Tandem Queues with Identical Service Times in Heavy Traffic

Bryce Ashley Terwilliger
San Diego, California

Bachelors of Science in Mathematics, University of California San Diego, 2011

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Mathematics

University of Virginia
May, 2016

Abstract

A *queueing network* consists of several nodes where at each node there is a server and a queue; jobs pass through the network receiving a random amount of service at each node. In classical Jackson networks, a particular job's random service requirements at each node are independent of one another. This independence enables the computation of various steady-state performance measures and scaling limits.

The independence assumption may not be very realistic, however. Consider instead a queueing network where, although service times are random, any particular job has identical service times at each server. In this situation much dependence is introduced and many classical results break down. Even for the simplest example of a two-node tandem queue, very little is known. In seminal work on this model, Boxma [4] found the steady state distribution for the workload in the second queue at special time-points, in the case arrivals are Poisson. The *workload* is the amount of time a newly arriving job would need to wait for service to begin and represents one of the most important measures of congestion in a queueing system. For the basic two-node model, the complicated dependencies exist only in the second queue. To expand on Boxma's result, we study the entire workload process in the second queue of the two-node tandem system. Unfortunately, this process does not converge under the same scaling as the workload in the first queue. To handle this, we introduce and study a related process M, called the plateau process, which encodes most of the information in the workload process. We show that under appropriate scaling,

workload in the first queue converges, and although the workload in the second queue does not converge, the plateau process does converges to a limit M^* that is a certain function of two independent Lévy processes.

Although the aforementioned result gives a characterization of the long-term dynamics of workload in the second queue, it is difficult to compute distributional quantities from this characterization explicitly. To this end, we find the one dimensional distribution of the plateau process on a certain subsequence of special time points, similar to Boxma's approach. For this more detailed analysis we restrict to the case of exponential interarrival times and regularly varying service times with infinite variance.

Contents

1 Introduction **1**

2 The plateau process $M(\cdot)$ **9**

2.0.1 Notation 9

2.1 Tandem queue model and main result 10

2.1.1 Definition of the model 11

2.1.2 Sequence of models, assumptions, and results 14

2.2 The plateau process as a function of U and V 15

2.2.1 The idleness process for the first queue 15

2.2.2 Workload in the second queue 18

2.3 Continuity properties of G, H, and F 22

2.4 Scaling limit of the plateau process 33

3 The distribution of successive maxima of $M(\cdot)$ **39**

3.1 Notation 40

3.2 The largest job in a busy period 41

3.3 Triangular array and Poisson arrivals 47

3.3.1 Bounds for $ny\bar{m}^{(n)}(ny)$ 49

3.4 Properties of κ 53

3.4.1 The equation that describes κ 53

3.4.2 Properties of $\kappa(y)$ 57

3.5 Representation for $W_2(\tilde{t}_k)$ 60

3.6 Distribution of $W_2^{(n)}(\tilde{t}^{(n)}_{[nt]})$ 63

Chapter 1

Introduction

Queueing theory is the study of waiting lines or queues. The general theory has been used to model diverse systems extending from telecommunication networks to the design of hospitals. A *queueing network* consists of several nodes where at each node there is a server and a queue; jobs pass through the network receiving a random amount of service at each node. Often, one is concerned with the amount of congestion in the system resulting from the configuration of the nodes. Given the primitive data for a queueing network one could compute the congestion in the system for the next few arrivals, but these computations quickly become overwhelming. Over longer periods of time it is reasonable to expect the distribution of the primitive elements of the system to characterize the congestion. This suggests congestion may be well approximated by simpler objects obtained via scaling limits, in much the same way as a large sum may be approximated by a normal random variable.

In classical Jackson networks, a particular job's random service requirements at each node are independent of one another. This independence enables the computation of various steady-state performance measures and scaling limits. For example, Jackson [8] found steady state distributions for queue length and Harrison and Williams [7] find necessary and sufficient conditions for the workload process in each queue to converge to independent reflected Brownian motions.

The independence assumption may not be very realistic, however. Consider the problem of moving files from point A to point B where at the midpoint an operation takes place that requires the presence of the entire file and does not significantly change the file's size. The amount of time required to move the file to the midpoint is about the same as the amount of time required to move the file from the midpoint to point B. In this example we may consider the time it takes to transmit a file as a service time and the two service times experienced by one file would not be independent, but rather depend on the file size.

The above model is an example of a system in which a job's service times are correlated at different nodes in the system. Very little is known about such models, even for the simplest case consisting of two queues in tandem with identical service times at each server. In this situation much dependence is introduced and many classical results break down.

This work concerns a tandem queueing model, consisting of two queues with a single server at each queue that serves jobs according to the First In First Out (FIFO) policy. Jobs arrive to the first queue according to an exogenous renewal process, where they wait to be served by the server in the first queue. When a job's service requirement is satisfied in the first queue it moves immediately to the second queue, where it waits to be served by the second server. Jobs only enter the second queue from the first queue, never from outside the system. When a job's service requirement is satisfied in the second queue it leaves the tandem queueing the system. Critical for this model, is that a job's service time at the second server is identical to it's service time at the first.

In seminal work, Boxma [4] found the steady state distribution for the workload at special time-points in this model, in the case arrivals are Poisson. The *workload* is the amount of time a newly arriving job would need to wait for service to begin and represents an important measure of congestion in a queueing system.

For the basic two-node model, the complicated dependencies exist only in the second queue. Expanding on Boxma's result, we study the entire workload process in the second queue of the two-node tandem system.

Our motivation is to obtain scaling limits for this process, under suitable asymptotic assumptions. Unfortunately, the workload process in the second queue does not always converge in the same setting as the workload in the first queue. That setting is known and we briefly outline the procedure for obtaining the scaling limit.

Consider a triangular array of stochastic primitives for the model: for a sequence $r \to \infty$ in $\mathbb{R}_+$, $\{v_i^r\}_{i=1}^{\infty}$ and $\{u_i^r\}_{i=1}^{\infty}$ are iid sequences of service times and interarrival times respectively to the first queue. Assume positive, finite means $\mathbb{E}\left[v_i^r\right] = \nu^r$ and $\mathbb{E}\left[u_i^r\right] = \mu^r$ for each $i = 1, 2, \ldots$. Let $U^r(t) = \sum_{k=1}^{\lfloor t \rfloor} u_k^r$, $V^r(t) = \sum_{k=1}^{\lfloor t \rfloor} v_k^r$, $\check{U}^r(t) =$

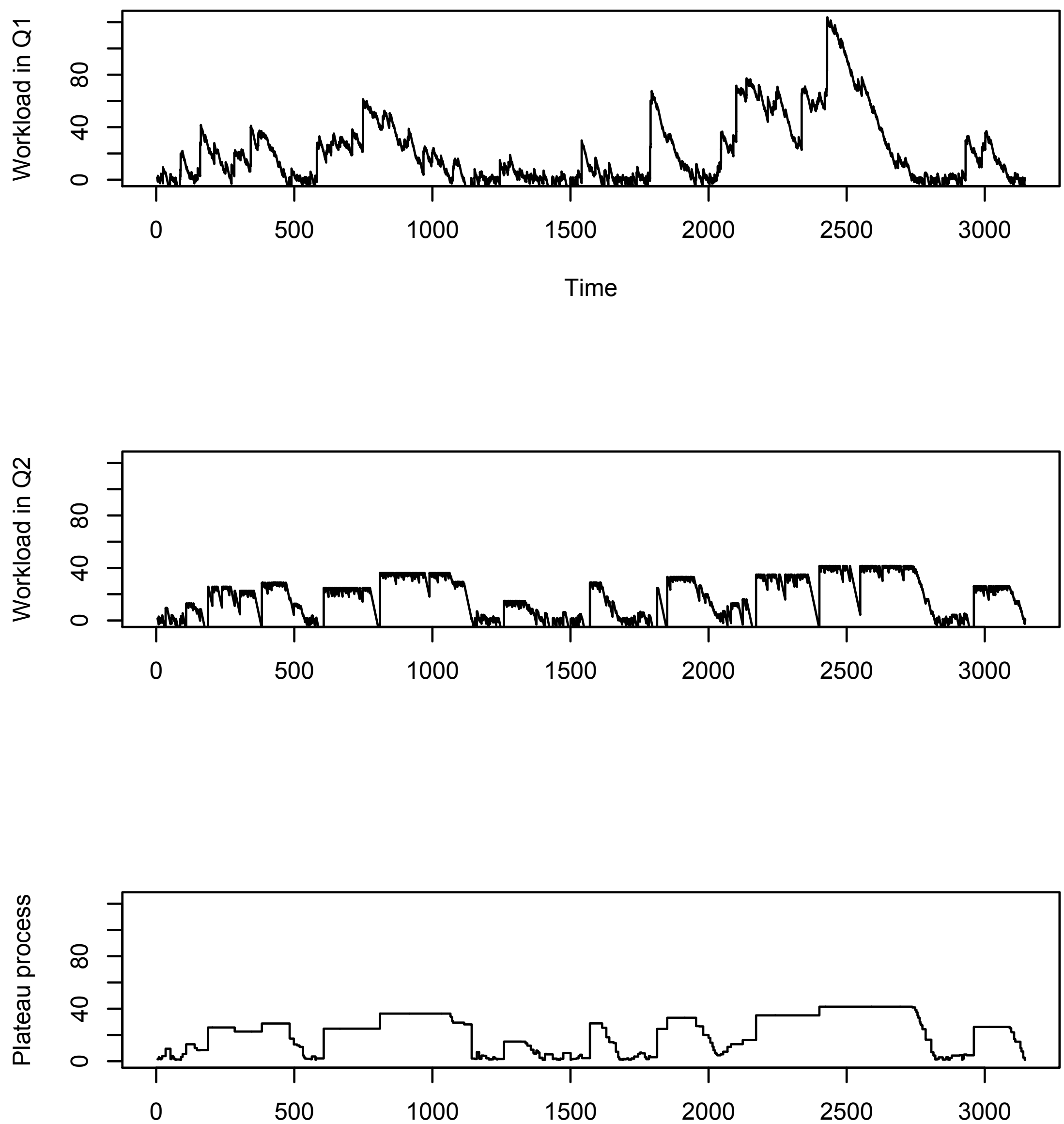

Figure 1.1: The workload in both queues with identical service times in each. 1000 Poisson arrivals with parameter 1/3.1 service times are Pareto(1,3/2).

$\frac{1}{a_r}(U^r(rt) - \mu^r r)$, and $\check{V}^r(t) = \frac{1}{a_r}(V^r(rt) - \nu^r r)$ for some sequence of positive constants a_r tending to infinity. Now, we make the asymptotic assumptions as $r \to \infty$ that $r/a_r \to \infty, \mu^r \to \mu, \nu^r \to \mu$,

$$(\nu^r - \mu^r)\frac{r}{a_r} = \mu^r(\rho^r - 1)\frac{r}{a_r} \to \mu\gamma,$$

$$\check{U}^r \Rightarrow U^*, \text{ and } \check{V}^r \Rightarrow V^*.$$

In this case U^* and V^* must be independent centered Lévy processes. Note that $\check{U}^r \Rightarrow U^*$ if and only if $\check{U}^r(1) \Rightarrow U^*(1)$ since for each r, $\{u_i^r\}$ is iid (See Whitt [12] supplement Theorem 2.4.1). Since $(\rho^r - 1)\frac{r}{a_r} \to \gamma$ we have $\rho^r \to 1$ at rate a_r/r. For example, if $a_r = \sqrt{r}$, $\check{U}^r(1) \Rightarrow N(0, \sigma_u^2)$ and $\check{V}^r(1) \Rightarrow N(0, \sigma_v^2)$ and the workload in the first queue converges to

$$\phi(V^* - U^* + \mu\gamma e)(t/\mu),$$

where $\phi(x)(t) = x(t) - \inf_{0 \le s \le t} x(s)$ is the reflection function and $e(t) = t$ is the identity function. Recall for a stable process B with parameter 2 we have, $B(t/\mu) \sim \frac{1}{\sqrt{\mu}}B(t)$. In this case we recognize $(V^* - U^*)/\sqrt{\mu} + \gamma e$ as a brownian motion with drift γ and variance $(\sigma_u^2 + \sigma_v^2)/\mu$, so the the workload is a reflected brownian motion with these characteristics. In this example, we have $\gamma \le 0$ when $\rho \uparrow 1$.

When V^* has continuous sample paths almost surely, our results show that the workload in the second queue is zero almost surely. On the other hand, if V^* is a Lévy process with parameter $1 < \alpha < 2$, then the workload in the second queue does not converge. To see this, consider a first-queue busy period; then the time between departures is a service time. The effect on the workload process in the second queue is the return to the height attained at the previous arrival unless the current job is larger. The frequency of return to the same height can be seen in the figure 1 where the workload in the second queue must hit zero before each increase. When compared to the workload in the first queue it is clear the behavior is very different because the workload in the second queue frequently has consecutive local maxima of the same value. If the new job is larger, then the workload in the second queue may be zero for a nonzero amount of time, but if a big job is not larger than the current workload in the second queue then the workload process decreases at rate r/a_r for a period long enough for the workload to return to its previous height when the big job arrives. Since $r/a_r \to \infty$ the workload process will fail to have a left limit at such a point.

Notice that the silhouette of the workload in the second queue seems to converge under the same scaling as the workload in the first queue. In contrast to the jagged peaks of the workload in the first queue, the silhouette is characterized by rolling hills. Much of the information about the workload is retained if we only keep track of these recurring levels or plateaus. In doing so we eliminate the oscillating behavior that prevents us from working directly with the workload in the second queue.

This is the strategy we follow. We introduce and study a process M, called the plateau process, which encodes most of the information in the workload process. The plateau process is defined to be the workload in the second queue at the time of the most recent arrival. This definition eliminates the difficulty with scaling described above. We show that under the scaling described above the plateau process converges to a limit M^* that is a certain function of the two independent Lévy processes U^* and V^*. More explicitly, the Nth job waits in the second queue for a period of time $F(U, V, 1)(N)$, where for two functions $x, y : [0, \infty) \to \mathbb{R}$

$$F(x, y, c)(t) = \sup_{0 \le s \le t} \left(y(s) - y(s-) + \sup_{0 \le r \le s} \left(x(r) - y([r - c]^+) \right) \right) - \sup_{0 \le s \le t} \left(x(s) - y([s - c]^+) \right).$$

At time t the number of jobs that have arrived to the second queue is $R(t)$, and the above functions are continuous on a relevant set in the Skorohod path space $\mathbb{D}$. For a sequence of models indexed by r, the plateau process in the rth model can be written

$$M^r(t) = F(U^r, V^r, 1)(R^r(t)).$$

Letting $\check{M}^r(t) = \frac{1}{a_r} M^r(rt)$, we show

$$\check{M}^r \Rightarrow M^*,$$

where $M^*(t) = F(U^* + \gamma\mu e, V^*, 0)(t/\mu)$, see Theorem 2.1.1. This is the subject of Chapter 2.

Although the aforementioned result gives a characterization of the long-term dynamics of workload in the second queue, it is difficult to compute distributional quantities from this characterization explicitly.

To this end, in chapter 3 we consider the sequence of times when the plateau process is large. That is the sequence of times when the last job in a busy period in the first queue arrives to the second queue. After each such time the workload in the second queue will be reduced by the duration of an idle period. We find the one dimensional distribution of the plateau process restricted to this special sequence of times in the case the exogenous arrival process is Poisson and the service times are regularly varying with parameter $1 < \alpha < 2$. We also use a slightly different scaling to compensate for the growing average length of a busy period as $\rho \uparrow 1$.

Chapter 2

The plateau process $M(\cdot)$

In this chapter we find the functions that describe the waiting time in the second queue in terms of the sums of interarrival times U and sums of service times V. We show that these functions are continuous in the Skorohod J_1 topology on relevant subset of $\mathbb{D}$ the space of right continuous functions with finite left limits. The continuous mapping theorem is used to show that the scaled waiting time function also converges on $\mathbb{D}$.

2.0.1 Notation

The following notation will be used throughout. Let $\mathbb{N} = \{1, 2, \ldots\}$ and let $\mathbb{R}$ denote the real numbers. Let $\mathbb{R}_+ = [0, \infty)$. For $a, b \in \mathbb{R}$, write $a \vee b$ for the maximum, and $a \wedge b$ for the minimum, $[a]^+ = 0 \vee a$, $[a]^- = 0 \vee -a$, $\lfloor a \rfloor$ for the integer part of a.

For $f : \mathbb{R}_+ \to \mathbb{R}$ let $f^{\uparrow} = g$ where $g(t) = \sup_{0 \le s \le t} f(s)$.

The arrival time of the nth job to the first queue is denoted t_n while the arrival time of the nth job to the second queue is the *transfer time* of the nth job D_n. D_n is the departure time of the nth job from the first queue in the model.

Following Ethier and Kurtz [5] let Λ' be the collection of strictly increasing functions mapping $\mathbb{R}_+$ onto $\mathbb{R}_+$. Let $\mathbb{D} = \mathbb{D}([0,\infty),\mathbb{R})$ be the space of real valued, right-continuous functions on $[0,\infty)$ with finite left limits. We endow $\mathbb{D}$ with the skorohod J_1-topology which makes $\mathbb{D}$ a Polish space [2]. For $T \geq 0$, let ρ_T be such that $\rho_T(x,y) = \sup_{s\in[0,T]} |x(s) - y(s)|$. Let $e \in \mathbb{D}$ be the identity function $e(t) = t$. For x with finite left limits, in particular $x \in \mathbb{D}$, let $x(t-) = \lim_{s\uparrow t} x(s)$, and $x^- = y$ where $y(t) = x(t-)$ for $t > 0$ and $y(0) = x(0)$.

Let Λ' be the collection of increasing functions λ mapping $\mathbb{R}_+ \to \mathbb{R}_+$. Let $\Lambda \subset \Lambda'$ be the set of Lipschitz continuous functions such that $\lambda \in \Lambda$ implies $\sup_{s>t\geq 0} \left|\log \frac{\lambda(s)-\lambda(t)}{s-t}\right| < \infty$. We will often use [5] proposition 3.5.3: Let $\{x_n\} \subset \mathbb{D}$ and $x \in \mathbb{D}$. Then $x_n \xrightarrow{J_1} x$ if and only if for each $T > 0$ there exists $\{\lambda_n\} \subset \Lambda'$ (possibly depending on T) such that $\lim_{n\to\infty} \sup_{0\leq t\leq T} |\lambda_n(t) - t| = 0$ and $\lim_{n\to\infty} \sup_{0\leq t\leq T} |x_n(t) - x(\lambda_n(t))| = 0$.

Weak convergence of random elements will be denoted by $\Rightarrow$. We adopt the convention that a sum of the form $\sum_{i=n}^{m}$ with $n > m$, or a sum over an empty set of indices equals zero.

2.1 Tandem queue model and main result

In this section we give a precise description of the tandem queue, specify our assumptions, and state our main result.

2.1.1 Definition of the model

We formulate a model equivalent to the one in Boxma [4]. The tandem queueing system consists of two queues Q1 and Q2 in series; both Q1 and Q2 are single-server queues with an unlimited buffer. Jobs enter the tandem system at Q1. After completion of service at Q1 a job immediately enters Q2, and when service at Q2, which is the exact same length as previously experienced in Q1, is completed it leaves the tandem system. Jobs are served individually and at both counters with the first in first out discipline. We assume the system is empty at time zero.

More precisely, Q1, the *exogenous arrival process* $E(\cdot)$ is a renewal process. Jump times of this process correspond to times at which jobs enter the system. This renewal process is defined from a sequence of interarrival times $\{u_i\}_{i=1}^{\infty}$, where u_1 denotes the time at which the first job to arrive after time zero enters the system and u_i, $i \geq 2$, denotes the time between the arrival of the $(i-1)$st and the ith jobs to enter the system after time zero. Thus, $U_i = \sum_{j=1}^{i} u_j$ is the time at which the ith arrival enters the system, which is interpreted as zero if $i = 0$, and $E(t) = \sup\{i \geq 0 : U_i \leq t\}$ is the number of exogenous arrivals by time t. We assume that the sequence $\{u_i\}_{i=1}^{\infty}$ is an i.i.d. sequence of nonnegative random variables with $\mathbb{E}[u_1] = \mu < \infty$.

At Q1, the service process, $\{V_i, i = 1, 2, \ldots\}$, is such that V_i records the total amount of service required from the server by the first i arrivals. More precisely, $\{v_i\}_{i=1}^{\infty}$ denotes an i.i.d. sequence of strictly positive random variables with common distribution function F independent of the collection $\{u_i\}_{i=1}^{\infty}$. We interpret v_i as the amount of processing time that the ith arrival requires from both servers. The v_i's are known as the *service times*. Then, $V_i = \sum_{j=1}^{i} v_j$, which is taken to be zero if $i = 0$. It is assumed that $\mathbb{E}[v_1] = \nu < \infty$.

Note that E and V are assumed to be independent.

For $t \geq 0$, let

$$I(t) = \sup_{s \leq t} \left[V_{E(s)} - s \right]^{-}.$$

We interpret $I(t)$ as the cumulative amount of time that the first server has been idle up to time t. For $n \geq 0$, let

$$I_n = I(U_n).$$

Then I_n is the cumulative amount of time that first server has be idle up to the arrival of the nth job in the first queue.

Let $W_i(t)$ denote the (immediate) workload at time t at Qi, $i = 1, 2$, which is the total amount of time that the server must work in order to satisfy the remaining service requirement of each job present in the system at time t, ignoring future arrivals. For $t \geq 0$ we define

$$W_1(t) = V_{E(t)} - t + I(t).$$

Let D_n be the *transfer time* of the nth job. So, the nth job exits Q1 and enters Q2 at time D_n. Let $d_1 = u_1 + v_1$ and $d_n = D_n - D_{n-1}$ for $n \geq 2$ be the *intertransfer time* between arrivals of $n-1$st and nth job to the second queue. For $n \geq 0$ we have

$$D_n = V_n + I_n.$$

Let $R(t)$ denote the number of transfers to Q2 by time t. For $t \geq 0$ we have

$$R(t) = \sup\{n \geq 0 : D_n \leq t\}. \tag{2.1.1}$$

Let $J(t)$ denote the cumulative amount of time that the second server has been idle up to time t, and $W_2(t)$ as the workload in Q2 at time t. That is, for $t \geq 0$ let

$$J(t) = \sup_{s \leq t} \left[V_{R(s)} - s\right]^-,$$

$$W_2(t) = V_{R(t)} - t + J(t).$$

If k is the index of the first job in a busy period of the first queue then $W_1(t_k) = v_k$. Similarly, $W_2(D_k) = v_k$ if the kth job arrives to the second queue at a time when the second queue is empty.

Finally, let M_n denote the workload in the second queue at the time of the arrival of the nth job to the second queue, which is just the sojourn time of the nth job in the second queue. Let $M(t)$ be the piecewise constant right continuous function that agrees with the work load in the second queue at each transfer time and whose discontinuities are contained in the transfer times. We call $M(t)$ the *plateau process*. For integers $n \geq 0$ and real numbers $t \geq 0$ we have

$$\begin{aligned} M_n &= W_2(D_n), \\ M(t) &= M_{R(t)}. \end{aligned} \tag{2.1.2}$$

The name plateau process comes from the tendency of M_{n+1} to be equal to M_n, although M_n may increase or decrease.

Definition 1. *For a real number $t \geq 0$,*

$$U(t) = U_{\lfloor t \rfloor} \qquad \textit{and} \qquad V(t) = V_{\lfloor t \rfloor}.$$

2.1.2 Sequence of models, assumptions, and results

We now specify a sequence of tandem queueing models indexed by $r \in \mathbb{R}$. Each model in the sequence is defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The rth model in the sequence is defined as in the previous section where we add a superscript r to each symbol. In particular, for $t \geq 0$ let $M^r(t)$ denote the plateau process in the rth system.

That is, a sequence of tandem queueing models indexed by r, where r increases to ∞ through a sequence in $(0, \infty)$, $\{v_i^r\}_{i=1}^{\infty}$ and $\{u_i^r\}_{i=1}^{\infty}$ are the service times and interarrival times to the first queue with positive, finite means $\mathbb{E}[v_i^r] = \nu^r$ and $\mathbb{E}[u_i^r] = \mu^r$ for each $i = 1, 2, \ldots$ independent of each other. Define the following scaled versions of processes in the rth model for a sequence of positive reals $a_r \to \infty$ and $t \geq 0$,

$$\begin{aligned} \bar{U}^r(t) &= r^{-1} U^r(rt) & \text{and} \qquad & \bar{V}^r(t) = r^{-1} V^r(rt) \\ \check{U}^r(t) &= a_r^{-1} \left(U^r(rt) - r\mu^r t \right) & \text{and} \qquad & \check{V}(t) = a_r^{-1} \left(V^r(rt) - r\nu^r t \right) \\ & \check{M}^r(t) = a_r^{-1} M^r(rt) \end{aligned} \tag{2.1.3}$$

Asymptotic assumptions We make the following asymptotic assumptions, as $r \to \infty$, about our sequence of models. Assume there is a sequence $\{a_r\}$ such that $r/a_r \to \infty$, $\check{U}^r(1) \Rightarrow \mathcal{U}^*$, $\check{V}^r(1) \Rightarrow \mathcal{V}^*$ in $\mathbb{R}$. In this case $\mathcal{U}^*$ and $\mathcal{V}^*$ are centered infinitely divisible random variables see Feller [6] XII.7. In this case we have $\check{U}^r \Rightarrow U^*$ and $\check{V}^r \Rightarrow V^*$ in $\mathbb{D}$, where U^* and V^* are Lévy stable motions with $U^*(1) \sim \mathcal{U}^*$ and $V^*(1) \sim \mathcal{V}^*$ by Whitt [12] supplement 2.4.1. We further assume $\lim_{r\to\infty} \mu^r = \lim_{r\to\infty} \nu^r = \mu$ and the traffic intensity parameter for the rth system $\rho^r = \frac{\mu^r}{\nu^r}$ satisfies

$$\frac{r}{a_r}\left(1-\rho^r\right) \to \gamma \in \mathbb{R}.$$

Definition 2. *Define the mapping $F : \mathbb{D} \times \mathbb{D} \times \mathbb{R} \to \mathbb{D}$ by*

$$F(x,y,c)(t) = \sup_{0\le s\le t}\left(y(s) - y(s-) + \sup_{0\le r\le s}\left(x(r) - y([r-c]^+)\right)\right. \\ \left. - \sup_{0\le s\le t}\left(x(s) - y([s-c]^+)\right)\right)$$

Theorem 2.1.1.

$$\check{M}^r \Rightarrow M^*,$$

where $M^*(t) = F(U^* + \gamma\mu e, V^*, 0)(t/\mu)$.

2.2 The plateau process as a function of U and V

In this section we derive various relationships namely the stochastic processes comprising the tandem queueing model. These relationships hold for any of the r indexed models, so we suppress superscripts referring to a particular model in sequence.

2.2.1 The idleness process for the first queue

This section is a prerequisite for understanding the arrival process in the second queue. If the cumulative idleness in the first queue is identically zero for all of time then the arrival process to the second queue is just a renewal process formed by the service times. Here we consider the cumulative idleness process in the first queue as a discrete time process. Consider the model defined in section 2.1.1

Lemma 2.2.1. *For each* $n \geq 1$,

$$I_n = u_1 + \max_{k=1}^{n} \left(\sum_{j=2}^{k} (u_j - v_{j-1}) \right), \tag{2.2.1}$$

for $n = 1, 2, \ldots$.

Proof. We proceed by induction. First observe that $\sum_{j=2}^{1}(u_j - v_{j-1}) = 0$, by convention, so

$$\max_{k=1}^{n} \left(\sum_{j=2}^{k} (u_j - v_{j-1}) \right) \geq 0$$

for $n \geq 1$. $I_1 = u_1 + \max_{k=1}^{1} \sum_{j=1}^{k} (u_j - v_{j-1}) = u_1$. For $n = 2$,

$$I_2 = u_1 + [u_2 - v_1]^+ = u_1 + \max_{k=1}^{2} \left(\sum_{j=2}^{k} (u_j - v_{j-1}) \right),$$

since there is no additional idleness if the second job arrives while the first job is in service. This is the base case for the induction.

For the inductive step, assume equation (2.2.1) holds for $n \geq 2$. There are two cases. In the first case the $(n+1)$st job arrives before the nth service is complete. In this case the first job in the current busy period had index $i \leq n$, arrived at time t_i, and the total amount of work that has arrived since t_i, $\sum_{k=i}^{n} v_k$ exceeds the amount of time $\sum_{k=i+1}^{n+1} u_k$ since t_i. That is,

$$\sum_{k=i+1}^{n+1} u_k - v_{k-1} < 0,$$

for some $i \leq n$. Thus

$$\max_{k=1}^{n+1}\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right) = \max_{k=1}^{n}\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right),$$

and the cumulative idle time has not increased

$$I_n = I_{n+1} = u_1 + \max_{k=1}^{n+1}\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right).$$

In the second case, the $(n+1)$st job arrives after the nth service is complete, so the total idle time just before the arrival of the $n+1$ job is $u_1 + \sum_{k=2}^{n+1} u_k - v_{k-1}$. In this case, for any job $i \leq n$, the total amount of time $\sum_{k=i+1}^{n+1} u_k$ exceeds the total amount of work $\sum_{k=i}^{n} v_k$ since t_i. That is,

$$\sum_{k=i+1}^{n+1} u_k - v_{k-1} \geq 0.$$

Thus,

$$\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right) \leq \left(\sum_{j=2}^{n+1}(u_j - v_{j-1})\right)$$

for each $k = 2, \ldots, n+1$, and we have $\displaystyle\sum_{j=2}^{n+1} u_j - v_{j-1} = \max_{k=1}^{n+1}\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right)$. ■

Note that the departure process of the first queue is equal to the arrival process $R(\cdot)$ of the second queue. Since the queueing discipline is FIFO, the number of jobs that have arrived to the second queue by time t is the greatest number N such that the total amount of time needed to complete the first N jobs, $\sum_{k=1}^{N} v_k$, is less than the amount of time spent working, t minus the cumulative idle time in the first queue.

2.2.2 Workload in the second queue

In this section we show how to write the plateau process $M(\cdot)$ as a function of the primitive arrival and service processes. The following formula relates sojourn times in the second queue to service times and idleness in the first queue. It comes from Lindley recursion [1] for a FIFO queue $W_2(D_{n+1}) = v_{n+1} + [W_2(D_n) - d_{n+1}]^+$, where no independence needs to be assumed about the intertransfer times d_k and service times v_k.

Lemma 2.2.2. *The sojourn time of the n^{th} job in the second queue is*

$$M_n = \max_{k=1}^{n} \{v_k + I_k\} - I_n.$$

Proof. Note that the sojourn time of the n^{th} job includes its service time. The second queue is initially empty and the service time of the nth job is the same in both queues. Clearly $I_1 = u_1$, since the first queue is empty until the arrival of the first job. So,

$$M_1 = v_1 = \max_{k=1}^{1} \{v_k + I_k\} - I_1.$$

The intertransfer time between the nth and $(n+1)$st job is $d_{n+1} = v_{n+1} + (I_{n+1} - I_n)$. Proceeding by induction, suppose $M_n = \max_{k=1}^{n} \{v_k + I_k\} - I_n$. Then, Lindley recursion gives

$$
\begin{aligned}
M_{n+1} &= v_{n+1} + [M_n - v_{n+1} - (I_{n+1} - I_n)]^+ \\
&= v_{n+1} \vee (M_n - (I_{n+1} - I_n)) \\
&= v_{n+1} \vee \left(\max_{k=1}^{n} (v_k + I_k) - I_n - (I_{n+1} - I_n) \right) \\
&= \left[(v_{n+1} + I_{n+1}) \vee \max_{k=1}^{n} (v_k + I_k) \right] - I_{n+1} \\
&= \max_{k=1}^{n+1} (v_k + I_k) - I_{n+1}.
\end{aligned}
$$

∎

Definition 3. *Define the translation function $G : \mathbb{D} \times \mathbb{R} \to \mathbb{D}$ by*

$$G(x, c)(t) = x([t - c]^+),$$

and define $H : \mathbb{D} \times \mathbb{D} \times \mathbb{R}_+ \to \mathbb{D}$ as the composition

$$H(x, y, c) = (x - G(y, c))^{\uparrow}.$$

More explicitly,

$$H(x, y, c)(t) = \sup_{0 \le s \le t} \left(x(s) - y([s - c]^+) \right).$$

We can write I_n in terms of V and U from definition 1.

Lemma 2.2.3. *For each $n \geq 1$,*

$$I_n = H(U, V, 1)(n),$$

Moreover H is constant on intervals of the form $[n, n + 1)$ where n is an integer, so for each integer n we have $H(U, V, n)(\lfloor t \rfloor) = H(U, V, n)(t)$ for all $t \geq 0$.

Proof. The processes V and U are constant between integers so H is constant on intervals of the form $[n, n+1)$, where n is an integer. For an integer k, $v_k = V(k) - V(k-)$ and $u_k = U(k) - U(k-)$. By lemma 2.2.1,

$$\begin{aligned} I_n &= u_1 + \max_{k=1}^{n}\left(\sum_{j=2}^{k}(u_j - v_{j-1})\right) \\ &= u_1 + \max_{k=1}^{n}\left(\sum_{j=2}^{k} u_j - \sum_{j=1}^{k-1} v_j\right) \\ &= \max_{k=1}^{n}\left(\sum_{j=1}^{k} u_j - \sum_{j=1}^{k-1} v_j\right) \\ &= \max_{k=1}^{n}\left(U(k) - V(k-1)\right) \\ &= \sup_{0\le s\le n}\big(U(s) - V([s-1]^+)\big) \\ &= \sup_{0\le s\le n}\left(U(s) - G(V,1)(s)\right) \\ &= H(U,V,1)(n). \end{aligned}$$

■

Now we can write R in terms of U and V.

Corollary 2.2.4.

$$R(t) = \max\left\{m \ge 0 : V(m) + H(U,V,1)(m) \le t\right\}.$$

Proof. From definition (2.1.1) we have $R(t) = \max\{N \ge 0 : \sum_{k=1}^{N} v_k + I_N \le t\}$. We have $\sum_{k=1}^{N} v_k = V(N)$ by definition 1 and $I_N = H(U,V,1)(N)$ by lemma 2.2.3 ■

We can now write the plateau process in terms of the function F defined in section

2.1.2. By definitions 2 and 3,

$$F(x,y,c) = \big(y - y^{-} + H(x,y,c)\big)^{\uparrow} - H(x,y,c),$$

or more explicitly,

$$F(x,y,c)(t) = \sup_{0\le s\le t} (y(s) - y(s-) + H(x,y,c)(s)) - H(x,y,c)(t).$$

Lemma 2.2.5. *For all $t \ge 0$,*

$$M_{\lfloor t\rfloor} = F(U,V,1)(t).$$

Proof. By lemma 2.2.2

$$\begin{aligned} M_{\lfloor t\rfloor} &= \max_{k=1}^{\lfloor t\rfloor} (v_k + I_k) - I_{\lfloor t\rfloor} \\ &= \max_{k=1}^{\lfloor t\rfloor} (V(k) - V(k-) + I_k) - I_{\lfloor t\rfloor} \\ &= \max_{k=1}^{\lfloor t\rfloor} (V(k) - V(k-) + H(U,V,1)(k)) - H(U,V,1)(\lfloor t\rfloor) \end{aligned}$$

by lemma 2.2.3. For a positive integer k we have $H(U,V,1)(t)$ is constant for t in $[k,k+1)$ and $V(k)-V(k-) \ge V(t)-V(t-)$ for t in $[k,k+1)$. Thus, $V(t)-V(t-)+H(U,V,1)(t)$ is maximized when t is an integer. Thus,

$$\begin{aligned} M_{\lfloor t\rfloor} &= \sup_{0\le s\le t} (V(s) - V(s-) + H(U,V,1)(s)) - H(U,V,1)(t) \\ &= F(U,V,1)(t). \end{aligned}$$

■

Finally we can express $M(\cdot)$ as function of U and V. By definition (2.1.2), $M(t)$

is the composition $M_{(\cdot)}$ with the arrival process to the second queue. That is,

$$\begin{aligned} M(t) &= M_{R(t)} \\ &= F(U,V,1)(\max\{m \geq 0 : V(m) + H(U,V,1)(m) \leq t\}). \end{aligned}$$

Notice that the plateau process is greater than or equal to the workload in the second queue at each time, that is $M(t) \geq W_2(t)$ for each $t \geq 0$.

2.3 Continuity properties of G, H, and F

In this section we identify a subset of the domain of F that contains the limits of the processes we are interested and where F is continuous. This result is obtained by treating F as a composition of continuous functions. The method of proof is similar to how Whitt showed addition is continuous on a large set in [11].

Lemma 2.3.1. *For any $x \in \mathbb{D}$, G is continuous at $(x,0)$ in the product topology on $\mathbb{D} \times \mathbb{R}$.*

Proof. Let c_n be a sequence in $\mathbb{R}$ with $c_n \to 0$, and let $x_n \to x$ in $\mathbb{D}$. Then for each $T > 0$ there exists $\{\lambda_n\} \subset \Lambda$ such that $\sup_{0 \leq t \leq T} |\lambda_n(t) - t| \to 0$ as $n \to \infty$ and $\sup_{0 \leq t \leq T} |x_n(t) - x(\lambda_n(t))| \to 0$ as $n \to \infty$.

For each $n = 1, 2, \ldots$ define

$$\tilde{\lambda}_n(t) = \begin{cases} \lambda_n(t - c_n), & \text{if } t \geq 2|c_n|, \\ \lambda_n\left(\left(1 - \frac{sgn(c_n)}{2}\right) t\right), & \text{if } t < 2|c_n|, \end{cases}$$

where $sgn(c_n) = -1$ if $c_n < 0$, $sgn(c_n) = 1$ if $c_n > 0$, and $sgn(c_n) = 0$ if $c_n = 0$.

We have $\{\tilde{\lambda}_n\} \subset \Lambda$ because each $\tilde{\lambda}_n$ is the composition of two functions in Λ. Now,

$$\begin{aligned}
\sup_{0\le t\le T}\left|\tilde{\lambda}_n(t)-t\right| &= \left(\sup_{0\le t<2|c_n|}\left|\tilde{\lambda}_n(t)-t\right|\right)\vee\left(\sup_{2|c_n|\le t\le T}\left|\tilde{\lambda}_n(t)-t\right|\right)\\
&= \left(\sup_{0\le t<2|c_n|}\left|\lambda_n\left(\left(1-\frac{sgn(c_n)}{2}\right)t\right)-t\right|\right)\vee\left(\sup_{2|c_n|\le t\le T}\left|\lambda_n(t-c_n)-t\right|\right)\\
&\le \left(\sup_{0\le t<2|c_n|}\left|\lambda_n\left(\left(1-\frac{sgn(c_n)}{2}\right)t\right)-\left(1-\frac{sgn(c_n)}{2}\right)t\right|\right.\\
&\quad\left.+\sup_{0\le t\le 2|c_n|}\left|\left(1-\frac{sgn(c_n)}{2}\right)t-t\right|\right)\vee\left(\sup_{2|c_n|\le t\le T}\left|\lambda_n(t-c_n)-(t-c_n)\right|+|c_n|\right).
\end{aligned}$$

When $0\le t<2|c_n|$ we have $0\le\left(1-\frac{sgn(c_n)}{2}\right)t\le 3|c_n|$, so

$$\begin{aligned}
\sup_{0\le t\le T}\left|\tilde{\lambda}_n(t)-t\right| &\le \left(\sup_{0\le t<3|c_n|}\left|\lambda_n(t)-t\right|+3|c_n|\right)\\
&\quad\vee\left(\sup_{2|c_n|-c_n\le t\le T-c_n}\left|\lambda_n(t)-t\right|+|c_n|\right)\\
&\le \sup_{0\le t\le T}\left|\lambda_n(t)-t\right|+3|c_n|,
\end{aligned}$$

so $\sup_{0\le t\le T}\left|\tilde{\lambda}_n(t)-t\right|\to 0$ as $n\to\infty$.

Now, it suffices to show $\sup_{0\le t\le T}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|\to 0$ by [5] Proposition 3.5.3. We have

$$\sup_{2|c_n|\le t\le T}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|$$

$$=\sup_{2|c_n|\le t\le T}\left|x_n([t-c_n]^+)-x(\tilde{\lambda}_n(t))\right|$$

$$=\sup_{2|c_n|\le t\le T}|x_n(t-c_n)-x(\lambda_n(t-c_n))|$$

$$=\sup_{2|c_n|-c_n\le t\le T-c_n}|x_n(t)-x(\lambda_n(t))|\to 0 \quad (2.3.1)$$

So it suffices to show $\sup_{0\le t<2|c_n|}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|\to 0$.

Fix $\epsilon>0$ and let $\eta>0$ such that $\sup_{0\le t\le\eta}|x(0)-x(t)|<\epsilon$ by right continuity of x at zero. Now, for n so large that $|c_n|<\min(T/3,\eta/6)$, $\sup_{0\le t\le T}|\lambda_n(t)-t|<\epsilon\wedge\eta/2$, and $\sup_{0\le t\le T}|x_n(t)-x(\lambda_n(t))|<\epsilon$ consider the $c_n<0$, $c_n>0$, and $c_n=0$ cases.

If $c_n<0$,

$$\sup_{0\le t<2|c_n|}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|$$

$$=\sup_{0\le t<2|c_n|}\left|x_n([t-c_n]^+)-x(\tilde{\lambda}_n(t))\right|$$

$$=\sup_{0\le t<-2c_n}|x_n(t-c_n)-x(\lambda_n(3t/2))|$$

$$\le\sup_{0\le t<-2c_n}|x_n(t-c_n)-x(\lambda_n(t-c_n))|+|x(\lambda_n(t-c_n))-x(\lambda_n(3t/2))|$$

$$\le\sup_{0\le t\le T}|x_n(t)-x(\lambda_n(t))|+\sup_{0\le t<-2c_n}|x(\lambda_n(t-c_n))-x(\lambda_n(3t/2))|$$

$$\le\sup_{0\le t\le T}|x_n(t)-x(\lambda_n(t))|+\sup_{0\le t<-2c_n}|x(\lambda_n(t-c_n))|+\sup_{0\le t<-2c_n}|x(\lambda_n(3t/2))|.$$

We have $(t-c_n)\vee(3t/2)\le -3c_n$ for $0\le t<-2c_n$, and so

$$\lambda_n(t-c_n)\vee\lambda_n(3t/2)\le\lambda_n(-3c_n)\le -3c_n+\eta/2\le\eta.$$

Thus,

$$\sup_{0\le t<2|c_n|}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|$$

$$\le \epsilon + \sup_{0\le t<-2c_n}|x(\lambda_n(t-c_n))| + \sup_{0\le t<-2c_n}|x(\lambda_n(3t/2))|$$

$$\le \epsilon + \sup_{0\le t\le\eta}|x(t)| + \sup_{0\le t\le\eta}|x(t)| \le 3\epsilon$$

If $c_n > 0$,

$$\sup_{0\le t<2|c_n|}\left|G(x_n,c_n)(t)-G(x,0)(\tilde{\lambda}_n(t))\right|$$

$$= \sup_{0\le t<2c_n}\left|x_n([t-c_n]^+)-x(\tilde{\lambda}_n(t))\right|$$

$$= \sup_{0\le t<2c_n}\left|x_n([t-c_n]^+)-x(\lambda_n(t/2))\right|$$

$$\le \sup_{0\le t<c_n}|x_n(0)-x(\lambda_n(t/2))| \vee \sup_{c_n\le t<2c_n}|x_n(t-c_n)-x(\lambda_n(t/2))|\,. \quad (2.3.2)$$

For the first term,

$$\sup_{0\le t\le c_n}|x_n(0)-x(\lambda_n(t/2))| \le \sup_{0\le t<c_n}|x_n(0)-x(0)|+|x(0)-x(\lambda_n(t/2))|$$

$$= |x_n(0)-x(\lambda_n(0))| + \sup_{0\le t<c_n}|x(0)-x(\lambda_n(t/2))|$$

$$\le \sup_{0\le t\le T}|x_n(t)-x(\lambda_n(t))| + \sup_{0\le t\le\eta}|x(0)-x(t)| \le 2\epsilon,$$

since $\lambda_n(t/2) \le \lambda_n(c_n/2) \le c_n/2 + \eta/2 \le \eta$ for $0 \le t \le c_n$. For the second term,

$$\begin{aligned}
\sup_{c_n \le t < 2c_n} |x_n(t - c_n) - x(\lambda_n(t/2))| &= \sup_{0 \le t < c_n} \left| x_n(t) - x\left(\lambda_n\left(\frac{t + c_n}{2}\right)\right)\right| \\
&\le \sup_{0 \le t < c_n} |x_n(t) - x(\lambda_n(t))| + \left| x(\lambda_n(t)) - x\left(\lambda_n\left(\frac{t + c_n}{2}\right)\right)\right| \\
&\le \epsilon + \sup_{0 \le t < c_n} \left| x(\lambda_n(t)) - x(0) + x(0) - x\left(\lambda_n\left(\frac{t + c_n}{2}\right)\right)\right| \\
&\le \epsilon + \sup_{0 \le t < c_n} |x(\lambda_n(t)) - x(0)| + \sup_{0 \le t < c_n} \left| x(0) - x\left(\lambda_n\left(\frac{t + c_n}{2}\right)\right)\right| \\
&\le \epsilon + 2 \sup_{0 \le t < \eta} |x(0) - x(t)| \le 3\epsilon,
\end{aligned}$$

since $\lambda_n(t) \vee \lambda_n(\frac{t+c_n}{2}) \le \lambda_n(c_n) \le c_n + \eta/2 \le \eta$ for $0 \le t \le c_n$.

If $c_n = 0$ then $\tilde{\lambda}_n = \lambda_n$ so $G(x_n, c_n)(t) - G(x, 0)(\tilde{\lambda}_n(t)) = x_n(t) - x(\lambda_n(t))$, which converges to zero uniformly by assumption.

So in all three cases we have

$$\sup_{0 \le t < 2|c_n|} \left| G(x_n, c_n)(t) - G(x, 0)(\tilde{\lambda}_n(t)) \right| \le 3\epsilon.$$

Together with (2.3.1) and since ϵ was arbitrary, we have

$$\sup_{0 \le t \le T} \left| G(x_n, c_n)(t) - G(x, 0)(\tilde{\lambda}_n(t)) \right| \to 0$$

as $n \to \infty$.

So we have $G(x_n, c_n) \to G(x, 0)$ on $\mathbb{D}$. ■

For $x \in \mathbb{D}$, let $\text{Disc}(x)$ denote the set of discontinuities of x.

Lemma 2.3.2. *H is continuous at $(x, y, 0)$ for all $x, y \in \mathbb{D}$ such that*

$$Disc(x) \cap Disc(y) = \varnothing.$$

Proof. Let $c_n \in \mathbb{R}$ with $c_n \to 0$ and let x_n and y_n be in $\mathbb{D}$ such that $x_n \to x$ and $y_n \to y$ and fix a time $T > 0$. Let $z_n = y_n - x_n$ and $z = y - x$. Since $\text{Disc}(x) \cap \text{Disc}(-y) = \varnothing$, [11] Theorem 4.1 tells us that there exists $\{\lambda_n\} \subset \Lambda'$ such that $\rho_T(\lambda_n, e) \to 0$ and $\rho_T(z_n, z \circ \lambda_n) \to 0$. Since G is continuous at $(z, 0)$ by lemma 2.3.1, and $(z_n, c_n) \to (z, 0)$ we have $\{\tilde{\lambda}_n\} \subset \Lambda'$ such that $\rho_T(\tilde{\lambda}_n, e) \to 0$ and $\rho_T(G(z_n, c_n), z \circ \tilde{\lambda}_n) \to 0$. In fact, we may construct $\tilde{\lambda}_n$ as in the proof of 2.3.1. Since $x \mapsto x^\uparrow$ is continuous on $\mathbb{D}$ and $(x)^\uparrow \circ \tilde{\lambda} = (x \circ \tilde{\lambda})^\uparrow$, we have $\rho_T(H(x_n, y_n, c_n), H(x, y, 0) \circ \tilde{\lambda}_n) \to 0$. Since T was arbitrary we have H is continuous $(x, y, 0)$. ∎

Lemma 2.3.3. *For all $x, y \in \mathbb{D}$,*

$$Disc(H(x, y, 0)) \subset \{t : y(t) - y(t-) > 0\} \cup \{t : x(t) - x(t-) < 0\}.$$

In particular, if $\{t : x(t) - x(t-) < 0\} = \varnothing$, then

$$Disc(H(x, y, 0)) \subset Disc(y).$$

Proof. $\text{Disc}(H(x, y, 0)) = \{t : H(x, y, 0)(t) - H(x, y, 0)(t-) \neq 0\} = \{t : H(x, y, 0)(t) - H(x, y, 0)(t-) > 0\}$ since $H(x, y, 0)$ is nondecreasing. Thus,

$$\begin{aligned} \text{Disc}(H(x, y, 0)) &\subset \{t : (y - x)(t) - (y - x)(t-) > 0\} \\ &\subset \{t : y(t) - y(t-) > 0\} \cup \{t : x(t) - x(t-) < 0\}. \end{aligned}$$

∎

Lemma 2.3.4. *Let λ_n and γ_n be strictly increasing homeomorphisms from $[0,T]$ onto $[0,T]$ and $x_n, x \in \mathbb{D}$ such that for some finite collection $\{t_j\}_{j=0}^N \subset [0,T]$ with*

(i) $0 = t_0 < t_1 < \cdots < t_N = T$ *we have* $\lambda_n^{-1}(t_j) = \gamma_n^{-1}(t_j)$ *for each* $j = 0, 1, 2, \ldots, N$,

(ii) $\rho_T(x_n, x \circ \lambda_n) < \epsilon$, *and*

(iii) $w(x, [t_{j-1}, t_j)) = \sup\left(|x(t) - x(s)| : t, s \in [t_{j-1}, t_j)\right) < \epsilon$ *for each* $j = 1, 2, \ldots, N$,

then

$$\rho_T(x_n, x \circ \gamma_n) < 3\epsilon.$$

Proof. Let $r_j = \gamma_n^{-1}(t_j) = \lambda_n^{-1}(t_j)$ for $j = 0, 1, \ldots, N$, so that $\cup_{j=1}^N [r_{j-1}, r_j) =$

$\cup_{j=1}^{N} [t_{j-1}, t_j) = [0, T)$.

$$
\begin{aligned}
\rho_T(x_n, x \circ \gamma_n) &= \sup_{0 \le t \le T} |x_n(t) - x(\gamma_n(t))| \\
&= \max_{k=1}^{N} \sup_{r_{j-1} \le t < r_j} |x_n(t) - x(\gamma_n(t))| \vee |x_n(T) - x(T)| \\
&= \max_{k=1}^{N} \sup_{t_{j-1} \le t < t_j} \left|x_n(\gamma_n^{-1}(t)) - x(t)\right| \vee |x_n(T) - x(T)| \\
&= \max_{k=1}^{N} \sup_{t_{j-1} \le t < t_j} \left|x_n(\gamma_n^{-1}(t)) - x(t_{j-1}) + x(t_{j-1}) - x(t)\right| \\
&\qquad \vee |x_n(T) - x(T)| \\
&\le \max_{k=1}^{N} \left(\sup_{t_{j-1} \le t < t_j} \left|x_n(\gamma_n^{-1}(t)) - x(t_{j-1})\right| + w(x, [t_{j-1}, t_j)) \right) \\
&\qquad \vee |x_n(T) - x(T)| \\
&\le \max_{k=1}^{N} \left(\sup_{r_{j-1} \le t < r_j} |x_n(t) - x(\lambda_n(r_{j-1}))| + \epsilon \right) \\
&\qquad \vee |x_n(T) - x(T)| \\
&\le \max_{k=1}^{N} \left(\sup_{r_{j-1} \le t < r_j} |x_n(t) - x(\lambda_n(t))| \right. \\
&\qquad \left. + \ |x(\lambda_n(t)) - x(\lambda_n(r_{j-1}))| + \epsilon \right) \vee |x_n(T) - x(T)| \\
&\le \max_{k=1}^{N} \left(\sup_{r_{j-1} \le t < r_j} |x_n(t) - x(\lambda_n(t))| + w(x, [t_{j-1}, t_j)) + \epsilon \right) \\
&\qquad \vee |x_n(T) - x(T)| \\
&\le \max_{k=1}^{N} \left(\sup_{r_{j-1} \le t < r_j} |x_n(t) - x(\lambda_n(t))| + 2\epsilon \right) \vee |x_n(T) - x(T)| \\
&\le \rho_T(x_n, x \circ \lambda_n) + 2\epsilon \\
&\le 3\epsilon.
\end{aligned}
$$

■

Finally, we prove that F is continuous on a relevant set.

Lemma 2.3.5. *F is continuous at $(x, y, 0)$ in the product topology on $\mathbb{D} \times \mathbb{D} \times \mathbb{R}$, for all x and $y \in \mathbb{D}$ with Disc$(x) \cap$ Disc$(y) = \varnothing$ and*

$$\{t : y(t) - y(t-) < 0\} = \varnothing.$$

Proof. Let $T > 0$, let ρ_T be the uniform metric on function from $[0, T]$ to $\mathbb{R}$, and fix $\epsilon > 0$. Apply Lemma 1 on page 110 of [2] to construct finite subsets $A_1 = \{t'_j\}$ and $A_2 = \{s_j\}$ of $[0, T]$ such that $0 = t'_0 < \cdots < t'_k = T$, $0 = s_0 < \cdots < s_m = T$, $w(y; [t'_{j-1}, t'_j)) = \sup\{|y(s) - y(t)| : s, t \in [t'_{j-1}, t'_j)\} < \epsilon$ and $w(H(x, y, 0); [s_{j-1}, s_j)) < \epsilon$ for all j. Since $\text{Disc}(y) \cap \text{Disc}(H(x, y, 0)) \subset \text{Disc}(x) \cap \text{Disc}(y) = \varnothing$, the two sets A_1 and A_2 can be chosen so that $A_1 \cap A_2 = \{0, T\}$. Note that $w(y; [t_{j-1}, t_j)) < \epsilon$ and $w(H(x, y, 0); [t_{j-1}, t_j)) < \epsilon$ for $\{t_j\} = A_1 \cup A_2$. Let 2δ be the distance between the closest two points in $A_1 \cup A_2$. Choose n_0 and homeomorphisms λ_n and μ_n in Λ so that

(*i*) $\rho_T(y_n, y \circ \lambda_n) < (\delta \wedge \epsilon)$,

(*ii*) $\rho_T(\lambda_n, e) < (\delta \wedge \epsilon)$,

(*iii*) $\rho_T(H(x_n, y_n, c_n), H(x, y, 0) \circ \mu_n) < (\delta \wedge \epsilon)$, and

(*iv*) $\rho_T(\mu_n, e) < (\delta \wedge \epsilon)$

for $n \geq n_0$. Thus for $n \geq n_0$

$$\lambda_n^{-1}(A_1) \cap \mu_n^{-1}(A_2) = \{0, T\}$$

and $\{r_j\} = \lambda_n^{-1}(A_1) \cup \mu_n^{-1}(A_2)$ has corresponding points in the same order as $\{t_j\} = A_1 \cup A_2$. Let γ_n be homeomorphisms of $[0, T]$ defined by

$$\gamma_n(r_j) = t_j$$

for corresponding points $r_j \in \lambda_n^{-1}(A_1) \cup \mu_n^{-1}(A_2)$ and $t_j \in A_1 \cup A_2$ and by linear interpolation elsewhere.

Note that for each $r_j \in \lambda_n^{-1}(A_1) \cup \mu_n^{-1}(A_2)$ either

$$\lambda_n(r_j) = t_j \qquad \text{or} \qquad \mu_n(r_j) = t_j.$$

Since $t \mapsto |\gamma_n(t) - t|$ is continuous the maximum is attained at some critical point (exposed point) r_j, so $\rho_T(\gamma_n, e) < \rho_T(\lambda_n, e) \vee \rho_T(\mu_n, e) < \epsilon$. Now,

$$\begin{aligned}\rho_T(F(x_n, y_n, c_n), &F(x, y, 0) \circ \gamma_n) \\ &\leq \rho_T\left(\left(y_n - y_n^- + H(x_n, y_n, c_n)\right)^\uparrow, \left(\left(y - y^- + H(x, y, 0)\right)^\uparrow\right) \circ \gamma_n\right) \\ &\quad + \rho_T\left(H(x_n, y_n, c_n), (H(x, y, 0)) \circ \gamma_n\right).\end{aligned}$$

For the first term we have

$$\begin{aligned}&\rho_T\left(\left(y_n - y_n^- + H(x_n, y_n, c_n)\right)^\uparrow, \left(\left(y - y^- + H(x, y, 0)\right)^\uparrow\right) \circ \gamma_n\right) \\ &\leq \rho_T\left(y_n - y_n^- + H(x_n, y_n, c_n), \left(y - y^- + H(x, y, 0)\right) \circ \gamma_n\right),\end{aligned} \tag{2.3.3}$$

since $(\cdot)^{\uparrow}$ is lipschitz, γ_n is increasing, and

$$\rho_T\left(y_n - y_n^- + H(x_n, y_n, c_n), \left(y - y^- + H(x, y, 0)\right) \circ \gamma_n\right)$$
$$\leq \rho_T\left(y_n, y \circ \gamma_n\right) + \rho_T\left(y_n^-, y^- \circ \gamma_n\right) + \rho_T\left(H(x_n, y_n, c_n), H(x, y, 0) \circ \gamma_n\right). \quad (2.3.4)$$

Since γ_n is strictly increasing,

$$\rho_T\left(y_n^-, y^- \circ \gamma_n\right) = \sup_{0 \leq t \leq T} \left| \lim_{s \nearrow t} y_n(s) - \lim_{r \nearrow \gamma_n(t)} y(r) \right|$$
$$= \sup_{0 \leq t \leq T} \left| \lim_{s \nearrow t} y_n(s) - \lim_{r \nearrow t} y(\gamma_n(r)) \right|,$$

and so

$$\rho_T\left(y_n^-, y^- \circ \gamma_n\right) \leq \sup_{0 \leq t \leq T} \left| y_n(t) - y(\gamma_n(t)) \right|,$$

since the left limit of y_n and $y \circ \gamma_n$ exist at each t. Therefore,

$$\rho_T\left(y_n^-, y^- \circ \gamma_n\right) \leq \rho_T\left(y_n, y \circ \gamma_n\right) \quad (2.3.5)$$

Combining (2.3, 2.3.3, 2.3.4, 2.3.5) we have,

$$\rho_T(F(x_n, y_n, c_n), F(x, y, 0) \circ \gamma_n)$$
$$\leq \rho_T\left(\left(y_n - y_n^- + H(x_n, y_n, c_n)\right)^{\uparrow}, \left(\left(y - y^- + H(x, y, 0)\right)^{\uparrow}\right) \circ \gamma_n\right)$$
$$+ \rho_T\left(H(x_n, y_n, c_n), H(x, y, 0) \circ \gamma_n\right)$$
$$\leq 2\rho_T\left(y_n, y \circ \gamma_n\right) + 2\rho_T\left(H(x_n, y_n, c_n), H(x, y, 0) \circ \gamma_n\right)$$
$$\leq 12\epsilon,$$

by lemma 2.3.4. ■

2.4 Scaling limit of the plateau process

In this section we prove several results concerning the sequence of models, and then combine these to prove theorem 2.1.1. We begin by showing that the function H scales nicely when no centering is required.

Lemma 2.4.1. *For positive constants a_n and n,*

$$a_n^{-1} H(x, y, c)(nt) = H(x^n, y^n, c/n)(t),$$

for all $t \geq 0$, where $x^n(t) = a_n^{-1} x(nt)$ and $y^n(t) = a_n^{-1} y(nt)$.

Proof. By definition,

$$\begin{aligned}
a_n^{-1} H(x, y, c)(nt) &= a_n^{-1} \sup_{0 \leq s \leq nt} \big(x(s) - y([s-c]^+) \big) \\
&= \sup_{0 \leq s \leq t} \big(a_n^{-1} x(ns) - a_n^{-1} y([ns - c]^+) \big) \\
&= \sup_{0 \leq s \leq t} \big(a_n^{-1} x(ns) - a_n^{-1} y(n\,[s - c/n]^+) \big) \\
&= \sup_{0 \leq s \leq t} \big(x^n(s) - y^n([s - c/n]^+) \big) \\
&= H(x^n, y^n, c/n)(t)
\end{aligned}$$

■

Lemma 2.4.2. *The set $\mathscr{K} = \{x \in \mathbb{D} : x(t) - x(t-) \geq 0$ for each $t \in (0, \infty)\}$ is closed in $\mathbb{D}$.*

Proof. Let $\{x_n\}$ be a sequence in $\mathscr{K}$ such that $x_n \to x$. Fix $t_0 \in (0, \infty)$ with $x(t_0) - x(t_0-) \neq 0$. There exists $t_n \to t_0$ with $x_n(t_n) - x_n(t_n-) \to x(t_0) - x(t_0-)$ by

[9] proposition VI.2.1. We have $x_n(t_n) - x_n(t_n-) \geq 0$ for each n since $x_n \in \mathscr{K}$, so $x(t_0) - x(t_0-) \geq 0$ and we must have $x \in \mathscr{K}$. ■

The next Lemma establishes a joint convergence involving the primitive input processes. Recall that $\check{U}^r \Rightarrow U^*$ and $\check{V}^r \Rightarrow V^*$ in $\mathbb{D}$.

Lemma 2.4.3. *For any sequence of real numbers $c_r \to c$,*

$$(\check{U}^r + c_r e, \check{V}^r, 1/r) \Rightarrow (U^* + ce, V^*, 0),$$

in the product topology on $\mathbb{D} \times \mathbb{D} \times \mathbb{R}$. Moreover,

$$Disc(U^* + ce) \cap Disc(V^*) = \varnothing \ a.s.$$

and $\{t : V^*(t) - V^*(t-) < 0\} = \varnothing a.s.$

Proof. Since ce is continuous, $\check{U}^r \Rightarrow U^*$, and $c_r e \Rightarrow ce$ we have $\check{U}^r + c_r e \Rightarrow U^* + ce$ by [11]. We have joint convergence $(\check{U}^r + c_r e, \check{V}^r) \Rightarrow (U^* + ce, V^*)$ since $\check{V}^r$ is independent of $\check{U}^r$ and therefore $\check{U}^r + c_r e$ is independent of $\check{V}^r$ because c_r is constant in ω, [12] Theorem 11.4.4, moreover U^* is independent of V^*. Since $1/r$ is constant in ω we have $1/r \to 0$ in probability so [2] Theorem 4.4 gives joint convergence

$$(\check{V}^r + c_r e, \check{U}^r, 1/n) \Rightarrow (U^* + ce, V^*, 0).$$

V^* is a stable Lévy motion by 2.4.1 of the online supplement to [12]. So V^* has no fixed discontinuities: $\mathbb{P}\{U^*(t) = U^*(t-)\} = 1$ for all $t \in (0, \infty)$. By [11] Lemma 4.3, gives $\mathbb{P}\{\text{Disc}(U^*) \cap \text{Disc}(V^*) = \varnothing\} = 1$ and since ce is continuous we have

$$\mathbb{P}\{\text{Disc}(U^* + ce) \cap \text{Disc}(V^*) = \varnothing\} = 1.$$

Finally, $\mathbb{P}\left\{\check{V}^r \in \mathscr{K}\right\} = 1$, $\check{V}^r \Rightarrow V^*$, and $\mathscr{K}$ is closed by Lemma 2.4.2, so the Portmanteau theorem gives

$$\mathbb{P}\left\{V^* \in \mathscr{K}\right\} \geq \limsup_{n\to\infty} \mathbb{P}\left\{\check{V}^r \in \mathscr{K}\right\} = 1.$$

■

For each $r > 0$ and $t \geq 0$ define $\bar{D}^r(t) = \frac{1}{r}D^r(rt)$. Using Corollary 2.2.4 under this fluid scaling, we have for all $t \geq 0$,

$$\bar{R}^r(t) = \frac{1}{r}R^r(rt).$$

We will need the fluid limit of $\bar{D}^r(\cdot)$.

Lemma 2.4.4. *As $r \to \infty$,*

$$\bar{R}^r \Rightarrow e/\mu$$

Proof. $\check{U}^r(1) \Rightarrow U^*(1)$ implies $\frac{r}{a_r}\left(\bar{U}^r(1) - \mu^r\right) \Rightarrow U^*(1)$, but $r/a_r \to \infty$ implies $\bar{U}^r(1) - \mu_r \Rightarrow 0$. Since $\mu^r \to \mu$ we have $\bar{U}^r(1) \Rightarrow \mu$. By Theorem 2.4.1 of the internet supplement to [12], we have $\bar{U}^r \Rightarrow \mu e$ in $\mathbb{D}$. Similarly, $\bar{V}^r \Rightarrow \mu e$ in $\mathbb{D}$. Now compute

$$\begin{aligned}
\bar{R}^r(t) &= \frac{1}{r}\sup\{m \geq 0 : V^r(m) + H(U^r, V^r, 1)(m) \leq rt\} \\
&= \sup\{x/r \geq 0 : V^r(x) + H(U^r, V^r, 1)(x) \leq rt\} \\
&= \sup\left\{x/r \geq 0 : \frac{V^r(x)}{r} + \frac{1}{r}H(U^r, V^r, 1)(x) \leq t\right\} \\
&= \sup\left\{y \geq 0 : \frac{V^r(ry)}{r} + \frac{1}{r}H(U^r, V^r, 1)(ry) \leq t\right\} \\
&= \sup\left\{y \geq 0 : \bar{V}^r(y) + H(\bar{U}^r, \bar{V}^r, 1/r)(y) \leq t\right\},
\end{aligned}$$

by lemma 2.4.1. We have $(\bar{U}^r, \bar{V}^r, 1/r) \Rightarrow (\mu e, \mu e, 0)$ in $\mathbb{D}$ since the processes are independent. The function H is continuous at $(\mu_u e, \mu_v e, 0)$, and addition is continuous at continuous elements of $\mathbb{D}$, so

$$\bar{V}^r + H(\bar{U}^r, \bar{V}^r, 1/r) \Rightarrow \mu e$$

in $\mathbb{D}$. The result follows because μe is in the set of continuity for the function $x \mapsto \sup\{y \geq 0 : x(y) \leq t\}$ by Corollary 13.6.4 in [12]. ■

We now prove the main result.

Proof of Theorem 2.1.1. By Lemma 2.2.5

$$M(t) = F(U^r, V^r, 1)(R^r(t)).$$

Under fluid scaling $\bar{R}^r \Rightarrow e/\mu$ by 2.4.4. We first consider the scaling limit for F, before composing with R^r.

$$\begin{aligned}
a_r^{-1} F(U^r, V^r, 1)(rt) &= a_r^{-1} \sup_{0 \leq s \leq rt} (V^r(s) - V^r(s-) + H(U^r, V^r, 1)(s)) \\
&\quad - a_r^{-1} H(U^r, V^r, 1)(rt) \\
&= \sup_{0 \leq s \leq rt} \left(a_r^{-1} V^r(s) - a_r^{-1} V^r(s-) + a_r^{-1} H(U^r, V^r, 1)(s)\right) \\
&\quad - a_r^{-1} H(U^r, V^r, 1)(rt) \\
&= \sup_{0 \leq s \leq t} \left(a_r^{-1} V^r(rs) - a_r^{-1} V^r(rs-) + a_r^{-1} H(U^r, V^r, 1)(rs)\right) \\
&\quad - a_r^{-1} H(U^r, V^r, 1)(rt).
\end{aligned}$$

$t \mapsto r\nu^r t$ is continuous so $r\nu^r(rs) - r\nu^r(rs-) = 0$ and

$$\begin{aligned} a_r^{-1} F(U^r, V^r, 1)(rt) = \sup_{0 \le s \le t} \big(\check{V}^r(s) - \check{V}^r(s-) + a_r^{-1} H(U^r, V^r, 1)(rs)\big) \\ - a_r^{-1} H(U^r, V^r, 1)(rt). \end{aligned} \tag{2.4.1}$$

Now, we address the idleness part of (2.4.1) that occurs twice.

$$\begin{aligned}
& a_r^{-1} H(U^r, V^r, 1)(rt) \\
&\quad = a_r^{-1} \sup_{0 \le s \le rt} \Bigg(U^r(s) - V^r([s-1]^+) \Bigg) \\
&\quad = \sup_{0 \le s \le t} \Bigg(a_r^{-1} U^r(rs) - a_r^{-1} V^r(r[s-1/r]^+) \Bigg) \\
&\quad = \sup_{0 \le s \le t} \Bigg(a_r^{-1} \left(U^r(rs) - r\mu^r s\right) + a_r^{-1} r\mu^r s \\
&\qquad\quad - a_r^{-1} \left(V^r(r[s-1/r]^+) - r\nu^r[s-1/r]^+\right) - a_r^{-1} r\nu^r[s-1/r]^+ \Bigg) \\
&\quad = \sup_{0 \le s \le t} \Bigg(\check{U}^r(s) + a_r^{-1} r\mu^r s - \check{V}^r([s-1/r]^+) - a_r^{-1} r\nu^r[s-1/r]^+ \Bigg) \\
&\quad = \sup_{0 \le s \le t} \Bigg(\check{U}^r(s) + a_r^{-1} r(\mu^r - \nu^r)s + a_r^{-1} r\nu^r(s - [s-1/r]^+) \\
&\qquad\quad - \check{V}^r([s-1/r]^+) \Bigg).
\end{aligned}$$

Since

$$a_r^{-1} r\nu^r(s - [s-1/r]^+) = a_r^{-1} r\nu^r(1/r \wedge s) = a_r^{-1} \nu^r(1 \wedge rs),$$

we have

$$\begin{aligned}
& a_r^{-1} H(U^r, V^r, 1)(rt) \\
&\qquad = H(\check{U}^r + a_r^{-1} r(\mu^r - \nu^r)e + a_r^{-1}\nu^r(1 \wedge re), \check{V}^r, 1/r)(t).
\end{aligned}$$

Putting this expression back into (2.4.1),

$$a_r^{-1}F(U^r, V^r, 1)(rt) = \sup_{0\le s\le t}\big[\check{V}^r(s) - \check{V}^r(s-)$$

$$+H(\check{U}^r + a_r^{-1}r(\mu^r - \nu^r)e + a_r^{-1}\nu^r(1\wedge re), \check{V}^r, 1/r)(s)\big]$$

$$- H(\check{U}^r + a_r^{-1}r(\mu^r - \nu^r)e + a_r^{-1}\nu^r(1\wedge re), \check{V}^r, 1/r)(t)$$

$$= F(\check{U}^r + a_r^{-1}r(\mu^r - \nu^r)e + a_r^{-1}\nu^r(1\wedge re), \check{V}^r, 1/r)(t).$$

By Lemma 2.4.3 we have $(U^* + \gamma\mu e, V^*, 0)$ satisfies the continuity criterion of Lemma 2.3.5. By the continuous mapping theorem

$$F(\check{U}^r + a_r^{-1}r(\mu^r - \nu^r)e + a_r^{-1}\nu^r(1\wedge re), \check{V}^r, 1/r) \Rightarrow F(U^* + \gamma\mu e, V^*, 0).$$

Finally, the scaled plateau process is a composition of F with R^r,

$$a_r^{-1}F(U^r, V^r, 1)(R^r(rt)) = a_r^{-1}F(U^r, V^r, 1)(r\bar{R}^r(t)).$$

Composition is continuous on $(\mathbb{D}\times C_0)$ by [11] Theorem 3.1, where $C_0 \subset \mathbb{D}$ denotes the strictly increasing, continuous functions. So the continuous mapping theorem yields

$$a_r^{-1}M^r(r\cdot) = \check{M}^r \Rightarrow M^* = F(U^* + \gamma\mu e, V^*, 0)(\cdot/\mu).$$

■

Chapter 3

The distribution of successive maxima of $M(\cdot)$

The plateau process only decreases when the first job in a Q1 busy period arrives to Q2. This suggests that studying the plateau process on this subsequence of points will reduce the complexity of the model. In fact, the value of the process at these points is closely related to the largest job in a busy period [4]. In this chapter we first study some general properties of these special points, then restrict our attention to Poisson arrivals and heavy tailed service times. Throughout this chapter we assume that $\rho \leq 1$ so that the largest job in a busy period is finite with probability 1. There are two key advantages of working with Poisson arrivals exploited in what follows. The first is that idle periods have the same distribution as interarrival times. The second advantage is the expected number of jobs in a busy period $\frac{1}{1-\rho}$ is given by the traffic intensity ρ. In this setting we find the one dimensional distribution of the workload in the second queue at these greatest values in terms of κ which is defined implicitly as a function of the primitives of the tandem queueing system.

3.1 Notation

In this chapter we return to the standard notation where capital letter are used for random variables. Let $\{V_i\}_{i=1}^{\infty}$ be a sequence of independent nonnegative random variables with common distribution function F. We do not superscript V_i because we choose the make V_i the same for each n. For $n = 1, 2, \ldots$ let $\{U_i^{(n)}\}_{i=1}^{\infty}$ be a sequence of independent exponential random variables with parameter $\lambda^{(n)}$ independent of the collection $\{V_i\}_{i=1}^{\infty}$. For each $n > 0$ let $M_i^{(n)}$ be the largest service time in the ith busy period. For example $M_1^{(n)} = \sup\{V_j : 1 \leq j \leq N$ where N is the smallest integer $\geq$ 1 such that $\sum_{j=1}^{N} \left(V_j - U_j^{(n)}\right) \leq 0\}$. Let $m^{(n)}$ be the distribution function for $M_1^{(n)}$. For $\rho^{(n)} = \lambda^{(n)}\mathbb{E}\left[V_1\right] \leq 1$, we have $\{M_i^{(n)}\}_{i=1}^{\infty}$ is a collection of proper random variables with distribution function m. We say a nonnegative function f is regularly varying with parameter ν if

$$\lim_{x\to\infty} f(\lambda x)/f(x) = \lambda^{\nu}$$

for each $\lambda > 0$, and a random variable V is a regularly varying with parameter ν if $x \mapsto \mathbb{P}\left\{V > x\right\}$ is regularly varying with parameter $-\nu$. Note that if a nonnegative random variable V is regularly varying with parameter ν then $\mathbb{E}\left[|V|^{\gamma}\right] < \infty$ if and only if $\gamma < \nu$. For a distribution function $F(x) = \mathbb{P}\left\{V \leq x\right\}$ we write $\bar{F}(x) = 1 - F(x)$. Let $\tilde{t}_n^{(n)}$ be the **transfer time** of the last customer from the nth busy period in the first queue in the nth system.

3.2 The largest job in a busy period

In this section we drop the superscript (n) because the results are true for each such system. In this section we show how the tail behavior of M the largest job in a busy period is related to the tail behavior of V the service times. First we state an important representation theorem by Boxma.

The following is a summary of a description of m a the solution to an equation found in [3].

Proposition 3.2.1. *Let $w > 0$, then*

$$m(w) = \int_0^w e^{-\lambda t(1-m(w))} dF(t).$$

Proof. The key step to the recursive formula is to write $m(w)$ in terms of the size of the first job to arrive V and an independent collection of maximum job sizes during a busy period $\{M^{(i)}\}_{i\in\mathbb{N}}$. For notation let $M^{(0)} = 0$. If the size of the first job V is t then there are $N(t)$ interruptions to work on V, during each interruption the probability that every job in the interruption is less than w is $m(w)$, before the queue begins to idle for the first time since V has arrived.

$$m(w) = P(V \leq w; \max_{i=0}^{N(V)} M^{(i)} \leq w).$$

Then conditioning on the size of the first job we have

$$\begin{aligned} m(w) &= \int_0^w \sum_{n=0}^{\infty} P(N(t)=n)P(\max_{i=0}^{n} M^{(i)} \leq w)dF(t) \\ &= \int_0^w e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}(m(w))^n dF(t) \\ &= \int_0^w e^{-\lambda t}e^{\lambda t m(w)}dF(t) \\ &= \int_0^w e^{-\lambda t(1-m(w))}dF(t) \end{aligned}$$

■

Let $\tau = \inf\{N \geq 1 : \sum_{n=1}^{N} U_{n+1} - V_n > 0\}$ so that the event

$$\{\tau = N\} = \left\{\max_{j=1}^{N-1}\sum_{k=1}^{j} U_{k+1} - V_k \leq 0 \text{ and } \sum_{k=1}^{N} U_{k+1} - V_k > 0\right\}.$$

Since τ is the hitting time of an integer valued stochastic process we have τ is a proper random variable if $\mathbb{E}[U-V] \geq 0$, and $\mathbb{E}[\tau] < \infty$ if $\mathbb{E}[U-V] > 0$ by Theorem XII.2.2 in [6]. In fact $\log \mathbb{E}[\tau] = \sum_{n=1}^{\infty}\frac{1}{n}\mathbb{P}\{\sum_{k=1}^{n} U_k - V_k \leq 0\}$ by [6] XII Theorem 3. When working in M/G/1 this expectation simplifies to $1/(1-\rho)$.

Lemma 3.2.2. *Suppose τ is a proper random variable and $x < x^* = \sup\{y : F(y) < 1\}$, then*

$$\frac{\bar{m}(x)}{\bar{F}(x)} = \sum_{k=1}^{\infty}\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau \geq k\right\}$$

Proof. The event

$$\{\tau \geq k\} = \left\{\max_{j=1}^{k-1}\sum_{i=1}^{j} U_{i+1} - V_i \leq 0\right\}$$

is independent of the event $\{V_k > x\}$. For each $N = 1, \ldots$ we may write $\{\max_{k=1}^{N} V_k > x\}$ as a disjoint union,

$$\left\{\max_{k=1}^{N} V_k > x\right\} = \bigsqcup_{k=1}^{N} \left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } V_k > x\right\},$$

where $\max_{j=1}^{0} V_j = -\infty$, so that $\left\{\max_{j=1}^{0} V_j \leq x \text{ and } V_1 > x\right\} = \{V_1 > x\}$. For x such that $0 < F(x) < 1$.

$$\begin{aligned}
\frac{\bar{m}(x)}{\bar{F}(x)} &= \frac{\mathbb{P}\left\{\max_{k=1}^{\tau} V_k > x\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{N=1}^{\infty} \frac{\mathbb{P}\left\{\max_{k=1}^{N} V_k > x \text{ and } \tau = N\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{N=1}^{\infty} \sum_{k=1}^{N} \frac{\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } V_k > x \text{ and } \tau = N\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{k=1}^{\infty} \sum_{N=k}^{\infty} \frac{\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } V_k > x \text{ and } \tau = N\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{k=1}^{\infty} \frac{\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } V_k > x \text{ and } \tau \geq k\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{k=1}^{\infty} \frac{\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau \geq k\right\} \mathbb{P}\left\{V_k > x\right\}}{\mathbb{P}\left\{V_1 > x\right\}} \\
&= \sum_{k=1}^{\infty} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau \geq k\right\}.
\end{aligned}$$

∎

Corollary 3.2.3. *Under the conditions of lemma 3.2.2 and* $\lim_{x\uparrow x^*} F(x) = 1$*.*

$$\lim_{x\uparrow x^*} \frac{\bar{m}(x)}{\bar{F}(x)} = \mathbb{E}\left[\tau\right].$$

Proof.

$$\lim_{x\uparrow x^*} \frac{\bar{m}(x)}{\bar{F}(x)} = \lim_{x\uparrow x^*} \sum_{k=1}^{\infty} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau \geq k\right\} = \sum_{k=1}^{\infty} \mathbb{P}\left\{\tau \geq k\right\}$$

by the monotone convergence theorem. The tail sum formula

$$\sum_{k=1}^{\infty} \mathbb{P}\left\{\tau \geq k\right\} = \mathbb{E}\left[\tau\right]$$

yields $\lim_{x\uparrow x^*} \frac{\bar{m}(x)}{\bar{F}(x)} = \mathbb{E}\left[\tau\right].$ ■

Corollary 3.2.4. *If $\mathbb{E}\left[\tau\right] < \infty$ and $\bar{F}$ is regularly varying at infinity with parameter $-\nu$, then $\bar{m}$ is regularly varying at infinity with parameter $-\nu$.*

Proof. By corollary 3.2.3

$$\lim_{x\to\infty} \frac{\bar{m}(yx)}{\bar{m}(x)} = \lim_{x\to\infty} \frac{\bar{m}(yx)}{\bar{F}(yx)} \frac{\bar{F}(x)}{\bar{m}(x)} \frac{\bar{F}(yx)}{\bar{F}(x)} = \frac{\mathbb{E}\left[\tau\right]}{\mathbb{E}\left[\tau\right]} y^{-\nu},$$

since $\bar{F}$ is regularly varying with parameter $-\nu$. ■

Clearly, the maximum increases as the set that the maximum is taken over increases. Indeed, as $U^{(n)}$ stochastically decreases, we have $M^{(n)}$ stochastically increases, $m^{(n)}$ decreases, and $\bar{m}^{(n)}$ increases.

Lemma 3.2.5. *Suppose for each $w \geq 0$, $\mathbb{P}\left\{U^{(1)} \leq w\right\} \leq \mathbb{P}\left\{U^{(2)} \leq w\right\}$, $\mathbb{P}\left\{U^{(1)} \leq 0\right\} = \mathbb{P}\left\{U^{(2)} \leq 0\right\} = 0$, $\tau^{(n)}$ are proper for $n = 1, 2$, and these distributions are continuous then $\bar{m}^{(1)}(w) \leq \bar{m}^{(2)}(w)$.*

Proof. By lemma 3.2.2 it suffices to show

$$\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau^{(1)} \geq k\right\} \leq \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau^{(2)} \geq k\right\}, \tag{3.2.1}$$

for each $k \geq 1$. When $k = 1$, $\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \tau^{(n)} \geq k\right\} = 1$ for $n = 1, 2$. Fix $y \geq 0$ we have

$$\begin{aligned}
\mathbb{P}\left\{V_1 \leq x \text{ and } U_2^{(1)} - V_1 \leq y\right\} &= \int_{w=0}^{x} \mathbb{P}\left\{U_2^{(1)} \leq w + y\right\} \mathbb{P}\left\{V \leq dw\right\} \\
&\leq \int_{w=0}^{x} \mathbb{P}\left\{U_2^{(2)} \leq w + y\right\} \mathbb{P}\left\{V \leq dw\right\} \\
&= \mathbb{P}\left\{V_1 \leq x \text{ and } U_2^{(2)} - V_1 \leq y\right\}.
\end{aligned}$$

Note that when $y = 0$ this implies equation (3.2.1) holds for $k = 2$. Now suppose for each $y \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(1)} - V_i \leq y\right\} \\
&\qquad \leq \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \leq y\right\}. \qquad (3.2.2)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}\left\{\max_{j=1}^{k} V_j \leq x \text{ and } \max_{j=1}^{k} \sum_{i=1}^{j} U_{i+1}^{(1)} - V_i \leq y\right\} \\
&\qquad - \mathbb{P}\left\{\max_{j=1}^{k} V_j \leq x \text{ and } \max_{j=1}^{k} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \leq y\right\} \\
&= \int_{w=0}^{x} \int_{z=0}^{w} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(1)} - V_i \leq y + w - z\right\} \mathbb{P}\left\{U^{(1)} \leq dz\right\} \\
&\qquad - \int_{z=0}^{w} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \leq y + w - z\right\} \\
&\qquad\qquad \mathbb{P}\left\{U^{(2)} \leq dz\right\} \mathbb{P}\left\{V \leq dw\right\} \\
&\leq \int_{w=0}^{x} \int_{z=0}^{w} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \leq y + w - z\right\} \mathbb{P}\left\{U^{(1)} \leq dz\right\} \\
&\qquad - \int_{z=0}^{w} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \leq x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \leq y + w - z\right\}
\end{aligned}$$

$$\mathbb{P}\left\{U^{(2)} \le dz\right\} \mathbb{P}\left\{V \le dw\right\}$$

$$= \int_{w=0}^{x} \mathbb{P}\left\{U^{(1)} \le w\right\} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y\right\}$$

$$- \int_{z=0}^{w} \mathbb{P}\left\{U^{(1)} \le z\right\} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y + w - dz\right\}$$

$$- \mathbb{P}\left\{U^{(2)} \le w\right\} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y\right\}$$

$$+ \int_{z=0}^{w} \mathbb{P}\left\{U^{(2)} \le z\right\} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y + w - dz\right\}$$

$$\mathbb{P}\left\{V \le dw\right\}$$

$$= \int_{w=0}^{x} \left(\mathbb{P}\left\{U^{(1)} \le w\right\} - \mathbb{P}\left\{U^{(2)} \le w\right\}\right)$$

$$\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y\right\}$$

$$+ \int_{z=0}^{w} \left(\mathbb{P}\left\{U^{(2)} \le z\right\} - \mathbb{P}\left\{U^{(1)} \le z\right\}\right)$$

$$\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y + w - dz\right\} \mathbb{P}\left\{V \le dw\right\},$$

by integration by parts and $\mathbb{P}\left\{U^{(1)} \le 0\right\} = \mathbb{P}\left\{U^{(2)} \le 0\right\} = 0$.

$$\left(\mathbb{P}\left\{U^{(1)} \le w\right\} - \mathbb{P}\left\{U^{(2)} \le w\right\}\right) \le 0$$

and

$$z \mapsto \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y + w - z\right\}$$

is decreasing imply

$$\int_{z=0}^{w} \left(\mathbb{P}\left\{U^{(2)} \le z\right\} - \mathbb{P}\left\{U^{(1)} \le z\right\}\right)$$

$$\mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \max_{j=1}^{k-1} \sum_{i=1}^{j} U_{i+1}^{(2)} - V_i \le y + w - dz\right\} \le 0$$

for each $w \ge 0$. Thus, equation (3.2.2) holds for k implies equation (3.2.2) holds for $k+1$. Since equation (3.2.2) reduces to equation (3.2.1) when $y = 0$, we have equation (3.2.1) holds for all $k \ge 1$ and we have

$$\begin{aligned}\bar{m}^{(1)}(x) &= \bar{F}(x) \sum_{k=1}^{\infty} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \tau^{(1)} \ge k\right\} \\ &\le \bar{F}(x) \sum_{k=1}^{\infty} \mathbb{P}\left\{\max_{j=1}^{k-1} V_j \le x \text{ and } \tau^{(2)} \ge k\right\} = \bar{m}^{(2)}(x).\end{aligned}$$

■

3.3 Triangular array and Poisson arrivals

Now we specialize to the $M/G/1$ queue. We will also need to generalize to the triangular array setup in order to see nonzero idleness as in the notation section.

Recall, $m^{(n)}(x) = \mathbb{P}\left\{\max_{i=1}^{\tau^{(n)}} V_i \le x\right\} = \mathbb{P}\left\{M^{(n)} \le x\right\}$.

The following lemma uses a Tauberian theorem.

Lemma 3.3.1. *Let* $\rho^{(n)} = \lambda^{(n)}\mathbb{E}[V]$ *for* $\beta \ge 0$. *Assume* $n^{\nu-1}\left(\frac{1-\rho^{(n)}}{l(n)}\right) \to \gamma$, *and* $1 - F(t) = \left(\frac{-1}{\Gamma(1-\nu)}\right) t^{-\nu} l(t)$ *for* $1 < \nu < 2$ *and* l *a slowly varying function. Fix* $y > 0$. *Then,*

$$\lim_{n\to\infty} \frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V} \middle| V > ny\right] - \frac{\bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}}{l(ny)}}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu}\left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right)} = 1.$$

Proof. Since arrivals are Poisson we have $m^{(n)}(ny) = \int_0^{ny} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t)$ when $\rho^{(n)} \leq 1$ by [3].

$$\bar{m}^{(n)}(ny) = 1 - \int_0^{\infty} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t) + \int_{ny}^{\infty} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t). \tag{3.3.1}$$

Fix $y > 0$,

$$\begin{aligned}
\int_{ny}^{\infty} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t) &= \int_0^{\infty} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} 1_{(ny,\infty)}(t) dF(t) \\
&= \mathbb{E}\left[e^{-\lambda^{(n)} V \bar{m}^{(n)}(ny)} 1_{(ny,\infty)}(V) \right] \\
&= \mathbb{P}\left\{ V > ny \right\} \mathbb{E}\left[e^{-\lambda^{(n)} \bar{m}^{(n)}(ny) V} \middle| V > ny \right] \\
&= \left(\frac{-1}{\Gamma(1-\nu)} \right) (ny)^{-\nu} l(ny) \mathbb{E}\left[e^{-\lambda^{(n)} \bar{m}^{(n)}(ny) V} \middle| V > ny \right]
\end{aligned}$$

Plug into equation (3.3.1),

$$\begin{aligned}
\bar{m}^{(n)}(ny) = 1 - \int_0^{\infty} & e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t) \\
& + \left(\frac{-1}{\Gamma(1-\nu)} \right) (ny)^{-\nu} l(ny) \mathbb{E}\left[e^{-\lambda^{(n)} \bar{m}^{(n)}(ny) V} \middle| V > ny \right].
\end{aligned}$$

Since $\lambda^{(n)} \mathbb{E}[V] = \rho^{(n)}$ we have

$$\begin{aligned}
&\int_0^{\infty} e^{-\lambda^{(n)} t \bar{m}^{(n)}(ny)} dF(t) - 1 + \lambda^{(n)} \bar{m}^{(n)}(ny) \mathbb{E}[V] \\
&= \left(\frac{-1}{\Gamma(1-\nu)} \right) (ny)^{-\nu} l(ny) \mathbb{E}\left[e^{-\lambda^{(n)} \bar{m}^{(n)}(ny) V} \middle| V > ny \right] - \bar{m}^{(n)}(ny)(1 - \rho^{(n)}). \quad (3.3.2)
\end{aligned}$$

Then dividing by $\left(\lambda^{(n)}\bar{m}^{(n)}(ny)\right)^{\nu} l(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)})$ we have

$$\frac{\int_0^\infty e^{-\lambda^{(n)}t\bar{m}^{(n)}(ny)}dF(t) - 1 + \lambda^{(n)}\bar{m}^{(n)}(ny)\mathbb{E}\left[V\right]}{\left(\lambda^{(n)}\bar{m}^{(n)}(ny)\right)^{\nu} l(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)})}$$
$$= \frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)(ny)^{-\nu}l(ny)\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V}\middle| V > ny\right] - \bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}}{\left(\lambda^{(n)}ny\bar{m}^{(n)}(ny)\right)^{\nu} l(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)})}. \tag{3.3.3}$$

The limit as $n \to \infty$ on the left hand side is 1 by [10] Theorem 8.1.6. To justify the use of theorem 8.1.6 we note the left hand side of equation (3.3.3) is, in the notation of used in Theorem 8.1.6, $(\hat{F}(s) - 1 + s\mathbb{E}\left[V\right])/(s^{\nu}l(1/s))$. So, we have $\bar{F}(x) \sim -1/\Gamma(1-\nu)x^{-\alpha}l(x)$ is equivalent to $(\hat{F}(s)-1+s\mathbb{E}\left[V\right])/(s^{\nu}l(1/s)) \to 1$ where $1 < \nu < 2$ and $s = s(n) = \lambda^{(n)}\bar{m}^{(n)}(ny)$. Since $\lambda^{(n)} \uparrow \lambda < \infty$, $\bar{m}^{(n)}(ny) \uparrow \bar{m}^{(\infty)}(ny)$ by lemma 3.2.5, $m^{(\infty)}$ is a proper probability distribution yields $s \leq \lambda\bar{m}^{(\infty)}(ny) \downarrow 0$ as $n \to \infty$.

■

3.3.1 Bounds for $ny\bar{m}^{(n)}(ny)$

For each $y > 0$, we need to show that $ny\bar{m}^{(n)}(ny)$ converges to something bounded away from 0 and infinity. The following lemmas provide these bounds.

Lemma 3.3.2. *Under the conditions of lemma 3.3.1,*

$$\limsup_{n\to\infty}(ny)\bar{m}^{(n)}(ny) \leq \max\left[2^{2/\nu}\mathbb{E}\left[V\right], 1\right].$$

Proof. If $\lambda^{(n)}(ny)\bar{m}(ny) \geq 1$, we take $A = 2$ and $\delta = \nu/2$ in Potter's Theorem [10] 1.5.6 so that for n sufficiently large

$$(1/2)\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{-\nu/2} \leq \left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right).$$

$\bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}$ and $\frac{-1}{\Gamma(1-\nu)}$ are nonnegative and $l(ny)$ is eventually positive, so

$$\begin{aligned}
&\frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V}\middle|V > ny\right] - \frac{\bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}}{l(ny)}}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu}\left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right)} \\
&\leq \frac{1}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu}\left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right)} \\
&\leq \frac{1}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu}\left((1/2)\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{-\nu/2}\right)} \\
&= \frac{2}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu/2}}.
\end{aligned}$$

Lemma 3.3.1 gives

$$\liminf_{n\to\infty} \frac{2}{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\nu/2}} \geq 1,$$

when $\limsup_{n\to\infty} \lambda^{(n)}(ny)\bar{m}(ny) \geq 1$. Since $\lambda^{(n)} \to 1/\mathbb{E}[V]$ we have

$$\limsup_{n\to\infty} ny\bar{m}^{(n)}(ny) \leq \max\left[2^{2/\nu}\mathbb{E}[V], 1\right].$$

■

To establish the lower bound, we first find the distribution that is the smallest among $\bar{m}^{(n)}$ is bounded above zero, that is the case $\mathbb{E}[U_1] = \mathbb{E}[V_1]$.

Lemma 3.3.3. *Let $\{U_i\}$ independent exponential with parameter λ independent of $\{V_i\}$ independent with distribution F, where $1-F$ is regularly varying with parameter $1<\nu<2$. Assume $\mathbb{E}[U_1] \geq \mathbb{E}[V_1]$, then*

$$\liminf_{w\to\infty} w\bar{m}(w) \geq \sqrt{\frac{4-2\nu}{\nu\lambda^2}}.$$

Proof. From Boxma's equation for exponential interarrival times (intensity $\lambda = \mathbb{E}[U]^{-1}$) we have the distribution of the largest job in a busy period is

$$\begin{aligned}
m(w) &= \int_0^w e^{-\lambda t\bar{m}(w)} F(dt) \\
&\leq \int_0^w 1 + \lambda t\bar{m}(w) + \lambda^2 t^2 (\bar{m}(w))^2 F(dt) \\
&= F(w) + \lambda\bar{m}(w)\int_0^w tF(dt) + \frac{1}{2}\lambda^2(\bar{m}(w))^2\int_0^w t^2F(dt),
\end{aligned}$$

because $e^{-u} \leq 1-u+\frac{1}{2}u^2$ when $0 \leq u < \infty$. Recall $\bar{m}(w) = 1-m(w)$,

$$\begin{aligned}
\bar{m}(w) &\geq 1 - F(w) - \lambda\bar{m}(w)\int_0^w tF(dt) - \frac{1}{2}\lambda^2(\bar{m}(w))^2\int_0^w t^2F(dt) \\
&= \bar{F}(w) - \lambda\bar{m}(w)\int_0^w tF(dt) - \lambda\bar{m}(w)\int_w^\infty tF(dt) \\
&\quad + \lambda\bar{m}(w)\int_w^\infty tF(dt) - \frac{1}{2}\lambda^2(\bar{m}(w))^2\int_0^w t^2F(dt) \\
&= \bar{F}(w) - \lambda\bar{m}(w)\mu + \lambda\bar{m}(w)\int_w^\infty tF(dt) - \frac{1}{2}\lambda^2(\bar{m}(w))^2\int_0^w t^2F(dt) \\
&\geq \bar{F}(w) - \lambda\bar{m}(w)\mu - \frac{1}{2}\lambda^2(\bar{m}(w))^2\int_0^w t^2F(dt).
\end{aligned}$$

Thus,

$$\bar{m}(w)\left(1+\rho+\frac{1}{2}\lambda^2\bar{m}(w)\int_0^w t^2F(dt)\right) \geq \bar{F}(w).$$

Since $1 + \rho + \frac{1}{2}\lambda^2 \bar{m}(w) \int_0^w t^2 F(dt) > 0$ we have

$$\begin{aligned} \bar{m}(w) &\geq \frac{\bar{F}(w)}{1 + \rho + \frac{1}{2}\lambda^2 \bar{m}(w) \int_0^w t^2 F(dt)} \\ w^2(\bar{m}(w))^2 &\geq \frac{w^2 \bar{m}(w) \bar{F}(w)}{1 + \rho + \frac{1}{2}\lambda^2 \bar{m}(w) \int_0^w t^2 F(dt)} \\ &\geq \frac{w^2 \bar{m}(w) \bar{F}(w)}{\frac{1}{2}\lambda^2 \bar{m}(w) \int_0^w t^2 F(dt)} \\ &\geq \frac{w^2 \bar{F}(w)}{\frac{1}{2}\lambda^2 \int_0^w t^2 F(dt)}. \end{aligned} \tag{3.3.4}$$

Now assume $\bar{F}$ is regularly varying with parameter $-\nu$ where $1 < \nu < 2$, then $2-\nu > 0$ so Theorem 1.6.4 in [10] gives

$$\lim_{w \to \infty} \frac{\int_0^w t^2 \bar{F}(dt)}{w^2 \bar{F}(w)} = \frac{-\nu}{2-\nu}.$$

So, the right hand side of (3.3.4) converges as $w \to \infty$ and we have

$$\frac{w^2 \bar{F}(w)}{-\frac{1}{2}\lambda^2 \int_0^w t^2 \bar{F}(dt)} \to \frac{2-\nu}{\nu \frac{1}{2}\lambda^2}.$$

Thus,

$$\liminf_{w \to \infty} w \bar{m}(w) \geq \sqrt{\frac{4-2\nu}{\nu \lambda^2}}.$$

∎

Corollary 3.3.4. *Under the conditions of lemma 3.3.1 and* $1 < \nu < 2$ *we have*

$$\liminf_{n \to \infty} ny \bar{m}^{(n)}(ny) \geq \mathbb{E}\left[V\right] \sqrt{\frac{4-2\nu}{\nu}}.$$

Proof. This follows from the assumption $\lambda^{(n)} \to 1/\mathbb{E}\left[V\right]$. ∎

3.4 Properties of κ

In this section we show that $ny\bar{m}^{(n)}(ny)$ converges to $\kappa(y)$ and we describe several properties of $\kappa(y)$ for fixed $1 < \nu < 2$, $\lambda > 0$, and $\gamma \geq 0$.

3.4.1 The equation that describes κ

In this section we reduce the limit in Lemma 3.3.1 using several technical lemmas.

Lemma 3.4.1. *If* $\lim_{n\to\infty} \bar{m}^{(n)}(ny)ny = \kappa$ *for finite, and* $n^{\nu-1}\left(\frac{1-\rho^{(n)}}{l(n)}\right) \to \gamma$ *we have*

$$\lim_{n\to\infty} \frac{\bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}}{l(ny)} = \kappa\gamma y^{\nu-1}.$$

Proof.

$$\begin{aligned}\frac{\bar{m}^{(n)}(ny)(1-\rho^{(n)})(ny)^{\nu}}{l(ny)} &= \left(\bar{m}^{(n)}(ny)ny\right)\left(\frac{n^{\nu-1}(1-\rho^{(n)})}{l(n)}\right)\left(\frac{l(n)}{l(yn)}\right)\left(y^{\nu-1}\right) \\ &\to (\kappa)\,(\gamma)\,(1)\,y^{\nu-1}.\end{aligned}$$

■

We can use Potters bound in the following lemma because $\bar{m}^{(n)}(ny)ny$ is nearly constant for fixed y and large n.

Lemma 3.4.2. *Fix* $y > 0$. *If* $\lim_{n\to\infty} \lambda^{(n)} = \lambda$ *and* $\lim_{n\to\infty} \bar{m}^{(n)}(ny)ny = \kappa$ *for* $0 < \kappa < \infty$ *we have*

$$\lim_{n\to\infty}\left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right) = 1.$$

Proof. By Potter's Theorem [10] Theorem 1.5.6 for any $A > 1$ and any $\delta > 0$ there exists T such that $ny > T$ and $\left(\frac{1}{\lambda^{(n)}\bar{m}(ny)}\right) > T$ implies

$$\left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right) \leq A \max\left\{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\delta}, \left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{-\delta}\right\},$$

$$\left(\frac{l(ny)}{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}\right) \leq A \max\left\{\left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{\delta}, \left(\lambda^{(n)} ny \bar{m}^{(n)}(ny)\right)^{-\delta}\right\}.$$

The right hand side converges as $n \to \infty$ since $\bar{m}^{(n)}(ny)ny \to \kappa$.

$$\limsup_{n\to\infty} \left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right) \leq A \max\left\{(\lambda\kappa)^{\delta}, (\lambda\kappa)^{-\delta}\right\},$$

$$\limsup_{n\to\infty} \left(\frac{l(ny)}{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}\right) \leq A \max\left\{(\lambda\kappa)^{\delta}, (\lambda\kappa)^{-\delta}\right\},$$

for each $\delta > 0$ and $A > 1$. Since $0 < \lambda\kappa < \infty$ we have

$$A \max\left\{(\lambda\kappa)^{\delta}, (\lambda\kappa)^{-\delta}\right\} \to 1,$$

as $\delta \downarrow 0$ and $A \downarrow 1$. Thus, $\lim_{n\to\infty} \left(\frac{l\left(\frac{1}{\lambda^{(n)}\bar{m}^{(n)}(ny)}\right)}{l(ny)}\right) = 1.$ ■

Definition 4. *We say T_ν is a Pareto ν random variable if*

$$\mathbb{P}\left\{T_\nu > x\right\} = \begin{cases} x^{-\nu} & \text{if } x \geq 1 \\ 1 & \text{if } x < 1 \end{cases}.$$

Clearly, T_ν is regularly varying with parameter ν.

Proposition 3.4.3. *Let $\lambda^{(n)} \to \lambda$, fix $y > 0$ and suppose $ny\bar{m}^{(n)}(ny) \to \kappa > 0$, and V is regularly varying with parameter ν then*

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V} \middle| V > ny\right] = \mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right].$$

Proof.

$$\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V}\,\middle|\,V>ny\right]=\int_0^\infty e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)t}1_{(ny,\infty)}(t)\,\frac{F(dt)}{1-F(ny)}$$

Substitute $u=\bar{m}^{(n)}(ny)t$,

$$\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V}\,\middle|\,V>ny\right]=\int_0^\infty e^{-\lambda^{(n)}u}1_{(ny\bar{m}^{(n)}(ny),\infty)}(u)\,\frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)}$$

We have

$$\lim_{n\to\infty}\mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V}\,\middle|\,V>ny\right]=\lim_{n\to\infty}\int_0^\infty e^{-\lambda^{(n)}u}1_{(\kappa,\infty)}(u)\,\frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)}$$

because $e^{-\lambda^{(n)}u}\leq 1$ and

$$\begin{aligned}
\lim_{n\to\infty}\int_0^\infty &\left|1_{(\kappa,\infty)}(u)-1_{(ny\bar{m}^{(n)}(ny),\infty)}(u)\right|\,\frac{F(\frac{du}{\bar{m}^{(n)}(ny)})}{1-F(ny)}\\
&=\lim_{n\to\infty}\left|\frac{1-F\left(\frac{\kappa}{\bar{m}^{(n)}(ny)}\right)}{1-F(ny)}-\frac{1-F\left(\frac{ny\bar{m}^{(n)}(ny)}{\bar{m}^{(n)}(ny)}\right)}{1-F(ny)}\right|\\
&=\lim_{n\to\infty}\left|\frac{1-F\left(\frac{\kappa ny}{ny\bar{m}^{(n)}(ny)}\right)}{1-F(ny)}-\frac{1-F\left(ny\right)}{1-F(ny)}\right|\\
&=\left(\frac{\kappa}{\kappa}\right)^{-\nu}-1=0
\end{aligned}$$

since $1-F$ is regularly varying with parameter $-\nu$.

The measure $\frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)}$ converges weakly to the measure $(du/\kappa)^{-\nu}$ as $n\to\infty$, since

$$\begin{aligned}
\lim_{n\to\infty}\int 1_{(a,b]}&\frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)}\\
&=\lim_{n\to\infty}\frac{F(b/\bar{m}^{(n)}(ny))}{1-F(ny)}-\frac{F(a/\bar{m}^{(n)}(ny))}{1-F(ny)}=\left(\frac{a}{\kappa}\right)^{-\nu}-\left(\frac{b}{\kappa}\right)^{-\nu}.
\end{aligned}$$

For all $\epsilon > 0$ there exists N such that $r > N$ implies $|e^{-\lambda^{(n)}u} - e^{-\lambda u}| < \epsilon$ thus $\lim_{n\to\infty} \int_0^\infty \left| e^{-\lambda^{(n)}u} - e^{-\lambda u} \right| 1_{(\kappa,\infty)}(u) \frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)} = 0$. So,

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V} \middle| V > ny \right] = \lim_{n\to\infty} \int_0^\infty e^{-\lambda u} 1_{(\kappa,\infty)}(u) \frac{F(du/\bar{m}^{(n)}(ny))}{1-F(ny)}.$$

Then weak convergence gives

$$\lim_{n\to\infty} \mathbb{E}\left[e^{-\lambda^{(n)}\bar{m}^{(n)}(ny)V} \middle| V > ny \right] = \kappa^\nu \int_\kappa^\infty e^{-\lambda t} \left(dt\right)^{-\nu}.$$

Then substitute $t = x/\kappa$,

$$\kappa^\nu \int_\kappa^\infty e^{-\lambda t} \left(dt\right)^{-\nu} = \int_1^\infty e^{-\lambda \kappa x} \left(dx\right)^{-\nu} = \mathbb{E}\left[e^{-\lambda \kappa T_\nu} \right].$$

■

The equation that describes $\kappa(y)$ in contained in the following Lemma.

Lemma 3.4.4. *Let T_ν be Pareto ν, $1 < \nu < 2$, $\gamma \geq 0$ and $\lambda > 0$. The equation in the variable $\kappa > 0$*

$$\left(\frac{-1}{\Gamma(1-\nu)} \right) \mathbb{E}\left[e^{-\lambda \kappa T_\nu} \right] - \kappa \gamma y^{\nu-1} = (\lambda \kappa)^\nu \tag{3.4.1}$$

has exactly one solution.

Proof. The left hand side is a strictly decreasing continuous function in κ and the right hand side is strictly increasing continuous function in κ. When $\kappa = 0$ the left hand side is $\left(\frac{-1}{\Gamma(1-\nu)} \right) > 0$ and the right hand side is 0. The left hand side goes to 0 if $\gamma = 0$ and $-\infty$ if $\gamma > 0$ as $\kappa \to \infty$ the right hand side goes to infinity as $\kappa \to \infty$. Thus (3.4.1) has exactly one solution. ■

Finally, we show that $ny\bar{m}^{(n)}(ny)$ converges to $\kappa(y)$.

Proposition 3.4.5. *Under the assumptions of lemma 3.3.1 we have*

$$\lim_{n\to\infty} n\bar{m}^{(n)}(ny) = \kappa/y, \tag{3.4.2}$$

where κ satisfies equation (3.4.1)*. κ is a function of y, λ, ν, and γ.*

Proof. Let $\tilde{\kappa}$ be a limit point of $ny\bar{m}^{(n)}(ny)$. Then $0 < \tilde{\kappa} < \infty$ by corollary 3.3.4 and lemma 3.3.2. Let n_r be a subsequence such that $\lim_{r\to\infty} n_r y\bar{m}^{(n_r)}(n_r y) = \tilde{\kappa}$. By lemma 3.3.1 we have

$$\lim_{r\to\infty} \frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda^{(n_r)}\bar{m}^{(n_r)}(n_r y)V}\middle|V > n_r y\right] - \frac{\bar{m}^{(n_r)}(n_r y)\frac{\beta}{n_r}(n_r y)^{\nu}}{l(n_r y)}}{\left(\lambda^{(n_r)} n_r y\bar{m}^{(n_r)}(n_r y)\right)^{\nu}\left(\frac{l\left(\frac{1}{\lambda^{(n_r)}\bar{m}^{(n_r)}(n_r y)}\right)}{l(n_r y)}\right)} = 1. \tag{3.4.3}$$

Lemmas 3.4.1, 3.4.3, and 3.4.2 reduce equation (3.4.3) to

$$\frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\tilde{\kappa}T_\nu}\right] - \tilde{\kappa}\gamma y^{\nu-1}}{(\lambda\tilde{\kappa})^{\nu}} = 1.$$

Thus, any limit point of $ny\bar{m}^{(n)}(ny)$ satisfies equation (3.4.1), so lemma 3.4.4 implies the limit point is unique, so $\lim_{n\to\infty} ny\bar{m}^{(n)}(ny) \to \kappa$. ■

3.4.2 Properties of $\kappa(y)$

In this section we describe several properties of κ. In particular $\kappa(y)$ is uniformly bounded above and regularly with parameter $\nu - 1$. First we need the a left inverse function.

Let $(\cdot)^{\rightharpoonup}$ be the map on $\{f : [0,\infty) \to [0,\infty)\}$ given by $f^{\rightharpoonup}(y) = \sup\{s : f(s) > y\}$ with the convention that the supremum of an empty set is $-\infty$. So, $(\cdot)^{\rightharpoonup}$ maps the set of positive valued Borel measurable nonincreasing functions on $(0,\infty)$ with $\lim_{t\downarrow 0} f(t) = \infty$ into the right continuous Borel measurable functions on $(0,\infty)$ with $f(\infty) = 0$ since $\cup_{t>y}\{s : f(s) > t\} = \{s : f(s) > y\}$.

Proposition 3.4.6. *Suppose G is nonincreasing, positive, $\lim_{t\downarrow 0} G(t) = \infty$, and G is regularly varying at zero with parameter $-\alpha$ for $0 \leq \alpha \leq \infty$. Then $G^{\rightharpoonup}$ is regularly varying at infinity with parameter $-1/\alpha$.*

Proof. Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ by $h(t) = 1/t$. We have $G \circ h$ is regularly varying at infinity with parameter α, $G \circ h(\infty) = \infty$, and $h \circ G$ is nondecreasing. Thus, Resnick Proposition 0.8 gives $h \circ G^{\rightharpoonup}$ is regularly varying at infinity with parameter $1/\alpha$ since $h \circ G^{\rightharpoonup} = h^{\rightharpoonup} \circ G^{\rightharpoonup} = (G \circ h)^{\leftharpoonup}$. f is regularly varying at infinity with parameter α implies that $h \circ f$ is regularly varying at infinity with parameter $-\alpha$, and $h \circ h$ is the identity function thus, $G^{\rightharpoonup} = h \circ h \circ G^{\rightharpoonup}$ is regularly varying at infinity with parameter $-1/\alpha$. ■

Lemma 3.4.7. *For fixed (λ, γ, ν), $\kappa(y)$ defined implicitly by equation (3.4.1) is continuous and regularly varying with parameter $-\nu$ if $\gamma > 0$ and $\kappa(y)$ is constant if $\gamma = 0$. Moreover, $\kappa(y) < \frac{1}{\lambda}\left(\frac{-1}{\gamma(1-\nu)}\right)^{1/\nu}$.*

Proof. If $\gamma = 0$, then κ satisfies $\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right] = (\lambda\kappa)^\nu$ so it does not depend on

y. If $\gamma > 0$, then κ satisfies

$$\left(\frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right]-(\lambda\kappa)^\nu}{\kappa\gamma}\right)^{1/(\nu-1)} = y.$$

Since

$$\kappa \mapsto \left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right] \text{ is strictly decreasing,}$$

$$\kappa \mapsto -(\lambda\kappa)^\nu \text{ is strictly decreasing and,}$$

$$\kappa \mapsto \kappa\gamma \text{ is strictly increasing,}$$

and each of these functions is continuous we have the $\kappa \mapsto y(\kappa)$ is strictly decreasing and continuous. So, $y \mapsto \kappa(y)$ is continuous. Again thinking of the inverse function, the inverse of $\kappa(y)$ is regularly varying at zero with parameter $-1/(\nu-1)$ since

$$\lim_{\kappa\to 0}\kappa^{1/(\nu-1)}\left(\frac{\left(\frac{-1}{\Gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right]-(\lambda\kappa)^\nu}{\kappa\gamma}\right)^{1/(\nu-1)} = \left(\frac{-1}{\Gamma(1-\nu)\gamma}\right)^{1/(\nu-1)}.$$

So, by proposition 3.4.6 we have $\kappa(y)$ is regularly varying with parameter $1-\nu$.

From equation (3.4.1) we have

$$\begin{aligned}\kappa &= \frac{1}{\lambda}\left(\left(\frac{-1}{\gamma(1-\nu)}\right)\mathbb{E}\left[e^{-\lambda\kappa T_\nu}\right]-\kappa\gamma y^{\nu-1}\right)^{1/\nu}\\ &\le \frac{1}{\lambda}\left(\frac{-1}{\gamma(1-\nu)}\right)^{1/\nu}.\end{aligned}$$

■

Note that although the limit points of $ny\bar{m}^{(n)}(ny)$ are uniformly bounded above 3.3.2 and below 3.3.4, this does not imply $\kappa(y)$ is bounded uniformly when κ is defined by (3.4.1)

Corollary 3.4.8. *Under the assumptions of lemma 3.3.1, let b be a real number then*

$$\lim_{n\to\infty} n\bar{m}^{(n)}(ny+b) = \kappa/y.$$

Proof. Fix $\epsilon > 0$. Let $N = \frac{|b|}{\epsilon}$ then for $n > N$ we have $n(y-\epsilon) \leq ny + b \leq n(y+\epsilon)$. For each $n > 0$, $n\bar{m}^{(n)}$ is increasing so

$$n\bar{m}^{(n)}(n(y-\epsilon)) \leq n\bar{m}^{(n)}(ny+b) \leq n\bar{m}^{(n)}(n(y+\epsilon)).$$

Since $y \mapsto \kappa/y$ is continuous letting ϵ go to zero in (3.4.2) gives $\lim_{n\to\infty} n\bar{m}^{(n)}(ny+b) = \kappa/y$. ■

3.5 Representation for $W_2(\tilde{t}_k)$

In this section we write the wait time in the second queue in terms of independent random variables. Here we are using the fact that for the M/G/1 queue the length of an idle period is independent of the service times in the preceding busy period. Let $\tilde{t}_n^{(n)}$ be the **transfer time** of the last customer from the nth busy period in the first queue in the nth system. Let W_2^n be the workload in the second queue of the nth system. Consider the actual waiting time of the a customer in the second of a tandem queue with identical service times. At the epoch of the arrival of the last job in the nth busy period to the second queue, the workload in the second queue is $W_2(\tilde{t}_n)$.

For completeness we include a proof for why each idle period in a particular $M/G/1$ queue is exponentially distributed.

Proposition 3.5.1. *Each idle period in the M/G/1 queue with $\rho \leq 1$ is exponentially distributed, moreover the sequence of idle periods is an independent collection.*

Proof. The durations of the free periods are independent random variables with the same distribution Feller 1 VI.9 [6].

Let $I_1^{(n)} = I_1$ be the duration of the first idle period in the nth system, assuming the first job arrives at epoch 0. Then

$$\mathbb{P}\{I_1 > x\} = \mathbb{P}\left\{\sum_{k=1}^{N}(U_k - V_k) > x \text{ where } N = \inf\left\{n \geq 1 : \sum_{k=1}^{n}(U_k - V_k) > 0\right\}\right\}.$$

For $\rho \leq 1$, the number of service times in a busy period N is a proper random variable taking values in $\{1, 2, \ldots\}$ since $\mathbb{P}\{N > n\} \leq \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(U_i - V_i) \leq 0\right\} \to 0$ by dominated convergence since $\frac{1}{n}\sum_{i=1}^{n}(U_i - V_i) \to \mathbb{E}[V_1]\frac{1-\rho}{\rho} > 0$ by the law of large numbers.

$$\mathbb{P}\{I_1 > x\} = \sum_{n=1}^{\infty}\mathbb{P}\left\{\sum_{k=1}^{n}(U_k - V_k) > x : n = \inf\left\{n \geq 1 : \sum_{k=1}^{n}(U_k - V_k) > 0\right\}\right\}.$$

For a fixed n, let $Y_n = V_n - \sum_{k=1}^{n-1}(U_k - V_k)$

$$\begin{aligned}
&\mathbb{P}\left\{\sum_{k=1}^{n}(U_k - V_k) > x : n = \inf\left\{n \geq 1 : \sum_{k=1}^{n}(U_k - V_k) > 0\right\}\right\} \\
&= \mathbb{P}\left\{U_n > x + Y_n \text{ and } U_n > Y_n > 0 \text{ and } \sup_{j=1,\ldots,n-1}\left\{\sum_{k=1}^{j} U_k - V_k\right\} \leq 0\right\} \\
&= \mathbb{P}\{U_n > x\}\,\mathbb{P}\left\{U_n > Y_n > 0 \text{ and } \sup_{j=1,\ldots,n-1}\left\{\sum_{k=1}^{j} U_k - V_k\right\} \leq 0\right\} \\
&= \mathbb{P}\{U_n > x\}\,\mathbb{P}\{N = n\}
\end{aligned}$$

by the memorylessness of the exponential random variable U_n since Y_n is independent of U_n. Since U_n are iid, $\mathbb{P}\{I_1 > x\} = \mathbb{P}\{U_1 > x\}$ for each $x > 0$. ■

Proposition 3.5.2. *For each $n \geq 1$.*

$$W_2(\tilde{t}_n) = \max_{k=1}^{n} \left(M_k - \sum_{j=k}^{n-1} I_j \right),$$

where M_k is the largest service time the in the kth busy period in the first queue and I_k is the length of the idle period in the first queue between the kth and $k+1$th busy period.

Proof. From the Lindley recursion see Asmussen III.7 [1] we have the actual waiting time of the nth customer in a first in first out queue is

$$W_n = [W_{n-1} - U_n]^+ + V_n,$$

for $n = 1, 2, \ldots$ where W_0 is the initial workload, U_n is the interarrival time between service V_{n-1} and V_n. For the second queue the period between the arrival of the $j-1$ and jth customer is $U_j = V_j + \xi_j$, where ξ_j is the length of the idle period in the first queue preceding the jth arrival to queue if V_j is the first customer in a busy period in the first queue and $\xi_j = 0$ otherwise. The actual wait time of the n customer in the second queue is

$$\begin{aligned}
W_2(\tilde{t}_n) &= \left[W_2(\tilde{t}_{n-1}) - (V_n + \xi_n)\right]^+ + V_n \\
&= \max\left\{W_2(\tilde{t}_{n-1}) - (V_n + \xi_n) + V_n, V_n\right\} \\
&= \max\left\{W_2(\tilde{t}_{n-1}) - \xi_n, V_n\right\} \\
&= \max\left\{W_2(\tilde{t}_{n-2}) - \xi_{n-1} - \xi_n, V_{n-1} - \xi_n, V_n\right\} \\
&= \max\left\{W_0 - \sum_{k=1}^{n} \xi_k, \max_{i=1}^{n}\left\{V_i - \sum_{j=i+1}^{n} \xi_j\right\}\right\}.
\end{aligned}$$

Let $\sigma(0) = 0$ and $\sigma(k)$ be the index of the last arrival in the kth busy period in the first queue for k in $\{1, 2, \ldots\}$. By partitioning $\{1, 2, \ldots, \sigma(n)\}$ into busy periods we have

$$\max_{i=1}^{\sigma(n)} \left\{ V_i - \sum_{j=i+1}^{\sigma(n)} \xi_j \right\} = \max_{k=1}^{n} \left\{ \max_{i=\sigma(k-1)+1}^{\sigma(k)} \left\{ V_i - \sum_{j=i+1}^{\sigma(n)} \xi_j \right\} \right\}.$$

$i \mapsto \sum_{j=i+1}^{\sigma(n)} \xi_j$ is constant for i ranging over a busy period and equal to total idleness that has accrued in the first queue from the epoch of the arrival of the i customer to the epoch of the arrival of the $\sigma(n)$ which is $\sum_{j=k}^{n-1} I_j$ for i in the kth busy period.

$$\begin{aligned} \max_{i=1}^{\sigma(n)} \left\{ V_i - \sum_{j=i+1}^{\sigma(n)} \xi_j \right\} &= \max_{k=1}^{n} \left\{ \max_{i=\sigma(k-1)+1}^{\sigma(k)} \{V_i\} - \sum_{j=k}^{n-1} I_j \right\} \\ &= \max_{k=1}^{n} \left\{ M_k - \sum_{j=k}^{n-1} I_j \right\}. \end{aligned}$$

■

3.6 Distribution of $W_2^{(n)}(\tilde{t}_{[nt]}^{(n)})$

The main result in this section is that distribution of the workload at particular times and appropriately scaled. The sequence of idle periods is iid exponential $\lambda^{(n)}$ in the nth system. Since the largest job in a busy period is independent of the idle period that follows, it is convenient to reindex the sequence of idle periods. This is why we write $\sum_{i=1}^{k-1} I_i^{(n)}$ instead of $\sum_{i=k}^{n-1} I_i^{(n)}$ in the following proposition.

Proposition 3.6.1. *Suppose in the nth system the sequence of idle times $\{I_i^{(n)}\}_{i=1}^{\infty}$*

is iid with $\mathbb{E}\left[I_1^{(n)}\right] = 1/\lambda^{(n)}$ *and* $\mathbf{Var}\left(I_1^{(n)}\right) = (\lambda^{(n)})^{-2}$ *such that* $\lambda^{(n)} \to \lambda > 0$.

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{n}\left(M_k^{(n)} - \sum_{i=1}^{k-1} I_i^{(n)}\right) \le x\right\} = \lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{n}\left(M_k^{(n)} - (k-1)/\lambda\right) \le x\right\}.$$

Proof. We have

$$\max_{k=1}^{n}\left(M_k^{(n)} - \frac{k-1}{\lambda}\right) - \frac{(n-1)|\lambda - \lambda^{(n)}|}{\lambda\lambda^{(n)}} - \max_{k=1}^{n}\left(\frac{k-1}{\lambda^{(n)}} - \sum_{i=1}^{k-1} I_i^{(n)}\right) \le \max_{k=1}^{n}\left(M_k^{(n)} - \sum_{i=1}^{k-1} I_i^{(n)}\right)$$

and

$$\max_{k=1}^{n}\left(M_k^{(n)} - \sum_{i=1}^{k-1} I_i^{(n)}\right) \le \max_{k=1}^{n}\left(M_k^{(n)} - \frac{k-1}{\lambda^{(n)}}\right) + \frac{(n-1)|\lambda - \lambda^{(n)}|}{\lambda\lambda^{(n)}} + \max_{k=1}^{n}\left(\frac{k-1}{\lambda^{(n)}} - \sum_{i=1}^{k-1} I_i^{(n)}\right),$$

so it suffices to show

$$\frac{1}{n}\max_{k=1}^{n}\left(\left|\sum_{i=1}^{k-1}\left(I_i^{(n)} - \frac{1}{\lambda^{(n)}}\right)\right|\right) \to 0 \qquad \text{in probability as } n \to \infty.$$

This follows from Kolmogorov's maximal inequality, for each $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{n}\left(\left|\sum_{i=1}^{k-1}\left(I_i^{(n)} - \frac{1}{\lambda^{(n)}}\right)\right|\right) \ge \epsilon\right\} &= \mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{n-1}\left(\left|\sum_{i=1}^{k}\left(I_i^{(n)} - \frac{1}{\lambda^{(n)}}\right)\right|\right) \ge \epsilon\right\} \\
&= \mathbb{P}\left\{\max_{k=1}^{n-1}\left(\left|\sum_{i=1}^{k}\left(I_i^{(n)} - \frac{1}{\lambda^{(n)}}\right)\right|\right) \ge n\epsilon\right\} \\
&\le \frac{1}{(n\epsilon)^2}\frac{n-1}{(\lambda^{(n)})^2} \\
&\to 0,
\end{aligned}$$

as $n \to \infty$ because $\lambda^{(n)} \to \lambda > 0$. ■

Proposition 3.6.2. *If* $nf(n, ny) \to g(y)$ *uniformly on* $[0, t]$*, and* g *is continuous on* $[0, t]$ *then*

$$\lim_{n\to\infty} \sum_{k=1}^{[nt]} f(n, k) = \int_0^t g(y)\, dy.$$

Proof. Define the measure $\mu = \sum_{k=1}^{\infty} \delta_{\{k\}}$ where $\delta_{\{k\}}(A) = 1$ if $k \in A$ and 0 otherwise.

$$\sum_{k=1}^{[nt]} f(n, k) = \int_{k\in\mathbb{R}} f(n, k) 1_{[0,[nt]]}(k)\, \mu(dk).$$

Substitute $y = k/n$.

$$\begin{aligned}\sum_{k=1}^{[nt]} f(n, k) &= \int_{y\in\mathbb{R}} f(n, ny) 1_{[0,t]}(y)\, \mu(ndy) \\ &= \int_{y\in\mathbb{R}} nf(n, ny) 1_{[0,t]}(y)\, \mu(ndy)/n.\end{aligned}$$

For n sufficiently large $|nf(n, ny) - g(y)| \leq \epsilon$ and $\int_{y\in\mathbb{R}} 1_{[0,t]}(y)\, \mu(ndy)/n \leq 1$, so $\mu(ndy)/n$ converges weakly to Lebesgue measure on $(0, \infty)$ implies

$$\int_{y\in\mathbb{R}} g(y) 1_{[0,t]}(y)\, \mu(ndy)/n \to \int_0^t g(y)\, dy.$$

■

Proposition 3.6.3. *Under the assumptions of lemma 3.3.1, for* $x > 0$ *we have*

$$\lim_{n\to\infty} \mathbb{P}\left\{ \frac{1}{n} \max_{k=1}^{[nt]} \left(M_k^{(n)} - \frac{k-1}{\lambda} \right) \leq x \right\} = \begin{cases} \left(1 + \frac{t}{x\lambda}\right)^{-\lambda\kappa} & \text{if } \gamma = 0 \\ \exp\left\{ -\lambda \int_x^{x+t/\lambda} \kappa(y)/y\, dy \right\} & \text{if } \gamma > 0 \end{cases}$$

Proof.

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{[nt]}\left(M_k^{(n)}-\frac{k-1}{\lambda}\right)\le x\right\} &= \mathbb{P}\left\{\max_{k=1}^{[nt]}\left(M_k^{(n)}-\frac{k-1}{\lambda}\right)\le nx\right\}\\ &= \prod_{k=1}^{[nt]}\mathbb{P}\left\{M_k^{(n)}-\frac{k-1}{\lambda}\le nx\right\}\\ &= \prod_{k=1}^{[nt]}\mathbb{P}\left\{M_k^{(n)}\le nx+\frac{k-1}{\lambda}\right\}\\ &= \prod_{k=1}^{[nt]} m^{(n)}\left(nx+\frac{k-1}{\lambda}\right)\\ &= \exp\left\{\sum_{k=1}^{[nt]}\ln\left(m^{(n)}\left(nx+\frac{k-1}{\lambda}\right)\right)\right\}.\end{aligned}$$

Let $f(n,k)=\ln\left(m^{(n)}\left(nx+\frac{k-1}{\lambda}\right)\right)$ then

$$\begin{aligned} nf(n,ny) &= n\ln\left(m^{(n)}\left(nx+\frac{ny-1}{\lambda}\right)\right)\\ &= \ln\left(\left(1-\frac{n\bar{m}^{(n)}\left(nx+ny/\lambda-\frac{1}{\lambda}\right)}{n}\right)^n\right).\end{aligned}$$

We have $\ln((1-x/n)^n)\to -x$ as $n\to\infty$ and $n\bar{m}^{(n)}\left(n\left(x+y/\lambda\right)-\frac{1}{\lambda}\right)\to\kappa(x+y/\lambda)/(x+y/\lambda)$ as $n\to\infty$ by corollary 3.4.8. Thus, $nf(n,ny)\to-\kappa(x+y/\lambda)/(x+y/\lambda)$ where the convergence is uniform for $y\in[0,t]$ since for each n, $m^{(n)}$ is nondecreasing and the limit is continuous. Now continuity of the exponential function and proposition 3.6.2 gives

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left\{\frac{1}{n}\max_{k=1}^{[nt]}\left(M_k^{(n)}-\frac{k-1}{\lambda}\right)\le x\right\} &= \exp\left\{-\int_0^t \kappa(x+y/\lambda)/(x+y/\lambda)\,dy\right\}\\ &= \exp\left\{-\lambda\int_x^{x+t/\lambda}\kappa(y)/y\,dy\right\}.\end{aligned}$$

■

Finally, we prove the main result for this chapter.

Theorem 3.6.4. *Let* $\rho^{(n)} = \lambda^{(n)}\mathbb{E}[V]$*, and assume* $\frac{1-\rho^{(n)}}{n(1-F(n))} \to \gamma \geq 0$ *as* $n \to \infty$*. Assume* $1 - F(t) = \left(\frac{-1}{\Gamma(1-\nu)}\right) t^{-\nu} l(t)$ *for* $1 < \nu < 2$ *and* l *a slowly varying function. Let* $\kappa(y)$ *be such that the parameters* $(\kappa, \lambda = \frac{1}{\mathbb{E}[V]}, \nu, \gamma, y)$ *satisfies equation* (3.4.1)*. For* $x > 0$ *we have*

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n} W_2^{(n)}\left(\tilde{t}_{[nt]}^{(n)}\right) \leq x\right\} = \begin{cases} \left(1 + \frac{t}{x\lambda}\right)^{-\lambda\kappa} & \text{if } \gamma = 0 \\ \exp\left\{-\lambda \int_x^{x+t/\lambda} \kappa(y)/y\, dy\right\} & \text{if } \gamma > 0 \end{cases}$$

Proof. By proposition 3.5.2

$$\mathbb{P}\left\{\frac{1}{n} W_2^{(n)}\left(\tilde{t}_{[nt]}^{(n)}\right) \leq x\right\} = \mathbb{P}\left\{\frac{1}{n} \max_{k=1}^{[nt]} \left(M_k^{(n)} - \sum_{j=k}^{[nt]-1} I_j^{(n)}\right) \leq x\right\}.$$

For each n, the iid collections $\{I_k^{(n)}\}$ and $\{M_k^{(n)}\}$ are independent so

$$\max_{k=1}^{[nt]} \left(M_k^{(n)} - \sum_{j=k}^{[nt]-1} I_j^{(n)}\right) \sim \max_{k=1}^{[nt]} \left(M_k^{(n)} - \sum_{j=1}^{k-1} I_j^{(n)}\right).$$

By proposition 3.6.1

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n} W_2^{(n)}\left(\tilde{t}_{[nt]}^{(n)}\right) \leq x\right\} = \lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n} \max_{k=1}^{[nt]} \left(M_k^{(n)} - (k-1)/\lambda\right) \leq x\right\}.$$

By proposition 3.6.3

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{1}{n} W_2^{(n)}\left(\tilde{t}_{[nt]}^{(n)}\right) \leq x\right\} = \begin{cases} \left(1 + \frac{t}{x\lambda}\right)^{-\lambda\kappa} & \text{if } \gamma = 0 \\ \exp\left\{-\lambda \int_x^{x+t/\lambda} \kappa(y)/y\, dy\right\} & \text{if } \gamma > 0 \end{cases}$$

■

Bibliography

[1] Asmussen, S. Applied probability and queues, 1987.

[2] Billingsley, P. *Convergence of probability measures.* Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics. Wiley, 1968.

[3] Boxma, O. On the longest service time in a busy period of the $M \backslash G \backslash 1$ queue. *Stochastic Processes and their Applications 8*, 1 (1978), 93–100.

[4] Boxma, O. On a tandem queueing model with identical service times at both counters, i. *Advances in Applied Probability* (1979), 616–643.

[5] Ethier, S. N., and Kurtz, T. G. *Markov processes: characterization and convergence*, vol. 282. John Wiley & Sons, 2009.

[6] Feller, W. An introduction to probability and its applications, vol. ii. *Wiley, New York* (1971).

[7] Harrison, J. M., and Williams, R. Brownian models of feedforward queueing networks: Quasireversibility and product form solutions. *The Annals of Applied Probability* (1992), 263–293.

[8] Jackson, J. R. Jobshop-like queueing systems. *Management science 10*, 1 (1963), 131–142.

[9] Jacod, J., and Shiryaev, A. N. *Limit theorems for stochastic processes. A Series of Comprehensive Studies in Mathematics 288.* Springer-Verlag, 1987.

[10] N. H. Bingham, C. M. Goldie, J. L. T. *Regular Variation.* Cambridge University Press, 1987.

[11] Whitt, W. Some useful functions for functional limit theorems. *Mathematics of operations research 5*, 1 (1980), 67–85.

[12] Whitt, W. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues.* Springer Science & Business Media, 2002.