EXACT ASYMPTOTICS FOR THE STATIONARY DISTRIBUTION OF A MARKOV CHAIN: A PRODUCTION MODEL

By Ivo Adan,
Robert D. Foley and David R. McDonald*

Eindhoven University of Technology

We derive rough and exact asymptotic expressions for the stationary distribution π of a Markov chain arising in a queueing/production context. The approach we develop can also handle "cascades", which are situations where the fluid limit of the large deviation path from the origin to the increasingly rare event is nonlinear. Our approach considers a process that starts at the rare event. In our production example, we can have two sequences of states that asymptotically lie on the same line, yet π has different asymptotics on the two sequences.

Key words and phrases: rare events, large deviations, exact asymptotics, change of measure, h transform, time reversal, Markov additive process, Markov chain, R-transient

AMS 2000 subject classifications: Primary 60K25; Secondary 60K20.

1. Introduction. We are interested in estimating the probability of rare events related to the stationary distribution π of Markov chains of the type that typically arise in modelling queueing networks. Unless π can be computed explicitly, such results are usually difficult to obtain—even through simulation. In this paper, we develop an approach to deriving the exact asymptotics of π that allows us to analyze situations where the fluid limit of excursions to the (increasingly) rare event is nonlinear. This nonlinear behavior can arise in a pair of stable, M/M/1 queues in tandem. Let (x, y) denote the joint queue length where x is the number in the downstream node, and let π denote the stationary distribution of the joint queue length. If we are interested in $\pi(\ell, y)$, think of ℓ as a large integer, we are interested in excursions from the origin to (ℓ, y). If the downstream server is substantially faster than the upstream server, and "substantially" can be determined from (2.3) in [5], it will be easier to initially accumulate a large number of customers in the upstream server while the number in the downstream server remains small. When a sufficient number have accumulated,

*Research supported in part by NSERC grant A4551

customers cascade from the upstream server to the downstream server. Thus, most excursions from the origin that reach (ℓ, y) will initially climb along the y-axis to a state near $(0, c\ell)$ with $c > 0$ before changing directions and heading towards (ℓ, y). The fluid limit or functional strong law, computed by dividing the joint queue length process by ℓ, speeding up time by ℓ, and letting $\ell \to \infty$, will be piecewise linear, first climbing the y-axis from the origin to $(0, c)$ and then changing direction and heading southeast to $(1, 0)$. On the other hand, if the downstream server were substantially slower than the upstream server, excursions to $(\ell, 0)$ would stay close to the x-axis, and the fluid limit would be a line segment going directly from the origin to $(1, 0)$.

Linear cases have been studied in [3, 4, 9]. The linear cases were connected to a particular transition matrix having convergence parameter $R = 1$ and being either 1-positive recurrent in the "jitter" case studied in [3, 9] or being either 1-null recurrent or 1-transient in the "bridge" cases studied in [4]. In the nonlinear case that transition matrix has convergence parameter $R > 1$. The approach developed here and in [6] can handle both linear and nonlinear situations. The essence of the approach is to consider a stochastic process that starts at the distant state and closely approximates the time reversal of the Markov chain. The primary methodological result is Lemma 6, which is further developed in [6].

To illustrate the power of the approach, we completely describe the exact asymptotics of π for a production model in all directions and for all stable parameter settings. The production model, described in the next section, has unbounded jumps; at every point in the state space, the boundaries influence the possible transitions. In addition, for certain regions of the parameters, the fluid limits of excursions to the rare events are nonlinear.

2. Production model & results for the production model. To illustrate the results, consider a production system consisting of two parallel machines, labeled m_1 and m_2. In front of the machines, there is a central buffer with infinite capacity where jobs await processing. The processing times are independent, exponential random variables, with rate μ_1 at machine m_1 and with rate μ_2 at machine m_2. There are two types of jobs: a and b. Type a jobs have the advantage that they can be processed at either machine. Type b jobs can only be processed at machine m_1; that is, machine m_2 only processes jobs of type a.

Such a situation can arise in a variety of contexts. In some situations, machine m_2 may have been restricted to processing jobs of type a jobs since type a jobs have a higher priority; in other situations, machine m_2 may

be incapable of processing certain jobs. For example, suppose the machines insert chips on circuit boards and the set of chip types available at machine m_2 is a proper subset of the set at m_1. A circuit board needing only chips available at machine m_2 would be a type a job, but any circuit board needing a chip that is only at machine m_1 would be a type b job.

We assume that the two job types arrive according to independent Poisson stream with rate λ_a and λ_b. We also assume that the service time and arrival processes are independent. Machine m_1 is never idle when there are jobs waiting in the system, and machine m_2 is never idle when there are type a jobs waiting for service.

We still need to describe the service discipline. Basically, the system tries to process the jobs in order of arrival except that machine m_2 can only process type a jobs. Thus, machine m_1 will always choose the job at the head of the buffer, but machine m_2 may have to search through the buffer for the oldest type a job. Lastly, when both machines are idle and a type a job arrives, assume that the job will be processed by m_1 with probability η. Hence, the system has five parameters: λ_a, λ_b, μ_1, μ_2, and η.

We will model the system as a Markov process, and we are interested in its stationary distribution π. If any one of the four parameters λ_a, λ_b, μ_1, and μ_2 are zero, the system becomes much simpler to analyze; hence, unless otherwise mentioned, we assume that

$$\lambda_a > 0, \qquad \lambda_b > 0, \qquad \mu_1 > 0, \qquad \mu_2 > 0. \tag{1}$$

We will also assume that

$$\alpha \equiv \frac{\lambda_a + \lambda_b}{\mu_1 + \mu_2} < 1, \qquad \beta \equiv \frac{\lambda_b}{\mu_1} < 1, \tag{2}$$

which are the necessary and sufficient conditions for stability, which in this paper is equivalent to having a unique stationary distribution π. It will be convenient to define $\gamma \equiv \lambda_b/(\lambda_a + \lambda_b)$, which is the probability that a job is of type b.

If the state of the process were simply the number of jobs of each type in the system, the process would not be Markovian. Instead, we model the system as a Markov process by delaying the discovery of a job's type until a machine needs to know the type, and only machine m_2 ever needs to discover a job's type. We represent the system as a Markov process with a two dimensional state space and define the states as:

- $(0, 0)$: the system is empty.

- $(1,0)$: there is one job in the system, and machine m_2 is working on that job.
- $(0,y)$ with $y>0$: machine m_2 is idle, machine m_1 is working on a job, and there are $y-1$ type b jobs waiting in the queue.
- (x,y) with $x>0$ and $y>0$: both machines are working, $x-1$ jobs of unknown type and $y-1$ jobs of type b are waiting in the queue.

Let S denote the state space; note that $(2,0),(3,0),\dots$ are not in S since these states would correspond to m_1 being idle, but m_1 is never idle when there is more than one job in the system. States on the y-axis correspond to states with m_2 idle. Also note that state (x,y) means that there are $x+y$ jobs in the system. Any type b jobs in the system must have arrived earlier than the rest of the jobs in the system, which are of unknown type, and machine m_2 must have inspected the type b jobs to have learned their type. The state space and transition rates from five selected states are depicted in Figure 1.

Under (1) and (2), the Markov process is irreducible and has a unique stationary distribution, which will be denoted by π; the argument is delayed until Subsection 2.1. Our first proposition gives bounds on π; the rough asymptotics of π follow directly from these bounds.

PROPOSITION 1. *There exists constants c_1 and c_2 such that*

$$0<c_1\alpha^x\beta^y\le\pi(x,y)\le c_2\alpha^x\beta^y<\infty. \tag{3}$$

Since the proof is tangential to our main interest, we have placed the proof in Appendix A

By $x_\ell\sim y_\ell$, we mean that $x_\ell/y_\ell\to 1$ where here and throughout this paper $\to$ means as $\ell\to\infty$. In this paper, "the exact asymptotics of π" means deriving an asymptotic expression for $\pi(x_\ell,y_\ell)$, that is, deriving an expression of the form $\pi(x_\ell,y_\ell)\sim f_\ell$, where (x_ℓ,y_ℓ) is some divergent sequence of states. The "rough asymptotics of π," means deriving an asymptotic expression for $\log\pi(x_\ell,y_\ell)$. From (3), it is straightforward to derive the rough asymptotics of π.

COROLLARY 1. *Let (x_ℓ,y_ℓ) be any sequence of states with $x_\ell+y_\ell\to\infty$ and $x_\ell/(x_\ell+y_\ell)$ converging to a constant. Then*

$$\log\pi(x_\ell,y_\ell)\sim x_\ell\log\alpha+y_\ell\log\beta.$$

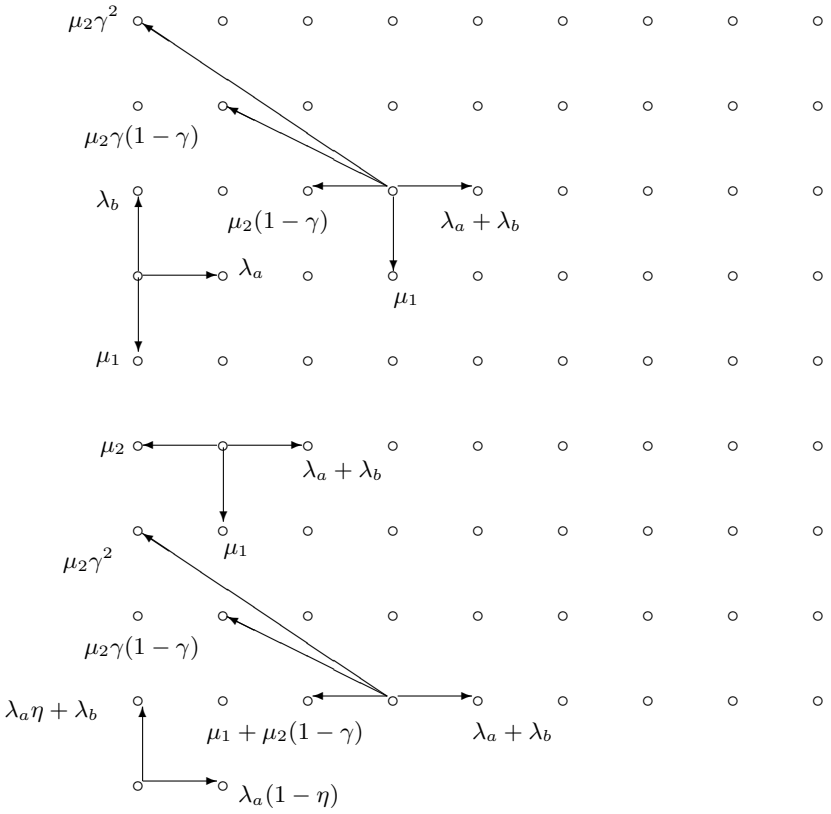

FIG 1. *Transition rates out of selected states (lower left node is* $(0,0)$*).*

The next two propositions give the exact asymptotics of π near an axis, assuming the "jitter" condition holds along that axis. A certain "drift vector"

$$d^* = (d_1^*, d_2^*) = ((\mu_1+\mu_2)(\lambda_a-\mu_2)/\mu_2, \mu_1(\lambda_b/\mu_1 - \lambda_a/\mu_2)), \tag{4}$$

will arise in the analysis, which gives insight into several aspects of the behavior of this system and can be used to define the jitter conditions. Under our assumptions, at least one of the components of the drift vector d^* will be negative. The following are equivalent:

- the jitter condition holds along the x-axis,
- $\lambda_b/\mu_1 < \lambda_a/\mu_2$,
- $\beta < \alpha$, and
- $d_2^* < 0$;

Similarly, the following are equivalent:

- the jitter condition holds along the y-axis,
- $\lambda_a/\mu_2 < 1$, and
- $d_1^* < 0$.

PROPOSITION 2. *If $d_2^* < 0$, then*

$$\pi(\ell, y) \sim \psi_v(\ell, y) \tag{5}$$

where

(6)
$$\psi_v(\ell,y) \equiv \begin{cases} c_v \frac{\alpha}{\mu_1+\mu_2-\lambda_a-\lambda_b} \frac{\lambda_a/\mu_2-\lambda_b/\mu_1}{\lambda_b/\mu_1} \alpha^{\ell-1} & \text{for } y=1 \\ c_v \frac{\alpha}{\mu_1+\mu_2-\lambda_a-\lambda_b} \frac{\lambda_a/\mu_2-\lambda_b/\mu_1}{\lambda_b/\mu_1} \frac{\mu_2}{\mu_1+\mu_2} \alpha^{\ell-1}\beta^{y-1} & \text{for } y=2,3,\ldots \end{cases}$$

and c_v is defined in (13).

PROPOSITION 3. *If $d_1^* < 0$, then*

$$\pi(x, \ell) \sim \psi_w(x, \ell) \tag{7}$$

where

$$\psi_w(x,\ell) \equiv \begin{cases} c_w \frac{\beta}{\mu_1-\lambda_b} \left(1-\frac{\lambda_a}{\mu_2}\right) \beta^{\ell-1} & \text{for } x=0 \\ c_w \frac{\beta}{\mu_1-\lambda_b} \left(1-\frac{\lambda_a}{\mu_2}\right) \frac{\lambda_a}{\mu_1+\mu_2} \alpha^{x-1}\beta^{\ell-1} & \text{for } x=1,2,\ldots \end{cases} \tag{8}$$

where c_w is defined in (26).

The proofs for Proposition 2 and 3 appear in Sections 3 and 5, respectively. Propositions 2 and 3 describe the exact asymptotics of π near an axis when the jitter condition for that axis holds. The next three propositions complete a rather general description of the exact asymptotics of π in the rest of the state space. The proofs of these three propositions use the approach to deriving exact asymptotics that is developed in Section 4.

Intuitively, our approach to obtaining asymptotic expressions for π considers an approximate time reversed process that *starts* at some distant state (x, y). The hypotheses of the next three propositions partition the stability region into three cases depending on whether (i) the x-axis jitter condition holds and the y-axis jitter condition does not ($d_1^* < 0, d_2^* \geq 0$), (ii) vice-versa ($d_1^* \geq 0, d_2^* < 0$), and (iii) the jitter conditions hold on both axes ($d_1^* < 0, d_2^* < 0$). If neither jitter condition holds ($d_1^* \geq 0, d_2^* \geq 0$), then the process is not stable.

Since Proposition 3 already gave the asymptotics near the y-axis under the conditions of the following proposition, we will let $x_\ell \to \infty$ in the following proposition.

PROPOSITION 4. *Let (x_ℓ, y_ℓ) be any sequence of states with $x_\ell \to \infty$. If $d_1^* < 0$ and $d_2^* \geq 0$, then $\pi(x_\ell, y_\ell) \sim \chi_v(x_\ell, y_\ell)$ where*

$$\chi_v(x_\ell, y_\ell) \equiv \begin{cases} c_w \frac{\beta}{\mu_1 - \lambda_b} \left(1 - \frac{\lambda_a}{\mu_2}\right) \frac{\lambda_a}{\mu_2} \alpha^{x_\ell - 1} & \text{for } y_\ell = 1, \\ c_w \frac{\beta}{\mu_1 - \lambda_b} \left(1 - \frac{\lambda_a}{\mu_2}\right) \frac{\lambda_a}{\mu_1 + \mu_2} \alpha^{x_\ell - 1} \beta^{y_\ell - 1} & \text{for } y_\ell = 2, 3, \ldots. \end{cases}$$

Since Proposition 2 already gave the asymptotics near the x-axis under the conditions of the following proposition, we will let $y_\ell \to \infty$ in the following proposition.

PROPOSITION 5. *Let (x_ℓ, y_ℓ) be any sequence of states with $y_\ell \to \infty$. If $d_1^* \geq 0$ and $d_2^* < 0$, then $\pi(x_\ell, y_\ell) \sim \chi_w(x_\ell, y_\ell)$ where*

$$\chi_w(x_\ell, y_\ell) \sim \begin{cases} c_v \frac{\mu_2}{\lambda_a} \frac{\alpha}{\mu_1 + \mu_2 - \lambda_a - \lambda_b} \frac{\lambda_a/\mu_2 - \lambda_b/\mu_1}{\lambda_b/\mu_1} \beta^{y_\ell - 1} & \text{for } x_\ell = 0, \\ c_v \frac{\alpha}{\mu_1 + \mu_2 - \lambda_a - \lambda_b} \frac{\lambda_a/\mu_2 - \lambda_b/\mu_1}{\lambda_b/\mu_1} \frac{\mu_2}{\mu_1 + \mu_2} \alpha^{x_\ell - 1} \beta^{y_\ell - 1} & \text{for } x_\ell = 2, 3, \ldots. \end{cases}$$

In the next proposition, the jitter conditions hold on both axes. Hence, Propositions 2 and 3 already give the asymptotics of π near the axes, and we have different asymptotics depending upon which axis the sequence is near. Thus, it should be no surprise that the asymptotics in the interior fall into different cases. There are three cases depending on whether the sequence (x_ℓ, y_ℓ) asymptotically lies above, below or on the line going through the

origin with slope d_2^*/d_1^*. When the sequence asymptotically lies on this "drift line," additional conditions are needed for the asymptotics to exist, and the exact asymptotics become a delicate mixture of the first two cases. In this last case, we assume that ℓ is asymptotically the number of jobs to simplify notation. By choosing the sequence of states appropriately, q can take any value in $[0,1]$. Thus, two different sequences of states may lie on same line asymptotically, yet the stationary distribution for these sequences of states can have different asymptotics. Assume that the sequence $(x_\ell/\ell, y_\ell/\ell)$ has a limit $(\bar{x}, \bar{y})$.

PROPOSITION 6. *Assume that $d_1^* < 0$ and $d_2^* < 0$. Let (x_ℓ, y_ℓ) be any sequence of states with $x_\ell > 1$ and $(x_\ell/\ell, y_\ell/\ell) \to (\bar{x}, \bar{y})$ where $0 < \max(\bar{x}, \bar{y}) < \infty$.*

1. *If $\bar{y}/\bar{x} > d_2^*/d_1^*$, then*

$$\pi(x_\ell, y_\ell) \sim \chi_v(x_\ell, y_\ell).$$

2. *If $\bar{y}/\bar{x} < d_2^*/d_1^*$, then*

$$\pi(x_\ell, y_\ell) \sim \psi_v(x_\ell, y_\ell).$$

3. *If $\bar{y}/\bar{x} = d_2^*/d_1^*$, and there exists (r, s) such that*

$$\sqrt{\ell}\left[\left(\frac{x_\ell}{\ell}, \frac{y_\ell}{\ell}\right) - (\bar{x}, \bar{y})\right] \to (r, s)$$

then

$$\pi(x_\ell, y_\ell) \sim q\chi_v(x_\ell, y_\ell) + (1-q)\psi_v(x_\ell, y_\ell)$$

where $q = \Phi[-(r - sd_1^*/d_2^*)/\sigma]$, Φ is the c.d.f. of a standard normal distribution, and σ^2 is given in (32).

The proofs of the first two propositions appear in Sections 3 and 5, respectively. All quantities in the exact asymptotic expressions for π in the last five propositions are explicitly known except for c_v and c_w, which are defined in (13) and (26). Under the conditions of Proposition 2, c_v is a finite, strictly positive constant that depends on the five system parameters, and similarly for c_w under the conditions of Proposition 3. Because of the form of (5) and (7), we know that there are no subexponential terms like $\ell^{-1/2}$ or $\ell^{-3/2}$ such as encountered in [4]. Both constants can be estimated through simulations that do not involve simulating rare events. In the special case when $\eta = \lambda_a/(2\lambda_a + \lambda_b)$, we know both c_v and c_w explicitly. Under the

conditions of Lemma 1, c_v is easily obtained from $\psi_w(0,\ell) = \pi^*(0,\ell)$ and c_w from $\psi_v(\ell,1) = \pi^*(\ell,1)$.

Our production system has one unusual property that we exploit. Even though π cannot be computed in general, π can be explicitly computed at one specific, non-degenerate value of the parameter η. By taking advantage of the explicit stationary distribution at that one parameter value, we can more easily prove two technical conditions needed in deriving our asymptotic results for the production model. However, even if we did not have this unusual property, we would still be able to obtain all of our exact asymptotic results provided we showed that $\sum_{y=1}^{\infty} \pi(1,y)\alpha^{-y} < \infty$ when $d_2^* < 0$ and that $\sum_{z=0}^{\infty} \pi(z,1)\beta^{-z} < \infty$ when $d_1^* < 0$. Such inequalities can be established by finding the appropriate Lyapounov functions as done in [3].

The unusual property of this production system is described in the following lemma.

LEMMA 1. *If $\eta = \eta^* \equiv \lambda_a/(2\lambda_a + \lambda_b)$, then $\pi(x,y) = \pi^*(x,y)$ where*

$$\pi^*(x,y) \equiv \begin{cases} c^* & \text{for } (x,y) = (0,0), \\ c^* \frac{(1-\eta^*)\lambda_a}{\mu_2} & \text{for } (x,y) = (1,0), \\ c^* \frac{\eta^*\lambda_a + \lambda_b}{\mu_1} \beta^{y-1} & \text{for } x = 0, y > 0, \\ c^* \frac{\eta^*\lambda_a(\lambda_a+\lambda_b)^2}{\mu_1\mu_2} \alpha^{x-1} & \text{for } x > 0, y = 1, \\ c^* \frac{\lambda_a}{\mu_1+\mu_2} \frac{\eta^*\lambda_a+\lambda_b}{\mu_1} \alpha^{x-1}\beta^{y-1} & \text{for } x > 0, y > 1, \end{cases}$$

and

$$c^* \equiv 1/\left(1 + \frac{(1-\eta^*)\lambda_a}{\mu_2} + \frac{\eta^*\lambda_a + \lambda_b}{\mu_1(1-\beta)} + \frac{\eta^*\lambda_a(\lambda_a+\lambda_b)^2}{\mu_1\mu_2(1-\alpha)} + \frac{\lambda_a(\eta^*\lambda_a+\lambda_b)\beta}{(\mu_1+\mu_2)\mu_1(1-\alpha)(1-\beta)}\right)$$

To prove Lemma 1, simply verify that π^* is a distribution satisfying the balance equations. By the way, with a finite capacity buffer, π^* is an invariant measure if $\eta = \eta^*$, but c^* is not the correct normalization constant.

2.1. *Uniformization and stability.* Generally, we will find it more convenient to work with the discrete time Markov chain obtained by uniformizing the continuous time process. The equivalent discrete time Markov chain will be denoted by $X_0, X_1, \ldots$, and we let K be its transition kernel. We refer to this Markov chain as the *uniformized chain*. Of course, π is the stationary distribution for the uniformized chain if and only if π is the stationary distribution for the continuous time process. For convenience and without loss of generality, assume that $\lambda_a + \lambda_b + \mu_1 + \mu_2 = 1$; thus, the transition

probabilities off the diagonal of K are just the corresponding transition rates of the continuous time Markov process. Figure 1 needs only minor changes to show the one step transition probabilities from the same 5 states. State $(0,0)$ needs to have an arrow going to itself with probability $\mu_1+\mu_2$. State $(0,6)$ needs to have an arrow going to itself with probability μ_2.

We stated that the necessary and sufficient conditions for stability are that $\alpha<1$ and $\beta<1$. In the special case when $\eta=\eta^*$, the result is obvious from Lemma 1. We claim that the same conditions hold for $\eta\neq\eta^*$. Note that the expected time from (0,1) to (0,0) and from (1,0) to (0,0) do not depend on η. Thus, if the expected return to (0,0) is finite (infinite) for $\eta=\eta^*$, then it is finite (infinite) for all η.

3. Exact asymptotics along the x-axis under x-jitter conditions. This section proves Proposition 2. We will derive exact asymptotic expressions for $\pi(\ell,y)$ using the same approach as in [3]. The condition needed for jittering along the x-axis will turn out to be $\lambda_b/\mu_1<\lambda_a/\mu_2$ or equivalently $d_2^*<0$. Again, we will work with the discrete time Markov chain $X_0,X_1,\ldots$ with one step transition kernel K obtained by uniformizing the continuous time Markov process and assume that $\lambda_a+\lambda_b+\mu_1+\mu_2=1$. In Section 5, we will do a similar analysis along the y-axis. To simplify notation, the definitions of Δ, K^∞, $\blacktriangle$, h, $\mathcal{K}^\infty$, and φ given in this section will be re-defined in Section 5 when we do the analogous analysis along the other axis. When we want to emphasize that a quantity is based on the definitions given in this section, we add a subscript "v"; we add a subscript "w" to emphasize that the quantity is based on the definitions in Section 5.

Our starting point is Orey's representation of π as

$$\pi(\ell,y)=\sum_{(x,z)\in\Delta}\pi(x,z)\mathrm{E}_{(x,z)}[N_\Delta(\ell,y)] \tag{9}$$

where Δ is some set of states and $N_\Delta(\ell,y)$ is the number of visits to (ℓ,y) before returning to Δ. More precisely, if $T_\Delta=\inf\{n>0:X_n\in\Delta\}$ and 1_A is the indicator of A, then $N_\Delta(\ell,y)=\sum_{n=1}^{T_\Delta}1_{\{X_n=(\ell,y)\}}$.

3.1. *The free process.* The next step is to define Δ and a Markov additive process dubbed the free process that captures the behavior of the uniformized chain on excursions from Δ to the rare events. The one-step transition kernel of the free process will be denoted by K^∞. Since we are interested in large deviations in the first coordinate, the first coordinate needs to be the additive part and the second coordinate the Markovian part. Let $\Delta=\Delta_v=\{(x,y)\in S:x\leq 1\}$. (The subscript "$v$" was chosen since Δ

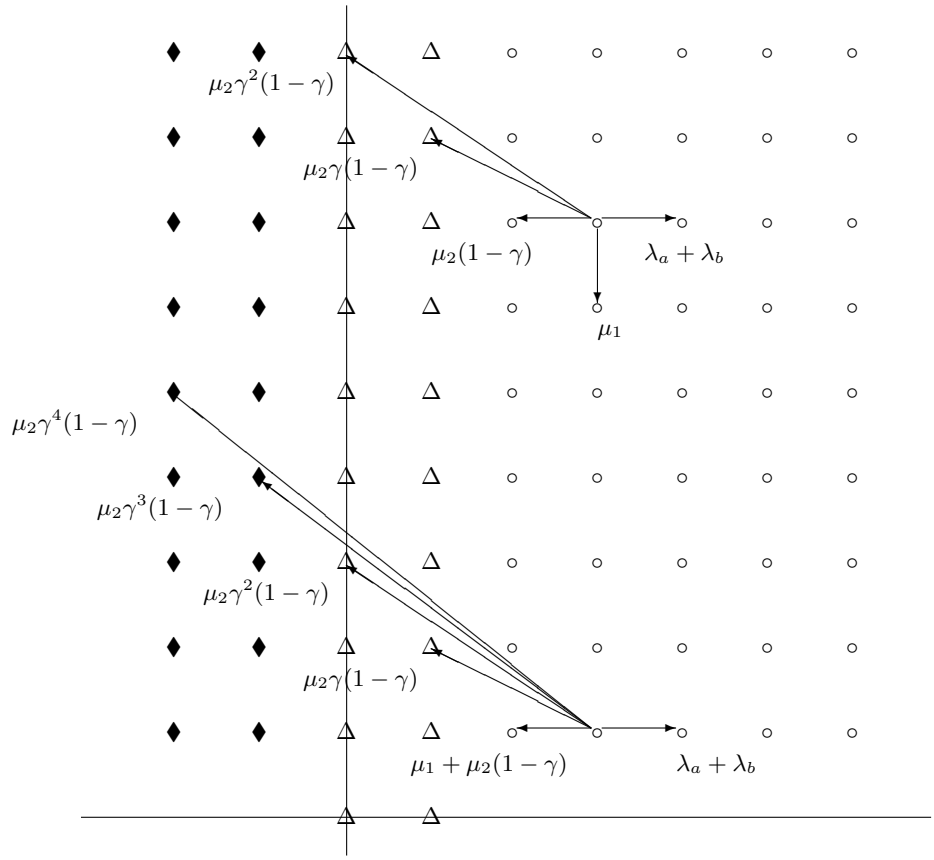

FIG 2. *Selected transition probabilities of the free process (lowest left* Δ *is* $(0,0)$*).*

looks like a *vertical* strip near the y-axis.) Let $K^\infty((m,z);(x+m,y)) = K^\infty((0,z);(x,y))$ where

$$K^\infty((0,z);(x,y)) = \begin{cases} \lambda_a + \lambda_b & \text{for } x=1,\ y=z, \\ \mu_1 & \text{for } x=0,\ y=z-1\geq 1, \\ \mu_1+\mu_2(1-\gamma) & \text{for } x=-1,\ y=z=1 \\ \mu_2(1-\gamma) & \text{for } x=-1 \text{ and } y=z>1 \\ \mu_2\gamma^m(1-\gamma) & \text{for } y-z=m>0,\ x=-(m+1) \end{cases}$$

Since we have removed the boundary Δ, the free process is free to wander over all of $\mathbb{Z}\times\mathbb{N}$ where $\mathbb{Z}\equiv\{\dots,-2,-1,0,1,2,\dots\}$ and $\mathbb{N}\equiv\{1,2,\dots\}$. Let $\blacktriangle = \Delta\cup\{(x,y): x<0, y\geq 1\}$, which is Δ plus all of the new states. Selected transition probabilities are shown from two states in Figure 2 where $\blacklozenge$ denotes a new state. We cannot show all transitions out of any state since there are an infinite number of transitions. The transition probabilities can be translated horizontally without change, which is the Markov additive property. States $(0,0)$ and $(1,0)$ are not accessible from any state (x,y) with $y\geq 1$; the transition probabilities out of these states will be unimportant, and we define them by $K^\infty((0,0),(0,1))=1$. Since K is positive recurrent, we would expect the free process to be transient and drift westerly.

If the free process starts in S, the free process and $X_0, X_1, \dots$ have the same transition probabilities until leaving $S\setminus\Delta$; that is, until hitting $\blacktriangle$. (In particular, K and K^∞ agree for transitions from states in Δ to $S\setminus\Delta$. In other examples, they might disagree; see [6].) For $(\ell,y)\in S\setminus\Delta$, we can rewrite (9)

$$\pi(\ell,y) = \sum_{(x,z)\in\Delta} \pi(x,z)\mathrm{E}_{(x,z)}[N_\blacktriangle(\ell,y)] \tag{10}$$

where $N_\blacktriangle(\ell,y)$ is the number of visits to (ℓ,y) by the free process until hitting $\blacktriangle$. Note that if either process starts in (0,0) or (1,0), there is no contribution to the expectation.

3.2. *The twisted free process.* To define the twisted free process, we need to find a harmonic function $h=h_v$ for the transition kernel K^∞ of the form $h(x,y)=a^x\hat{h}(y)$, where in a temporary abuse of notation a has nothing to do with customer type. By *harmonic*, we mean that h satisfies $K^\infty h = h$. The rough asymptotics will be given by $1/a$, but we already know from Corollary 1 that the rough asymptotics for states (ℓ,y) are given by α; hence, $1/a=\alpha$. If, in a second abuse of notation, we guess that $\hat{h}(y)=b^{y-1}$,

our guess for h is $h(x,y) = \alpha^{-x} b^{y-1}$. If we insert our guess into $K^\infty h = h$ and solve, we discover that $\hat{h}(y) = \alpha^{-(y-1)}$ and $h(x,y) = \alpha^{-(x+y-1)}$.

With this harmonic function we can define the h-transform or twisted kernel $\mathcal{K}((0,z);(x,y)) \equiv K^\infty((0,z);(x,y))h(x,y)/h(0,z)$ yielding

$$\mathcal{K}((0,z);(x,y)) = \begin{cases} \mu_1 + \mu_2 & \text{for } x = 1 \text{ and } y = z, \\ \mu_1\alpha & \text{for } x = 0 \text{ and } y = z - 1 \geq 1, \\ (\mu_1 + \mu_2(1-\gamma))\alpha & \text{for } x = -1 \text{ and } y = z = 1 \\ \mu_2(1-\gamma)\alpha & \text{for } x = -1 \text{ and } y = z > 1 \\ \mu_2\gamma^m(1-\gamma)\alpha & \text{for } y - z = m > 0 \text{ and } x = -(m+1) \end{cases}$$

We refer to this Markov additive process with kernel $\mathcal{K}$ as the *twisted free process*. The transition diagram is simply a reweighting of the arcs in Figure 2. Let $\mathcal{N}_\blacktriangle(\ell,y)$ denote the number of visits to (ℓ,y) by the twisted free process. As in [3], $\mathrm{E}_{(x,z)}[N_\blacktriangle(\ell,y)] = (h(x,z)/h(\ell,y))\mathrm{E}_{(x,z)}[\mathcal{N}_\blacktriangle(\ell,y)]$; hence, for $(\ell,y) \in S \setminus \Delta$, we can rewrite (10) as

$$\pi(\ell,y)h(\ell,y) = \sum_{(x,z)\in\Delta} \pi(x,z)h(x,z)\mathrm{E}_{(x,z)}[\mathcal{N}_\blacktriangle(\ell,y)], \tag{11}$$

where we are taking advantage of the fact that transitions from Δ to $S \setminus \Delta$ also follow K^∞.

It will be important to know whether the Markovian part of the twisted free process is positive recurrent or not. From Foster's criteria, we can simply check whether the vertical drift for any state $(0,z)$ with $z > 1$ is negative or not. The only downward jumps are to state $(0,z-1)$ with probability $\mu_1\alpha$. The process jumps up $m > 0$ levels with probability $\mu_2\gamma^m(1-\gamma)\alpha$. The expected change in the second component is $\sum_{m=1}^\infty m\mu_2\gamma^m(1-\gamma)\alpha - \mu_1\alpha$, which simplifies to $(\mu_2\lambda_b/\lambda_a - \mu_1)\alpha$. Thus, the process is positive recurrent if and only if $\lambda_b/\mu_1 < \lambda_a/\mu_2$; that is, if and only if the load on machine m_2 from type a jobs is greater than the load on server one from type b jobs alone. This jitter condition along the x-axis can also be expressed as $\beta < \alpha$ or as $d_2^* < 0$. Consequently, we need to split the analysis into two cases. In the remainder of this section, we assume that the jitter condition along the x-axis holds. The other case will be handled in Section 4.

Since the Markovian part is positive recurrent, the twisted free process will tend to hug or jitter near the bottom of the state space. Let $\varphi = \varphi_v$ denote the stationary distribution of the Markovian part of the twisted free

process. Solving for φ yields

$$\varphi(y) = \begin{cases} \frac{\lambda_a/\mu_2 - \lambda_b/\mu_1}{\lambda_b/\mu_1} & \text{for } y = 1 \\ \frac{\lambda_a/\mu_2 - \lambda_b/\mu_1}{\lambda_b/\mu_1} \frac{\mu_2}{\mu_1+\mu_2} \left(\frac{\beta}{\alpha}\right)^{y-1} & \text{for } y = 2, 3, \ldots \end{cases}$$

Next we compute the stationary horizontal drift of the twisted free process. Let $\tilde{d}_v$ be the stationary horizontal drift of the twisted free process. Conditioned on the Markovian part being in state y, the horizontal drift is almost independent of y except for a slight extra probability of going one step to the left when $y = 1$. Thus,

$$\begin{aligned} \tilde{d}_v &= \mu_1 + \mu_2 - \sum_{m=1}^{\infty} m\mu_2(1-\gamma)\gamma^{m-1}\alpha - \varphi(1)\mu_1\alpha \\ &= \mu_1 + \mu_2 - \lambda_a - \lambda_b, \end{aligned}$$

which is greater than 0 so the twisted free process drifts to the right, while bouncing along the bottom of the state space. Kesten's Theorem 2 in [7] suggests that as $\ell \to \infty$, the expected number of visits to (ℓ, y) by the (aperiodic) twisted free process starting from any fixed state (x, z) converges to the intuitively reasonable quantity $\varphi(y)/\tilde{d}_v$. Kesten's Theorem 2 has a non-lattice condition (I.3) on the additive part, which does not hold in our example. However, Kesten's coupling argument is simpler in this discrete case; see also Lemma 5 in [3].

Let $\mathcal{H}_v(x, z)$ be the probability that the twisted free process starting from (x, z) as defined in this section never hits $\blacktriangle$. As in the proof of Lemma 1.2 in [9], escaping from (x, z) and the number of visits to (ℓ, y) are asymptotically independent, so $\mathrm{E}_{(x,z)}[\mathcal{N}_{\blacktriangle}(\ell, y)] \to \mathcal{H}_v(x, z)\varphi(y)/\tilde{d}_v$. Since $\beta < \alpha$, Proposition 1 implies that $\pi h 1_{\Delta} < \infty$; furthermore, since the expectation is bounded,

$$\sum_{(x,z)\in\Delta} \pi(x, z)h(x, z)\mathrm{E}_{(x,z)}[\mathcal{N}_{\blacktriangle}(\ell, y)] \to c_v \frac{\varphi(y)}{\tilde{d}_v} \tag{12}$$

where

$$\begin{aligned} c_v &\equiv \sum_{(x,z)\in\Delta} \pi(x, z)h(x, z)\mathcal{H}_v(x, z) \\ &= \sum_{z=1}^{\infty} \pi(1, z)\alpha^{-z}\mathcal{H}_v(1, z) \end{aligned} \tag{13}$$

since $\mathcal{H}_v(x,z) > 0$ only if $x = 1$ and $z > 0$. Note that $0 < c_v < \infty$. Hence, the r.h.s. of (12) is a finite, positive function of y, and from (11), we have the asymptotic result

$$\pi(\ell, y) \sim \frac{c_v \varphi_v(y)}{\tilde{d}_v h_v(\ell, y)}$$

Putting in our expressions for $\tilde{d}_v$, φ_v, and h_v shows that $\pi(\ell, y) \sim \psi_v(\ell, y)$, which is (5).

REMARK 1. *Showing that $\sum_{(x,z)\in\Delta} \pi(x,z)h(x,z)$ is finite can be the difficult part in many applications; see pp. 597—600 or 604—606 of [3] for examples. For the model analyzed in this paper, the bounds on π made the condition easy to verify.*

4. Cascades, bridges and a new approach to deriving exact asymptotics. At this point, we could derive analogous results for $\pi(x, \ell)$ when the jitter condition holds along the y-axis. This would entail redefining Δ and all related quantities. Instead, we delay the analogous analysis to Section 5, leave the definition of $\Delta = \Delta_v$ unchanged, and pursue asymptotics for $\pi(\ell, y)$ when the jitter condition along the x-axis fails to hold. This requires an alternative approach since when $d_2^* > 0$, the important paths to (ℓ, y) turn out to be "cascades" where the initial segment "jitters" up the y-axis to a height roughly proportional to ℓ before turning and heading in a southeasterly direction to (ℓ, y).

To introduce the alternative approach, let us re-obtain the exact asymptotic results for $\pi(\ell, y)$ in the jitter case of Section 3 when $d_2^* < 0$; that is, when the important paths jitter near the x-axis. Then we modify this alternative approach to handle the cascade (and bridge) cases, which avoid the x-axis. The previous approach studied the twisted process starting from Δ as it wandered out to the rare event (ℓ, y). The essence of the new approach is that instead of studying the twisted free process starting on Δ, we study the *time reversal* of the twisted free process *starting from the rare event* (ℓ, y).

4.1. *An alternative derivation of Proposition 2.* Temporarily, assume that $\beta < \alpha$ so that φ, the stationary distribution of the twisted free process, exists. The definitions of Δ, K^∞, h and $\mathcal{K}$ are the same as in Section 3. Let $\mathcal{X} \equiv \mathcal{X}(0), \mathcal{X}(0), \ldots$ be the twisted free process, which has transition kernel $\mathcal{K}$. Let $\overleftarrow{\mathcal{X}}$ be the time reversal of the twisted free process with respect to φ. The process $\overleftarrow{\mathcal{X}}$ is also a Markov additive process with kernel $\overleftarrow{\mathcal{K}}$ that satisfies

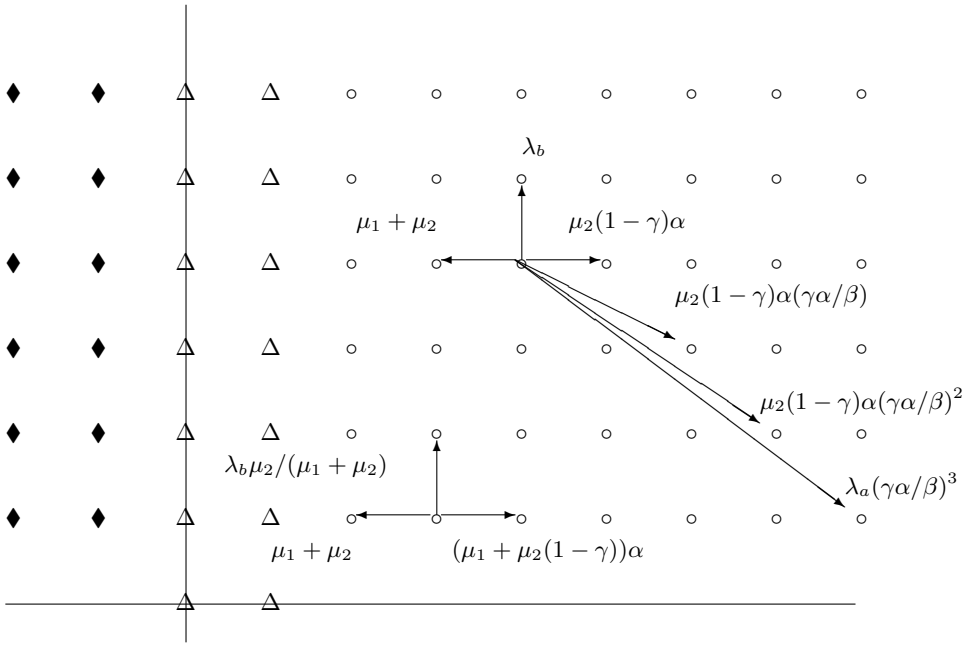

FIG 3. *The time reversal of the twisted free process (lowest left* Δ *is* $(0,0)$*).*

the relationship

$$\varphi(z)\mathcal{K}((0,z);(x,y)) = \varphi(y)\overleftarrow{\mathcal{K}}((0,y);(-x,z)) \tag{14}$$

The transition structure is shown in Figure 3. Notice that the long jumps are towards the southeast; however, they are no longer unbounded. Instead, they are truncated by the bottom of the state space.

If we assume that the twisted free process has initial distribution π, (11) can be written as

$$\pi(\ell,y)h(\ell,y) = \mathrm{E}\left[1_\Delta(\mathcal{X}(0))h(\mathcal{X}(0))\mathcal{N}_\blacktriangle(\ell,y)\right] \quad \text{for any } (\ell,y)\in S\setminus\Delta \tag{15}$$

Let $\mathcal{T}$ be the number of steps until $\overleftarrow{\mathcal{X}}$ hits $\blacktriangle$ at $\overleftarrow{\mathcal{X}}(\mathcal{T}) \equiv (\overleftarrow{\mathcal{X}}_1(\mathcal{T})), \overleftarrow{\mathcal{X}}_2(\mathcal{T}) = (1, \overleftarrow{\mathcal{X}}_2(\mathcal{T}))$. By looking at the time reversal, (15) can be written as

$$\pi(\ell,y)\frac{h(\ell,y)}{\varphi(y)} = \mathrm{E}_{(\ell,y)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] \tag{16}$$

where $1_\Delta(x) \equiv 1_{\{x\in\Delta\}}$; see the appendix for a more detailed justification of (16). If the r.h.s. of (16) converges to a finite positive constant as ℓ tends

to infinity, then we have an exact asymptotic expression for $\pi(\ell, y)$. Now we investigate the convergence of the r.h.s. under different conditions.

When the Markovian part of the Markov additive process is positive recurrent with stationary distribution φ, the time reversal drifts west and hits ▲. The limiting distribution as $\ell \to \infty$ of the hitting location on ▲ can be obtained from Kesten's Theorem 1 of [7]; see also Proposition 2.4 in [9] and Appendix A of [1]. For our situation, Kesten's result simplifies to

$$\Pr{}_{(\ell,y)}\left\{\overleftarrow{\mathcal{X}}(\mathcal{T}) = (x,z)\right\} \to \frac{\varphi(z)\mathcal{H}_v(x,z)}{\tilde{d}_v} \tag{17}$$

where the probability $\mathcal{H}_v(x,z)$ is the probability the time reversal of $\overleftarrow{\mathcal{X}}$ (which is just $\mathcal{X}$) leaving from (x,z) never returns to ▲. If we can justify the convergence in the following,

$$\mathrm{E}_{(\ell,y)}\left[1_{\Delta_v}\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_v(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_v(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] \to \sum_{(x,z)\in\Delta} \frac{\varphi(z)\mathcal{H}_v(x,z)}{\tilde{d}_v}\pi(x,z)\frac{h_v(x,z)}{\varphi_v(z)} \tag{18}$$

$$= c_v/\tilde{d}_v, \tag{19}$$

then we have an alterative derivation of (12) and Proposition 2.

To justify the convergence in (18), the l.h.s. can be expressed as $\mathrm{E}_{(0,y)}[g_v(X^{\#}(\ell))]$ where $g_v(x,z) \equiv 1_{\Delta_v}(x,z)\pi(x,z)h_v(x,z)/\varphi_v(z)$ and $X^{\#}$ is Kesten's overshoot Markov chain (see between (3.2) and (3.3) of [7] or Section 2.2 of [9]), which has stationary distribution given by the r.h.s. of (17). The convergence follows if g_v is integrable with respect to the stationary distribution of $X^{\#}$ (e.g., Theorem 14.0.1 of [10]), but the integrability follows from $\pi h 1_{\Delta} < \infty$.

4.2. *Cascades, bridges, and a proof of Proposition 4.* Now we modify the argument in the previous subsection to handle the non-jitter cases. Assume that $d_2^* > 0$. The first difficulty appears to be that the Markovian part of the twisted free process does not have a stationary distribution φ. However, the Markovian part does possess an invariant measure that is unique up to rescaling. Re-define $\varphi = \varphi_v$ to be the invariant measure

$$\varphi_v(y) = \begin{cases} 1 & \text{for } y = 1 \\ \frac{\mu_2}{\mu_1+\mu_2}\left(\frac{\beta}{\alpha}\right)^{y-1} & \text{for } y = 2,3,\ldots \end{cases}$$

This invariant measure φ_v can be used in (14) to define the kernel for the time reversal of the twisted free process. The argument continues without changes until just after (16). Under the jitter conditions, the location

$\overleftarrow{\mathcal{X}}(\mathcal{T})$ where the time reversal hits Δ_v converged to a proper distribution as $\ell \to \infty$. However, when $d_2^* > 0$, the Markovian part of the twisted free process is not stable, and the time reversal of the twisted free process drifts northwesterly. To see this, compute the drift ignoring the truncation along the x-axis (or compute the expected drift from state (x, y) as $y \to \infty$), to obtain d^* as given in (4). Since $1 > \beta > \alpha$, d_1^* is negative, and d_2^* is positive. Including the truncation would push the process even more northwesterly. Consequently, conditioned on starting in (ℓ, y), the hitting location $\overleftarrow{\mathcal{X}}(\mathcal{T}) \equiv (\overleftarrow{\mathcal{X}}_1(\mathcal{T}), \overleftarrow{\mathcal{X}}_2(\mathcal{T})) = (1, \overleftarrow{\mathcal{X}}_2(\mathcal{T}))) \to (1, \infty)$ a.s. as $\ell \to \infty$.

The fact that $\overleftarrow{\mathcal{X}}(\mathcal{T})$ diverges would seem to sound the death knell for (16) converging to a finite positive constant. However, just when all appears lost, note that $\pi(1,\ell)h(1,\ell)/\varphi(\ell) = \frac{(\mu_1+\mu_2)}{\mu_2}\frac{\beta}{\alpha}\pi(1,\ell)\beta^{-\ell}$ for $\ell > 0$; furthermore, the hypothesis of Proposition 3 holds, so $\pi(1,\ell)\beta^{-\ell}$ converges to the constant $\psi_w(1,0)$. (Even though the proof of Proposition 3 appears in a later section, the proof does not rely on results from this section. Also, even though the convergence of $\pi(1,\ell)h_v(1,\ell)/\varphi_v(\ell)$ appears to be a fluke arising in this example, [6] shows that this property holds much more generally.) Since the integrand is bounded,

$$\begin{aligned} \mathrm{E}_{(\ell,y)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_v(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_v(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] &\to \frac{(\mu_1+\mu_2)}{\mu_2}\frac{\beta}{\alpha}\psi_w(1,0) \\ &= c_w\frac{\beta}{\alpha}\frac{1}{\mu_1-\lambda_b}\left(1-\frac{\lambda_a}{\mu_2}\right)\frac{\lambda_a}{\mu_2}, \end{aligned} \tag{20}$$

which is enough to determine the exact asymptotics of $\pi(\ell, y)$ when $\beta > \alpha$. However, rather than giving the exact asymptotics of $\pi(\ell, y)$, we extend the argument in two different ways: first to include sequences of states away from the axes, and second to include $\beta = \alpha$.

To extend the argument to include sequences of states away from the x-axis, use (37) to write

$$\pi(x_\ell, y_\ell)\frac{h_v(x_\ell, y_\ell)}{\varphi_v(y_\ell)} = \mathrm{E}_{(x_\ell,y_\ell)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_v(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_v(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] \tag{21}$$

where (x_ℓ, y_ℓ) are states in $S \setminus \Delta$ with (x_ℓ, y_ℓ) where $x_\ell \to \infty$. Now, conditioned on starting in (x_ℓ, y_ℓ), we still have the hitting location $\overleftarrow{\mathcal{X}}(\mathcal{T}) \to (1, \infty)$ a.s. as $\ell \to \infty$ so that we still have

$$\mathrm{E}_{(x_\ell,y_\ell)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_v(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_v(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] \to \frac{(\mu_1+\mu_2)}{\mu_2}\frac{\beta}{\alpha}\psi_w(1,0). \tag{22}$$

When $\beta = \alpha$, the Markovian part is not positive recurrent, and we are in a bridge case. However, if the Markovian part is not positive recurrent, then $(1, \overleftarrow{\mathcal{X}}_2(\mathcal{T}))) \to (1, \infty)$ in probability, which is enough to obtain (20). Thus, we have that

$$\pi(x_\ell, y_\ell) \sim \frac{(\mu_1 + \mu_2)}{\mu_2} \frac{\beta}{\alpha} \psi_w(1,0) \frac{\varphi(y_\ell)}{h(x_\ell, y_\ell)} = \chi_v(x_\ell, y_\ell).$$

This completes the proof of Proposition 4

5. A proof of Proposition 3. We derive the exact asymptotics of $\pi(x, \ell)$ under jitter conditions along the y-axis using the same approach as in Section 3. The conditions needed for a jitter along the y-axis is $d_1^* < 0$. Again, we work with the discrete time Markov chain $X_0, X_1, \dots$ with transition kernel K obtained by uniformizing the continuous time Markov process. We re-define many of the terms introduced in the previous section including Δ, K^∞, $\blacktriangle$, h, $\mathcal{K}$, and φ. When we want to emphasize that a quantity is based on the definitions given in this section, we add a subscript "w".

Since we are interested in large deviations in the second coordinate, we look for a Markov additive process where the first coordinate is the Markovian part and the second coordinate is the additive part. Let the boundary $\Delta = \Delta_w = \{(x, y) \in S : y \leq 1\}$. (The subscript w comes from Δ in this section looks like a *wide* set.) Let K^∞ denote the transition kernel of the Markov additive process. Define $K^\infty((z, m); (x, y + m)) = K^\infty((z, 0); (x, y))$ where

$$K^\infty((z,0);(x,y)) = \begin{cases} \lambda_a + \lambda_b & \text{for } z > 0,\ x = z+1 \text{ and } y = 0, \\ \mu_1 & \text{for } x = z \text{ and } y = -1, \\ \mu_2(1-\gamma) & \text{for } z > 1,\ x = z-1 \text{ and } y = 0, \\ \mu_2 & \text{for } z = 1,\ x = z-1 \text{ and } y = 0 \\ \mu_2 & \text{for } z = x = y = 0, \\ \lambda_b & \text{for } z = 0,\ x = 0,\ y = 1, \\ \lambda_a & \text{for } z = 0,\ x = 1,\ y = 0, \\ \mu_2\gamma^y(1-\gamma) & \text{for } 0 < x = z - (y+1),\ y = 1, 2, \dots, z-2, \\ \mu_2\gamma^y & \text{for } z > 1,\ x = 0,\ y = z-1, \end{cases}$$

Again, we refer to the Markov additive process with kernel K^∞ as the *free process*, since we have removed the boundary Δ. In this case, the process is free to wander over the right two quadrants $\mathbb{Z}_+ \times \mathbb{Z}$ where $\mathbb{Z}_+ \equiv \{0, 1, 2, \dots\}$. Let $\blacktriangle = \{(x, y) : x \geq 0, y \leq 1\}$, which is Δ and all of the new states for the

free process. Since K is positive recurrent, we would expect the free process to be transient.

The next step is to find a harmonic function $h = h_w$ for the transition kernel K^∞ of the form $h(x,y) = \hat{h}(x)b^y$. From Corollary 1, the rough asymptotics for states (x,ℓ) are given by β; hence, $1/b = \beta$. Let $\hat{h}(0) = 1$. If we insert our guess into $K^\infty h = h$, we discover that

$$\hat{h}(x) = \begin{cases} 1 & \text{for } x = 0, \\ \left(\frac{\lambda_a+\mu_1}{\lambda_a+\lambda_b}\right)^{x-1} & \text{for } x = 1, 2, \ldots \end{cases} \tag{23}$$

We use the harmonic function $h(x,y) = \hat{h}(x)\beta^{-y}$ to define the h-transform or twisted kernel $\mathcal{K}((0,z);(x,y)) \equiv K^\infty((z,0);(x,y))h(x,y)/h(z,0)$ yielding

$$\mathcal{K}((z,0);(x,y)) = \begin{cases} \lambda_a + \mu_1 & \text{for } z > 0,\ x = z+1 \text{ and } y = 0, \\ \lambda_b & \text{for } x = z \text{ and } y = -1, \\ \mu_2\lambda_a/(\lambda_a+\mu_1) & \text{for } z > 1, x = z-1 \text{ and } y = 0, \\ \mu_2 & \text{for } z = 1,\ x = 0 \text{ and } y = 0 \\ \mu_2 & \text{for } z = x = y = 0, \\ \mu_1 & \text{for } z = 0,\ x = 0,\ y = 1, \\ \lambda_a & \text{for } z = 0,\ x = 1,\ y = 0, \\ \mu_2 \frac{\lambda_a}{\lambda_a+\mu_1}\left(\frac{\mu_1}{\lambda_a+\mu_1}\right)^y & \text{for } 0 < x = z-(y+1),\ y = 1, 2, \ldots, z-2, \\ \mu_2\left(\frac{\mu_1}{\lambda_a+\mu_1}\right)^y & \text{for } z > 1,\ x = 0,\ y = z-1, \end{cases}$$

If $\lambda_a/\mu_2 < 1$, then the stationary distribution φ_w for the Markovian part of the twisted free process exists and is given by

$$\varphi_w(x) = \begin{cases} \left(1-\frac{\lambda_a}{\mu_2}\right) & \text{for } x = 0, \\ \left(1-\frac{\lambda_a}{\mu_2}\right)\frac{\lambda_a}{\mu_1+\mu_2}\left(\frac{\lambda_a+\mu_1}{\mu_1+\mu_2}\right)^{x-1} & \text{for } x = 1, 2, \ldots \end{cases} \tag{24}$$

When the Markovian part is positive recurrent and in steady state, the vertical drift of the process is

$$\begin{aligned} \tilde{d}_w &= -\lambda_b + \mu_1\varphi(0) + \sum_{x=2}^{\infty}\varphi(x)\sum_{n=1}^{x-2} n\mu_2\frac{\lambda_a}{\lambda_a+\mu_1}\left(\frac{\mu_1}{\lambda_a+\mu_1}\right)^n + (x-1)\mu_2\left(\frac{\mu_1}{\lambda_a+\mu_1}\right)^{x-1} \\ &= \mu_1 - \lambda_b \\ &> 0, \end{aligned}$$

which makes sense since the twisted free process, in essence, has interchanged two rates: the arrival rate of type b customers with the service rate at machine m_1.

Starting from Orey's representation of π and using arguments similar to those in Section 3, we end up with

$$\pi(x,\ell)h_w(x,\ell) = \sum_{(z,y)\in\Delta} \pi(z,y)h_w(z,y)\mathrm{E}_{(z,y)}[\mathcal{N}_{\blacktriangle}(x,\ell)] \tag{25}$$
$$\to c_w \frac{\varphi_w(x)}{\tilde{d}_w}$$

where

$$c_w \equiv \sum_{(z,y)\in\Delta_w} \pi(z,y)h_w(z,y)\mathcal{H}_w(z,y)$$
$$= \sum_{z=0}^{\infty} \pi(z,1)\beta^{-z}\mathcal{H}_w(z,1). \tag{26}$$

and $\mathcal{H}_w(z,y)$ is the probability that the twisted free process *as defined in this section* starting from (z,y) never returns to $\blacktriangle$. The "similar arguments" need $\pi h_w 1_{\Delta_w} < \infty$, which follows from the bounds on π given in (3).

Using the bounds on π and noting that $\mathcal{H}_w(z,1)$ is strictly positive for $z \neq 1$, it follows that c_w is a finite, strictly positive constant. By putting in our expressions for $\tilde{d}_w$, h and φ, we obtain

$$\pi(x,\ell) \sim \frac{c_w\,\varphi_w(x)}{\tilde{d}_w\,h_w(x,\ell)} = \psi_w(x,\ell)$$

This completes the proof of Proposition 3.

6. Using the time reversal to prove Proposition 5. We use the new approach to derive the exact asymptotics of $\pi(x_\ell, y_\ell)$ when the jitter condition along the y-axis does not hold; thus, we assume that $\lambda_a/\mu_2 \geq 1$ or equivalently $d_1^* \geq 0$. For stability the jitter condition along the x-axis must hold; that is, $d_2^* < 0$.

The definitions of $\Delta = \Delta_w$, K^∞, $\blacktriangle$, h and $\mathcal{K}$ are still those of Section 5. Since the Markovian part of the twisted free process does not have a stationary distribution, we re-define $\varphi = \varphi_w$ to be the invariant measure

$$\varphi_w(x) = \begin{cases} 1 & \text{for } x = 0, \\ \frac{\lambda_a}{\mu_1+\mu_2}\left(\frac{\lambda_a+\mu_1}{\mu_1+\mu_2}\right)^{x-1} & \text{for } x = 1, 2, \ldots, \end{cases}$$

Using φ_w we define the transition kernel $\overleftarrow{\mathcal{K}}$ of the time reversal of the twisted free process from the relationship

$$\varphi_w(x)\mathcal{K}((x,y);(z,0)) = \varphi_w(z)\overleftarrow{\mathcal{K}}((z,0);(x,y))$$

This time reversal is remarkably similar to the time reversal depicted in Figure 3 (though Δ is different and this time reversal lives in the right two quadrants instead of the upper two quadrants). For example, the transition probabilities out of state (4,4) are identical, except that there is no truncation of the geometric distribution of jumps to the southeast. Hence, the drift vector from (4,4) and any state (x,y) with $x > 2$ is the same as (4), which means that the time reversal will be drifting southeasterly in the direction d^* when $x > 2$. In our production example, d^* turns out to be the direction of drift for both time reversals asymptotically as their Markovian parts become large. This similarity does not hold in general. In other examples, the two may have different asymptotic drifts as discussed in [6].

From (25) and Lemma 6, we can obtain the analog of (21). Thus, for $(x_\ell, y_\ell) \in S \setminus \Delta$

$$\pi(x_\ell,y_\ell)\frac{h_w(x_\ell,y_\ell)}{\varphi_w(x_\ell)} = \mathrm{E}_{(x_\ell,y_\ell)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_w(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_w(\overleftarrow{\mathcal{X}}_1(\mathcal{T}))}\right] \tag{27}$$

There is one wrinkle to using the time reversal of the twisted free process defined in Section 4 that we did not encounter previously. When leaving $S \setminus \Delta$, the time reversal may leap completely over Δ landing in $\blacktriangle \setminus \Delta$ at $\overleftarrow{\mathcal{X}}(\mathcal{T})$. Now note that

$$\frac{h_w(x,y)}{\varphi_w(x)} = \begin{cases} \beta^{-y} & \text{for } x = 0, \\ \frac{\mu_1+\mu_2}{\lambda_a}\alpha^{-(x-1)}\beta^{-y} & \text{for } x = 1,2,\ldots, \end{cases}$$

Hence, for $\ell > 0$, we have from Proposition 2

$$\pi(\ell,1)\frac{h_w(\ell,1)}{\varphi_w(\ell)} = \frac{\lambda_a+\lambda_b}{\beta\lambda_a}\pi(\ell,1)\alpha^{-\ell} \tag{28}$$

$$\to \frac{\lambda_a+\lambda_b}{\beta\lambda_a}\psi_v(0,1) \tag{29}$$

The appropriate condition for the Markovian part $\overleftarrow{\mathcal{X}}_1(\mathcal{T})$ to diverge (at least in probability) is that $d_1^* \geq 0$.

To handle the added wrinkle that $\overleftarrow{\mathcal{X}}(\mathcal{T})$ might land in $\blacktriangle \setminus \Delta$, we need the $\Pr_{(x_\ell,y_\ell)}\left\{\overleftarrow{\mathcal{X}}_2(\mathcal{T}) = 1\right\}$, which is the probability of landing on Δ starting

from (x_ℓ, y_ℓ). There is more than one way to compute this quantity. The one-step transition probabilities of the time reversed twisted free process jumping downwards an amount $y > 0$ is

$$\lambda_a \left(\frac{\mu_1}{\mu_1+\mu_2}\right)^y \frac{\mu_2}{\mu_1+\mu_2}$$

Thus, conditioned on jumping downwards, the distance jumped has a geometric distribution. Hence, the first time the process goes below 2, the probability of stopping at 1 is $\mu_2/(\mu_1+\mu_2)$. Another way of computing this quantity is to use the expression for (27) applied to the already derived asymptotic expression for $\pi(\ell, y)$ with $y > 1$ given in Proposition 2.

Let (x_ℓ, y_ℓ) be a sequence of states in S such that $y_\ell > 1$ and $x_\ell + y_\ell \to \infty$. For such a sequence,

$$\begin{aligned}\mathrm{E}_{(x_\ell,y_\ell)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_w(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_w(\overleftarrow{\mathcal{X}}_1(\mathcal{T}))}\right] &\to \frac{\lambda_a+\lambda_b}{\beta\lambda_a}\psi_v(0,1)\frac{\mu_2}{\mu_1+\mu_2} \\ &= c_v\frac{\alpha\mu_2}{\beta\lambda_a}\frac{1}{\mu_1+\mu_2-\lambda_a-\lambda_b}\frac{\lambda_a/\mu_2-\lambda_b/\mu_1}{\lambda_b/\mu_1},\end{aligned}$$

which is a finite positive constant; Proposition 5 follows from knowing this constant and (27).

7. Asymptotics of $\pi(x_\ell, y_\ell)$ when jitter conditions hold on both axes and a proof of Propostion 6. To complete the description of the asymptotics of π, we need to consider the asymptotics when $d_1^* < 0$ and $d_2^* < 0$, which is assumed to hold in this section. Under these conditions, Propositions 2 and 3 already give the asymptotics near the axes. Hence, we only need consider sequences of states (x_ℓ, y_ℓ) where both $x_\ell \to \infty$ and $y_\ell \to \infty$. To obtain asymptotics away from the axes, we need to use one of the time reversals. Let us call the time reversal studied in Subsection 4.1 where $\Delta = \Delta_v$ was a vertical strip near the y-axis as the *v-time reversal*; the time reversal where $\Delta = \Delta_w$ was a horizontal strip near the x-axis will be the *w-time reversal*. Since the jitter condition holds for both time reversals, we assume that φ_v and φ_w are the stationary distribution of the Markovian part of the v-time reversal and w-time reversal, respectively.

Let τ be the time that a time reversal exits the transient set $\{2, 3, \dots\}^2$. At time τ, the time reversal can hit either $V = \{1\} \times \{2, 3, \dots\}$ or $W = \{(i, j) : j \le 1, i + j \ge 3\}$. Let q_ℓ be the probability that the time reversal starting from $(x_\ell, y_\ell) \in \{2, 3, \dots\}^2$ hits V at time τ. The probability does not depend upon which of the two time reversals we choose. Furthermore,

the probability of hitting any particular state in V is the same for both time reversals. This latter property does not hold on states in W since the southeastern jumps are truncated for the v-time reversal, but not for the w-time reversal. The next lemma says that the time reversal is far from the origin at time τ.

LEMMA 2. *Let $\overleftarrow{\mathcal{X}}$ be either time reversal of either twisted free process starting from state $(x_\ell, y_\ell) \in \{2, 3, \dots\}^2$ where $x_\ell + y_\ell \to \infty$. Then* $\Pr_{(x_\ell,y_\ell)}\left\{\max\{\overleftarrow{\mathcal{X}}_1(\tau), \overleftarrow{\mathcal{X}}_2(\tau)\} \le k\right\} \to 0$ *for all k.*

PROOF. Consider a random walk that uses the same transition probabilities as the w-time reversal when the time reversal is in some state $(z, 0)$ with $z > 1$. Suppose the random walk starts in state (x_ℓ, y_ℓ). Let $\tau(n)$ be the first time that the random walk hits the line with elements (w, z) where $w + z = x_\ell + y_\ell - n$; that is, the total has decreased by n. Note that the random walk cannot jump over this line. Also note that the position on this line is the sum of n i.i.d. displacements where the mean can be determined from the drift vector d^*. Lastly, note that the probability that a time reversals hits within k of the origin is smaller than the probability that the random walk at time $\tau(x_\ell + y_\ell - k)$ is between $(k, 0)$ and $(0, k)$. If the variance of the displacements is finite, then this probability goes to zero as $\ell \to \infty$ by the central limit theorem. If the variance is infinite, this probability also goes to zero, which completes the argument. □

Since the process is far from the origin at time τ, the r.h.s. of (21) will be a mixture of two cases depending on which coordinate of the time reversal is big at time τ. The next lemma merely gives the mixture under the assumption that q_ℓ converges. We delay determining $\lim q_\ell$ until Lemmas 4 and 5.

LEMMA 3. *If $q_\ell \to q$, then*

(30)
$$\mathrm{E}_{(x_\ell,y_\ell)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h_v(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi_v(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right] \to q\frac{(\mu_1+\mu_2)}{\mu_2}\frac{\beta}{\alpha}\psi_w(1,0)\frac{1}{\varphi(1)} + (1-q)\frac{c_v}{\tilde{d}_v}$$

PROOF. Consider the v-time reversal; thus, we are interested in the asymptotic hitting distribution on Δ_v as $\ell \to \infty$ assuming that the process started at (x_ℓ, y_ℓ). The v-process drifts southwest. From Lemma 2, the probability that the v-process is close to the origin at time τ goes to zero as $\ell \to \infty$. The v-process will either be in Δ far above the origin with probability converging to q, or with probability converging to $(1 - q)$ at some distant state

$(\ell', 1)$. In both cases, we can compute the r.h.s. of (21). In the former case, we have (22) except that φ is the invariant distribution; in the latter case, the process jitters in giving (19), which completes the proof. □

To finish the proof of Proposition 6, we need to investigate $\lim q_\ell$. The following lemma will show that if (x_ℓ, y_ℓ) asymptotically lies above the drift line, then the time reversal will hit Δ_v far above the origin, which means $q_\ell \to 1$. In this case, the r.h.s. of (20) behaves as in the cascade case analyzed in Subsection 4.2. On the other hand, if (x_ℓ, y_ℓ) asymptotically lies below the drift line, then the process hits some distant point of W near the x-axis ($q_\ell \to 0$) and then jitters along the lower edge of the state space until hitting Δ_v near the origin, which can be analyzed as in (18). Lemma 5 considers the delicate case when (x_ℓ, y_ℓ) asymptotically lies on the drift line, which can be a mixture of the previous two cases. The mixing probability will be the $\lim q_\ell$, provided the limit exists.

LEMMA 4. *Assume $d_1^* < 0$ and $d_2^* < 0$. Let (x_ℓ, y_ℓ) be any sequence of states with $x_\ell > 1$ and $(x_\ell/\ell, y_\ell/\ell) \to (\bar{x}, \bar{y})$ where $0 < \max(\bar{x}, \bar{y}) < \infty$. If $\bar{y}/\bar{x} > d_2^*/d_1^*$, then $q_\ell \to 1$. If $\bar{y}/\bar{x} < d_2^*/d_1^*$, then $q_\ell \to 0$.*

LEMMA 5. *Assume $d_1^* < 0$ and $d_2^* < 0$. Let (x_ℓ, y_ℓ) be any sequence of states with $(x_\ell/\ell, y_\ell/\ell) \to (\bar{x}, \bar{y})$ where $0 < \max(\bar{x}, \bar{y}) < \infty$. If there exists (r, s) such that*

$$\sqrt{\ell}\left[\left(\frac{x_\ell}{\ell}, \frac{y_\ell}{\ell}\right) - \left(\frac{d_1^*}{d_1^* + d_2^*}, \frac{d_2^*}{d_1^* + d_2^*}\right)\right] \to (r, s)$$

then $q_\ell \to \Phi[-(s - r d_2^*/d_1^*)/\sigma]$, Φ is the c.d.f. of a standard normal distribution, and σ is defined in (32).

The proofs of Lemmas 4 and 5 will rely on Sections 2 and 4 in Chapter 11 of Ethier and Kurtz [2]. To use these results, we closely follow [2] by considering a family of continuous time Markov processes

$$\hat{Y}_\ell(t) = \hat{Y}_\ell(0) + \sum_{i=1}^{\infty} v_i N_i(\ell \beta_i t)$$

where $\hat{Y}_\ell(0) = (x_\ell - 1, y_\ell - 1)$, $v_1 = (-1, 0)$, $v_2 = (0, 1)$, $v_3 = (1, 0)$, $v_4 = (2, -1), \ldots$, the N_i's are independent Poisson processes with rate 1, $\beta_1 = (\mu_1 + \mu_2)$, $\beta_2 = \lambda_b$, and $\beta_k = \mu_2(1-\gamma)\alpha(\gamma\alpha/\beta)^{k-3}$ for $k = 3, 4, \ldots$. A glance at Figure 3 should help to motivate the definitions of the v_i's and β_i's. The

process $\hat{Y}_\ell(\cdot)$ is a continuous time Markov process with transitions that occur at rate ℓ. For the first $\tau-1$ jumps, the continuous time process $\hat{Y}_\ell(\cdot)$ and the discrete time process $\overleftarrow{\mathcal{X}}(\cdot)$ starting from (x_ℓ, y_ℓ) can be coupled so that the jump sizes (one of the v_k's) are identical. Thus, until exiting, when the latter is in state (x,y), the former is in state $(x-1,y-1)$. The reason for shifting the state is so that the exit time τ of $\overleftarrow{\mathcal{X}}(\cdot)$ from $\{2,3,\dots\}^2$ can be expressed as the exit time from $\{1,2,3,\dots\}^2$, which will make things slightly nicer. The coupling can also be constructed so that if $\hat{Y}_\ell(\cdot)$ hits state (x,y) when exiting $\{1,2,3,\dots\}^2$, then

$$\overleftarrow{\mathcal{X}}(\tau) = \begin{cases} (x+1,y+1) & \text{for } x \le 0, y > 0 \\ (x+1+y,1) & \text{for } x > 0, y \le 0. \end{cases}$$

This follows from the fact that westward exiting jump can only be of size $v_1 = (-1,0)$ but southward exiting jumps of $\hat{Y}_\ell(\cdot)$ can be quite large and may need to be truncated.

Note that $\sum_{i>0} v_i\beta_i = d^*$. Let $Y_\ell(t) = \hat{Y}_\ell(t)/\ell$. Thus,

$$Y_\ell(t) = Y_\ell(0) + \sum_{i>0} \frac{v_i}{\ell} \tilde{N}_i(\ell\beta_i t) + d^* t,$$

where $\tilde{N}_i(t) \equiv N_i(t) - t$ is the Poisson process centered at its mean. Let

$$\tau_\ell \equiv \inf\{t > 0 : \min(Y_\ell(t)) \le 0\}.$$

Hence, q_ℓ is the probability that the first coordinate of $Y_\ell(\tau_\ell)$ is zero. From Theorem 2.1 of [2],

$$\lim_{\ell\to\infty} \sup_{s\le t} |Y_\ell(s) - ((\bar{x},\bar{y}) + d^* t)| = 0 \text{ a.s..}$$

If $(\bar{x},\bar{y}) + d^*\bar{\tau}$ exits the upper quadrant anywhere except at the origin, we immediately know $\lim_{\ell\to\infty} q_\ell$.

PROOF. of Lemma 4. If $\bar{y}/\bar{x} > d_2^*/d_y^*$, then $(\bar{x},\bar{y}) + d^*t$ for $t \ge 0$ exits the upper quadrant through the y-axis above the origin implying $\lim_{\ell\to\infty} q_\ell = 1$. If $\bar{y}/\bar{x} < d_2^*/d_y^*$, then $(\bar{x},\bar{y}) + d^*t$ for $t \ge 0$ exits the upper quadrant through the x-axis to the right of the origin implying $\lim_{\ell\to\infty} q_\ell = 0$. □

When $\bar{y}/\bar{x} = d_2^*/d_y^*$, then the fluid limit exits the upper quadrant at the origin at time

$$\bar{\tau} \equiv -\bar{y}/d_2^* = -\bar{x}/d_1^*. \tag{31}$$

To determine $\lim_{\ell\to\infty} q_\ell = 1$ in this case, we need a finer analysis, which will be provided by applying the results in Sections 11.2 and 11.4 of [2] to $Z_\ell(t) \equiv \sqrt{\ell}\,(Y_\ell(s) - ((\bar{x},\bar{y}) + d^*t))$. In particular, from Theorem 2.3 of [2], $Z_\ell \Rightarrow Z$ where

$$Z(t) = (r,s) + \sum_{i>0} v_i W_i(\beta_i t)$$

and $W_1, W_2, \ldots$ are independent, standard Brownian motions. We are interested in the hitting location $Z_\ell(\tau_\ell)$ since q_ℓ is the probability that the first coordinate of $Z_\ell(\tau_\ell)$ is zero (and that the second coordinate is greater than zero). Unfortunately, we cannot appeal to Theorem 4.1 of [2] since $\min(x,y)$ is not differentiable. Instead, we bound q_ℓ as follows.

PROOF. of Lemma 5 Let $\varphi_x(s,t) \equiv s$ and $\varphi_y(s,t) \equiv t$. Define the exit time from the right two quadrants as $\tau_\ell^x = \inf\{t > 0 : \varphi_x(Z_\ell(t)) \le 0\}$. Note that $\bar{q}_\ell \equiv \Pr\{\varphi_y(Z_\ell(t)) > 0\} \ge q_\ell$. The reason for the latter inequality is that Z_ℓ could have exited the upper right hand quadrant by entering the lower right hand quadrant, which would be time τ_ℓ, and then reenter the upper right hand quadrant before exiting the right two quadrants and entering the upper left quadrant at time τ_ℓ^x.

Since φ_x is continuously differentiable and the rest of the conditions of Theorem 4.1 in Chapter 11 of [2] hold,

$$\bar{q}_\ell \to \Pr\{\varphi_y(Z(\bar{\tau})) - (d_2^*/d_1^*)\varphi_x(Z(\bar{\tau})) > 0\}$$

where $Z(\bar{\tau})$ has a bivariate normal distribution. Before computing the parameters of this distribution, we derive a lower bound $\underline{q}_\ell \le q_\ell$. The procedure is similar starting with $\tau_\ell^y = \inf\{t > 0 : \varphi_y(Z_\ell(t)) \le 0\}$ and ending with

$$\underline{q}_\ell \to \Pr\{\varphi_x(Z(\bar{\tau})) - (d_1^*/d_2^*)\varphi_y(Z(\bar{\tau})) < 0\}.$$

Fortunately, the upper and lower bounds converge to the same value so we know $\lim_{\ell\to\infty} q_\ell \to q \equiv \Pr\{\tilde{Z} < 0\}$ where $\tilde{Z} \equiv \varphi_x(Z(\bar{\tau})) - (d_1^*/d_2^*)\varphi_y(Z(\bar{\tau}))$.

Under the assumptions of Lemma 5, the mean of $Z(\bar{\tau}) \equiv (Z_1(\bar{\tau}), Z_1(\bar{\tau}))$ is (r,s); hence, the normal random variable $\tilde{Z}$ has mean $r - (d_1^*/d_2^*)s$ and variance

$$\sigma^2 = \mathrm{Var}[Z_1(\bar{\tau})] + (d_1^*/d_2^*)^2\mathrm{Var}[Z_2(\bar{\tau})] - 2(d_1^*/d_2^*)\mathrm{Cov}[Z_1(\bar{\tau}), Z_2(\bar{\tau})]$$

where

$$\begin{aligned}
\mathrm{Var}[Z_1(\bar{\tau})] &= \bar{\tau}\left[(\mu_1+\mu_2)+\lambda_a\sum_{k\geq 0}(k+1)^2p(1-p)^k\right]\\
\mathrm{Var}[Z_2(\bar{\tau})] &= \bar{\tau}\left[\lambda_b+\lambda_a\sum_{k\geq 0}k^2p(1-p)^k\right]\\
\mathrm{Cov}[Z_1(\bar{\tau}),Z_2(\bar{\tau})] &= -\bar{\tau}\lambda_a\sum_{k\geq 0}k(k+1)p(1-p)^k
\end{aligned}$$

with $p=\mu_2/(\mu_1+\mu_2)$. After simplifying,

$$\begin{aligned}
(32)\quad \sigma^2 = \bar{\tau}[&\mu_1+\mu_2+\lambda_a(\frac{\mu_1}{\mu_2})^2+4\lambda_a(\frac{\mu_1}{\mu_2})+\lambda_a+\lambda_b(\frac{d_1^*}{d_2^*})^2+\lambda_a(\frac{d_1^*}{d_2^*})^2(\frac{\mu_1}{\mu_2})^2\\
&+2\lambda_a(\frac{d_1^*}{d_2^*})^2(\frac{\mu_1}{\mu_2})+2\lambda_a(\frac{d_1^*}{d_2^*})(\frac{\mu_1}{\mu_2})^2+6\lambda_a(\frac{d_1^*}{d_2^*})(\frac{\mu_1}{\mu_2})]
\end{aligned}$$

where from (31) $\bar{\tau}\equiv -\bar{y}/d_2^*$. Thus, $q=\Phi[-(r-sd_1^*/d_2^*)/\sigma]$ where Φ is the cumulative distribution function of a standard normal random variable. □

8. Concluding remarks. Theorem 2.2.1 in [8] is somewhat similar to our results though the Markov chain analyzed (QBD process) is different. The approach in [8] does not seem to give any information about the large deviation path. Theorem 2.2.1 does not distinguish among the jitter, bridge and cascade situations; consequently, c and the r.h.s. of (2.14) in [8] may be zero, which limits the usefulness of the result. The important question of the existence and construction of a positive left invariant vector $\mathbf{x}$ giving a finite $c>0$ in (2.13) of [8] for cascades is answered in some generality in [6]. Nevertheless, Theorem 2.2.1 in [8] does handle a cascade situation; furthermore, their proof does seem to involve a time reversal. The idea of starting at some (increasingly) distant state and using some sort of time reversal seems quite useful.

Without the tools in our paper, it appears difficult to obtain the exact asymptotics of π for the production model. There does not seem to be a natural quasi-birth death structure for the Markov chain shown in Figure 1 allowing an approach similar to [8]. Even finding the rough asymptotics of π using the traditional theory of large deviations [11] does not appear straightforward. We are not aware of any existing results that would establish a large deviations principle in the production problem with the range of jumps being unbounded. Even if a large deviation principle could be established and a good rate function found, the optimization problem appears

more formidable than in [5]. Because of the homogeneous transition structure in the interior of the model studied in [5], attention could be restricted to paths that were piecewise linear with at most one change of direction and speed and that change could only occur on an axis. Such paths could be described by a finite dimensional vector making the optimization problem tractable. The transition structure of the production model does not have the same homogeneity on the interior; the y-axis influences the transition structure at every state. Without a similar result restricting attention to some nice set of paths, the optimization would have to consider all continuous paths, including paths that might curve and change speed in the interior, which would be far more difficult.

APPENDIX A: A PROOF OF PROPOSITION 1

If we show that the bounds in (3) hold for all but a finite number of states, then c_1 and c_2 can be adjusted so that the bounds hold for all states. Consequently, Proposition 1 holds trivially if the central buffer were finite. Even so, the following argument could be refined to derive substantive bounds for the finite buffer system.

As described in Section 2.1, let $X_0, X_1, \ldots$ be the uniformized Markov chain with transition kernel K and stationary distribution π. Without loss of generality assume $\lambda_a + \lambda_b + \mu_1 + \mu_2 = 1$. Let π^* be the distribution given in Lemma 1. Under the conditions of Lemma 1, we know $\pi = \pi^*$. Now define $\tau \equiv \inf\{n > 0 | X_n = (0,0)\}$ to be the hitting time of the origin, and let N be the number of visits to state (x,y) during $0, 1, \ldots, \tau - 1$. From standard regenerative arguments, $\pi(x,y) = \pi(0,0)\mathrm{E}_{(0,0)}[N]$ where $\mathrm{E}_{(i,j)}$ denotes the conditional expectation given that $X_0 = (i,j)$. By conditioning on the first step, we have for any state (x,y) other than $(0,0)$ that

$$\pi(x,y) = \pi(0,0)\lambda_a(1-\eta)\mathrm{E}_{(1,0)}[N] + \pi(0,0)(\lambda_b + \lambda_a\eta)\mathrm{E}_{(1,0)}[N].$$

Now assume that (x,y) is not $(0,0)$, $(1,0)$, or $(0,1)$, and let $H(i,j)$ be the probability of hitting state (x,y) before hitting the origin conditioned on the process starting in state (i,j). Thus,

$$\begin{aligned} \pi(x,y) &= \pi(0,0)\mathrm{E}_{(x,y)}[N]\left[\lambda_a(1-\eta)H(1,0) + (\lambda_b + \lambda_a\eta)H(0,1)\right] \\ \pi^*(x,y) &= \pi^*(0,0)\mathrm{E}_{(x,y)}[N][\lambda_a(1-\eta^*)H(1,0) + (\lambda_b + \lambda_a\eta^*)H(0,1)] \end{aligned} \tag{33}$$

Notice that $H(1,0)$ is weighted more heavily, and $H(0,1)$ less heavily, when $\eta < \eta^*$. Let $r \equiv \lambda_a(1-\eta^*)/(\lambda_b + \lambda_a\eta^*) = \lambda_a/(\lambda_a + \lambda_b)$ be the ratio of the

weights when $\eta = \eta^*$. Temporarily assume that $\eta \leq \eta^*$. By either increasing both weights or decreasing both weights, we can obtain upper or lower bounds. In particular,

$$\begin{aligned}
&\pi(0,0)\mathrm{E}_{(x,y)}[N][r(\lambda_b + \lambda_a\eta)H(1,0) + (\lambda_b + \lambda_a\eta)H(0,1)] \\
&\leq \pi(x,y) \\
&\leq \pi(0,0)\mathrm{E}_{(x,y)}[N][\lambda_a(1-\eta)H(1,0) + \frac{1}{r}\lambda_a(1-\eta)H(0,1)],
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
&\frac{\pi(0,0)}{\pi^*(0,0)}\frac{(\lambda_b + \lambda_a\eta)}{(\lambda_b + \lambda_a\eta*)}\pi^*(x,y) \\
&\leq \pi(x,y) \\
&\leq \frac{\pi(0,0)}{\pi^*(0,0)}\frac{\lambda_a(1-\eta)}{\lambda_a(1-\eta^*)}\pi^*(x,y)
\end{aligned}$$

From the form of π^*, it is clear that finite constants $c_1 > 0$ and $c_2 > 0$ can be calculated so that (3) holds for $\eta \leq \eta^*$. On the other hand, if $\eta \geq \eta^*$, the same argument goes through after reversing the inequalities except that there is a slight problem in the last step when $\eta = 1$. When $\eta = 1$, the lower bound has a factor $\lambda_a(1-\eta)$ implying that the lower bound is 0 at that point.

To obtain the lower bound when $\eta = 1$, consider states (x,y) with $x+y \geq 4$. Let $q(i,j)$ be the probability of going from $(0,0)$ to state (i,j) in three steps while avoiding $(0,0)$. Note that $q(0,1)$ and $q(1,0)$ are both positive.

$$\begin{aligned}
\pi(x,y) &> \pi(0,0)\mathrm{E}_{(x,y)}[N][q(1,0)H(1,0) + q(0,1)H(0,1)] \\
&\geq \pi(0,0)\mathrm{E}_{(x,y)}[N][\min[q(1,0), rq(0,1)]H(1,0) + \min[q(1,0)/r, q(0,1)]H(0,1)] \\
&= \pi(0,0)\mathrm{E}_{(x,y)}[N]\frac{\min[q(1,0)/r, q(0,1)]}{(\lambda_b + \lambda_a\eta*)}[r(\lambda_b + \lambda_a\eta*)H(1,0) + (\lambda_b + \lambda_a\eta*)H(0,1)] \\
&= \frac{\pi(0,0)}{\pi^*(0,0)}\frac{\min[q(1,0)/r, q(0,1)]}{(\lambda_b + \lambda_a\eta*)}\pi^*(x,y),
\end{aligned}$$

which makes it clear that a constant $c_1 > 0$ can be selected even when $\eta = 1$.

APPENDIX B: TIME REVERSALS AND REPRESENTING π

In this section, we describe the representation of the stationary distribution π of a Markov chain using the time reversal of associated Markov

additive process. We use this representation in several places: (16), (21), and (27).

Let K be the transition kernel of an irreducible, positive recurrent Markov chain on a countable state space S with stationary distribution π. Let K^∞ be the transition kernel of a Markov additive process on $S^\infty \supset S$. Partition S into two sets: Δ and Θ. In this section, We assume that

$$K^\infty(x,y) = K(x,y) \text{ for } x \in \Delta \text{ and } y \in \Theta, \text{ and} \tag{34}$$

$$K^\infty(x,y) = K(x,y) \text{ for } x \in \Theta \text{ and } y \in \Theta. \tag{35}$$

Starting from Orey's representation and using the arguments in Section 3, we can often represent π as

$$\pi(x,y)h(x,y) = \mathrm{E}\left[1_\Delta(\mathcal{X}(0))h(\mathcal{X}(0))\mathcal{N}_\blacktriangle(x,y)\right] \tag{36}$$

where $\mathcal{N}_\blacktriangle(x,y)$ is the number of visits to (x,y) by the Markov additive process $\{\mathcal{X}(n) = (\mathcal{X}_1(n), \mathcal{X}_2(n)); n = 0,1,2,\dots\}$ with state space $\blacktriangle \cup S$, $\blacktriangle \cap S = \Delta$, and initial distribution π until $\mathcal{X}$ hits $\blacktriangle$ at time $\mathcal{T}$.

Let φ be an invariant measure for the Markovian part of $\mathcal{X}$, which we arbitrarily choose to be the second component. Let $\overleftarrow{\mathcal{X}}$ be the time reversal of $\mathcal{X}$; see (14). The next lemma shows that if (36) holds for some $(x,y) \in S \setminus \Delta$, then we can also represent $\pi(x,y)$ as a function of the hitting distribution on $\Delta \subset \blacktriangle$.

LEMMA 6. *If (36) holds for some $(x,y) \in S \setminus \Delta$, then*

$$\pi(x,y)\frac{h(x,y)}{\varphi(y)} = \mathrm{E}_{(x,y)}\left[1_\Delta(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right], \tag{37}$$

where π is extended to $\blacktriangle \cup S$ by defining $\pi(\blacktriangle \setminus \Delta) = 0$.

PROOF. Let

$$\mathcal{K}^n_\blacktriangle((w,z),(x,y)) \equiv \Pr{}_{(w,z)}\{\mathcal{X}_n = (x,y), \mathcal{X}_j \notin \blacktriangle, 0 < j < n\}, \text{ and}$$
$$\overleftarrow{\mathcal{K}}^n_\blacktriangle((x,y),(w,z)) \equiv \Pr{}_{(x,y)}\left\{\overleftarrow{\mathcal{X}}_n = (w,x), \overleftarrow{\mathcal{X}}_j \notin \blacktriangle, 0 < j < n\right\}.$$

Using (14), it is straightforward to show that

$$\mathcal{K}^n_\blacktriangle((w,z),(x,y)) = \frac{\varphi(y)}{\varphi(z)}\overleftarrow{\mathcal{K}}^n_\blacktriangle((x,y),(w,z)).$$

Starting from (36),

$$\begin{aligned}\pi(x,y)h(x,y) &= \mathrm{E}\left[1_{\blacktriangle}(\mathcal{X}(0))h(\mathcal{X}(0))\mathcal{N}_{\blacktriangle}(x,y)\right] \\ &= \sum_{(w,z)\in\blacktriangle} 1_{\Delta}((w,z))\pi(w,z)h(w,z)\sum_{n=1}^{\infty}\mathcal{K}_{\blacktriangle}^{n}((w,z),(x,y)) \\ &= \sum_{(w,z)\in\blacktriangle} 1_{\Delta}((w,z))\pi(w,z)h(w,z)\sum_{n=1}^{\infty}\frac{\varphi(y)}{\varphi(z)}\overleftarrow{\mathcal{K}}_{\blacktriangle}^{n}((x,y),(w,z)) \\ &= \sum_{(w,z)\in\blacktriangle} 1_{\Delta}((w,z))\pi(w,z)h(w,z)\frac{\varphi(y)}{\varphi(z)}\Pr{}_{(x,y)}\left\{\overleftarrow{\mathcal{X}}(\mathcal{T})=(w,z)\right\}.\end{aligned}$$

Hence,

$$\pi(x,y)\frac{h(x,y)}{\varphi(y)} = \mathrm{E}_{(x,y)}\left[1_{\Delta}(\overleftarrow{\mathcal{X}}(\mathcal{T}))\pi(\overleftarrow{\mathcal{X}}(\mathcal{T}))\frac{h(\overleftarrow{\mathcal{X}}(\mathcal{T}))}{\varphi(\overleftarrow{\mathcal{X}}_2(\mathcal{T}))}\right].$$

□

REFERENCES

[1] Dabrowski, A., Lee, J., and McDonald, D. R. (2008). Large deviations of mulit-class M/G/1 queues. Submitted to the Canadian Journal of Statistics.

[2] Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley & Sons, 534 p.

[3] Foley, R. D. and McDonald, D. R. (2001). Join the shortest queue: stability and exact asymptotics. *Annals Of Applied Probability* **11**, 3 (Aug.), 569–607.

[4] Foley, R. D. and McDonald, D. R. (2005a). Bridges and networks: Exact asymptotics. *Annals of Applied Probability 15*, 542. doi:10.1214/105051604000000675.

[5] Foley, R. D. and McDonald, D. R. (2005b). Large deviations of a modified Jackson network: stability and rough asymptotics. *Annals of Applied Probability 15*, 519. doi:10.1214/105051604000000666.

[6] Foley, R. D. and McDonald, D. R. (2008). Rare events and exact asymptotics: a constructive theory. In preperation.

[7] Kesten, H. (1974). Renewal theory for functionals of a Markov-chain with general state space. *Annals Of Probability* **2**, 3, 355–386.

[8] Li, H., Miyazawa, M., and Zhao, Y. Q. (2007). Geometric decay in a QBD process with countable background states with applications to a join-the-shortest-queue model. *Stochastic Models* **23**, 3, 413–438.

[9] McDonald, D. R. (1999). Asymptotics of first passage times for random walk in an orthant. *Annals of Applied Probability* **9**, 1 (Feb.), 110–145.

[10] Meyn, S. and Tweedie, R. (1993). *Markov Chains and Stochastic Stability.* citeseer.ist.psu.edu/meyn93crash.html.

[11] Schwartz, A. and Weiss, A. (1995). *Large Deviations for Performance Analysis: Queues, Communication, and Computing.* Chapman and Hall, London, UK.

Eindhoven University of Technology
Department of Mathematics and Computer Science
HG 9.07
P.O. Box 513
5600 MB Eindhoven
the Netherlands
printeade1