in (4.27) contains each time only one term, and they give the asymptotic of this term.

Note that in [22], Borst, Mandjes and Van Uitert study a GPS system with 2 classes. One is light-tailed and the other is heavy-tailed. Moreover the light-tailed class is still stable when the other queue is backlogged. They show that in these conditions, a large workload in the light-tailed class is due to a large service in the other queue and a change of drift in the light-tailed class.

4.5 Towards an Extension of the Single Big Event Theorem

The single big event theorem is always efficient to obtain asymptotics of first order state variables, like the maximal dater. In some cases, it also gives the asymptotics of second order state variables like the workload at each node. For example, in the case of the tandem queues, it allows to get the exact asymptotics of the delay at the second queue, this is done in [14]. This result extends to more general (max,plus)-linear networks (this is a work in progress with Ton Dieker). With the results we obtained for the generalized Jackson networks, it seems that the computation of the asymptotics of the stationary workload of an individual queue in the network is doable. One has to compute carefully the corresponding fluid limit.

But in some cases, either one big jump is not sufficient to get a local instability or the one big jump scenario has to be compared with other scenario, to see if it is still the most probable scenario (like in the GPS case). Typically, this situation arises when the fluid limit is zero.

Consider the following framework : we have independent sequences of i.i.d. random variables $\{X^1, X^1_i\}$, $\{X^2, X^2_i\}$ and $\{X^3, X^3_i\}$. We take the following notation

$$\begin{aligned} S^1_n &= \sum_{i=1}^n F(X^1_i, X^3_i), \\ S^2_n &= \sum_{i=1}^n G(X^2_i, X^3_i), \\ W^1 &= \sup_{n\geq 0} S^1_n, \\ W^2 &= \sup_{n\geq 0} S^2_n, \end{aligned}$$

where the function F and G are deterministic an such that

$$\begin{aligned}\mathbb{E}\left[F(X^1, X^3)\right] &= -a^1 < 0,\\ \mathbb{E}\left[G(X^2, X^3)\right] &= -a^2 < 0.\end{aligned}$$

We are interested in the tail of a random variable Z such that

$$Z \leq \min(W^1, W^2).$$

Since $-a^1, -a^2 < 0$, the right hand side is finite and the interpretation is that Z can not be big through only one big X^1 or one big X^2.

If one can define a typical event for the random variable $\min(W^1, W^2)$, then the same kind of techniques could apply. Hence the first result to obtain is the tail asymptotics of $\min(W^1, W^2)$. It seems that the method given by Zachary [85] to prove Veraverbeke's Theorem could extend to this kind of framework. If so this would give the whole picture of the asymptotics of GPS queues. Namely in the GPS framework presented above, we should be able to prove the following property.

We assume that each input process is M/GI with the following assumption on the service times

$$\mathbb{P}(\sigma_1^j > x) = \overline{F}_j(x),$$

such that $\overline{F}_j^s \in \mathcal{R}(-\alpha_j)$, with $0 < \alpha_j \leq \infty$. We denote

$$E(i) = \left\{D \subset \{1, \dots, N\},\ \rho^i > R(D)\phi^i\right\}.$$

For any set $D \in E(i)$, the queue i is unstable if all queues in D are backlogged. We denote

$$\begin{aligned}\alpha(i) &= \min_{\{j\} \in E(i)} \alpha_j,\\ \beta(i) &= \min_{\substack{D \in E(i)\\ |D| \geq 2}} \sum_{\ell \in D} \alpha_\ell.\end{aligned}$$

The following property might be correct !

Proposition 19. *If $\alpha(i) < \beta(i) < \infty$, then we have as $x \to \infty$,*

$$\mathbb{P}(W^i > x) \sim \sum_{j=1}^{N} \lambda^j \int\int_{\{(\sigma,t) \in \Delta^{i,\{j\}}(x)\}} \mathbb{P}\left(\sigma^j \in d\sigma\right) dt$$

Chapitre 5

Large Deviations for Monotone Separable Networks

5.1 Introduction

In this chapter, we consider a monotone separable network as described in Section 2.1. We are interested in large deviation results for such queueing system in equilibrium. Equilibrium systems have generally been treated on a case-by-case basis. For a general overview of applications of large deviations theory to queueing problems we refer to the book of Ganesh O'Connell and Wischik [46].

The case of the single server queue has received extensive attention in the literature. See for example the work of Glynn and Whitt in [50] or Duffield and O'Connell in [34] which gives results in a very general framework.

The extension of these ideas to networks appears to be rather challenging problem. Ganesh and Anantharam in [45] derive large deviations results for two queues in tandem, with renewal arrivals and exponential service times. In [31] De Veciana, Courcoubetis and Walrand characterize the departure process from a single $G/D/1$ queue in the large deviation regime. They show that there is a region over which the large deviation rate functions for the cumulative departures and arrivals agree and bounds are given outside that region. Chang and Zajic [24] consider the case of a single arrival stream and stochastic service rate. In [76], O'Connell gives a full description of the rate function for the cumulative departures under the hypothesis that the arrival processes jointly satisfy a sample path large deviations principle with linear geodesics. Roughly speaking, this means that the most likely path to an extreme value is a straight line. A natural question is then : do the departures also satisfy this hypothesis ? If so, then one could treat quite complicated networks by successive iteration of the single-buffer results in [76]. Indeed if the service process is deterministic, then the departure process has linear geodesics. So a recursive analysis of networks of such queues is possible as in [23]. Even if the service process is stochastic, it is shown in [47] that conditional on the departure rate from a queue exceeding its mean, the departure process has linear geodesics. We are typically interested in the probability of queue lengths exceeding some large threshold and in well-designed networks this requires departure rates exceeding their mean. Therefore, we have linear geodesics in the region of interest and so the study of networks of queues using a recursive approach is again feasible. With such an approach Bertsimas, Paschalidis and Tsitsiklis compute in [18] the decay rate of the stationary waiting time and queue length distributions at each node in an acyclic network in the context of quite general arrival and service processes.

To the best of our knowledge, the large deviations analysis of queueing systems with any kind of feedback is restricted to some specific cases. In [5] Anantharam, Heidelberger and Tsoucas use quasi-reversibility arguments (see Kelly [64]) to study rare event in the case of Jackson networks. Using a different approach Ignatiouk-Robert derives the rate function for sample path large deviation of such networks in [59]. In this section we choose a different approach inspired from works of Ganesh [44] and Toomey [83].

In this chapter, we study large deviations asymptotics of the form

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(Z > x) = -\theta^*, \tag{5.1}$$

where the random variable Z corresponds to a "global" state variable of a random process. We only deal with exponentially decaying distributions.

This chapter is made of four parts, each of them building up on the results of the previous one. Here is a brief overview :

- In Section 5.2, we derive tail asymptotics of the form (5.1) where Z is the global maxima of an independent subadditive process. In particular, we show that the associated θ^* is positive and give an explicit way of computing its value.
- In Section 5.3, we derive the tail asymptotics (5.1) where Z corresponds to the "time to empty" a queueing network in its stationary regime. This definition will be made precise in the framework of monotone-separable networks.
- In Section 5.4 , we concentrate on a sub-class of the monotone-separable networks, namely the (max,plus)-linear networks. We derive for the stationary solution of a (max,plus)-linear recursion the associated θ^* in an explicit way.
- In Section 5.5, we concentrate on the case of generalized Jackson networks. Our results are partial in the sense that we obtain sample path large deviations result for the transient process and the connection with the stationary version has still to be made. However, we choose to include here these results because they are original and the general methodology could be used for other queueing networks.

Tail asymptotics for the supremum of an independent subadditive process

Let $S_n = X_1 + \cdots + X_n$ be a random walk where the sequence $\{X, X_i,\ 1 \le i\}$ is a sequence of independent identically distributed (i.i.d) random variables, whit $\mathbb{E}[X] < 0$. Define $M := \sup_{n\ge1} S_n < \infty$ a.s. Then we have

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M > x) = -\theta^*, \quad \text{where } \theta^* = \sup\left\{\theta > 0,\ \log \mathbb{E}\left[e^{\theta X}\right] < 0\right\}, \tag{5.2}$$

with the convention that the supremum of the empty set is $-\infty$. Note that this case has been extensively studied in the literature and much finer estimates are available, see the complementary works of Iglehart [58] and Pakes [77].

In the first part, we extend this result by considering instead of the additive process S_n, a subadditive process $Y_{[1,n]}$. Our main result is that the tail asymptotics (5.2) remains valid when one replace the logarithmic moment generating function of the X_i's by the properly scaled logarithmic moment generating function of the process $Y_{[1,n]}$. In particular, all the information needed to establish (5.2) is contained in the scaled logarithmic moment generating function. We do not require any large deviation principle for the process $\{\frac{Y_{[1,n]}}{n}\}$. This is a surprising fact in the field of large deviations for subadditive processes. Indeed studying the random variable M is much simpler than trying to get a large deviations principle for the process $\{Y_{[1,n]}/n\}$ (which remains an open question in the independent subadditive case) and we give an example of two subadditive processes with the same scaled logarithmic moment generating function but satisfying large deviation principle with different rate functions.

Large deviations for monotone separable networks

Literature on large deviations of queueing networks with feedback is rare and confined to the setting of networks described by finite-dimensional Markov processes, see Dupuis and Ellis [36], Dupuis, Ellis and Weiss [37] and the recent work of Igniatiouk-Robert [59], [60]. Moreover, these works concentrate on local large deviations and cannot handle the large deviations of the network in its stationary regime. The large deviation asymptotics of queueing systems are difficult to analyze because they are dynamical systems with discontinuities. To the best of our knowledge, there is no rigorous result on the large deviations of non-exponential networks with feedback in their stationary regime.

We will show in the second part that the monotone-separable framework allows us to derive the tail asymptotics for ”global” variable of the stationary version of such networks. This framework was first introduced by Baccelli and Foss [13] to study the stability condition of these networks. In particular, this framework includes generalized Jackson networks, stochastic Petri Nets and polling systems. The main theorem of this second part is Theorem 13 that gives the exponential decay of the stationary maximal dater (which will be defined latter) for such networks in term of the asymptotic logarithmic moment generating function.

Case of study I : (max, plus)-linear systems

To apply our Theorem 13 we consider the sub-class of the monotone separable networks consisting of the (max,plus)-linear networks. From a queueing point of view, these networks include for example the single server queue, tandem queues, fork-join systems and the maximal dater corresponds to the end-to-end delay. Our work extends the analysis of tandem queues done by Ganesh [44].

More generally we study in the third part the stationary solution of a (max,plus)-linear recursion. Results concerning large deviations of products of random topical operators have been obtained by Toomey in [83]. In rough words, these results would correspond to large deviations of the process $Y_{[1,n]}$ (i.e. before taking the supremum). However very restrictive conditions are required on the coefficients of the matrix. Here we do not assume these requirements to be fulfilled but we show that under mild assumptions on the matrix structure, the tail behavior of $\sup_n Y_{[1,n]}$ is explicitly given and can be computed (or approximated) in practical cases.

Case of study II : generalized Jackson networks

This section is independent of the preceding ones. We derive a sample path large deviation principle for the arrival and departure processes associated to the nodes of a generalized Jackson network. In particular, we obtain an explicit rate function under rather weak stochastic assumptions (i.i.d.). Even in the exponential case, the formulation of the rate function is original and seems more explicit than the rate function derived in [59].

5.2 Tail asymptotics for the supremum of an independent subadditive process

5.2.1 Framework and main result

Assume the variables $\{\xi_n\}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P}, \theta)$, where θ is an ergodic, measure-preserving shift transformation, such that $\xi_n \circ \theta = \xi_{n+1}$. We assume that there exists a set of functions $\{g_\ell\}$, $g_\ell : \mathcal{K}^\ell \to \mathbb{R}$, such that :

$$Y_{[m,n]} = g_{n-m+1}\{\xi_\ell,\ m \leq \ell \leq n\}, \tag{5.3}$$

for all $m \leq n$. The functions g_n are deterministic and we assume that they are such that the family of random variables $Y = \{Y_{[m,n]},\ m \leq n, m, n \in \mathbb{Z}\}$ is a subadditive process, i.e. satisfies the following three conditions :

1. subadditivity : $Y_{[m,n]} \leq Y_{[m,\ell]} + Y_{[\ell+1,n]}$, for all $m \leq \ell < n$;
2. stationarity : the joint distributions of $\{Y_{[m,n]}, m \leq n\}$ are the same as the joint distributions of $\{Y_{[m+1,n+1]},\ m \leq n\}$;
3. moment condition : $\mathbb{E}[|Y_{[0,n]}|] < \infty$ for each $n \geq 0$ and $\mathbb{E}[Y_{[0,n]}] > -\alpha n$ for some $\alpha \in \mathbb{R}$ and all $n \geq 0$.

Under the foregoing ergodic assumption, there exists a constant μ such that (see Kingman [65])

$$\lim_{n\to\infty} \frac{Y_{[1,n]}}{n} = \lim_{n\to\infty} \frac{\mathbb{E}[Y_{[1,n]}]}{n} = \mu \quad a.s. \tag{5.4}$$

In what follows, we will make the following assumptions :

(A1) the constant μ defined in (5.4) is negative ;

(A2) the sequence $\{\xi_n\}$ is a sequence of i.i.d. random variables ;

(A3) There exists $\eta > 0$ such that, $\mathbb{E}\left[e^{\eta Y_{[1,1]}}\right] < \infty$, and for $\theta > 0$, if $\mathbb{E}\left[e^{\theta Y_{[1,1]}}\right] = \infty$, then $\mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] = \infty$ for all n.

In view of assumption **(A1)**, one can define the following random variable :

$$M := \sup_{n\geq 1} Y_{[1,n]} < \infty \quad a.s.$$

Note that the random variables $Y_{[a,b]}, Y_{[c,d]}, \ldots, Y_{[e,f]}$ are independent whenever $a \leq b < c \leq d < \cdots < e \leq f$, we say that the subadditive process Y is independent.

We know that a subadditive independent process is superconvolutive and the existence of the following moment generating function follows [55] (see Lemma 27 for a proof),

$$\Lambda(\theta) \quad = \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right],$$

Let

$$\theta^* = \sup\left\{\theta > 0,\ \Lambda(\theta) < 0\right\}, \tag{5.5}$$

where the supremum of the empty set is $-\infty$.

Theorem 11. *Under previous assumptions, we have $\theta^* > 0$ and*

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M > x) = -\theta^*.$$

Theorem 11 extends a well-known result in the case of random walks to the case of independent subadditive processes. One important point is that we do not require any large deviations principle for the process $\{Y_{[1,n]}/n\}$. The existence of the constant θ^* is ensured by the moment condition **(A3)** and its value is explicitly given by (5.5).

Example 5.

– *A first difference with the additive case is that it is possible that $\mathbb{P}(M > x) > 0$ for any $x > 0$ while $\theta^* = \infty$: consider the following subadditive process,*

$$Y_{[m,n]} = Z \prod_{i=m}^{n} X_i - (n - m + 1),$$

where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0) < 1$ and $Z \sim Normal(0, 1)$ is independent of everything else. We have clearly $\lim_{n\to\infty} Y_{[1,n]}/n = -1$ and

$$\log \left(e^{\theta Y_{[1,n]}} \right) = \log \left(p^n (e^{\theta^2/2} - 1) + 1 \right) - n\theta.$$

Hence we have $\Lambda(\theta) = -\theta < 0$, for all $\theta > 0$ and $\mathbb{P}(M > x) \geq \mathbb{P}(X_1 = 1)\mathbb{P}(Z > x + 1) > 0$ for all x.

– *Note that in the additive case, the fact that $\mathbb{E}\left[e^{\theta Y_{[1,1]}}\right] = \infty$ implies that $\mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] = \infty$ for all n. In the subadditive case, this not anymore true and Assumption* **(A3)** *is needed for Theorem 11 to hold. Consider a sequence of i.i.d. exponentially distributed (with mean 1) random variables $\{X_n\}_n$ and consider the subadditive process ($n \leq m$) :*

$$Y_{[n,m]} := X_n \mathbb{1}_{\{n=m\}} + (n - m)\mathbb{1}_{\{n<m\}}.$$

In this case, we clearly have $M = X_1$, hence $\mathbb{P}(M > x) = e^{-x}$ and

$$\forall n > 1, \quad \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] = e^{\theta(1-n)} \Rightarrow \theta^* = \infty.$$

– *Consider the case $Y_{[m,n]} = Z_{[m,n]} - S_{[m,n]}$, where the processes Z and S are independent, S is a non-negative additive process (i.e. a random walk) and Z is a subadditive process with*

$$0 \leq Z_{[1,n]} \leq Z_{[1,n+1]}.$$

Then we have for $\theta > 0$,

$$\begin{aligned} \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] &= \mathbb{E}\left[e^{\theta Z_{[1,n]}}\right] \mathbb{E}\left[e^{-\theta S_{[1,n]}}\right] \\ &\geq \mathbb{E}\left[e^{\theta Y_{[1,1]}}\right] \left(\mathbb{E}\left[e^{-\theta S_{[1,1]}}\right] \right)^{n-1}, \end{aligned}$$

and Assumption **(A3)** *is satisfied as soon as $\mathbb{E}\left[e^{\eta Y_{[1,1]}}\right] < \infty$ for some $\eta > 0$.*

To make the connection with the existing literature, we state the following result (which proof is given in Section 5.2.3.0) :

Corollary 3. *Under previous assumptions and if*

1. *the sequence* $\{Y_{[1,n]}/n\}$ *satisfies a large deviation principle (LDP) with a good rate function* I*;*
2. *there exists* $\epsilon > 0$ *such that* $\Lambda(\theta^* + \epsilon) < \infty$,

where θ^* *is defined as in (5.5). Then we have*

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M > x) = -\theta^* = -\inf_{\alpha>0} \frac{I(\alpha)}{\alpha}. \tag{5.6}$$

Without the assumption that the process $Y_{[1,n]}$ is subadditive, this kind of result has been extensively studied in the queueing literature (we refer to the work of Duffy, Lewis and Sullivan [35]). However, we see that considering the moment generating function instead of the rate function allows us to get a more general result than (5.6) since we do not require the assumption on the tail (see the example of section 5.4.2). Indeed this assumption ensures that the tail asymptotics of $\mathbb{P}(Y_{[1,n]} > nc)$ for a single n value cannot dominate those of $\mathbb{P}(M > x)$. In this case, equation (5.6) has a nice interpretation : the natural drift of the process $Y_{[1,n]}$ is μn, where $\mu < 0$. The quantity $I(\alpha)$ can be seen as the cost for changing the drift of this process to $\alpha > 0$. Now in order to reach level x, this drift has to last for a time x/α. Hence the total cost for reaching level x with drift α is $xI(\alpha)/\alpha$ and the process naturally choose the drift with the minimal associated cost. We will see how this non-rigorous heuristic can be made more precise in what follows.

5.2.2 Beyond the Gärtner-Ellis theorem

In this section, we discuss the relations between Theorem 11 and Gärtner-Ellis Theorem.

If the origin belongs to the interior of the domain $D_\Lambda = \{\theta,\ \Lambda(\theta) < \infty\}$ (which is not required here), we see that Assumption 2.3.2 of [32] is satisfied. In which case, the upper bound of the Gärtner-Ellis Theorem holds (see Theorem 2.3.6 in [32]), hence for $\alpha > 0$ we have,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{Y_{[1,n]}}{n} \geq \alpha\right) \leq -\inf_{x\geq\alpha} \Lambda^*(x) = -\Lambda^*(\alpha), \tag{5.7}$$

where $\Lambda^*(x) = \sup_{\theta\geq 0}\{\theta x - \Lambda(\theta)\}$ is the Fenchel-Legendre transform of $\Lambda(\theta)$. Note that we restrict the supremum over the set $\theta \geq 0$ and the function $x \mapsto \Lambda^*(x)$ is non-decreasing in $x > 0$ (see the following Section 5.2.3.0 for a justification).

We give now an example of a subadditive independent process for which the upper bound (5.7) given by Gärtner-Ellis Theorem is not tight.

Consider the following independent sequences $\{\sigma_i^1\}$ and $\{\sigma_i^2\}$ of i.i.d. random variables :

$$\mathbb{P}(\sigma_i^1 = 1) = 1 - p, \qquad \mathbb{P}(\sigma_i^1 = 2) = p,$$
$$\mathbb{P}(\sigma_i^2 = 0) = 1 - p, \qquad \mathbb{P}(\sigma_i^1 = 3) = p.$$

For $\ell = 1, 2$ and $u \leq v$, we denote $S^\ell_{[u,v]} = \sum_{i=u}^v \sigma_i^\ell$ and we define the random variable $Z_{[1,n]} = \max_{1\leq k\leq n}\{S^1_{[1,k]} + S^2_{[k,n]}\}$. With $\max(1+p, 3p) < a < 3$, we can define $M = \sup_n(Z_{[1,n]} - na) < \infty$ which is the supremum of an independent subadditive process with negative drift. We denote the moment generating functions as follows $\Lambda^\ell(\theta) = \log \mathbb{E}\left[e^{\theta\sigma_1^\ell}\right]$. It is easy to compute (this will be done in a much more general context in Part 3),

$$\begin{aligned}\Lambda(\theta) &= \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{\theta Z_{[1,n]}}\right] - a\theta \\ &= \max(\Lambda^1(\theta),\Lambda^2(\theta)) - a\theta.\end{aligned}$$

Thanks to Theorem 11, we see that we have

$$\theta^* = \min\left\{\theta^1,\, \theta^2\right\}, \quad \text{where } \theta^\ell = \sup\{\theta > 0, \Lambda^\ell(\theta) < a\theta\}.$$

Note that our example corresponds to a system of 2 queues in tandem and that this result follows directly from the work of Ganesh [44].

The rate functions for $S^1_{[1,n]}$ and $S^2_{[1,n]}$ are

$$J^1(x) = \begin{cases} (x-1)\log[(x-1)/p] + (2-x)\log[(2-x)/(1-p)], & x\in[1,2], \\ +\infty, & x\notin[1,2]. \end{cases}$$

$$J^2(x) = \begin{cases} x/3\log[x/3p] + (1-x/3)\log[(3-x)/(3-3p)], & x\in[0,3], \\ +\infty, & x\notin[0,3]. \end{cases}$$

On the exponential scale a deviation happens in the most likely way. Hence, we have for all $x \geq \max(1+p, 3p)$:

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Z_{[1,n]} > nx\right) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(Y_{[1,n]} > n(x-a)\right) = -I(x), \tag{5.8}$$

where $I(x) = \min(J^1(x), J^2(x))$. This function is clearly not convex as shown on Figure 5.1 for $p = 1/3$.

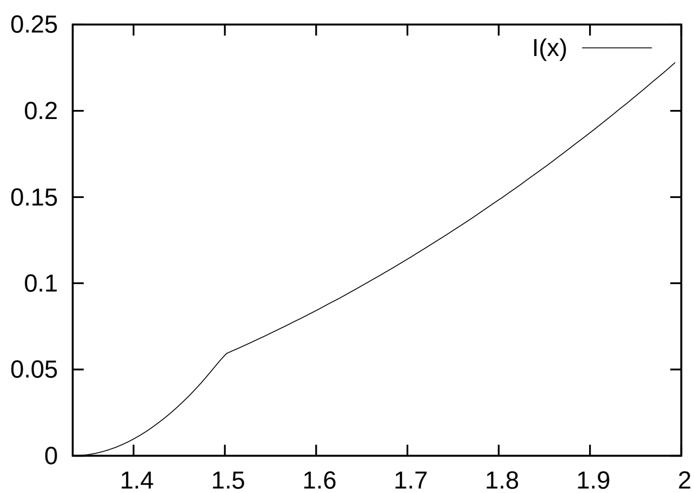


FIG. 5.1 – Nonconvex rate function

We have $I(x+a) \geq \Lambda^*(x)$ and these functions are distinct. Hence, in this case, the upper bound given by Gärtner-Ellis Theorem is not tight.

There is no hope to find a general large deviation theory of subadditive processes where the rate function would be given by the convex conjugate of the logarithmic moment generating function. Another example (leading to the same conclusion) can be found in [80]. In their work, Seppäläinen and Yukich consider subadditive Euclidean functionals that are regular nearly additive processes. This property allows them to derive a LDP for such functionals. In our framework, such an approach is not valid since it cannot handle previous example.

Our example provides a simple illustration of a limitation inherent in the convex methodology : the upper rate function is the best possible convex upper bound and does not necessary coincide with the actual rate function. A similar phenomena in the context of mixture of probability measures was observed by Dinwoodie and Zabell in [33].

We end this section by showing that the information given by the scaled moment generating function is not enough to prove a LDP. We modify our example in order to get two independent subadditive processes with the same scaled moment generating function but with different rate functions.

Consider the following sequence $\{\sigma_i^3\}$ of i.i.d. random variables independent of previously defined random variables,

$$\sigma_i^3 = kX_i^q + y,$$

with $\{X_i^q\}$ a sequence of i.i.d. Bernoulli random variables $\mathbb{P}(X_i^q = 1) = q = 1 - \mathbb{P}(X_i^q = 0)$.

With the same notation as above, we take $\bar{Z}_{[1,n]} = \max_{1\leq \ell \leq j \leq n}\left\{S^1_{[1,\ell]} + S^2_{[\ell,j]} + S^3_{[j,n]}\right\}$. With the following choice of parameters : $p = 1/3$, $q = 1/2$, $k = 2$ and $y = 1/5$ we have $\Lambda^3(\theta) \leq \max(\Lambda^1(\theta), \Lambda^2(\theta))$. Hence the processes $Z_{[1,n]}$ and $\bar{Z}_{[1,n]}$ have the same scaled moment generating function but they clearly have different rate functions as shown on Figure 5.2.

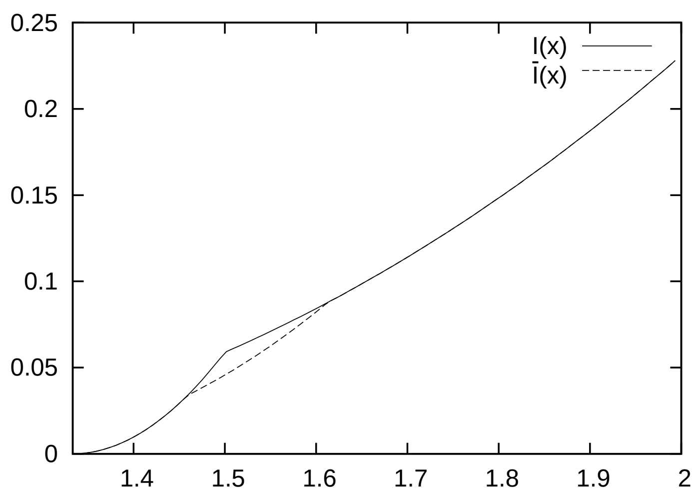


FIG. 5.2 – Same moment generating function with different rate functions

This very simple example shows the asymmetry in the large deviation behavior of the upper and lower tails of a subadditive process. In particular, the following result follows directly from Hammersley's work [55] :

Proposition 20. *Under assumption* **(A2)** *and if*

$$\mathbb{E}\left[\exp(\theta Y_{[1,n]})\right] < \infty,$$

for some $0 \geq \theta \geq \tau$, $\tau < 0$ *and all* r. *Then the limits,*

$$\psi(x) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Y_{[1,n]} \leq nx) \quad \textit{and,} \quad \Lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp(\theta Y_{[1,n]}),$$

exist for all x *and all* $\theta \leq 0$ *and satisfy*

$$\psi(x) = \inf_{\theta} \{\Lambda(\theta) - \theta x\} \quad \textit{and,} \quad \Lambda(\theta) = \sup_{x} \{\psi(x) + \theta x\}.$$

This results allows Grossmann and Yakir [54] to prove a similar result to ours but for the large deviations of the global maxima of independent super-additive processes. We should stress that Proposition 20 leaves open the question : for what values of x is it the case that $\psi(x) < 0$? In particular, Grimmett gives in [53], an example of a subadditive process for which it is not the case that $\psi(\mu - \epsilon) < 0$ (where $\mu = \lim_n \frac{Y_{[1,n]}}{n}$).

5.2.3 Proofs

Moment generating function

Lemma 27. *Under the foregoing assumption, the following limit*

$$\Lambda(\theta) \quad = \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right],$$

exists in $\mathbb{R} \cup \{+\infty\}$ *for all* $\theta \geq 0$. $\Lambda(.)$ *is a proper convex function. Further the domain of* Λ *is given by* $\{\theta \geq 0,\ \Lambda(\theta) < \infty\} = \{\theta \geq 0,\ \mathbb{E}[e^{\theta Y_{[1,1]}}] < \infty\} \supset [0, \eta)$ *where* η *is defined in* **(A3)**.

Proof. Let

$$\Lambda_n(\theta) \quad = \quad \log \mathbb{E}\left[e^{\theta \frac{Y_{[1,n]}}{n}}\right]. \tag{5.9}$$

Thanks to the subadditive property of Y, we have,

$$Y_{[1,n+m]} \leq Y_{[1,n]} + Y_{[n+1,n+m]},$$

and $Y_{[1,n]}$ and $Y_{[n+1,n+m]}$ are independent. Hence for $\theta \geq 0$, we have,

$$\Lambda_{n+m}((n+m)\theta) \leq \Lambda_n(n\theta) + \Lambda_m(m\theta).$$

Hence we can define for any $\theta \geq 0$,

$$\Lambda(\theta) \quad = \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] = \lim_{n\to\infty} \frac{\Lambda_n(n\theta)}{n} = \inf_{n\geq 1} \frac{\Lambda_n(n\theta)}{n},$$

as an extended real number. The fact that Λ is a proper convex function follows from Lemma 2.3.9 of [32]. The last fact follows from Assumption **(A3)** and,

$$\frac{1}{n} \log \mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] \leq \log \mathbb{E}\left[e^{\theta Y_{[1,1]}}\right] \quad \text{for } \theta \geq 0.$$

□

Lemma 28. *Under the foregoing assumptions, we have* $\theta^* > 0$ *and*

$$\begin{aligned} \Lambda(\theta) < 0 \quad & \textit{if} \quad \theta \in (0, \theta^*), \\ \Lambda(\theta) > 0 \quad & \textit{if} \quad \theta > \theta^*. \end{aligned}$$

Proof. Let

$$\theta_n = \sup\{\theta > 0,\ \Lambda_n(n\theta) < 0\}. \tag{5.10}$$

We fix n such that

$$\mathbb{E}[Y_{[1,n]}] < 0.$$

We first show that $\theta_n > 0$ and

$$\Lambda_n(n\theta) < 0 \quad \text{if} \quad \theta \in (0, \theta_n), \tag{5.11}$$

$$\Lambda_n(n\theta) > 0 \quad \text{if} \quad \theta > \theta_n \tag{5.12}$$

The function $\theta \mapsto \Lambda_n(n\theta)$ is convex, continuous and differentiable on $[0, \eta)$. Hence we have

$$\Lambda_n(n\delta) = \delta \mathbb{E}[Y_{[1,n]}] + o(\delta),$$

which is less than zero for sufficiently small $\delta > 0$. Hence, the set over which the supremum in the definition of θ_n is taken is not empty and $\theta_n > 0$. Now (5.11) and (5.12) follow from the definition of θ_n, the convexity of $\theta \mapsto \Lambda_n(n\theta)$ and the fact that $\Lambda_n(0) = 0$.

We now show that $\theta_n \to \theta^*$ as $n \to \infty$. We have for $\theta \geq 0$

$$\lim_{n\to\infty} \frac{\Lambda_n(n\theta)}{n} = \inf_{n\geq 1} \frac{\Lambda_n(n\theta)}{n} = \Lambda(\theta).$$

Hence for $\theta \geq 0$, we have $\frac{\Lambda_n(n\theta)}{n} \geq \Lambda(\theta)$ and

$$\forall \theta \in (0, \theta_n), \quad \Lambda(\theta) \leq \frac{\Lambda_n(n\theta)}{n} < 0.$$

This implies that $\theta^* \geq \theta_n > 0$. If $\theta^* < \infty$, we can choose $\epsilon > 0$ such that $\theta^* - \epsilon > 0$ and then we have $\Lambda_n(n(\theta^* - \epsilon))/n \to \Lambda(\theta^* - \epsilon) < 0$. Hence for sufficiently large n, we have $\frac{\Lambda_n(n(\theta^* - \epsilon))}{n} < 0$, hence $\theta^* - \epsilon \leq \theta_n$, and we proved that $\theta_n \to \theta^*$. $\Lambda(.)$ is a convex function and since $\Lambda(0) = 0$, the lemma follows.

If $\theta^* = \infty$, we still have $\theta_n \to \infty$ (that will be needed in proof of Lemma 29) by the same argument as above with $\theta^* - \epsilon$ replaced by any real number.

□

Upper bound

Lemma 29. *Under the foregoing assumptions, we have*

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(M > x) \leq -\theta^*.$$

Proof. For any $L \geq 1$, we denote for $n \geq 0$,

$$V_n(L) := \max\{Y_{[nL+1,nL+1]}, Y_{[nL+1,nL+2]}, \dots Y_{[nL+1,(n+1)L]}\},$$

and we have, for L such that $\mathbb{E}[Y_{[1,L]}] < 0$,

$$M \leq \max\left\{V_0(L), \sup_{n\geq 0}\left(\sum_{i=0}^{n} Y_{[iL+1,(i+1)L]} + V_{n+1}(L)\right)\right\} =: U(L).$$

and the right-hand term is almost surely finite.

We will show that under previous assumptions, we have

$$\limsup_{x\to\infty} \frac{1}{x}\log \mathbb{P}(U(L) > x) \leq -\theta_L, \tag{5.13}$$

where θ_L is defined as in (5.10).

Thanks to Lemma 28 we know that $\theta_L \to \theta^*$ as L tends to infinity, hence the lemma will follow.

We now prove (5.13). We define

$$S_{[0,n]}(L) = \sum_{i=0}^{n} Y_{[iL+1,(i+1)L]}.$$

For all θ and $\epsilon > 0$, there is a finite positive constant A such that,

$$\mathbb{E}\left[e^{\theta S_{[0,n]}(L)}\right] \leq Ae^{(n+1)(\Lambda_L(L\theta)+\epsilon)}.$$

The constant A depends on θ and ϵ, but this is suppressed in the notation.

Let $\theta \in (0, \theta_L)$, we have $\Lambda_L(L\theta) < 0$ see proof of Lemma 28.

We have (with the convention that the constant A differs from line to line but is always finite),

$$\begin{aligned}
\mathbb{E}\left[e^{\theta U(L)}\right] &= \mathbb{E}\left[\max\left\{e^{\theta V_0(L)}, \sup_{n\geq 0} e^{\theta S_{[0,n]}(L)+V_{n+1}(L)}\right\}\right] \\
&\leq \mathbb{E}\left[e^{\theta V_0(L)}\right] + \mathbb{E}\left[\sup_{n\geq 0} e^{\theta S_{[0,n]}(L)+V_{n+1}(L)}\right] \\
&\leq \mathbb{E}\left[e^{\theta V_0(L)}\right]\left(1 + \sum_{n\geq 0}\mathbb{E}\left[e^{\theta S_{[0,n]}(L)}\right]\right) \\
&\leq \mathbb{E}\left[e^{\theta V_0(L)}\right]\left(1 + \sum_{n\geq 1} Ae^{n(\Lambda_L(L\theta)+\epsilon)}\right).
\end{aligned}$$

Since $\theta \in (0, \theta_L)$, we can choose $\epsilon > 0$ such that

$$\Lambda_L(L\theta) + \epsilon < 0,$$

and thanks to Lemma 27, we have

$$\mathbb{E}\left[e^{\theta V_0(L)}\right] \leq \mathbb{E}\left[e^{\theta Y_{[1,1]}}\right] + \left(\mathbb{E}\left[e^{\theta Y_{[1,1]}}\right]\right)^2 + \cdots + \left(\mathbb{E}\left[e^{\theta Y_{[1,1]}}\right]\right)^L < \infty.$$

Therefore, $\mathbb{E}\left[e^{\theta U(L)}\right] \leq A$ for some finite constant A. Hence by Chernoff's inequality,

$$\mathbb{P}\left(U(L) > x\right) \leq e^{-\theta x}\mathbb{E}\left[e^{\theta U(L)}\right] \leq Ae^{-\theta x}.$$

Since the above holds for all $0 < \theta < \theta_L$, we get

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}\left(U(L) > x\right) \leq -\theta_L.$$

□

Lower Bound

We take the following notation : $\mathbb{P}_n(A) = \mathbb{P}(Y_{[1,n]} \in A)$, and the same convention for $\hat{\mathbb{P}}_{\theta,n}$ defined for θ such that $\Lambda_n(n\theta) < \infty$, as the transformed measure :

$$\hat{\mathbb{P}}_{\theta,n} := e^{\theta Y_{[1,n]}} e^{-\Lambda_n(n\theta)}\mathbb{P}_n.$$

The function $\theta \mapsto \Lambda(\theta)$ is convex, hence the left-hand derivatives $\Lambda'(\theta-)$ and the right-hand derivatives $\Lambda'(\theta+)$ exist for all $\theta > 0$. Moreover, we have $\Lambda'(\theta-) \leq \Lambda'(\theta+)$ and the function $\theta \mapsto \frac{1}{2}(\Lambda'(\theta-) + \Lambda'(\theta+))$ is non-decreasing, hence $\Lambda'(\theta) = \Lambda'(\theta-) = \Lambda'(\theta+)$ except for $\theta \in \Delta$, where Δ is at most countable.

The following lemma is similar to Lemma 10 of Zerner [86],

Lemma 30. *Let $\theta > 0$, and $u < v$ such that*

$$u < \Lambda'(\theta-) \leq \Lambda'(\theta+) < v < \infty.$$

Then

$$\lim_{n\to\infty} \hat{\mathbb{P}}_\theta\left(\frac{Y_{[1,n]}}{n} \in (u,v)\right) = 1. \tag{5.14}$$

Proof. First note that $\Lambda'(\theta+)$ is well defined hence there exists $y > 0$ such that $\Lambda(\theta+y) < \infty$, hence $\Lambda_n(n(\theta+y)) < \infty$, for all n sufficiently large. We have for all $0 < x < y$,

$$\begin{aligned} \frac{1}{n}\log \hat{\mathbb{P}}_\theta\left(Y_{[1,n]} \geq nv\right) &= -\frac{\Lambda_n(n\theta)}{n} + \frac{1}{n}\log \mathbb{E}\left[e^{(\theta+x)Y_{[1,n]}}e^{-xY_{[1,n]}}, Y_{[1,n]} \geq nv\right] \\ &\leq -\frac{\Lambda_n(n\theta)}{n} + \frac{\Lambda_n(n(\theta+x))}{n} - xv, \end{aligned}$$

Hence, we have

$$\limsup_{n\to\infty} \frac{1}{n}\log \hat{\mathbb{P}}_\theta\left(Y_{[1,n]} \geq nv\right) \leq -\left(v - \frac{\Lambda(\theta+x) - \Lambda(\theta)}{x}\right)x,$$

which is negative for small x. A corresponding statement holds for the event $\{Y_{[1,n]} \leq nu\}$, this implies (5.14). □

Lemma 31. *Under the foregoing assumptions, we have*

$$\liminf_{x\to\infty} \frac{1}{x}\log \mathbb{P}(M > x) \geq -\theta^*.$$

Proof.

We consider first the case where there exists $\theta > \theta^*$ such that $\Lambda(\theta) < \infty$. In this case we have $\Lambda(\theta^*) = 0$ and $\Lambda'(\theta^*+) > 0$. To prove this, assume that $\Lambda'(\theta^*+) = 0$. Take $\theta < \theta^*$, thanks to Lemma 28, we have $\Lambda(\theta) < 0$. Choose $\epsilon > 0$ such that $0 < \Lambda(\theta^* + \epsilon) < \epsilon|\Lambda(\theta)|$. We have

$$\frac{\Lambda(\theta^* + \epsilon)}{\epsilon} < \frac{-\Lambda(\theta)}{\theta^* - \theta},$$

which contradicts the convexity of $\Lambda(\theta)$.

Hence, we can find $t \leq \theta^* + \epsilon$ such that

$$0 < \Lambda(t), \quad t \notin \Delta.$$

Note that these conditions imply $t > \theta^*$ and $\Lambda'(t) \geq \Lambda'(\theta^*+) > 0$.

Moreover for any $\alpha > 0$, $\epsilon > 0$, we have

$$\mathbb{P}\left(Y_{[1,n]} > n\alpha\right) \geq e^{\Lambda_n(nt)} e^{-nt(\alpha+\epsilon)} \hat{\mathbb{P}}_t\left(\frac{Y_{[1,n]}}{n} \in (\alpha, \alpha+\epsilon)\right).$$

Fix $\alpha := \Lambda'(t) - \epsilon/2 > 0$. Given $x > 0$, define $n := \lfloor x/\alpha \rfloor$. We have

$$\mathbb{P}(M > x) \geq \mathbb{P}(Y_{[1,n]} \geq n\alpha),$$

hence we have

$$\frac{1}{x}\log \mathbb{P}(M > x) \geq \frac{1}{n\alpha}\left(\Lambda_n(nt) - nt(\alpha+\epsilon) + \log \hat{\mathbb{P}}_t\left(\frac{Y_{[1,n]}}{n} \in (\alpha, \alpha+\epsilon)\right)\right)$$

Taking the limit in x and n (while α is fixed) gives thanks to Lemma 5.14

$$\begin{aligned} \liminf_{x\to\infty} \frac{1}{x}\log \mathbb{P}(M > x) &\geq \frac{1}{\alpha}\left(\Lambda(t) - t(\alpha+\epsilon)\right) \\ &\geq -(\theta^* + \epsilon)\frac{(\alpha+\epsilon)}{\alpha}. \end{aligned}$$

We consider now the case where for all $\theta > \theta^*$, we have $\Lambda(\theta) = \infty$.

Fix $K > 0$ and let $\tilde{\mathbb{P}}_n^K$ be the law of $Y_{[1,n]}$ conditioned on $\{Y_i \leq K,\ i = 1, \ldots, n\}$, where we denote $Y_i = Y_{[i,i]}$. Then for all n, we have

$$\mathbb{P}_n(A) \geq \tilde{\mathbb{P}}_n^K(A)\mathbb{P}(Y_1 \leq K)^n.$$

Thanks to subadditivity, we have $\tilde{\mathbb{P}}^K(Y_{[1,n]} \leq nK) = 1$ and the following moment generating function is bounded, for $\theta \geq 0$,

$$\tilde{\Lambda}_n^K(\theta) := \log \tilde{\mathbb{E}}^K\left[e^{\theta Y_{[1,n]}}\right] \leq n\theta K.$$

Moreover, we have

$$\begin{aligned} \tilde{\Lambda}_n^K(\theta) &= \log \mathbb{E}\left[e^{\theta Y_{[1,n]}},\ Y_i \leq K,\ i = 1, \ldots, n\right] - n\log \mathbb{P}(Y_1 \leq K) \\ &=: \Lambda_n^K(\theta) - n\log \mathbb{P}(Y_1 \leq K). \end{aligned}$$

Thanks to subadditivity, we have

$$e^{\theta Y_{[1,n+m]}}\mathbb{1}_{\{Y_i\le K,\, i=1,\dots,n+m\}} \le e^{\theta Y_{[1,n]}}\mathbb{1}_{\{Y_i\le K,\, i=1,\dots,n\}} e^{\theta Y_{[n+1,n+m]}}\mathbb{1}_{\{Y_i\le K,\, i=n+1,\dots,n+m\}},$$

hence thanks to the independence, we can define,

$$\begin{aligned}\tilde{\Lambda}^K(\theta) &:= \lim_{n\to\infty}\frac{\tilde{\Lambda}_n^K(n\theta)}{n}\\ &= \lim_{n\to\infty}\frac{\Lambda_n^K(n\theta)}{n} - \log\mathbb{P}(Y_1\le K)\\ &=: \Lambda^K(\theta) - \log\mathbb{P}(Y_1\le K).\end{aligned}$$

Let $\tilde{\theta}^K = \sup\{\theta>0,\ \tilde{\Lambda}^K(\theta)<0\}$.

Thanks to the preceding proof, there exists $\alpha>0$ such that

$$\liminf_{n\to\infty}\frac{1}{n}\log\tilde{\mathbb{P}}_n^K([\alpha,\infty)) \ge -(\tilde{\theta}^K+\epsilon)(\alpha+\epsilon).$$

Hence we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n([\alpha,\infty)) \quad\ge\quad -(\tilde{\theta}^K+\epsilon)(\alpha+\epsilon)+\log\mathbb{P}(Y_1\le K). \tag{5.15}$$

Note that for any fixed θ, the function $\Lambda^K(\theta)$ is nondecresing in K and $\lim_{K\to\infty}\Lambda^K(\theta) = \Lambda(\theta)$. Hence we have $\tilde{\Lambda}^K(\theta^*+\epsilon)\to\infty$ as K tends to infinity. Hence for sufficiently large K, we have $\tilde{\Lambda}^K(\theta^*+\epsilon)>0$ and this implies that $\tilde{\theta}^K\le\theta^*+\epsilon$.

Hence dividing by α and taking the limit $K\to\infty$ in (5.15) gives :

$$\liminf_{x\to\infty}\frac{1}{x}\log\mathbb{P}(M>x) \ge \frac{1}{\alpha}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n([\alpha,\infty)) \quad\ge\quad -(\theta^*+2\epsilon)\frac{\alpha+\epsilon}{\alpha},$$

and the lemma follows. □

Proof of Corollary 3

Proof.

We have only to show that $\theta^* = \inf_{\alpha>0}\frac{I(\alpha)}{\alpha}$. Thanks to Varadhan's Integral Lemma (see Theorem 4.3.1 in [32]), we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{\theta Y_{[1,n]}}\right] = \Lambda(\theta) = \sup_x\{\theta x - I(x)\} =: I^*(\theta),$$

for $\theta<\theta^*+\epsilon$.

Thanks to Lemma 28, we have for ϵ positive, $0>\Lambda(\theta^*-\epsilon)\ge(\theta^*-\epsilon)x-I(x)$, from which we get

$$\inf_{\alpha>0}\frac{I(\alpha)}{\alpha}\ge\theta^*.$$

For $\theta^*+\epsilon>\theta>\theta^*$, we have $\Lambda(\theta)>0$ thanks to Lemma 28. Hence there exists $\alpha^*\in\mathbb{R}$ such that $\theta\alpha^*-I(\alpha^*)>0$. Since I is non-negative and $\theta>0$, we have $\alpha^*>0$ and,

$$\inf_{\alpha>0}\frac{I(\alpha)}{\alpha}\le\frac{I(\alpha^*)}{\alpha^*}<\theta.$$

Since we took any $\theta^* + \epsilon > \theta > \theta^*$, we proved

$$\inf_{\alpha>0} \frac{I(\alpha)}{\alpha} \le \theta^*.$$

$\square$

Estimating tails

In this paper, we are interested in estimating tail probabilities. As in [18], we introduce (Λ^* denotes the convex conjugate of Λ),

$$\Lambda^+(\theta) = \begin{cases} \Lambda(\theta), & \text{if } \theta \ge 0, \\ +\infty, & \text{if } \theta < 0. \end{cases}$$
$$\Lambda^{*+}(x) = \begin{cases} \Lambda^*(x), & \text{if } x \ge \gamma(0) - a, \\ 0, & \text{if } x < \gamma(0) - a. \end{cases}$$

Lemma 32. *We have*

$$\Lambda^{*+}(x) = \sup_{\theta\in\mathbb{R}} \left\{\theta x - \Lambda^+(\theta)\right\},$$

and the function $\Lambda^{*+}(x)$ is non-decreasing in x.

Proof. We show that $\Lambda^*(\gamma(0) - a) = 0$ and for all $x \ge \gamma(0) - a$, we have $\Lambda^*(x) = \sup_{\theta\ge 0} \{\theta x - \Lambda(\theta)\}$, from which the lemma follows. Since Λ_n (defined in (5.9)) is convex and differentiable in 0, we have $\Lambda_n(\theta) \ge \Lambda_n'(0)\theta$ and taking the limit on both sides, we get

$$\Lambda(\theta) \ge \theta(\gamma(0) - a).$$

Hence for all $x \ge \gamma(0) - a$, we have for $\theta < 0$

$$\theta x - \Lambda(\theta) \le \theta(\gamma(0) - a) - \Lambda(\theta) \le \Lambda^*(\gamma(0) - a) = 0.$$

The monotonicity of Λ^{*+} follows from the monotonicity of $\theta x - \Lambda(\theta)$ in x as θ is fixed. $\square$

5.3 Large deviations for monotone-separable networks

In this part, we consider a stochastic network described by the following framework

- The network has a single input point process N, with points $\{T_n\}$; for all $m \le n \in N$, let $N_{[m,n]}$ be the restriction of N, namely the point process with points $\{T_\ell\}_{m\le\ell\le n}$.
- The network has a.s. finite activity for all finite restrictions of N : for all $m \le n \in N$, let $X_{[m,n]}(N)$ be the time of last activity in the network, when this one starts empty and is fed by $N_{[m,n]}$. We assume that for all finite m and n as above, $X_{[m,n]}$ is finite.

We assume that there exists a set of functions $\{f_\ell\}$, $f_\ell : \mathbb{R}^\ell \times K^\ell \to \mathbb{R}$, such that :

$$X_{[m,n]}(N) = f_{n-m+1}\{(T_\ell, \zeta_\ell),\ m \le \ell \le n\}, \tag{5.16}$$

for all n, m and N, where the sequence $\{\zeta_n\}$ is that describing service times and routing decisions.

We say that a network described as above is monotone-separable if the functions f_n are such that the following properties hold for all N :

1. **Causality :** for all $m \leq n$,
$$X_{[m,n]}(N) \geq T_n;$$
2. **External monotonicity :** for all $m \leq n$,
$$X_{[m,n]}(N') \geq X_{[m,n]}(N),$$
whenever $N' := \{T'_n\}$ is such that $T'_n \geq T_n$ for all n, a property which we will write $N' \geq N$ for short ;
3. **Homogeneity :** for all $c \in \mathbb{R}$ and for all $m \leq n$
$$X_{[m,n]}(N + c) = X_{[m,n]}(N) + c;$$
4. **Separability :** if for all $m \leq \ell < n$, $X_{[m,\ell]}(N) \leq T_{\ell+1}$, then
$$X_{[m,n]}(N) = X_{[\ell+1,n]}(N).$$

5.3.1 Tail asymptotics of the maximal dater

Stability and stationary maximal daters

In this section, we summarize the main results of Baccelli and Foss [13].

By definition, for $m \leq n$, the $[m,n]$ maximal dater is
$$Z_{[m,n]}(N) := X_{[m,n]}(N) - T_n = X_{[m,n]}(N - T_n).$$
Note that $Z_{[m,n]}(N)$ is a function of $\{\zeta_l\}_{m \leq \ell \leq n}$ and $\{\tau_l\}_{m \leq \ell \leq n}$ only, where $\tau_n = T_{n+1} - T_n$. In particular, $Z_n := Z_{[n,n]}(N)$ is not a function of N (which makes the notation consistent).

Under the above conditions, the variables $X_{[m,n]}$ and $Z_{[m,n]}$ satisfy the internal monotonicity property : for all N, $m \leq n$,
$$\begin{aligned} X_{[m-1,n]}(N) &\geq X_{[m,n]}(N), \\ Z_{[m-1,n]}(N) &\geq Z_{[m,n]}(N). \end{aligned}$$
In particular, the sequence $\{Z_{[-n,0]}(N)\}$ is non-decreasing in n. Put
$$Z := Z_{(-\infty,0]} = \lim_{n\to\infty} Z_{[-n,0]}(N) \leq \infty.$$
Lemma 33. *[13]* **Subadditive property of** Z
Under the above conditions, $\{Z_{[m,n]}\}$ satisfies the following subadditive property : for all $m \leq \ell < n$, for all N,
$$Z_{[m,n]}(N) \leq Z_{[m,\ell]}(N) + Z_{[\ell+1,n]}(N).$$
Assume the variables $\{\tau_n, \zeta_n\}$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P}, \theta)$, where θ is an ergodic, measure-preserving shift transformation, such that $(\tau_n, \zeta_n) \circ \theta = (\tau_{n+1}, \zeta_{n+1})$. The following integrability assumptions are also assumed to hold :
$$\mathbb{E}[\tau_n] := \lambda^{-1} = a < \infty, \quad \mathbb{E}[Z_n] < \infty.$$
Denote by $Q = \{T'_n\}$ the degenerate input process with $T'_n = 0$ a.s. for all n.

Lemma 34. *[13] Under the foregoing ergodic assumption, there exists a non-negative constant* $\gamma(0)$ *such that*

$$\lim_{n\to\infty} \frac{Z_{[-n,-1]}(Q)}{n} = \lim_{n\to\infty} \frac{\mathbb{E}\left[Z_{[-n,-1]}(Q)\right]}{n} = \gamma(0)\ a.s.$$

The main result on the stability region is the following :

Theorem 12. *[13] Under the foregoing ergodic assumptions, either* $Z = \infty$ *a.s. or* $Z < \infty$ *a.s.*

(a) If $\lambda\gamma(0) < 1$*, then* $Z < \infty$ *a.s.*

(b) If $Z < \infty$ *a.s., then* $\lambda\gamma(0) \leq 1$*.*

A proof of this result can be found in [10] see Theorem 2.11.3. We give in Section 5.3.2 an upper bound and a lower bound that allow to prove Theorem 12. These bounds will be used for the study of large deviations.

Moment generating function and tail asymptotics

In the rest of the paper, we will make the following assumptions (that are of course compatible with previous stationary ergodic assumptions) :

– Assumption **(AA)** on the arrival process into the network $\{T_n\}$:
  $\{T_n\}$ is a renewal process independent of the service time and routing sequences $\{\zeta_n\}$. Moreover for all real θ, the function

$$\Lambda_T(\theta) \quad = \quad \log \mathbb{E}\left[e^{\theta(T_1 - T_0)}\right]$$

  is finite in a neighborhood of 0.

– Assumption **(AZ)** : the sequence $\{\zeta_n\}$ is a sequence of i.i.d. random variables, such that the random variable $Z_0 := Z_{[0,0]}$ is light-tailed, i.e. for θ in a neighborhood of 0,

$$\mathbb{E}[e^{\theta Z_0}] < +\infty.$$

– Stability : $\gamma(0) < a := \mathbb{E}[T_1 - T_0]$ see Theorem 12.

The subadditive property of Z directly implies the following property (its proof follows the lines of the proof of Lemma 27) : for any monotone separable network that satisfies assumption **(AZ)**, the following limit

$$\Lambda_Z(\theta) \quad = \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta Z_{[1,n]}(Q)}\right],$$

exists in $\mathbb{R} \cup \{+\infty\}$ for all θ. Further, the origin belongs to the interior of its domain $D_{\Lambda_Z} = \{\theta,\ \Lambda_Z(\theta) < \infty\}$. $\Lambda_Z(.)$ is a proper convex function.

Note that the subadditive property of Z is valid regardless of the point process N (see Lemma 33). Like in the study of the stability of the network, it turns out that the right quantity to look at is $Z_{[m,n]}(Q)$ where Q is the degenerate input point process with all its point equal to 0.

Theorem 13. *Under previous assumptions, we have*

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(Z > x) = -\theta^* < 0,$$

where $\theta^* = \sup\{\theta > 0,\ \Lambda_T(-\theta) + \Lambda_Z(\theta) < 0\}$.

It is relatively easy to see that under our light-tailed assumption the stationary maximal dater Z will be light-tailed (see Corollary 3 in [14]). Theorem 13 shows that the tail distribution of Z is indeed exponentially decaying for any monotone-separable network. But the main contribution of this theorem is to give an explicit way of computing this rate of decay. It is the goal of the third part of this paper to show that it is actually possible to calculate the logarithmic moment generating function Λ_Z for various categories of networks.

In the context of heavy-tailed asymptotics (and more precisely for subexponential distributions), the moment generating function is infinite for all $\theta > 0$. There is no general result for the tail asymptotics of the maximal dater of a monotone separable network. However the methodology derived by Baccelli and Foss [14] allows to get exact asymptotics for (max,plus)-linear networks [16] and generalized Jackson networks [15].

5.3.2 Upper $G/G/1/\infty$ queue and lower bound for the maximal dater

The material of this section is not new and may be found in various references (that are given in what follows). For the sake of completeness, we include all the proofs. We derive now upper and lower bounds for the stationary maximal dater Z. These bounds allow to prove Theorem 12 and will be the main tools for the study of large deviations.

We first derive a lower bound that will give us part (b) of Theorem 12.

Proposition 21. *We have the following lower bound*

$$Z \geq \sup_{n\geq 0} \left(Z_{[-n,0]}(Q) + T_{-n} - T_0 \right).$$

Proof.
For n fixed, let N^n be the point process with point $T_j^n = T_{-n} - T_0$, for all j. Then

$$\begin{aligned} Z_{[-n,0]} &= X_{[-n,0]}(N) - T_0 \geq X_{[-n,0]}(N^n) \\ &= X_{[-n,0]}(Q) + T_{-n} - T_0 = Z_{[-n,0]}(Q) + T_{-n} - T_0, \end{aligned}$$

where we used external monotonicity in the first inequality and homogeneity between the first and second line. □

Proof.of Theorem 12 part (b)
Suppose that $\lambda\gamma(0) > 1$, then we have

$$\liminf_{n\to\infty} \frac{Z_{[-n,0]}(N)}{n} \geq \gamma(0) - a > 0,$$

which concludes the proof of part (b). □

We assume now that $\gamma(0) < a$. We pick an integer $L \geq 1$ such that

$$\mathbb{E}\left[Z_{[-L,-1]}(Q)\right] < La, \tag{5.17}$$

which is possible in view of Lemma 34. Without loss of generality, we assume that $T_0 = 0$. Part (a) of Theorem 12 will follow from the following proposition (that can be found in [14]) :

Proposition 22. *The stationary maximal dater Z is bounded from above by the stationary response time $\hat{R}$ in the $G/G/1/\infty$ queue with service times*

$$\hat{s}_n := Z_{[L(n-1)+1,Ln]}(Q)$$

and inter-arrival times $\hat{\tau}_n := T_{Ln} - T_{L(n-1)}$, *where* L *is the integer defined in (5.17). Since* $\mathbb{E}[\hat{s}_1] < \mathbb{E}[\hat{\tau}_1] = La$, *this queue is stable. With the convention* $\sum_0^{-1} = 0$, *we have,*

$$Z \leq \hat{s}_0 + \sup_{k \geq 0} \sum_{i=-k}^{-1} (\hat{s}_i - \hat{\tau}_{i+1}) \, .$$

Proof.

To an input process N, we associate the following upper bound process, $N^+ = \{T_n^+\} \geq N$, where $T_n^+ = T_{kL}$ if $n = (k-1)L + 1, \ldots, kL$. Then for all n, since we assumed $T_0 = 0$, we have thanks to the external monotonicity,

$$X_{[-n,0]}(N) = Z_{[-n,0]}(N) \leq X_{[-n,0]}(N^+) = Z_{[-n,0]}(N^+). \tag{5.18}$$

We show that for all $k \geq 1$,

$$Z_{[-kL+1,0]}(N^+) \leq \hat{s}_0 + \sup_{-k+1 \leq i \leq 0} \sum_{j=-i}^{-1} (\hat{s}_j - \hat{\tau}_{j+1}). \tag{5.19}$$

This inequality will follow from the two next lemmas

Lemma 35. *Assume* $T_0 = 0$. *For any* $m < n \leq 0$,

$$Z_{[m,0]}(N) \leq Z_{[n,0]}(N) + (Z_{[m,n-1]}(N) - \tau_{n-1})^+.$$

Proof.

Assume first that $Z_{[m,n-1]}(N) - \tau_{n-1} \leq 0$, which is exactly $X_{[m,n-1]}(N) \leq T_n$. Then by the separability property, we have

$$Z_{[m,0]}(N) = X_{[m,0]}(N) = X_{[n,0]}(N) = Z_{[n,0]}(N).$$

Assume now that $Z_{[m,n-1]}(N) - \tau_{n-1} > 0$. Let $N' = \{T'_j\}$ be the input process defined as follows

$$\begin{aligned} \forall j \leq n-1, \quad T'_j &= T_j, \\ \forall j \geq n, \quad T'_j &= T_j + Z_{[m,n-1]}(N) - \tau_{n-1}. \end{aligned}$$

Then we have $N' \geq N$ and $X_{[m,n-1]}(N') \leq T'_n$, hence by the external monotonicity, the separability and the homogeneity properties, we have

$$\begin{aligned} Z_{[m,0]}(N) &= X_{[m,0]}(N) \leq X_{[m,0]}(N') \\ &= X_{[n,0]}(N') = X_{[n,0]}(N) + Z_{[m,n-1]}(N) - \tau_{n-1} \\ &= Z_{[n,0]}(N) + Z_{[m,n-1]}(N) - \tau_{n-1}. \end{aligned}$$

□

From this lemma we derive directly

Lemma 36. *Assume* $T_0 = 0$. *For any* $n < 0$,

$$Z_{[n,0]}(N) \leq \sup_{n \leq k \leq 0} \left(\sum_{i=k}^{-1} (Z_i - \tau_{i+1}) \right) + Z_0,$$

with the convention $\sum_0^{-1} = 0$

Applying Lemma 36 to $Z_{[-kL+1,0]}(N^+)$ gives (5.19). We now return to the proof of Proposition 22. We have

$$\begin{aligned} Z &= \lim_{k\to\infty} Z_{[-kL+1,0]} \\ &= \sup_{k\geq 0} Z_{[-kL+1,0]}(N) \\ &\leq \sup_{k\geq 0} Z_{[-kL+1,0]}(N^+) \quad \text{thanks to (5.18)} \\ &\leq \sup_{k\geq 0} \left(\hat{s}_0 + \sup_{-k+1\leq i\leq 0} \sum_{j=-i}^{-1} (\hat{s}_j - \hat{\tau}_{j+1}) \right) = \hat{R}, \quad \text{thanks to (5.19).} \end{aligned}$$

from Lemma 36. □

5.3.3 Proofs of the tail asymptotics

Recall that we defined

$$\Lambda(\theta) = \Lambda_T(-\theta) + \Lambda_Z(\theta).$$

Note that $\Lambda_Z(.)$ and $\Lambda_T(.)$ are proper convex functions, hence $\Lambda(.)$ is a well defined convex function. It is the scaled moment generating function of the process $\{Y_{[0,n]} := Z_{[-n,0]}(Q) + T_{-n} - T_0\}$ which satisfies the assumptions of the Part 5.2. Note in particular that by the monotonicity property, we have for $n \geq 0$,

$$Z_{[-n,0]}(Q) \geq Z_{[0,0]}(Q),$$

which directly implies Assumption **(A3)**. The fact that $\theta^* > 0$ follows directly from Lemma 28.

Lower Bound

Lemma 37. *Under previous assumptions, we have*

$$\liminf_{x\to\infty} \frac{1}{x} \log \mathbb{P}(Z > x) \geq -\theta^*.$$

Proof. We have (see Proposition 21)

$$Z \geq \sup_n \left\{ Z_{[-n,0]}(Q) + T_{-n} - T_0 \right\} = \sup_n Y_{[0,n]}. \tag{5.20}$$

Hence the lemma follows directly from Theorem 11. □

Upper bound

Lemma 38. *Under previous assumptions, we have*

$$\limsup_{x\to\infty} \frac{1}{x} \log \mathbb{P}(Z > x) \leq -\theta^*.$$

Proof. For L sufficiently large, we have with the convention $\sum_0^{-1} = 0$ (see Proposition 22),

$$Z \leq \sup_{n \geq 0} \left(\sum_{i=-n}^{-1} \hat{s}_i(L) - \hat{\tau}_{i+1}(L) \right) + \hat{s}_0(L) =: V(L) + \hat{s}_0(L).$$

We will show that under previous assumptions, we have

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}(V(L) + \hat{s}_0(L) > x) \leq -\theta_L, \tag{5.21}$$

where θ_L is defined as in (5.10) and the lemma will follow since $\theta_L \to \theta^*$ as L tends to infinity (see Lemma 27).

As in the proof of Lemma 29, for all $\theta \in (0, \theta_L)$, we have

$$\max \left\{ \mathbb{E}\left[e^{\theta \hat{s}_0(L)} \right], \mathbb{E}\left[e^{\theta V(L)} \right] \right\} < \infty.$$

Hence for $\theta \in (0, \theta_L)$, we have $\mathbb{E}\left[e^{\theta(W(L)+\hat{s}_0(L))} \right] = \mathbb{E}\left[e^{\theta W(L)} \right] \mathbb{E}\left[e^{\theta \hat{s}_0(L)} \right] \leq A$ for some finite constant A. Hence by Chernoff's inequality,

$$\mathbb{P}\left(W(L) + \hat{s}_0(L) \geq x \right) \leq e^{-\theta x} \mathbb{E}\left[e^{\theta(W(L)+\hat{s}_0(L))} \right] \leq A e^{-\theta x}.$$

Since the above holds for all $0 < \theta < \theta^L$, we get

$$\limsup_{x \to \infty} \frac{1}{x} \log \mathbb{P}\left(W(L) + \hat{s}_0(L) \geq x \right) \leq -\theta^L.$$

□

5.4 Case of study I : (max, plus)-linear systems

5.4.1 (Max, plus)-linear systems and monotone-separable networks

We now study in more details a specific class of monotone-separable networks.

Framework

The (max, plus) semi-ring $\mathbb{R}_{\max}$ is the set $\mathbb{R} \cup \{-\infty\}$, equipped with $\max$, written additively (i.e., $a \oplus b = \max(a, b)$) and the usual sum, written multiplicatively (i.e., $a \otimes b = a + b$). The zero element is $-\infty$.

For matrices of appropriate sizes, we define $(A \oplus B)^{(i,j)} = A^{(i,j)} \oplus B^{(i,j)} := \max(A^{(i,j)}, B^{(i,j)})$, $(A \otimes B)^{(i,j)} = \bigoplus_k A^{(i,k)} \otimes B^{(k,j)} := \max_k (A^{(i,k)} + B^{(k,j)})$.

Let s and m be arbitrary fixed natural numbers such that $m \leq s$. We assume that two matrix-valued maps $\mathcal{A}$ and $\mathcal{B}$ are given :

$$\begin{array}{rccc} \mathcal{A}: & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,s)}\left(\mathbb{R}_{\max}\right) \\ & \zeta = (\zeta^{(1)}, \ldots, \zeta^{(m)}) & \mapsto & \mathcal{A}(\zeta), \end{array}$$

$$\begin{array}{rccc} \mathcal{B}: & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,1)}\left(\mathbb{R}_{\max}\right) \\ & \zeta = (\zeta^{(1)}, \ldots, \zeta^{(m)}) & \mapsto & \mathcal{B}(\zeta), \end{array}$$

where the matrix $\mathcal{A} = \mathcal{A}(\zeta)$ has the following block structure :

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}(1,1) & | & -\infty & | & -\infty & | & -\infty \\ - & - & - & - & - & - & - \\ \mathcal{A}(2,1) & | & \mathcal{A}(2,2) & | & -\infty & | & -\infty \\ - & - & - & - & - & - & - \\ & \vdots & & \vdots & & \vdots & \\ - & - & - & - & - & - & - \\ \mathcal{A}(d,1) & | & \mathcal{A}(d,2) & | & & | & \mathcal{A}(d,d) \end{pmatrix},$$

where each $\mathcal{A}(\ell,\ell)$ is an irreducible matrix.

The (max, plus)-linear system associated to $\mathcal{A}$ and $\mathcal{B}$

Given a marked point process $N = \{(T_n, \zeta_n)\}_{-\infty<n<\infty}$, with $\zeta_n = (\zeta_n^{(1)}, \dots, \zeta_n^{(m)}) \in \mathbb{R}_+^m$, we can define the sequence of matrices $\{A_n\}$ and $\{B_n\}$ by

$$A_n := \mathcal{A}(\zeta_n), \quad B_n := \mathcal{B}(\zeta_n).$$

To the sequences $\{A_n\}$, $\{B_n\}$, and $\{T_n\}$, we associate the following (max, plus)-linear recurrence :

$$\mathcal{X}_{n+1} = A_{n+1} \otimes \mathcal{X}_n \oplus B_{n+1} \otimes T_{n+1}, \tag{5.22}$$

where $\{\mathcal{X}_n, n \in \mathbb{Z}\}$ is a sequence of state variables of dimension s. The stationary solution to this equation is constructed as follows. We write

$$Y_{[m,n]} := \bigoplus_{m\le k\le n} D_{[k+1,n]} \otimes B_k \otimes T_k = \max_{m\le k\le n} \left(D_{[k+1,n]} \otimes B_k + T_k \right), \tag{5.23}$$

where for $k < n$, $D_{[k+1,n]} = \bigotimes_{j=n}^{k+1} A_j = A_n \otimes \cdots \otimes A_{k+1}$ and $D_{[n+1,n]} = E$, the identity matrix (the matrix with all its diagonal elements equal to 0 and all its non-diagonal elements equal to $-\infty$). It is easy to check that $Y_{[m,m]} = B_m \otimes T_m$, and for all $n \ge m$,

$$Y_{[m,n+1]} = A_{n+1} \otimes Y_{[m,n]} \oplus B_{n+1} \otimes T_{n+1}.$$

In view of (5.23), the sequence $\{Y_{[-n,0]}\}$ is non-decreasing in n, so that we can define the stationary solution of (5.22),

$$Y_{(-\infty,0]} := \lim_{n\to\infty} Y_{[-n,0]} \le \infty.$$

The mapping $N = \{(T_n, \zeta_n)\} \mapsto X_{[m,n]}(N) = \bigoplus_{1\le i\le s} Y_{[m,n]}^{(i)}$ defines a stochastic network. By definition the $[m,n]$ maximal dater is

$$Z_{[m,n]}(N) = \bigoplus_{1\le i\le s} Y_{[m,n]}^{(i)} - T_n.$$

We give in the next section the assumptions on $\mathcal{A}$ and $\mathcal{B}$ under which this network is monotone-separable.

Conditions for a monotone-separable network

We now give the assumptions on $\mathcal{A}$ and $\mathcal{B}$:

(MS1) For all i, there exists k such that $\zeta^i = \mathcal{A}^{(k,k)}(\zeta)$. And each submatrix $\mathcal{A}(\ell,\ell)$ has at least one diagonal coefficient which is not $-\infty$.

(MS2) Any coefficient $\mathcal{A}^{(i,j)}(\zeta)$ or $\mathcal{B}^{(i)}(\zeta)$ that is not $-\infty$ is of the form :

$$\bigoplus_u \bigotimes_{k\in\mathcal{K}_u} \zeta^{(k)},$$

for some sets $\mathcal{K}_u \subset [1,m]$.

(MS3) We have for all $\zeta \in \mathbb{R}_+^m$,

$$\mathcal{A}(\zeta)\otimes \mathbf{0} = \mathcal{B}(\zeta)\oplus \mathbf{0},$$

where **0** is the vector with all its entries equal to 0.

We stress that any FIFO event graph with a single input fits into our framework ; see [11] for details on this class.

Note that the random sequence of matrices $\{A_n, B_n\}$ has fixed structure, i.e. for each i, j, $A_n^{(i,j)}$ (resp. $B_n^{(i)}$) is equal to $-\infty$ for all n or is non-negative for all n. Moreover, each irreducible matrix $\mathcal{A}(\ell,\ell)$ is aperiodic, i.e. there exists $N < \infty$ such that $\mathcal{A}(\ell,\ell)^N$ has all entries finite, because of Assumption **(MS1)**.

The following lemma shows that the conditions above define a monotone-separable network.

Lemma 39. *The network associated with a (max,plus)-linear recurrence is monotone-separable provided $\{A_n, B_n\}$ has fixed structure and $A_n\otimes \mathbf{0} \le B_n \oplus \mathbf{0}$ for all n.*

Proof.

The first three properties are immediate. Let us prove that separability holds. If $X_{[m,l]}(N) \le T_{l+1}$, then $Y_{[m,l]} \le \mathbf{0}\otimes T_{l+1}$.

So by monotonicity,

$$\begin{aligned} A_{l+1}\otimes Y_{[m,l]} &\le A_{l+1}\otimes \mathbf{0}\otimes T_{l+1} \\ &\le B_{l+1}\otimes T_{l+1} \oplus \mathbf{0}\otimes T_{l+1}. \end{aligned}$$

Hence we have

$$\begin{aligned} A_{l+1}\otimes Y_{[m,l]} \oplus B_{l+1}\otimes T_{l+1} &\le B_{l+1}\otimes T_{l+1}\oplus \mathbf{0}\otimes T_{l+1} \\ Y_{[m,l+1]} &\le Y_{[l+1,l+1]} \oplus \mathbf{0}\otimes T_{l+1}. \end{aligned} \tag{5.24}$$

But $\max_i B_{l+1}^{(i)} \ge 0$, hence we have $\max_i Y_{[l+1,l+1]}^{(i)} \ge T_{l+1}$. And then

$$X_{[m,l+1]}(N) = \max_i Y_{[m,l+1]}^{(i)} \le \max_i Y_{[l+1,l+1]}^{(i)} = X_{[l+1,l+1]}(N).$$

We show by induction that for all $n \ge l+1$,

$$Y_{[m,n]} \le Y_{[l+1,n]} \oplus \mathbf{0}\otimes T_{l+1}. \tag{5.25}$$

In view of (5.24), it is true for $n = l + 1$. Suppose it is true for n, then we have by monotonicity,

$$\begin{array}{rcl} A_{n+1} \otimes Y_{[m,n]} & \leq & A_{n+1} \otimes Y_{[l+1,n]} \oplus B_{n+1} \otimes T_{l+1} \oplus \mathbf{0} \otimes T_{l+1} \\ Y_{[m,n+1]} & \leq & Y_{[l+1,n+1]} \oplus \mathbf{0} \otimes T_{l+1}, \quad \text{since } T_{n+1} \geq T_{l+1}. \end{array}$$

Now taking the maximum over the indices in (5.25) gives $X_{[m,n]}(N) \leq X_{[l+1,n]}(N)$, but the converse inequality is clearly true in view of the definition of the mapping $X(.)$. Hence we have finally

$$X_{[m,n]}(N) = X_{[l+1,n]}(N).$$

□

5.4.2 Tail asymptotics for (max,plus)-linear networks

We consider now a (max,plus)-linear network as described in the above section (which is a monotone-separable network). We assume moreover that the stochastic assumptions of Section 5.3.1.0 are valid. Namely stability holds and we can define the *stationary maximal dater* by

$$0 \leq Z := Z_{(-\infty,0]} = \bigoplus_{1 \leq i \leq s} Y^{(i)}_{(-\infty,0]} - T_0 < \infty. \tag{5.26}$$

Moreover the sequence $\{\zeta_n\}$ is a sequence of i.i.d. random variables and we make the additional assumption that each component of the vector ζ_n is independent of each other and that for all i, for θ in a neighborhood of the origin,

$$\mathbb{E}\left[e^{\theta \zeta_0^{(i)}}\right] < \infty.$$

Note that we have $Z_0 = \bigoplus_{i=1}^s B_0^{(i)} \leq \zeta_0^{(1)} + \cdots + \zeta_0^{(d)}$, hence this ensures that assumption **(AZ)** holds.

Theorem 14. *Let Z be the stationary maximal dater of a (max,plus)-linear network. Associated to the irreducible matrices $\{A_n(\ell,\ell)\}$, we define the following function :*

$$\Lambda_\ell(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta (A_n(\ell,\ell)\otimes \cdots \otimes A_1(\ell,\ell))^{(u,v)}}\right],$$

where the limit exists in $\mathbb{R} \cup \{\infty\}$ and is independent of u, v. Then we have

$$\lim_{x\to\infty} \frac{1}{x} \log \mathbb{P}(Z > x) = -\theta^* < 0, \tag{5.27}$$

where $\theta^* = \min\{\theta^\ell\}$ and the θ^ℓ's are defined as follows

$$\theta^\ell = \sup\{\theta > 0,\ \Lambda_\ell(\theta) + \Lambda_T(-\theta) < 0\}.$$

In a queueing context, the sequence of matrices $\{A_n(\ell,\ell)\}$ corresponds to a specific ”component” of the network. It is well-known that the stability of such a network is constraint by the ”slowest” component. Here we see that in a large deviations regime, the ”bad” behavior of the network is due to a "bottleneck" component (which is not necessarily the same as the "slowest" component in average).

The computation of the function $\Lambda_\ell(\theta)$ is not easy in general and will not be discussed here. One practical question of interest would be to find good ways to estimate this function from the statistics made on the traffic. We should stress that we made the assumptions that each component of the vector ζ_n are independent of each other. This is of course not required to get the asymptotics (5.27), however removing this assumptions will change the moment generating function Λ_Z (given here in Lemma 42) and hence the value of θ^* (see the example below). Note that if one removes the assumptions of independence (in n) of the sequence of matrices (A_n, B_n), it is still possible to get some results. In [68], specific techniques on gaussian processes allow to get some asymptotics when the sequence ζ_n is driven by a fractional Brownian motion.

Example 6. *Consider the (max,plus)-linear recursion associated with the following sequence of matrices :*

$$A_n = \begin{pmatrix} \boxed{\zeta_n^{(1)}} & -\infty & -\infty & -\infty \\ \zeta_n^{(1+2)} & \boxed{\zeta_n^{(2)}} & -\infty & -\infty \\ \zeta_n^{(1+3)} & -\infty & \boxed{\zeta_n^{(3)}} & -\infty \\ \zeta_n^{(1+2\oplus 3)} & \zeta_n^{(2)} & \zeta_n^{(3)} & \boxed{0} \end{pmatrix}, B_n = \begin{pmatrix} \zeta_n^{(1)} \\ \zeta_n^{(1+2)} \\ \zeta_n^{(1+3)} \\ \zeta_n^{(1+2\oplus 3)} \end{pmatrix}, \tag{5.28}$$

where we used the shorthand notations, $\zeta_n^{(i+j)} = \zeta_n^{(i)} + \zeta_n^{(j)}$ and $\zeta_n^{(i+j\oplus k)} = \zeta_n^{(i)} + \max\{\zeta_n^{(j)}, \zeta_n^{(k)}\}$.

It is clear that these matrices satisfy the required assumptions to belong to the monotone-separable framework. We refer to Section 2.2.4 to see that this system corresponds to a tree queueing network.

The associated irreducible matrices are of size one and boxed in (5.28), hence we have for $\ell = 1, 2, 3$,

$$\Lambda_\ell(\theta) = \log \mathbb{E}\left[e^{\theta \zeta_1^{(\ell)}}\right],$$

and $\Lambda_4(\theta) = 0$. Hence for $\ell = 1, 2, 3$, θ^ℓ corresponds to the exponential rate of decay for the supremum of the random walk : $\sup_n \sum_{i=0}^n \left(\zeta_n^{(\ell)} - \tau_n\right)$, i.e. the stationary workload of a single server queue with arrival process N and service times given by the sequence $\{\zeta_n^{(\ell)}\}_n$. As a special case if each $\zeta^{(\ell)}$ has the same exponential distribution with mean $1/\mu$ and if the arrival process is Poisson with rate $\lambda < \mu$, we have $\theta^\ell = \mu - \lambda = \theta^*$.

Now assume that we have $\zeta_n^{(1)} = \zeta_n^{(2)} = \zeta_n^{(3)}$ for each n and the sequence $\{\zeta_n^{(1)}\}$ is a sequence of i.i.d random variables exponentially distributed with mean $1/\mu$. We have the same marginal probabilities as above but we are clearly not anymore in the framework of Theorem 14. However the system is still monotone separable and we can apply results from Part 2 but we have to compute the moment generating function Λ_Z corresponding to these stochastic assumptions. In this simple case, it is easy to see that

$$\Lambda_Z(\theta) = \log \frac{\mu}{\mu - \theta},$$

for $\theta < \mu/2$ and $\Lambda_Z(\theta) = \infty$ otherwise. Hence if we assume that the arrival process is Poisson *with rate $\lambda < \mu$, then we have*

$$\begin{aligned} \lambda \leq \mu/2 &\Rightarrow \theta^* = \mu/2, \\ \lambda > \mu/2 &\Rightarrow \theta^* = \mu - \lambda. \end{aligned}$$

In particular note that in the case $\lambda \leq \mu/2$, the condition on the tail (2) of Corollary 3 fails whereas Theorem 11 still holds. For small values of λ, the tail of the sojourn time is determined by the total service requirement of a single customer.

5.4.3 Computation of the moment generating function

Auxiliary result

Lemma 40. *We have*

$$Z_{[0,n]}(Q) = \bigoplus_{1\leq i\leq s} \left(D_{[1,n]} \otimes B_0\right)^{(i)}.$$

Proof.

From the definition, we have

$$Z_{[0,n]}(Q) = \bigoplus_{1\leq i\leq s} \bigoplus_{0\leq k\leq n} \left(D_{[k+1,n]} \otimes B_k\right)^{(i)}.$$

We will prove that for all $0 \leq k \leq n$,

$$\bigoplus_{1\leq i\leq s} \left(D_{[1,n]} \otimes B_0\right)^{(i)} \geq \bigoplus_{1\leq i\leq s} \left(D_{[k+1,n]} \otimes B_k\right)^{(i)}, \tag{5.29}$$

from which the lemma follows.

We have

$$\begin{aligned} B_0 &\geq \mathbf{0} \\ A_1 \otimes B_0 &\geq A_1 \otimes \mathbf{0} = B_1 \oplus \mathbf{0}, \end{aligned}$$

iterating we get

$$D_{[1,n]} \otimes B_0 \geq \mathbf{0} \oplus B_n \oplus A_n \otimes B_{n-1} \oplus D_{[n-1,n]} \otimes B_{n-2} \oplus \cdots \oplus D_{[2,n]} \otimes B_1.$$

Taking the supremum of all the components of the vector gives (5.29). □

(Max,plus) algebra and computation of the moment generating function

We begin with a general result showing the existence of the function Λ_ℓ. Let $\{M_n\}$ be an i.i.d. sequence of irreducible aperiodic (max,plus)-matrices with fixed structure. We denote

$$M_{[1,n]}^{(i,j)} = \left(M_n \otimes \cdots \otimes M_1\right)^{(i,j)}.$$

Lemma 41. *The following limit exists in $\mathbb{R} \cup \{+\infty\}$ and is independent of i and j,*

$$\Lambda_M(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta M_{[1,n]}^{(i,j)}}\right].$$

Proof.
We denote

$$\Lambda^{(i,j)}(\theta,n) \quad = \quad \log \mathbb{E}\left[e^{\theta M^{(i,j)}_{[1,n]}}\right].$$

We first take $\theta \geq 0$. We have

$$\begin{aligned}
\Lambda^{(i,j)}(\theta,n+m) &= \log \mathbb{E}\left[e^{\theta M^{(i,j)}_{[1,n+m]}}\right] \\
&= \log \mathbb{E}\left[\max_k e^{\theta M^{(i,k)}_{[n+1,n+m]}} e^{\theta M^{(k,j)}_{[1,n]}}\right] \\
&\geq \max_k \left\{\log \mathbb{E}\left[e^{\theta M^{(i,k)}_{[n+1,n+m]}}\right] + \log \mathbb{E}\left[e^{\theta M^{(k,j)}_{[1,n]}}\right]\right\} \\
&= \max_k \left\{\Lambda^{(i,k)}(\theta,m) + \Lambda^{(k,j)}(\theta,n)\right\}.
\end{aligned}$$

In particular for $j = i$, we have

$$\Lambda^{(i,i)}(\theta,n+m) \quad \geq \quad \Lambda^{(i,i)}(\theta,m) + \Lambda^{(i,i)}(\theta,n).$$

Moreover thanks to the fixed structure assumption, there exists N such that for $n \geq N$, we have $M^{(i,j)}_{[1,n]} > -\infty$ for all i and j, hence $\Lambda^{(i,j)}(\theta,n) > -\infty$ and we have

$$\lim_{n\to\infty} \frac{1}{n}\Lambda^{(i,i)}(\theta,n) = \sup_{n\geq N} \frac{1}{n}\Lambda^{(i,i)}(\theta,n) > -\infty.$$

For arbitrary i and j, choose $n, m \geq N$ and note that

$$\begin{aligned}
\Lambda^{(i,j)}(\theta,n+m) &\geq \Lambda^{(i,i)}(\theta,n) + \Lambda^{(i,j)}(\theta,m), \\
\Lambda^{(i,i)}(\theta,n+m) &\geq \Lambda^{(i,j)}(\theta,n) + \Lambda^{(j,i)}(\theta,m),
\end{aligned}$$

where all terms are in $\mathbb{R} \cup \{+\infty\}$. Letting $n \to \infty$ while keeping m fixed, it follows that

$$\lim_{n\to\infty} \frac{1}{n}\Lambda^{(i,j)}(\theta,n) = \lim_{n\to\infty} \frac{1}{n}\Lambda^{(i,i)}(\theta,n).$$

The arguments for the case $\theta < 0$ exactly parallels the one just given, but exploits (min,plus)-inequalities rather than (max,plus)-inequalities. □

We now compute $\Lambda_Z(\theta)$ for a (max,plus)-linear system. We introduce first some notations,

$$\begin{aligned}
D_{[1,n]}(\ell) &= \bigotimes_{j=n}^{1} A_j(\ell,\ell) = A_n(\ell,\ell) \otimes \cdots \otimes A_1(\ell,\ell) \\
\Lambda_\ell(\theta) &= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[e^{\theta D^{(i,j)}_{[1,n]}(\ell)}\right],
\end{aligned}$$

which does not depend on i and j as shown above.

Lemma 42. *We have for $\theta \geq 0$*

$$\Lambda_Z(\theta) = \sup_\ell \Lambda_\ell(\theta).$$

Proof.

The lower bound follows directly from the following inequality : for all ℓ, we have

$$\mathbb{E}\left[e^{\theta Z_{[0,n]}(Q)}\right] \quad \geq \quad \mathbb{E}\left[e^{\theta (D^{\ell}_{[1,n]})^{(i,j)}}\right].$$

We now derive the upper bound.

We first introduce some notations :

1. Let $\Upsilon = \{(n_1, \ldots, n_d) \in \mathbb{N}^d, n_1+n_2+\cdots+n_d = n+1\}$ and denote $n(i,j) = n_i+\cdots+n_j$ for $i \leq j$.
2. Let Δ_ℓ denotes the size of the irreducible matrix $A(\ell,\ell)$ and $\Delta(i,j) = \Delta_i + \cdots + \Delta_j$ for $i \leq j$.

We take the convention that if $i > j+1$ then $n(i,j) = 0$, $\Delta(i,j) = 0$ and $D^{(\alpha,\beta)}_{[i,j]} = 0$.

We decompose the product $D_{[1,n]} \otimes B_0$ as follows :

$$\begin{aligned}\bigoplus_{1\leq i\leq s} \left(D_{[1,n]} \otimes B_0\right)^{(i)} \quad = \quad & \max\Big\{ B_0^{(\alpha_0)} + D^{(\beta_1,\alpha_0)}_{[1,n_1-1]} + A^{(\alpha_1,\beta_1)}_{n_1} + \\ \ldots \quad & +D^{(\beta_{i+1},\alpha_i)}_{[n(1,i)+1,n(1,i+1)-1]} + A^{(\alpha_{i+1},\beta_{i+1})}_{n(1,i+1)} + \ldots \\ & +D^{(\beta_{d+1},\alpha_d)}_{[n(1,d-1)+1,n(1,d)]}\Big\},\end{aligned}$$

where the maximum is taken over all $(n_1, \ldots, n_d) \in \Upsilon$ and with the following constraints for the α_i's and β_i's : let j be the smallest integer such that $n_j \neq 0$, then

$$\begin{aligned}\alpha_0 &\in [\Delta(1,j-1)+1, \Delta(1,j)] \\ \alpha_\ell &\in [\Delta(1,\ell-1)+1, \Delta(1,\ell)],\ \forall \ell, \\ \beta_\ell &\in [\Delta(1,\ell)+1, \Delta(1,\ell+1)],\ \forall \ell.\end{aligned}$$

Hence we have

$$\begin{aligned}\bigoplus_{1\leq i\leq s} \left(D_{[1,n]} \otimes B_0\right)^{(i)} \quad \leq \quad & \max_{(n_1,\ldots,n_d)\in\Upsilon} \max\Big\{ \max_i B_0^{(i)} + D^{(\beta_1,\alpha_0)}_{[1,n_1-1]} + \max_{i,j} A^{(i,j)}_{n_1} + \\ \ldots \quad & +D^{(\beta_{i+1},\alpha_i)}_{[n(1,i)+1,n(1,i+1)-1]} + \max_{i,j} A^{(i,j)}_{n(1,i+1)} + \ldots \\ & +D^{(\beta_{d+1},\alpha_d)}_{[n(1,d-1)+1,n(1,d)]}\Big\},\end{aligned}$$

where the maximum is taken with the same constraints as above for the α_i's and β_i's. We can rewrite it as follows :

$$\begin{aligned}Z_{[1,n]}(Q) = \bigoplus_{1\leq i\leq s} \left(D_{[1,n]} \otimes B_0\right)^{(i)} \leq \\ \max_{(n_1,\ldots,n_d)\in\Upsilon} \max_i B_0^{(i)} + \sum_{\ell=1}^{d} \max_{i,j} A^{(i,j)}_{n(1,\ell)} + \max \sum_{i=1}^{d} D^{(\beta_{i+1},\alpha_i)}_{[n(1,i)+1,n(1,i+1)-1]}.\end{aligned}$$

Hence we have

$$\begin{aligned}
\mathbb{E}\left[e^{\theta Z_{[0,n]}(Q)}\right] &= \mathbb{E}\left[\exp\left\{\theta \max_{1\le i\le s}(D_{[1,n]}\otimes B_0)^{(i)}\right\}\right] \\
&\le \mathbb{E}\left[\max_{(n_1,\ldots,n_d)\in\Upsilon}\exp\theta\left\{\max_i B_0^{(i)}+\sum_{\ell=1}^{d}\max_{i,j}A_{n(1,\ell)}^{(i,j)}+\max\sum_{i=1}^{d}D_{[n(1,i)+1,n(1,i+1)-1]}^{(\beta_{i+1},\alpha_i)}\right\}\right] \\
&\le \sum_{(n_1,\ldots,n_d)\in\Upsilon}\mathbb{E}\left[\exp\theta\left\{\max_i B_0^{(i)}+\sum_{\ell=1}^{d}\max_{i,j}A_{n(1,\ell)}^{(i,j)}+\max\sum_{i=1}^{d}D_{[n(1,i)+1,n(1,i+1)-1]}^{(\beta_{i+1},\alpha_i)}\right\}\right] \\
&= \left(\sum_{(n_1,\ldots,n_d)\in\Upsilon}\mathbb{E}\left[\exp\left\{\theta\max\sum_{i=1}^{d}D_{[n(1,i)+1,n(1,i+1)-1]}^{(\beta_{i+1},\alpha_i)}\right\}\right]\right) \\
&\quad \underbrace{\mathbb{E}\left[\exp\theta\left\{\max_i B_0^{(i)}+\sum_{\ell=1}^{d}\max_{i,j}A_{\ell}^{(i,j)}\right\}\right]}_{\delta(\theta)},
\end{aligned}$$

where we used independence in the last equality. Now observe that $|\Upsilon| = \binom{n+1}{d} \le (n+1)^d$ and that

$$\max\sum_{i=1}^{d}D_{[n(1,i)+1,n(1,i+1)-1]}^{(\beta_{i+1},\alpha_i)} = \sum_{\ell=1}^{d}\max_{i,j}D_{[n(1,i)+1,n(1,i+1)-1]}^{(i,j)}(\ell).$$

Hence we have

$$\mathbb{E}\left[e^{\theta Z_{[0,n]}(Q)}\right] \le \delta(\theta)(n+1)^d \sup_{\{(n_1,\ldots,n_d)\in\Upsilon\}}\prod_{\ell=1}^{d}\mathbb{E}\left[\exp\theta\max_{i,j}D_{[1,n_\ell]}^{(i,j)}(\ell)\right]$$

and taking the $\log$, we obtain

$$\begin{aligned}
\log\mathbb{E}\left[e^{\theta Z_{[0,n]}(Q)}\right] &\le \log\left(\delta(\theta)(n+1)^d\right) \\
&\quad + \sup_{\{(n_1,\ldots,n_d)\in\Upsilon\}}\sum_{\ell=1}^{d}\log\mathbb{E}\left[\exp\theta\max_{i,j}D_{[1,n_\ell]}^{(i,j)}(\ell)\right].
\end{aligned}$$

Assume that $\Lambda_\ell(\theta) < \infty$ for all ℓ implies $\log(\delta(\theta)) < \infty$. Then for such θ, there are positive constants such that

$$\log\mathbb{E}\left[\exp\theta\max_{i,j}D_{[1,n_\ell]}^{(i,j)}(\ell)\right] \le n_\ell(\Lambda_\ell(\theta)+\epsilon)+K_\ell.$$

Hence in this case, we have

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\left[e^{\theta Z_{[0,n]}(Q)}\right] \le \sup_\ell\Lambda_\ell(\theta) < \infty.$$

We now show that $\Lambda_\ell(\theta) < \infty$ for all ℓ implies $\log(\delta(\theta)) < \infty$. In this case we have for all $i \in [1,m]$, $\mathbb{E}\left[e^{\theta\zeta_1^{(i)}}\right] < \infty$ thanks to **(A1)**, and then thanks to **(A2)**, we have

$$\max\left(\max_i B_1^{(i)}, \max_{\alpha,\beta}A_1^{(\alpha,\beta)}\right) \le \sum_{i=1}^{m}\zeta_1^{(i)},$$

hence the upper bound follows. □

5.5 Case of study II : Large Deviations for generalized Jackson networks

We first introduce some notations.

For $(E, d, \leq)$ a complete, separable metric space with partial order $\leq$, we denote by $\mathbb{D}(E)$ the space of cadlag non-decreasing E-valued functions defined on $\mathbb{R}_+$ with Skorohod (J_1) topology and by $\mathbb{C}(E)$ the space of continuous non-decreasing E-valued functions defined on $\mathbb{R}_+$. Restricted to $\mathbb{C}(E)$ the Skorohod topology is just the compact uniform topology.

For $x, y \in \mathbb{R}^K$, we write $x \leq y$ if $x^{(i)} \leq y^{(i)}$ for all i. We denote by $\wedge$ the minimum and by $\vee$ the maximum in $\mathbb{R}^K$. For $\mathbf{X}, \mathbf{Y} \in \mathbb{D}(\mathbb{R}_+^K)$, we write $\mathbf{X} \leq \mathbf{Y}$ if $\mathbf{X}(t) \leq \mathbf{Y}(t)$ for all $t \geq 0$ and for maps $F, G \in \mathbb{D}(\mathbb{R}_+^K)^2$, we denote $F \leq G$ if $F(\mathbf{X}) \leq G(\mathbf{X})$ for all $\mathbf{X} \in \mathbb{D}(\mathbb{R}_+^K)$. For $x \in \mathbb{R}^K$, we denote $\|x\| = \vee_{i=1}^K x^{(i)}$ and for $\mathbf{X} \in \mathbb{D}(\mathbb{R}_+^K)$, we denote $\|\mathbf{X}\| = \sup_t \|\mathbf{X}(t)\|$. We denote $\mathbb{D}_0(E) = \{f \in \mathbb{D}(E),\ f(0) = 0\}$ and $\mathbb{C}_0(E) = \{f \in \mathbb{C}(E),\ f(0) = 0\}$.

A piecewise linear function is a continuous function such that there exists a partition $\tau = (t_0 = 0 < t_1 < \dots)$ with $t_k \to \infty$ and such that the function is linear on each interval (t_k, t_{k+1}). For any function $f \in \mathbb{D}(\mathbb{R}_+^K)$, we define the polygonal approximation of f with step $1/n$ as the (piecewise linear) function

$$f_n(t) = f\left(\frac{\lfloor nt \rfloor}{n}\right) + (nt - \lfloor nt \rfloor)\left(f\left(\frac{\lfloor nt \rfloor + 1}{n}\right) - f\left(\frac{\lfloor nt \rfloor}{n}\right)\right)$$

$\mathbb{M}^K$ is the set of substochastic matrices of size $K \times K$. For $M \in \mathbb{M}^K$, we denote by $\rho(M)$ its spectral radius, by M^t its transpose and $M^{(i)}$ denotes the line $M^{(i)} = (M^{(i,1)}, \dots M^{(i,K)})$. In particular, we will identify a function $\mathbf{P} \in \mathbb{D}(\mathbb{M}^K)$ with its K components $\mathbf{P}^{(i)} \in \mathbb{D}(\mathbb{R}_+^K)$, where $\mathbf{P}^{(i)}(t) = (\mathbf{P}^{(i,1)}(t), \dots \mathbf{P}^{(i,K)}(t))$. Note that for $M, N \in \mathbb{M}^K$, we have $M \leq N$ if $M^{(i,j)} \leq N^{(i,j)}$ for all i and j.

We will use the Kullback-Leibler information divergence, which is a nonsymmetric measure of distance between distributions in the sense that for any two distributions P and R on $\mathcal{X}^k$ where $\mathcal{X}$ is a finite set,

$$D(P\|R) = \sum_{x \in \mathcal{X}^k} P(x) \log\left(\frac{P(x)}{R(x)}\right),$$

is nonnegative and equals 0 if and only if $P = R$. We use the standard notational conventions $\log 0 = -\infty$, $\log \frac{1}{0} = \infty$ and $0 \log 0 = 0 \log \frac{0}{0} = 0$. For any fixed R, the divergence $D(P\|R)$ is a continuous function of P restricted to $\{P,\ S(P) \subset S(Q)\}$ where $S(P)$ denotes the support of P.

For $P \in \mathbb{M}^K$, we denote by $\tilde{P}$ the $K \times (K+1)$ stochastic matrix obtained as follows : for all $i, j \leq K$, $\tilde{P}^{(i,j)} = P^{(i,j)}$ and $\tilde{P}^{(i,K+1)} = 1 - \sum_{k=1}^K P^{(i,k)}$. For $P, R \in \mathbb{M}^K$, we will denote

$$\begin{aligned}
\tilde{D}(P\|R) &:= D(\tilde{P}\|\tilde{R}) \\
&= \sum_{i,j \leq K} P^{(i,j)} \log\left(\frac{P^{(i,j)}}{R^{i,j)}}\right) + \sum_{i \leq K}\left(1 - \sum_k P^{(i,k)}\right) \log\left(\frac{1 - \sum_k P^{(i,k)}}{1 - \sum_k R^{(i,k)}}\right) \\
&= \sum_{i=1}^K \tilde{D}(P^{(i)}\|R^{(i)}).
\end{aligned}$$

5.5.1 Generalized Jackson networks

General setting and notation

We recall here the notation introduced in [12] to describe a generalized Jackson network with K nodes.

The networks we consider are characterized by the fact that service times and routing decisions are associated with stations and not with customers. This means that the j-th service on station k takes $\sigma_j^{(k)}$ units of time, where $\{\sigma_j^{(k)}\}_{j\geq 1}$ is a predefined sequence. In the same way, when this service is completed, the leaving customer is sent to station $\nu_j^{(k)}$ (or leaves the network if $\nu_j^{(k)} = K+1$) and is put at the end of the queue on this station, where $\{\nu_j^{(k)}\}_{j\geq 1}$ is also a predefined sequence, called the routing sequence. The sequences $\{\sigma_j^{(k)}\}_{j\geq 1}$ and $\{\nu_j^{(k)}\}_{j\geq 1}$, where k ranges over the set of stations, are called the driving sequences of the network. A generalized Jackson network will be defined by $\Big\{\{\sigma_j^{(k)}\}_{j\geq 1},\ \{\nu_j^{(k)}\}_{j\geq 1},\ n^{(k)},\ 0\leq k\leq K\Big\}$, where $(n^{(0)}, n^{(1)}, \ldots, n^{(K)})$ describes the initial condition. The interpretation is as follows : for $k\neq 0$, at time $t=0$, in node k, there are $n^{(k)}$ customers with service times $\sigma_1^{(k)}, \ldots, \sigma_{n^{(k)}}^{(k)}$ (if appropriate, $\sigma_1^{(k)}$ may be interpreted as a residual service time). In particular at time 0, the total number of customers in the network is $n^{(1,K)} = n^{(1)} + \ldots n^{(K)}$. Node 0 models the external arrival of customers in the network. Hence,

- if $n^{(0)}=0$ and $n^{(1,K)}$ is positive, there is a bulk arrival at time 0 of $n^{(j)}$ customers to node j, for all j, and no external arrival after time 0 ;
- if $\infty \geq n^{(0)} \geq 1$, then for all $1\leq j\leq n^{(0)}$, the arrival time of the $n^{(1,K)}+j$-th customer in the network takes place at $\sigma_1^{(0)} + \cdots + \sigma_j^{(0)}$ and it joins the end of the queue of station $\nu_{j+n^{(1,K)}}^{(0)}$. Hence $\sigma_j^{(0)}$ is the $n^{(1,K)}+j$-th inter-arrival time in the network.

In what follows, we will describe the driving sequences thanks to their associated counting functions. Consider the case $n^{(0)}=\infty$, we will use the following notation for each of these counting functions :

- $\sigma^{(k)}(1,n) = \sum_{j=1}^{n}\sigma_j^{(k)}$, for $0\leq k\leq K$;
- for $n\leq n^{(1,K)}$, we define $T_n=0$ and for $n>n^{(1,K)}$, we define $T_n = n^{(1,K)}+\sigma^{(0)}(1,n)$. Then T_n is the n-th exogenous arrival time in the network ;
- for $n\leq n^{(i)}$, we define $T_n^{(i)}=0$ and for $n>n^{(i)}$, we define $T_n^{(i)} = n^{(i)} + \sum_{j=1}^{k_n^{(i)}}\sigma_j^{(0)}\mathbb{1}_{\{\nu_j^{(0)}=i\}}$, with $k_n^{(i)} = \inf\{k,\ \sum_{j=1}^{k}\mathbb{1}_{\{\nu_j^{(0)}=i\}} \geq n-n^{(i)}\}$. $T_n^{(i)}$ is the n-th exogenous arrival time at node i.

We define the sequence of Jackson networks $\mathbf{JN}_n = \{\mathbf{S}_n(t), \mathbf{P}_n(t), \mathbf{N}_n(t)\}$ with

$$\begin{aligned}
\mathbf{N}_n^{(i)}(t) &= \frac{1}{n}\sum_k \mathbb{1}_{\{T_k^{(i)}\leq nt\}},\\
\mathbf{S}_n^{(i)}(t) &= \frac{1}{n}\sum_k \mathbb{1}_{\{\sigma^{(i)}(1,k)\leq nt\}},\\
\mathbf{P}_n^{(i,j)}(t) &= \frac{1}{n}\sum_{k\leq nt} \mathbb{1}_{\{\nu_k^{(i)}=j\}}.
\end{aligned}$$

Note that $\mathbf{N}_n^{(i)}(0) = n^{(i)}$ and that we have for all $0 \le u \le v$,

$$\sum_{j=1}^{K} \mathbf{P}_n^{(i,j)}(u) - \mathbf{P}_n^{(i,j)}(v) \le (u - v),$$

in particular, we have $\mathbf{P}_n \in \mathbb{D}_0(\mathbb{M}^K)$.

We denote the input and output processes of each queue k of the networks by $\mathbf{A}^{(k)}$ and $\mathbf{D}^{(k)}$ respectively. We will use the following notation $\mathbf{A} = (\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(K)})$ and $\mathbf{D} = (\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(K)})$. We now describe how the processes $\mathbf{A}$ and $\mathbf{D}$ are obtained form $\mathbf{JN}$.

We define the map $\Gamma : \mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K) \to \mathbb{D}(\mathbb{R}_+^K)$ as follows :

$$\Gamma(\mathbf{X}, \mathbf{P}, \mathbf{N})^{(i)}(t) := \mathbf{N}^{(i)}(t) + \sum_{j=1}^{K} \mathbf{P}^{(j,i)}(\mathbf{X}^{(j)}(t)).$$

The following lemma is straightforward.

Lemma 43. *The map Γ is continuous for the compact uniform topology and non-decreasing in its first argument.*

We define the map $\Phi : \mathbb{D}(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{R}_+^K) \to \mathbb{D}_0(\mathbb{R}_+^K)$ as follows :

$$\Phi(\mathbf{X}, \mathbf{Y})^{(i)}(t) := \inf_{0 \le s \le t} \left\{ \mathbf{Y}^{(i)}(t) - \mathbf{Y}^{(i)}(s) + \mathbf{X}^{(i)}(s) \right\} \wedge \mathbf{Y}^{(i)}(t).$$

Lemma 44. *The map Φ is continuous for the compact uniform topology and non-decreasing in its first argument.*

Proof. We can clearly consider the mapping Φ with $K = 1$ only. Let $\mathcal{R} : \mathbb{D}(\mathbb{R}) \to \mathbb{D}(\mathbb{R}_+)$ be the one-dimensional reflection map defined by $\mathcal{R}(\mathbf{X})(t) := \sup_{0 \le s \le t} \{\mathbf{X}(t) - \mathbf{X}(s)\} \vee \mathbf{X}(t)$. We have $\Phi(\mathbf{X}, \mathbf{Y}) = \mathbf{X} - \mathcal{R}(\mathbf{X} - \mathbf{Y})$. It is easy to see that for any $T > 0$,

$$\sup_{0 \le t \le T} |\mathcal{R}(\mathbf{X})(t) - \mathcal{R}(\mathbf{X}')(t)| \le 2 \sup_{0 \le t \le T} |\mathbf{X}(t) - \mathbf{X}'(t)|,$$

from which the continuity of Φ follows. Its monotonicity is obvious. □

Remark 18. Consider the mapping Φ with $K = 1$ and $\mathbf{Y}(t) = \mu t$, with $\mu \ge 0$. If $\mu = 0$, since $\Phi(\mathbf{X}, \mathbf{Y}) \le \mathbf{Y}$, we have $\Phi(\mathbf{X}, \mathbf{Y})(t) = 0$ for all t. If $\mu \ne 0$, we have $\Phi(\mathbf{X}, \mathbf{Y})(t) = \inf_{0 \le s \le t}\{\mathbf{X}(s) + \mu(t - s)\}$. Moreover if $\mathbf{X}$ is a concave function, then this equation reduces to $\Phi(\mathbf{X}, \mathbf{Y})(t) = \mathbf{X}(t) \wedge \mu t$. Hence we can write

$$\begin{aligned} \mathbf{Y}(t) = \mu t, \text{ with } \mu \ge 0 \quad &\Rightarrow \quad \Phi(\mathbf{X}, \mathbf{Y})(t) = \mu t \wedge \inf_{0 \le s \le t}\{\mathbf{X}(s) + \mu(t - s)\}, \\ \text{if moreover } \mathbf{X} \text{ is a concave function} \quad &\Rightarrow \quad \Phi(\mathbf{X}, \mathbf{Y})(t) = \mu t \wedge \mathbf{X}(t). \end{aligned}$$

Proposition 2.1 of [67] shows that the following fixed-point equation :

$$\left\{ \begin{array}{lcl} \mathbf{A}_n & = & \Gamma(\mathbf{D}_n, \mathbf{P}_n, \mathbf{N}_n) = \Gamma(\mathbf{D}_n, \mathbf{JN}_n), \\ \mathbf{D}_n & = & \Phi(\mathbf{A}_n, \mathbf{S}_n) = \Phi(\mathbf{A}_n, \mathbf{JN}_n), \end{array} \right. \tag{5.30}$$

has an unique solution when each component of $n\mathbf{S}_n$, $n\mathbf{P}_n$ and $n\mathbf{N}_n$ are counting functions (i.e. non-decreasing functions of $\mathbb{D}(\mathbb{R}_+^K)$ or $\mathbb{D}(\mathbb{M}^K)$ that are piece-wise constant with jumps of size one). In this case the corresponding functions $n\mathbf{A}_n$ and $n\mathbf{D}_n$ are also counting functions and we denote the solution of (5.30) by $\Psi(\mathbf{S}_n, \mathbf{P}_n, \mathbf{N}_n) = \Psi(\mathbf{JN}_n)$.

Stochastic assumptions

In what follows, it will be important to distinguish the nodes of the network that do not receive any exogenous customer, i.e. the nodes $i \notin \mathcal{S}$ with $\mathcal{S} = \{i,\ \exists k \geq 1,\ \nu_k^{(0)} = i\}$. A generalized Jackson network $\mathbf{JN} = \{\mathbf{S}, \mathbf{P}, \mathbf{N}\}$ is an object in $\mathcal{E} \subset \mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K)$, with the additional constraint $\mathbf{N}^{(i)}(t) = 0$ for all t, for $i \notin \mathcal{S}$. Note that $\mathcal{E}$ is closed in $\mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K)$.

We define for $(s^{(1)}, \dots, s^{(K)}) \in \mathbb{R}_+^K$ and $(n^{(1)}, \dots, n^{(K)}) \in \mathbb{R}_+^K$, the functions

$$\begin{aligned} I^{\mathbf{S}}(s^{(1)}, \dots, s^{(K)}) &= \sum_{i=1}^K I^{\mathbf{S}^{(i)}}(s^{(i)}), \\ I^{\mathbf{N}}(n^{(1)}, \dots, n^{(K)}) &= \sum_{i \in \mathcal{S}} I^{\mathbf{N}^{(i)}}(n^{(i)}) + \infty \mathbb{1}_{\{n^{(i)}>0,\ i \notin \mathcal{S}\}}, \end{aligned}$$

where each $I^{\mathbf{S}^{(i)}} : \mathbb{R}_+ \mapsto \mathbb{R}_+ \cup \{+\infty\}$ (resp. $I^{\mathbf{N}^{(i)}}$ for $i \in \mathcal{S}$) is a $[0, \infty]$-valued convex good rate function, attaining zero on $\mathbb{R}_+$ admitting a unique minimum at the point $\mu^{(i)}$ (resp. $\lambda^{(i)}$ for $i \in \mathcal{S}$) and with a domain open on the right.

We assume that the sequence $\mathbf{JN}_n = \{\mathbf{S}_n(t), \mathbf{P}_n(t), \mathbf{N}_n(t)\}$ satisfies a LDP in the space $\mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K)$ with a good rate function $I^{\mathbf{JN}}$ given by

$$I^{\mathbf{JN}}(\mathbf{S}, \mathbf{P}, \mathbf{N}) := I^0(\mathbf{N}(0)) + \int_0^\infty I^{\mathbf{S}}(\dot{\mathbf{S}}(t)) + D(\dot{\mathbf{P}}(t) \| R) + I^{\mathbf{N}}(\dot{\mathbf{N}}(t)) dt, \tag{5.31}$$

if the argument functions are absolutely continuous and equal to infinity otherwise.

Assumptions on the matrix R :

1. We assume that $\rho(R) < 1$.
2. We assume that for all $1 \leq i \leq K$, we have

$$(\mathcal{N} + \mathcal{N}R + \cdots + \mathcal{N}R^K)^{(i)} > 0, \tag{5.32}$$

   where $\mathcal{N}^{(i)} = \mathbb{1}_{\{i \in \mathcal{S}\}}$.

We recall here some results of [78] concerning large deviations of renewal processes and show that our assumptions on the rate function (5.31) are satisfied in the i.i.d case. Denote by $\{\zeta_i,\ i \geq 1\}$ a sequence of non-negative i.i.d. random variables with positive mean. Let

$$\begin{aligned} \alpha(\theta) &= \log \mathbb{E}\left[e^{\theta \zeta_1}\right], \\ \theta^* &= \sup\{\theta > 0,\ \alpha(\theta) < \infty\}, \\ \alpha^*(x) &= \sup_\theta \{\theta x - \alpha(\theta)\} = \sup_{\theta < \theta^*} \{\theta x - \alpha(\theta)\}, \\ g(x) &= x\alpha^*(1/x) = \sup_{\theta < \theta^*} \{\theta - x\alpha(\theta)\}. \end{aligned}$$

The function α is a convex function and differentiable on $(-\infty, \theta^*)$ with $\alpha'(\theta) = \mathbb{E}[\zeta_1] > 0$. In particular, we have $\lim_{\theta \uparrow \theta^*} \alpha(\theta) = \infty$. Thanks to [49], we know that α^* and g are convex rate functions. Moreover if $\theta^* > 0$ then α^* is a good rate function. Introduce the sequence of processes $\{\mathbf{C}_n\}_n$:

$$\mathbf{C}_n(t) = \frac{1}{n} \sum_i \mathbb{1}_{\{\sum_{j=1}^i \zeta_j \leq nt\}}.$$

Then Theorem 3.1 of [78] gives : If $\mathbb{P}(\zeta_1 > 0) = 1$, then the sequence $\{\mathbf{C}_n\}_n$ satisfies a LDP in $\mathbb{D}(\mathbb{R}_+)$ with the good rate function

$$I^{\mathbf{C}}(\mathbf{x}) = \begin{cases} \int_0^\infty g(\dot{\mathbf{x}}(t))dt, & \text{if } \mathbf{x} \in \mathbb{C}(\mathbb{R}_+) \text{ tends to infinity and is absolutely continuous,} \\ \infty, & \text{otherwise.} \end{cases}$$

General methodology

As in [67], we see a generalized Jackson network as a solution of the fixed point equation

$$\begin{cases} \mathbf{A} &= \Gamma(\mathbf{D}, \mathbf{JN}), \\ \mathbf{D} &= \Phi(\mathbf{A}, \mathbf{JN}), \end{cases} \quad \Leftrightarrow \quad (\mathbf{A}, \mathbf{D}) =: \Psi(\mathbf{JN}). \tag{5.33}$$

It is known that Ψ is well defined for counting processes, see [26] or [67]. It is natural to ask whether Ψ is well defined for processes in $\mathbb{D}$ or at least for absolutely continuous processes. If this was true, and if Ψ was shown to be continuous, then we would get thanks to the contraction principle that the process $(\mathbf{A}, \mathbf{D})$ satisfies a LDP with good rate function

$$\text{"} I^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D}) = \inf \left\{ I^{\mathbf{JN}}(\mathbf{JN}),\ \Psi(\mathbf{JN}) = (\mathbf{A}, \mathbf{D}) \right\}. \text{"} \tag{5.34}$$

However, the map Ψ turns out not to be well defined for all possible limits of a sequence of Jackson networks $\{\mathbf{JN}_n\}_n$ as defined previously. In particular, the fixed point equation (5.33) can very well be stated for processes in $\mathbb{D}$ but then may have several different solutions. We refer to the appendix for a simple example.

To circumvent this difficulty, we adopt the following strategy. We find a domain $\mathcal{D}_{\mathbf{JN}} \subset \mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K)$ satisfying the following constraints :

– the map Ψ is well defined on $\mathcal{D}_{\mathbf{JN}}$;
– any solution $(\mathbf{A}, \mathbf{D})$ of the fixed point equation (5.33) associated with a "continuous" Jackson network $\mathbf{JN}$ can be approximated by a sequence $\{\mathbf{JN}_n\} \in \mathcal{D}_{\mathbf{JN}}^{\mathbb{N}}$ such that

$$\begin{aligned} \mathbf{JN}_n &\to \mathbf{JN}, \\ \Psi(\mathbf{JN}_n) &\to (\mathbf{A}, \mathbf{D}), \\ I^{\mathbf{JN}}(\mathbf{JN}_n) &\to I^{\mathbf{JN}}(\mathbf{JN}). \end{aligned}$$

Hence in order to remove the quote from (5.34), we follow a quite standard method of proofs for large deviations of stochastic processes analogue with the theory of weak convergence : it consists of first verifying a compactness condition and then showing that there is only one possible limit. In our context, we proceed as follows :

1. we show that our sequence of processes is exponentially tight ;
2. we use $\mathcal{D}_{\mathbf{JN}}$ to determine the rate function.

In Section 5.5.2, we give the theoretical framework that shows how $\mathcal{D}_{\mathbf{JN}}$ determines the rate function. This result is stated in great generality (without any reference to our specific problem) and could be of independent interest since this method of proof could be applied to other dynamical systems, with discontinuous statistics.

5.5.2 An extension of the contraction principle

Let $\mathcal{E}, \mathcal{F}$ be complete separable metric spaces. Let $G : \mathcal{E} \times \mathcal{F} \to \mathbb{R}$ be a continuous function. We assume that there exists $\mathcal{D} \subset \mathcal{E}$, such that for all $x \in \mathcal{D}$, there exists an unique $y \in \mathcal{F}$ such that $G(x, y) = 0$. We denote it by, $y = H(x)$ where $H : \mathcal{D} \to \mathcal{F}$,

$$\forall x \in \mathcal{D}, \quad G(x, y) = 0 \iff y = H(x).$$

Proposition 23. *Let $\{X_n\}_n$ be a sequence of $\mathcal{E}$-valued random variables and $\{Y_n\}_n$ be a sequence of $\mathcal{F}$-valued random variables, where $\mathcal{E}$ and $\mathcal{F}$ are metric spaces. We assume that each sequence is exponentially tight.*

Assume that the sequence $\{X_n\}_n$ satisfies a LDP with good rate function I^X and that $G(X_n, Y_n) = 0$ a.s. for all n.

We assume that for all (x, y) such that $G(x, y) = 0$ and $I^X(x) < \infty$, there exists a sequence $x_n \to x$, such that $x_n \in \mathcal{D}$ for all n, $H(x_n) \to y$ and $I^X(x_n) \to I^X(x)$. We denote by $\mathcal{S}(x, y) = \{x_n\}_n$ this sequence. If $G(x, y) \neq 0$ or $I^X(x) = \infty$, we take $\mathcal{S}(x, y) = \emptyset$ and we denote $\mathcal{S}(y) = \cup_x \{\mathcal{S}(x, y)\}$.

Then the sequence $\{X_n, Y_n\}_n$ satisfies a LDP with good rate function :

$$I^{X,Y}(x, y) := \begin{cases} I^X(x), & G(x, y) = 0, \\ \infty, & \textit{otherwise.} \end{cases} \tag{5.35}$$

In particular, if $X_n \in \mathcal{D}$ for all n and if the sequence $\{H(X_n)\}_n$ is exponentially tight, then it satisfies a LDP in $\mathcal{F}$ with good rate function :

$$I^{H(X)}(y) := \inf\{\lim_{n\to\infty} I^X(x_n),\ \{x_n\}_n \in \mathcal{S}(y)\}. \tag{5.36}$$

Remark 19.

– There are alternative ways of expressing the rate function,

$$I^{H(X)}(y) = \inf\{I^X(x),\ y \in H^x\},$$

where $H^x := \{y \in \mathcal{F},\ \exists x_n \to x,\ H(x_n) \to y\}$. $I^{H(X)}$ is the lower semicontinuous regularization of the following function defined for $y \in H(\mathcal{D}) \subset \mathcal{F}$,

$$\tilde{I}^{H(X)}(y) := \inf\{I^X(x),\ y = H(x)\}.$$

– The main interest of the definition (5.36) is that the rate function is computed only thanks to the sequences $\mathcal{S}(x, y) \in \mathcal{D}^{\mathbb{N}}$.

– Note that if we assume that $H(\mathcal{D})$ is closed (in particular if $\mathcal{D} = \mathcal{E}$) then this proposition follows from the contraction principle (for an extensive discussion of this principle, see the work of Garcia [48]).

Proof. Thanks to Lemma 3.6 of [42], the sequence $\{X_n, Y_n\}$ is exponentially tight. Then by Theorem 3.7 of [42], there exists a subsequence $\{n_k\}$ along which the sequence $\{X_{n_k}, Y_{n_k}\}$ satisfies a LDP with a good rate function. If we can prove that there is a unique possible rate function (that does not depend on the subsequence $\{n_k\}$) then the proposition will follow.

Hence, for simplicity of notations, we still denote the extracted subsequence by $\{X_n, Y_n\}$ and we assume that $\{X_n, Y_n\}$ satisfies a LDP with good rate function $\tilde{I}^{X,Y}$. We will show that $\tilde{I}^{X,Y} = I^{X,Y}$ given by (5.35).

Consider the continuous mappings H_1 and H_2 from $\mathcal{E} \times \mathcal{F}$ to $\mathcal{E} \times \mathcal{F} \times \mathbb{R}$,

$$H_1(x,y) := (x,y,G(x,y)), \quad H_2(x,y) := (x,y,0).$$

We have clearly $H_1(X_n,Y_n) = H_2(X_n,Y_n)$ a.s. Moreover thanks to the contraction principle, $\{H_1(X_n,Y_n)\}_n$ and $\{H_2(X_n,Y_n)\}_n$ satisfy LDPs with the good rate functions

$$I^{H_1}(x,y,z) = \inf\{\tilde{I}^{X,Y}(x,y), z = G(x,y)\} \quad I^{H_2}(x,y,z) = \inf\{\tilde{I}^{X,Y}(x,y), z = 0\},$$

where $\inf \emptyset = \infty$. Since $H_1(X_n,Y_n) = H_2(X_n,Y_n)$, we have $I^{H_1} = I^{H_2}$. Now we have,

$$\tilde{I}^{X,Y}(x,y) = \inf_z\{I^{H_1}(x,y,z)\} = \inf\{\tilde{I}^{X,Y}(x,y), G(x,y) = 0\},$$

hence $\tilde{I}^{X,Y}(x,y) = \infty$ as soon as $G(x,y) \neq 0$. It remains to show that $G(x,y) = 0$ implies $\tilde{I}^{X,Y}(x,y) = I^X(x)$. We have clearly $I^X(x) \leq \tilde{I}^{X,Y}(x,y)$ for all (x,y) since $\{X_n\}$ satisfies a LDP with good rate function

$$I^X(x) = \inf\{\tilde{I}^{X,Y}(x,y),\ y \in \mathcal{F},\ G(x,y) = 0\}.$$

In particular, the definition of $\mathcal{D}$ implies $I^X(x) = \tilde{I}^{X,Y}(x,H(x))$ for $x \in \mathcal{D}$.

Take (x,y) such that $G(x,y) = 0$ and $I^X(x) < \infty$. There exists $x_n^* \to x$ with $x_n^* \in \mathcal{D}$, $H(x_n^*) \to y$ and $I^X(x_n^*) \to I^X(x)$. Thanks to the lower semicontinuity property of $\tilde{I}^{X,Y}$, we can find for any $\delta > 0$, an $\epsilon > 0$ such that

$$\frac{1}{\delta} \wedge \left(\tilde{I}^{X,Y}(x,y) - \delta\right) \leq \inf_{z \in B(y,\epsilon)} \tilde{I}^{X,Y}(x,z),$$

where $B(y,\epsilon)$ is the closed ball in $\mathcal{F}$ of center y and radius ϵ.

Thanks to the lower semicontinuity of the function $x \mapsto \inf_{z \in B(y,\epsilon)} \tilde{I}^{X,Y}(x,z)$, we have

$$\begin{aligned}
\inf_{z \in B(y,\epsilon)} \tilde{I}^{X,Y}(x,z) &\leq \liminf_{x_n \to x} \inf_{z \in B(y,\epsilon)} \tilde{I}^{X,Y}(x_n,z) \\
&\leq \liminf_{n \to \infty} \inf_{z \in B(y,\epsilon)} \tilde{I}^{X,Y}(x_n^*,z) \\
&\leq \lim_{n \to \infty} I^X(x_n^*) = I^X(x),
\end{aligned}$$

because $H(x_n^*) \in B(y,\epsilon)$ for sufficiently large n. Hence we proved that for any $\delta > 0$, $\frac{1}{\delta} \wedge \left(\tilde{I}^{X,Y}(x,y) - \delta\right) \leq I^X(x)$ for (x,y) such that $G(x,y) = 0$ and $I^X(x) < \infty$, this concludes the proof of (5.35).

The various expressions of $I^{H(X)}$ are now quite easy to obtain from

$$I^{H(X)}(y) = \inf\{I^X(x),\ G(x,y) = 0\}. \tag{5.37}$$

For (5.36), note that since the set $\{x,\ G(x,y) = 0\}$ is closed the minimum in (5.37) (if it is finite) is attained for a certain x^* with $G(x^*,y) = 0$ and $I^X(x^*) < \infty$.

We prove now that

$$\inf\{I^X(x),\ y \in H^x\} = \inf\{I^X(x),\ G(x,y) = 0\}.$$

If $y \in H^x$, then there exists $x_n \to x$ such that $H(x_n) \to y$. Hence by continuity of G, we have $G(x, y) = 0$. Now if $G(x, y) = 0$ and $I^X(x) < \infty$, it follows from the assumptions that $y \in H^x$.

To see that the last assertion is true, we show that for any open set $O \subset \mathcal{F}$, we have,

$$\inf_{y \in O} I^{H(X)}(y) = \inf_{y \in O} \{I^X(x),\ y = H(x)\}. \tag{5.38}$$

For $y \in O$ and any x such that $G(x, y) = 0$, there exists $x_n \to x$, such that $H(x_n) \to y$ and $I^X(x_n) \to I^X(x)$. Hence for n sufficiently large, we have $H(x_n) \in O$ and then

$$\inf_{y \in O} \{I^X(x),\ y = H(x)\} \le \inf_n I^X(x_n) \le I^X(x).$$

Taking the minimum over all x such that $G(x, y) = 0$ gives the $\ge$ inequality in (5.38), the converse inequality is obvious. □

5.5.3 Extension of Ψ to piece-wise linear Jackson networks

In this section we consider processes that are continuous, i.e. in $\mathbb{C}(E)$, hence topological concepts refer to the compact uniform topology.

We first recall Proposition 3.2 of [67],

Proposition 24. *Given a $K \times K$ substochastic matrix P with $\rho(P) < 1$ and vectors $(\alpha, y) \in \mathbb{R}_+^{2K}$, the fixed point equation*

$$x^{(i)} \ = \ \alpha^{(i)} + \sum_{j=1}^{K} P^{(j,i)} \left(x^{(j)} \wedge y^{(j)} \right),$$

has a unique solution $x(y, P, \alpha)$. Moreover, $(y, \alpha) \mapsto x(y, P, \alpha)$ is a continuous non-decreasing function.

We first consider a linear Jackson network $\mathbf{JN}$ and show that the mapping Ψ is well defined for such a network. By linear, we mean the following $\mathbf{N}^{(i)}(t) = N^{(i)} + \lambda^{(i)} t$, with $\lambda^{(i)} \ge 0$ and $N^{(i)} \in \mathbb{R}_+$, $\mathbf{S}^{(i)}(t) = \mu^{(i)} t$, with $\mu^{(i)} \ge 0$, and $\mathbf{P}^{(i,j)}(t) = P^{(i,j)} t$. We assume that $\rho(P) < 1$.

Lemma 45. *Under previous assumptions, the fixed point equation (5.33) has an unique solution $\mathbf{X}_f[\mu, P, N, \lambda](t) = x(\mu t, P, N + \lambda t)$, where $\mu = (\mu^{(i)})_i$, $N = (N^{(i)})_i$ and $\lambda = (\lambda^{(i)})_i$.*

Proof. Since μ, P, N, λ are fixed here, we omit to explicitly write the dependence in these variables. In this case, the fixed point equation (5.33) reduces to (see Remark 18)

$$\left\{ \begin{array}{rcl} \mathbf{A}^{(i)}(t) & = & N^{(i)} + \lambda^{(i)} t + \sum_{j=1}^{K} P^{(j,i)} \mathbf{D}^{(j)}(t), \\ \mathbf{D}^{(i)}(t) & = & \mu^{(i)} t \wedge \inf_{0 \le s \le t} \{\mathbf{A}^{(i)}(s) + \mu^{(i)}(t - s)\}. \end{array} \right. \tag{5.39}$$

Thanks to Proposition 24, $\mathbf{X}_f(t) = x(\mu t, P, N + \lambda t)$ is the unique solution of the fixed point equation

$$\left\{ \begin{array}{rcl} \mathbf{A}^{(i)}(t) & = & N^{(i)} + \lambda^{(i)} t + \sum_{j=1}^{K} P^{(j,i)} \mathbf{D}^{(j)}(t), \\ \mathbf{D}^{(i)}(t) & = & \mathbf{A}^{(i)}(t) \wedge \mu^{(i)} t. \end{array} \right. \tag{5.40}$$

We prove now that $\mathbf{X}_f$ is the unique solution of the fixed point equation (5.39).

For simplicity, we denote the fixed point equation (5.39), resp. (5.40), by $\mathbf{A} = F(\mathbf{A})$, resp. by $\mathbf{A} = \tilde{F}(\mathbf{A})$. Note that these functions are non-decreasing, continuous and such that $F \leq \tilde{F}$.

From $\mathbf{0} \leq \mathbf{X}_f$, we get $\mathbf{0} \leq F(\mathbf{0}) \leq \tilde{F}(\mathbf{0}) \leq \tilde{F}(\mathbf{X}_f)$. Hence $F^n(\mathbf{0}) \nearrow \mathbf{L} \leq \mathbf{X}_f$ and $F(\mathbf{L}) = \mathbf{L}$. Moreover for any solution $\mathbf{Y}$ of the fixed point equation (5.39), we have $\mathbf{L} \leq \mathbf{Y} \leq \mathbf{X}_f$ because $\mathbf{Y} = F(\mathbf{Y}) \leq \tilde{F}(\mathbf{Y})$ and $\tilde{F}^n(\mathbf{Y}) \nearrow \mathbf{X}_f$.

Since $\mathbf{0}$ is a concave function, we have $F(\mathbf{0}) = \tilde{F}(\mathbf{0})$ and hence it is still a concave function. Hence we have $\tilde{F}^n(\mathbf{0}) = F^n(\mathbf{0})$ since the image by $\tilde{F}$ of a concave function is a concave function and $F = \tilde{F}$ on the subspace of concave functions. Hence we have $\mathbf{L} = \mathbf{X}_f$ which concludes the proof. $\square$

In order, to extend Ψ to piece-wise linear Jackson networks, we proceed step by step on each interval where the driving functions $\mathbf{S}, \mathbf{P}, \mathbf{N}$ are linear. The following lemma allows to glue the constructed solution on each adjacent interval. In a queueing context, this lemma says that the output of a single server queue fed by the arrival process $\mathbf{A}$ and service time process $\mathbf{S}$ viewed from time u is just the same as the output process of a single server queue that we start at time u with arrival process $\tilde{\mathbf{A}}(t) = \mathbf{A}(t+u) - \mathbf{A}(u) + \mathbf{A}(u) - \mathbf{D}(u)$ (i.e. with the same increment as the original process on this period of time plus an additional bulk corresponding to the queue length at time u) and with service time process $\tilde{\mathbf{S}}(t)$.

Lemma 46. *Let $\mathbf{A}, \mathbf{S} \in \mathbb{D}(\mathbb{R}_+) \times \mathbb{D}_0(\mathbb{R}_+)$ and $\mathbf{D} = \Phi(\mathbf{A}, \mathbf{S})$. Define $\tilde{\mathbf{A}}, \tilde{\mathbf{S}} \in \mathbb{D}(\mathbb{R}_+) \times \mathbb{D}_0(\mathbb{R}_+)$ as follows*

$$\begin{aligned}\tilde{\mathbf{A}}(t) &:= \mathbf{A}(t+u) - \mathbf{D}(u),\\ \tilde{\mathbf{S}}(t) &:= \mathbf{S}(t+u) - \mathbf{S}(u).\end{aligned}$$

Let $\tilde{\mathbf{D}} = \Phi(\tilde{\mathbf{A}}, \tilde{\mathbf{S}})$, then we have

$$\tilde{\mathbf{D}}(t) = \mathbf{D}(t+u) - \mathbf{D}(u).$$

Proof. We show that for $\mathbf{D} = \Phi(\mathbf{A}, \mathbf{S})$, we have

$$\mathbf{D}(t+u) - \mathbf{D}(u) = \inf_{u \leq s \leq t+u} \left\{\mathbf{S}(t+u) - \mathbf{S}(s) + \mathbf{A}(s) - \mathbf{D}(u)\right\} \wedge \left\{\mathbf{S}(t+u) - \mathbf{S}(u)\right\},$$

from which the lemma follows.

We write

$$\begin{aligned}\mathbf{D}(t+u) - \mathbf{D}(u) &= \inf_{0 \leq s \leq u} \left\{\mathbf{S}(t+u) - \mathbf{S}(s) + \mathbf{A}(s) - \mathbf{D}(u)\right\}\\ &\quad \wedge \inf_{u \leq s \leq t+u} \left\{\mathbf{S}(t+u) - \mathbf{S}(s) + \mathbf{A}(s) - \mathbf{D}(u)\right\} \wedge \left\{\mathbf{S}(t+u) - \mathbf{D}(u)\right\},\end{aligned}$$

Since $\mathbf{D}(u) \leq \mathbf{S}(u)$, we have to prove that

$$\mathbf{S}(t+u) - \mathbf{S}(u) \geq \inf_{0 \leq s \leq u} \left\{\mathbf{S}(t+u) - \mathbf{S}(s) + \mathbf{A}(s) - \mathbf{D}(u)\right\} \wedge \left\{\mathbf{S}(t+u) - \mathbf{D}(u)\right\}.$$

This will follow from,

$$\begin{aligned}\inf_{0 \leq s \leq u} \left\{\mathbf{S}(t+u) - \mathbf{S}(s) + \mathbf{A}(s) - \mathbf{D}(u)\right\} &= \mathbf{S}(t+u) - \mathbf{S}(u) + \inf_{0 \leq s \leq u} \left\{\mathbf{S}(u) - \mathbf{S}(s) + \mathbf{A}(s)\right\} - \mathbf{D}(u)\\ &\leq \mathbf{S}(t+u) - \mathbf{S}(u).\end{aligned}$$

□

We consider now piece-wise linear Jackson networks : the functions $u \mapsto \mathbf{N}^{(i)}(u), u \mapsto \mathbf{S}^{(i)}(u)$ and $u \mapsto \mathbf{P}^{(i,j)}(u)$ are continuous piece-wise linear functions such that $\mathbf{N}^{(i)}(0) \in \mathbb{R}_+$ and $\mathbf{S}^{(i)}(0) = \mathbf{P}^{(i,j)}(0) = 0$ and $\rho(\dot{\mathbf{P}}(t)) < 1$ for all $t \geq 0$.

Proposition 25. *For a piece-wise linear Jackson network, there exists an unique solution of the fixed point equation (5.33). We still denote by Ψ the mapping that to any piece-wise linear Jackson network $\mathbf{JN}$ associates the corresponding couple $(\mathbf{A}, \mathbf{D})$.*

Proof. The existence is a direct consequence of monotonicity properties and continuity of the maps Γ and Φ. We define the sequence of processes $\{\mathbf{A}[k], \mathbf{D}[k]\}_{k \geq 0}$ with the recurrence equation :

$$\begin{cases} \mathbf{A}[k+1] = \Gamma(\mathbf{D}[k], \mathbf{JN}), \\ \mathbf{D}[k+1] = \Phi(\mathbf{A}[k+1], \mathbf{JN}), \end{cases}$$

and with initial condition $\mathbf{D}[0] = \mathbf{0}$. By the monotonicity properties of Φ and Γ, we have

$$\begin{aligned} \mathbf{0} \leq \mathbf{A}[1] \quad &\Rightarrow \quad \Phi(\mathbf{0}, \mathbf{JN}) = \mathbf{0} = \mathbf{D}[0] \leq \Phi(\mathbf{A}[1], \mathbf{JN}) = \mathbf{D}[1] \\ &\Rightarrow \quad \Gamma(\mathbf{D}[0], \mathbf{JN}) = \mathbf{A}[1] \leq \Gamma(\mathbf{D}[1], \mathbf{JN}) = \mathbf{A}[2], \end{aligned}$$

and the sequence $\{\mathbf{A}[k], \mathbf{D}[k]\}_{k \geq 0}$ is increasing. Note that $\mathbf{D}[k] \leq \mathbf{S}$ and hence the following limits are well defined

$$\lim_{k \to \infty} \mathbf{A}[k] = \mathbf{A} \quad \text{and,} \quad \lim_{k \to \infty} \mathbf{D}[k] = \mathbf{D}.$$

Since Γ and Φ are continuous, $(\mathbf{A}, \mathbf{D})$ is a solution of the fixed point equation (5.33).

We now prove uniqueness. First recall that we call α, a partition of $\mathbb{R}_+$, any sequence of points $\alpha = \{a_n\}_n$ with $a_0 = 0$ and $a_n \to \infty$. For two partitions $\alpha = \{a_n\}_n$ and $\beta = \{b_n\}_n$, we say that $\gamma = \{g_n\}_n$ is the union of α and β if γ is a partition such that for all n there exists m such that either $g_n = a_m$ or $g_n = b_m$.

Let $\tau = \{t_n\}_n$ be the union of the partitions associated with each function $\mathbf{S}, \mathbf{P}, \mathbf{N}$. We define for $x \in \mathbb{R}_+$, $d(x, \tau) = \min_n \{t_n - x,\ t_n > x\} > 0$.

Assume that we are given two solutions of the fixed point equation (5.33) : $(\mathbf{A}_1, \mathbf{D}_1)$ and $(\mathbf{A}_2, \mathbf{D}_2)$. First note that thanks to Lemmas 49 and 50, any solution of (5.33) is absolutely continuous. Let $z = \inf\{t, \mathbf{A}_1(t) \neq \mathbf{A}_2(t)\}$, in particular, we have $\mathbf{A}_1(t) = \mathbf{A}_2(t)$ and $\mathbf{D}_1(t) = \mathbf{D}_2(t)$ for all $t \leq z$.

Define $u = \min_i d(\mathbf{D}_\bullet^{(i)}(z), \tau) \wedge d(z, \tau) > 0$, where the notation $_\bullet$ can be replaced either by $_1$ or by $_2$. We have that for $t \in [0, u]$,

$$\begin{aligned} \tilde{\mathbf{S}}^{(i)}(t) &:= \mathbf{S}^{(i)}(z+t) - \mathbf{S}^{(i)}(z) = t\mu^{(i)}, \\ \tilde{\mathbf{P}}^{(i,j)}(t) &:= \mathbf{P}^{(i,j)}(\mathbf{D}_\bullet^{(i)}(z) + t) - \mathbf{P}^{(i,j)}(\mathbf{D}_\bullet^{(i)}(z)) = tP^{(i,j)}, \\ \tilde{\mathbf{N}}^{(i)} &:= \mathbf{N}^{(i)}(z+t) - \mathbf{N}^{(i)}(z) + \mathbf{A}_\bullet^{(i)}(z) - \mathbf{D}_\bullet^{(i)}(z) = t\lambda^{(i)} + \mathbf{A}_\bullet^{(i)}(z) - \mathbf{D}_\bullet^{(i)}(z), \end{aligned}$$

Let $\tilde{\mathbf{A}}(t) = \mathbf{X}_f[\mu, P, \mathbf{A}_\bullet(z) - \mathbf{D}_\bullet(z), \lambda](t)$ be the unique solution associated to the infinite horizon linear Jackson network defined above. The associated departure process is $\tilde{\mathbf{D}}(t) = \tilde{\mathbf{A}}(t) \wedge \mu t$. Let

$v = \inf\{t,\ \inf_i \tilde{\mathbf{D}}^{(i)}(t) = u\}$, in particular since $\tilde{\mathbf{D}}^{(i)}(t) \leq \mu^{(i)} t$, we have $v > 0$. In view of Lemma 46, we have for $t \in (0, v)$,

$$\mathbf{A}_{\bullet}(t+z) = \tilde{\mathbf{A}}(t) + \mathbf{D}(z), \quad \mathbf{D}_{\bullet}(t+z) = \tilde{\mathbf{D}}(t) + \mathbf{D}(z)$$

this contradicts the fact that $z < \infty$ and concludes the proof. □

Let $\mathcal{E} = \mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{M}^K) \times \mathbb{D}(\mathbb{R}_+^K)$ and $\mathcal{F} = \mathbb{D}(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{R}_+^K)$.

For $\mathbf{JN} \in \mathcal{E}$ and $(\mathbf{A}, \mathbf{D}) \in \mathcal{F}$, we define the function

$$G(\mathbf{JN}, \mathbf{A}, \mathbf{D}) = \|(\mathbf{A} - \Gamma(\mathbf{D}, \mathbf{JN}), \mathbf{D} - \Phi(\mathbf{A}, \mathbf{JN}))\|.$$

The function G is continuous and such that

$$G(\mathbf{JN}, \mathbf{A}, \mathbf{D}) = 0 \Leftrightarrow \begin{cases} \mathbf{A} &= \Gamma(\mathbf{D}, \mathbf{JN}), \\ \mathbf{D} &= \Phi(\mathbf{A}, \mathbf{JN}). \end{cases}$$

Let $\mathcal{D}_{\mathbf{JN}}$ be the subspace of $\mathcal{E}$ of piecewise linear Jackson networks : namely $\mathbf{JN} = (\mathbf{S}, \mathbf{P}, \mathbf{N}) \in \mathcal{D}_{\mathbf{JN}}$ if the functions $u \mapsto \mathbf{N}^{(i)}(u)$, $u \mapsto \mathbf{S}^{(i)}(u)$ and $u \mapsto \mathbf{P}^{(i,j)}(u)$ are piecewise linear functions such that $\rho(\dot{\mathbf{P}}(t)) < 1$ for all $t \geq 0$ and $\mathbf{N}^{(i)} = \mathbf{0}$ for $i \notin \mathcal{S}$. We denote $\dot{\mathbf{JN}} = (\dot{\mathbf{S}}, \dot{\mathbf{P}}, \dot{\mathbf{N}})$.

We proved that

$$\forall \mathbf{JN} \in \mathcal{D}_{\mathbf{JN}}, \quad G(\mathbf{JN}, \mathbf{A}, \mathbf{D}) = 0 \Leftrightarrow (\mathbf{A}, \mathbf{D}) = \Psi(\mathbf{JN}),$$

where Ψ has been explicitly defined above. In the next section we will define the mapping $\mathcal{S} : \mathcal{E} \times \mathcal{F} \to \mathcal{D}_{\mathbf{JN}}^{\mathbb{N}}$.

5.5.4 Sample path large deviations

In order to simplify the notations, we assume that $\mathbf{N}_n(0) = 0$ for all n. This condition can be weakened to the standard condition :

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathbf{N}_n(0) > \epsilon) = 0,$$

for all $\epsilon > 0$. In this case, we have $I^0(x) = \infty$ for all $x \neq 0$ and $I^0(0) = 0$.

Construction of the approximating sequence

Proposition 26. *We consider $\mathbf{JN} = (\mathbf{S}, \mathbf{P}, \mathbf{N}) \in \mathcal{E}$ such that $I^{\mathbf{JN}}(\mathbf{JN}) < \infty$ and such that there exists $(\mathbf{A}, \mathbf{D}) \in \mathcal{F}$ that satisfies the fixed point equation (5.33),*

$$\begin{cases} \mathbf{A} &= \Gamma(\mathbf{D}, \mathbf{JN}), \\ \mathbf{D} &= \Phi(\mathbf{A}, \mathbf{JN}). \end{cases}$$

There exists a sequence $\{\mathbf{JN}_n\}_n = \mathcal{S}(\mathbf{JN}, \mathbf{A}, \mathbf{D})$ such that

$$\mathbf{JN}_n \in \mathcal{D}_{\mathbf{JN}} \quad \textit{for all } n\,; \tag{5.41}$$

$$\mathbf{JN}_n \to \mathbf{JN}; \tag{5.42}$$

$$\Psi(\mathbf{JN}_n) \to (\mathbf{A}, \mathbf{D}); \tag{5.43}$$

$$I^{\mathbf{JN}}(\mathbf{JN}_n) \to I^{\mathbf{JN}}(\mathbf{JN}). \tag{5.44}$$

First note that since $I^{\mathbf{JN}}(\mathbf{JN}) < \infty$, each process $\mathbf{S}, \mathbf{P}, \mathbf{N}$ is absolutely continuous and $\dot{\mathbf{JN}}$ is well-defined. Moreover thanks to Lemma 51, the processes $\mathbf{A}$ and $\mathbf{D}$ are absolutely continuous too.

The idea to construct the sequence $\{\mathbf{JN}_n\}_n$ is to consider the piecewise approximation of the fixed point equation (5.33). First consider the routing equation $\mathbf{A} = \Gamma(\mathbf{D}, \mathbf{JN})$ for times t such that $nt \in \mathbb{N}$,

$$\underbrace{\mathbf{A}^{(i)}(t+1/n) - \mathbf{A}^{(i)}(t)}_{\Delta_n^{(i)}(\mathbf{A})(t)} = \underbrace{\mathbf{N}^{(i)}(t+1/n) - \mathbf{N}^{(i)}(t)}_{\Delta_n^{(i)}(\mathbf{N})(t)} + \sum_{j=1}^{K} \dot{\tilde{\mathbf{P}}}_n^{(j,i)}(\mathbf{D}^{(j)}(t+)) \underbrace{(\mathbf{D}^{(j)}(t+1/n) - \mathbf{D}^{(j)}(t))}_{\Delta_n^{(j)}(\mathbf{D})(t)},$$

where we define the piece-wise linear process $\tilde{\mathbf{P}}_n^{(j,i)}(t)$ as follows, for $s \in (\mathbf{D}^{(j)}(t), \mathbf{D}^{(j)}(t+1/n))$,

$$\dot{\tilde{\mathbf{P}}}_n^{(j,i)}(s) := \frac{\mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t+1/n)) - \mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t))}{\mathbf{D}^{(j)}(t+1/n)) - \mathbf{D}^{(j)}(t)},$$

if $\mathbf{D}^{(j)}(t+1/n) \neq \mathbf{D}^{(j)}(t)$, and we take $\dot{\tilde{\mathbf{P}}}_n^{(j,i)}(\mathbf{D}^{(j)}(t)) = 0$ otherwise. In other words, we have

$$\begin{aligned}\tilde{\mathbf{P}}_n^{(j,i)}(\mathbf{D}^{(j)}(t+1/n)) - \tilde{\mathbf{P}}_n^{(j,i)}(\mathbf{D}^{(j)}(t)) &= \dot{\tilde{\mathbf{P}}}_n^{(j,i)}(\mathbf{D}^{(j)}(t+))(\mathbf{D}^{(j)}(t+1/n) - \mathbf{D}^{(j)}(t)) \\ &= \mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t+1/n)) - \mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t))\end{aligned}$$

Note that $\{\dot{\tilde{\mathbf{P}}}_n^{(j,i)}(t)\}_{i,j} \in \mathbb{M}^K$ since

$$\sum_i \mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t+1/n)) - \mathbf{P}^{(j,i)}(\mathbf{D}^{(j)}(t)) \leq \mathbf{D}^{(j)}(t+1/n) - \mathbf{D}^{(j)}(t).$$

but the matrix $(\dot{\tilde{\mathbf{P}}}_n^{(j,i)}(\mathbf{D}^{(j)}(t+)))_{i,j}$ may not be of spectral radius less than 1. To circumvent this difficulty, we will modify slightly the processes as follows, (the variables η, ϵ_n, δ will be made precise latter)

$$\begin{aligned}\Delta_n^{(i)}(\mathbf{A}) + \frac{\eta^{(i)}}{n} &= \Delta_n^{(i)}(\mathbf{N}) + \frac{\delta^{(i)}}{n} \\ &+ \sum_{j=1}^{K} \left((1-\epsilon_n^{(j)})\dot{\tilde{\mathbf{P}}}_n^{(j,i)} + \epsilon_n^{(j)} R^{(j,i)}\right)\left(\Delta_n^{(j)}(\mathbf{D}) + \frac{\eta^{(j)}}{n}\right),\end{aligned} \tag{5.45}$$

where we omit to write the time t and use the simplified notation $\dot{\tilde{\mathbf{P}}}_n^{(j,i)} = \dot{\tilde{\mathbf{P}}}_n^{(j,i)}(\mathbf{D}^{(j)}(t+))$.

We have to find η, ϵ_n, δ such that (5.45) holds with $\eta^{(i)}, \epsilon_n^{(i)}, \delta^{(i)}$ non-negative and $\delta^{(i)} = 0$ for $i \notin \mathbb{S}$. These constraints are satisfied by the following choice : first take δ such that $\delta^{(i)} > 0$ for all $i \in \mathbb{S}$ and $\delta^{(i)} = 0$ for $i \notin \mathbb{S}$. Let $\eta(\delta) = \eta$ be the unique solution in $\mathbb{R}_+^K$ of the following equation (recall that $\rho(R) < 1$),

$$\eta^{(i)} = \delta^{(i)} + \sum_{j=1}^{K} \eta^{(j)} R^{(j,i)}.$$

Note that $\eta^{(i)} > 0$ for all i thanks to (5.32). Finally let define $\epsilon_n(\delta) = \epsilon_n$ as follows $\epsilon_n^{(i)} = \frac{\eta^{(i)}}{n\Delta_n^{(i)}(\mathbf{D})+\eta^{(i)}} \in (0,1]$ (note that $\epsilon_n^{(i)} = 1$ iff $\Delta_n^{(i)}(\mathbf{D}) = 0$).

It is easy to see that (5.45) holds since we have

$$(1-\epsilon_n^{(j)})\left(\Delta_n^{(j)}(\mathbf{D}) + \frac{\eta^{(j)}}{n}\right) = \Delta_n^{(j)}(\mathbf{D}), \quad \text{or,} \quad \epsilon_n^{(j)}\left(\Delta_n^{(j)}(\mathbf{D}) + \frac{\eta^{(j)}}{n}\right) = \frac{\eta^{(j)}}{n},$$

which imply that

$$\begin{aligned}\Delta_n^{(i)}(\mathbf{A}) &= \Delta_n^{(i)}(\mathbf{N}) + \sum_{j=1}^K (1-\epsilon_n^{(j)})\dot{\hat{\mathbf{P}}}_n^{(j,i)}\left(\Delta_n^{(j)}(\mathbf{D}) + \frac{\eta^{(j)}}{n}\right) \quad \text{and,}\\ \frac{\eta^{(i)}}{n} &= \frac{\delta^{(i)}}{n} + \sum_{j=1}^K \epsilon_n^{(j)} R^{(j,i)}\left(\Delta_n^{(j)}(\mathbf{D}) + \frac{\eta^{(j)}}{n}\right),\end{aligned}$$

and summing these two equalities gives (5.45).

For δ fixed, we define for $s \in (\mathbf{D}^{(j)}(t) + t\eta(\delta), \mathbf{D}^{(j)}(t+1/n) + (t+1/n)\eta(\delta))$,

$$\dot{\mathbf{P}}_{n,\delta}^{(j,i)}(s) = (1-\epsilon_n^{(j)})\dot{\hat{\mathbf{P}}}_n^{(j,i)}(s) + \epsilon_n^{(j)} R^{(j,i)},$$

where $\epsilon_n(\delta)$ is defined as above. In view of Lemma 52, the matrix $\dot{\mathbf{P}}_{n,\delta}^{(j,i)}(s)$ is of spectral radius less than one since $\epsilon_n^{(j)} > 0$ for all j. Then as a direct consequence of (5.45), we have for $nt \in \mathbb{N}$,

$$\mathbf{A}^{(i)}(t) + t\eta(\delta) = \mathbf{N}^{(i)}(t) + t\delta + \sum_{j=1}^K \mathbf{P}_{n,\delta}^{(j,i)}(\mathbf{D}^{(j)}(t) + t\eta(\delta)). \tag{5.46}$$

If $\mathbf{N}_{n,\delta}$ is the polygonal approximation of $t \to \mathbf{N}(t) + t\delta$ with step $1/n$, we have clearly $\dot{\mathbf{N}}_{n,\delta} \to \dot{\mathbf{N}} + \delta$ as n tends to infinity. Similarly, we have as n tends to infinity,

$$\dot{\mathbf{P}}_{n,\delta}^{(j,i)}(\mathbf{D}^{(j)}(t)) \to \begin{cases} (1-\epsilon^{(j)}(t))\dot{\mathbf{P}}^{(j,i)}(\mathbf{D}^{(j)}(t))\frac{(\dot{\mathbf{D}}^{(j)}(t)+\eta(\delta))}{\dot{\mathbf{D}}^{(j)}(t)} & \\ \quad +\epsilon^{(j)}(t)R^{(j,i)}(\dot{\mathbf{D}}^{(j)}(t)+\eta(\delta)) & \text{if } \dot{\mathbf{D}}^{(j)}(t) > 0,\\ R^{(j,i)}\eta(\delta) & \text{otherwise,}\end{cases}$$

where $\epsilon^{(j)}(t) = \eta^{(j)}(\delta)/(\eta^{(j)}(\delta) + \dot{\mathbf{D}}^{(j)}(t)) < 1$. Hence when n tends to infinity and δ tends to 0, we have $\dot{\mathbf{N}}_{n,\delta} \to \dot{\mathbf{N}}$ and $\dot{\mathbf{P}}_{n,\delta}^{(j,i)} \to \dot{\mathbf{P}}^{(j,i)}$.

We consider now the queueing equation $\mathbf{D} = \Phi(\mathbf{A}, \mathbf{S})$ and construct the approximating sequence for $\mathbf{S}$.

We begin with a first general lemma : given two processes $\mathbf{A}$ and $\mathbf{D}$, we construct a piecewise linear function $\mathbf{S}_n$ (with step $1/n$) as follows (with $nt \in \mathbb{N}$) :

– if $\mathbf{A}(t) = \mathbf{D}(t)$ and $\mathbf{A}(t+1/n) = \mathbf{D}(t+1/n)$, then $\mathbf{S}_n(t+1/n) - \mathbf{S}_n(t) = \mathbf{S}(t+1/n) - \mathbf{S}(t)$;

– otherwise, $\mathbf{S}_n(t+1/n) - \mathbf{S}_n(t) = \mathbf{D}(t+1/n) - \mathbf{D}(t)$.

We will write in short $\mathbf{S}_n = \Upsilon_n(\mathbf{A}, \mathbf{D}, \mathbf{S})$.

Lemma 47. *Let $(\mathbf{A}, \mathbf{D}, \mathbf{S})$ be absolutely continuous functions of $\mathbb{D}(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{R}_+^K) \times \mathbb{D}_0(\mathbb{R}_+^K)$ such that $\Phi(\mathbf{A}, \mathbf{S}) = \mathbf{D}$. We denote $\mathbf{S}_n = \Upsilon_n(\mathbf{A}, \mathbf{D}, \mathbf{S})$. We have $\mathbf{D}_n = \Phi(\mathbf{A}_n, \mathbf{S}_n)$ where $(\mathbf{A}_n, \mathbf{D}_n)$ is the polygonal approximation of $(\mathbf{A}, \mathbf{D})$ with step $1/n$ and we have $\dot{\mathbf{S}}_n \to \dot{\mathbf{S}}$ as n tends to infinity.*

Proof. We denote $\tilde{\mathbf{D}}_n = \Phi(\mathbf{A}_n, \mathbf{S}_n)$. From the proof of Lemma 46, we have

$$\tilde{\mathbf{D}}_n(t+1/n) - \tilde{\mathbf{D}}_n(t) = \inf_{t \le s \le t+1/n} \left\{ \mathbf{S}_n(t+1/n) - \mathbf{S}_n(s) + \mathbf{A}_n(s) - \tilde{\mathbf{D}}_n(t) \right\} \wedge \left\{ \mathbf{S}_n(t+1/n) - \mathbf{S}_n(t) \right\},$$

since all the functions are linear on the interval $(t, t+1/n)$, we have

$$\tilde{\mathbf{D}}_n(t+1/n) - \tilde{\mathbf{D}}_n(t) \;=\; \left\{ \mathbf{A}_n(t+1/n) - \tilde{\mathbf{D}}_n(t) \right\} \wedge \left\{ \mathbf{S}_n(t+1/n) - \mathbf{S}_n(t) \right\}.$$

If $\tilde{\mathbf{D}}_n(t) = \mathbf{D}_n(t)$, then we have clearly $\tilde{\mathbf{D}}_n(t+1/n) = \mathbf{D}_n(t+1/n)$ since

- if $\mathbf{A}_n(t) = \mathbf{D}_n(t)$ and $\mathbf{A}_n(t+1/n) = \mathbf{D}_n(t+1/n)$, then we have $\mathbf{S}(t+1/n) - \mathbf{S}(t) \ge \mathbf{D}_n(t+1/n) - \mathbf{D}_n(t) = \mathbf{A}_n(t+1/n) - \tilde{\mathbf{D}}_n(t)$ see (5.52) for the inequality ;
- otherwise, $\mathbf{S}_n(t+1/n) - \mathbf{S}_n(t) = \mathbf{D}_n(t+1/n) - \mathbf{D}_n(t)$ by definition and $\mathbf{A}_n(t+1/n) \ge \mathbf{D}_n(t+1/n)$.

This proves the first part of the lemma.

To see that the second part holds, let $C = \{t,\ \mathbf{A}(t) = \mathbf{D}(t)\}$. C is a closed set and according to Lemma 51, we have for all $t \in C^c$, $\dot{\mathbf{S}}(t) = \dot{\mathbf{D}}(t)$. For such $t \in C^c$, we have for $\epsilon > 0$ sufficiently small and for sufficiently large n, $\mathbf{A}_n(u) \neq \mathbf{D}_n(u)$ for all $|u-t| \le \epsilon$. Hence we have $\dot{\mathbf{S}}_n(t) = \dot{\mathbf{D}}_n(t) \to \dot{\mathbf{D}}(t)$.

Now for $t \in C^o$ in the interior of C, we have clearly $\dot{\mathbf{S}}_n(t) \to \dot{\mathbf{S}}(t)$. Hence we have $\dot{\mathbf{S}}_n(t) \to \dot{\mathbf{S}}(t)$ for $t \in C^o \cup C^c$ which concludes the proof. □

We define the sequence $\mathbf{JN}_{n,\delta} = (\mathbf{S}_{n,\delta}, \mathbf{P}_{n,\delta}, \mathbf{N}_{n,\delta})$ where $\mathbf{S}_{n,\delta}(t) = \Upsilon_n(\mathbf{A}(t)+\eta t, \mathbf{D}(t)+\eta t, \mathbf{S}(t)+\eta t)$. Note that we have $\mathbf{D}(t)+\eta t = \Phi(\mathbf{A}(t)+\eta t, \mathbf{S}(t)+\eta t)$, hence Lemma 47 applies, in particular, we have $\dot{\mathbf{S}}_{n,\delta}(t) \to \dot{\mathbf{S}}(t) + \eta(\delta)$ as n tends to infinity.

We have $\mathbf{JN}_{n,\delta} \in \mathcal{D}_{\mathbf{JN}}$ by construction and the sequence $\{\mathbf{JN}_{n,\delta_n}\}_n$ satisfies (5.42) for some $\delta_n \to 0$. Moreover, we have thanks to (5.46) and Lemma 47,

$$\left\{ \begin{array}{rcl} \mathbf{A}_{n,\delta} &=& \Gamma(\mathbf{D}_{n,\delta}, \mathbf{JN}_{n,\delta}), \\ \mathbf{D}_{n,\delta} &=& \Phi(\mathbf{A}_{n,\delta}, \mathbf{JN}_{n,\delta}), \end{array} \right. \Leftrightarrow (\mathbf{A}_{n,\delta}, \mathbf{D}_{n,\delta}) = \Psi(\mathbf{JN}_{n,\delta}),$$

where $\mathbf{A}_{n,\delta}$ and $\mathbf{D}_{n,\delta}$ are the polygonal approximation of $\mathbf{A}(t)+\eta t$ and $\mathbf{D}(t)+\eta t$ with step $1/n$.

For $n \to \infty$ and $\delta \to 0$, we have $(\mathbf{A}_{n,\delta}, \mathbf{D}_{n,\delta}) \to (\mathbf{A}, \mathbf{D})$, hence we have $\Psi(\mathbf{JN}_{n,\delta}) \to (\mathbf{A}, \mathbf{D})$, i.e. the sequence $\{\mathbf{JN}_{n,\delta_n}\}_n$ satisfies (5.43).

The following lemma shows that (5.44) is satisfied too.

Lemma 48. *For any* $\mathbf{JN}$ *such that* $I^{\mathbf{JN}}(\mathbf{JN}) < \infty$, *if* $\dot{\mathbf{JN}}_n \to \dot{\mathbf{JN}}$ *as* n *tends to infinity and* $\{\mathbf{JN}_n\}_n \in \mathcal{D}_{\mathbf{JN}}^{\mathbb{N}}$, *then we have* $I^{\mathbf{JN}}(\mathbf{JN}_n) \to I^{\mathbf{JN}}(\mathbf{JN})$.

Proof. Let $\{\mathbf{JN}_n = (\mathbf{S}_n, \mathbf{P}_n, \mathbf{N}_n)\}_n$ be a sequence in $\mathcal{D}_{\mathbf{JN}}^{\mathbb{N}}$ such that $\dot{\mathbf{JN}}_n$ converges to $\dot{\mathbf{JN}}$.

We consider the case of the sequence of processes $\{\mathbf{S}_n\}_n$ in details (we can restrict ourselves to the one dimensional case).

We first take $T > 0$. We define $\varsigma = ess\sup\{\dot{\mathbf{S}}(t),\ t \le T\} = \inf\{\alpha,\ leb[t \le T,\ \dot{\mathbf{S}}(t) > \alpha] = 0\}$, where leb is for the Lebesgue measure. Since $\int_0^T I^{\mathbf{S}}(\dot{\mathbf{S}}(t))dt < \infty$, ς belongs to the domain of $I^{\mathbf{S}}$ which is open on the right. Hence we can find $\epsilon > 0$ such that $\varsigma + \epsilon$ still belongs to this domain. Moreover, since $I^{\mathbf{S}}$ is convex, it is uniformly continuous on $[0, \varsigma+\epsilon]$. Hence, for $\delta > 0$, we can assume that

$$\forall x, y \in [0, \varsigma+\epsilon],\ |x-y| < \epsilon \Rightarrow |I^{\mathbf{S}}(x) - I^{\mathbf{S}}(y)| \le \delta.$$

There exists $N(\epsilon,T)$ such that for $n\geq N$, we have

$$\sup_{t\leq T}|\dot{\mathbf{S}}_n(t)-\dot{\mathbf{S}}(t)|<\epsilon,$$

where ϵ has been chosen above.

Let $\{x_k^n\}_k$ be the partition associated to $\mathbf{S}_n$ (in which for simplicity we add $T\in\cup_k\{x_k^n\}$). Then we have

$$\int_0^T I^S(\dot{\mathbf{S}}_n(t))dt \;=\; \sum_{\{k,\,x_k^n<T\}}(x_{k+1}^n-x_k^n)I^{\mathbf{S}}\left(\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}\dot{\mathbf{S}}_n(t)dt\right)$$

Now we have for each term in the sum of the right-hand term,

$$I^{\mathbf{S}}\left(\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}\dot{\mathbf{S}}_n(t)dt\right) \;=\; I^{\mathbf{S}}\left(\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}\dot{\mathbf{S}}(t)dt+\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}(\dot{\mathbf{S}}_n(t)-\dot{\mathbf{S}}(t))dt\right),$$

hence we have

$$\begin{aligned}I^{\mathbf{S}}\left(\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}\dot{\mathbf{S}}_n(t)dt\right) &\leq I^{\mathbf{S}}\left(\frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}\dot{\mathbf{S}}(t)dt\right)+\delta\\ &\leq \frac{1}{x_{k+1}^n-x_k^n}\int_{x_k^n}^{x_{k+1}^n}I^S(\dot{\mathbf{S}}(t))dt+\delta,\end{aligned}$$

where the last inequality follows from Jensen's inequality. Hence for any $\delta>0$, we showed that for sufficiently large n, we have,

$$\int_0^T I^{\mathbf{S}}(\dot{\mathbf{S}}_n(t))dt\leq\int_0^T I^{\mathbf{S}}(\dot{\mathbf{S}}(t))dt+\delta T,$$

hence we have for any $T>0$

$$\limsup_{n\to\infty}\int_0^T I^{\mathbf{S}}(\dot{\mathbf{S}}_n(t))dt\leq\int_0^T I^{\mathbf{S}}(\dot{\mathbf{S}}(t))dt, \tag{5.47}$$

and then the result is true for $T=\infty$ by monotonicity. The converse inequality follows from

$$\begin{aligned}\liminf_{n\to\infty}\int_0^\infty I^{\mathbf{S}}(\dot{\mathbf{S}}_n(t))dt &\geq \int_0^\infty \liminf_{n\to\infty} I^{\mathbf{S}}(\dot{\mathbf{S}}_n(t))dt\\ &\geq \int_0^\infty I^{\mathbf{S}}(\dot{\mathbf{S}}(t))dt,\end{aligned}$$

where the first inequality is due to Fatou's Lemma and the second one to the lower semicontinuity of $I^{\mathbf{S}}$. Hence we proved that

$$\lim_{n\to\infty}\int_0^\infty I^{\mathbf{S}}(\dot{\mathbf{S}}_n(t))dt=\int_0^\infty I^{\mathbf{S}}(\dot{\mathbf{S}}(t))dt.$$

The same argument can be repeated for $\mathbf{N}_n$. Note that $\{\mathbf{JN}_n\}_n\in\mathcal{D}_{\mathbf{JN}}^{\mathbb{N}}$ implies that $\mathbf{N}_n^{(i)}(t)=0$ for all $i\notin\mathcal{S}$. For $i\in\mathcal{S}$, we can use the fact tha the domain of $I^{\mathbf{N}^{(i)}}$ is open as previously. In the case of $\mathbf{P}_n$, we can not use the argument on the openness of the domain, but we have $\tilde{D}(R^{(i)}\|R^{(i)})=0$ and then the convexity of D directly implies that $\tilde{D}(\dot{\mathbf{P}}_n^{(i)}\|R^{(i)})\leq\tilde{D}(\dot{\mathbf{P}}^{(i)}\|R^{(i)})$, from which we derive an equivalent of (5.47). □

Exponential tightness

We first recall some definitions (see [42]). A sequence of random variables $\{X_n\}_n \in (\mathbb{R}^K)^{\mathbb{N}}$ is exponentially tight if

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|X_n\| > M) = -\infty.$$

For $\delta > 0$ and $T > 0$, define the modulus of continuity in $\mathbb{D}(E)$ by

$$w'(\mathbf{X}, \delta, T) := \inf_{\{t_i\}} \max_i \sup_{s,t\in[t_{i-1},t_i)} d(\mathbf{X}(s), \mathbf{X}(t)),$$

where the infimum is over $\{t_i\}$ satisfying

$$0 = t_0 < t_1 < \cdots < t_{m-1} < T \leq t_m$$

and $\min_{1\leq i\leq n}(t_i - t_{i-1}) > \delta$.

Theorem 4.1 of [42] tells us : let $\mathcal{T}_0$ be a dense subset of $\mathbb{R}_+$. Suppose that for each $t \in \mathcal{T}_0$, $\{\mathbf{X}_n(t)\}_n$ is exponentially tight. Then $\{\mathbf{X}_n\}_n$ is exponentially tight in $\mathbb{D}(E)$ if and only if for each $\epsilon > 0$ and $T > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(w'(\mathbf{X}_n, \delta, T) > \epsilon) = -\infty. \tag{5.48}$$

A sequence of stochastic processes $\{\mathbf{X}_n\}_n$ that is exponentially tight in $\mathbb{D}(E)$ is C-exponentially tight if for each $\eta > 0$ and $T > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sup_{s\leq T} d(\mathbf{X}_n(s), \mathbf{X}_n(s-)) \geq \eta) = -\infty. \tag{5.49}$$

Then Theorem 4.13 of [42] gives : an exponentially tight sequence $\{\mathbf{X}_n\}_n$ in $\mathbb{D}(E)$ is C-exponentially tight if and only if each rate function I that gives the LDP for a subsequence $\{\mathbf{X}_{n(k)}\}_{n(k)}$, satisfies $I(\mathbf{x}) = \infty$ for each $\mathbf{x} \in \mathbb{D}(E)$ such that $\mathbf{x} \notin \mathbb{C}(E)$.

The stochastic assumptions of Section 6.2 ensure that the sequence of processes $\{\mathbf{JN}_n\}_n$ satisfies a LDP with good rate function (this implies that the sequence is exponentially tight) giving an infinite mass to discontinuous path. Hence the sequence of processes $\{\mathbf{JN}_n\}_n$ is C-exponentially tight.

We have to show that the sequence of processes $\{(\mathbf{A}_n, \mathbf{D}_n)\}_n$ is exponentially tight. The fact of dealing with non-decreasing processes simplifies the definitions. For $\mathbf{X} \in \mathbb{D}(\mathbb{R}_+^K)$ (or $\mathbb{D}(\mathbb{M}^K)$) non-decreasing, $\delta > 0$ and $T > 0$, we define $w_\delta(\mathbf{X}, T) = \sup_{t\in[0,T]} \|\mathbf{X}(t+\delta) - \mathbf{X}(t)\|$. We have clearly $w'(\mathbf{X}, \delta, T) = w_\delta(\mathbf{X}, T)$ and if $\{\mathbf{X}_n(0)\}_n$ is exponentially tight then (5.48) implies that $\{\mathbf{X}_n(t)\}_n$ is exponentially tight for each $t > 0$. Lemmas 49 and 50 show that conditions (5.48) and (5.49) are satisfied for the sequence of processes $\{(\mathbf{A}_n, \mathbf{D}_n)\}_n$. The exponential tightness of $\{(\mathbf{A}_n(0), \mathbf{D}_n(0))\}_n$ is clear since $\mathbf{A}_n(0) = \mathbf{D}_n(0) = 0$.

Large deviations results

Proposition 27. *The sequence of processes $\{(\mathbf{A}_n, \mathbf{D}_n)\}_n$ satisfies a LDP in $\mathbb{D}(\mathbb{R}_+^K) \times \mathbb{D}(\mathbb{R}_+^K)$ with good rate function $I^{\mathbf{A},\mathbf{D}}$. For $\mathbf{A}, \mathbf{D}$ absolutely continuous and such that $\mathbf{A}(0) = \mathbf{D}(0) = 0$, $I^{\mathbf{A},\mathbf{D}}$ is given by*

$$I^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D}) = \int_0^\infty H(\mathbf{A}(s), \mathbf{D}(s), \dot{\mathbf{A}}(s), \dot{\mathbf{D}}(s))ds, \tag{5.50}$$

where $H(A,D,\dot A,\dot D) := \inf_{P,N} h(A,D,\dot A,\dot D,P,N)$*, with* h *given by,*

$$h(A,D,\dot A,\dot D,P,N) := \sum_{i\in E(A,D)} I^{\mathbf{S}^{(i)}}(\dot D^{(i)})\mathbb{1}_{\{\dot D^{(i)}>\mu^{(i)}\}} + \sum_{i\notin E(A,D)} I^{\mathbf{S}^{(i)}}(\dot D^{(i)}) + \sum_i \dot D^{(i)} D(P^{(i)}\|R^{(i)}) + I^{\mathbf{N}}(N)$$

where $E(A,D)=\{i,\, A^{(i)}=D^{(i)}\}$ *and with the infimum taken over the set of* $(P,N)\in\mathbb{M}^K\times\mathbb{R}_+^K$ *such that*

$$\dot A = N + P^t\dot D.$$

For all other $\mathbf{A},\mathbf{D}$*, we have* $I^{\mathbf{A},\mathbf{D}}(\mathbf{A},\mathbf{D})=\infty$*.*

Proof. We define

$$\tilde I^{\mathbf{A},\mathbf{D}}(\mathbf{A},\mathbf{D}) = \inf\left\{\lim_{n\to\infty} I^{\mathbf{JN}}(\mathbf{JN}_n),\ \{\mathbf{JN}_n\}_n\in\mathcal{S}(\mathbf{A},\mathbf{D})\right\}, \tag{5.51}$$

where we recall that $\mathcal{S}(\mathbf{A},\mathbf{D}) = \cup_{\mathbf{JN}}\mathcal{S}(\mathbf{JN},\mathbf{A},\mathbf{D})$, where $\mathcal{S}(\mathbf{JN},\mathbf{A},\mathbf{D})$ is defined in Proposition 26. We have to show that $\tilde I^{\mathbf{A},\mathbf{D}} = I^{\mathbf{A},\mathbf{D}}$ given by (5.50).

Consider $\mathbf{JN}\in\mathcal{D}_{\mathbf{JN}}$ and let $(\mathbf{A},\mathbf{D})=\Psi(\mathbf{JN})$. Let $\tau=\{0=t_0<t_1<\ldots\}$ be such that the processes $\mathbf{A},\mathbf{D},\mathbf{S},\mathbf{N}$ and $\mathbf{D}\circ\mathbf{P}$ have a constant derivative on each (t_k,t_{k+1}). Then from $\mathbf{A}=\Gamma(\mathbf{D},\mathbf{JN})$, we derive

$$\dot{\mathbf{A}}^{(i)}(t) = \dot{\mathbf{N}}^{(i)}(t) + \sum_j \dot{\mathbf{D}}^{(j)}(t)\dot{\mathbf{P}}^{(j,i)}(\mathbf{D}^{(j)}(t)).$$

From $\mathbf{D}=\Phi(\mathbf{A},\mathbf{S})$, we get the following constraints :

– if $\mathbf{A}^{(i)}(t_k) > \mathbf{D}^{(i)}(t_k)$ or $\mathbf{A}^{(i)}(t_{k+1}) > \mathbf{D}^{(i)}(t_{k+1})$, then we have $\dot{\mathbf{D}}^{(i)}(t) = \dot{\mathbf{S}}^{(i)}(t)$ for $t\in(t_k,t_{k+1})$;

– otherwise $\mathbf{A}^{(i)}(t)=\mathbf{D}^{(i)}(t)$ for $t\in(t_k,t_{k+1})$ and we have $\dot{\mathbf{S}}^{(i)}(t)\geq\dot{\mathbf{A}}^{(i)}(t)=\dot{\mathbf{D}}^{(i)}(t)$ for $t\in(t_k,t_{k+1})$.

Now we can compute $I^{\mathbf{JN}}(\mathbf{JN})$ as follows

$$\begin{aligned} I^{\mathbf{JN}}(\mathbf{JN}) &= \int_0^\infty \sum_{i\in E(A,D)} I^{\mathbf{S}^{(i)}}(\dot{\mathbf{S}}^{(i)}(s)) + \sum_{i\notin E(A,D)} I^{\mathbf{S}^{(i)}}(\dot{\mathbf{D}}^{(i)}(s))ds \\ &\quad + \int_0^\infty \sum_j \dot{\mathbf{D}}^{(j)}(s) D(\dot{\mathbf{P}}^{(j)}(s)\|R^{(j)}) + I^{\mathbf{N}}(\dot{\mathbf{N}}(s))ds \\ &\geq \int_0^\infty h(\mathbf{A}(s),\mathbf{D}(s),\dot{\mathbf{A}}(s),\dot{\mathbf{D}}(s),\dot{\mathbf{P}}(s),\dot{\mathbf{N}}(s))ds \geq I^{\mathbf{A},\mathbf{D}}(\Psi(\mathbf{JN})), \end{aligned}$$

since for $i\in E(\mathbf{A}(s),\mathbf{D}(s))$, we have $I^{\mathbf{S}^{(i)}}(\dot{\mathbf{S}}^{(i)}(s)) \geq I^{\mathbf{S}^{(i)}}(\dot{\mathbf{D}}^{(i)}(s))\mathbb{1}_{\{\dot{\mathbf{D}}^{(i)}(s)>\mu^{(i)}\}}$ because $\dot{\mathbf{S}}^{(i)}(s)\geq\dot{\mathbf{D}}^{(i)}(s)$ and $I^{\mathbf{S}^{(i)}}$ is non-negative, convex with $\mu^{(i)}$ as unique zero. Hence, we have $\tilde I^{\mathbf{A},\mathbf{D}} \geq I^{\mathbf{A},\mathbf{D}}$.

Consider now $(\mathbf{A},\mathbf{D})$ such that $I^{\mathbf{A},\mathbf{D}}(\mathbf{A},\mathbf{D})<\infty$, then we denote by $(\mathbf{p}(s),\mathbf{n}(s))$ the argument that achieves the minimum in $H(\mathbf{A}(s),\mathbf{D}(s),\dot{\mathbf{A}}(s),\dot{\mathbf{D}}(s))$ for any fixed s (note that h is a good rate function). Let $\mathbf{P}(\mathbf{D}(t)) = \int_0^t \mathbf{p}(s)ds$ and $\mathbf{N}(t)=\int_0^t\mathbf{n}(s)ds$. We have $\mathbf{A} = \Gamma(\mathbf{D},\mathbf{P},\mathbf{N})$. Now define $\mathbf{s}(s)$ as follows :

– if $\mathbf{A}^{(i)}(s) = \mathbf{D}^{(i)}(s)$ then $\mathbf{s}^{(i)}(s) = \dot{\mathbf{D}}^{(i)}(s) \vee \mu^{(i)}$;
– if $\mathbf{A}^{(i)}(s) > \mathbf{D}^{(i)}(s)$ then $\mathbf{s}^{(i)}(s) = \dot{\mathbf{D}}^{(i)}(s)$.

We have $\mathbf{D} = \Phi(\mathbf{A}, \mathbf{S})$ with $\mathbf{S}(t) = \int_0^t \mathbf{s}(s)ds$. Hence we have $(\mathbf{A}, \mathbf{D}) = (\Gamma(\mathbf{D}, \mathbf{JN}), \Phi(\mathbf{A}, \mathbf{JN}))$ for $\mathbf{JN} = (\mathbf{S}, \mathbf{P}, \mathbf{N})$ and $I^{\mathbf{JN}}(\mathbf{JN}) = I^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D}) < \infty$ by construction. Hence the sequence $\mathcal{S}(\mathbf{JN}, \mathbf{A}, \mathbf{D}) = \{\mathbf{JN}_n\}_n$ is well-defined and we have $\tilde{I}^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D}) \leq \lim_{n\to\infty} I^{\mathbf{JN}}(\mathbf{JN}_n) = I^{\mathbf{JN}}(\mathbf{JN}) = I^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D})$. □

From this proposition, it is quite easy to derive a LDP for the process $\mathbf{Q}_n(t) := \mathbf{A}_n(t) - \mathbf{D}_n(t)$ counting the number of customer in each queue.

Corollary 4. *The sequence of processes $\{\mathbf{Q}_n\}_n$ satisfies a LDP in $\mathbb{D}(\mathbb{R}_+^K)$ with good rate function that is finite for $\mathbf{Q}$ absolutely continuous and such that $\mathbf{Q}(0) = 0$ and given by :*

$$I^{\mathbf{Q}}(\mathbf{Q}) := \int_0^\infty H^{\mathbf{Q}}(\mathbf{Q}(s), \dot{\mathbf{Q}}(s))ds,$$

where $H^{\mathbf{Q}}$ is given by,

$$H^{\mathbf{Q}}(Q, \dot{Q}) := \inf\left\{\sum_{i\in E(Q)} I^{\mathbf{S}^{(i)}}(D^{(i)})\mathbb{1}_{\{D^{(i)}>\mu^{(i)}\}} + \sum_{i\notin E(Q)} I^{\mathbf{S}^{(i)}}(D^{(i)}) + \sum_i D^{(i)} D(P^{(i)}\|R^{(i)}) + I^{\mathbf{N}}(N)\right\}$$

where $E(Q) = \{i, Q^{(i)} = 0\}$ and the infimum is taken over the set of $(D, P, N) \in \mathbb{R}_+^K \times \mathbb{M}^K \times \mathbb{R}_+^K$ such that

$$\dot{Q} = N + (P^t - Id)D.$$

Proof. Thanks to the contraction principle, we have

$$I^{\mathbf{Q}}(\mathbf{Q}) = \inf\{I^{\mathbf{A},\mathbf{D}}(\mathbf{A}, \mathbf{D}), \mathbf{Q} = \mathbf{A} - \mathbf{D}\},$$

which gives directly the corollary. □

5.5.5 Appendix

Properties of the map Γ and Φ

For $\mathbf{X} \in \mathbb{D}(E)$, $\delta > 0$ and $T > 0$, we define $w_\delta(\mathbf{X}, T) = \sup_{t\in[0,T]} d(\mathbf{X}(t+\delta), \mathbf{X}(t))$.

Lemma 49. *We have*

$$w_\delta(\Phi(\mathbf{X}, \mathbf{Y}), T) \leq w_\delta(\mathbf{Y}, T).$$

Proof. It is clearly sufficient to consider the case $K = 1$. We will prove that

$$\Phi(\mathbf{X}, \mathbf{Y})(t+\delta) - \Phi(\mathbf{X}, \mathbf{Y})(t) \leq \mathbf{Y}(t+\delta) - \mathbf{Y}(t), \quad (5.52)$$

from which the lemma follows.

If $\Phi(\mathbf{X}, \mathbf{Y})(t) = \mathbf{Y}(t)$, then we have $\Phi(\mathbf{X}, \mathbf{Y})(t+\delta) \leq \mathbf{Y}(t+\delta)$ and (5.52) is clear.

Assume now that $\Phi(\mathbf{X}, \mathbf{Y})(t) = \inf_{0\leq s<t} \{\mathbf{Y}(t) - \mathbf{Y}(s) + \mathbf{X}(s)\} < \mathbf{Y}(t)$. We have

$$\mathbf{Y}(t+\delta) - \mathbf{Y}(s) + \mathbf{X}(s) = \mathbf{Y}(t) - \mathbf{Y}(s) + \mathbf{X}(s) + \mathbf{Y}(t+\delta) - \mathbf{Y}(t),$$

and (5.52) follows by taking the minimum in $s \in [0, t]$ and observing that $\Phi(\mathbf{X}, \mathbf{Y})(t+\delta) \leq \inf_{0\leq s<t} \{\mathbf{Y}(t+\delta) - \mathbf{Y}(s) + \mathbf{X}(s)\}$. □

The following lemma is clear :

Lemma 50. *We have*

$$w_\delta(\Gamma(\mathbf{X},\mathbf{P},\mathbf{N}),T) \le w_\delta(\mathbf{N},T) + w_\delta(\mathbf{P},\|\mathbf{X}(T)\|).$$

Lemma 51. *Assume $\mathbf{S} \in \mathbb{D}_0(\mathbb{R}_+)$ is absolutely continuous, then for any $\mathbf{A} \in \mathbb{D}(\mathbb{R}_+)$, we have $\mathbf{D} := \Phi(\mathbf{A},\mathbf{S})$ is absolutely continuous and,*

- *for all t such that $\mathbf{A}(t) > \mathbf{D}(t)$, we have $\dot{\mathbf{D}}(t) = \dot{\mathbf{S}}(t)$;*
- *if $\mathbf{A}(t) = \mathbf{D}(t)$ for $t \in (u,v)$ with $u < v$, then we have $\dot{\mathbf{S}}(t) \ge \dot{\mathbf{A}}(t) = \dot{\mathbf{D}}(t)$ for $t \in (u,v)$.*

Proof. It follows directly form (5.52) that if $\mathbf{S}$ is absolutely continuous, then $\Phi(\mathbf{X},\mathbf{S})$ is absolutely continuous for any $\mathbf{X}$. The rest of the lemma is obvious. □

Auxiliary results

Lemma 52. *Given a substochastic matrix R such that $\rho(R) < 1$ and a substochastic matrix P such that the support of P is included in the support of R, i.e. $R^{(i,j)} = 0 \Rightarrow P^{(i,j)} = 0$. Then for any ϵ such that $0 < \epsilon^{(i)} \le 1$ for all i, the matrix with coefficients $M^{(i,j)} = (1-\epsilon^{(i)})P^{(i,j)} + \epsilon^{(i)}R^{(i,j)}$ is of spectral radius less than 1.*

Proof. By a suitable permutation of rows and columns, we can assume that R is given in its canonical form

$$R = \begin{pmatrix} S_1(R) & * & * & * \\ 0 & S_2(R) & * & * \\ 0 & 0 & \ddots & * \\ 0 & 0 & 0 & S_n(R) \end{pmatrix}, \tag{5.53}$$

where each $S_i(R)$ is an irreducible matrix. We have $\rho(R) < 1$ if and only if each $S_i(R)$ is not a stochastic matrix.

In view of the assumption on the support of P, the matrix P has the same structure as (5.53) and we have with the same notation as above, $S_i(M)$ which is an irreducible and not stochastic matrix. □

An example

In this section, we construct 2 different sequences of Jackson networks $\mathbf{JN}_n^1$ and $\mathbf{JN}_n^2$ such that their fluid limits are the same

$$\mathbf{JN}_n^1 \to \mathbf{JN} \quad \text{and} \quad \mathbf{JN}_n^2 \to \mathbf{JN},$$

but such that

$$\begin{aligned} (\mathbf{A}_n^1,\mathbf{D}_n^1) = \Psi(\mathbf{JN}_n^1) &\;\to\; (\mathbf{A}^1,\mathbf{D}^1), \\ (\mathbf{A}_n^2,\mathbf{D}_n^2) = \Psi(\mathbf{JN}_n^2) &\;\to\; (\mathbf{A}^2,\mathbf{D}^2), \end{aligned}$$

with $(\mathbf{A}^1,\mathbf{D}^1) \ne (\mathbf{A}^2,\mathbf{D}^2)$.

We consider a toy example with only one node. Once a customer is served, he can either go out of the network or go back to this same node. We define the following driving sequences :

$$\begin{aligned} \sigma^{(0),n} &= (\underbrace{1,\dots,1}_{n}, n, \underbrace{1,\dots,1}_{n}, n, \dots), \\ \sigma^{(1),n} &= \alpha(1,1,\dots), \end{aligned}$$

with $\alpha < 1$. We define now two different routing sequences

$$\begin{aligned} \nu^{(1),n} &= (\underbrace{2,\dots,2}_{n+1}, \underbrace{1,\dots,1}_{n+1}, \dots), \\ \nu^{(1),n}(x) &= (\underbrace{2,\dots,2}_{\lfloor xn \rfloor}, 1, \underbrace{2,\dots,2}_{n-\lfloor xn \rfloor}, \underbrace{1,\dots,1}_{\lfloor xn \rfloor}, 2, \underbrace{1,\dots,1}_{n-\lfloor xn \rfloor}, \dots), \end{aligned}$$

where $x < 1$. We denote by $\mathbf{JN}_n^1 = \{\sigma^n, \nu^n, 0\}$ and $\mathbf{JN}_n^2 = \{\sigma^n, \nu^n(x), 0\}$. $\nu^n(x)$ is obtained from ν^n by only interchanging a 1 and a 2. Hence we have

$$\mathbf{JN}_n^1 \to \mathbf{JN} \quad \text{and} \quad \mathbf{JN}_n^2 \to \mathbf{JN}.$$

Indeed the fluid Jackson network $\mathbf{JN}$ is given on Figure 5.3.

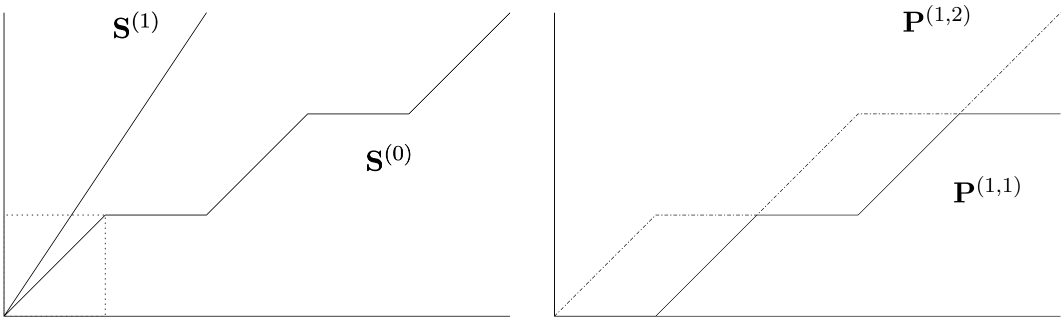


FIG. 5.3 – Fluid Jackson networks : $\mathbf{JN}$

The fluid limit of the departure processes are given on Figure 5.4 (for 2 different values of α).

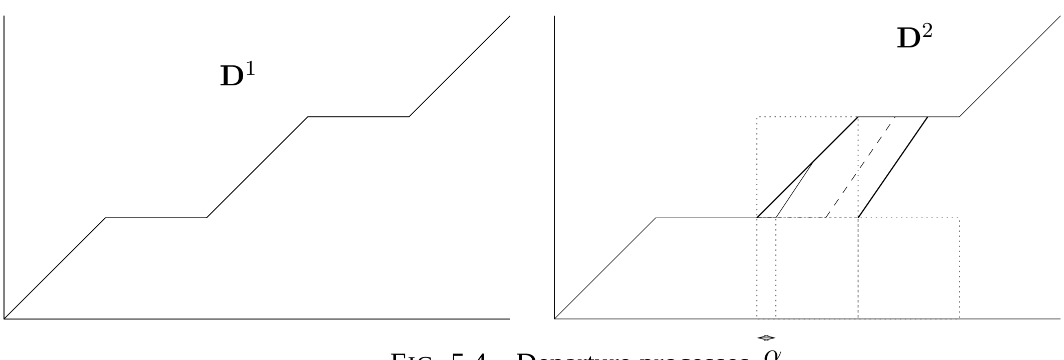


FIG. 5.4 – Departure processes

To explain $\mathbf{D}^2$, we write for each arrival (number on the left) the couple corresponding to :
the inter-arrival time | the routing decision (2 means that the customer exits the network) :

$$\begin{array}{ccccc}
1 & \to & 1 & | & 2 \\
2 & \to & 1 & | & 2 \\
3 & \to & 1 & | & 2 \\
 & \vdots & & & \\
\lfloor xn \rfloor & \to & 1 & | & 2 \\
\lfloor xn \rfloor + 1 & \to & 1 & | & 1,2 \\
\lfloor xn \rfloor + 2 & \to & 1 & | & 2 \\
 & \vdots & & & \\
n & \to & 1 & | & 2 \\
n+1 & \to & n & | & \underbrace{1,\dots,1}_{\lfloor xn \rfloor},2 \\
n+2 & \to & 1 & | & \underbrace{1,\dots,1}_{n-\lfloor xn \rfloor},2 \\
n+3 & \to & 1 & | & 2 \\
 & \vdots & & & \\
n+\lfloor xn \rfloor + 1 & \to & 1 & | & 2 \\
n+\lfloor xn \rfloor + 2 & \to & 1 & | & 1,2 \\
n+\lfloor xn \rfloor + 3 & \to & 1 & | & 2 \\
 & \vdots & & & \\
2n+1 & \to & 1 & | & 2 \\
2(n+1) & \to & n & | & \underbrace{1,\dots,1}_{\lfloor xn \rfloor},2 \\
 & \vdots & & &
\end{array}$$

Chapitre 6

Asymptotics of Fractional Brownian Max Plus Networks

6.1 Introduction

Recall that a standard fractional Brownian motion (FBM) process with Hurst parameter $H \in [1/2, 1)$ is a Gaussian centered process with stationary increments, continuous paths and such that

$$E[F(s)F(t)] = \frac{1}{2}\left(s^{2H} + t^{2H} - |s-t|^{2H}\right),$$

for all $s, t \in \mathbb{R}$.

Queues with FBM input process has received much attention in the literature. Studies [74, 34, 73, 75] have focused primarily on the workload W of a *single* server queue, where $W := \sup_{t>0}(\rho t + \sigma Z_t - t)$, with mean input rate ρ, standard deviation σ, and server capacity 1. A lower bound $\mathbb{P}(W > x)$ was first obtained by Norros in [74], this lower bound has been later shown by Duffield and O'Connell in [34] to be asymptotically exact in logarithm using large deviation principle, further extensions on deriving exact expression and stronger asymptotic estimates are developed in [75] and [73]. All these studies assert that the workload W of a single server queue is asymptotically Weibullian, namely,

$$\log \mathbb{P}(W > x) \sim -\frac{1}{2\sigma^2} x^{2(1-H)} \frac{(1-\rho)^{2H}(1-H)^{2(H-1)}}{H^{2H}}. \tag{6.1}$$

In this paper, we focus on the end-to-end delay in a network setting.

To the best of our knowledge, there exist few results on the tail asymptotic of the end-to-end delay in a network setting. Under the assumptions of independent and identically distributed (i.i.d.) service times and of the existence of moment generating functions, large deviation results were derived in [83] and [44] for stochastic event graphs. In case when the service times are i.i.d. and subexponential, exact asymptotics were obtained in Chapter 3 for stochastic event graphs, where the end-to-end delay has subexponential tail distribution. In the current section, we focus on another cause of heavytailness for the end-to-end delay, namely LRD, what has not been done in a network context.

We consider the steady state distribution of the end-to-end delay of a tagged flow in queueing networks where some of the queues have self-similar cross traffic. We assume that such cross traffic, say at queue i, is modeled by Fractional Brownian Motion (FBM) with Hurst parameter $H_i \in [1/2, 1)$, and is independent of other queues. Note that when $H_i = 0.5$, we have an ordinary Brownian motion model. We assume that at least one of the queues have the Hurst parameter that is strictly greater than 0.5. The arrival process of the tagged flow is renewal. Two types of queueing networks are considered.

We show that the end-to-end delay of the tagged flow in a tandem queueing network is completely dominated by one of the queues. The dominant queue is the one with the maximal Hurst parameter. If several queues have the same maximal Hurst parameter, then we have to compare the ratio $\frac{(1-\rho)^H}{\sigma}$ to determine the dominant queue, where ρ is the load of the queue. We have then

$$\log \mathbb{P}(D > x) \sim \log \mathbb{P}(W > x),$$

where W is the steady-state workload of a single server queue with the same FBM inputs as the dominant queue, which is known to be asymptotically Weibullian.

We also consider general structure of networks that belongs to the event graph framework. We show that the end-to-end delay is still asymptotically Weibullian with the same shape parameter. We also provide upper and lower bounds on the constant that determines the scale parameter of the corresponding Weibull distribution.

This section is based on [68].

6.2 Stochastic Assumptions

6.2.1 Taking Cross Traffic into Account

Consider a network of queues with cross traffic, the case of queues in tandem is illustrated in Figure 6.1. We assume that the service times of customers of the tagged flow are negligible compared to the queueing delays. We see that the time spent in a server is mainly due to the cross traffic. Thus, in our model, in order to analyze the delay of the tagged customers, we define the virtual service times for each tagged customer to be the amount of cross traffic arrived between two successive arrivals of the tagged customers. This (virtual) service time is denoted as σ_n^i (for server i). The resulting queueing system (with such virtual service times) is a single class FIFO queueing networks. In the sequel, we shall thus consider only such FIFO queues, with the (virtual) service times to be possibly self-similar.

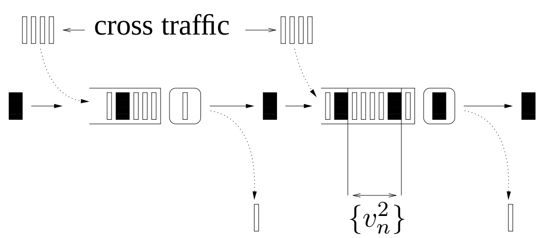


FIG. 6.1 – Queues with cross traffic

6.2.2 Model

For now on, we consider an event graph as described in Section 2.2.2, with $m \leq K$ timed transitions, namely $\mathcal{T}_{timed} = \{t(1), \ldots, t(m)\}$, satisfying the assumptions in Property 3, and with associated recursion :

$$X_n = A_n \otimes X_{n-1} \oplus B_n \otimes T_n$$

of dimension $s \leq KL$. This means that the matrices $\{A_n, B_n\}$ and vectors that are used in the recursion are obtained via two applications f and g such that :

$$\begin{array}{rccc} \mathcal{A}: & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,s)}(\mathbb{R}_{\max}) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{A}(\sigma), \end{array}$$

$$\begin{array}{rccc} \mathcal{B}: & \mathbb{R}_+^m & \to & \mathbb{M}_{(s,1)}(\mathbb{R}_{\max}) \\ & \sigma = (\sigma^1, \ldots, \sigma^m) & \mapsto & \mathcal{B}(\sigma), \end{array}$$

via the formula

$$\begin{aligned} \mathcal{A}(\zeta_n) &= A_n, \\ \mathcal{B}(\zeta_n) &= B_n. \end{aligned}$$

with $\zeta_n = (\sigma_n^{t(1)}, \ldots, \sigma_n^{t(m)})$.

In what follows by feed-forward network, we will understand an event graph such that each communication class is made of only one timed transition.

6.2.3 Model Description and Stochastic Assumptions

We always implicitly assumed that the σ_n^i were non-negative to get a dynamical interpretation of the (max,plus) equations. Nevertheless, the construction of recurrence (2.5) does not require any assumption on the sign of the σ_n^i. We will use the notation $\{\beta_n^i\}$ instead of $\{\sigma_n^i\}$ to make a clear difference if the β_n^i do not have to be non-negative.
In what follows, we will consider :

- a sequence of arrival times $N = \{T_n\}_{n \in \mathbb{N}}$ that is a renewal process : inter-arrival times $\{\tau_n = T_{n+1} - T_n\}$ are i.i.d. We assume moreover that $\mathbb{E}[\tau_0] = 1$ and $T_0 = 0$ (under Palm probability).
- sequences $\{\beta_n^i\}_{n \in \mathbb{N}}, i \in \mathcal{T}$ that are constructed as follows

$$\beta_n^i = S_i(T_{n+1}) - S_i(T_n), \quad \text{with} \quad S_i(t) = \rho_i t + \sigma_i F^i(t), \tag{6.2}$$

  where F^i is a FBM with Hurst parameter $1/2 \leq H_i < 1$. The FBM F^i are independent of each other and $H = \max\{H_i\} > 1/2$. If i is an untimed transition, we take $\rho_1 = \sigma_i = 0$.

Remark 20. The condition on the mean of τ_0 is not restrictive, we can take any renewal process with positive intensity. Moreover, we see that our virtual service times β_n^i are not non-negative but each sequence is self-similar and long range dependent if $H_i > 1/2$.

Stability of the system :
Each sequence $\{\beta_n^i\}_{n \in \mathbb{N}}$ is stationary and ergodic (see [28] Theorem 14.2.1), hence we have :

$$\lim_{k \to \infty} \frac{(D_{[-k+1,0]})^{(s,i)}}{k} = \lim_{k \to \infty} \frac{(D_{[-k+1,0]} \otimes B_{-k})^{(s)}}{k} = \gamma, \tag{6.3}$$

where γ is the top Lyapunov exponent of the sequence $\{A_n\}$ (see [11]). We will always assume that

$$\gamma < 1. \tag{6.4}$$

Under this condition we know that the maximal dater of the event graph is finite a.s. since $(D_{[-k+1,0]}\otimes B_{-k})^{(s)} + T_{-k} \to -\infty$ a.s. We refer to [11] for the following lemma :

Lemma 53. *Consider the matrix $P = A(\rho)$, with $\rho = (\rho_{t(1)}, \ldots, \rho_t(m))$. We denote by $\overline{\rho}$ the maximal (max, plus)-eigenvalue of P. We have :*

$$\max(\rho_{t(i)}) \leq \overline{\rho} \leq \gamma.$$

6.3 Logarithmic Tail Asymptotic : the General Case

6.3.1 Main result

Theorem 15 (Main Result)**.** *Let Z be the stationary maximal dater of the event graph. Consider the set of transitions with maximal Hurst parameter denoted H,*

$$\mathcal{S} := \arg\max\{H_i\},$$

now define the subset of dominant transitions as follows

$$\mathcal{D} := \arg\min_{i\in\mathcal{S}} \left\{ \frac{(1-\rho_i)^H}{\sigma_i} \right\}.$$

We denote by W the workload of a single server queue with the same parameter as one of the transitions in $\mathcal{D}$, then we have

$$C(1-\gamma)^2 \leq \underline{\lim} \frac{\log \mathbb{P}(Z > x)}{\log \mathbb{P}(W > x)} \leq \overline{\lim} \frac{\log \mathbb{P}(Z > x)}{\log \mathbb{P}(W > x)} \leq 1, \tag{6.5}$$

where γ is the top Lyapunov exponent associated to the network, see (6.3), and the constant C satisfies :

$$\frac{(1-\gamma)^{2H}}{\sum_{j\in\mathcal{S}}(\sigma_j)^2} \max_i \frac{(\sigma_i)^2}{(1-\rho_i)^{2H}} \leq C.$$

6.3.2 First result with deterministic arrival times

In this section we construct a graph $\mathcal{G}$ which is slightly different from the graph $\mathcal{G}_{A\cup B}$ of Section 2.2.3. Moreover we introduce weights that are not standard.

Applications $\mathcal{A}$ and $\mathcal{B}$ of section 6.2.3 can be viewed as purely algebraic objects. Following Section 2.3 of [11], we can associate to each application $\mathcal{A}$ and $\mathcal{B}$ a directed graph, respectively $\mathcal{G}_A$ and $\mathcal{G}_B$. For f, the set of nodes is $\{1, \ldots, s\}$ and an arc from i to j is introduced in $\mathcal{G}_A$ if $\mathcal{A}(0)^{(j,i)} \neq \epsilon$. For g, the set of nodes is $\{0, 1, \ldots, s\}$ and an arc form 0 to i is introduced in $\mathcal{G}_B$ if $\mathcal{B}(0)^{(i)} \neq \epsilon$. We denote $\mathcal{G} = \mathcal{G}_A \cup \mathcal{G}_B$. Each coefficient of A and B is a (max,plus)-expression $expr := \bigoplus_{j=1}^{d} \bigotimes_{k\in\mathcal{K}_j} \beta^k$, and we put in $\mathcal{G}$, d copies of the original arc and give to each of them a weight that is the associated set $\mathcal{K}_j$. We obtain a weighted graph $\mathcal{G}_w$. For each arc $e \in \mathcal{G}_w$, $\mathcal{W}(e)$

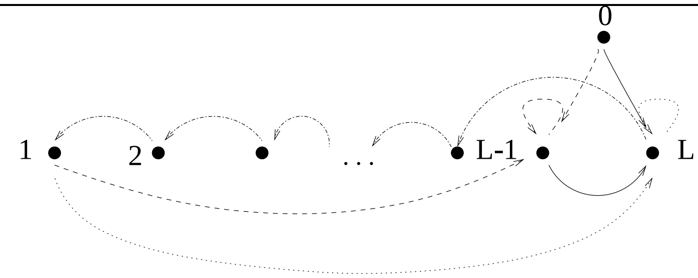


FIG. 6.2 – Graph $\mathcal{G}_w$ for Tandem Queueing Network with Fixed Window Control

denotes the weight of e (i.e. a set of indices). We give here the graph corresponding to queues in tandem with window control (section 2.2.4.0), line style corresponds to the mark : dashed=$\{1\}$, dotted=$\{2\}$, solid=$\{1,2\}$ and dash-dot=$\emptyset$ (observe that in this case $s = L$).

We denote by Ξ the set of paths in $\mathcal{G}_w$ going from node 0 to node s. For $\xi = (e_0, e_1, \dots, e_l) \in \Xi$, we denote :

$$\begin{aligned} |\xi|_l &= l+1, \quad \rho(\xi) = \sum_{i=0}^{l} \sum_{j \in \mathcal{W}(e_i)} \rho_j, \\ \mathcal{F}(\xi) &= \sum_{i=0}^{l} \sum_{j \in \mathcal{W}(e_i)} \sigma_j (F^j(i+1) - F^j(i)), \\ |\xi|_w &= \mathbb{E}\left[\mathcal{F}(\xi)^2\right]. \end{aligned}$$

In the special case $T_n = n$, the maximal dater can be expressed as :

$$\begin{aligned} Z &= \max_{k \geq 0} \left[(D_{[-k,0]} \otimes B_{-k})^{(s)} - k \right] \\ &\overset{\text{dist}}{=} \sup_{\xi \in \Xi} \left[\rho(\xi) + \mathcal{F}(\xi) - (|\xi|_l - 1) \right]. \end{aligned}$$

First we rewrite the event $\{Z > x\}$:

$$\begin{aligned} \{Z > x\} &= \{\exists \xi \in \Xi,\ \rho(\xi) + \mathcal{F}(\xi) - |\xi|_l + 1 > x\} \\ &= \{\sup_{\xi \in \Xi} \frac{\mathcal{F}(\xi)}{x - 1 + |\xi|_l - \rho(\xi)} > 1\}. \end{aligned} \tag{6.6}$$

To consider the event $\{Z > x\}$ or $\{Z > x+1\}$ does not change the asymptotic. For the simplicity of notations, we consider the latter in what follows.
Based on (6.6), to study the tail asymptotic for Z, it suffices to focus on the supremum of the following centered Gaussian process :

$$\left\{ X_\xi^x = \frac{\mathcal{F}(\xi)}{x + |\xi|_l - \rho(\xi)} \right\}_{\xi \in \Xi}.$$

Define :

$$m_x = \mathbb{E}\left[\sup_{\xi \in \Xi} X_\xi^x\right], \quad \text{and} \quad \sigma_x^2 = \sup_{\xi \in \Xi} \mathbb{E}\left(X_\xi^x\right)^2. \tag{6.7}$$

Notice that

$$\sigma_x^2 = \sup_{\xi \in \Xi} \frac{|\xi|_w}{(x + |\xi|_l - \rho(\xi))^2}.$$

We claim the following logarithmic tail asymptotic for Z.

Proposition 28. *Consider Z the stationary maximal dater of the event graph. We assume deterministic arrival times, $T_n = n$. Then we have*

$$(1-\gamma)^2 \le \underline{\lim}(-2\sigma_x^2)\log \mathbb{P}(Z > x) \le \overline{\lim}(-2\sigma_x^2)\log \mathbb{P}(Z > x) \le 1, \tag{6.8}$$

where γ is the top Lyapunov exponent associated to the network.

Remark 21. We will show that for feed-forward networks, the upper bound is indeed tight.

To prove the above main result, we shall need the so-called "Borell's inequality"[1, p.43,p.47] for the supremum of a Gaussian process which we recall below.

6.3.3 Borell's Inequality

In what follows, we shall always assume that T has a countable dense subset and the processes we consider are always separable. We recall that a real stochastic process $\{X_t\}_{t\in T}$ is separable if there is a sequence $\{t_j\}$ of parameter values and a set Λ of probability 0 such that, if A is any closed interval and I is any open interval, the sets

$$\{X_t(\omega) \in A, t \in I \cap T\}, \quad \{X_{t_j}(\omega) \in A, t_j \in I \cap T\},$$

differ by at most a subset of Λ.
The following property can be found in [1], Theorem 2.1 :

Proposition 29. *Let $\{X_t\}_{t\in T}$ be a centered Gaussian process with sample paths bounded a.s. Let $\|X\| = \sup_{t\in T} X_t$. Then $\mathbb{E}\|X\| < \infty$, and for all $\lambda > 0$*

$$\mathbb{P}\{|\|X\| - \mathbb{E}\|X\|| > \lambda\} \le 2\exp\left(-\frac{1}{2}\lambda^2/\sigma_T^2\right), \tag{6.9}$$

where $\sigma_T^2 := \sup_{t\in T} \mathbb{E}X_t^2$. In particular, for all $\lambda > \mathbb{E}\|X\|$, equation (6.9) may be rewritten as follows :

$$\mathbb{P}\{\|X\| > \lambda\} \le 2\exp\left(-\frac{(\lambda - \mathbb{E}\|X\|)^2}{2\sigma_T^2}\right). \tag{6.10}$$

The only assumption made on the parameter space T is that T is totally bounded in the canonical metric. We recall that the canonical metric is defined as follows

$$d(s,t) := \sqrt{\mathbb{E}(X_s - X_t)^2}. \tag{6.11}$$

We denote by $N(\epsilon)$ the smallest number of closed d-balls of radius ϵ that cover T. T is totally bounded if the function $N(\epsilon)$ is finite for all $\epsilon > 0$.
In fact, following proof of theorem 2.1 in [1], we see that this assumption may be relaxed.

Consider a centered Gaussian process with sample paths bounded a.s. $\{X_t\}_{t\in T}$. Let $\{T_n\}_{n\geq 1}$ be an increasing sequence of subsets of T that tends to a dense subset of T containing the sequence $\{t_j\}$ of points satisfying the conditions of the separability definition. We suppose that each T_n is totally bounded in the canonical metric, and we denote $\|X\|_n = \sup_{t\in T_n} X_t$. Then for any n, thanks to property 29, we have $\mathbb{E}\|X\|_n < \infty$, and for all $\lambda > 0$

$$\mathbb{P}\{|\|X\|_n - \mathbb{E}\|X\|_n| > \lambda\} \leq 2\exp\left(-\frac{1}{2}\lambda^2/\sigma_n^2\right), \tag{6.12}$$

where $\sigma_n^2 := \sup_{t\in T_n} \mathbb{E}X_t^2$. Moreover, we have $\sigma_n^2 \uparrow \sigma_T^2$. We consider the case $\sigma_T^2 < \infty$ and first show that $\mathbb{E}\|X\| < \infty$ like in [1].
Suppose $\mathbb{E}\|X\| = \infty$ and choose $\lambda_0 > 0$ such that

$$e^{-\lambda_0^2/2\sigma_T^2} \leq 1/4, \quad \mathbb{P}\left[\sup_{t\in T} X_t < \lambda_0\right] \geq 3/4.$$

Such a constant exists since σ_T is finite and the random variable $\sup_{t\in T} X_t$ is finite a.s.
Now since $\mathbb{E}\|X\|_n \uparrow \mathbb{E}\|X\| = \infty$, we can find n such that $\mathbb{E}\|X\|_n > 2\lambda_0$. Borell's inequality on the space T_n then gives

$$\begin{aligned} \frac{1}{2} \geq 2e^{-\lambda_0^2/2\sigma_T^2} &\geq 2e^{-\lambda_0^2/2\sigma_n^2} \\ &\geq \mathbb{P}[|\|X\|_n - \mathbb{E}\|X\|_n| > \lambda_0] \\ &\geq \mathbb{P}[\mathbb{E}\|X\|_n - \|X\| > \lambda_0] \\ &\geq \mathbb{P}[\lambda_0 > \|X\|] \geq 3/4. \end{aligned}$$

This proved the required contradiction, and so $\mathbb{E}\|X\| < \infty$. Since $\|X\|_n \uparrow \|X\|$ a.s.(separability condition), we have for all $\lambda > 0$

$$\mathbb{P}\{|\|X\| - \mathbb{E}\|X\|| > \lambda\} \leq 2\exp\left(-\frac{1}{2}\lambda^2/\sigma_T^2\right). \tag{6.13}$$

Application 1. *Consider the process* $\{G_t = \frac{Z(t)}{1+t}\}_{t\in[0,\infty)}$. *Since* $\lim_{t\to\infty} Z(t)/t = 0$, *this process is a.s. bounded. Here we take* $T_n = [0,n]$, *and* $T = [0,\infty)$. *Each* T_n *is totally bounded (see [73]) and* $\sigma_T^2 = \sup_{t\geq 0} \mathbb{E}G_t^2 = H^{2H}(1-H)^{2(1-H)}$. *Hence Borell's inequality applies for this process on the whole interval* $[0,\infty)$.

Application 2. *If* T *is countable, then Borell's inequality applies. Just take* T_n *finite and hence totally bounded.*

6.3.4 Auxiliary Results

In this section, we derive some necessary auxiliary results before we prove the main results as claimed in Property 28. Recall that

$$\left\{X_\xi^x = \frac{\mathcal{F}(\xi)}{x + |\xi|_l - \rho(\xi)}\right\}_{\xi\in\Xi},$$

and

$$m_x := \mathbb{E}\left[\sup_{\xi\in\Xi} X_\xi^x\right], \text{ and } \sigma_x^2 = \sup_{\xi\in\Xi} \mathbb{E}\left(X_\xi^x\right)^2. \tag{6.14}$$

The process $\{X_\xi^x\}$ is a centered Gaussian process. The stability condition (6.4) $\gamma < 1$ ensures that $Z < \infty$ almost surely, from which the boundedness of the sample path of process $\{X_\xi^x\}$ follows. In our context, the parameter set Ξ is countable as the countable union of the finite sets : $\Xi_n = \{\xi \in \Xi,\ |\xi|_l = n\}$. Hence Borell's inequality applies (see Application 2 in previous section) and if $m_x \leq 1$ (which is shown in the next lemma), we will have

$$\mathbb{P}\left(\sup_{\xi\in\Xi} X_\xi^x > 1\right) \leq 2\exp\left(-\frac{(1-m_x)^2}{2\sigma_x^2}\right). \tag{6.15}$$

Lemma 54. *We have* $\limsup_{x\to\infty} m_x \leq \gamma < 1$.

Proof.
The function $x \mapsto \sup_\xi \frac{\mathcal{F}(\xi)^+}{x+|\xi|_l-\rho(\xi)}$ is non-increasing since

$$\begin{aligned} x \leq y \quad &\Rightarrow \quad \frac{\mathcal{F}(\xi)^+}{x+|\xi|_l-\rho(\xi)} \geq \frac{\mathcal{F}(\xi)^+}{y+|\xi|_l-\rho(\xi)} \\ &\Rightarrow \quad \sup_\xi \frac{\mathcal{F}(\xi)^+}{x+|\xi|_l-\rho(\xi)} \geq \sup_\xi \frac{\mathcal{F}(\xi)^+}{y+|\xi|_l-\rho(\xi)}. \end{aligned}$$

Thanks to Borell's inequality, we have $\mathbb{E}\left[\sup_{\xi\in\Xi} X_\xi^1\right] < +\infty$ and by symmetry, $\mathbb{P}(\sup_\xi |X_\xi^1| > \lambda) \leq 2\mathbb{P}(\sup_\xi X_\xi^1 > \lambda)$, hence we have $\mathbb{E}\left[\sup_{\xi\in\Xi}(X_\xi^1)^+\right] \leq \mathbb{E}\left[\sup_{\xi\in\Xi}|X_\xi^1|\right] < +\infty$. Then we can use monotone convergence to derive

$$\lim_{x\to\infty} \mathbb{E}\left[\sup_{\xi\in\Xi}(X_\xi^x)^+\right] = \mathbb{E}\left[\lim_{x\to\infty}\sup_{\xi\in\Xi}(X_\xi^x)^+\right].$$

Thanks to (6.3), we know that for any $0 < \epsilon < 1-\gamma$, there exists a finite random variable L, such that

$$|\xi|_l \geq L \quad \Rightarrow \quad \mathcal{F}(\xi) + \rho(\xi) \leq (\gamma+\epsilon)|\xi|_l.$$

Hence for such a path, we have

$$\begin{aligned} \frac{\mathcal{F}(\xi)^+}{x+|\xi|_l-\rho(\xi)} \quad &\leq \quad \frac{(\gamma+\epsilon)|\xi|_l-\rho(\xi)}{x+|\xi|_l-\rho(\xi)} \\ &\leq \quad \gamma+\epsilon. \end{aligned}$$

We define the random variable $M = \sup_{|\xi|_l\leq L}\mathcal{F}(\xi)^+$. We have $M < +\infty$ a.s. and

$$\sup_{\xi\in\Xi}(X_\xi^x)^+ \quad \leq \quad \frac{M}{x} + \gamma + \epsilon.$$

Hence we have $\lim_{x\to\infty}\sup_{\xi\in\Xi}(X_\xi^x)^+ \leq \gamma$, and the result follows since $m_x \leq \mathbb{E}\left[\sup_{\xi\in\Xi}(X_\xi^x)^+\right]$.
□

Remark 22. The bound of Lemma 54 is tight in the sense that there are cases for which we have

$$\lim_{x\to\infty} m_x = \gamma.$$

We take the example of two queues in tandem with window control of size $L = 1$. We recall the recursion equations with the notation of section 2 ($v_n^{1,2} = v_n^1 + v_n^2$) :

$$\begin{pmatrix} x_n^1 \\ x_n^2 \end{pmatrix} = \begin{pmatrix} v_n^1 & v_n^1 \\ v_n^{1,2} & v_n^{1,2} \end{pmatrix} \otimes \begin{pmatrix} x_{n-1}^1 \\ x_{n-1}^2 \end{pmatrix} \oplus \begin{pmatrix} v_n^1 \\ v_n^{1,2} \end{pmatrix} \otimes T_n.$$

Take $\rho_1 = \sigma_1 = 0$ (service in station 1 is instantaneous) and $\rho_2 = 0$. We have

$$(D_{[-k+1,0]} \otimes B_{-k})^{(2)} \quad \stackrel{\text{dist}}{=} \quad \sum_{i=0}^{k} \sigma_2 \left((F^2(i+1) - F^2(i))\right)^+ .$$

Hence we have

$$\gamma = \sigma_2 \mathbb{E}\left[\left(F^2(1)\right)^+\right] > 0.$$

We have $\mathbb{E}\left[\left(F^2(1)\right)^+\right] \le \mathbb{E}\left[1 + (F^2(1))^2\right] = 2$, hence we can choose $\sigma_2 = 1/3$ and we have $\gamma < 1$. Now we see that for $\xi \in \Xi_n$, we have

$$X_\xi^x \ge 1/3 \frac{\sum_{i=0}^n \left(F^2(i+1) - F^2(i)\right)^+}{x+n},$$

hence

$$\begin{aligned} \sup_\xi X_\xi^x &\ge \sup_n 1/3 \frac{\sum_{i=0}^n \left(F^2(i+1) - F^2(i)\right)^+}{x+n} \\ &\ge \lim_{n\to\infty} 1/3 \frac{\sum_{i=0}^n \left(F^2(i+1) - F^2(i)\right)^+}{x+n} \\ &= \gamma > 0. \end{aligned}$$

In this specific case, thanks to Lemma 54, we have $\lim_{x\to\infty} m_x = \gamma$.

If X and Y are centered Gaussian random variables with respective variances σ_X^2 and σ_Y^2, we will write $X \le_{\text{var}} Y \Leftrightarrow \sigma_X \le \sigma_Y$.

Lemma 55. *We have*

$$X_\xi^x \le_{var} \frac{\sum_{i=1}^m \sigma_i F^i(|\xi|_l)}{x + |\xi|_l (1-\gamma)}.$$

Proof.
We first prove that

$$\mathcal{F}(\xi) \le_{\text{var}} \sum_{i=1}^{m} \sigma_i F^i(|\xi|_l). \tag{6.16}$$

Take $t_1 < t_2 < t_3 < t_4$, we use the notation : $\Delta_1 = t_2 - t_1$, $\Delta_2 = t_3 - t_2$, $\Delta_3 = t_4 - t_3$, $\Delta = t_4 - t_1$ and $F(\Delta_1) = F(t_2) - F(t_1), \dots$ We have $F(\Delta_3) + F(\Delta_1) \leq_{\text{var}} F(\Delta)$. This follows from the following inequalities with $1/2 \leq H < 1$ (recall that $a^{2H} + b^{2H} \leq (a+b)^{2H}$) :

$$\begin{aligned}
\mathbb{E}(F(\Delta_3) + F(\Delta_1))^2 &= \mathbb{E}F(\Delta_3)^2 + \mathbb{E}F(\Delta_1)^2 + 2\mathbb{E}F(\Delta_3)F(\Delta_1) \\
&= \Delta_3^{2H} + \Delta_1^{2H} + \Delta^{2H} - (\Delta_1 + \Delta_2)^{2H} + \Delta_2^{2H} - (\Delta_2 + \Delta_3)^{2H} \\
&\leq \Delta^{2H} + \Delta_1^{2H} + \Delta_2^{2H} + \Delta_3^{2H} - \Delta_1^{2H} - \Delta_2^{2H} - \Delta_2^{2H} - \Delta_3^{2H} \\
&\leq \Delta^{2H} = \mathbb{E}F(\Delta)^2.
\end{aligned}$$

We have then $\sum_{i=0}^{|\xi|_l - 1} \mathbb{1}_{j \in \mathcal{W}(e_i)}(F^j(i+1) - F^j(i)) \leq_{\text{var}} F^j(|\xi|_l)$, hence $\mathcal{F}(\xi) \leq_{\text{var}} \sum_{i=1}^m \sigma_i F^i(|\xi|_l)$. By definition, we have $\rho(\xi) \leq |\xi|_l \overline{\rho} \leq |\xi|_l \gamma$ for any $\xi \in \Xi$, and we get $x + |\xi|_l - \rho(\xi) \geq x + |\xi|_l(1-\gamma)$. Now thanks to (6.16), we have

$$\frac{\mathcal{F}(\xi)}{x + |\xi|_l - \rho(\xi)} \leq_{\text{var}} \frac{\sum_{i=1}^m \sigma_i F^i(|\xi|_l)}{x + |\xi|_l(1-\gamma)}$$

□

From lemma 55, we derive

Lemma 56. *We have* $\sigma_x^2 := \sup_{\xi \in \Xi} \mathbb{E}(X_\xi^x)^2 \to 0$ *as* $x \to \infty$. *If we denote* $H := \max\{H_j\}$, *we have* $\sigma_x^2 = \mathcal{O}(x^{2(H-1)})$.

Proof.

Consider the process $C_t^x = \frac{\sum_{i=1}^m \sigma_i F^i(t)}{x + t(1-\gamma)}$, by a change of variable, we have

$$C_{xt/(1-\gamma)}^x = \frac{\sum_{i=1}^m \sigma_i F^i(xt/(1-\gamma))}{x + xt},$$

and the self-similarity of the FBM $F^i(t)$ ensures that the process $\{C_{xt/(1-\gamma)}^x\}$ has the same distribution as the process

$$\frac{\sum_{i=1}^m \sigma_i (x/(1-\gamma))^{H_i} F^i(t)}{x + xt} = \sum_{i=1}^m \frac{\sigma_i x^{H_i - 1}}{(1-\gamma)^{H_i}} G_t^i,$$

with $G_t^i = \frac{F^i(t)}{1+t}$. Thanks to previous lemma, we have $\mathbb{E}(X_\xi^x)^2 \leq \mathbb{E}(C_{|\xi|_l}^x)^2$, hence

$$\sup_{\xi \in \Xi} \mathbb{E}(X_\xi^x)^2 \leq \sup_{t > 0} \mathbb{E}(C_t^x)^2,$$

but we have

$$\sup_{t>0} \mathbb{E}(C_t^x)^2 = \sum_{i=1}^m \frac{(\sigma_i)^2 x^{2(H_i - 1)}}{(1-\gamma)^{2H_i}} \sup_{t>0} \mathbb{E}(G_t^i)^2.$$

A simple calculation gives : $\sup_{t>0} \mathbb{E}(G_t^i)^2 = (H_i)^{2H_i}(1-H_i)^{2(1-H_i)}$, then

$$\sigma_x^2 \leq \sum_{i=1}^m \frac{(\sigma_i)^2 x^{2(H_i-1)}}{(1-\gamma)^{2H_i}} (H_i)^{2H_i}(1-H_i)^{2(1-H_i)}.$$

Thanks to Lemma 3 of [16], we know that each β_n^j is on the diagonal of the matrix A_n. Hence for any $l \geq 1$, there exists a path in Ξ such that $|\xi_0|_l = l, \quad \rho(\xi_0) \geq \rho_j l, \quad |\xi_0|_w \geq (\sigma_j)^2 l^{2H_j}$.
Hence we have

$$\mathbb{E}(X_{\xi_0}^x)^2 \geq \frac{(\sigma_j)^2 l^{2H_j}}{(x + l(1-\rho_j))^2}.$$

Taking an index j, such that $H_j = H$, we have

$$\begin{aligned} \sigma_x^2 &\geq \sup_{l \geq 1} \frac{(\sigma_j)^2 l^{2H}}{(x + l(1-\rho_j))^2} \\ &= (\sigma_j)^2 x^{2(H-1)} \sup_{l \geq 1} \frac{(l/x)^{2H}}{(1 + l/x(1-\rho_j))^2} \\ &\sim (\sigma_j)^2 x^{2(H-1)} \sup_{t > 0} \frac{t^{2H}}{(1 + t(1-\rho_j))^2} \\ &= (\sigma_j)^2 x^{2(H-1)} \frac{H^{2H}}{(1-\rho_j)^{2H}(1-H)^{2(H-1)}}. \end{aligned}$$

This gives the last result. □

6.3.5 Proof of Property 28

Upper bound :
Taking the logarithm of equation (6.15), we obtain

$$2\sigma_x^2 \log \mathbb{P}(Z > x) \leq 2\sigma_x^2 \log(2) - (1 - m_x)^2 .$$

Thanks to lemmas 54 and 56, we have

$$\limsup_{x \to \infty} 2\sigma_x^2 \log \mathbb{P}(Z > x) \leq -(1-\gamma)^2. \tag{6.17}$$

Lower bound :
We denote :

$$\overline{\Phi}(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty e^{-x^2/2} dx.$$

We have :

$$\begin{aligned} \mathbb{P}(Z > x) &= \mathbb{P}\left(\sup_{\xi \in \Xi} \rho(\xi) + \mathcal{F}(\xi) - |\xi|_l > x \right) \\ &\geq \sup_{\xi \in \Xi} \mathbb{P}\left(\mathcal{F}(\xi) > x + |\xi|_l - \rho(\xi) \right) \\ &= \sup_{\xi \in \Xi} \overline{\Phi}\left(\frac{x + |\xi|_l - \rho(\xi)}{\sqrt{\mathbb{E}[\mathcal{F}(\xi)^2]}} \right) \\ &= \overline{\Phi}\left(\inf_{\xi \in \Xi} \frac{x + |\xi|_l - \rho(\xi)}{\sqrt{|\xi|_w}} \right). \end{aligned}$$

Using the approximation $\log \overline{\Phi}(y) \sim -y^2/2$, we obtain

$$\log \mathbb{P}(Z > x) \geq -\inf_{\xi \in \Xi} \frac{(x + |\xi|_l - \rho(\xi))^2}{2|\xi|_w},$$

hence

$$\liminf_{x \to \infty} 2\sigma_x^2 \log \mathbb{P}(Z > x) \geq -1. \tag{6.18}$$

Equations (6.17) and (6.18) give the desired asymptotic for deterministic arrival times.
□

Remark 23. The fact that the bound of Lemma 54is tight, shows the limits of our approach. Even if we can compute the variance σ_x^2, the technique used here can not give an exact asymptotic for the quantity $\log \mathbb{P}(Z > x)$ in these particular cases.

From Property 28, we need, in order to prove Theorem 15, to compute the asymptotic of σ_x^2 and to show that the result still holds with random arrival times. This is done in the two next sections.

6.3.6 Bounds on σ_x^2

To prove Theorem 15 (first with deterministic arrival times), we derive from Property 28,

$$\begin{aligned} \liminf \frac{\log \mathbb{P}(Z > x)}{\log \mathbb{P}(W > x)} &= \frac{\liminf -2\sigma_x^2 \log \mathbb{P}(Z > x)}{\limsup -2\sigma_x^2 \log \mathbb{P}(W > x)} \\ &\geq \frac{(1-\gamma^2)}{\limsup -2\sigma_x^2 \log \mathbb{P}(W > x)}, \end{aligned}$$

and similarly,

$$\limsup \frac{\log \mathbb{P}(Z > x)}{\log \mathbb{P}(W > x)} \leq \frac{1}{\liminf -2\sigma_x^2 \log \mathbb{P}(W > x)}.$$

We have now to compare the quantity σ_x^2 and $\log \mathbb{P}(W > x)$ when x tends to infinity. We recall that

$$\log \mathbb{P}(W > x) \sim -\frac{1}{2\sigma^2} x^{2(1-H)} \frac{(1-\rho)^{2H}(1-H)^{2(H-1)}}{H^{2H}},$$

where $\frac{(1-\rho)^H}{\sigma} = \frac{(1-\rho_j)^H}{\sigma_j}$ for any $j \in \mathcal{S}$.
Thanks to Lemma 56, we know that for an index $j \in \mathcal{S}$, for sufficiently large x,

$$\sigma_x^2 = (\sigma_j)^2 x^{2(H-1)} \frac{H^{2H}}{(1-\rho_j)^{2H}(1-H)^{2(H-1)}}.$$

Hence we have

$$\liminf -2\sigma_x^2 \log \mathbb{P}(W > x) \geq 1. \tag{6.19}$$

In proof of Lemma 56, we showed that

$$\sigma_x^2 \leq \sum_{i=1}^{m} \frac{(\sigma_i)^2 x^{2(H_i-1)}}{(1-\gamma)^{2H_i}} (H_i)^{2H_i} (1-H_i)^{2(1-H_i)}.$$

Hence, we have

$$\limsup -2\sigma_x^2 \log \mathbb{P}(W > x) \leq \frac{(1-\rho)^{2H} \sum_{j\in\mathcal{S}}(\sigma_j)^2}{(1-\gamma)^{2H}\sigma^2}. \quad (6.20)$$

Thanks to inequalities (6.20) and (6.19), we have proved Theorem 15 in the specific case of deterministic arrival times.

6.3.7 From deterministic times to random arrival times

We prove now that the result extends to random arrival times.
We denote $\Psi(\epsilon) = \sup_n (T_{-n} + n(1-\epsilon))$. There exist K and λ such that for sufficiently large x, we have

$$\mathbb{P}[\Psi(\epsilon) \geq x] \leq K e^{-\lambda x}.$$

We have the following decomposition :

$$\begin{aligned} Z &= \sup_\xi \left\{\rho(\xi) + \mathcal{F}(\xi) + T_{-|\xi|_l}\right\} \\ &\leq \sup_\xi \left\{\rho(\xi) + \mathcal{F}(\xi) - |\xi|_l(1-\epsilon)\right\} + \sup_n \left\{T_{-n} + n(1-\epsilon)\right\} \\ &= Z^{(1-\epsilon)} + \Psi(\epsilon). \end{aligned}$$

Notice that provided that $\epsilon < 1-\gamma$, we have $Z^{(1-\epsilon)} < +\infty$ a.s. Therefore,

$$\begin{aligned} \mathbb{P}[Z > x] &\leq \mathbb{P}[Z^{(1-\epsilon)} + \Psi(\epsilon) > x] \\ &= \mathbb{P}[Z^{(1-\epsilon)} + \Psi(\epsilon) > x, \Psi(\epsilon) < \alpha x] + \mathbb{P}[Z^{(1-\epsilon)} + \Psi(\epsilon) > x, \Psi(\epsilon) \geq \alpha x] \\ &\leq \mathbb{P}[\Psi(\epsilon) < \alpha x]\mathbb{P}[Z^{(1-\epsilon)} > (1-\alpha)x] + K e^{-\lambda \alpha x}. \end{aligned}$$

Hence for $\epsilon < 1-\gamma$, we have

$$\log \mathbb{P}[Z > x] \leq \log\left\{\mathbb{P}[\Psi(\epsilon) < \alpha x]\mathbb{P}[Z^{(1-\epsilon)} > (1-\alpha)x] + K e^{-\lambda \alpha x}\right\}. \quad (6.21)$$

We can write $Z^{(1-\epsilon)} = (1-\epsilon)\sup_\xi \{\rho(\xi)/(1-\epsilon) + \mathcal{F}(\xi)/(1-\epsilon) - |\xi|_l\}$. We can apply Property 28 to $Z^{(1-\epsilon)}$ provided that we take for the definition of the β's the process $S_i^{(1-\epsilon)}(t) = \frac{\rho_i}{1-\epsilon}t - \frac{\sigma_i}{(1-\epsilon)}F^i(t)$ in equation (6.2). Hence we have :

$$\limsup \frac{\mathbb{P}(Z^{(1-\epsilon)} > (1-\alpha)x)}{U(x)} \leq 1,$$

with $U(x) = [\frac{1-\alpha}{1-\epsilon}x]^{2(1-H)} \frac{((1-\rho/(1-\epsilon))^{2H}(1-H)^{2(H-1)}}{2(\sigma/(1-\epsilon))^2 H^{2H}}$. Since $H > 1/2$, we have thanks to (6.21)

$$\limsup \frac{\mathbb{P}(Z > x)}{\mathbb{P}(W > x)} \leq (1-\alpha)^{2(1-H)} \left(\frac{1-\epsilon-\rho}{1-\rho}\right)^{2H}.$$

Letting α and ϵ go to 0, we get the desired asymptotic.
Now due to the strong law of large number, we can choose $G \equiv G(\epsilon)$, such that

$$\mathbb{P}(T_{-n} \geq -n(1+\epsilon) - G, \ \forall n \geq 0) \geq 1-\epsilon.$$

We denote this event by K_ϵ. Then, on the event K_ϵ, we have

$$\begin{aligned} Z &= \sup_\xi \{\rho(\xi) + \mathcal{F}(\xi) - |\xi|_l(1+\epsilon) \\ &\quad + T_{-|\xi|_l} - |\xi|_l(1+\epsilon)\} \\ &\geq \sup_\xi \{\rho(\xi) + \mathcal{F}(\xi) - |\xi|_l(1+\epsilon)\} - G \\ &= Z^{(1+\epsilon)} - G. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbb{P}(Z > x) &\geq \mathbb{P}(Z > x, K_\epsilon) \\ &\geq \mathbb{P}(Z^{(1+\epsilon)} > x + G)(1-\epsilon). \end{aligned}$$

and using the same kind of technique as before, we obtain the lower bound. This concludes the proof of Theorem 15.
□

6.4 Logarithmic Tail Asymptotic : the Feed-Forward Case

6.4.1 Main result

Theorem 16 (Main Result)**.** *Let Z be the stationary end-to-end delay associated to a tree network. Consider the set of transitions with maximal Hurst parameter denoted H,*

$$\mathcal{S} := \arg\max\{H_i\},$$

now define the subset of dominant transitions as follows

$$\mathcal{D} := \arg\min_{i\in\mathcal{S}} \left\{ \frac{(1-\rho_i)^H}{\sigma_i} \right\}.$$

We denote by W the workload of a single server queue with the same parameter as one of the transitions in $\mathcal{D}$, then we have

$$\log \mathbb{P}(Z > x) \sim \log \mathbb{P}(W > x). \tag{6.22}$$

Note that the tandem queueing network is a special case of tree networks. The result of Theorem 16 thus holds for tandem queueing network as well.

Lemma 57. *In the case of tree networks, we have $\lim_{x\to\infty} m_x = 0$.*

Proof.
For any path $\xi = (e_0, \ldots, e_l)$, we write $t^0 = 0$ and $t^k = t^{k-1} + \sum_{i=0}^l \mathbb{1}_{\{k\in\mathcal{W}(e_i)\}}$, then we have

$$\begin{aligned} \frac{|\mathcal{F}(\xi)|}{x + |\xi|_l - \rho(\xi)} &= \frac{\left|\sum_{k=1}^m \sigma_k (F^k(t^k) - F^k(t^{k-1}))\right|}{x + |\xi|_l - \rho(\xi)} \\ &\leq \frac{\sum_k \sigma_k |F^k(t^k) - F^k(t^{k-1})|}{x + (1-\overline{\rho})|\xi|_l} \\ &\leq \sum_k \frac{\sigma_k |F^k(t^k) - F^k(t^{k-1}))|}{x + (1-\overline{\rho})(t^k - t^{k-1})}, \end{aligned}$$

where first inequality follows from $\forall \xi,\ \rho(\xi) \leq \overline{\rho}|\xi|_l$. Hence, we have

$$\begin{aligned} \left|\mathbb{E}\left[\sup_{\xi\in\Xi} X_\xi^x\right]\right| &\leq \mathbb{E}\left[\sup_{\xi\in\Xi} |X_\xi^x|\right] \\ &\leq \mathbb{E}\left[\sup_{t^k} \sum_k \frac{\sigma_k |F^k(t^k) - F^k(t^{k-1}))|}{x + (1-\overline{\rho})(t^k - t^{k-1})}\right] \\ &\leq \sum_k \mathbb{E}\left[\sup_u \frac{\sigma_k |F^k(u)|}{x + (1-\overline{\rho})u}\right]. \end{aligned}$$

But we know that $\lim_{u\to\infty} F^k(u)/u = 0$, hence

$$\sup_u \frac{\sigma_k |F^k(u)|}{x + (1-\overline{\rho})u} \to 0 \quad \text{as } x \to \infty,$$

and each term of the sum goes to zero as $x \to \infty$ by monotone convergence. □

6.4.2 Computation of σ_x^2

The case of Single Server Queue

Equation (6.8) takes the simple from :

$$\begin{aligned} &\inf_{\xi\in\Xi} \frac{(x + |\xi|_l - \rho(\xi))^2}{2|\xi|_w} \\ &= \inf_{n\geq 1} \frac{(x + n(1-\rho))^2}{2\sigma^2 n^{2H}} \\ &= \frac{1}{2\sigma^2} x^{2(1-H)} \inf_{n\geq 1} \frac{(1 + n/x(1-\rho))^2}{(n/x)^{2H}} \\ &\sim \frac{1}{2\sigma^2} x^{2(1-H)} \inf_{t>0} \frac{(1 + t(1-\rho))^2}{t^{2H}}. \end{aligned}$$

The infimum is attained in $t^*_{single} = \frac{H}{(1-\rho)(1-H)}$ and we have

$$\log \mathbb{P}(W > x) \sim -\frac{1}{2\sigma^2} x^{2(1-H)} \frac{(1-\rho)^{2H}(1-H)^{2(H-1)}}{H^{2H}}.$$

The case of 2 Queues in Tandem

For $\xi = (e_0, \ldots, e_l) \in \Xi$, we define :

$$m = \sum_{i=0}^{l} \mathbb{1}_{\{1\in\mathcal{W}(e_i)\}}, \quad n = \sum_{i=0}^{l} \mathbb{1}_{\{2\in\mathcal{W}(e_i)\}}.$$

Then we have

$$\mathbb{E}(X_\xi^x)^2 = \frac{(\sigma_1)^2 m^{2H_1} + (\sigma_2)^2 n^{2H_2}}{(x + m + n - \rho_1 m - \rho_2 n)^2}. \tag{6.23}$$

We first suppose that $H_1 > H_2$, then we have

$$\begin{aligned}
\sigma_x^2 &= x^{2(H_1-1)} \cdot \\
&\quad \sup_{m,n} \frac{(\sigma_1)^2(\frac{m}{x})^{2H_1} + (\sigma_2)^2(\frac{n}{x})^{2H_2} x^{2(H_2-H_1)}}{(1+\frac{m+n}{x} - \rho_1 \cdot \frac{m}{x} - \rho_2 \cdot \frac{n}{x})^2} \\
&\sim x^{2(H_1-1)} \sup_{t>0, u\in[0,1]} \frac{(\sigma_1)^2 (ut)^{2H_1}}{(1+(1-\rho(u))t)^2}
\end{aligned}$$

with $\rho(u) = \rho_1 u + \rho_2(1-u)$. The supremum is attained in $u = 1$ and $t^* = \frac{H}{(1-\rho)(1-H)}$, and we obtain

$$\sigma_x^2 \sim x^{2(H-1)} \frac{\sigma^2 H^{2H}}{(1-\rho)^{2H}(1-H)^{2(H-1)}},$$

with $H := H_1$, $\rho := \rho_1$ and $\sigma := \sigma_1$.
The case $H_2 > H_1$ is exactly the same. Hence $H := H_2, \rho := \rho_2$ and $\sigma := \sigma_2$.
We suppose now that $H_1 = H_2 = H$, then we have

$$\begin{aligned}
&\sup_{m,n} \frac{(\sigma_1)^2 m^{2H} + (\sigma_2)^2 n^{2H}}{(x+m+n-\rho_1 m - \rho_2 n)^2} \\
&= x^{2(H-1)} \sup_{m,n} \frac{(\sigma_1)^2(\frac{m}{x})^{2H} + (\sigma_2)^2(\frac{n}{x})^{2H}}{(1+\frac{m+n}{x} - \rho_1 \cdot \frac{m}{x} - \rho_2 \cdot \frac{n}{x})^2} \\
&\sim x^{2(H-1)} \sup_{t>0, u\in[0,1]} \frac{(\sigma_1)^2(ut)^{2H} + (\sigma_2)^2((1-u)t)^{2H}}{(1+(1-\rho(u))t)^2} \\
&= x^{2(H-1)} \sup_{u\in[0,1]} \Big\{ \big[(\sigma_1)^2 u^{2H} + (\sigma_2)^2(1-u)^{2H}\big] \cdot \\
&\qquad \cdot \sup_{t>0} \frac{t^{2H}}{(1+(1-\rho(u))t)^2} \Big\} \\
&= x^{2(H-1)} \sup_{u\in[0,1]} \Big\{ \big[(\sigma_1)^2 u^{2H} + (\sigma_2)^2(1-u)^{2H}\big] \cdot \\
&\qquad \cdot \frac{H^{2H}}{(1-\rho(u))^{2H}(1-H)^{2(H-1)}} \Big\}
\end{aligned}$$

But the function $u \mapsto \frac{(\sigma_1)^2 u^{2H} + (\sigma_2)^2(1-u)^{2H}}{(1-\rho(u))^{2H}}$ is either monotone on $[0,1]$, or non-increasing on $[0,u^*]$ and non-decreasing on $[u^*,1]$ for a certain u^*. Thus the supremum is attained either at 0 or at 1, and we have

$$\sigma_x^2 \sim x^{2(H-1)} \frac{\sigma^2 H^{2H}}{(1-\rho)^{2H}(1-H)^{2(H-1)}}$$

with

$$\frac{(1-\rho)^{2H}}{\sigma^2} = \min\left\{ \frac{(1-\rho_1)^{2H}}{(\sigma_1)^2}, \frac{(1-\rho_2)^{2H}}{(\sigma_2)^2} \right\}.$$

This gives the desired result for 2 queues in tandem.

Remark 24. We recall that we always assume that $H > 1/2$. Nevertheless, for now on, we never use this assumption and in fact for deterministic arrival times, Theorem 1 is still true with $H = 1/2$. This corresponds to Brownian queues and in the case of tandem queues the large deviation technic used in [44] apply and it is straightforward that Theorem 1 of [44] gives exactly the same result as our Theorem 1 for deterministic arrival times.

General Tree Networks

First observe that previous result holds true for k queues in tandem and then we have

$$\sigma_x^2(k) = x^{2(H-1)} \max_{j \in \mathcal{D}^k} \frac{(\sigma_j)^2 H^{2H}}{(1-\rho_j)^{2H}(1-H)^{2(H-1)}},$$

where H is the maximal Hurst parameter and $\mathcal{D}^k$ is defined as in Theorem 15 for the k queues. But for a general tree network, the variance σ_x^2 is the maximum of the variance corresponding to a path going from the root of the tree to any leaf, i.e. a network of queues in tandem, hence we have directly

$$\sigma_x^2 = x^{2(H-1)} \max_{k \in \mathcal{D}} \frac{(\sigma_k)^2 H^{2H}}{(1-\rho_k)^{2H}(1-H)^{2(H-1)}}.$$

Bibliographie

[1] R. Adler. *Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Math. Stat., Lecture notes, 1990.

[2] R. Agrawal, A. Makowski, and P. Nain. On a reduced load equivalence for fluid queues under subexponentiality. *Queueing Systems*, 33 :5–41, 1999.

[3] V. Anantharam. How large delays build up in a $GI/G/1$ queue. *Queueing Systems Theory Appl.*, 5(4) :345–367, 1989.

[4] V. Anantharam. Scheduling strategies and long-range dependence. *Queueing Systems*, 33 :73–89, 1999.

[5] V. Anantharam, P. Heidelberger, and P. Tsoucas. Analysis of rare events in continuous time markov chains via time reversal and fluid approximation. *Rep. RC 16280, IBM*, 1990.

[6] S. Asmussen. A probabilistic look at the wiener-hopf equation. *SIAM Review*, 40 :189–201, 1998.

[7] S. Asmussen and C. Klüppelberg. Large deviations results in the presence of heavy tails. *Stoc. Proc. Appl.*, 64 :265–286, 1996.

[8] S. Asmussen, H. Schmidli, and V. Schmidt. Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability*, 31 :422–447, 1999.

[9] K. Athreya and P. Ney. *Branching Processes*. Springer, Berlin, 1972.

[10] F. Baccelli and P. Brémaud. *Elements of Queueing Theory*. Springer-Verlag, 2003.

[11] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity*. Wiley, 1992. Available at http ://www-rocq.inria.fr/metalau/cohen/SED/book-online.html.

[12] F. Baccelli and S. Foss. Ergodicity of Jackson-type queueing networks. *Queueing Systems Theory Appl.*, 17(1-2) :5–72, 1994.

[13] F. Baccelli and S. Foss. On the saturation rule for the stability of queues. *Journal of Applied Probability*, 32 :494–507, 1995.

[14] F. Baccelli and S. Foss. Moments and tails in monotone-separable stochastic networks. *Ann. Appl. Probab.*, 14(2) :612–650, 2004.

[15] F. Baccelli, S. Foss, and M. Lelarge. Tails in generalized Jackson networks with subexponential service-time distributions. *J. Appl. Probab.*, 42(2) :513–530, 2005.

[16] F. Baccelli, M. Lelarge, and S. Foss. Asymptotics of subexponential max plus networks : the stochastic event graph case. *Queueing Syst.*, 46(1-2) :75–96, 2004.

[17] F. Baccelli, S. Schlegel, and V. Scmidt. Asymptotics of stochastic networks with subexponential service times. *QUESTA*, 33 :205–232, 1999.

[18] D. Bertsimas, I. C. Paschalidis, and J. N. Tsitsiklis. On the large deviations behavior of acyclic networks of $G/G/1$ queues. *Ann. Appl. Probab.*, 8(4) :1027–1069, 1998.

[19] P. Billingsley. *Probability and Measure*. Wiley, 1979.

[20] A. A. Borovkov. Some limit theorems in queueing theory. *Teor. Verojatnost. i Primenen.*, 9 :608–625, 1964.

[21] S. Borst, O. Boxma, and P. Jelenković. Reduced-load equivalence and induced burstiness in GPS queues with long-tailed traffic flows. *Queueing Syst.*, 43(4) :273–306, 2003.

[22] S. Borst, M. Mandjes, and M. van Uitert. Generalized processor sharing queues with heterogeneous traffic classes. *Adv. in Appl. Probab.*, 35(3) :806–845, 2003.

[23] C. S. Chang. Sample path large deviations and intree networks. *Queueing Systems Theory Appl.*, 20 :7–36, 1995.

[24] C. S. Chang and T. Zajic. Effective bandwidths of departure processes from queues with time varying capacities. *IEEE INFOCOM Proceedings*, 1995.

[25] H. Chen. Fluid approximations and stability of multiclass queueing networks : Work-conserving disciplines. *Annals of Applied Probability*, 5 :637–665, 1995.

[26] H. Chen and A. Mandelbaum. Discrete flow networks : Bottlenecks analysis and fluid approximations. *Math. Op. Res.*, 16 :408–446, 1991.

[27] J. W. Cohen. Some results on regular variation for distributions in queueing and fluctuation theory. *J. Appl. Probability*, 10 :343–353, 1973.

[28] I. Cornfeld, S. Fomin, and Y. Sinai. *Ergodic Theory*. Springer, Berlin, 1982.

[29] J. Dai. On positive harris recurrence of multiclass queueing networks : a unified approach via fluid limit models. *Annals of Applied Probability*, 5 :49–77, 1995.

[30] J. Dai. A fluid limit model criterion for the instability of multiclass queueing networks. *Annals of Applied Probability*, 6 :751–757, 1996.

[31] G. de Veciana, C. Courcoubetis, and J. Walrand. Decoupling bandwidths for networks : a decomposition approach to ressource management. *IEEE INFOCOM Proceedings*, 1994.

[32] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag, 1998.

[33] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Probab.*, 20(3) :1147–1166, 1992.

[34] N. Duffield and N. O'Connell. Large deviation and overflow probabilities for the general single-server queue, with applications. *Mathematical Proceedings of the Cambridge Philosophical Society*, 118 :363–375, 1995.

[35] K. Duffy, J. T. Lewis, and W. G. Sullivan. Logarithmic asymptotics for the supremum of a stochastic process. *Ann. Appl. Probab.*, 13(2) :430–445, 2003.

[36] P. Dupuis and R. S. Ellis. The large deviation principle for a general class of queueing systems. I. *Trans. Amer. Math. Soc.*, 347(8) :2689–2751, 1995.

[37] P. Dupuis, R. S. Ellis, and A. Weiss. Large deviations for Markov processes with discontinuous statistics. I. General upper bounds. *Ann. Probab.*, 19(3) :1280–1297, 1991.

[38] P. Dupuis and K. Ramanan. A Skorokhod problem formulation and large deviation analysis of a processor sharing model. *Queueing Systems Theory Appl.*, 28 :109–124, 1998.

[39] P. Dupuis and K. Ramanan. Convex duality and the Skorokhod problem. I, II. *Probab. Theory Related Fields*, 115 :153–195, 197–236, 1999.

[40] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, 2003.

[41] P. Embrechts and N. Veraverbeke. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance : Math and Economics*, 1 :55–72, 1982.

[42] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*. 2005.

[43] S. Foss. Ergodicity of queueing networks. *Sib. Math. J.*, 32 :184–203, 1991.

[44] A. Ganesh. Large deviations of the sojourn time for queues in series. *Annals of Operations Research*, 79 :3–26, 1998.

[45] A. Ganesh and V. Anantharam. Stationary tail probabilities in exponential server tandems with renewal arrivals. *Queueing Systems Theory Appl.*, 22(3-4) :203–247, 1996.

[46] A. Ganesh, N. O'Connell, and D. Wischik. *Big Queues*. Springer-Verlag, 2004. available online.

[47] A. J. Ganesh and N. O'Connell. The linear geodesic property is not generally preserved by a fifo queue. *Annals of Applied Probability*, 8 :98–111, 1998.

[48] J. Garcia. An extension of the contraction principle. *J. Theoret. Probab.*, 17(2) :403–434, 2004.

[49] P. W. Glynn and W. Whitt. Large deviations behavior of counting processes and their inverses. *Queueing Systems Theory Appl.*, 17 :107–128, 1994.

[50] P. W. Glynn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.*, 31A :131–156, 1994. Studies in applied probability.

[51] C. M. Goldie and C. Klüppelberg. Subexponential distributions. In *A practical guide to heavy tails (Santa Barbara, CA, 1995)*, pages 435–459. Birkhäuser Boston, Boston, MA, 1998.

[52] W. Gordon and G. Newell. Closed queueing systems with exponential servers. *Oper. Res.*, 15 :254–265, 1967.

[53] G. Grimmett. Large deviations in subadditive processes and first-passage percolation. In *Particle systems, random media and large deviations (Brunswick, Maine, 1984)*, volume 41 of *Contemp. Math.*, pages 175–194. Amer. Math. Soc., Providence, RI, 1985.

[54] S. Grossmann and B. Yakir. Large deviations for global maxima of independent superadditive processes with negative drift and an application to optimal sequence alignments. *Bernoulli*, 10(5) :829–845, 2004.

[55] J. M. Hammersley. Postulates for subadditive processes. *Ann. Probability*, 2 :652–680, 1974.

[56] J. Harrison and M. Reiman. Reflected brownian motion on an orthant. *The Annals of Prob.*, 9 :302–308, 1981.

[57] T. Huang and K. Sigman. Steady-state asymptotics for tandem, split-match and other feedforward queues with heavy-tailed service. *QUESTA*, 33 :233–260, 1999.

[58] D. L. Iglehart. Extreme values in the $GI/G/1$ queue. *Ann. Math. Statist.*, 43 :627–635, 1972.

[59] I. Ignatiouk-Robert. Large deviations of Jackson networks. *Ann. Appl. Probab.*, 10(3) :962–1001, 2000.

[60] I. Ignatiouk-Robert. Large deviations for processes with discontinuous statistics. *Ann. Probab.*, 33(4) :1479–1508, 2005.

[61] J. Jackson. Jobshop-like queueing systems. *Management Sci.*, 10 :518–527, 1963.

[62] P. Jelenković and A. Lazar. Asymptotic results for multiplexing subexponential on-off processes. *Adv. Appl. Prob.*, 31 :394–421, 1999.

[63] P. Jelenković, P. Momčilović, and B. Zwart. Reduced load equivalence under subexponentiality. *Queueing Systems*, 43 :273–306, 2003.

[64] F. P. Kelly. *Reversibility and stochastic networks*. John Wiley & Sons Ltd., Chichester, 1979. Wiley Series in Probability and Mathematical Statistics.

[65] J. F. C. Kingman. The ergodic theory of subadditive stochastic processes. *J. Roy. Statist. Soc. Ser. B*, 30 :499–510, 1968.

[66] D. Korshunov. On distribution tail of the maximum of a random walk. *Stoch. Proc. Appl.*, 72 :97–103, 1997.

[67] M. Lelarge. Fluid limit of generalized Jackson queueing networks with stationary and ergodic arrivals and service times. *J. Appl. Probab.*, 42(2) :491–512, 2005.

[68] M. Lelarge, Z. Liu, and C. Xia. Asymptotics tail distribution of end-to-end delay in networks of queues with self-similar cross traffic. *IEEE INFOCOM Proceedings*, 2004.

[69] N. Likhanov and R. Mazumdar. Cell loss asymptotics for buffers fed with a large number of independent stationary sources. *Journal of Applied Probability*, 36 :86–96, 1999.

[70] N. Likhanov and R. Mazumdar. Loss asymptotics in large buffers fed by heterogeneous long-tailed sources. *Advances in Applied Probability*, 32 :1168–1189, 2000.

[71] K. Majewski. Single class queueing networks with discrete and fluid customers on the time interval $\mathbb{R}$. *Queueing Syst. Theory Appl.*, 36 :405–435, 2000.

[72] W. Massey. *Single Class Queueing Networks with Discrete and Fluid Customers on the Time Interval* $\mathbb{R}$. PhD thesis, Department of Mathematics, Stanford University, 1981.

[73] L. Massoulié and A. Simonian. Large buffer asymptotics for the queue with fbm input. *Journal of Applied Probability*, 36 :894–906, 1999.

[74] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16 :387–396, 1994.

[75] I. Norros, A. Simonian, D. Veitch, and J. Virtamo. A benes formula for a buffer with fractional brownian input. *9th ITC Specialists Seminar'95 : Teletraffic Modelling and Measurement*, 1995.

[76] N. O'Connell. Large deviations for departures from a shared buffer. *J.Appl.Probab.*, 34 :753–766, 1997.

[77] A. G. Pakes. On the tails of waiting-time distributions. *J. Appl. Probability*, 12(3) :555–564, 1975.

[78] A. Puhalskii. Large deviation analysis of the single server queue. *Queueing Systems Theory Appl.*, 21(1-2) :5–66, 1995.

[79] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer Series in Statistics, 1981.

[80] T. Seppäläinen and J. E. Yukich. Large deviation principles for Euclidean functionals and other nearly additive processes. *Probab. Theory Related Fields*, 120(3) :309–345, 2001.

[81] A. Skorokhod. Stochastic equations for diffusions in a bounded region. *Theor. of Prob. and its Appl.*, 6 :264–274, 1961.

[82] A. Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific J. Math.*, 5 :285–309, 1955.

[83] F. Toomey. Large deviations of products of random topical operators. *Ann. Appl. Probab.*, 12(1) :317–333, 2002.

[84] N. Veraverbeke. Asymptotic behavior of wiener-hopf factors of a random walk. *Stochastic Processes and Applications*, 5 :27–37, 1977.

[85] S. Zachary. A note on veraverbeke's theorem. *Queueing Systems*, 46 :9–14, 2004.

[86] M. P. W. Zerner. Lyapounov exponents and quenched large deviations for multidimensional random walk in random environment. *Ann. Probab.*, 26(4) :1446–1476, 1998.

[87] A. Zwart. *Queueins Systems with Heavy Tails*. PhD thesis, Eindhoven University of Technology, 2001.