RC23591 (W0504-116) April 25, 2005
Electrical Engineering

IBM Research Report

Tails for (max,plus) Recursions under Subexponentiality

A. B. Dieker
CWI
PO Box 217
1090 GB Amsterdam
The Netherlands
and
University of Twente
PO Box 217
7500 AE Enschede
The Netherlands

Marc Lelarge
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
and
ENS-INRIA
45 rue d'Ulm
75005 Paris
France

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

Tails for (max, plus) recursions under subexponentiality

A. B. Dieker[†,⋆] and M. Lelarge[‡,§]

[†]CWI
P.O. Box 94079, 1090 GB Amsterdam, the Netherlands

[⋆]University of Twente
Faculty of Mathematical Sciences
P.O. Box 217, 7500 AE Enschede, the Netherlands

[‡]ENS-INRIA
45 rue d'Ulm, 75005, Paris, France

[§]IBM
T. J. Watson Research Center
Yorktown Heights, NY 10598, USA

Abstract

We study the stationary solution to a (max, plus)-linear recursion. Under subexponentiality assumptions on the input to the recursion, we obtain the tail asymptotics of certain (max, plus)-linear functionals of this solution.

(Max, plus)-linear recursions arise in FIFO queueing networks; more specifically, in stochastic event graphs. In the event graph setting, two special cases of our results are of particular interest and have already been investigated in the literature. Firstly, the functional may correspond to the end-to-end response time of the event graph. Secondly, for two queues in tandem, the functional may correspond to the sojourn time in the second queue. Our results allow for more general networks; we illustrate this by studying the asymptotics of the resequencing delay due to multi-path routing.

1 Introduction

Ever since the derivation of the tail asymptotics of the waiting time in a FIFO $GI/GI/1$ queue under subexponentiality (Pakes [22] or Veraverbeke [24]), a vast body of literature has been devoted to asymptotics for isolated systems. For instance, Asmussen, Schmidli and Schmidt [4] allow for more general arrival process than renewal processes. Moreover, under certain conditions, when different inputs are multiplexed, the 'heaviest' one dominates the tail asymptotics (see [1, 15, 19]). From the point of view of asymptotics, the queue length behaves qualitatively different from the waiting time; see Asmussen, Klüppelberg, and Sigman [3] and Foss and Korshunov [16]. For surveys on the state of the art for this kind of asymptotics, see the special issues of Queueing Systems, 33:1–3 (1999), and 46:1–2 (2004).

In recent years, there has been some interest in extending the FIFO $GI/GI/1$ result to networks of queues. Among the first contributions in this area are the papers by Huang and Sigman [17] and Baccelli, Schlegel and Schmidt [13]. Huang and Sigman consider fork-join systems (of which tandem queues are special cases), for which one service time tail is heavier than all others. Baccelli *et al.* consider the stationary sojourn time up to some node in a network that can be represented by so-called event graphs. They solve the case of irreducible event graphs, and derive upper and lower bounds for the reducible case. Recently, the asymptotics of the response time (end-to-end sojourn time) were found in the general reducible case; see [10].

An interesting question is whether such asymptotics can also be established for other characteristics of the system than sojourn times up to a specific node. This question has already been answered negatively in the case of the multi-server queue. Scheller-Wolf and Sigman [23] show that the tail of the response time of the whole system is typically strictly heavier than the tail of individual waiting times. On the other hand, for queues in tandem, Baccelli and Foss show in [9] that the tail of the total sojourn time in the system and the sojourn time in the second queue can be of the same order.

In this paper, we investigate the tail asymptotics for functionals of the stationary solution of (max, plus)-linear recursions. This choice is motivated by the fact that single input, FIFO event graphs (which can represent queues in tandem, for instance) admit a representation as (max, plus)-linear systems in a random medium [7]. Therefore, our results can be used to study the tail of several characteristics of queueing networks. Our work can thus be regarded as a continuation of [10, 13]. Like in [10], we rely on results of Baccelli and Foss [9]. They show that for a wide class of networks, a large sojourn time must be caused by a single large service time in the distant past. This theorem is also the basis for the recent results obtained for generalized Jackson networks [11].

The main motivation for our work stems from a queueing problem. Some applications or transport protocols require packets to be delivered to the destination in the order of transmission at the sender. Packets that arrive at the receiving host may be mis-ordered because of multi-path routing. The transport layer at the receiver is responsible to temporarily buffer out-of-order packets and to resequence all packets: by doing so, some of them are delayed. Thanks to our results, we are able to study the asymptotics of this delay for quite general disordering networks. We should stress that our results are not restricted to a queueing context; they may also be relevant for modeling production systems, for instance.

As for the technical aspects of our work, we use a limit theorem that characterizes the most likely way for a rare event to occur. A related limit theorem for Jackson networks (see [20]) served as the basis for the analysis in [11]. One contribution of our work is to give new proofs of [10] that provide probabilistic intuition of the asymptotics in relation with the "single-big-event" of [9]. A significant part of this paper, however, deals with the situation where this theorem cannot be used, i.e., when the limit vanishes. Different techniques must then be used, relying on the extensive use of (max, plus)-algebra.

The paper is organized as follows. Section 2 gives the general (max, plus) framework and the stochastic assumptions; we also give explicit conditions for a particular system to fall within the scope of this paper. In Section 3, we give our main results: a limit theorem that was referred to earlier, and a theorem that yields the tail asymptotics of the stationary solution of the recursion in terms of an integral. We also relate our results to the literature. In Section 4, we apply our results to the resequencing problem. Sections 5–7 are devoted to the proofs of our results.

2 General framework and stochastic assumptions

In this paper we consider *open* systems that belong to the general class of monotone separable networks. The system has a single input marked point process $N = \{(T_n, \zeta_n)\}_{-\infty<n<\infty}$, where in a queueing context the sequence $\{\zeta_n\}$ describes the service times and routing decisions. We refer to [6] and [8] for a precise definition of this monotone separable framework.

We focus on a subclass of these systems that we describe in the next section.

2.1 (Max, plus)-linear systems

Notation

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:

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

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

It is the aim of this section to show how one can associate a (max, plus)-linear system to these maps. To do so, we first introduce some notation, notions, and assumptions associated to the two maps.

There is a natural way (see Section 2.3 of [7]) to associate a graph $\mathcal{G}_{\mathcal{A}} = (\mathcal{V}_{\mathcal{A}}, \mathcal{E}_{\mathcal{A}})$ to the map $\mathcal{A}$. Let $\mathcal{V}_{\mathcal{A}} = \{1, \ldots, s\}$, which we abbreviate as $[1, s]$. If $\mathcal{A}^{(j,i)} > -\infty$, then the edge (i, j) belongs to $\mathcal{E}_{\mathcal{A}}$. Two nodes of $\mathcal{V}_{\mathcal{A}}$ are said to belong to the same communication class if there is a directed path from the first to the second and another one from the second to the first. Let $\mathcal{C}_1, \ldots, \mathcal{C}_d$ be the communication classes of $\mathcal{G}_{\mathcal{A}}$ and $\lessdot$ the associated partial order, namely $\mathcal{C}_\ell \lessdot \mathcal{C}_m$ if there is a path from any vertex in $\mathcal{C}_\ell$ to any vertex in $\mathcal{C}_m$. Without loss of generality, we assume that $\mathcal{C}_\ell \lessdot \mathcal{C}_m$ implies $\ell \leq m$; this is a notationally convenient restriction on the numbering of the communication classes.

We use the following notation:

- for any coordinate $i \in \mathcal{V}_{\mathcal{A}}$, its communication class is denoted by $[i]$,
- for any coordinate i (resp. communication class $\mathcal{C}$), the subset of coordinates j such that $[j] \lessdot [i]$ (resp. $[j] \lessdot \mathcal{C}$) is denoted by $[\leq i]$ (resp. $[\leq \mathcal{C}]$;
- for any coordinate i (resp. communication class $\mathcal{C}$), the subset of coordinates j such that $[i] \lessdot [j]$ (resp. $\mathcal{C} \lessdot [j]$) is denoted by $[i \leq]$ (resp. $[\mathcal{C} \leq]$);
- for any coordinate i (resp. communication class $\mathcal{C}$) and $j \in [i \leq]$ (resp. $j \in [\mathcal{C} \leq]$), we write
$$[i \leq j] := [i \leq] \cap [\leq j] \quad (\text{resp. } [i \leq \mathcal{C}] := [i \leq] \cap [\leq \mathcal{C}]),$$
and similarly for $[\mathcal{C} \leq i]$.

It is convenient to also impose some structure on the numbering of the coordinates. Indeed, again without loss of generality, we may assume that $[i] \lessdot [j]$ implies $i \leq j$. This means that we have the following block structure for the matrix $\mathcal{A} = \mathcal{A}(\zeta)$:

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}(1,1) & | & -\infty & | & -\infty & | & -\infty \\ - & - & - & - & - & - & - \\ \mathcal{A}(2,1) & | & \mathcal{A}(2,2) & | & -\infty & | & -\infty \\ - & - & - & - & - & - & - \\ & \vdots & & \vdots & & \vdots & \\ - & - & - & - & - & - & - \\ \mathcal{A}(d,1) & | & \mathcal{A}(d,2) & | & & | & \mathcal{A}(d,d) \end{pmatrix},$$

where each $\mathcal{A}(\ell,\ell)$ is an irreducible matrix (corresponding to communication class $\mathcal{C}_\ell$).

Assumptions on $\mathcal{A}$ and $\mathcal{B}$

We now formulate the assumptions on the maps $\mathcal{A}$ and $\mathcal{B}$. Given a vector $v = (v^{(1)}, \ldots, v^{(K)})$, we call a (max, plus) expression $\mathcal{P}$ a polynomial in v of unit maximum degree if it has the form

$$\mathcal{P} = \bigoplus_j \bigotimes_{k \in \mathcal{K}_j} v^{(k)},$$

where $\mathcal{K}_j \subset [1, K]$.

Assumption (M) (structure of the maps)

(M1) *Write $I = \{i \in [1,s] : \mathcal{B}^{(i)} > -\infty\}$ and $J = [1,s] \backslash I$. Then (the following properties do not depend on the argument of the maps)*

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

(M2) *For all i, $\mathcal{A}^{(i,i)} > -\infty$. Moreover, for all $k \in [1,m]$, there exists a unique j such that $\mathcal{A}^{(j,j)}(\zeta) = \zeta^{(k)}$; this defines a map $c(\cdot)$ from $[1,m]$ to $\{\mathcal{C}_1, \ldots \mathcal{C}_d\}$ by $c(k) = [j]$;*

(M3) *For any $\ell_1 \leq \ell_2$, each of the coefficients of $A_n(\ell_2, \ell_1)$ (that are not 0 or $-\infty$) are a polynomial (in $\mathbb{R}_{\max}$) in $\left\{\zeta_n^{(k)} : \mathcal{C}_{\ell_1} \lessdot c(k) \lessdot \mathcal{C}_{\ell_2}\right\}$ of unit maximum degree.*

A discussion of this assumption is deferred to the end of this section, after the introduction of the (max, plus)-linear system associated to $\mathcal{A}$ and $\mathcal{B}$. In Section 4, we give a non-trivial example of these maps.

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 $\zeta_n = (\zeta_n^{(1)}, \ldots, \zeta_n^{(m)})$, we can define the 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:

$$\mathcal{X}_{n+1} = A_{n+1} \otimes \mathcal{X}_n \oplus B_{n+1} \otimes T_{n+1}, \tag{1}$$

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

$$Y_{[m,n]} := \bigoplus_{m \le k \le n} D_{[k+1,n]} \otimes B_k \otimes T_k = \max_{m \le k \le n} \left(D_{[k+1,n]} \otimes B_k + T_k \right), \qquad (2)$$

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 (2), the sequence $\{Y_{[-n,0]}\}$ is non-decreasing in n, so that we can define the stationary solution of (1),

$$Y_{(-\infty,0]} := \lim_{n \to \infty} Y_{[-n,0]} \le \infty.$$

We define the *stationary maximal dater* by

$$0 \le Z := Z_{(-\infty,0]} = \bigoplus_{1 \le i \le s} Y^{(i)}_{(-\infty,0]} - T_0 \le \infty, \qquad (3)$$

In what follows, we assume that $T_0 = 0$ and we give a condition for this limit to be almost surely finite in Section 2.2.

We now comment on Assumption **(M)**. First, we stress that any FIFO event graph with a single input fits into our framework; see [7] and [21] for details on this class. Some examples are given in Section 3.2.

Assumption **(M1)** implies that $A_n \otimes \mathbf{0} = B_n \oplus \mathbf{0}$, where $\mathbf{0}$ denotes the vector with all its entries equal to 0. By Lemma 7 of [21], this shows that the mapping $N = \{(T_n, \zeta_n)\} \mapsto X_{[m,n]}(N) = \bigoplus_{1 \le i \le s} Y^{(i)}_{[m,n]}$ defines a monotone separable network. This property is crucial for our analysis, as it enables us to use the results of Baccelli and Foss [9].

Assumption **(M2)** is best understood in the context of event graphs. Each $\zeta^{(k)}$, $k \in [1, m]$ then typically corresponds to one timed transisition, that is recycled in order to make the event graph FIFO. In other words, the FIFO property translates into **(M2)** for the corresponding matrix.

Assumption **(M3)** intuitively entails that each $\zeta^{(k)}$, $k \in [1, m]$ is associated with only one communication class $c(k)$ in $\mathcal{G}_{\mathcal{A}}$, since the coefficients of $A_n(\ell, \ell)$ (that are not 0 or $-\infty$) are polynomials of unit maximum degree in the variables $\left\{ \zeta_n^{(k)} : c(k) = \mathcal{C}_\ell \right\}$ only. Note that for the non-diagonal matrices $A_n(\ell_2, \ell_1)$, $\ell_1 \ne \ell_1$, the coefficients are upper bounded by $\sum_{k \in \mathcal{K}} \zeta_n^{(k)}$, where $\mathcal{K} = \{k : \mathcal{C}_{\ell_1} \lessdot c(k) \lessdot \mathcal{C}_{\ell_2}\}$. We have then for all $u \le s$, by **(M1)**,

$$\max_{k:\, c(k) \lessdot [u]} \zeta_n^{(k)} \le \max_{i,j \le u} A_n^{(i,j)} = \max_{i \le u} B_n^{(i)} \le \sum_{k:\, c(k) \lessdot [u]} \zeta_n^{(k)},$$

which is used repeatedly in what follows. In [13], this property allows the authors to verify their assumption **S4** and in [10], to show that Assumption **(AA')** (that extends slightly Assumption **(AA)** of [9]) is verified (see Lemma 3 of [10]).

2.2 Stochastic assumptions

We formulate the stochastic assumptions underlying our results.

Assumption (IA) (independence assumption)
We suppose that the sequences $\{\zeta_n\}$ *and* $\{\tau_n := T_{n+1} - T_n\}$ *are mutually independent and each of them consists of i.i.d. random variables with finite means.*

Supposing that $\mathbb{E}\tau_0 =: a < \infty$ and $\mathbb{E}\zeta_0^{(i)} =: b^{(i)} < \infty$ for $i \in \mathcal{T}$, 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 of irreducible matrices $\{A_n(k,k)\}$ (corresponding to class k), see Theorem 7.27 in [7].

Assumption (S) (stability)
We assume that $\gamma := \max_k \gamma_k < a$.

Then in view of Theorem 7.36 of [7], we have that under **(S)** the maximal dater Z defined in (3) is almost surely finite.

We use the following notation: if $j \in \mathcal{C}_i$, we write $\gamma_{[j]} := \gamma_i$. Then we define for any subset $E \subset \{1, \dots, s\}$, $\gamma_E := \max_{j \in E} \gamma_{[j]}$. The quantities $\gamma_{[\leq i]}$, $\gamma_{[i \leq]}$ and $\gamma_{[i \leq j]}$ are of special interest.

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 is written as $f(x) = o(g(x))$.

We now give the subexponential assumptions concerning the input to the system. Recall that a distribution function G on $\mathbb{R}_+$ is called subexponential if $\overline{G^{*2}}(x) \sim 2\overline{G}(x)$, where $\overline{G^{*2}}$ is the tail of the twofold convolution of G.

Throughout, we let F be a distribution function on $\mathbb{R}_+$ such that:

- F is subexponential, with finite first moment;
- The integrated distribution F^s of F with the tail

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

  is subexponential.

Note that since $\overline{F}^s$ is long-tailed, we can find a non-decreasing integer-valued function N_x tending to infinity such that, for all finite real number b,

$$\sum_{n=0}^{N_x} \overline{F}(x + nb) = o(\overline{F}^s(x)). \tag{4}$$

We now give the assumptions on the distribution of $\zeta_n = (\zeta_n^{(1)}, \dots, \zeta_n^{(m)})$. Note that these assumptions correspond to Assumptions **(SE)** and **(H)** of [9] when replacing $Y_i^{(j)}$ by $\zeta_i^{(j)}$.

Assumption (SE) (subexponentiality)
The following equivalence holds when x tends to infinity (with $d^{(j)} \geq 0$):

$$\mathbb{P}(\zeta_1^{(j)} > x) \sim d^{(j)} \overline{F}(x),$$

for all $j = 1, \dots, m$ *with* $\sum_{j=1}^m d^{(j)} > 0$.

Assumption (H) (hypothesis on tails)
We have as x tends to infinity:

$$\mathbb{P}\left(\sum_{j=1}^{m} \zeta_1^{(j)} > x\right) \sim \mathbb{P}\left(\max_{j=1}^{m} \zeta_1^{(j)} > x\right) \sim \sum_{j=1}^{m} \mathbb{P}\left(\zeta_1^{(j)} > x\right) \sim \sum_{j=1}^{m} d^{(j)}\overline{F}(x).$$

Assumption (TA) (technical assumption)
There exists some sequence $z_n \to \infty$ such that $z_n = o(n)$, so that for any $j = 1, \ldots, m$

$$\sum_{n=0}^{\infty} \mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma), \sum_{k\neq j} \zeta_0^{(k)} > z_n\right) = o(\overline{F}^s(x)).$$

Note that if we assume that the sequences $\{\zeta_n^{(i)}\}_n$ are mutually independent in i, then Assumptions **(IA)** and **(SE)** imply directly Assumptions **(H)** and **(TA)**. In this case we are in the framework of [10]. However, Section 4 shows that it can be useful to allow for a weak dependence.

3 Main results and examples

In this section, we present our main results, and show that they generalize several results in the literature.

3.1 Limit theorem and subexponential asymptotics

Our first result is a limit theorem which is essential to our approach. More precisely, we study how $Y_{[-n,0]}$ scales with n if one element of ζ_{-n} is extraordinarily large, i.e., of the same order as n.

To formalize this, we need some new notation: for $\zeta \in \mathbb{R}_+^m$, we define

$$Y_{[-n,0]}(\zeta) := \max_{0\leq p\leq n-1} \left(D_{[-p+1,0]} \otimes B_{-p} + T_{-p}\right) \vee \left(D_{[-n+1,0]} \otimes \mathcal{B}(\zeta) + T_{-n}\right). \tag{5}$$

It is the aim to describe the limit of $Y_{[-n,0]}(\zeta)$ if one component of ζ, say $\zeta^{(j)}$, is large. In general, this limit depends on $j, \zeta^{(j)}$ and n, and we denote it by $f(j, \zeta^{(j)}, n)$. Element ℓ of $f(j, \zeta^{(j)}, n)$ is denoted by $f^{(\ell)}(j, \zeta^{(j)}, n)$.

Theorem 1 *Under Assumptions* **(M)**, **(IA)** *and* **(S)**, *we have for any sequence $z_n \to \infty$ with $z_n = o(n)$ the following limit,*

$$\lim_{n\to\infty} \sup_{\substack{\zeta:\zeta^{(j)}\geq n(a-\gamma)\\ \zeta^{(k)}\leq z_n, k\neq j}} \left\| \frac{Y_{[-n,0]}(\zeta) - f(j,\zeta^{(j)}, n)}{n} \right\| = 0 \quad a.s.,$$

where

$$f^{(\ell)}(j,\zeta,n) := \begin{cases} (\zeta + n(\gamma_{[c(j)\leq\ell]} - a))^+ & \text{if } c(j) \lessdot [\ell]; \\ 0 & \text{otherwise.} \end{cases}$$

In this paper, we study the asymptotics of $P(\Phi(Y_{(-\infty,0]}) > x)$, with $\Phi : \mathbb{R}_+^s \to \mathbb{R}_+$ given by

$$\Phi(y^{(1)}, \ldots, y^{(s)}) := \max_{p=1,\ldots,L} \sum_{i=1}^{C_p} \left(y^{(\ell_i^p)} - y^{(k_i^p)}\right), \tag{6}$$

where the ℓ_i^p, k_i^p are chosen such that for all p,

$$[k_1^p] \lessdot [\ell_1^p] \lessdot \cdots \lessdot [k_{C_p}^p] \lessdot [\ell_{C_p}^p].$$

The choice $k=0$ is allowed, with the convention that $0 \lessdot [\ell]$ for all ℓ, and $y^{(0)}=0$. We set

$$W := \Phi(Y_{(-\infty,0]}); \quad W_{[-n,0]} := \Phi(Y_{[-n,0]}),$$

so that $\lim_{n\to\infty} W_{[-n,0]} = W$. Note that $W \le Z$ by construction. We write $C := \sum_p C_p$.

Associated to Φ and $f(j,\cdot,\cdot)$, we define the following domain

$$\Delta^j(x) := \left\{(\sigma,t) \in \mathbb{R}_+^2 : \Phi\left(f(j,\sigma,t)\right) > x\right\}. \tag{7}$$

Before giving a more precise description of $\Delta^j(x)$ in Lemma 1, we note that Δ^j clearly satisfies a scaling property: $\Delta^j(x) = x\Delta^j(1) := x\Delta^j$ for $x>0$. In particular, the property $\Delta^j(x) \neq \emptyset$ does not depend on x, hence we write $\Delta^j \neq \emptyset$.

Theorem 2 *Under Assumptions* **(M)**, **(IA)**, **(S)**, **(SE)**, **(H)**, *and* **(TA)**, *we have*

- *if $\Delta^j \neq \emptyset$ for every j with $d^{(j)}>0$,*

$$\mathbb{P}(W > x) = \sum_{\substack{j=1 \\ d^{(j)}>0}}^{m} \int_{(\sigma,t)\in\Delta^j(x)} \mathbb{P}(\zeta^{(j)} \in d\sigma)dt + o(\overline{F}^s(x)).$$

- *if $\Delta^j = \emptyset$ for some j with $d^{(j)}>0$,*

$$\mathbb{P}(W > x) = \sum_{\substack{j=1 \\ d^{(j)}>0, \Delta^j\neq\emptyset}}^{m} \int_{(\sigma,t)\in\Delta^j(x)} \mathbb{P}(\zeta^{(j)} \in d\sigma)dt + o(\overline{F}^s(x)).$$

  provided the γ_ℓ are all distinct.

Note that we are not always able to derive the exact asymptotics of W. We are dealing in this paper with tails of the order of $\overline{F}^s(x)$. In particular, if for fixed p each pair (ℓ_i^p, k_i^p) are coordinates that belong to the same communication class $\mathcal{C}_i$, it follows from our proof that the tail distribution of W is of the order $\overline{F}(x) = o(\overline{F}^s(x))$ (see proof of Lemma 4). This should be compared to the work of Ayhan, Palmowski and Schlegel [5] on closed networks. However, we do not address this issue in the present paper.

The following lemma shows that $\Delta^j(x)$ has a special geometry by definition of Φ in (6).

Fix j and write for any $p=1,\dots,L$ and $m=1,\dots,C_p$,

$$b_m^{j,p} := \begin{cases} \sum_{s=m}^{C_p} \left(\gamma_{[c(j)\le\ell_s^p]} - \gamma_{[c(j)\le k_s^p]}\right) & \text{if } c(j) \lessdot [\ell_m^p] \text{ and } c(j) \lessdot [k_m^p]; \\ \infty & \text{otherwise,} \end{cases}$$

and set $b_{C_p+1}^{j,p} = 0$. Note that in particular $b_1^{j,p} = \infty$ if $k_1^p = 0$.

Lemma 1 *We have for all j and x,*

$$\Delta^j(x) = \bigcup_{p=1}^{L} \bigcup_{\substack{i=1 \\ b_i^{j,p} > b_{i+1}^{j,p}, c(j)\lessdot[\ell_i^p]}}^{C_p} \left\{(\sigma,t) : t \in \left[\frac{x}{b_i^{j,p}}, \frac{x}{b_{i+1}^{j,p}}\right), \sigma > x + t\left(a - \gamma_{[c(j)\le\ell_i^p]} - b_{i+1}^{j,p}\right)\right\}, \tag{8}$$

where $x/0$ and x/∞ should be interpreted as ∞ and 0 respectively, and an empty union should be interpreted as the empty set.

3.2 Examples

In this subsection, we work out two special cases of Theorem 2 that can be found in the queueing literature. First, W corresponds to the (stationary) sojourn time up to some node, i.e., for some ℓ, $W = Y^{(\ell)}_{(-\infty,0]}$. In the second example, W corresponds to the sojourn time in the second queue in a two station tandem system.

Sojourn times in event graphs

Suppose that $W = Y^{(\ell)}_{(-\infty,0]}$ for some given ℓ, and that the maps $\mathcal{A}$ and $\mathcal{B}$ correspond to a FIFO event graph with a single input. As already pointed out, any such event graph fits into our framework. We also refer to [21] for a description of the construction of the matrices corresponding to such an event graph, and for the proof that these matrices verify Assumption **(M)**.

Special cases of the above situation are considered in [13, 10]. In [13], upper and lower bounds are derived, which are shown to be tight in the irreducible case (i.e., $d = 1$). In [10], the tail asymptotics for W have been found if ℓ is the output transition, so that W is the maximal dater; see (3).

In the case $W = Y^{(\ell)}_{(-\infty,0]}$, the computation of $\Delta^j(x)$ is straightforward:

$$\Delta^j(x) = \begin{cases} \{(\sigma, t) : \sigma > x + t(a - \gamma_{[c(j)\leq \ell]})\} & \text{if } c(j) \lessdot [\ell], \\ \emptyset & \text{otherwise.} \end{cases}$$

As a result, the integral given in Theorem 2 has the formulation given in the following corollary. It slightly generalizes the main results of [13] and [10], since we allow for a (weak) dependence between the service times.

Corollary 1 *Under Assumptions* **(M)**, **(IA)**, **(S)**, **(SE)**, **(H)**, *and* **(TA)**, *we have*

$$\mathbb{P}(Y^{(\ell)}_{(-\infty,0]} > x) \sim \left(\sum_{j:\, c(j) \lessdot [\ell]} \frac{d^{(j)}}{a - \gamma_{[c(j)\leq \ell]}} \right) \overline{F}^s(x), \tag{9}$$

where $\gamma_{[c(j)\leq \ell]} = \max_{i \in [c(j) \leq \ell]} \gamma_i$.

If the sum on the right-hand side of (9) is empty, it means in view of Assumption **(M)** that the matrices A_n, B_n restricted to coordinates in $[\leq \ell]$ are deterministic (equal to 0 or $-\infty$) and we have $Y^{(\ell)}_{(-\infty,0]} \leq 0$. A more interesting case is $\sum_{j:\, c(j) \lessdot [\ell]} d^{(j)} = 0$, in which the previous equation should be understood as $\mathbb{P}(Y^{(\ell)}_{(-\infty,0]} > x) = o(\overline{F}^s(x))$, and does not yield the exact asymptotics. However, one can then restrict the (max, plus)-linear recursion to the coordinates $[\leq \ell]$ and find an appropriate function F' that satisfies Assumptions **(SE)**, **(H)**, and **(TA)**. Then the corollary yields the exact asymptotics in this case too, which is of the order $\overline{F'}^s(x) = o(\overline{F}^s(x))$.

Queues in tandem

Consider two single server queues in tandem, with an unlimited waiting space at each of the queues. It is easy to see that the corresponding matrices are:

$$A_n = \begin{pmatrix} \zeta_n^{(1)} & -\infty \\ \zeta_n^{(1)} + \zeta_n^{(2)} & \zeta_n^{(2)} \end{pmatrix}, \quad B_n = \begin{pmatrix} \zeta_n^{(1)} \\ \zeta_n^{(1)} + \zeta_n^{(2)} \end{pmatrix},$$

and that $Y^{(i)}_{(-\infty,0]}$ corresponds to the time that customer 0 leaves queue i. In this case, we have $\gamma_i = \mathbb{E}[\zeta^{(i)}]$ and $\gamma = \max_i \gamma_i$. This example is solved by Baccelli and Foss [9], who study the case where the service times at both stations are independent. Completely different behavior is obtained in the presence of strong dependence, cf. Boxma and Deng [14].

The following corollary deals with the tail of the sojourn time at the second queue. Define $W = Y^{(2)}_{(-\infty,0]} - Y^{(1)}_{(-\infty,0]}$. A direct application of Theorem 2 gives

1. in the case $\gamma = \gamma_1 > \gamma_2$,

$$\mathbb{P}(W > x) = \frac{d^{(2)}}{a - \gamma^{(2)}} \overline{F}^s(x) + o(\overline{F}^s(x)).$$

2. in the case $\gamma = \gamma_2 > \gamma_1$,

$$\mathbb{P}(W > x) = \frac{d^{(1)}}{a - \gamma} \overline{F}^s\left(x \frac{a - \gamma_1}{\gamma - \gamma_1}\right) + \frac{d^{(2)}}{a - \gamma^{(2)}} \overline{F}^s(x) + o(\overline{F}^s(x)).$$

This result corresponds exactly to Theorems 10 and 12 of [9].

The assumption that the mean service times are distinct may seem unnatural, but it is essential in order to prevent second-order effects from appearing in the formula. The situation with equal means is studied for two stations in [9], and a central limit-type term shows up if the variances of $\zeta^{(1)}$ and $\zeta^{(2)}$ are finite; see Theorem 11 of [9]. We do not deal with this case in the present paper. Our main contribution here is to give a proof that extends Theorems 10 and 12 of [9] to more complex systems: for example, the previous result still holds if we replace each single server by a FIFO event graph as described in previous section. Such a result requires a different proof technique than in [9], where the specific structure of tandem queues is used extensively.

4 Application: resequencing delay

The aim of this section is to apply the results of this paper to a somewhat more realistic problem. For the sake of clarity, we consider a given particular network and its corresponding (max, plus) representation. While this example is quite representative for the kind of problems that our results cover, we keep it relatively simple.

To present our model, we use the formalism of Petri nets. We refer to [7] for a detailed explanation of Petri nets and their (max, plus) representations.

4.1 Motivation

In many distributed applications (e.g., remote computations, database manipulations, or data transmission over a computer network), information integrity requires that data exchanges between different nodes of a system be performed in a specific order. However, due to random delays over different paths in a system, the packets or updates may arrive at the receiver in a different order than their chronological order. In such a case, a buffer (with infinite capacity) at the receiver has to store misordered packets temporarily. We refer to this buffer as the *resequencing buffer*.

In this section, we analyze the waiting time of a packet in the resequencing buffer; this is referred to as the *resequencing delay*. Insights into this delay can then be used for dimensioning the resequencing buffer. In our model, misordering is caused by (random) multi-path routing. A similar framework has been studied by Jean-Marie and Gün [18], where the misordering network consists of K parallel $M/GI/1$ queues and the corresponding distribution of the resequencing delay is derived. More recently, Xia and Tse [26] consider a $2-M/M/1$ queueing system and derive large deviation results for the resequencing queue size. In the present subexponential framework, we derive the exact asymptotics for the resequencing delay and allow for a more general network with possible feedback mechanisms.

For more background, references, and other approaches to the resequencing problem, we refer to [2, 12, 25].

4.2 Model description

We start with a description of the model. We first recall the example that appeared in [13]; it is a tandem queue with blocking (TQB), see Figure 1.

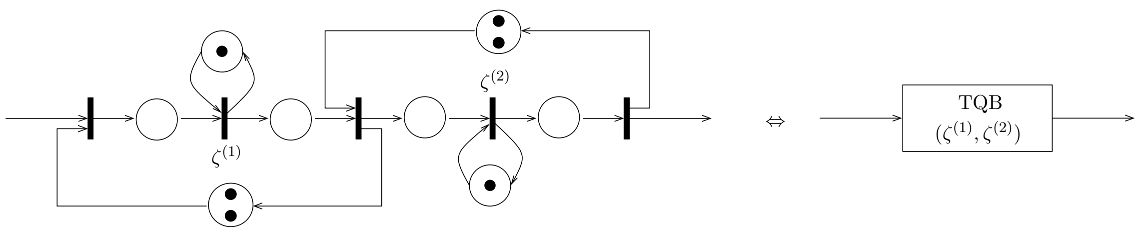

Figure 1: Tandem queues with blocking.

This system corresponds to two single server stations in tandem. Each station has both an input buffer and an output buffer. There is a blocking mechanism which prevents that more than two packets are present in station 1 at any time, with a similar constraint on the total number of packets in station 2. This implemented as follows: as long as there are two packets 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 packets in station 2, transfer from station 1 to station 2 is forbidden, and packets 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, kanban blocking in manufacturing.

We use this system as a subsystem of our network, cf. Figure 2.

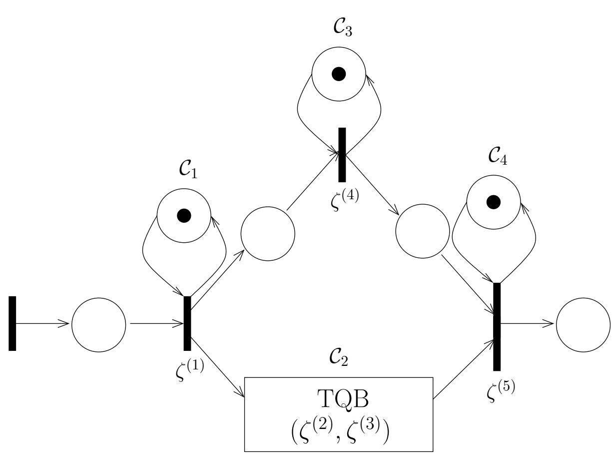

Figure 2: Network model

In this case, the mappings $\mathcal{A}$ and $\mathcal{B}$ are given by the following expressions (a proof is left to the reader):

$$
\mathcal{A}(\zeta) = \left(\begin{array}{c|ccccc|c|c}
\zeta^{(1)} & -\infty & -\infty & -\infty & -\infty & -\infty & -\infty & -\infty \\
\hline
-\infty & -\infty & -\infty & -\infty & 0 & -\infty & -\infty & -\infty \\
-\infty & -\infty & -\infty & -\infty & -\infty & 0 & -\infty & -\infty \\
\zeta^{(1,2)} & \zeta^{(2)} & -\infty & \zeta^{(2)} & -\infty & -\infty & -\infty & -\infty \\
\zeta^{(1,2)} & \zeta^{(2)} & 0 & \zeta^{(2)} & -\infty & -\infty & -\infty & -\infty \\
\zeta^{(1,2,3)} & \zeta^{(2,3)} & \zeta^{(3)} & \zeta^{(2,3)} & -\infty & \zeta^{(3)} & -\infty & -\infty \\
\hline
\zeta^{(1,4)} & -\infty & -\infty & -\infty & -\infty & -\infty & \zeta^{(4)} & -\infty \\
\hline
\zeta^{(1,23\vee 4,5)} & \zeta^{(2,3,5)} & \zeta^{(3,5)} & \zeta^{(2,3,5)} & -\infty & \zeta^{(3,5)} & \zeta^{(4,5)} & \zeta^{(5)}
\end{array}\right),
$$

$$
\mathcal{B}(\zeta) = \left(\begin{array}{c}
\zeta^{(1)} \\
\hline
-\infty \\
-\infty \\
\zeta^{(1,2)} \\
\zeta^{(1,2)} \\
\zeta^{(1,2,3)} \\
\hline
\zeta^{(1,4)} \\
\hline
\zeta^{(1,23\vee 4,5)}
\end{array}\right),
$$

where $\zeta^{(i,j)} = \zeta^{(i)} + \zeta^{(j)}$ and $\zeta^{(1,23\vee 4,5)} = \zeta^{(1)} + \max\{\zeta^{(2)} + \zeta^{(3)}, \zeta^{(4)}\} + \zeta^{(5)}$. This network belongs to the framework of [10] and hence the maps $\mathcal{A}$ and $\mathcal{B}$ automatically satisfy Assumption **(M)**. We have four communication classes:

$$\mathcal{C}_1 = [c(1)] = \{1\},\ \mathcal{C}_2 = [c(2)] = [c(3)] = \{2, \ldots, 6\},\ \mathcal{C}_3 = [c(4)] = \{7\},\ \mathcal{C}_4 = [c(5)] = \{8\}.$$

Note that the Petri net of Figure 2 does not have a unique (max, plus) representation; a different representation is found by permutating the coordinates $\{2, \ldots, 6\}$. However, if the coordinates are associated with transitions as we did in the above maps $\mathcal{A}$ and $\mathcal{B}$, $Y^{(1)}_{(-\infty,k]}$ corresponds to departure time of packet k from node 1 (corresponding to communication class $\mathcal{C}_1$), $Y^{(6)}_{(-\infty,k]}$ from the TQB system, $Y^{(7)}_{(-\infty,k]}$ from node 3 and $Y^{(8)}_{(-\infty,k]}$ from node 4.

In order to take into account routing decisions, we take a slightly different stochastic framework than [10]. In our example, we want to model a situation where a packet that leaves the first node is randomly routed up (to node 3) or down (to the TQB system). Once packet k reaches the receiver, it leaves the system if all packets j with $j < k$ have already left the system. Otherwise it stays in the resequencing buffer, where it waits for the packets with number less than k.

A different kind of routing is described by the Petri net of Figure 2. There, each time a packet (say k) finishes its service $\zeta^{(1)}_k$ in node 1, there is one packet sent up and one packet sent down simultaneously (by definition of a Petri net). The 'up'-packet ('down'-packet) is then also the k-th packet for node 3 (for the TQB system). The k-th packet joins the queue of node 4 once both packets have left node 3 and the TQB system respectively, i.e., it joins at epoch $\max\left\{Y^{(6)}_{(-\infty,k]}, Y^{(7)}_{(-\infty,k]}\right\}$. Node 4 is then a standard $\cdot/G/1/\infty$ queue.

Since the routing mechanism of our model is different from Figure 2, we need a trick to still apply our results. The idea is to introduce *clones*, i.e., packets that behave like real packets except that they never require any service time: their service time is null. Suppose that the real route of packet k is up. Then at the end of its service in the first node, a clone is sent to the TQB system. Since $\zeta_k^{(2)} = \zeta_k^{(3)} = 0$, it is clear that packets k and $k-1$ leave the TQB system at the same time: $Y^{(6)}_{(-\infty,k]} = Y^{(6)}_{(-\infty,k-1]}$. Similarly, if the real route of packet k is down, then a clone is sent up. Therefore, in both cases the "real" packet k joins the queue of node 4 once "real" packet $k-1$ has joined it (and not before). In particular packets are ordered when they leave node 4. This shows that the stationary resequencing delay is given by

$$R = \max\left\{Y^{(8)}_{(-\infty,0]} - Y^{(6)}_{(-\infty,0]}, Y^{(8)}_{(-\infty,0]} - Y^{(7)}_{(-\infty,0]}\right\}.$$

It corresponds to the time spent by "real" packet in the resequencing buffer. In particular, if we take $\zeta_k^{(5)} = 0$ for all k, this delay is purely due to multi-path routing.

The cloning procedure clearly has an impact on the distributions of the service times. Indeed, a dependence structure has been introduced (given the route of packet k, all other services are artificially set to zero). It turns out that we can still apply our results, as shown in the next subsection.

4.3 Asymptotics for the resequencing delay

Let $\sigma_n = (\sigma_n^{(1)}, \dots \sigma_n^{(5)})$ be an i.i.d. sequence of mutually independent random variables satisfying Assumption **(SE)** with a distribution F. Let $\{r_n\}_{n\in\mathbb{Z}}$ be a sequence of i.i.d. random variables, independent of everything else, with values in $\{\text{up}, \text{down}\}$. We write $\mathbb{P}(r_n = \text{up}) = 1 - \mathbb{P}(r_n = \text{down}) =: p$, and assume that $0 < p < 1$. We also define

$$\zeta_n^{(1)} := \sigma_n^{(1)},\ \zeta_n^{(2)} := \sigma_n^{(2)} 1\!\!1_{\{r_n=\text{down}\}},\ \zeta_n^{(3)} := \sigma_n^{(3)} 1\!\!1_{\{r_n=\text{down}\}},\ \zeta_n^{(4)} := \sigma_n^{(4)} 1\!\!1_{\{r_n=\text{up}\}},\ \zeta_n^{(5)} := \sigma_n^{(5)}.$$

In order to apply our main theorem, we first argue that our set of assumptions are satisfied. We have

$$\begin{aligned}
\mathbb{P}(\zeta_1^{(j)} > x) &\sim d^{(j)}\overline{F}(x), \quad \text{for } j = 1, 5,\\
\mathbb{P}(\zeta_1^{(j)} > x) &\sim (1-p)d^{(j)}\overline{F}(x), \quad \text{for } j = 2, 3,\\
\mathbb{P}(\zeta_1^{(4)} > x) &\sim pd^{(4)}\overline{F}(x).
\end{aligned}$$

Since $0 < p < 1$, Assumption **(SE)** is satisfied (although the $d^{(j)}$ are different). Now we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=1}^m \zeta_1^{(j)} > x\right) &= p\mathbb{P}\left(\sigma_1^{(1)} + \sigma_1^{(4)} + \sigma_1^{(5)} > x\right) + (1-p)\mathbb{P}\left(\sigma_1^{(1)} + \sigma_1^{(2)} + \sigma_1^{(3)} + \sigma_1^{(5)} > x\right)\\
&\sim \left[d^{(1)} + (1-p)\left(d^{(2)} + d^{(3)}\right) + pd^{(4)} + d^{(5)}\right]\overline{F}(x),
\end{aligned}$$

and similarly for the maximum instead of the sum; therefore, Assumption **(H)** is satisfied. To show that Assumption **(TA)** holds, we note that

$$\mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma), \sum_{k\neq j}\zeta_0^{(k)} > z_n\right) \le \mathbb{P}\left(\sigma_0^{(j)} > x + n(a-\gamma)\right)\mathbb{P}\left(\sum_k \sigma_0^{(k)} > z_n\right).$$

After summing over n, we see that this is majorized by (recall that N_x is defined in (4) and set $z_x := z_{N_x}$),

$$\sum_{n<N_x}\mathbb{P}\left(\sigma_0^{(j)} > x + n(a-\gamma)\right) + \mathbb{P}\left(\sum_k \sigma_0^{(k)} > z_x\right)\sum_{n\ge N_x}\mathbb{P}\left(\sigma_0^{(j)} > x + n(a-\gamma)\right).$$

Both terms are readily seen to be $o(\overline{F}^s(x))$.

If we assume that our system is stable, we are in position to apply our main Theorem 2. Note that we have

$$\gamma_1 = \mathbb{E}[\zeta_1^{(1)}],\ \gamma_3 = \mathbb{E}[\zeta_1^{(4)}],\ \gamma_4 = \mathbb{E}[\zeta_1^{(5)}],$$

but γ_2 is much more intricate to compute (see [7] for more references on Lyapunov exponents). We consider now the case where $\zeta^{(5)} = 0$ and $\gamma_1 < \gamma_2 < \gamma_3 < a$.

The computation of the domains $\Delta^j(x)$ is straightforward:

$$\begin{aligned}
\Delta^1(x) &= \left\{(\sigma, t) : \sigma > x + t(a-\gamma_3),\ t > \frac{x}{\gamma_3 - \gamma_2}\right\} \\
\Delta^j(x) &= \{(\sigma, t) : \sigma > x + t(a-\gamma_2)\} \text{ for } j = 2,3 \\
\Delta^4(x) &= \{(\sigma, t) : \sigma > x + t(a-\gamma_3)\}\,.
\end{aligned}$$

Therefore, Theorem 2 gives the following result:

$$\mathbb{P}(R > x) = \frac{d^{(1)}}{a-\gamma_3}\overline{F}^s\left(\frac{a-\gamma_2}{\gamma_3-\gamma_2}x\right) + \left(\frac{(1-p)(d^{(2)}+d^{(3)})}{a-\gamma_2} + \frac{pd^{(4)}}{a-\gamma_3}\right)\overline{F}^s(x) + o(\overline{F}^s(x)).$$

In particular, if $\max(d^{(2)}, d^{(3)}, d^{(4)}) > 0$, then we obtain the exact asymptotics. If $\max(d^{(2)}, d^{(3)}, d^{(4)}) = 0$ (in which case automatically $d^{(1)} > 0$ by **(SE)**), we still have the exact asymptotics if we assume in addition to the subexponentiality the dominated variation of F^s, namely

$$\liminf_{x\to\infty} \frac{\overline{F}^s(2x)}{\overline{F}^s(x)} > 0,$$

which is the case if F is regularly varying, for instance.

5 The limit theorem: proof of Theorem 1

Before proving Theorem 1, we first recall a first-order ergodic theorem that can be found in Section 7.2 of [7]. Under Assumption **(IA)**, we have the following limits,

$$\begin{aligned}
\forall j,\ \lim_{n\to\infty} \frac{(D_{[-n+1,0]} \otimes B_{-n})^{(j)}}{n} &= \gamma_{[\le j]} \quad \text{a.s.} \\
\lim_{n\to\infty} \frac{T_{-n}}{n} &= -a \quad \text{a.s.}
\end{aligned}$$

Since $D_{[-\lfloor nt\rfloor+1,0]} \otimes B_{-\lfloor nt\rfloor})^{(j)}/n$ and $-T_{-\lfloor nt\rfloor}/n$ are both nondecreasing as functions of t for any n, they converge locally uniformly to $\gamma_{[\le j]}t$ and at respectively; this is Dini's theorem.

Recall that we have

$$Y^{(j)}_{[-n,0]} = \max_{0\le p\le n}\left((D_{[-p+1,0]} \otimes B_{-p})^{(j)} + T_{-p}\right),$$

and thanks to Assumption **(S)**, we have $\gamma_{[\le j]} \le \gamma < a$. The following lemma follows.

Lemma 2 *Under Assumptions* **(IA)** *and* **(S)**, *we have for all* j,

$$\lim_{n\to\infty} \frac{Y^{(j)}_{[-n,0]}}{n} = 0 \quad a.s.$$

We can now prove Theorem 1.

Proof of Theorem 1. Let R_n be the region defined by (we may fix j),

$$R_n = \{\zeta : \zeta^{(k)} \le z_n, \quad k \neq j\}. \tag{10}$$

The idea is to derive bounds on $Y^{(\ell)}_{[-n,0]}(\zeta)$ that hold uniformly for ζ in R_n.

First consider the case $c(j) \not\lessdot [\ell]$. The computation of $Y^{(\ell)}_{[-n,0]}(\zeta)$ involves only the restriction of the matrices to coordinates $i \le [\ell]$ and does not involve $\zeta^{(j)}$, cf. Assumption **(M3)**. Hence the same argument that leads to Lemma 2 gives $\sup_{\zeta\in R_n} Y^{(\ell)}_{[-n,0]}(\zeta) = o(n)$ almost surely.

We consider now the case $c(j) \lessdot [\ell]$. We have the decomposition $Y^{(\ell)}_{[-n,0]}(\zeta) = Y^{(\ell)}_{[-n+1,0]} \vee H(1,n,\zeta) \vee H(2,n,\zeta)$, where

$$\begin{aligned} H(1,n,\zeta) &:= \max_{i\in[\le \ell]\cap[c(j)\le]^c} D^{(\ell,i)}_{[-n+1,0]} + \mathcal{B}^{(i)}(\zeta) + T_{-n} \\ H(2,n,\zeta) &:= \max_{i\in[c(j)\le \ell]} D^{(\ell,i)}_{[-n+1,0]} + \mathcal{B}^{(i)}(\zeta) + T_{-n}. \end{aligned}$$

For the first term, which does not depend on ζ, Lemma 2 yields $Y^{(\ell)}_{[-n+1,0]} = o(n)$. We now turn our attention to the other terms, for which we make use of the facts that

$$\begin{aligned} \forall i \in [c(j) \le \ell],\ \lim_{n\to\infty} \frac{D^{(\ell,i)}_{[-n+1,0]}}{n} &= \gamma_{[c(j)\le \ell]} \le \gamma_{[\le \ell]} < a, \\ \forall i \in [\le \ell],\ \lim_{n\to\infty} \frac{D^{(\ell,i)}_{[-n+1,0]}}{n} &\le \gamma_{[\le \ell]} < a. \end{aligned}$$

Thanks to **(M1)**, we have for any $i \in [c(j)\le]^c$ and $\zeta \in R_n$,

$$\mathcal{B}^{(i)}(\zeta) \le \max_j \mathcal{A}^{(i,j)}(\zeta) \le (m-1) z_n,$$

where the last statement follows from **(M3)**. Hence we have

$$\lim_{n\to\infty} \frac{\sup_{\zeta\in R_n} H(1,n,\zeta)}{n} \le \gamma_{[\le \ell]} - a < 0.$$

Now we consider $H(2,n,\zeta)$. Assumption **(M)** implies that there exists some $\kappa \in c(j)$ such that

$$\zeta^{(j)} = \mathcal{A}^{(\kappa,\kappa)}(\zeta) \le \mathcal{B}^{(\kappa)}(\zeta) = \max_j \mathcal{A}^{(\kappa,j)}(\zeta) \le \zeta^{(j)} + (m-1) z_n. \tag{11}$$

The first inequality leads to

$$H(2,n,\zeta) - \left[\zeta^{(j)} + n(\gamma_{[c(j)\le \ell]} - a)\right] \ \ge \ D^{(\ell,\kappa)}_{[-n+1,0]} - n\gamma_{[c(j)\le \ell]} + T_{-n} + na,$$

and since the second inequality in (11) is valid for any κ, we have

$$\begin{aligned} & H(2,n,\zeta) - \left[\zeta^{(j)} + n(\gamma_{[c(j)\le \ell]} - a)\right] \\ \le\ & \max_{i\in[c(j)\le \ell]} D^{(\ell,i)}_{[-n+1,0]} - n\gamma_{[c(j)\le \ell]} + \max_i \mathcal{B}^{(i)}(\zeta) - \zeta^{(j)} + T_{-n} + na \\ \le\ & \max_{i\in[c(j)\le \ell]} D^{(\ell,i)}_{[-n+1,0]} - n\gamma_{[c(j)\le \ell]} + (m-1) z_n + T_{-n} + na. \end{aligned}$$

In conclusion, we showed that there exists a sequence $\{\eta_n\}$ tending to 0 such that,

$$\begin{aligned}\sup_{\zeta\in R_n}\left|Y^{(\ell)}_{[-n+1,0]}\vee H(1,n,\zeta)\right| &\leq n\eta_n\\ \sup_{\zeta\in R_n}\left|H(2,n,\zeta)-\left[\zeta^{(j)}+n(\gamma_{[c(j)\leq\ell]}-a)\right]\right| &\leq n\eta_n.\end{aligned}$$

It is easy to see that this implies the theorem. □

6 Subexponential asymptotics: proof of Theorem 2

6.1 Preliminaries and idea of the proof

Before giving the intuition behind our proof, we need an auxiliary result for which the foundations were laid in Section 5.

Assumption **(TA)** gives a sequence $\{z_n\}$ tending to ∞ such that $z_n = o(n)$. Thanks to Theorem 1, there exists a sequence ϵ_n tending to 0 such that for $j = 1,\ldots,m$, the probability of the event

$$K^j_n := \left\{\sup_{\substack{\zeta:\zeta^{(j)}\geq n(a-\gamma)\\ \zeta^{(k)}\leq z_n, k\neq j}}\left\|\frac{Y_{[-n,0]}(\zeta)-f(j,\zeta^{(j)},n)}{n}\right\|<\epsilon_n\right\}$$

tends to one as $n\to\infty$. For notational convenience, we also define $\bar{K}^j_n := K^j_n\cap\left\{\zeta^{(k)}_{-n}\leq z_n, k\neq j\right\}$.

The following proposition entails that a large value of W is caused by a large service requirement somewhere in the distant past.

Write

$$G^j(x) := \sum_{n\geq N_x}\mathbb{P}\left(W_{[-n,0]}>x,\zeta^{(j)}_{-n}>x+n(a-\gamma),\bar{K}^j_n\right),$$

and $G(x) := \sum_{j=1,\ldots,m:\, d^{(j)}>0} G^j(x)$.

Proposition 1 *Suppose that Assumptions* **(M)**, **(IA)**, **(S)**, **(SE)**, **(H)**, *and* **(TA)** *hold.*
For any $y_x\to\infty$ such that $y_x = o(x)$, we have for $x\to\infty$,

$$G(x+y_x)+o(\overline{F}^s(x))\leq\mathbb{P}(W>x)\leq G(x-y_x)+o(\overline{F}^s(x)).$$

In particular, if G is long-tailed, we have $\mathbb{P}(W>x) = G(x)+o(\overline{F}^s(x))$.

We now have all the basic elements for the proof of Theorem 2. Indeed, on the event $\bar{K}^j_n\cap\left\{\zeta^{(j)}_{-n}>x+n(a-\gamma)\right\}$, we can replace $Y_{[-n,0]}$ by its approximation from our limit theorem in the expression of $W_{[-n,0]}$, and we get

$$\mathbb{P}\left(W_{[-n,0]}>x,\zeta^{(j)}_{-n}>x+n(a-\gamma),\bar{K}^j_n\right)\approx\mathbb{P}\left((\zeta^{(j)}_{-n},n)\in\Delta^j(x)\right),$$

where $\Delta^j(x)$ is defined in (7). Since this approximation cannot be used when $\Delta^j(x)=\emptyset$, this case requires a separate analysis.

The rest of the proof is divided into two parts. First, in Section 6.2, we use the above approximation to show that if $\Delta^j\neq\emptyset$, then $G^j(x) = H^j(x)+o(\overline{F}^s(x))$, where

$$H^j(x) = \int_{(\sigma,t)\in\Delta^j(x)}\mathbb{P}(\zeta^{(j)}\in d\sigma)dt.$$

Then, in Section 6.3, we prove that $G^j(x) = o(\overline{F}^s(x))$ in the complementary case under the additional assumption that the Lyapunov exponents γ_ℓ are all distinct. This proves Theorem 2.

Proof of Proposition 1. The proof relies on Theorem 8 of [9]. Assumptions **(IA)**, **(H)** and **(SE)** are the same as in [9] when replacing $Y_i^{(j)}$ by $\zeta_i^{(j)}$. It has been proved in [10] that Assumption **(AA)** of [9] can be relaxed and that Theorem 8 is still valid under the so-called Assumption **(AA')**, which is satisfied in our framework, see Lemma 3 of [10]. Now define the events

$$K^j_{n,x} := K^j_n \cap \{Z_{(-\infty,-n-1]} \le y_x/C\},$$

which are independent of ζ_{-n}. Moreover, the probability of this event tends to one (uniformly in $n \ge N_x$) since

$$\mathbb{P}(K^j_{n,x}) \ge \mathbb{P}(K^j_n) + \mathbb{P}\left(Z_{(-\infty,-n-1]} \le y_x/C\right) - 1,$$

and the second term on the right hand side does not depend on n. Finally, we have $W \le Z_{(-\infty,0]} = Z$, where Z is defined in (3) and is the stationary maximal dater of some monotone separable network (see Section 2.1). Hence Theorem 8 of [9] gives

$$\mathbb{P}(W > x) = \sum_{j=1}^{m} \sum_{n \ge N_x} \mathbb{P}\left(W > x, \zeta^{(j)}_{-n} > x + n(a-\gamma), K^j_{n,x}\right) + o(\overline{F}^s(x)). \tag{12}$$

We first deal with the case $d^{(j)} = 0$. We have

$$\begin{aligned}\sum_{n \ge N_x} \mathbb{P}\left(W > x, \zeta^{(j)}_{-n} > x + n(a-\gamma), K^j_{n,x}\right) &\le \sum_{n \ge N_x} \mathbb{P}\left(\zeta^{(j)}_{-n} > x + n(a-\gamma)\right) \\ &\le \frac{1}{a-\gamma} \int_x^{\infty} \mathbb{P}\left(\zeta^{(j)}_1 > y\right) dy = o(\overline{F}^s(x)).\end{aligned}$$

Hence the sum in (12) can be restricted to the j such that $d^{(j)} > 0$.

Now note that we have the following inequalities for any i:

$$Y^{(i)}_{[-n,0]} \le Y^{(i)}_{(-\infty,0]} \le Y^{(i)}_{(-\infty,-n-1]} + Y^{(i)}_{[-n,0]},$$

from which we derive (recall that $C = \sum_p C_p$)

$$W_{[-n,0]} - CZ_{(-\infty,-n-1]} \le W \le W_{[-n,0]} + CZ_{(-\infty,-n-1]}.$$

Hence on $K^j_{n,x}$, we have $|W - W_{[-n,0]}| \le y_x$ and then

$$\begin{aligned}&\sum_{n \ge N_x} \mathbb{P}\left(W > x, \zeta^{(j)}_{-n} > x + n(a-\gamma), K^j_{n,x}\right) \\ \le\ &\sum_{n \ge N_x} \mathbb{P}\left(W_{[-n,0]} > x - y_x, \zeta^{(j)}_{-n} > x + n(a-\gamma), K^j_{n,x}\right) \\ \le\ &G^j(x - y_x) + \sum_{n \ge N_x} \mathbb{P}\left(\zeta^{(j)}_{-n} > x + n(a-\gamma), \sum_{k \ne j} \zeta^{(k)}_{-n} > z_n\right).\end{aligned}$$

The last term is $o(\overline{F}^s(x))$ by **(TA)** and we get the right inequality of the proposition.

For the lower bound, we again use (12), and observe that

$$\mathbb{P}\left(W > x, \zeta_{-n}^{(j)} > x + n(a-\gamma), K_{n,x}^{j}\right) \geq G^{j}(x+y_x)\mathbb{P}(Z_{(-\infty,-n-1]} \leq y_x/C).$$

Since $\mathbb{P}\left(Z_{(-\infty,-n-1]} \leq y_x/C\right) = \mathbb{P}\left(Z \leq y_x/C\right)$ does not depend on n and tends to one as x tends to infinity, we have

$$\begin{aligned}\mathbb{P}\left(Z_{(-\infty,-n-1]} > y_x/C\right) G^{j}(x+y_x) &\leq \mathbb{P}\left(Z > y_x/C\right) G^{j}(x) \\ &\leq \mathbb{P}\left(Z > y_x/C\right) \sum_{n \geq N_x} \mathbb{P}\left(\zeta_{-n}^{(j)} > x + n(a-\gamma)\right).\end{aligned}$$

The desired lower bound follows from the fact that this is $o(\overline{F}^{s}(x))$.

The last statement of the proposition is a direct consequence of the fact that if G is long-tailed, one can choose y_x such that $G(x \pm y_x) \sim G(x)$. □

6.2 The case $\Delta^{j} \neq \emptyset$

In this subsection, we fix some j such that $\Delta^{j} \neq \emptyset$ and $d^{(j)} > 0$, and we prove that $G^{j}(x) = H^{j}(x) + o(\overline{F}^{s}(x))$. This gives us the behavior of $P(W > x)$ with Proposition 1, since we still have freedom to choose y_x, as long as it tends to infinity and $y_x = o(x)$. Since both H^{j} and $\overline{F}^{s}$ are long-tailed (for the first, we refer to Lemma 1), we may select y_x such that $H^{j}(x \pm y_x) = H^{j}(x) + o(\overline{F}^{s}(x))$ and $\overline{F}^{s}(x \pm y_x) \sim \overline{F}^{s}(x)$. Both the upper and lower bound of Proposition 1 then reduce to $H(x) + o(\overline{F}^{s}(x))$.

We start with the upper bound: by definition of $\bar{K}_{n}^{j}$ and Δ^{j}, we have

$$G^{j}(x) \leq \sum_{n \geq N_x} \mathbb{P}\left((\sigma_{0}^{(j)}, n) \in \Delta^{j}(x - n\epsilon_n)\right).$$

To see that this does not exceed $H^{j}(x) + o(\overline{F}^{s}(x))$, the key ingredient is Lemma 1. The details are left to the reader, as one can mimic the proof in Section 3.3 of [11].

We now find a lower bound on $G^{j}(x)$ of the form $H^{j}(x) + o(\overline{F}^{s}(x))$. To this end, we use a similar reasoning as for the upper bound:

$$\begin{aligned}&\mathbb{P}\left(W_{[-n,0]} > x, \bar{K}_{n}^{j}, \zeta_{-n}^{(j)} > x + n(a-\gamma)\right) \\ \geq\ &\mathbb{P}\left((\zeta_{-n}^{(j)}, n) \in \Delta^{j}(x + n\epsilon_n), \bar{K}_{n}^{j}, \zeta_{-n}^{(j)} > x + n(a-\gamma)\right) \\ \geq\ &\mathbb{P}\left((\zeta_{-n}^{(j)}, n) \in \Delta^{j}(x + n\epsilon_n)\right) - \mathbb{P}\left(\zeta_{-n}^{(j)} > x + n(a-\gamma), (\bar{K}_{n}^{j})^{c}\right),\end{aligned}$$

as $(\zeta_{-n}^{(j)}, n) \in \Delta^{j}(x + n\epsilon_n)$ implies $\zeta_{-n}^{(j)} > x + n\epsilon_n + n(a-\gamma)$ by Lemma 1. Now we have

$$\begin{aligned}&\mathbb{P}\left(\zeta_{-n}^{(j)} > x + n(a-\gamma), (\bar{K}_{n}^{j})^{c}\right) \\ \leq\ &\mathbb{P}\left(\zeta_{-n}^{(j)} > x + n(a-\gamma), \sum_{k \neq j} \zeta_{-n}^{(k)} \geq z_n\right) + \left[1 - \mathbb{P}(K_{n}^{j})\right] \mathbb{P}\left(\zeta_{-n}^{(j)} > x + n(a-\gamma)\right).\end{aligned}$$

After summing over n, we see that the first term is $o(\overline{F}^{s}(x))$ thanks to Assumption **(TA)**, and that the second term is $o(\overline{F}^{s}(x))$ thanks to $1 - \mathbb{P}(K_{n}^{j}) \leq 1 - \mathbb{P}(K_{N_x}^{j})$. It remains to show that $\mathbb{P}\left((\zeta_{-n}^{(j)}, n) \in \Delta^{j}(x + n\epsilon_n)\right) = H^{j}(x) + o(\overline{F}^{s}(x))$, for which we again refer to [11].

6.3 The case $\Delta^j = \emptyset$

In this subsection, we fix some j such that $\Delta^j = \emptyset$ and $d^{(j)} > 0$, and we prove that $G^j(x) = o(\overline{F}^s(x))$ under the additional assumption that $\gamma_1, \ldots, \gamma_d$ are all distinct. This suffices to obtain $G^j(x - y_x) = o(\overline{F}^s(x))$ by choosing y_x appropriately. In fact, one cannot hope that always $G^j(x) = o(\overline{F}^s(x))$ if $\Delta^j = \emptyset$. For instance, Theorem 11 of [9] shows that a second-order phenomenon plays a role.

Let us now briefly outline the idea of all proofs in this subsection. On the event $\bar{K}_n^j$, we derive upper bounds $\tilde{W}_{[-n,0]}$ on $W_{[-n,0]}$, so that $\tilde{W}_{[-n,0]}$ and $\zeta_{-n}^{(j)}$ are independent. Then

$$G^j(x) \leq \sum_{n \geq N_x} \mathbb{P}\left(\tilde{W}_{[-n,0]} > x\right) \mathbb{P}\left(\zeta^{(j)} > x + n(a-\gamma)\right).$$

Then we derive a further upper bound $\tilde{W}$ on $\tilde{W}_{[-n,0]}$ that holds uniformly in n. The claim then follows after proving that $\tilde{W}$ is almost surely finite and noting that the sum is of order $\overline{F}^s(x)$.

We may restrict ourselves to the case $L = 1$, $C_1 = 1$ without loss of generality. Indeed, $G^j(x)$ is upper bounded by

$$\sum_{p=1}^{L} \sum_{i=1}^{C_p} \sum_{n \geq N_x} \mathbb{P}\left(Y_{[-n,0]}^{(\ell_i^p)} - Y_{[-n,0]}^{(k_i^p)} > x/C, \zeta_{-n}^{(j)} > x + n(a-\gamma), \bar{K}_n^j\right).$$

For every p and i, the above argument for $N = 1$ provides a finite random variable $\tilde{W}_{p,i}$ such that this does not exceed

$$\sum_{p=1}^{L} \sum_{i=1}^{C_p} \mathbb{P}\left(\tilde{W}_{p,i} > x/C\right) \sum_{n \geq N_x} \mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma)\right).$$

The third sum is of the order $\overline{F}^s(x)$, as desired.

The above argument shows that the value of C is irrelevant in the analysis; therefore, we focus on upper bounding

$$\sum_{n \geq N_x} \mathbb{P}\left(Y_{[-n,0]}^{(\ell)} - Y_{[-n,0]}^{(k)} > x, \zeta_{-n}^{(j)} > x + n(a-\gamma)\right).$$

In this case, there are three reasons for Δ^j to be empty: firstly, the fluid limit may vanish because $\ell \in [\leq c(j)] \backslash c(j)$, secondly because $[k] = [\ell]$, and, thirdly, because $c(j) \lessdot [k] \neq [\ell]$ and $\gamma_{[c(j) \leq \ell]} = \gamma_{[c(j) \leq k]}$. The following lemma deals with the first case.

Lemma 3 *Let $L = C_1 = 1$. Suppose $\ell \in [\leq c(j)] \backslash c(j)$. Then we have $G^j(x) = o(\overline{F}^s(x))$.*

Proof. The computation of $Y_{[-n,0]}^{(\ell)}$ involves only the restriction of the matrices to classes $[\leq \ell]$ and does not involve $\zeta_{-n}^{(j)}$ in view of Assumption **(M3)**. Hence we have by **(M1)**

$$Y_{[-n,0]}^{(\ell)} \leq \max_{i,k \in [\leq \ell]} A_{-n}^{(i,k)} + Y_{[-n+1,0]}^{(\ell)} \leq \sum_{k \neq j} \zeta_{-n}^{(k)} + Y_{[-n+1,0]}^{(\ell)}.$$

Note that the term $Y_{[-n+1,0]}^{(\ell)}$ is independent of ζ_{-n}. This yields

$$\begin{aligned}
&\mathbb{P}\left(Y_{[-n,0]}^{(\ell)} - Y_{[-n,0]}^{(k)} > x, \zeta_{-n}^{(j)} > x + n(a-\gamma), \bar{K}_n^j\right) \\
\leq\ &\mathbb{P}\left(Y_{[-n+1,0]}^{(\ell)} > x - mz_n\right) \mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma)\right) \\
\leq\ &\mathbb{P}\left(Y_{(-\infty,0]}^{(\ell)} > x - mz_{N_x}\right) \mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma)\right),
\end{aligned}$$

for $n \geq N_x$. Since we can assume that $N_x \leq x$, we have $z_{N_x} = o(x)$. The claim follows after summing over n. □

The following lemma deals with the second case.

Lemma 4 *Let $L = C_1 = 1$. Suppose that $[k] = [\ell]$. Then we have $G^j(x) = o(\overline{F}^s(x))$.*

Proof. Recall that s is the dimension of the matrices A_n. Since $\mathcal{A}^{(i,i)} > -\infty$ by **(M2)**, we have for all p, and all m_1, m_2 with $[m_1] \lessdot [m_2]$,

$$\left(D_{[p,p+s-1]}\right)^{(m_2,m_1)} > -\infty.$$

Observe that for $p \geq s$,

$$\left(D_{[-p+1,0]} \otimes B_{-p}\right)^{(\ell)} = \max_m D^{(\ell,m)}_{[-s+1,0]} + \left(D_{[-p+1,-s]} \otimes B_{-p}\right)^{(m)},$$

and denote the optimizing argument by $m^* \in [\leq \ell]$. Clearly,

$$\left(D_{[-p+1,0]} \otimes B_{-p}\right)^{(k)} \geq D^{(k,m^*)}_{[-s+1,0]} + \left(D_{[-p+1,-s]} \otimes B_{-p}\right)^{(m^*)},$$

and the previous observation gives $D^{(k,m^*)}_{[-s+1,0]} > -\infty$ because $[\ell] = [k]$. Hence, we have for $p \geq s$,

$$\left(D_{[-p+1,0]} \otimes B_{-p}\right)^{(\ell)} - \left(D_{[-p+1,0]} \otimes B_{-p}\right)^{(k)} \quad \leq \quad D^{(\ell,m^*)}_{[-s+1,0]} \leq \sum_{i=-s+1}^{0} \sum_j \zeta_i^{(j)},$$

where the last inequality follows from **(M3)**. The latter inequality holds trivially if $p < s$, so that

$$Y^{(\ell)}_{[-n,0]} - Y^{(k)}_{[-n,0]} \leq \sum_{i=-s+1}^{0} \sum_j \zeta_i^{(j)}. \tag{13}$$

Since this upper bound is independent of ζ_{-n} if n is large enough, we have for $x \to \infty$,

$$\begin{aligned}
&\sum_{n \geq N_x} \mathbb{P}\left(Y^{(\ell)}_{[-n,0]} - Y^{(k)}_{[-n,0]} > x, \zeta^{(j)}_{-n} > x + n(a-\gamma)\right) \\
\leq\ & \mathbb{P}\left(\sum_{i=-s+1}^{0} \sum_j \zeta_i^{(j)} > x\right) \sum_{n \geq N_x} \mathbb{P}\left(\zeta_0^{(j)} > x + n(a-\gamma)\right) \\
\sim\ & \frac{\overline{F}^s(x)}{a-\gamma} \mathbb{P}\left(\sum_{i=1-s}^{0} \sum_j \zeta_i^{(j)} > x\right) = o(\overline{F}^s(x)),
\end{aligned}$$

as claimed. □

One specific consequence of the proof is worth pointing out. Since the upper bound in (13) is independent of n, we have

$$\mathbb{P}\left(Y^{(\ell)}_{(-\infty,0]} - Y^{(k)}_{(-\infty,0]} > x\right) \leq \mathbb{P}\left(\sum_{i=-s+1}^{0} \sum_j \zeta_i^{(j)} > x\right) \sim sd\overline{F}(x).$$

As pointed out already in Section 3, this is related to the results in [5] for closed networks.

Next we deal with the case that the fluid limit is zero while the big event takes place in a class $c(j) \lessdot [k]$.

Lemma 5 *Let $L = C_1 = 1$. Suppose that the Lyapunov exponents γ_ℓ are all distinct. Assume that $c(j) \lessdot [k]$, $[k] \neq [\ell]$ and that $\gamma_{[c(j)\leq \ell]} = \gamma_{[c(j)\leq k]}$. Then we have $G^j(x) = o(\overline{F}^s(x))$.*

Proof. We start with some notation, using the graph $\mathcal{G}_\mathcal{A}$ as introduced in Section 2.1. For a path $\xi = (i_0, i_1, \ldots, i_p)$ in $\mathcal{G}_\mathcal{A}$, we set

$$D_{[a+1,a+p]}(\xi) = A_{a+p}^{(i_p,i_{p-1})} + \cdots + A_{a+2}^{(i_2,i_1)} + A_{a+1}^{(i_1,i_0)}.$$

Define also $\Xi(i, i', p)$ as the set of paths that start in i, end in i', and have length p, i.e., paths of the form $(i = i_0, i_1, \ldots, i_p = i')$. We can express $Y_{[-n,0]}^{(\ell)}$ in this notation as follows:

$$Y_{[-n,0]}^{(\ell)} = \bigoplus_{0\leq p\leq n} \bigoplus_{i} \max_{\xi\in\Xi(i,\ell,p)} D_{[-p+1,0]}(\xi) + B_{-p}^{(i)} + T_{-p}.$$

If a path $\xi = (i_0, \ldots, i_p)$ goes through communication class $\mathcal{C}_\ell$, i.e. if there exists $i_k \in \mathcal{C}_\ell$, we write $\ell \in [\xi]$.

It is useful to introduce

$$\hat{Y}_{[-n,0]}^{(\ell)} = \bigoplus_{0\leq p\leq n} \bigoplus_{\substack{1\leq i\leq s \\ i\notin[c(j)\leq] \text{ if } p=n}} \max_{\xi\in\Xi(i,\ell,p)} D_{[-p+1,0]}(\xi) + B_{-p}^{(i)} + T_{-p}.$$

Note that the only difference with $Y_{[-n,0]}^{(\ell)}$ is the additional requirement that $i \notin [c(j) \leq]$ if $p = n$ in the maximization. We also define

$$\overline{Y}_{[-n,0]}^{(\ell)} := \bigoplus_{i\in[c(j)\leq \ell]} \max_{\xi\in\Xi(i,\ell,n)} D_{[-n+1,0]}(\xi) + B_{-n}^{(i)} + T_{-n},$$

so that $Y_{[-n,0]}^{(\ell)} = \hat{Y}_{[-n,0]}^{(\ell)} \vee \overline{Y}_{[-n,0]}^{(\ell)}$.

The assumptions imply that there exists a unique "bottleneck" communication class $\mathcal{C}_b$ with the property that

$$c(j) \lessdot \mathcal{C}_b \lessdot [k] \lessdot [\ell], \quad \gamma_b = \gamma_{[c(j)\leq \ell]} = \gamma_{[c(j)\leq k]}, \quad \text{and} \quad b \neq [\ell].$$

We write $Y_{[-n,0]}^{(\ell);b\in[\xi]}$ and $Y_{[-n,0]}^{(\ell);b\notin[\xi]}$ if the maximization over ξ is only done over $\xi \in \Xi(i, \ell, p)$ with $b \in [\xi]$ and $b \notin [\xi]$ respectively. The quantities with hats and bars are defined similarly.

The following key representation of $Y_{[-n,0]}^{(\ell)}$ is immediate:

$$Y_{[-n,0]}^{(\ell)} = \max\left(Y_{[-n,0]}^{(\ell);b\in[\xi]}, \hat{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}, \overline{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}\right),$$

with the convention that the maximum over the empty set is $-\infty$ (in the definitions of $\overline{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}$ and $\hat{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}$).

It suffices to prove

$$\sum_{n\geq N_x} \mathbb{P}\left(Y_{[-n,0]}^{(\ell);b\in[\xi]} - Y_{[-n,0]}^{(k)} > x; \zeta_{-n}^{(j)} > x + n(a-\gamma), \bar{K}_n^j\right) = o(\overline{F}^s(x)), \tag{14}$$

and similarly for $Y_{[-n,0]}^{(\ell);b\in[\xi]}$ replaced by $\hat{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}$ and $\overline{Y}_{[-n,0]}^{(\ell);b\notin[\xi]}$. This is proven in three steps, one for each term.

Step 1: $Y_{[-n,0]}^{(\ell);b\in[\xi]}$

We start with the proof of (14). Note that for $0 \le p \le n$,

$$
\begin{aligned}
& \bigoplus_i \max_{\substack{\xi\in\Xi(i,\ell,p)\\ b\in[\xi]}} D_{[-p+1,0]}(\xi) + B^{(i)}_{-p} \\
= \; & \bigoplus_{m\in\mathcal{C}_b} \bigoplus_{0\le q\le p} (D_{[-p+1,-q]} \otimes B_{-p})^{(m)} + D^{(\ell,m)}_{[-q+1,0]} \\
\le \; & \bigoplus_{m\in\mathcal{C}_b} \bigoplus_{0\le q\le p} (D_{[-p+1,0]} \otimes B_{-p})^{(m)} + D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} \\
\le \; & \bigoplus_{m\in\mathcal{C}_b} (D_{[-p+1,0]} \otimes B_{-p})^{(m)} + \bigoplus_{m\in\mathcal{C}_b} \bigoplus_{0\le q\le n} \left(D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]}\right).
\end{aligned}
$$

This implies the inequality

$$
Y^{(\ell);b\in[\xi]}_{[-n,0]} - Y^{(k)}_{[-n,0]} \le \bigoplus_{m\in\mathcal{C}_b} Y^{(m)}_{[-n,0]} - Y^{(k)}_{[-n,0]} + \bigoplus_{m\in\mathcal{C}_b} \bigoplus_{0\le q\le n} \left(D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]}\right), \tag{15}
$$

so that

$$
\begin{aligned}
& \mathbb{P}\left(\bigoplus_{m\in\mathcal{C}_b} Y^{(m)}_{[-n,0]} - Y^{(k)}_{[-n,0]} + \bigoplus_{m\in\mathcal{C}_b} \bigoplus_{0\le q\le n} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} > x, \zeta^{(j)}_{-n} > x + n(a-\gamma)\right) \\
\le \; & \sum_{m\in\mathcal{C}_b} \mathbb{P}\left(Y^{(m)}_{[-n,0]} - Y^{(k)}_{[-n,0]} > x/2, \zeta^{(j)}_{-n} > x + n(a-\gamma)\right) \\
+ \; & \sum_{m\in\mathcal{C}_b} \mathbb{P}\left(\bigoplus_{0\le q\le n} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} > x/2, \zeta^{(j)}_{-n} > x + n(a-\gamma)\right).
\end{aligned}
$$

Therefore, we may fix $m \in \mathcal{C}_b$ for the remainder of this step, and show that the two probabilities are $o(\overline{F}^s(x))$ after summing over n.

Since $\mathcal{C}_b \lessdot [k]$, there exists $u \in [k], v \in \mathcal{C}_b$ such that $\mathcal{A}^{(u,v)} > -\infty$. Following the same argument as in Lemma 4, we have (for $n \ge 2s$)

$$
Y^{(m)}_{[-n,0]} - \sum_{i=-2s}^{-s-1} \sum_j \zeta^{(j)}_i - A^{(u,v)}_{-s} - \sum_{i=-s+1}^{0} \sum_j \zeta^{(j)}_i \le Y^{(k)}_{[-n,0]}.
$$

Hence we have $Y^{(m)}_{[-n,0]} \le Y^{(k)}_{[-n,0]} + \sum_{i=-2s}^{0} \sum_j \zeta^{(j)}_i$. It implies that the first term above is $o(\overline{F}^s(x))$.

We now show that the second term is $o(\overline{F}^s(x))$. Since

$$
\bigoplus_{0\le q\le n} \left(D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]}\right)
$$

is independent of ζ_{-n} and

$$
\mathbb{P}\left(\bigoplus_{0\le q\le n} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} > x\right) \le \mathbb{P}\left(\bigoplus_{q\ge 0} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} > x\right),
$$

it is sufficient to prove that the probability in the previous display tends to zero as $x \to \infty$.

For this, write $\Xi'(m,\ell,q)$ for the subset of $\Xi(m,\ell,q)$ with paths that do not visit any coordinate of $\mathcal{C}_b$, except for the starting point. For any $\xi = (i_0,\ldots,i_q) \in \Xi(m,\ell,q)$, we define $p(\xi)$ by the requirement that $i_\kappa \in \mathcal{C}_b$ for $\kappa \le p(\xi)$ and $i_\kappa \in [\le \ell]\backslash\mathcal{C}_b$ for $\kappa > p(\xi)$. We also write $m(\xi) := i_{p(\xi)}$. Using the optimizing path $\xi^*_{q,m}$ in $\max_{\xi\in\Xi(m,\ell,q)} D_{[-q+1,0]}(\xi)$ as the argument to p and m, we get

$$\begin{aligned} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} &= D^{(m(\xi^*_{q,m}),m)}_{[-q+1,-q+p(\xi^*_{q,m})]} + D^{(\ell,m(\xi^*_{q,m}))}_{[-q+p(\xi^*_{q,m})+1,0]} - D^{(m,m)}_{[-q+1,0]} \\ &\le D^{(\ell,m(\xi^*_{q,m}))}_{[-q+p(\xi^*_{q,m})+1,0]} - D^{(m,m(\xi^*_{q,m}))}_{[-q+p(\xi^*_{q,m})+1,0]}. \end{aligned}$$

Since $p(\xi^*_{q,m}) \ge 0$ and $m(\xi^*_{q,m}) \in \mathcal{C}_b$, this implies

$$\bigoplus_{m\in\mathcal{C}_b}\bigoplus_{q\ge 0} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} \le \bigoplus_{m\in\mathcal{C}_b}\bigoplus_{m'\in\mathcal{C}_b}\bigoplus_{q\ge 0} \max_{\xi\in\Xi'(m',\ell,q)} D_{[-q+1,0]}(\xi) - D^{(m,m')}_{[-q+1,0]}.$$

Since for every $m, m' \in \mathcal{C}_b$, with probability one,

$$\lim_{q\to\infty} \frac{\max_{\xi\in\Xi'(m',\ell,q)} D_{[-q,0]}(\xi) - D^{(m,m')}_{[-q,0]}}{q} = \gamma_{[\mathcal{C}_b \le \ell]\backslash\mathcal{C}_b} - \gamma_b < 0,$$

we have $\bigoplus_{q\ge 0} D^{(\ell,m)}_{[-q+1,0]} - D^{(m,m)}_{[-q+1,0]} < \infty$ almost surely, showing (14).

Step 2: $\hat{Y}^{(\ell);b\notin[\xi]}_{[-n,0]}$

Now we turn to (14) with $Y^{(\ell);b\in[\xi]}_{[-n,0]}$ replaced by $\hat{Y}^{(\ell);b\notin[\xi]}_{[-n,0]}$. We have

$$\hat{Y}^{(\ell);b\notin[\xi]}_{[-n,0]} \le \hat{Y}^{(\ell)}_{[-n,0]} \le \max_{i,\kappa\notin[c(j)\le]} A^{(i,\kappa)}_{-n} + Y^{(\ell)}_{[-n+1,0]} \le \sum_{\kappa\ne j} \zeta^{(\kappa)}_{-n} + Y^{(\ell)}_{[-n+1,0]},$$

where we used **(M1)** and **(M3)**. The rest of the proof of the second step is similar to the case treated in Lemma 3; therefore, we omit the proof.

Step 3: $\overline{Y}^{(\ell);b\notin[\xi]}_{[-n,0]}$

It remains to show that (14) holds with $\overline{Y}^{(\ell);b\notin[\xi]}_{[-n,0]}$ instead of $Y^{(\ell);b\in[\xi]}_{[-n,0]}$. Thanks to Assumption **(M)**, there exists some (non-random) $u \in c(j)$ such that

$$Y^{(k)}_{[-n,0]} \ge D^{(k,u)}_{[-n+1,0]} + B^{(u)}_{-n} + T_{-n} \ge D^{(k,u)}_{[-n+1,0]} + \zeta^{(j)}_{-n} + T_{-n}.$$

Moreover, we have

$$\overline{Y}^{(\ell);b\notin[\xi]}_{[-n,0]} \le \sum_{k=1}^{m} \zeta^{(k)}_{-n} + \bigoplus_{i\in[c(j)\le]} \max_{\substack{\xi\in\Xi(i,\ell,n)\\ b\notin[\xi]}} D_{[-n+1,0]}(\xi) + T_{-n},$$

so that

$$\overline{Y}^{(\ell);b\notin[\xi]}_{[-n,0]} - Y^{(k)}_{[-n,0]} \le \sum_{k\ne j} \zeta^{(k)}_{-n} + \bigoplus_{i\in[c(j)\le]} \max_{\substack{\xi\in\Xi(i,\ell,n)\\ b\notin[\xi]}} D_{[-n+1,0]}(\xi) - D^{(k,u)}_{[-n+1,0]}.$$

By the same argument that was used in Lemma 3, it suffices to show that for every $i \in [c(j) \le \ell]$,

$$\lim_{x\to\infty} \mathbb{P}\left(\bigoplus_{n\ge 0} \max_{\substack{\xi\in\Xi(i,\ell,n)\\ b\notin[\xi]}} D_{[-n+1,0]}(\xi) - D^{(k,u)}_{[-n+1,0]} > x - m z_{N_x} \right) = 0.$$

This follows from the fact that for $i \in [c(j) \leq \ell]$,

$$\limsup_{n\to\infty} \frac{\max_{\xi\in\Xi(i,\ell,n)_{b\notin[\xi]}} D_{[-n+1,0]}(\xi)}{n} \leq \gamma_{[c(j)\leq\ell]\setminus\mathcal{C}_b} < \gamma_b = \lim_{n\to\infty} \frac{D^{(k,u)}_{[-n+1,0]}}{n}.$$

□

The assumption of distinct Lyapunov exponents is convenient, but can be slightly relaxed. Indeed, if $c(j) \lessdot [k]$, $[k] \neq [\ell]$ and $\gamma_{[c(j)\leq\ell]} = \gamma_{[c(j)\leq k]}$, it follows from the above proof that it suffices to have the existence of a unique bottleneck class $\mathcal{C}_b$ such that $c(j) \lessdot \mathcal{C}_b \lessdot [k]$, $\gamma_b = \gamma_{[c(j)\leq\ell]} = \gamma_{[c(j)\leq k]}$, and $\gamma_{[c(j)\leq\ell]\setminus\mathcal{C}_b} < \gamma_b$.

7 Characterization of Δ^j: proof of Lemma 1

Thanks to the specific form of Φ, it suffices to prove the claim for $L = 1$. For simplicity, we write $k_j := k_j^1$, $\ell_j := \ell_j^1$, $C := C^1$, and $b_i^j := b_i^{j,1}$.

We start by partitioning the set $\Delta^j(x)$ conveniently; for this, we exploit the fact that $\gamma_{[c(j)\leq k_i]}$ and $\gamma_{[c(j)\leq \ell_i]}$ increase in i and that $\gamma_{[c(j)\leq\ell_{i-1}]} \leq \gamma_{[c(j)\leq k_i]}$. Hence, since the definitions of Δ^j and f imply that

$$\Delta^j(x) \cap \{\sigma/t < a - \gamma_{[c(j)\leq\ell_C]}\} = \emptyset,$$

we see that $\Delta^j(x)$ can be written as

$$\begin{aligned}&\left(\Delta^j(x) \cap \{\sigma/t \geq a - \gamma_{[c(j)\leq k_1]}\}\right) \cup \bigcup_{i=2}^{C} \left(\Delta^j(x) \cap \left\{\sigma/t \in \left[a - \gamma_{[c(j)\leq k_i]}, a - \gamma_{[c(j)\leq \ell_{i-1}]}\right)\right\}\right)\\ \cup\; &\bigcup_{i=1}^{C} \left(\Delta^j(x) \cap \left\{\sigma/t \in \left[a - \gamma_{[c(j)\leq \ell_i]}, a - \gamma_{[c(j)\leq k_i]}\right)\right\}\right),\end{aligned}$$

where all sets should be interpreted as part of the (σ, t)-plane. We write the latter decomposition of $\Delta^j(x)$ as $S_I \cup \bigcup_{i=2}^C S_{II}(i) \cup \bigcup_{i=1}^C S_{III}(i)$. Moreover, we define for $\kappa = 1, \ldots, C$ the strip

$$U_\kappa := \left\{ t \in \left[\frac{x}{b^j_\kappa}, \frac{x}{b^j_{\kappa+1}}\right) \right\},$$

provided $b^j_\kappa > b^j_{\kappa+1}$; otherwise $U_\kappa = \emptyset$. We also set $S_I^\kappa := S_I \cap U_\kappa$, $S_{II}^\kappa(i) = S_{II}(i) \cap U_\kappa$, and $S_{III}^\kappa(i) = S_{III}(i) \cap U_\kappa$ for $i, \kappa = 1, \ldots, C$.

The remainder of the proof is split into three parts. First we suppose that either $[c(j) \leq k_1] \neq \emptyset$, or there exists some $i^* = 2, \ldots, C$ such that $[c(j) \leq k_{i^*-1}] = [c(j) \leq \ell_{i^*-1}] = \emptyset$ and $[c(j) \leq k_{i^*}] \neq \emptyset$. In the second part, we suppose that there is some i^* such that $[c(j) \leq k_{i^*}] = \emptyset$, while $[c(j) \leq \ell_{i^*}] \neq \emptyset$. The last part consists of the case that $\ell_C \in [\leq c(j)]\setminus c(j)$.

We start with the first part. Set $i^* = 1$ if $[c(j) \leq k_1] \neq \emptyset$. The union in (8) is taken from $i = i^*$ to C (provided $b^j_i > b^j_{i+1}$). If $i^* > 1$, we have by definition of f^j,

$$\Delta^j(x) = \left\{ (\sigma, n) \in \mathbb{R}_+ : \sum_{i=i^*}^{C} \left(f^j_{\ell_i}(\sigma, n) - f^j_{k_i}(\sigma, n)\right) > x \right\},$$

so that it can be reduced to the case with $C^{\text{new}} = C - i^* + 1$ by choosing $k_i^{\text{new}} = k_{i+i^*-1}$ and $\ell_i^{\text{new}} = \ell_{i+i^*-1}$. In the new scenario, we have $[c(j) \leq k_1^{\text{new}}] \neq \emptyset$, so that it suffices to suppose that $i^* = 1$. As an aside, observe that all b^j_i are finite in this case, so that $\Delta^j(x)$ and the σ-axis are disjoint.

Hence, suppose that $[c(j) \leq k_1] \neq \emptyset$. It is left to the reader to check that for $i = 1, \ldots, C$,

$$
\begin{aligned}
S_I^\kappa &= \left\{\sigma/t \geq a - \gamma_{[c(j)\leq k_1]}\right\} \cap U_\kappa, \\
S_{II}^\kappa(i) &= \begin{cases} \left\{\sigma/t \in \left[a - \gamma_{[c(j)\leq k_i]}, a - \gamma_{[c(j)\leq \ell_{i-1}]}\right)\right\} \cap U_\kappa & \text{if } \kappa \geq i; \\ \emptyset & \text{otherwise,} \end{cases} \\
S_{III}^\kappa(i) &= \begin{cases} \left\{\sigma/t \in \left[a - \gamma_{[c(j)\leq \ell_i]}, a - \gamma_{[c(j)\leq k_i]}\right)\right\} \cap U_\kappa & \text{if } \kappa > i; \\ \left\{\sigma/t < a - \gamma_{[c(j)\leq k_i]}\right\} \cap \left\{\sigma > x + t\left(a - \gamma_{[c(j)\leq \ell_i]} - b_{i+1}^j\right)\right\} \cap U_i & \text{if } \kappa = i; \\ \emptyset & \text{otherwise.} \end{cases}
\end{aligned}
$$

This implies that $\Delta^j(x) = \bigcup_{\kappa=1}^C \Delta_\kappa^j(x)$, with

$$
\Delta_\kappa^j(x) := S_I^\kappa \cup S_{III}^\kappa(1) \cup \bigcup_{i=2}^{\kappa} \left(S_{II}^\kappa(i) \cup S_{III}^\kappa(i)\right),
$$

where an empty union should be interpreted as an empty set. It is not difficult to see that $\Delta_\kappa^j(x)$ can be identified with $\left\{\sigma > x + t\left(a - \gamma_{[c(j)\leq \ell_\kappa]} - b_{\kappa+1}^j\right)\right\} \cap U_\kappa$, as desired.

We now turn to the second part of the proof, where we suppose that there is some i^* such that $[c(j) \leq k_{i^*}] = \emptyset$, while $[c(j) \leq \ell_{i^*}] \neq \emptyset$. Again, we may suppose without loss of generality that $i^* = 1$. Set $\underline{b}_1^j = \gamma_{[c(j)\leq \ell_1]} - \gamma_{[c(j)\leq k_1]} + b_2^j$ and note that $\underline{b}_1^j$ was called b_1^j in the first part of the proof (while now $b_1^j = \infty$ by definition). Therefore, the only difference is that the S_I are now slightly changed; indeed, denoting the old S_I by $\underline{S}_I$, we have

$$
\begin{aligned}
S_I &= \left\{\sigma/t \geq a - \gamma_{[c(j)\leq k_1]}\right\} \cap \left\{\sigma > x + t\left(a - \gamma_{[c(j)\leq \ell_1]} - b_2^j\right)\right\} \\
&= \underline{S}_I \cup \left(\left\{\sigma > x + t\left(a - \gamma_{[c(j)\leq \ell_1]} - b_2^j\right)\right\} \cap \left\{t \in \left[0, \underline{b}_1^j\right)\right\}\right).
\end{aligned}
$$

Hence, $\Delta^j(x)$ is exactly the same as in the first setting, but part of the strip $\left\{t \in \left[0, \underline{b}_1^j\right)\right\}$ is now added. This is accomplished in (8) by definition of b_1^j.

For the third and last part of the proof, it remains to investigate the case that $\ell_C \in [\leq c(j)]\backslash c(j)$. Since the fluid limit then vanishes, we have $\Delta^j(x) = \emptyset$. This is accomplished in (8) by the convention on empty intersections.

Acknowledgments

The research of ABD was supported by the Netherlands Organization for Scientific Research (NWO) under grant 631.000.002. Part of this work was done while he was visiting ENS Paris, and he thanks F. Baccelli for many helpful discussions.

Part of this work was done while ML was visiting Institut Mittag-Leffler during the program *Queueing Theory and Teletraffic Theory*.

References

[1] R. Agrawal, A. Makowski, and P. Nain, *On a reduced load equivalence for fluid queues under subexponentiality*, Queueing Syst., 33 (1999), pp. 5–41.

[2] S. Agrawal and R. Ramaswamy, *Analysis of the resequencing delay for M/M/m systems*, in Proceedings of the 1987 ACM SIGMETRICS conference on measurement and modeling of computer systems, Banff, Alberta, Canada, 1987, pp. 27–35.

[3] S. Asmussen, C. Klüppelberg, and K. Sigman, *Sampling at subexponential times, with queueing applications*, Stochastic Process. Appl., 79 (1999), pp. 265–286.

[4] S. Asmussen, H. Schmidli, and V. Schmidt, *Tail probabilities for non-standard risk and queueing processes with subexponential jumps*, Adv. in Appl. Probab., 31 (1999), pp. 422–447.

[5] H. Ayhan, Z. Palmowski, and S. Schlegel, *Cyclic queueing networks with subexponential service times*, J. Appl. Probab., 41 (2004), pp. 791–801.

[6] F. Baccelli and P. Brémaud, *Elements of queueing theory*, Springer-Verlag, Berlin, 2003.

[7] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity*, John Wiley & Sons Ltd., Chichester, 1992. Available from `http://www.maxplus.org`.

[8] F. Baccelli and S. Foss, *On the saturation rule for the stability of queues*, J. Appl. Probab., 32 (1995), pp. 494–507.

[9] ———, *Moments and tails in monotone-separable stochastic networks*, Ann. Appl. Probab., 14 (2004), pp. 612–650.

[10] F. Baccelli, S. Foss, and M. Lelarge, *Asymptotics of subexponential max plus networks: the stochastic event graph case*, Queueing Syst., 46 (2004), pp. 75–96.

[11] ———, *Tails in generalized Jackson networks with subexponential service distributions*. To appear in *J. Appl. Probab.*, 2005.

[12] F. Baccelli, E. Gelenbe, and B. Plateau, *An end-to-end approach to the resequencing problem*, Journal of the ACM, 31 (1984), pp. 474–485.

[13] F. Baccelli, S. Schlegel, and V. Schmidt, *Asymptotics of stochastic networks with subexponential service times*, Queueing Syst., 33 (1999), pp. 205–232.

[14] O. J. Boxma and Q. Deng, *Asymptotic behaviour of the tandem queueing system with identical service times at both queues*, Math. Methods Oper. Res., 52 (2000), pp. 307–323.

[15] A. B. Dieker, *Reduced-load equivalence for queues with Gaussian input*, Queueing Syst., 49 (2005), pp. 405–414.

[16] S. Foss and D. Korshunov, *Sampling at a random time with a heavy-tailed distribution*, Markov Process. Related Fields, 6 (2000), pp. 543–568.

[17] T. Huang and K. Sigman, *Steady-state asymptotics for tandem, split-match and other feedforward queues with heavy tailed service*, Queueing Syst., 33 (1999), pp. 233–259.

[18] A. Jean-Marie and L. Gün, *Parallel queues with resequencing*, Journal of the ACM, 40 (1993), pp. 1188–1208.

[19] P. Jelenković, P. Momčilović, and B. Zwart, *Reduced load equivalence under subexponentiality*, Queueing Syst., 46 (2004), pp. 97–112.

[20] M. Lelarge, *Fluid limit of generalized Jackson queueing networks with stationary and ergodic arrivals and service times*. To appear in *J. Appl. Probab.*, 2005.

[21] ———, *Rare events in networks*, PhD thesis, École polytechnique, 2005.

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

[23] A. Scheller-Wolf and K. Sigman, *Delay moments for FIFO $GI/GI/s$ queues*, Queueing Syst., 25 (1997), pp. 77–95.

[24] N. Veraverbeke, *Asymptotic behaviour of Wiener-Hopf factors of a random walk*, Stochastic Processes Appl., 5 (1977), pp. 27–37.

[25] Y. Xia and D. Tse, *Analysis on packet resequencing for reliable network protocols*, in Proc. of IEEE INFOCOM, 2003.

[26] ———, *On the large deviation of resequencing queue size: 2-$M/M/1$ case*, in Proc. of IEEE INFOCOM, 2004.