Asymptotics of Stochastic Networks with Subexponential Service Times

François Baccelli [1] Sabine Schlegel [2] Volker Schmidt [2]

February 29, 2000

Abstract

We analyse the tail behaviour of stationary response times in the class of open stochastic networks with renewal input admitting a representation as $(\max,+)$–linear systems.

For a K-station tandem network of single server queues with infinite buffer capacity, which is one of the simplest models in this class, we first show that if the tail of the service time distribution of one server, say server $i_0 \in \{1, \ldots, K\}$, is subexponential and heavier than those of the other servers, then the stationary distribution of the response time until the completion of service at server $j \geq i_0$ asymptotically behaves like the stationary response time distribution in an isolated single-server queue with server i_0. Similar asymptotics are given in the case when several service time distributions are subexponential and asymptotically tail-equivalent.

This result is then extended to the asymptotics of general $(\max,+)$–linear systems associated with i.i.d. driving matrices having one (or more) dominant diagonal entry in the subexponential class. In the irreducible case, the asymptotics are surprisingly simple, in comparison with results of the same kind in the Cramér case: the asymptotics only involve the excess distribution of the dominant diagonal entry, the mean value of this entry, the intensity of the arrival process, and the Lyapunov exponent of the sequence of driving matrices. In the reducible case, asymptotics of the same kind, though somewhat more complex, are also obtained.

As a direct application, we give the asymptotics of stationary response times in a class of stochastic Petri nets called event graphs.

[1] ENS, DMI, 45 rue d'Ulm Paris 75005, France

[2] Institute of Stochastics, University of Ulm, D-89069 Ulm, Germany

This is based on the assumption that the firing times are independent and that the tail of the firing times of one of the transitions is subexponential and heavier than those of the others. An extension of these results to non-renewal input processes is discussed. Asymptotics of queue size processes are also considered.

1 Introduction

In the single-station case, the asymptotical tail behaviour of stationary waiting and response time distributions is known; see, for example, [3], [11], [14], [19], [21] and [22]. The multi-station case has recently been studied in several papers. For instance, [7] considers the stationary waiting time and queue length distributions at each node in an acyclic queueing network, [1] deals with long-range dependence in queueing networks with infinite server stations and in [15] tandem queues are studied using large deviation techniques. However, all the results for the multi-server case obtained in these papers are based on the assumption that the service time distributions have exponentially bounded tails.

For tandem queues with heavy-tailed service times there are only a few results so far. To the best of our knowledge, the first results on the matter concern waiting times in tandem queues (see [5]). In this paper, the tail behaviour of stationary total waiting time (i.e. the time between the arrival of a customer in the network and its service in a given station) was analyzed. Further results were then obtained on the asymptotics of the delays in individual nodes of tandem and other feedforward queues in [17]. Asymptotic properties in two-server tandem queues with identical service times at both servers are studied in [8]. For more general queueing networks with heavy-tailed service times, there seem to be no results in the literature.

The present paper is structured as follows: §2 contains all definitions and lemmas on subexponential distributions to be used here.

In §3, we study the asymptotics of the response times in tandem networks of single server queues. When the tail of service times of one server, say server $i_0 \in \{1, \ldots, K\}$, is subexponential and heavier than those of the other servers, we show that the stationary distribution of the response time until the completion of service at server $j \geq i_0$, asymptotically behaves like the stationary waiting time distribution in an isolated single-server queue with server i_0. For sake of clarity, this result is first presented in the two-station network case (§3.2), and then in the general dimension case (§3.3).

In §4 we then extend the results of §3 to networks with a more general structure. For this purpose, in §4.1 we introduce the notion of a $(\max,+)$–linear system which will be used to describe the evolution of state variables in several classes of stochastic networks.

In §4.2 we first give a general result on the asymptotics of state variables in irreducible $(\max,+)$–linear systems with i.i.d. driving matrices, a class of systems introduced in [4]. We then relax the irreducibility assumption in §4.3. Finally, in §4.4 we show how this allows one to compute the asymptotics of response times in stochastic event graphs where the firing times of each transition are i.i.d. and one of the transition firing times is heavier than the others.

Two extensions are discussed in §5. In §5.1 we outline how the assumption of independent interarrival times can be relaxed. In §5.2 we discuss the type of results which can be derived for the asymptotics queue size processes.

2 Subexponential Distributions

Before stating our main results, we recall some properties of the class $\mathcal{S}$ of subexponential distributions introduced in [9]. For convenience we introduce the following notation. For two functions $f(x), g(x)$ the symbol $f(x) \sim g(x)$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$.

Definition 2.1 *A distribution function F on $\mathbb{R}_+ = [0,\infty)$ is called subexponential ($F \in \mathcal{S}$) if $\overline{F^{*2}}(x) \sim 2\overline{F}(x)$, where $\overline{F}(x) = 1 - F(x)$ and F^{*2} denotes the convolution $F * F$.*

For a random variable X with distribution function $F \in \mathcal{S}$ we also write $X \in \mathcal{S}$. Prominent examples of subexponential distributions are Pareto-type distributions (with regularly varying tail function), Weibull distributions or lognormal distributions.

In the proof of our main results, we will make use of the following important properties of the class $\mathcal{S}$.

Lemma 2.1 *Let F, G be two distributions on $\mathbb{R}_+$ and assume that there exists a constant $c \in (0,\infty)$ with $\overline{G}(x) \sim c\overline{F}(x)$. Then, $F \in \mathcal{S}$ if and only if $G \in \mathcal{S}$.*

Lemma 2.2 *Let F, G, H be distributions on $\mathbb{R}_+$ such that $F \in \mathcal{S}$, $\overline{G}(x) \sim c_1\overline{F}(x)$ and $\overline{H}(x) \sim c_2\overline{F}(x)$, where $c_i \in (0,\infty)$ for $i = 1, 2$. Then,*

$$\overline{G * H}(x) \sim (c_1 + c_2)\overline{F}(x)$$

and in particular $\overline{F * H}(x) \sim (1 + c_2)\overline{F}(x)$.

Lemma 2.3 *Let $F \in \mathcal{S}$ and let G and H be distributions on $\mathbb{R}_+$ such that $\overline{G}(x) \sim c_1\overline{F}(x)$ and $\overline{H}(x) \sim c_2\overline{F}(x)$ as $x \to \infty$; $c_i \geq 0$. Then $\overline{(G \times H)}(x) \sim (c_1 + c_2)\overline{F}(x)$, where $(G \times H)(x) = G(x)H(x)$.*

Lemma 2.4 *Let $F \in \mathcal{S}$ and let G be a distribution on $\mathbb{R}_+$ such that $\overline{G}(x) = o(\overline{F}(x))$ as $x \to \infty$. Then $\overline{F * G}(x) \sim \overline{F}(x)$.*

Lemma 2.5 *Let $F \in \mathcal{S}$ and let G be a distribution on $\mathbb{R}_- = (-\infty, 0)$ Then $\overline{F * G}(x) \sim \overline{F}(x)$.*

The results stated in Lemmas 2.1 to 2.5 are well-known. For a proof of Lemmas 2.1 and 2.2, see [10] and [13]. For Lemmas 2.4 and 2.5, see [20]. Furthermore, notice that the result of Lemma 2.3 follows from the following observation. Since $\overline{(G \times H)}(x) = \overline{G}(x) + G(x)\overline{H}(x)$, the assumptions of Lemma 2.3 imply that

$$\lim_{x\to\infty} \frac{\overline{(G \times H)}(x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{\overline{G}(x)}{\overline{F}(x)} + \lim_{x\to\infty} \frac{G(x)\overline{H}(x)}{\overline{G}(x)} = c_1 + c_2 \ .$$

The following result is an immediate consequence of Lemmas 2.1 and 2.4.

Lemma 2.6 *Let $F \in \mathcal{S}$ and let G and H be distributions on $\mathbb{R}_+$ such that $\overline{G}(x) \sim \overline{F}(x)$ and $\overline{H}(x) = o(\overline{F}(x))$ as $x \to \infty$. Then $\overline{H * G}(x) \sim \overline{F}(x)$.*

Proof Note that $\overline{G}(x) \sim \overline{F}(x)$ implies $G \in \mathcal{S}$ by Lemma 2.1. Now,

$$\lim_{x\to\infty} \frac{1 - H(x)}{1 - G(x)} = \lim_{x\to\infty} \frac{1 - H(x)}{1 - F(x)} \frac{1 - F(x)}{1 - G(x)} = 0$$

and thus, from Lemma 2.4, $\overline{H * G}(x) \sim \overline{G}(x)$, which yields

$$\lim_{x\to\infty} \frac{1 - H * G(x)}{1 - F(x)} = \lim_{x\to\infty} \frac{1 - H * G(x)}{1 - G(x)} \frac{1 - G(x)}{1 - F(x)} = 1 \ . \qquad \square$$

Lemma 2.7 *Let $F \in \mathcal{S}$. Then, $\overline{F^{*n}}(x) \sim \overline{F^n}$ for each $n \in \mathbb{N}$, where $F^{*n} = F^{*(n-1)} * F$.*

For a *proof* of Lemma 2.7, see [13].

Lemma 2.8 *If $F \in \mathcal{S}$ then $F^{*n}, F^n \in \mathcal{S}$ for each $n \in \mathbb{N}$.*

Proof Note that, by Lemma 2.7,

$$\lim_{x\to\infty} \frac{1 - F^{*2n}(x)}{1 - F^{*n}(x)} = \lim_{x\to\infty} \frac{1 - F^{2n}(x)}{1 - F^{n}(x)} = \lim_{x\to\infty} (1 + F^n(x)) = 2.$$

Thus, the statement follows from Lemmas 2.1 and 2.7. □

For the proofs of the statements given later in §3 and §4 we need further generalizations of Lemma 2.4.

Lemma 2.9 *Let $F \in \mathcal{S}$ and let G be a distribution on $\mathbb{R}_+$ such that $\overline{G}(x) = o(\overline{F}(x))$ as $x \to \infty$. Then, for each $k \in \mathbb{N}$ and each $n \in \mathbb{N}$,*

$$\overline{F^{*k} * G^{*n}}(x) \sim \overline{F^{*k}}(x) \ .$$

Proof Let $k \in \mathbb{N}$ be fixed. Then, due to Lemmas 2.1, 2.7 and 2.8, we have $F^{*k} \in \mathcal{S}$. Now,

$$\lim_{x\to\infty} \frac{1 - G(x)}{1 - F^{*k}(x)} = \lim_{x\to\infty} \frac{1 - G(x)}{1 - F(x)} \, \frac{1 - F(x)}{1 - F^{*k}(x)} = 0 \ ,$$

and thus Lemma 2.4 implies $\overline{F^{*k} * G}(x) \sim \overline{F^{*k}}(x)$, which shows that the assertion is true for $n = 1$. Assume that the assertion holds for some $n \in \mathbb{N}$. Then,

$$\lim_{x\to\infty} \frac{1 - G(x)}{1 - F^{*k} * G^{*n}(x)} = \lim_{x\to\infty} \frac{1 - G(x)}{1 - F(x)} \, \frac{1 - F(x)}{1 - F^{*k}(x)} \, \frac{1 - F^{*k}(x)}{1 - F^{*k} * G^{*n}(x)} = 0 \ ,$$

i.e. $\overline{G}(x) = o(\overline{F^{*k} * G^{*n}}(x))$. Since $F^{*k} * G^{*n} \in \mathcal{S}$ by hypothesis of induction and Lemma 2.1, we finally obtain

$$\begin{aligned} &\lim_{x\to\infty} \frac{1 - F^{*k} * G^{*(n+1)}(x)}{1 - F^{*k}(x)} \\ &= \lim_{x\to\infty} \frac{1 - F^{*k} * G^{*n} * G(x)}{1 - F^{*k} * G^{*n}(x)} \, \frac{1 - F^{*k} * G^{*n}(x)}{1 - F^{*k}(x)} = 1 \end{aligned}$$

by Lemma 2.4 and hypothesis of induction. □

In the same vein, we have the following result.

Lemma 2.10 *Let $F \in \mathcal{S}$ and let $G_1, \ldots, G_n$, $n \geq 1$, be distributions on $\mathbb{R}_+$ such that $\overline{G}_i(x) = o(\overline{F}(x))$ as $x \to \infty$ for each $i \in \{1, \ldots, n\}$. Then, for each $k \in \mathbb{N}$, $l_1, \ldots, l_n \in \mathbb{N}$*

$$\overline{F^{*k} * G_1^{*l_1} * \ldots * G_n^{*l_n}}(x) \sim \overline{F^{*k}}(x) \ .$$

The *proof* of Lemma 2.10 is omitted since it is similar to the proof of Lemma 2.9; see also [5].

Let F be a distribution on $\mathbb{R}_+$ with positive and finite mean m. We use the notation F^s for the *integrated tail distribution* of F defined by

$$F^s(x) = \frac{1}{m} \int_0^x (1 - F(y))dy \ .$$

As shown below, if $F \in \mathcal{S}$, then the tail of F^s is heavier than that of F.

Lemma 2.11 *If $F \in \mathcal{S}$, then $\overline{F}(x) = o(\overline{F^s}(x))$ as $x \to \infty$.*

Proof Let $F \in \mathcal{S}$. Then for each $h > 0$,

$$\lim_{x \to \infty} \frac{1 - F(x)}{1 - F(x + h)} = 1, \tag{1}$$

as a consequence of Lemma 2.4, where we interpret h as a constant random variable. So for all $h > 0$,

$$\begin{aligned} 0 \ &\leq \ \limsup_{x \to \infty} \frac{1 - F(x)}{\int_x^\infty (1 - F(y))dy} \leq \limsup_{x \to \infty} \frac{1 - F(x)}{\int_x^{x+h} (1 - F(y))dy} \\ &\leq \ \limsup_{x \to \infty} \frac{1 - F(x)}{h(1 - F(x + h))} = \frac{1}{h} \ . \end{aligned}$$

Now, letting $h \to \infty$, the assertion follows. □

We now only consider distributions F on $\mathbb{R}_+$ such that $F(0) = 0$ and $F(x) < 1$ for all $x \in \mathbb{R}_+$ and introduce two further classes of distributions. The definition of class $\mathcal{S}^*$ is motivated by the need for $F^s \in \mathcal{S}$, which does not follow from $F \in \mathcal{S}$. For more on this, see [18].

Definition 2.2 (a) *We say that F belongs to the class $\mathcal{S}^*$ if F has finite mean m and*

$$\lim_{x \to \infty} \int_0^x \frac{\overline{F}(x - y)}{\overline{F}(x)} \overline{F}(y) \, dy = 2m. \tag{2}$$

(b) *We say that F belongs to $\mathcal{L}$ (or that F is long-tailed) if for all $y \in \mathbb{R}$,*

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1 \ . \tag{3}$$

It follows from (1) that $\mathcal{S} \subset \mathcal{L}$. Furthermore, the following results hold true.

Lemma 2.12 *Let $F, G \in \mathcal{L}$. Suppose that there exist constants $c_-, c_+ \in (0,\infty)$ such that*

$$c_- \le \frac{\overline{G}(x)}{\overline{F}(x)} \le c_+ \tag{4}$$

for all $x \ge 0$. Then, $F \in \mathcal{S}^*$ if and only if $G \in \mathcal{S}^*$.

Lemma 2.13 *If $F \in \mathcal{S}^*$, then $F \in \mathcal{S}$ and $F^s \in \mathcal{S}$.*

For a *proof* of Lemmas 2.12 and 2.13, see [18].

Lemma 2.14 *If $F \in \mathcal{S}^*$ then $F^n \in \mathcal{S}^*$ and $F^{*n} \in \mathcal{S}^*$ for each $n \in \mathbb{N}$.*

Proof From Lemma 2.13 we have that $F \in \mathcal{L}$. But this implies $F^n \in \mathcal{L}$, because for all $y \in \mathbb{R}$

$$\begin{aligned}
&\lim_{x\to\infty} \frac{1-F^n(x-y)}{1-F^n(x)} \\
&= \lim_{x\to\infty} \frac{1-F(x-y)}{1-F(x)} \frac{1+F(x-y)+F^2(x-y)+\cdots+F^{n-1}(x-y)}{1+F(x)+F^2(x)+\cdots+F^{n-1}(x)} \\
&= 1 \ ,
\end{aligned}$$

where we have used that $\lim_{x\to\infty} F(x) = 1$. In addition,

$$\frac{1}{n} \le \frac{1-F(x)}{(1-F(x))(1+F(x)+\cdots+F^{n-1}(x))} = \frac{1-F(x)}{1-F^n(x)} \le 1 \ ,$$

and this completes the proof due to Lemmas 2.12 and 2.7. □

Standard examples of distributions belonging to class S^* are Weibull distributions, lognormal distributions and Pareto-type distributions.

We further need the following asymptotic result on the tail distribution of the supremum of a random walk with negative drift and subexponential increments. Let $\{\xi_n\}$ be a sequence of positive i.i.d. random variables with distribution function F and let $\{\tau_n\}$ be a sequence of positive i.i.d. random variables which are independent of $\{\xi_n\}$. Consider the random walk $\{\sum_{k=1}^n (\xi_k - \tau_k), n \ge 0\}$.

Lemma 2.15 *If $F^s \in \mathcal{S}$ and $\rho = \mathbb{E}\xi/\mathbb{E}\tau < 1$, then*

$$\mathbb{P}\Big(\sup_{n\geq 1} \sum_{k=1}^{n} (\xi_k - \tau_k) > x\Big) \sim \frac{\rho}{1-\rho}\overline{F^s}(x) \ .$$

Lemma 2.15 is in essence due to Veraverbeke [23]. Other proofs can also be found in [14] and [21]. Earlier proofs based on more specific assumptions can also be found in [11] and [19].

A thorough treatment of subexponential distributions and their properties is given, for example, in [21].

3 Response Times in Tandem Networks

We first show how the response times in tandem queues can be described by a set of recurrence equations involving only max and + operations. This example is presented first so as to introduce the reader to the more abstract framework of $(\max, +)$–linear systems of which it is a particular case. We also remark that the results of the present section can be easily concluded from the asymptotics which have been derived in [5] for stationary waiting times in tandem queues. The reason for this is the fact that the total waiting time from external arrival to the tandem network until the beginning of service in a given node and the service time itself are independent random variables. Note that this passage from waiting times to response times given cannot be reversed, since response times and service times are not independent.

However, in §3.2 below we give a separate proof for the asymptotics of response times in 2–station tandem queues, since we are convinced that this will help the reader to become familiar with the general idea of our proving technique which is then extended in §4 to the framework of $(\max, +)$–linear systems.

3.1 Queues in Tandem

Consider a system of K single-server FIFO queues in tandem. Assume that all servers have an infinite capacity buffer initially empty. Let server one be fed by the (Palm) stationary process $\{T_n\}$ of external arrival epochs. This process is assumed to be a renewal process with finite and positive intensity λ. We assume that the service times $\{\sigma_n^i\}$ at server i ($i = 1, \ldots, K$) are i.i.d. with distribution F_i and finite mean m_i, that the sequences $\{\sigma_n^1\}, \ldots, \{\sigma_n^K\}$

and the sequence of interarrival times $\{\tau_n\} = \{T_{n+1} - T_n\}$ are independent and that the *stability condition*

$$\max\{m_1, \dots, m_K\} < \mathbb{E}\tau \tag{5}$$

holds. We also assume that there exists a distribution $F \in \mathcal{S}^*$ with finite mean m, and constants $c_i \in [0, \infty)$ with $\sum_{i=1}^K c_i > 0$ such that for all $i \in \{1, \dots, K\}$

$$\lim_{x\to\infty} \frac{\overline{F}_i(x)}{\overline{F}(x)} = c_i \; . \tag{6}$$

Let $i_0 = \min\{1 \le i \le K : c_i > 0\}$.

Let R_n^i denote the time which elapses between the n-th arrival epoch T_n and the completion of the n-th service in queue i $(i = 1, \dots, K)$ under the assumption that the system is initially empty. Then, under the above assumptions, R_n^i converges in total variation to the distribution of the random variable:

$$R^i = \sup_{n\ge 0}(D_n^i + T_{-n}) \; , \tag{7}$$

where

$$D_n^i = \max_{1\le l_n \le \dots \le l_1 \le i} \Big\{ \sum_{k=l_1}^{i} \sigma_0^k + \sum_{k=l_2}^{l_1} \sigma_{-1}^k + \dots + \sum_{k=l_n}^{l_{n-1}} \sigma_{-(n-1)}^k + \sum_{k=1}^{l_n} \sigma_{-n}^k \Big\} \tag{8}$$

(see, for example, [6]).

Note that R^1 is just the stationary response time in a single-server FIFO queue. The tail behaviour of its distribution can easily be obtained from the stationary waiting time distribution which is well-known if $F_1 \in \mathcal{S}^*$; see [3], [14] and [21]. We will thus concentrate on the case $i \ge 2$.

3.2 Two Queues in Tandem

We first consider the case of only two queues in tandem and investigate the tail of the stationary response time R^2 until the completion of service at server two.

Theorem 3.1 *Let $K = 2$, $i_0 \in \{1, 2\}$. Then, under Conditions* (5) *and* (6),

$$\begin{aligned} \max\Big\{\frac{c_1\lambda m}{1-\lambda m_1}, \frac{c_2\lambda m}{1-\lambda m_2}\Big\} &\le \liminf_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s}(x)} \\ \le \limsup_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s}(x)} &\le \frac{(c_1 + c_2)\lambda m}{1 - \lambda \max\{m_1, m_2\}} \; . \end{aligned} \tag{9}$$

Proof We derive an upper and a lower bound for $\mathbb{P}(R^2 > x)$. For $i = 2$, Formula (8) can be rewritten as

$$D_n^2 = \max\Big\{ \max_{0\le j\le n} \{\sigma_0^2 + \sigma_{-1}^2 + \cdots + \sigma_{-j}^2 + \sigma_{-j}^1 + \sigma_{-j-1}^1 + \cdots + \sigma_{-n}^1\} \Big\} . \quad (10)$$

Now, let $k \le n$ be fixed. D_n^2 can be bounded from above as follows:

$$D_n^2 \le \sum_{j=1}^{[\frac{n+2}{k}]+1} I_j^{(k)},$$

where $[x]$ denotes the largest integer smaller than x and where $I_j^{(k)}$ is the random variable defined by

$$\begin{aligned} I_j^{(k)} \quad = \quad & \max\{\sigma^1_{-(j-1)k} + \sigma^1_{-((j-1)k+1)} + \cdots + \sigma^1_{-(jk-1)}, \\ & \sigma^2_{-((j-1)k+1)} + \sigma^1_{-((j-1)k+1)} + \cdots + \sigma^1_{-(jk-1)}, \\ & \cdots, \\ & \sigma^2_{-((j-1)k+1)} + \sigma^2_{-((j-1)k+2)} + \cdots + \sigma^2_{-jk}\} . \end{aligned}$$

These random variables are i.i.d. in j, and such that $I_j^{(k)}$ is equivalent in law to $D_k^2 - \sigma_0^2 - \sigma_{-k}^1$. Let $\varepsilon > 0$ be such that $\max\{m_1, m_2\} + \varepsilon < \mathbb{E}\tau$. Since

$$\lim_{k\to\infty} \frac{\mathbb{E}I^{(k)}}{k} = \lim_{k\to\infty} \Big(\frac{\mathbb{E}D_k^2}{k} - \frac{\mathbb{E}\sigma^2}{k} - \frac{\mathbb{E}\sigma^1}{k} \Big) = \max\{m_1, m_2\} \ ,$$

we can choose $k > 2,\ k \in \mathbb{N}$ large enough to have

$$\frac{\mathbb{E}I^{(k)}}{k} < \max\{m_1, m_2\} + \varepsilon < \mathbb{E}\tau \ . \quad (11)$$

We then obtain the following upper bound for R^2:

$$\begin{aligned} R^2 &= \sup_{n\ge 0}(D_n^2 - T_{-n}) \\ &\le \ \max\Big\{ \max_{1\le n\le k} D_n^2, \sup_{n>k}(\max_{0\le j\le n} \{\sigma_0^2 + \sigma_{-1}^2 \\ &\qquad + \cdots + \sigma_{-j}^2 + \sigma_{-j}^1 + \cdots + \sigma_{-n}^1\} - T_{-n})\Big\} \\ &\le \ \max\Big\{ D_k^2, \sup_{n>k}(\max_{0\le j\le k} \{\sigma_0^2 + \sigma_{-1}^2 + \cdots + \sigma_{-j}^2 + \sigma_{-j}^1 + \cdots + \sigma_{-k}^1\} \end{aligned}$$

$$
\begin{aligned}
&\quad + \max_{k+1\le j\le n}\{\sigma^2_{-(k+1)}+\cdots+\sigma^2_{-j}+\sigma^1_{-j}+\cdots+\sigma^1_{-n}\}-T_{-n})\Big\} \\
&= \max\Big\{D_k^2, D_k^2+\sup_{n>k}(\max_{k+1\le j\le n}\{\sigma^2_{-(k+1)}+\cdots+\sigma^2_{-j}+\sigma^1_{-j} \\
&\qquad\qquad +\cdots+\sigma^1_{-n}\}-T_{-n})\Big\} \\
&= D_k^2+\max\Big\{0,\sup_{n>k}(\max_{k+1\le j\le n}\{\sigma^2_{-(k+1)}+\cdots+\sigma^2_{-j}+\sigma^1_{-j} \\
&\qquad\qquad +\cdots+\sigma^1_{-n}\}-T_{-n})\Big\} \\
&\le D_k^2+\max\Big\{0,\sup_{n>k}(\max_{k+1\le j\le k+n+1}\{\sigma^2_{-(k+1)}+\cdots+\sigma^2_{-j}+\sigma^1_{-j} \\
&\qquad +\cdots+\sigma^1_{-n}+\sigma^1_{-(n+1)}+\cdots+\sigma^1_{-(n+k+1)}\}-T_{-n})\Big\} \\
&= D_k^2+\tilde R^2
\end{aligned}
$$

with D_k^2 and $\tilde R^2$ independent, where

$$
\begin{aligned}
\tilde R^2 &\overset{\mathrm{d}}{=} \max\Big\{0,\sup_{n>k}(D_n^2-T_{-n})\Big\} \\
&\le \max\Big\{0,\sup_{n>k}(I_1^{(k)}+I_2^{(k)}+I_3^{(k)} \\
&\qquad\qquad + \sum_{j=1}^{[\frac{n+2}{k}]-2}(I_{j+3}^{(k)}-(T_{-kj}-T_{-k(j-1)})))\Big\} \\
&= \max\Big\{0, I_1^{(k)}+I_2^{(k)}+I_3^{(k)} \\
&\qquad\qquad +\sup_{n>k}\sum_{j=1}^{[\frac{n+2}{k}]-2}(I_{j+3}^{(k)}-(T_{-kj}-T_{-k(j-1)}))\Big\} \\
&= I_1^{(k)}+I_2^{(k)}+I_3^{(k)}+\sup_n\sum_{j=1}^{n}(I_{j+2}^{(k)}-(T_{-kj}-T_{-k(j-1)}))\,.
\end{aligned}
$$

Note that the random walk $\{\sum_{j=1}^n(I_{j+3}^{(k)}-(T_{-kj}-T_{-k(j-1)})),\, n\ge 1\}$ has a negative drift and thus we obtain from Lemma 2.15 that

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_n\sum_{j=1}^{n}(I_{j+3}^{(k)}-(T_{-kj}-T_{-k(j-1)}))>x\Big) \qquad (12)\\
&\sim \frac{c_1+c_2}{k/\lambda-\mathbb{E}I^{(k)}}\int_x^\infty \overline{F^{*k}}(y)dy \sim (c_1+c_2)\frac{\lambda m}{1-\lambda\frac{\mathbb{E}I^{(k)}}{k}}\overline{F^s}(x)\ ,
\end{aligned}
$$

as $x \to \infty$, where we have used that

$$\mathbb{P}(I^{(k)} > x) \sim (c_1 + c_2)\overline{F^{*k}}(x) \ . \tag{13}$$

This can be seen as follows:

$$\begin{aligned} &\limsup_{x\to\infty} \frac{\mathbb{P}(I^{(k)} > x)}{\overline{F^{*k}}(x)} \\ \le \ & \limsup_{x\to\infty} \frac{\mathbb{P}(\sigma_{-1}^2 + \cdots + \sigma_{-k}^2 + \sigma_0^1 + \cdots + \sigma_{-(k-1)}^1 > x)}{\overline{F^{*k}}(x)} \\ = \ & \limsup_{x\to\infty} \frac{1 - F_2^{*k} * F_1^{*k}(x)}{1 - F^{*k}(x)} = c_1 + c_2 \ . \end{aligned}$$

The last step follows from Lemma 2.2 if both $c_1, c_2 > 0$ and from Lemma 2.9 otherwise. On the other hand,

$$\begin{aligned} &\liminf_{x\to\infty} \frac{\mathbb{P}(I^{(k)} > x)}{\overline{F^{*k}}(x)} \\ \ge \ & \liminf_{x\to\infty} \frac{\mathbb{P}(\max\{\sigma_{-1}^2 + \cdots + \sigma_{-k}^2, \sigma_0^1 + \cdots + \sigma_{-(k-1)}^1\} > x)}{1 - F^{*k}(x)} = c_1 + c_2 \end{aligned}$$

and thus (13) follows. In particular, this implies $\mathbb{P}(I^{(k)} \le x) \in \mathcal{S}^*$ by Lemmas 2.12 and 2.14. Further, due to Lemma 2.11,

$$\lim_{x\to\infty} \frac{\mathbb{P}(I_1^{(k)} + I_2^{(k)} + I_3^{(k)} > x)}{\overline{F^s}(x)} = \lim_{x\to\infty} \frac{\mathbb{P}(I_1^{(k)} + I_2^{(k)} + I_3^{(k)} > x)}{\overline{F}(x)} \frac{\overline{F}(x)}{\overline{F^s}(x)} = 0$$

and

$$\begin{aligned} &\limsup_{x\to\infty} \frac{\mathbb{P}(D_k^2 > x)}{\overline{F^s}(x)} \\ \le \ & \limsup_{x\to\infty} \frac{\mathbb{P}(\sigma_0^2 + \sigma_{-1}^2 + \cdots + \sigma_{-k}^2 + \sigma_0^1 + \cdots + \sigma_{-k}^1 > x)}{\overline{F^s}(x)} \\ = \ & \limsup_{x\to\infty} \frac{1 - F_2^{*(k+1)} * F_1^{*(k+1)}(x)}{1 - F(x)} \frac{1 - F(x)}{1 - F^s(x)} = 0 \end{aligned}$$

and therefore, using Lemma 2.10,

$$\begin{aligned} \limsup_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s}(x)} \ &\le \ \limsup_{x\to\infty} \frac{\mathbb{P}(D_k^2 + \tilde{R}^2 > x)}{\overline{F^s}(x)} \\ &= \ \frac{(c_1 + c_2)\lambda m}{1 - \lambda \frac{\mathbb{E} I^{(k)}}{k}} < \frac{(c_1 + c_2)\lambda m}{1 - \lambda \max\{m_1, m_2\} - \varepsilon} \ . \end{aligned}$$

Letting $\varepsilon \downarrow 0$, we finally obtain

$$\limsup_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s}(x)} \le (c_1 + c_2)\frac{\lambda m}{1 - \lambda \max\{m_1, m_2\}} \; .$$

For the lower bound, note that for $i = 1, 2$

$$\begin{aligned} \mathbb{P}(R^2 > x) &\ge \mathbb{P}\Big(\sup_n (\sigma^i_{-1} + \sigma^i_{-2} + \cdots + \sigma^i_{-n}) - T_{-n}) > x\Big) \\ &= \mathbb{P}\Big(\sup_n \sum_{j=1}^{n} (\sigma^i_{-j} - \tau_{-j}) > x\Big) \end{aligned}$$

and this implies

$$\liminf_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s}(x)} \ge c_i \frac{\lambda m}{1 - \lambda m_i} \; . \qquad \square$$

Corollary 3.1 *Let $K = 2$, $i_0 \in \{1, 2\}$, $F_{i_0} \in \mathcal{S}^*$ and $\overline{F}_i(x) = o(\overline{F_{i_0}}(x))$ for $i \neq i_0$. Then, under Condition* (5),

$$\frac{\lambda m_{i_0}}{1 - \lambda m_{i_0}} \le \liminf_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s_{i_0}}(x)} \le \limsup_{x\to\infty} \frac{\mathbb{P}(R^2 > x)}{\overline{F^s_{i_0}}(x)} \le \frac{\lambda m_{i_0}}{1 - \lambda \max_{1\le j\le 2} m_j} \; .$$

In particular, if $m_{i_0} = \max\{m_1, m_2\}$, then the above inequalities imply the asymptotical tail equivalence $\mathbb{P}(R^2 > x) \sim \lambda m_{i_0}(1 - \lambda m_{i_0})^{-1}\overline{F^s_{i_0}}(x)$, and the distribution of R^2 is in $\mathcal{S}$.

3.3 K-Station Tandem Network

Here is the extension to K queues in tandem.

Theorem 3.2 *Assume that Conditions* (5) *and* (6) *hold. Then, for each $i \in \{i_0, \dots, K\}$,*

$$\begin{aligned} \max_{1\le j\le i}\Big\{\frac{c_j \lambda m}{1 - \lambda m_j}\Big\} &\le \liminf_{x\to\infty} \frac{\mathbb{P}(R^i > x)}{\overline{F^s}(x)} \\ \le \; \limsup_{x\to\infty} \frac{\mathbb{P}(R^i > x)}{\overline{F^s}(x)} &\le \frac{(c_1 + \cdots + c_i)\lambda m}{1 - \lambda \max_{1\le j\le i} m_j} \; . \end{aligned} \qquad (14)$$

The *proof* of Theorem 3.2 is omitted since it is similar to the proof of Theorem 3.1; see also [5]. Besides this, notice that Theorem 3.2 is a special case of Theorem 4.2 given below.

Corollary 3.2 *Let $i_0 \in \{1, \ldots, K\}$, $F_{i_0} \in \mathcal{S}^*$ and $\overline{F}_i(x) = o(\overline{F_{i_0}}(x))$ for all $i \neq i_0$. Then, under Condition* (5), *for each $i \in \{i_0, \ldots, K\}$,*

$$\frac{\lambda m_{i_0}}{1 - \lambda m_{i_0}} \leq \liminf_{x \to \infty} \frac{\mathbb{P}(R^i > x)}{\overline{F^s_{i_0}}(x)} \leq \limsup_{x \to \infty} \frac{\mathbb{P}(R^i > x)}{\overline{F^s_{i_0}}(x)} \leq \frac{\lambda m_{i_0}}{1 - \lambda \max_{1 \leq j \leq i} m_j} .$$

In particular, if $m_{i_0} = \max_{1 \leq j \leq K} m_j$, then the above inequalities imply the asymptotical tail equivalence $\mathbb{P}(R^i > x) \sim \lambda m_{i_0}(1 - \lambda m_{i_0})^{-1} \overline{F^s_{i_0}}(x)$, and the distribution of R^i is in $\mathcal{S}$.

4 Tail Behaviour of $(\max, +)$-Linear Systems

4.1 Model Description and Notation

Let K be an arbitrary fixed natural number. We consider a class of stochastic networks with a (Palm) stationary process $\{T_n,\ n \in \mathbb{Z}\}$ of external arrival epochs, with positive and finite intensity $\lambda > 0$, where $T_0 = 0$ and $\lambda^{-1} = \mathbb{E}(T_{n+1} - T_n)$.

The state variables of the network are the K-dimensional random vectors $X_1, X_2, \ldots$, which are assumed to satisfy an evolution equation of the form

$$X_{n+1} = A_n \otimes X_n \oplus B_{n+1} \otimes T_{n+1} . \tag{15}$$

Here, the coordinates X_n^i represent absolute times like the beginning or completion of the n-th service in the i-th queue. The other notations are those of [4]: $\otimes$ denotes the product of matrices in the $(\max, +)$-algebra, and $\oplus$ the sum of matrices in this algebra. Furthermore, A_n is a $K \times K$-square random matrix, B_n is a K-dimensional random (column) vector; all entries of A_n and B_n are either a.s. non-negative or a.s. equal to $-\infty$, an assumption which we will refer to as the *support assumption* in what follows. The vector B_n will in addition be assumed to have at least one coordinate not a.s. equal to $-\infty$. Assume that the interarrival times $\tau_n = T_{n+1} - T_n$ and the matrices (A_n, B_n) are jointly stationary and ergodic. The matrix $C(x)$, $x \in \mathbb{R}$, is the $K \times K$-square matrix defined as follows

$$C_{i,j}(x) = \begin{cases} -x & \text{if } i = j; \\ -\infty & \text{otherwise.} \end{cases}$$

Starting from (15), we will investigate the recursion

$$S_{n+1} = A_n \otimes C(\tau_n) \otimes S_n \oplus B_{n+1}, \tag{16}$$

for the state variables S_n, where $S_n = X_n - T_n$ is a vector of either waiting times or response times. In this case, whenever the interarrival times τ_n and the matrices (A_n, B_n) are jointly stationary, the stationary solution $S = \lim_{n\to\infty} S_n$ of (16), where the limit means convergence in distribution, admits a representation of the form

$$S = \sup_{n\geq 0}(D_n + T_{-n}) , \tag{17}$$

where $D_0 = B_0$, $T_0 = 0$ and

$$D_n = A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n} \otimes B_{-n} ; \tag{18}$$

see, for example, [4] and [6].

In this section, we investigate such $(\max, +)$-linear systems with a general topological structure. We also show (see §4.3) how response times in tandem queues fit into this framework. The analysis will be subdivided into two cases of increasing generality. We first consider the irreducible case.

4.2 Irreducible Case with i.i.d. Matrices

Under the assumptions of §4.1, consider the equivalence relation over the set of coordinates $\{1, \ldots, K\}$: $i \equiv j$ if there exist n and n' such that

$$(A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n})_{i,j} \geq 0 \quad \text{and } (A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n'})_{j,i} \geq 0.$$

Notice that due to the support assumption, when they exist, the integers n and n' can always be taken smaller than or equal to K.

Denote the equivalence classes of the above equivalence relation by $\mathcal{C}_1, \ldots, \mathcal{C}_d$ for some $d \in \{1, \ldots, K\}$. We first consider the case where there is only one such equivalence class $\mathcal{C} = \{1, \ldots, K\}$. In this case, we call the network *irreducible*. Notice that then, for all i and j,

$$\frac{(A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n})_{i,j}}{n} \to \gamma \quad \text{both } a.s. \text{ and in } L_1 \tag{19}$$

and

$$\frac{\max_{i,j}(A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n})_{i,j}}{n} \to \gamma \quad \text{both } a.s. \text{ and in } L_1, \tag{20}$$

where γ is a constant referred to as the $(\max, +)$–*Lyapunov exponent* of the sequence $\{A_n\}$. For more on this, see [4].

The following further assumption will be made on the topology of the network:

T There exists a coordinate i_0 such that $(A_0)_{i_0,i_0} \geq 0$ a.s. and, for all $i \in \{1, \dots, K\}$ and $n \geq 1$,

$$(A_{-1} \otimes \cdots \otimes A_{-n} \otimes B_{-n})^i \leq (A_{-1} \otimes \cdots \otimes A_{-n} \otimes A_{-(n+1)})_{i,i_0} \quad a.s. \tag{21}$$

Notice that due to the stationarity assumption, **T** implies that for all s,

$$(A_{s-1} \otimes \cdots \otimes A_{s-n} \otimes B_{s-n})^i \leq (A_{s-1} \otimes \cdots \otimes A_{s-n} \otimes A_{s-(n+1)})_{i,i_0} \quad a.s. \tag{22}$$

The stochastic assumptions (in addition to the above stationarity and ergodicity) are that

S1 $\gamma < \mathbb{E}\tau$, which generalizes the stability condition (5);

S2 the matrices $\{A_n\}$ are i.i.d.;

S3 the arrival process $\{T_n\}$ is a renewal process, independent of $\{A_n\}$;

S4 there exists a non-empty subset $\mathcal{H} \subset \{1, \dots, K\}$ such that

- $B_0^i \geq 0$ and $(A_0)_{i,i} \geq 0$, for all $i \in \mathcal{H}$;
- the random variables $\{(A_0)_{i,i},\ i \in \mathcal{H}\}$ are independent;
- there exists a distribution function $F \in \mathcal{S}^*$ with finite mean m such that for all $i \in \mathcal{H}$, $\overline{F}_i(x) = \mathbb{P}((A_0)_{i,i} > x) \sim c_i\overline{F}(x)$, with $c_i \geq 0$ and $\sum_{i\in\mathcal{H}} c_i > 0$;
- for each pair (s, n) of positive integers with $s \leq n$, there exists a random variable $H_{s,n}$, which is independent of the sequence $\{(A_{-s})_{i,i}, \dots, (A_{-n})_{i,i},\ i \in \mathcal{H}\}$, such that

$$\max_{i,j}\{(A_{-s} \otimes \cdots \otimes A_{-n})_{i,j}\} \leq H_{s,n} + Z_{\mathcal{H}} \tag{23}$$

  and

$$\mathbb{P}(H_{s,n} + Z_{\mathcal{H}} > x) \sim \mathbb{P}(Z_{\mathcal{H}} > x)\,, \tag{24}$$

  where $Z_{\mathcal{H}} = \sum_{i\in\mathcal{H}}((A_{-s})_{i,i} + \cdots + (A_{-n})_{i,i})$.

Observe that the model of tandem queues considered in §3 does not fit into this framework, since it is not irreducible. However, in the following modified model of a tandem network with blocking, there is only one equivalence class of coordinates.

Example Suppose that there are two single server stations in tandem fed by an input stream $\{T_n\}$. Each station has both an input buffer and an output buffer. There is a blocking mechanism which prevents that more than two customers are present in station 1 at any time, with a similar constraint on the total number of customers in station 2. This is implemented as follows: as long as there are two customers in station 1, entrance into the input buffer of station 1 is forbidden, and arrivals of the input stream are buffered in an external buffer of infinite capacity. Similarly, as long as there are two customers in station 2, transfer from station 1 to station 2 is forbiden, and customers of station 1 are buffered in the output buffer of station 1. This basic blocking mechanism can be found in several applications: it is referred to as *window flow control* in communications (e.g. in the TCP protocol), *kanban blocking* in manufacturing etc. The Petri net description of this system is given in Figure 1. As we will show below, when denoting the

Figure 1: Tandem queues with blocking

service times in station i by $\{\sigma_n^i\}$, there exists a representation of type (16) for the response times R_n^i, (here again R_n^i represents the time between the arrival of customer n and the time this customer completes its service time in station i, $i = 1, 2$), with the following matrix for A_n (here $\varepsilon = -\infty$, the neutral element for $\oplus$)

$$A_n = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \\ \sigma_{n+1}^1 & \sigma_{n+1}^1 & \varepsilon & \varepsilon & \varepsilon & \sigma_{n+1}^1 & \varepsilon \\ \sigma_{n+1}^1 & \sigma_{n+1}^1 & 0 & \varepsilon & \varepsilon & \sigma_{n+1}^1 & 0 \\ \sigma_{n+1}^{1,2} & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 & \sigma_{n+1}^2 & \varepsilon & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 \\ \sigma_{n+1}^{1,2} & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 & \sigma_{n+1}^2 & 0 & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \end{pmatrix}$$

and with the vector B_n given by $B_n^\top = (0,\ \sigma_n^1,\ \sigma_n^1,\ \sigma_n^{1,2},\ \sigma_n^{1,2},\ \varepsilon,\ \varepsilon)$, where $\sigma_n^{1,2} = \sigma_n^1 + \sigma_n^2$. Thus, in this case $K = 7$. Notice that the second coordinate of S_n is equal to R_n^1, and the fourth is equal to R_n^2. It is not difficult to see that when service times are i.i.d. and independent of $\{T_n\}$, the above assumptions are satisfied for this model. Indeed, it follows from the specific form of A_n that this model is irreducible. The topological assumption **T** is also satisfied: this is easy to check that the first column of A_{n-1} is an upper bound for B_n, so that for all $i \in \{1, \ldots, K\}$ and $1 \leq s \leq n$

$$(A_{-s} \otimes \cdots \otimes A_{-n} \otimes B_{-n})^i \leq (A_{-s} \otimes \cdots \otimes A_{-n} \otimes A_{-(n+1)})_{i,1} .$$

Suppose that **S1** and **S3** hold. We now prove that the other stochastic assumptions are also satisfied. It is easy to check that if the service times are i.i.d., then the matrices A_n are i.i.d. indeed. In addition, using the fact that each entry of A_n is bounded from above by $\sigma^1_{n+1} + \sigma^2_{n+1}$, it is not difficult to show that for all n,

$$\max_{i,j}\{(A_{-s} \otimes \cdots \otimes A_{-n})_{i,j}\} \leq \sum_{k=s-1}^{n-1} (\sigma^2_{-k} + \sigma^1_{-k})$$

so that (23) holds for the following instances of set $\mathcal{H}$:

- $\mathcal{H} = \{4\}$; in this case, we have $(A_{-k})_{4,4} = \sigma^2_{-k+1}$ and we can take $H_{s,n} = \sum_{k=s-1}^{n-1} \sigma^1_{-k}$. Let $F_i(x) = \mathbb{P}(\sigma^i_0 \leq x)$, $i = 1, 2$. If we assume that $F_2 \in S^*$, and that $\overline{F_1}(x) = o(\overline{F_2}(x))$ as $x \to \infty$, then, by Lemma 2.9, (24) will hold. Of course, a symmetrical situation exchanging the roles of F_1 and F_2 would lead to similar conclusions for the case $\mathcal{H} = \{2\}$.

- $\mathcal{H} = \{2, 4\}$; in this case, we have $(A_{-k})_{2,2} = \sigma^1_{-k+1}$, $(A_{-k})_{4,4} = \sigma^2_{-k+1}$ and we can take $H_{s,n} = 0$. With the same notations as above, if we assume that $\overline{F}_i(x) \sim c_i \overline{F}(x)$ with $F \in \mathcal{S}^*$, and with $c_i > 0$, then (24) trivially holds.

Note that blocking networks of the same type with an arbitrary number of stations would satisfy the above assumptions in case of i.i.d. service times, when some of them have a tail heavier than the others in the above sense.

Theorem 4.1 *Consider the stationary state variables S given in* (17)*. Under the above topological and stochastic assumptions, for all $j \in \{1, \ldots, K\}$, $S^j \in \mathcal{S}$ and*

$$\mathbb{P}(S^j > x) \sim \frac{\lambda m \sum_{i \in \mathcal{H}} c_i}{1 - \lambda\gamma} \overline{F^s}(x) \tag{25}$$

where γ is the $(\max, +)$*–Lyapunov exponent of the sequence $\{A_n\}$ given in* (19) *and* (20)*. In particular, if $\mathcal{H}$ has a single element, say i_0, such that $c_{i_0} > 0$ and $c_i = 0$ for all $i \in \mathcal{H} \setminus \{i_0\}$, then for all j*

$$\mathbb{P}(S^j > x) \sim \frac{\lambda m_{i_0}}{1 - \lambda\gamma} \overline{F^s_{i_0}}(x), \tag{26}$$

where m_{i_0} denotes the mean of F_{i_0}.

Remark The constant γ in (25) is the expected stationary cycle time of the associated autonomous (max, +)–linear system $X_{n+1} = A_n \otimes X_n$ (see [4]). In more concrete terms, in our blocking network model, γ^{-1} is the stationary throughput of the system with saturated input, or equivalently the stationary throughput of the associated closed system with 4 customers.

Proof of Theorem 4.1 We first show the lower bound corresponding to (25). Let $i \in \mathcal{H}$ be fixed. We will need the following variables: for $p \geq q$, let

$$g^i_{p,q} = (A_{p-1} \otimes A_{p-2} \otimes \cdots \otimes A_{q-1})_{i,i}$$

and

$$g_{p,q} = \max_{i \in \mathcal{H}} \{(A_{p-1} \otimes A_{p-2} \otimes \cdots \otimes A_{q-1})_{i,i}\} .$$

Note that if $p \geq 2K + 2 + q$, then for all i as above

$$g^i_{p,q} \geq g_{p-K-1,q+K+1} . \tag{27}$$

This follows from the relation

$$\begin{aligned} g^i_{p,q} \quad \geq \quad & \max_{i' \in \mathcal{H}} \{(A_{p-1} \otimes \cdots \otimes A_{p-K-1})_{i,i'} + g^{i'}_{p-K-1,q+K+1} \\ & \qquad + (A_{q+K-1} \otimes \cdots \otimes A_{q-1})_{i',i}\} \end{aligned}$$

and from the fact that for all $j \in \{1, \ldots, K\}$, for all $i' \in \mathcal{H}$, and for all integers h

$$\begin{aligned} (A_h \otimes A_{h-1} \otimes \cdots \otimes A_{h-K})_{j,i'} &\geq 0 \, , \\ (A_h \otimes A_{h-1} \otimes \cdots \otimes A_{h-K})_{i',j} &\geq 0 \, , \end{aligned}$$

which is deduced from the irreducibility assumption and the assumption that $(A_n)_{i',i'} \geq 0$ for $i' \in \mathcal{H}$.

We now show that under the foregoing assumptions, for each $k > K$, the following bound holds for all coordinates j of the vector D_n defined in (18) and for all $i \in \mathcal{H}$:

$$D^j_n \geq \sum_{l=1}^{[\frac{n}{k}]-1} g^i_{-lk,-((l+1)k-1)}, \tag{28}$$

where $\sum_{l=1}^{r} = 0$, for $r < 1$ by convention. This follows from the bound

$$\begin{aligned} D^j_n \quad \geq \quad & (A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-k})_{j,i} + (A_{-k-1} \otimes \cdots \otimes A_{-2k})_{i,i} + \cdots \\ & + (A_{-k([\frac{n}{k}]-1)-1} \otimes \cdots \otimes A_{-k[\frac{n}{k}]})_{i,i} \\ & + (A_{-k[\frac{n}{k}]-1} \otimes \cdots \otimes A_{-n} \otimes B_{-n})^i, \end{aligned}$$

from the fact that

$$(A_{-1} \otimes \cdots \otimes A_{-k})_{j,i} \geq 0$$

(see above) and from the fact that

$$(A_{-k[\frac{n}{k}]-1} \otimes \cdots \otimes A_{-n} \otimes B_{-n})^i \geq 0$$

(which we obtain from the assumption that both $(A_n)_{i,i} \geq 0$ and $B_n^i \geq 0$).

Let $a(l,k) = g_{-lk-K-1,-(l+1)k+K+2}$. Combining (27) and (28), we get for $k > 2K+3$

$$D_n^j \geq \sum_{l=1}^{[\frac{n}{k}]-1} a(l,k)\,, \tag{29}$$

and using (17) we deduce from this that for all j,

$$S^j \geq T_{-k} + \sup_{p \geq 1} \Big(\sum_{l=1}^{p} a(l,k) + (T_{-(p+1)k} - T_{-k}) \Big)\,.$$

Let $G(x)$ denote the distribution function of $a(0,k)$, which turns out to be such that $G^s \in \mathcal{S}$. Indeed, we can prove that

$$\overline{G}(x) \sim (k - 2K - 3) \sum_{i \in \mathcal{H}} c_i \overline{F}(x)\,. \tag{30}$$

This follows from the two bounds

$$\begin{aligned}
a(0,k) &\geq \max_{i \in \mathcal{H}} \Big\{ \sum_{l=K+2}^{-k-K-1} (A_{-l})_{i,i} \Big\} \\
a(0,k) &\leq \sum_{i \in \mathcal{H}} \sum_{l=K+2}^{-k-K-1} (A_{-l})_{i,i} + H_{K+2,-k-K-1}
\end{aligned}$$

(where the last bound is obtained from **S4**), using **S4** together with a repeated use of Lemmas 2.2 and 2.4 for the upper bound and Lemma 2.3 for the lower bound. Furthermore, from (30) we obtain

$$\begin{aligned}
\overline{G^s}(x) &= \frac{1}{\mathbb{E}a(0,k)} \int_x^\infty \overline{G}(y)\,dy \\
&\sim \frac{(k-2K-3)\sum_{i \in \mathcal{H}} c_i}{\mathbb{E}a(0,k)} \int_x^\infty \overline{F}(y)\,dy \\
&= \frac{(k-2K-3)m \sum_{i \in \mathcal{H}} c_i}{\mathbb{E}a(0,k)} \overline{F^s}(x)\,,
\end{aligned}$$

and thus $G^s \in \mathcal{S}$ by Lemma 2.1.

Using now the fact that the random variables $a(l,k)$ are i.i.d., the renewal assumption on the input point process, and Lemmas 2.5 and 2.15, we obtain that for k large such that $\mathbb{E}a(0,k) < k/\lambda$,

$$\liminf_{x\to\infty} \frac{\mathbb{P}(S^j > x)}{\overline{G^s}(x)} \geq \frac{\frac{\lambda}{k}\mathbb{E}a(0,k)}{1-\frac{\lambda}{k}\mathbb{E}a(0,k)} .$$

Thus,

$$\liminf_{x\to\infty} \frac{\mathbb{P}(S^j > x)}{\overline{F^s}(x)} \geq \frac{\frac{\lambda}{k}\mathbb{E}a(0,k)}{1-\frac{\lambda}{k}\mathbb{E}a(0,k)} \frac{(k-2K-3)m\sum_{i\in\mathcal{H}} c_i}{\mathbb{E}a(0,k)} .$$

But in view of (19), we have $\lim_{k\to\infty}(\mathbb{E}a(0,k)/k) = \gamma$, so that

$$\liminf_{x\to\infty} \frac{\mathbb{P}(S^j > x)}{\overline{F^s}(x)} \geq \frac{\lambda m \sum_{i\in\mathcal{H}} c_i}{1-\lambda\gamma} . \tag{31}$$

We next derive an upper bound corresponding to (31). For $p \geq q$, let

$$G_{p,q} = \max_{i,j}\{(A_{p-1}\otimes A_{p-2}\otimes\cdots\otimes A_{q-1})_{i,j}\} .$$

Let k be a fixed integer, and let h be the smallest integer such that $n+1 \leq hk$. We get from $\mathbf{T}$ that

$$\begin{aligned} D_n^j &\leq (A_{-1}\otimes\cdots\otimes A_{-n-1})_{j,i_0} \\ &\leq (A_{-1}\otimes\cdots\otimes A_{-n-1})_{j,i_0} + \sum_{d=-n-2}^{-hk} (A_d)_{i_0,i_0} \\ &\leq \max_{j,i}(A_{-1}\otimes\cdots\otimes A_{-hk})_{j,i} \\ &\leq \sum_{l=0}^{h-1} G_{-lk,-((l+1)k-1)}. \end{aligned}$$

We deduce from this and from the monotonicity of T_{-n} that

$$S^j \leq \sup_{p\geq 0}\Big\{\sum_{l=0}^{p} G_{-lk,-((l+1)k-1)} + T_{-pk}\Big\} .$$

For the same type of reasons as above, (23) and (24) yield

$$\mathbb{P}(G_{0,-(k-1)} > x) \sim k\sum_{i\in\mathcal{H}} c_i\overline{F}(x). \tag{32}$$

Since the random variables $G_{-lk,-((l+1)k-1)}$ are i.i.d., we prove as above that

$$\limsup_{x\to\infty} \frac{\mathbb{P}(S^j > x)}{\overline{F^s}(x)} \leq \frac{\frac{\lambda}{k}\mathbb{E}G_{0,-(k-1)}}{1-\frac{\lambda}{k}\mathbb{E}G_{0,-(k-1)}} \frac{km\sum_{i\in\mathcal{H}} c_i}{\mathbb{E}G_{0,-(k-1)}} ,$$

for all sufficiently large k. But in view of (20),

$$\lim_{k\to\infty} \frac{\mathbb{E}G_{0,-(k-1)}}{k} = \gamma,$$

so that

$$\limsup_{x\to\infty} \frac{\mathbb{P}(S^j > x)}{\overline{F^s}(x)} \leq \frac{\lambda m \sum_{i\in\mathcal{H}} c_i}{1-\lambda\gamma} . \qquad \square$$

4.3 General Case with i.i.d. Matrices

In the general case where the network is not necessarily irreducible, there exists

- a partition of the set of coordinates $\{1,\ldots,K\}$ into d equivalence classes $\mathcal{C}_1,\ldots,\mathcal{C}_d$, for some $d \geq 1$;
- a partial order $\prec$ on the set $\{1,\ldots,d\}$, which will be assumed to be such that $l \prec k$ only for l and k such that $1 \leq l \leq k \leq d$; we will denote $\pi(k)$ (resp. $\sigma(k)$) the set of all predecessors (resp. successors) of k with respect to this partial order:

  $$\pi(k) = \{l \leq k \text{ such that } l \prec k\}, \quad \sigma(k) = \{l \geq k \text{ such that } k \prec l\};$$

- non-negative constants γ_p, $p = 1,\ldots,d$;

such that

- for all $i \in \mathcal{C}_k$ and $j \in \mathcal{C}_l$ with $k \in \pi(l)$, there exists n $leqK$ such that $(A_{-1}\otimes A_{-2}\otimes\cdots\otimes A_{-n})_{j,i} \geq 0$; this follows from the support assumption;
- for all $i \in \mathcal{C}_k$ and $j \in \mathcal{C}_l$ with $k \notin \pi(l)$, and for all n,

  $$(A_{-1}\otimes A_{-2}\otimes\cdots\otimes A_{-n})_{i,j} = \varepsilon; \qquad (33)$$

- for all $i \in \mathcal{C}_k$ and $j \in \mathcal{C}_l$, with $l \in \pi(k)$,

$$\frac{(A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n})_{i,j}}{n} \rightarrow_{n\to\infty} \eta_{k,l} \quad \text{both } a.s. \text{ and in } L_1, \tag{34}$$

where

$$\eta_{k,l} = \max_{h \in \pi(k) \cap \sigma(l} \gamma_h \tag{35}$$

and γ_h is defined as in (19) with $\mathcal{C}$ replaced by $\mathcal{C}_h$. More generally, for each non-empty subset D of the set of coordinates and for each mapping $f : \{1, \ldots, K\} \to 2^{\{1,\ldots,K\}}$

$$\max_{i\in D} \max_{j \in f(i)} \frac{(A_{-1} \otimes A_{-2} \otimes \cdots \otimes A_{-n})_{i,j}}{n} \rightarrow_{n} \Delta_{D,f} \quad \text{both } a.s. \text{ and in } L_1, \tag{36}$$

with $\Delta_{D,f} = \max_{i\in D} \max_{j\in f(i)} \eta_{c(i),c(j)}$, where $c(i)$ denotes the equivalence class of i.

For more on this, see the discussion on strongly connected components in [4]. The constant γ_p is often referred to as the $(\max, +)$-Lyapunov exponent of class $\mathcal{C}_p$, and $\gamma = \max_{1\le p\le d} \gamma_p$ is referred to as the *maximal Lyapunov exponent* of $\{A_n\}$.

The network will be assumed to satisfy the same topological assumption **T**, with the requirement that $i_0 \in \mathcal{C}_1$ and that $1 \in \pi(r)$ for all $1 \le r \le d$, and the same stochastic assumptions **S1**-**S4** as in §4.2 with **S4** replaced by

S4′ there exists a subset $\mathcal{H}$ of $\{1, \ldots, K\}$ with elements $i^p_1, \ldots, i^p_{n_p} \in \mathcal{C}_p$, $p = 1, \ldots, d$ and such that

- $B^i_0 \ge 0$ and $(A_0)_{i,i} \ge 0$, for all $i \in \mathcal{H}$;
- the random variables $\{(A_0)_{i,i},\ i \in \mathcal{H}\}$ are independent;
- there exists a distribution function $F \in \mathcal{S}^*$ with finite mean m such that for all $i \in \mathcal{H}$, $F_i(x) = \mathbb{P}((A_0)_{i,i} \le x)$ is such that $\overline{F}_i(x) \sim c_i \overline{F}(x)$, with $c_i \ge 0$ and $\sum_{i\in\mathcal{H}} c_i > 0$;
- for each pair (s, n) of positive integers with $s \le n$, there exists a random variable $H_{s,n}$, which is independent of the sequences $\{(A_{-s})_{i,i}, \ldots, (A_{-n})_{i,i},\ i \in \mathcal{H}\}$, such that for all $1 \le d_0 \le d$

$$\max_{\substack{i\in\mathcal{C}_l, j\in\mathcal{C}_k,\\ k,l\in\pi(d_0)}} \{(A_{-s} \otimes \cdots \otimes A_{-n})_{i,j}\} \quad \le \quad H_{s,n} + Z_{\mathcal{H},d_0} \tag{37}$$

and

$$\mathbb{P}(H_{s,n} + Z_{\mathcal{H},d_0} > x) \sim \mathbb{P}(Z_{\mathcal{H},d_0} > x)\,, \tag{38}$$

where $Z_{\mathcal{H},d_0} = \sum_{r\in\pi(d_0)} \sum_{i\in\mathcal{H}^r} ((A_{-s})_{i,i} + \cdots + (A_{-n})_{i,i})$ and $\mathcal{H}^r$ denotes the subset of elements of $\mathcal{H}$ which belong to $\mathcal{C}_r$, $1 \le r \le d$.

We are now in a position to state the following result for the more general case considered in this subsection.

Theorem 4.2 *Let $d_0 \in \{1,\ldots,d\}$ be such that there is an $r \in \pi(d_0)$ with at least one element i_0 of $\mathcal{C}_r$ in $\mathcal{H}$ such that $c_{i_0} > 0$. Under the foregoing topological and stochastic assumptions, for each $i \in \mathcal{C}_{d_0}$*

$$\max_{r\in\pi(d_0)} \frac{\lambda m \sum_{i\in\mathcal{H}^r} c_i}{1-\lambda\gamma_r} \le \liminf_{x\to\infty} \frac{\mathbb{P}(S^i > x)}{\overline{F^s}(x)} \tag{39}$$

$$\le \limsup_{x\to\infty} \frac{\mathbb{P}(S^i > x)}{\overline{F^s}(x)} \le \frac{\lambda m (\sum_{r\in\pi(d_0)} \sum_{i\in\mathcal{H}^r} c_i)}{1-\lambda\Gamma_{d_0}},$$

where $\Gamma_{d_0} = \max_{p\in\pi(d_0)} \gamma_p$.

Proof Let $d_0 \in \{1,\ldots,d\}$ fulfil the conditions of the theorem. We first derive the lower bound. For $p \ge q$, let

$$g^{d_0}_{p,q} = \max_{i\in\mathcal{H}^{d_0}} \{(A_{p-1} \otimes A_{p-2} \otimes \cdots \otimes A_{q-1})_{i,i}\}\,.$$

In view of (36), we have $\lim_{k\to\infty}(\mathbb{E} g^{d_0}_{0,-(k-1)}/k) = \gamma_{d_0}$. For the same type of reasons as in the irreducible case, for all $i \in \mathcal{C}_{d_0}$, and for all $r \in \pi(d_0)$,

$$D^i_n \ \ \ge \ \ \sum_{l=1}^{[\frac{n}{k}]-1} g^r_{-lk,-((l+1)k-1)}\,,$$

and by the same type of proof as in §4.2, we deduce from this that

$$\liminf_{x\to\infty} \frac{\mathbb{P}(S^i > x)}{\overline{F^s}(x)} \ge \max_{r\in\pi(d_0)} \frac{\lambda m \sum_{i\in\mathcal{H}^r} c_i}{1-\lambda\gamma_r}\,. \tag{40}$$

To derive the upper bound, let for $p \ge q$

$$G_{p,q} = \max_{i\in\mathcal{C}_l,\ j\in\mathcal{C}_k,\ k,l\in\pi(d_0)} \{(A_{p-1} \otimes \cdots \otimes A_{q-1})_{i,j}\}\,.$$

In view of (36), $\lim_{k\to\infty}(\mathbb{E}G_{0,-(k-1)}/k) = \Gamma_{d_0}$. Using $\mathbf{T}$, we still obtain as in the irreducible case that

$$D_n^i \leq \sum_{l=0}^{h-1} G_{-lk,-((l+1)k-1)} \,.$$

So by the same reasoning as in §4.2,

$$\limsup_{x\to\infty} \frac{\mathbb{P}(S^i > x)}{\overline{F^s}(x)} \leq \frac{\lambda m(\sum_{r\in\pi(d_0)}\sum_{i\in\mathcal{H}^r} c_i)}{1-\lambda\Gamma_{d_0}} \,. \qquad \square$$

Corollary 4.1 *Let $\mathcal{H}$ have a single element i_0 such that $c_{i_0} > 0$ and $c_i = 0$ for all $i \in \mathcal{H}\setminus\{i_0\}$, where $i_0 \in \mathcal{C}_{d_0}$ for some $d_0 \in \{1,\dots,d\}$. If in addition $\gamma_{d_0} = \Gamma_{d_0}$, which is equivalent to the condition that $\gamma_{d_0} \geq \gamma_l$ for all $l \in \pi(d_0)$, then for all $i \in \mathcal{C}_{d_0}$, $S^i \in S$ and*

$$\mathbb{P}(S^i > x) \sim \frac{\lambda m_{i_0}}{1-\lambda\gamma_{d_0}} \overline{F^s_{i_0}}(x) \,. \tag{41}$$

Example We revisit the example of tandem queues considered in §3. As before, let R_n^i be the time which elapses between the n-th external arrival time T_n and the n-th departure from station i. This model falls in the above class of (non-irreducible) networks with more than one equivalence class of coordinates. For instance, in the case with two stations, it is easy to check that whenever the initial condition is that of an empty system, then the random vectors $R_n = (R_n^1, R_n^2)^\top$ satisfy a recurrence equation of the form $R_{n+1} = A_n \otimes C(\tau_n) \otimes R_n \oplus B_{n+1}$ with

$$A_n = \begin{pmatrix} \sigma_{n+1}^1 & \varepsilon \\ \sigma_{n+1}^1 + \sigma_{n+1}^2 & \sigma_{n+1}^2 \end{pmatrix}$$

and with the vector B_n given by $B_n^\top = (\sigma_n^1,\ \sigma_n^1 + \sigma_n^2)$. There are two equivalence classes of coordinates, which are both reduced to one element: $\mathcal{C}_1 = \{1\}$ and $\mathcal{C}_2 = \{2\}$. In addition, $\gamma_1 = m_1$ and $\gamma_2 = m_2$ and $\gamma = \max\{m_1, m_2\}$. If the random sequences $\{\sigma_n^1\}$ and $\{\sigma_n^2\}$ are independent sequences of i.i.d. random variables whose distributions F_1 and F_2 satisfy (5) and (6), both the topological and the stochastic assumptions are satisfied. The only non-trivial property to check is $\mathbf{T}$, which follows from the fact that here too for the same reason as in the blocking network example considered in §4.2,

$$(A_{-1} \otimes \cdots \otimes A_{-n} \otimes B_{-n})^i \leq (A_{-1} \otimes \cdots \otimes A_{-n} \otimes A_{-(n+1)})_{i,1} \,.$$

4.4 Stochastic Event Graphs

Stochastic event graphs are a class of discrete event systems whose evolution is known to be captured by a $(\max, +)$–linear recurrence equation (see [4] – we use the terminology of this reference throughout this subsection). Networks of tandem queues without and with blocking are two simple instances of stochastic event graphs. The aim of this subsection is to analyse the conditions under which the technical assumptions of the two preceding subsections are satisfied. As a result, when these conditions are satisfied, the asymptotic tail distribution of response times in stochastic event graphs can be characterized using Theorems 4.1 and 4.2.

Each of the K transitions of the stochastic event graph that we consider is assumed to be FIFO and with an i.i.d. firing time sequence $\{\sigma_n^i\}$ for transition i, where the random variable σ_n^i is non-negative and integrable. The event graph is assumed to have a single input transition, which is connected to the internal transitions via places with 0 initial marking.

If we take as state variables the vectors $X_n = (X_n^i)$, where i ranges over the set $\{1, \ldots, K\}$, and where X_n^i denotes the epoch when transition i *completes* its n-th firing, then it is easy to check that under the above assumptions, the basic recurrence equation for $\{X_n\}$ is of the form

$$\begin{aligned} X_{n+1} &= A_{n+1,0} \otimes X_{n+1} \oplus A_{n+1,1} \otimes X_n \oplus \cdots \oplus A_{n+1,M} \otimes X_{n+1-M} \\ &\quad \oplus B_{n+1} \otimes T_{n+1}, \end{aligned} \tag{42}$$

where

- M is the maximal initial marking;
- the matrices $A_{n+1,j}$, $j = 0, \ldots, M$ and B_{n+1} are measurable with respect to $\sigma(\{\sigma_{n+1}^i, i = 1, \ldots, K\})$;

For more on this, see [16], p. 43.

Example (continued) In the K-station tandem queueing model considered in §3, if all queues are initially empty, we have $M = 1$. The $K \times K$-matrices $A_{n,0}$ and $A_{n,1}$ are given by

$$(A_{n,0})_{i,j} = \begin{cases} \sigma_n^i & \text{if } i = j+1, \\ \varepsilon & \text{else,} \end{cases} \qquad (A_{n,1})_{i,j} = \begin{cases} \sigma_n^i & \text{if } i = j, \\ \varepsilon & \text{else,} \end{cases}$$

and $B_n^\top = (\sigma_n^1, \varepsilon, \ldots, \varepsilon)$.

Example (continued) The blocking network of §4.2 is a stochastic event graph with $M = 2$, $K = 5$, and with the following matrices (see [4] for the method to derive the matrices):

$$A_{n,0} = \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \sigma_n^1 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \sigma_n^2 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon \end{pmatrix}, \qquad A_{n,1} = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \sigma_n^1 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \sigma_n^2 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \end{pmatrix},$$

and

$$A_{n,2} = \begin{pmatrix} \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \end{pmatrix}, \qquad B_n = \begin{pmatrix} 0 \\ \varepsilon \\ \varepsilon \\ \varepsilon \\ \varepsilon \end{pmatrix}.$$

Following the methodology described in [4], Chapter 2, it is easy to prove that if the event graph is *live* (liveness is equivalent to the fact that $A_{n,0}$ is strictly lower triangular, up to a permutation of the coordinates), then (42) can be written as an explicit recurrence equation for $\{X_n\}$ as follows:

$$\begin{aligned} X_{n+1} &= \widetilde{A}_{n+1,n} \otimes X_n \oplus \widetilde{A}_{n+1,n-1} \otimes X_{n-1} \oplus \cdots \oplus \widetilde{A}_{n+1,n+1-M} \otimes X_{n+1-M} \\ &\quad \oplus A^*_{n+1,0} \otimes B_{n+1} \otimes T_{n+1} \;, \end{aligned}$$

where $\widetilde{A}_{n,n-i}$, $i = 1, \ldots, M$, is the square matrix of dimension K defined by the relations:

$$\widetilde{A}_{n,n-i} = A^*_{n,0} \otimes A_{n,i} \,, \qquad A^*_{n,0} = \bigoplus_{m \geq 0} A^m_{n,0} \,.$$

This M-step recursion can then be reduced to a one step recursion of the form $\overline{X}_{n+1} = \overline{A}_n \otimes \overline{X}_n \oplus \overline{B}_{n+1} \otimes T_{n+1}$, where

- the vectors $\overline{X}_n$, of dimension MK, have the block structure:

$$\overline{X}_n^\top = (X_n^\top \; X_{n-1}^\top \; \ldots \; X_{n-M+1}^\top) \,;$$

- the square matrices $\overline{A}_n$, of dimension MK, have the block structure:

$$\overline{A}_n = \begin{pmatrix} \widetilde{A}_{n+1,n} & \widetilde{A}_{n+1,n-1} & \ldots & \ldots & \widetilde{A}_{n+1,n+1-M} \\ I & \mathcal{E} & \ldots & \mathcal{E} & \mathcal{E} \\ \mathcal{E} & I & \ddots & \vdots & \mathcal{E} \\ \vdots & \ddots & I & \mathcal{E} & \mathcal{E} \\ \mathcal{E} & \ldots & \mathcal{E} & I & \mathcal{E} \end{pmatrix},$$

where

- I and $\mathcal{E}$ denote the $K \times K$ identity and null matrices of $(\mathbb{R}^{K,K}, \oplus, \otimes)$ respectively;
- the vectors $\overline{B}_n$, of dimension MK, have the block structure:

$$\overline{B}_n^\top = ((A_{n,0}^* \otimes B_n)^\top \ \varepsilon^\top \ \ldots \ \varepsilon^\top),$$

  where ε here denotes the null vector of $(\mathbb{R}^K, \oplus, \otimes)$.

It follows from this that the response time vector R_n, of dimension MK, with coordinates $R_n^i = \overline{X}_n^i - T_n$ then satisfies the equation

$$R_{n+1} = \overline{A}_n \otimes C(\tau_n) \otimes R_n \oplus \overline{B}_{n+1}.$$

The key observation is now that the matrices $\{\overline{A}_n\}$ are i.i.d. and integrable under the above assumptions.

Example (continued) Let us show how to derive the equations of the blocking network of §4.2. It is easily checked that

$$A_{n,0}^* = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \sigma_n^1 & 0 & \varepsilon & \varepsilon & \varepsilon \\ \sigma_n^1 & 0 & 0 & \varepsilon & \varepsilon \\ \sigma_n^{1,2} & \sigma_n^2 & \sigma_n^2 & 0 & \varepsilon \\ \sigma_n^{1,2} & \sigma_n^2 & \sigma_n^2 & 0 & 0 \end{pmatrix}.$$

where $\sigma_n^{1,2} = \sigma_n^1 + \sigma_n^2$. Direct computations lead to the following matrix of dimension 10 for $\overline{A}_n$:

$$\overline{A}_n = \begin{pmatrix} 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon \\ \sigma_{n+1}^1 & \sigma_{n+1}^1 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \sigma_{n+1}^1 & \varepsilon & \varepsilon \\ \sigma_{n+1}^1 & \sigma_{n+1}^1 & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \sigma_{n+1}^1 & \varepsilon & 0 \\ \sigma_{n+1}^{1,2} & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 & \sigma_{n+1}^2 & \varepsilon & \varepsilon & \varepsilon & \sigma_{n+1}^{1,2} & \varepsilon & \sigma_{n+1}^2 \\ \sigma_{n+1}^{1,2} & \sigma_{n+1}^{1,2} & \sigma_{n+1}^2 & \sigma_{n+1}^2 & 0 & \varepsilon & \varepsilon & \sigma_{n+1}^{1,2} & \varepsilon & \sigma_{n+1}^2 \\ 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & 0 & \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon \end{pmatrix}.$$

The final step consists in dropping coordinate i if column i has only ε entries (indeed, in this case coordinate i is never used in the recursion). Acting so, one can successively drop coordinates 6,7 and 9 to get the matrix A_n of dimension 7 given in §4, which is irreducible as can be checked when considering the cycle 1,2,3,4,5,7,3,6,1.

This shows that, under the above i.i.d. assumptions, when taking *response times* as state variables, then assumption **S2** is always satisfied. If in addition **T**, **S1**, **S3** and **S4**$'$ are satisfied, then one can always reduce the analysis to that considered in §4.3. This is illustrated by the two treatments that were proposed for tandem queues, in §3 and §4.3.

5 Extensions

5.1 Asymptotics for Non-Renewal Input

Note that for the results of Theorems 3.1, 3.2, 4.1 and 4.2 to hold, it is not necessary that $\{T_n\}$ be a renewal process. Indeed, the key tool to obtain the asymptotics in (9), (14), (25) and (39) is Lemma 2.15, and this is the only place where we use the renewal assumption. In [3], Lemma 2.15 has been generalized in several directions which allows us to relax the renewal assumption on $\{T_n\}$. To give an example we mention the following result derived in [3].

Let $\{\xi_n\}$ be a sequence of positive i.i.d. random variables with distribution F and let $\{\tau_n\}$ be an arbitrary stationary ergodic sequence of positive random variables with intensity $\lambda^{-1} = \mathbb{E}\tau_n < \infty$ independent of $\{\xi_n\}$. Consider the random walk $\{\sum_{k=1}^n (\xi_k - \tau_k), n \geq 0\}$ and define the random variable $Z_\varepsilon = \sup_{n\geq 1}\{n(\lambda^{-1} - \varepsilon) - \sum_{i=1}^n \tau_i\}$. Let G denote the distribution of Z_ε. Then the following theorem holds.

Theorem 5.1 *If $F^s \in \mathcal{S}$, $\rho = \mathbb{E}\xi/\mathbb{E}\tau < 1$ and $\overline{G}(x) = o(\overline{F^s}(x)), x \to \infty$, for all sufficiently small $\varepsilon > 0$, then*

$$\mathbb{P}\Big(\sup_{n\geq 1}\sum_{k=1}^{n}(\xi_k - \tau_k) > x\Big) \sim \frac{\rho}{1-\rho}\overline{F^s}(x)\ .$$

The *proof* of Theorem 5.1 can be found in [3]. Theorem 5.1 covers e.g. Poisson cluster processes, superpositions of renewal processes and φ-mixing sequences, for details see [3] and [21]. Thus, the results of §3 and §4 hold whenever the input process $\{T_n\}$ falls into one of the above classes of stationary ergodic sequences.

5.2 Queue Size Distributions

The aim of this section is to show an example of application of our results to the asymptotics of queue sizes. Let us consider the example of Figure 1, where we assume that the input process is Poisson with intensity λ. Under this assumption, the time-stationary distribution of the total number of customers Q in the system coincides with that at arrival or departure epochs (PASTA property). Since the system is FIFO, the stationary distribution of the total number of customers in the system at departure epochs also coincides with the total number of arrivals during the end to end response time of a customer in the system. By standard arguments we therefore obtain the following representation for the tail of the time-stationary queue size distribution:

$$\mathbb{P}(Q > x) = \mathbb{P}(N(R) > x), \tag{43}$$

where N is a Poisson process with intensity λ and R is a random variable which is independent of N and distributed like the stationary end to end response time in the system. We can then use the results that we obtained on the asymptotics of the tail of the distribution of stationary response times in §4.2 and recent results in [2] to compute the asymptotics of $\mathbb{P}(Q > x)$.

Here is a more concrete instance. Assume that we are in the case where $\mathcal{H} = \{4\}$, and where the tail of F_2 is Weibull with parameter $0 < \beta < 1/2$, that is $\overline{F}_2(x) = \exp(-x^\beta)$ (see the example of §4.2 for the definitions of $\mathcal{H}$ and F_2). Then it follows from (43), from Theorem 4.1 in the present paper, from Lemma 3.2 and from Proposition 3.1 in [2] that

$$\mathbb{P}(Q > x) \sim \frac{\lambda m_2}{1 - \lambda\gamma} \beta^{-1} \left(\frac{x}{\lambda}\right)^{1-\beta} \exp\left(-\left(\frac{x}{\lambda}\right)^{\beta}\right), \quad x \to \infty.$$

This is just one particular example. Several results of the same nature, for example for the time-stationary distribution of the total number of customers in tandem queues, can be derived from the results of the present paper, from equations linking waiting or response times and queue sizes (like Equation(43)) and from the results of [2].

References

[1] Anantharam, V. (1996) Networks of queues with long-range dependent traffic streams. *Stochastic Networks: Stability and Rare Events*, Lecture Notes in Statistics **117**, 237–256, Springer–Verlag, Berlin.

[2] Asmussen, S., C. Klüppelberg and K. Sigman (1999) Sampling at subexponential times, with queueing applications. *Stoch. Proc. Appl.* (to appear).

[3] Asmussen, S., H. Schmidli and V. Schmidt (1999) Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Adv. Appl. Probab.* **31** (to appear).

[4] Baccelli, F., G. Cohen, G.J. Olsder and J.P. Quadrat (1992) *Synchronization and Linearity*, J. Wiley & Sons, Chichester.

[5] Baccelli, F., Schlegel, S. and V. Schmidt (1998) Asymptotics of tandem queues with subexponential service times. *Proceedings of the Workshop on Stochastic Modelling of High-Speed Networks.* Technical University of Munich, June 1998 (to appear in CS Press, Munich).

[6] Baccelli, F. and V. Schmidt (1996) Taylor series expansions for Poisson-driven (max,+)–linear systems. *Ann. Appl. Probab.* **6**, 138–185.

[7] Bertsimas, D., I. Paschalides and J. Tsitsiklis (1998) On the large deviations behaviour of acyclic networks of G/G/1 queues. *Ann. Appl. Prob.* (to appear)

[8] Boxma, O.J. and Q. Deng (1998) Asymptotic properties in the tandem queueing system with identical service times at both queues. *Working paper.*

[9] Chistyakov, V.P. (1964) A theorem on sums of independent, positive random variables and its applications to branching processes. *Theory Probab. Appl.* **9**, 640–648.

[10] Cline, D.B.H. (1986) Convolution tails, product tails and domains of attraction. *Prob. Theory Rel. Fields* **72**, 529–557.

[11] Cohen, J.W. (1973) Some results on regular variation for distributions in queueing and fluctuation theory. *J. Appl. Probab.* **10**, 343–353.

[12] Embrechts, P. and C.M. Goldie (1980) On closure and factorization theorems for subexponential and related distributions. *J. Austral. Math. Soc. Ser. A* **29**, 243–256.

[13] Embrechts, P. and C.M. Goldie (1982) On convolution tails. *Stoch. Proc. Appl.* **13**, 263–278.

[14] Embrechts, P. and N. Veraverbeke (1982) Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* **1**, 55–72.

[15] Ganesh, A. and V. Anantharam (1996) Stationary tail probabilities in exponential server tandems with renewal arrivals. *Queueing Systems* **22**, 203–248.

[16] Gaujal, B. (1994) "Parallélisme et simulation des systèmes à événements discrets", Thèse de Doctorat, Université de Nice Sophia Antipolis.

[17] Huang, T. and K. Sigman (1998) Delay asymptotics for tandem, split & match, and other feed forward queues with heavy tailed service. *Queueing Systems* (to appear).

[18] Klüppelberg, C. (1988) Subexponential distributions and integrated tails. *J. Appl. Probab.* **25**, 132–141.

[19] Pakes, A.G. (1975) On the tails of waiting-time distributions. *J. Appl. Probab.* **12**, 555–564.

[20] Pitman, E.J.G. (1980) Subexponential distribution functions. *J. Austral. Math. Soc.* (Series A) **29**, 337–347.

[21] Rolski, T., H. Schmidli, V. Schmidt and J.L. Teugels (1999) *Stochastic Processes for Insurance and Finance.* J. Wiley & Sons, Chichester (to appear).

[22] Teugels, J.L. and N. Veraverbeke (1973) Cramér-type estimates for the probability of ruin. *CORE Discussion Paper* No. 7316

[23] N. Veraverbeke (1977) Asymptotic behavior of Wiener–Hopf factors of a random walk, *Stoch. Proc. Appl.*, **5**, 27-37.