Queueing Syst (2007) 57: 129–145
DOI 10.1007/s11134-007-9046-5

Stability of join-the-shortest-queue networks

J.G. Dai · John J. Hasenbein · Bara Kim

Received: 27 May 2007 / Revised: 27 September 2007 / Published online: 10 October 2007

Abstract This paper investigates stability behavior in a variant of a generalized Jackson queueing network. In our network, some customers use a join-the-shortest-queue policy when entering the network or moving to the next station. Furthermore, we allow interarrival and service times to have general distributions. For networks with two stations we derive necessary and sufficient conditions for positive Harris recurrence of the network process. These conditions involve only the mean values of the network primitives. We also provide counterexamples showing that more information on distributions and tie-breaking probabilities is needed for networks with more than two stations, in order to characterize the stability of such systems. However, if the routing probabilities in the network satisfy a certain homogeneity condition, then we show that the stability behavior can be explicitly determined, again using the mean value parameters of the network. A byproduct of our analysis is a new method for using the fluid model of a queueing network to show *non-positive recurrence* of a process. In previous work, the fluid model was only used to show either positive Harris recurrence or transience of a network process.

Keywords Jackson networks · Join-the-shortest-queue policy · Positive Harris recurrence · Fluid model · Stability

Mathematics Subject Classification (2000) Primary 90B15 · Secondary 60K25

J.G. Dai's research supported in part by National Science Foundation grants DMI-0300599, CMMI-0727400 and CNS-0718701, and by an IBM Faculty Award.
J.J. Hasenbein's research supported in part by National Science Foundation grant DMI-0132038.
B. Kim's research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment).

J.G. Dai
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: dai@gatech.edu

J.J. Hasenbein
Graduate Program in Operations Research and Industrial Engineering, Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712, USA
e-mail: jhas@mail.utexas.edu

B. Kim (✉)
Department of Mathematics and Telecommunication Mathematics Research Center, Korea University, Anam-dong, Sungbuk-ku, Seoul, 136-701, Republic of Korea
e-mail: bara@korea.ac.kr

1 Queueing network models

We consider a variant of the classical Jackson queueing network [9, 10]. The main added feature is that an arriving customer may have several routes to choose from at its arrival time. We assume that the customer always chooses to join the shortest queue among a set of allowed queues. In addition, we allow the interarrival and service times to have general distributions, rather than being restricted to the exponential case.

Our queueing network model is assumed to have $J \geq 1$ stations, with each station consisting of a single server. Each station has a dedicated queue or buffer that holds customers waiting to be served by the station. Let $\mathcal{J} = \{1, \ldots, J\}$ be the set of stations. For each station $i \in \mathcal{J}$, let $\eta_i(n)$ be the service time of the nth customer to be served by station i. We assume that each station is non-idling, that customers within a buffer are served on a first-in–first-out basis, and that no service is preempted. To describe the external arrival processes, let $\mathcal{P}$ be the class of nonempty subsets of $\mathcal{J}$. For each subset $C \in \mathcal{P}$ of queues, there is an associated exogenous arrival process with interarrival times $\{\xi_C(n) : n \geq 1\}$. We call this arrival process a type-C external arrival process. Upon arriving to the network, each type-C customer joins the shortest queue among all the queues in C, using a tie-breaking rule to be specified shortly. After being served by station i, $i \in \mathcal{J}$, a customer leaves the system with probability $1 - p_i^*$, and becomes a type-C customer with probability p_{iC}, independent of the customer's entire history, where $\sum_{C \in \mathcal{P}} p_{iC} = p_i^*$. When multiple queues are tied for the shortest queue, a tie-breaking rule is needed. We assume that for each subset $B \in \mathcal{P}$ of queues, there is a distribution $\gamma_B = \{\gamma_{B,j} : j \in B\}$. When a customer is to join a shortest queue that is tied by a set B of queues, the customer joins queue j with probability $\gamma_{B,j}$ independently of its history. This type of routing behavior on the part of arriving customers is called Join-the-Shortest-Queue (JSQ) in the literature.

We allow $\xi_C(n) = \infty$ for all n for some C. In this case, the type-C-external arrival process is null. Let

$$\mathcal{E} = \{C \in \mathcal{P} : \text{the type-}C\text{-external arrival process is non-null}\}.$$

For each $C \in \mathcal{E}$, we assume that $\xi_C = \{\xi_C(n) : n \geq 1\}$ is an independent and identically distributed (i.i.d.) sequence with mean $1/\lambda_C$, and for each station i, $\eta_i = \{\eta_i(n) : n \geq 1\}$ is an i.i.d. sequence with mean $1/\mu_i$. We further assume that the interarrival time sequences, service time sequences, feedback decisions, and tie-breaking decisions are all independent. Additional distributional assumptions on the interarrival times will be specified in Sect. 2. We call λ_C the arrival rates, μ_i the service rates, p_{iC} the feedback probabilities, and $\gamma_{B,j}$ the tie-breaking probabilities of the network. From now on, for purposes of discussion, and stating results, we will refer to the network described above as a *JSQ Network*.

The dynamics of the JSQ network can be described by a continuous time Markov process $X = \{X(t) : t \geq 0\}$, as long as the state space is chosen appropriately. When the interarrival and service time distributions are exponential, $Z = \{(Z_1(t), \ldots, Z_J(t)) : t \geq 0\}$ is such a process, where $Z_i(t)$ is the total number of customers that are either waiting in queue i or being served by station i at time t. This paper is primarily concerned with the stability of the queueing network. The network is said to be stable if the Markov process X is positive Harris recurrent. When $\lambda_C = 0$ and $p_{i,C} = 0$ for all $C \in \mathcal{P}$ with more than one element, i.e., customers are never offered a choice of queues to join, the corresponding network is called a generalized Jackson network. Under some minor conditions on the interarrival time distributions, it is known that such a network is stable if and only if the traffic intensity at each station is less than one (see, for example, Dai [2]). The traffic intensity is defined through the first order data of the network, i.e., arrival rates, service rates and feedback probabilities. In particular, the stability of a generalized Jackson network does not depend on the distributions of interarrival and service times. One might hope that for the model introduced in this paper, the positive Harris recurrence can again be determined by the arrival rates, service rates and the feedback probabilities. Our first result, Theorem 1, shows that this assertion is indeed true when $J = 2$ by describing explicit recurrence conditions in terms of arrival rates, service rates and feedback probabilities. In particular, the stability of a 2-station network does not depend on the distributions of interarrival and service times, nor does it depend on the tie-breaking probabilities. Unfortunately, when $J \geq 3$, an analogous result does not hold. Specifically, as our second result, we provide two counterexamples which demonstrate that Theorem 1 cannot be generalized to larger networks. In the first example with $J = 3$, we show that the positive Harris recurrence of the process depends on the tie-breaking probabilities $\gamma_{B,i}$. In the second example, again with $J = 3$, we show that the positive Harris recurrence of the process depends on the distributions of the service times. As a final contribution of the paper, we prove that when all the stations have homogeneous feedback probabilities, i.e., p_{iC} does not depend on queue i, the positive Harris recurrence is again determined by the arrival rates, services and feedback probabilities, and not on the distributions of the interarrival and service times or the tie-breaking probabilities. In this case, we give explicit recurrence conditions in terms of arrival rates, service rates and feedback probabilities. The first and the last result will be used in a companion paper to prove recent conjectures of Suhov and Vvedenskaya [14].

Queueing systems with JSQ type routing have a long history in the literature. We only mention the papers in which there is stability analysis of JSQ networks. Kurkova [11] treated a special system when $J = 2$, the interarrival and service times distributions are exponential, and a fair coin is flipped to break a tie. She represented the system as a continuous time Markov chain with a countable state space and obtained an explicit recurrence condition for the Markov chain by using Lyapunov functions. Stability of JSQ networks, when there is no feedback, was studied by Foss and Chernova [8] and Foley and McDonald [6, 7]. A quite general JSQ network with feedback was treated by Suhov and Vvedenskaya [14]. However, their stability analysis was limited to a few special cases.

Queueing networks with alternate routes arise in many telecommunication and service systems. A customer call center is an example of such a service system. The myopic join-the-shortest-queue routing decision is often employed in practice. The stability of these networks is essential to the capacity planning of these systems.

We employ the standard fluid model tool in our stability analysis. Whenever appropriate, we do not go through every detail of using the tool; readers may consult, for example, Dai [4] for additional details. Fluid models are commonly used to prove the positive recurrence of queueing networks and/or the transience of such systems, but here we are also able to use the fluid model approach to prove non-positive recurrence. As such we are able to identify the stability behavior on the boundary of the stability region. The behavior on the boundary is often left as an open question in stability analysis via fluid models, and to the best of our knowledge, this paper is the first to use fluid models to prove the non-positive recurrence of a queueing network.

The paper is organized as follows: In Sect. 2, we provide the Markov process characterization of the network. This section also gives the necessary and sufficient conditions for stability in terms of arrival rates, service rates and feedback probabilities for systems with two stations ($J=2$), and for systems with more than two stations ($J\geq 3$) under an additional assumption on the network structure. In Sect. 3, two examples with three stations ($J=3$) are given, the first of which shows that the stability depends on the tie-breaking probabilities, and the second of which shows that the stability depends on not only the first order data but also the distributions of the service times. In Sect. 4, a fluid model for the system is defined and criteria for stability and instability of the system are given. Most of the proofs are collected in Sect. 5 and the Appendices 1–3.

Now we collect some mathematical notation used in the rest of the paper. For a set C, $|C|$ indicates the cardinality of C. However, for $x\in\mathbb{R}^N$, we use $|x|$ to denote the l^1-norm. For random variables X and Y, $X\geq_{\text{st}} Y$ indicates that X is stochastically larger than Y. When a probability operator appears with a subscript π, this indicates the probability is the one generated by initial distribution π (this may include a degenerate initial distribution consisting of only one state).

2 Network definitions and main results

We use

$$X(t)=(Z(t),U(t),V(t)) \tag{1}$$

to denote the state of our queueing network at time t. The first component $Z(t)=(Z_1(t),\dots,Z_J(t))$ is J-dimensional, where, as before, $Z_i(t)$ is the total number of customers that are either waiting in queue i or being served by station i at time t. The second component $U(t)=(U_C(t): C\in\mathcal{E})$ is $|\mathcal{E}|$-dimensional, where $U_C(t)$ is the remaining interarrival time of the type-C external arrival process at time t. The last component $V(t)=(V_1(t),\dots,V_J(t))$ is J-dimensional, where $V_i(t)$ is the remaining service time of the customer who is in service at station i at time t. ($V_i(t)$ is set to be zero if there is no customer in service at station i at time t.) The process $X=\{X(t):t\geq 0\}$ is taken to be right continuous with left limits. It follows from Dai [2] that X is a strong Markov process whose state space $\mathcal{S}$ is a subset of $\mathbb{R}^{2J+|\mathcal{E}|}$.

The Markov process X is said to be positive Harris recurrent if it possesses a unique stationary distribution. To state the main results of this paper, we make the following additional assumptions on interarrival times. We assume that, for any $C\in\mathcal{E}$, the distribution of $\xi_C(1)$ is unbounded, i.e.,

$$\mathbb{P}(\xi_C(1)\geq x)>0,\quad \text{for any } x>0. \tag{2}$$

We also assume that, for any $C\in\mathcal{E}$, the distribution of $\xi_C(1)$ is spread out, i.e., there exists an integer $n_C>0$ and a function $q_C(x)\geq 0$ on $(0,\infty)$ with $\int_0^\infty q_C(x)dx>0$, such that

$$\mathbb{P}(a\leq \xi_C(1)+\dots+\xi_C(n_C)\leq b)\geq \int_a^b q_C(x)dx,$$

for any $0\leq a<b$.

Our first result is for queueing networks with $J=2$. For simplicity of notation we use λ_1, λ_2 and λ instead of $\lambda_{\{1\}}$, $\lambda_{\{2\}}$ and $\lambda_{\{1,2\}}$, respectively, in the two station case. We also use p_{ij} instead of $p_{i\{j\}}$, $i,j=1,2$. To avoid trivial cases, we assume that $p_{11}<1$, $p_{22}<1$ and at least one of p_1^* and p_2^* is less than 1. However, we make no assumptions on $p_{i,\{1,2\}}$, $i=1,2$.

Theorem 1 *Consider a JSQ network with $J=2$. The Markov process X is positive Harris recurrent if and only if the following three conditions hold*:

(i) $\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2<0$;
(ii) *if* $p_2^*<1$, *then* $p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1p_1^*-\mu_1)+(1-p_2^*)(\lambda_1+\mu_1p_{11}-\mu_1)<0$;
(iii) *if* $p_1^*<1$, *then* $p_{12}(\lambda_1+\lambda_2+\lambda+\mu_2p_2^*-\mu_2)+(1-p_1^*)(\lambda_2+\mu_2p_{22}-\mu_2)<0$.

We leave the proof of the theorem to Sect. 5. Kurkova [11] obtained a necessary and sufficient condition that is equivalent to ours (see Appendix 2). Her paper examines the special case when the exogenous arrival processes are Poisson, all service times have an exponential distribution with mean 1, and $\gamma_{\{1,2\},j}=\frac{1}{2}$, $j=1,2$. The following theorem is proved in Sect. 5.

Theorem 2 *Consider a JSQ network with $J=2$. The process X is unstable in the sense that $|Z(t)|\to\infty$ as $t\to\infty$ with probability* 1 *if*

$$\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2>0,\quad \textit{or} \tag{3}$$

$$\begin{aligned} &p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1p_1^*-\mu_1)\\ &\quad+(1-p_2^*)(\lambda_1+\mu_1p_{11}-\mu_1)>0,\quad \textit{or}\end{aligned} \tag{4}$$

$$p_{12}(\lambda_1+\lambda_2+\lambda+\mu_2 p_2^*-\mu_2) + (1-p_1^*)(\lambda_2+\mu_2 p_{22}-\mu_2) > 0. \quad (5)$$

We now provide an interpretation of the conditions in Theorem 1. The first condition is the most straightforward. First, partition the state space of the number of customers in the system, say (z_1, z_2) into two regions. Let region I be $\{(z_1,z_2): z_1 < z_2\}$ and region II be $\{(z_1,z_2): z_1 > z_2\}$ (we ignore the boundary set for now). In region I all type-$\{1,2\}$ customers join the queue at station 1. Then, if station 1 is busy the net rate at which it eliminates jobs from the system is

$$r_1 \equiv \mu_1+\mu_1 p_{12}-\mu_1 p_1^*+\mu_2 p_{22}-\mu_2 p_2^*-\lambda_1-\lambda.$$

Similarly, in region I the net rate at which station 2 eliminates customers from the system is

$$r_2 \equiv \mu_2-\mu_1 p_{12}-\mu_2 p_{22}-\lambda_2.$$

Notice that the left-hand side of condition (i) is simply $-(r_1+r_2)$, i.e. condition (i) implies that the total net rate at which customers are eliminated must be positive. One can check that the left-hand side of (i) also corresponds to the net customer elimination rate in region II. On the boundary between the two regions, the elimination rate seemingly should depend on the tie-breaking probability. However, since the rates are the same in either region, we see that the tie-breaking probability is immaterial to this rate condition.

Condition (i) is a type of drift condition on the interior of the state space. The other two conditions are drift rate conditions on the boundaries. To see this suppose $z_1 = 0$, i.e. station 1 is idle. In this case, the net drift rate of the number of jobs is given by

$$(s_1,s_2) \equiv (\lambda_1+\lambda+\mu_2(p_2^*-p_{22}), \lambda_2+\mu_2 p_{22}-\mu_2).$$

Then condition (iii) is equivalent to $(-r_2, r_1)\cdot(s_1,s_2) < 0$, i.e. the normal to the interior drift and the reflection vector must form an acute angle. This is the usual stability condition for a process with (constant) oblique reflection at the boundaries. Condition (ii) has an analogous interpretation for the boundary defined by $z_2 = 0$.

Theorem 1 implies that the stability of a 2-station network does not depend on the distributions of interarrival and service times or the tie-breaking probabilities. Unfortunately, when $J \ge 3$, the analogous result does not hold as we will see in Sect. 3. However, for such networks we can identify stability conditions in terms of λ_C, μ_C and $p_{iC}, i \in \mathcal{J}, C \in \mathcal{P}$, under an additional assumption on network structure.

Assumption 1 For any $C \in \mathcal{P}$, p_{iC} does not depend on $i \in \mathcal{J}$. Namely, all stations have the same feedback probabilities. For $C \in \mathcal{P}$, let

$$\Lambda_C \equiv \sum_{B:\phi\neq B\subseteq C} \lambda_B,$$

$$P_C \equiv \sum_{B:\phi\neq B\subseteq C} p_{iB} \quad \text{and} \quad \mu_C \equiv \sum_{i\in C}\mu_i.$$

Let $\lambda^* \equiv \Lambda_{\mathcal{J}}$ be the total external arrival rate to the network and $p^* \equiv p_i^*$, which is independent of station $i \in \mathcal{J}$. To avoid triviality, further assume that $p^* < 1$.

Under this assumption, the stability of larger networks can be determined directly from the first-order network parameters, as the next two results demonstrate.

Theorem 3 *Consider a JSQ network with $J \ge 3$ whose parameters are in concordance with Assumption 1. The Markov process $\{X(t): t \ge 0\}$ is positive Harris recurrent if and only if*

$$\Lambda_C + \frac{\lambda^*}{1-p^*}P_C < \mu_C, \quad \textit{for all } C \in \mathcal{P}. \quad (6)$$

Theorem 4 *Consider a JSQ network with $J \ge 3$ whose parameters are in concordance with Assumption 1. The process X is unstable in the sense that $|Z(t)| \to \infty$ as $t \to \infty$ with probability 1 if there exists a $C \in \mathcal{P}$ such that*

$$\Lambda_C + \frac{\lambda^*}{1-p^*}P_C > \mu_C. \quad (7)$$

We leave the proofs of both theorems to Sect. 5.

3 Two counterexamples

In this section, we consider the case $J = 3$ and give two examples which show that λ_C, μ_i and p_{iC}, $i \in \mathcal{J}, C \in \mathcal{P}$, are not sufficient to determine the stability of the system. The first example shows that the stability of the system may depend on the tie-breaking rule $\gamma_{C,i}, C \in \mathcal{P}, i \in \mathcal{J}$. The second example shows that the stability of the system may depend not only on the mean service times but also on the distributions of the service times.

Both examples fit into a class of networks, whose structure is pictured in Fig. 1. The network has three stations, each represented by a circle. Each station serves customers in its queue, which is represented by an open rectangle. In each example, there are potentially four types of exogenous arrival processes, which are assumed to be four independent Poisson processes. The first three processes correspond to arrivals which are dedicated to queues 1, 2, and 3 respectively. The fourth Poisson process corresponds to arrivals which join the shorter of the two queues 1 and 2. If the

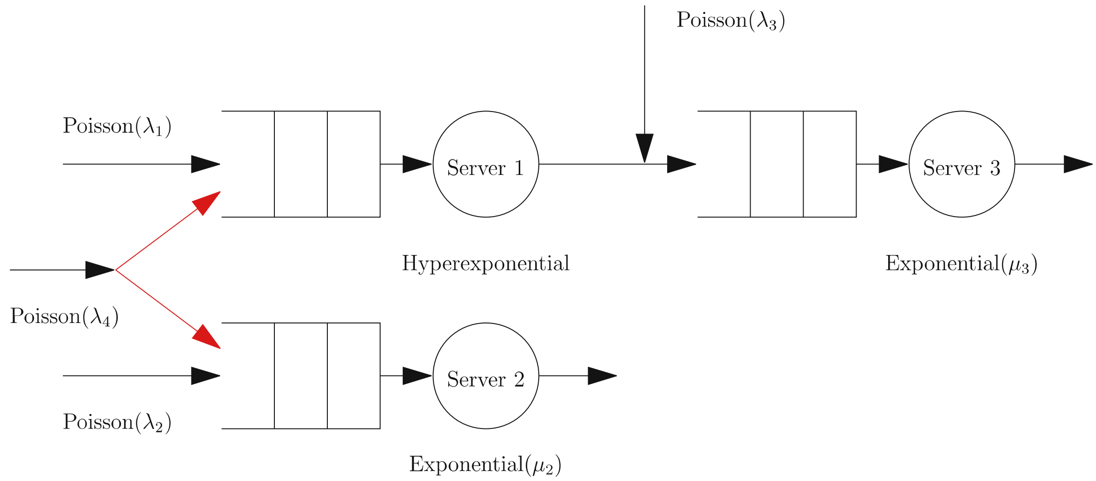

Fig. 1 A JSQ network

queue lengths are equal at the time of an arrival, the customer breaks the tie using a Bernoulli(r) random variable which is independent of all the other primitive processes, with a success indicating that the customer joins queue 1. The four Poisson processes have rates λ_i, $i = 1, 2, 3, 4$.

The service times at stations 2 and 3 are assumed to be i.i.d. exponential random variables with rates μ_2 and μ_3, respectively. The service times at station 1 are assumed to be i.i.d. random variables which are hyperexponential. We assume that the hyperexponential is generated by mixing independent exponential(a) and exponential(b) random variables, with the first component being chosen with probability ν. With these assumptions, the natural definition of the rate of service at station 1 is then:

$$\mu_1 = (\nu a^{-1} + (1-\nu)b^{-1})^{-1}.$$

Now, in such a network let

$$Y(t) = \begin{cases} 0, & \text{if no job is in service at station 1;} \\ 1, & \text{if the current job in service at station 1 is assigned an exponential}(a) \text{ service;} \\ 2, & \text{if the current job in service at station 1 is assigned an exponential}(b) \text{ service.} \end{cases}$$

Then, for this class of networks both $\{(Z_1(t), Z_2(t), Y(t)) : t \geq 0\}$ and $\{(Z_1(t), Z_2(t), Z_3(t), Y(t)), t \geq 0\}$ are irreducible continuous time Markov chains (CTMCs). When

$$\lambda_1 + \lambda_2 + \lambda_4 < \mu_1 + \mu_2, \tag{8}$$

it follows from Theorem 1 that the continuous time Markov chain $\{(Z_1(t), Z_2(t), Y(t)) : t \geq 0\}$ is positive recurrent. We use $\mathbb{P}\{(Z_1(\infty), Z_2(\infty)) \in \cdot\}$ to denote the stationary distribution of $\{(Z_1(t), Z_2(t)) : t \geq 0\}$. We use $A_1(t)$ to denote the number of customers that have entered either the queue or service at station 1 in $[0, t]$, and we use $D_1(t)$ to denote the number of service completions by station 1 in $[0, t]$. Note that $A_1(t)/t$ and $D_1(t)/t$ are the arrival rate at station 1 and the departure rate from station 1, respectively, in $[0, t]$. For a fixed time t, both of these rates are random. Our next proposition shows that, when (8) is satisfied, these rates converge to constants as $t \to \infty$.

Proposition 1 *Assume that condition (8) holds.*

(a) *Set* $d_1 = \mu_1 \mathbb{P}\{Z_1(\infty) > 0\}$. *For each initial state* x,

$$\mathbb{P}_x\left\{\lim_{t\to\infty} D_1(t)/t = d_1\right\} = 1. \tag{9}$$

(b) *Set* $a_1 = \lambda_1 + \lambda_4(\mathbb{P}\{Z_1(\infty) < Z_2(\infty)\} + r\mathbb{P}\{Z_1(\infty) = Z_2(\infty)\})$. *For each initial state* x,

$$\mathbb{P}_x\left\{\lim_{t\to\infty} A_1(t)/t = a_1\right\} = 1. \tag{10}$$

(c) $a_1 = d_1$.

Proof The proofs of both (a) and (b) follow by applying standard sample path versions of PASTA as in Wolff [15, Chap. 5, Theorem 6 and Example 5–23]. We outline the proof for (a), the proof for (b) uses similar arguments. All arguments hold for the probability measure generated by a fixed, but arbitrary initial state x.

Let $\{N(t), t \geq 0\}$ be a Poisson process with rate μ_1. This process generates departures from station 1 whenever there is a job present at the station, otherwise an event in $N(\cdot)$ is ignored. Recall that $D_1(t)$ is the number of departures from station 1 in $[0, t]$. Then sample path PASTA and standard results for ergodic CTMC's yield:

$$\lim_{t\to\infty} \frac{D_1(t)}{N(t)} = \mathbb{P}\{Z_1(\infty) > 0\} \quad \text{a.s.}$$

The strong law of large numbers for renewal processes gives:

$$\lim_{t\to\infty}\frac{N(t)}{t}=\mu_1 \quad \text{a.s.}$$

Thus

$$\lim_{t\to\infty}\frac{D_1(t)}{t}=\frac{D_1(t)}{N(t)}\frac{N(t)}{t}=\mu_1\mathbb{P}\{Z_1(\infty)>0\} \quad \text{a.s.}$$

To prove (c), we note that from the proof of Theorem 1, the fluid model of the network consisting of the first two queues is stable. Thus, the network is rate stable, see for example, Dai [4]. Rate stability implies that $d_1=a_1$, proving part (c). □

When condition (8) holds, Proposition 1 asserts that the long-run departure rate from station 1 exists and is equal to d_1, a component of our next proposition.

Proposition 2 *For the network in Fig. 1, the Markov chain* $\{(Z_1(t),Z_2(t),Z_3(t),Y(t)):t\geq 0\}$ *is positive recurrent iff*

$$\lambda_1+\lambda_2+\lambda_4<\mu_1+\mu_2 \quad \textit{and} \quad \lambda_3+d_1<\mu_3.$$

Proposition 2 is proven in the Appendix 1. We are now ready to analyze a set of examples which give further insight into the stability behavior of JSQ networks.

Example 1 We now consider a special case of the three station network introduced above. Let $\lambda_1=\lambda_2=0$ and let λ_3 and λ_4 be arbitrary. Furthermore, assume $\nu=1$ and $\mu_1:=a=\mu_2$. Thus, there are exponential service times at all stations, with stations 1 and 2 having the same service rates. For simplicity, we then drop the component $Y(t)$ from the state descriptor X, since it is not needed for X to be a Markov chain.

We now argue that the positive recurrence of X depends on the tie-breaking parameter r. The first condition in Proposition 2 reduces to $\lambda_4<2\mu_1$. Under this condition, by Theorem 1, the process $\{(Z_1(t),Z_2(t)):t\geq 0\}$ is positive recurrent. Let $\{\kappa_{ij}(r):i,j\geq 0\}$ be the stationary distribution of this process. Then applying Proposition 2 we immediately obtain:

Claim 1 *If* $\lambda_4\geq 2\mu_1$ *then the Markov chain* X *is not positive recurrent. If* $\lambda_4<2\mu_1$, *then* X *is positive recurrent if and only if*

$$\lambda_3+\mu_1\left(1-\sum_{j=0}^{\infty}\kappa_{0j}(r)\right)<\mu_3. \tag{11}$$

By Lemma 7 in Appendix 3, it is seen that $\sum_{j=0}^{\infty}\kappa_{0j}(r)$ decreases strictly as r increases. Thus it is clear that one can choose fixed parameters $\lambda_3,\lambda_4,\mu_1$, and μ_3 for which the stability conditions will hold for some r and not hold for another choice of r. *In particular, the necessary and sufficient conditions for the positive recurrence of* X *depend on the tie-breaking parameter* r.

Example 2 Consider now another special case of the network depicted in Fig. 1. In particular let $\lambda_1=\lambda_2=0.8$, $\lambda_3=0.17$, $\lambda_4=0.1$, $r=1/2$ and $\mu_1=\mu_2=\mu_3=1$, where station 1's service time distribution remains to be chosen. We now argue that the positive recurrence of X depends on the distribution of the service times for station 1 even if the mean is fixed.

First suppose $\nu=1$ and $a=1$. Thus, all service times for station 1 are exponentially distributed with mean 1. For this case, we have the following claim:

Claim 2 *If the service times for station* 1 *are exponentially distributed with mean* 1 *then the process* X *is not positive recurrent.*

Proof For the set of parameters under consideration, condition (8) holds and we can apply Proposition 1, which implies that the departure rate from station 1 (and station 2) exists with probability 1. As argued earlier, from Theorem 1, condition (8) also implies that the fluid model of the network consisting of the first two queues is stable, and so the network itself is rate stable. Hence, so the total departure rate from the first two queues must equal the total arrival rate of 1.7. Furthermore, by symmetry, the departures rates from station 1 and station 2 must be equal. Thus, $d_1=0.85$ and applying Proposition 2, we infer that X is not positive recurrent. □

Now suppose we alter the distribution, but not the mean service time, at station 1. In particular, let $0<\nu<1$ and $a=\frac{\nu}{1-\nu+\nu^2}$ and $b=1/\nu$. Then services at station 1 are hyperexponential with the following c.d.f.:

$$\begin{aligned}F(x)=\nu\left(1-\exp\left(\frac{-\nu x}{1-\nu+\nu^2}\right)\right)\\+(1-\nu)\left(1-\exp\left(-\frac{x}{\nu}\right)\right),\quad 0\leq x<\infty.\end{aligned}\tag{12}$$

Note that for any $0<\nu<1$ the mean service time is 1. In this case we need the component Y for X to be a CTMC.

Claim 3 *If the service times for station* 1 *are hyperexponential as described above, then the process* X *is positive recurrent.*

Proof In this case, Proposition 1 gives:

$$d_1 = 0.8 + 0.1[\mathbb{P}\{Z_1(\infty) < Z_2(\infty)\} + 0.5\mathbb{P}\{Z_1(\infty) = Z_2(\infty)\}]. \tag{13}$$

Then, by Proposition 2, X is positive recurrent iff

$$0.17 + 0.8 + 0.1\mathbb{P}\{Z_1(\infty) < Z_2(\infty)\} + 0.05\mathbb{P}\{Z_1(\infty) = Z_2(\infty)\} < 1,$$

or equivalently

$$10\mathbb{P}\{Z_1(\infty) < Z_2(\infty)\} + 5\mathbb{P}\{Z_1(\infty) = Z_2(\infty)\} < 3.$$

A sufficient condition for the inequality above to hold is

$$g(\nu) \equiv \mathbb{P}\{Z_1(\infty) \le Z_2(\infty)\} \; < \; 0.3. \tag{14}$$

We will show that this is true for ν sufficiently small. Observe that

$$Z_1(\infty) \ge_{\mathrm{st}} Z^{\nu}_{M/G/1} \quad \text{and} \quad Z_2(\infty) \le_{\mathrm{st}} Z_{M/M/1}, \tag{15}$$

where $Z^{\nu}_{M/G/1}$ denotes a random variable whose distribution is the stationary distribution of the number of customers in an ordinary $M/G/1$ queue with arrival rate 0.8 and service time distribution function F given by (12), and $Z_{M/M/1}$ denotes a random variable whose distribution is the stationary distribution of the number of customers in an ordinary $M/M/1$ queue with arrival rate 0.9 and service rate 1.

Since the Laplace-Stieltjes transform (LST) of service times in the $M/G/1$ queue is

$$\int_0^{\infty} e^{-sx}dF(x) = \frac{\nu^2}{(1-\nu+\nu^2)s+\nu} + \frac{1-\nu}{s\nu+1}, \quad \mathrm{Re}(s) > 0,$$

the Pollaczek-Khintchine (see, e.g., p. 260 in [12]) formula yields

$$\mathbb{E}[z^{Z^{\nu}_{M/G/1}}] = \frac{(4-3\nu+8\nu^2)-4(1-2\nu+2\nu^2)z}{5(4-3\nu+8\nu^2)-4(5-\nu+6\nu^2+4\nu^3)z+16\nu(1-\nu+\nu^2)z^2},$$

which is the probability generating function for the number of customers in the $M/G/1$ queue, in stationarity. Therefore

$$\lim_{\nu\to 0+} \mathbb{E}[z^{Z^{\nu}_{M/G/1}}] = 0.2, \quad |z| < 1.$$

Now applying the continuity theorem for probability generating functions (cf. Theorem 1.5.1 in [13]) we have

$$\lim_{\nu\to 0+} \mathbb{P}(Z^{\nu}_{M/G/1} \le x) = 0.2 \quad \text{for all } 0 < x < \infty. \tag{16}$$

By (15),

$$\begin{aligned} g(\nu) &= 1 - \mathbb{P}\{Z_1(\infty) > Z_2(\infty)\} \\ &\le 1 - \mathbb{P}\{Z_1(\infty) > x > Z_2(\infty)\} \\ &= 1 - \mathbb{P}(\{Z_2(\infty) < x\} - \{Z_1(\infty) \le x\}) \\ &\le 1 - \mathbb{P}\{Z_2(\infty) < x\} + \mathbb{P}\{Z_1(\infty) \le x\} \\ &\le 1 - \mathbb{P}(Z_{M/M/1} < x) + \mathbb{P}(Z^{\nu}_{M/G/1} \le x), \end{aligned}$$

for any $0 < x < \infty$. Hence, by (16),

$$\limsup_{\nu\to 0+} g(\nu) \le 1.2 - \mathbb{P}(Z_{M/M/1} < x).$$

Letting $x \to \infty$ leads to

$$\limsup_{\nu\to 0+} g(\nu) \le 0.2.$$

Therefore (14) holds for sufficiently small ν. Hence, for sufficiently small ν, X is positive recurrent when the service times for station 1 have the hyperexponential distribution function (12). □

Claims 2 and 3, taken together, show that the positive recurrence of X depends on more than just the mean values of the primitive distributions in the network.

4 The fluid model and stability

In this section we introduce the queueing and fluid dynamical equations, and provide results which relate the queueing model and fluid models defined by these equations. This framework allows us to use fluid model techniques to prove the main results in Sect. 5.

We now define a number of processes related to the queueing network:

$E(t) = \{E_C(t) : C \in \mathcal{E}\}, t \ge 0$, where $E_C(t)$ is the number of customers which arrive during $[0, t]$ due to the type-C external arrival process.

$A(t) = \{A_i(t) : i \in \mathcal{J}\}, t \ge 0$, where $A_i(t)$ is the number of arrivals to buffer i during $[0, t]$ (including exogenous arrivals and feedback arrivals).

$D(t) = \{D_i(t) : i \in \mathcal{J}\}, t \ge 0$, where $D_i(t)$ is the number of customers which complete service at station i during $[0, t]$.

$S(t) = \{S_i(t) : i \in \mathcal{J}\}, t \ge 0$, where $S_i(t)$ is the number of customers station i completes if it spends t units of time working on such customers.

$\Phi(n) = \{\Phi_{iC}(n) : i \in \mathcal{J}, C \in \mathcal{P}\}$, $n = 0, 1, 2, \ldots,$ where $\Phi_{iC}(n)$ is the number of customers, among the first n who depart station i, which become type-C customers.

$T(t)=\{T_i(t):i\in\mathcal{J}\}, t\geq 0$, where $T_i(t)$ is the amount of time spent working on customers at station i during $[0,t]$.

$I(t)=\{I_i(t):i\in\mathcal{J}\}, t\geq 0$, where $I_i(t)$ is the amount of time station i idles during $[0,t]$.

Then, the following equations define the dynamics of a JSQ network:

$$Z_i(t)=Z_i(0)+A_i(t)-D_i(t),\quad i\in\mathcal{J},\ t\geq 0, \tag{17}$$

$$Z_i(t)\geq 0,\quad i\in\mathcal{J},\ t\geq 0, \tag{18}$$

$$T_i(\cdot) \text{ and } I_i(\cdot) \text{ are nondecreasing},\quad i\in\mathcal{J}, \tag{19}$$

$$T_i(t)+I_i(t)=t,\quad i\in\mathcal{J},\ t\geq 0, \tag{20}$$

If $Z_i(u)>0$, for $u\in(s,t)$ then

$$I_i(s)=I_i(t),\ i\in\mathcal{J},\ 0\leq s\leq t, \tag{21}$$

$$D_i(t)=S_i(T_i(t)),\quad i\in\mathcal{J},\ t\geq 0, \tag{22}$$

$$\begin{aligned}\sum_{i\in C}(A_i(t)-A_i(s)) &\geq \sum_{B:B\subseteq C}\Big\{(E_B(t)-E_B(s)) \\ &\quad +\sum_{i\in\mathcal{J}}(\Phi_{iB}(D_i(t))-\Phi_{iB}(D_i(s)))\Big\}, \\ &\quad C\in\mathcal{P}, 0\leq s\leq t.\end{aligned} \tag{23}$$

$$\begin{aligned}\sum_{i\in C}(A_i(t)-A_i(s)) &= \sum_{B:B\subseteq C}\Big\{(E_B(t)-E_B(s)) \\ &\quad +\sum_{i\in\mathcal{J}}(\Phi_{iB}(D_i(t))-\Phi_{iB}(D_i(s)))\Big\}, \\ &\quad C\in\mathcal{P}, 0\leq s\leq t, \\ &\quad \text{if } Z_i(u)>Z_j(u) \text{ for all } i\in C,\ j\in\mathcal{J}-C \text{ and } u\in(s,t).\end{aligned} \tag{24}$$

Equations (17–22) are standard equations for generalized Jackson networks operating under an arbitrary non-idling policy. The last two equations however, are new, and they enforce the JSQ routing behavior of the customers.

Using the dynamical equations (17–24) we derive the corresponding fluid model equations. Our methodology closely follows a now standard procedure and we only outline the general steps. By the strong law of large numbers, for almost all sample paths ω, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_C(k,\omega)=\lambda_C^{-1},\quad C\in\mathcal{E}, \tag{25}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\eta_i(k,\omega)=\mu_i^{-1},\quad i\in\mathcal{J}, \tag{26}$$

$$\lim_{n\to\infty}\frac{1}{n}\Phi_{iC}(n,\omega)=p_{iC},\quad i\in\mathcal{J},\ C\in\mathcal{P}. \tag{27}$$

Let $\mathbb{X}\equiv\{(A(t),T(t),I(t),Z(t)), t\geq 0\}$ be a network process governed by (17–24), and $\mathbb{X}_x$ be such a process with initial state $x=(z,u,v)$. By taking $C=\mathcal{J}$ in (24), one has for each station k and each $0\leq s<t$ that

$$\begin{aligned}A_k(t)-A_k(s) &\leq \sum_{i\in\mathcal{J}}(A_i(t)-A_i(s)) \\ &= \sum_{B:B\subseteq\mathcal{J}}\Big\{(E_B(t)-E_B(s)) \\ &\quad +\sum_{i\in\mathcal{J}}(\Phi_{iB}(D_i(t))-\Phi_{iB}(D_i(s)))\Big\}.\end{aligned}$$

It follows from the same argument as in Dai [2] that for every sample path ω satisfying (25–27) and every collection $\{x_r:r>0\}$ of initial states such that $\{|x_r|/r:r>0\}$ is bounded, there exists a subsequence $r_n\to\infty$ such that $\frac{1}{r_n}\mathbb{X}_{x_{r_n}}(r_n\cdot,\omega)$ converges uniformly on any compact subset of $[0,\infty)$ to some limit say $\bar{\mathbb{X}}=(\bar{A}(\cdot),\bar{T}(\cdot),\bar{I}(\cdot),\bar{Z}(\cdot))$. Each such limit $\bar{\mathbb{X}}$ is called a *fluid limit*. In the special case where the sequence of initial states $\{x_r:r>0\}$ is independent of r, we call the limit a *fluid limit with fixed initial state*. Both types of fluid limits are used in our subsequent stability analysis of the process $\mathbb{X}$.

As shown in Bramson [1], in the analysis of stability via fluid limits, it is sufficient to consider the so-called undelayed fluid limit, i.e. when

$$\lim_{r\to\infty}\frac{1}{r}(|u_r|+|v_r|)=0, \tag{28}$$

where u_r and v_r are subvectors of the initial state $x_r=(z_r,u_r,v_r)$. Thus, from now on we only consider undelayed fluid limits.

Now, let $\bar{\mathbb{X}}$ be a fluid limit obtained from a sequence of initial states $\{x_r\}$ satisfying (28). It is readily seen that all of $\bar{A}_i(\cdot)$, $\bar{T}_i(\cdot)$, $\bar{I}_i(\cdot)$ and $\bar{Z}_i(\cdot)$, $i\in\mathcal{J}$, are Lipschitz continuous. Hence they are absolutely continuous and thus differentiable almost everywhere with respect to the Lebesgue measure. We say that t is a regular point of $\bar{\mathbb{X}}$ if all components of $\bar{\mathbb{X}}$ are differentiable at t. From now on, we implicitly assume that t is a regular point whenever the derivative of a component of $\bar{\mathbb{X}}$ is involved. Applying fluid limits to (17–28), we obtain the equations:

$$\bar{Z}_i(t)=\bar{Z}_i(0)+\bar{A}_i(t)-\mu_i\bar{T}_i(t),\quad i\in\mathcal{J},\ t\geq 0, \tag{29}$$

$$\bar{Z}_i(t)\geq 0,\quad i\in\mathcal{J},\ t\geq 0, \tag{30}$$

$$\bar{T}_i(\cdot) \text{ and } \bar{I}_i(\cdot) \text{ are nondecreasing},\quad i\in\mathcal{J}, \tag{31}$$

$$\bar{T}_i(t) + \bar{I}_i(t) = t, \quad i \in \mathcal{J},\ t \geq 0, \tag{32}$$

$$\text{If } \bar{Z}_i(t) > 0, \quad \text{then } \dot{\bar{I}}_i(t) = 0, \quad i \in \mathcal{J},\ t \geq 0, \tag{33}$$

$$\sum_{i \in C} \dot{\bar{A}}_i(t) \geq \Lambda_C + \sum_{i \in \mathcal{J}} P_{iC}\mu_i \dot{\bar{T}}_i(t), \quad C \in \mathcal{P},\ t \geq 0, \tag{34}$$

$$\sum_{i \in C} \dot{\bar{A}}_i(t) = \Lambda_C + \sum_{i \in \mathcal{J}} P_{iC}\mu_i \dot{\bar{T}}_i(t), \quad C \in \mathcal{P}, t \geq 0,$$
$$\text{if } Z_i(t) > Z_j(t) \quad \text{for all } i \in C \text{ and } j \in \mathcal{J} - C, \tag{35}$$

where

$$\Lambda_C \equiv \sum_{B:\phi \neq B \subset C} \lambda_B \quad \text{and} \quad P_{iC} \equiv \sum_{B:\phi \neq B \subset C} p_{iB}.$$

We call the equations (29–35) the *fluid model equations* and call a solution $\bar{\mathbb{X}} = \{(\bar{A}(t), \bar{T}(t), \bar{I}(t), \bar{Z}(t)), t \geq 0\}$, of the fluid model equations a *fluid model solution*. Note that any fluid limit with fixed initial state necessarily has $\bar{Z}(0) = 0$. Thus these fluid limits form a subset of fluid solutions with $\bar{Z}(0) = 0$. The following definitions and lemmas indicate the usefulness of different types of fluid limits.

Definition 1 (i) The fluid model is *stable* if there exists a $\delta > 0$ such that for each fluid model solution $\bar{\mathbb{X}}$, with $|\bar{Z}(0)| \leq 1$, $\bar{Z}(t) = 0$ for $t \geq \delta$.

(ii) The fluid model is *weakly unstable* if there exists a $\delta > 0$ such that for each fluid model solution $\bar{\mathbb{X}}$, with $\bar{Z}(0) = 0$, $\bar{Z}(\delta) \neq 0$.

The same reasoning used in Dai [2, 3], can be applied to the class of networks we consider here to give the following criteria.

Lemma 1 (Dai [2]) *If the fluid model is stable, then the Markov process X is positive Harris recurrent.*

Lemma 2 (Dai [3]) *If the fluid model is weakly unstable, then the process X is unstable in the sense that, for each fixed initial state x, $|Z(t)| \to \infty$ as $t \to \infty$ with probability* 1.

If we assume a priori that the process X is positive recurrent, then any fluid limit with fixed initial state must obey an extra dynamical equation, which augments the fluid model equations presented in (29–35). It turns out that the augmented set of equations will be quite useful for proving non-positive recurrence using fluid model analysis.

So, suppose that X is positive Harris recurrent and let π be its stationary distribution. Since every station is non-idling, for each fixed initial state x,

$$\lim_{t\to\infty} \frac{T_i(t)}{t} = \lim_{t\to\infty} \frac{1}{t} \int_0^t 1_{\{Z_i(s)>0\}} ds$$
$$= \pi\left(\{(z,u,v) \in \mathcal{S} : z_i > 0\}\right) \quad \mathbb{P}_x\text{-a.s., } i \in \mathcal{J}.$$

Therefore,

$$\bar{T}_i(t) = t\pi\left(\{(z,u,v) \in \mathcal{S} : z_i > 0\}\right), \quad t \geq 0, i \in \mathcal{J}, \tag{36}$$

for every fluid limit $\bar{\mathbb{X}}$, which is a limit of scaled sample paths with a fixed initial state. Choose a compact set $\mathcal{K} \subset \mathcal{S}$ such that $\pi(\mathcal{K}) > 0$. By (2), there exists a $t_0 > 0$ such that for each $(z,u,v) \in \mathcal{K}$,

$$\mathbb{P}_{(z,u,v)}(|Z(t_0)| = 0) > 0.$$

Therefore

$$\begin{aligned}
&\pi(\{(z,u,v) \in \mathcal{S} : |z| = 0\}) \\
&\quad = \mathbb{P}_\pi(|Z(0)| = 0) = \mathbb{P}_\pi(|Z(t_0)| = 0) \\
&\quad = \int_{\mathcal{S}} \mathbb{P}_{(z,u,v)}(|Z(t_0)| = 0)\, d\pi(z,u,v) \\
&\quad \geq \int_{\mathcal{K}} \mathbb{P}_{(z,u,v)}(|Z(t_0)| = 0)\, d\pi(z,u,v) > 0.
\end{aligned}$$

Combining this with (36) yields

$$\dot{\bar{T}}_i(t) < 1, \quad i \in \mathcal{J}, \tag{37}$$

for every fluid limit $\bar{\mathbb{X}}$, which is a limit of scaled sample paths with fixed initial state.

We call the equations (29–35) plus (37) the *augmented fluid model equations* and call a solution $\bar{\mathbb{X}}$, to these equations an *augmented fluid model solution*.

Definition 2 The augmented fluid model is *weakly unstable* if there exists a $\delta > 0$ such that for each augmented fluid model solution $\bar{\mathbb{X}}$, with $\bar{Z}(0) = 0$, $\bar{Z}(\delta) \neq 0$.

Suppose that the augmented fluid model is weakly unstable but the Markov process X is positive Harris recurrent. Since the augmented fluid model equations are satisfied by every fluid limit which is a limit of scaled sample paths with fixed initial state, the argument in Dai [3] implies that the process is unstable in the sense that, $|Z(t)| \to \infty$ as $t \to \infty$ with probability 1, which is a contradiction. Therefore we obtain the following instability criterion.

Lemma 3 *If the augmented fluid model is weakly unstable, then the Markov process $\{X(t) : t \geq 0\}$ is not positive Harris recurrent.*

5 Proofs

In this section we prove the main results of the paper.

5.1 Proof of Theorem 1

Sufficiency: Suppose that $\bar{\mathbb{X}}$ is a fluid model solution. Let $f(t)=|\bar{Z}(t)|$. It is readily seen that the fluid model is stable if there exists an $\epsilon>0$ such that

$$\dot{f}(t)\le -\epsilon \quad \text{if } f(t)>0. \tag{38}$$

Hence, by Lemma 1, X is positive Harris recurrent if there exists an $\epsilon>0$ satisfying (38). By (29), $\dot{f}(t)$ can be written as

$$\dot{f}(t)=\dot{\bar{A}}_1(t)+\dot{\bar{A}}_2(t)-\mu_1\dot{\bar{T}}_1(t)-\mu_2\dot{\bar{T}}_2(t).$$

Employing (35) with $C=\{1,2\}$ we obtain,

$$\dot{f}(t)=\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1\dot{\bar{T}}_1(t) + (p_2^*-1)\mu_2\dot{\bar{T}}_2(t). \tag{39}$$

Now we show that (38) holds for some $\epsilon>0$ by considering three cases separately.

Case 1 Suppose $\bar{Z}_1(t)>0$ and $\bar{Z}_2(t)>0$. Then by (32) and (33), $\dot{\bar{T}}_i=1$, $i=1,2$. So (39) becomes

$$\dot{f}(t)=\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2,$$

which is negative by (i).

Case 2 $\bar{Z}_1(t)>0$ and $\bar{Z}_2(t)=0$.

By (32) and (33),

$$\dot{\bar{T}}_1(t)=1. \tag{40}$$

Substituting (40) into (39) gives,

$$\dot{f}(t)=\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2\dot{\bar{T}}_2(t). \tag{41}$$

Next, evaluating (35) with $C=\{1,2\}$ and using (40) yields

$$\dot{\bar{A}}_1(t)+\dot{\bar{A}}_2(t)=\lambda_1+\lambda_2+\lambda+p_1^*\mu_1+p_2^*\mu_2\dot{\bar{T}}_2(t). \tag{42}$$

Similarly, evaluating (35) with $C=\{1\}$, along with (40) yields

$$\dot{\bar{A}}_1(t)=\lambda_1+p_{11}\mu_1+p_{21}\mu_2\dot{\bar{T}}_2(t). \tag{43}$$

We subtract (43) from (42) to obtain

$$\dot{\bar{A}}_2(t)=\lambda_2+\lambda+(p_1^*-p_{11})\mu_1+(p_2^*-p_{21})\mu_2\dot{\bar{T}}_2(t). \tag{44}$$

By assumption $\bar{Z}_2(t)=0$ which implies $\dot{\bar{Z}}_2(t)=0$. Hence by (29),

$$\dot{\bar{A}}_2(t)=\mu_2\dot{\bar{T}}_2(t). \tag{45}$$

Therefore, substituting (45) into (44) gives

$$\mu_2(1-p_2^*+p_{21})\dot{\bar{T}}_2(t)=\lambda_2+\lambda+(p_1^*-p_{11})\mu_1. \tag{46}$$

Now, if $1-p_2^*+p_{21}=0$ then $p_2^*=1$ and so, by (41),

$$\dot{f}(t)=\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1,$$

which is negative by (i).

Otherwise, suppose $1-p_2^*+p_{21}>0$. Then, by (46),

$$\dot{\bar{T}}_2(t)=\frac{\lambda_2+\lambda+(p_1^*-p_{11})\mu_1}{\mu_2(1-p_2^*+p_{21})}. \tag{47}$$

In this case, (47) and (41) imply

$$\dot{f}(t)=\frac{p_{21}[\lambda_1+\lambda_2+\lambda+\mu_1(p_1^*-1)]}{1-p_2^*+p_{21}} + \frac{(1-p_2^*)[\lambda_1+\mu_1(p_{11}-1)]}{1-p_2^*+p_{21}},$$

which by (i) is negative when $p_2^*=1$ and by (ii) is negative when $p_2^*<1$.

Case 3 $\bar{Z}_1(t)=0$ and $\bar{Z}_2(t)>0$. The argument in this case is analogous to that of case 2.

Necessity: Lemma 3 implies that we need only show that the augmented fluid model is weakly unstable if any of (i)–(iii) of Theorem 1 does not hold. By symmetry, it is sufficient to analyze the three cases examined below. Let $\bar{\mathbb{X}}$ be an augmented fluid model solution with $\bar{Z}(0)=0$ and let

$$f(t)=|\bar{Z}(t)|, \quad t\ge 0.$$

Considering three cases separately, we show that $\dot{f}(t)>0$ for all regular $t>0$, which completes the proof.

Case 1 Suppose (i) does not hold. By (39) and (37),

$$\dot{f}(t)>\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2\ \ge\ 0,$$

which proves the result for this case.

Case 2 Suppose (i) holds and (ii) does not hold. If $\bar{Z}_2(t)>0$, then by (32) and (33), $\dot{\bar{T}}_2(t)=1$, which contradicts (37). Hence $\bar{Z}_2(t)=0$ and $\dot{\bar{Z}}_2(t)=0$. As before, by (29),

$$\dot{\bar{A}}_2(t)=\mu_2\dot{\bar{T}}_2(t). \tag{48}$$

By subtracting (34) evaluated at $C=\{1\}$ from (35) evaluated at $C=\{1,2\}$, we have

$$\dot{\bar{A}}_2(t)\le\lambda_2+\lambda+(p_1^*-p_{11})\mu_1\dot{\bar{T}}_1(t) + (p_2^*-p_{21})\mu_2\dot{\bar{T}}_2(t).$$

Hence by (37) and (48),

$$\dot{\bar{T}}_2(t) < \frac{\lambda_2+\lambda+(p_1^*-p_{11})\mu_1}{\mu_2(1-p_2^*+p_{21})}. \qquad (49)$$

Substituting (49) into (39) and applying $\dot{\bar{T}}_1(t)<1$ lead to

$$\dot{f}(t) > \frac{p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1 p_1^*-\mu_1)}{1-p_2^*+p_{21}} + \frac{(1-p_2^*)(\lambda_1+\mu_1 p_{11}-\mu_1)}{1-p_2^*+p_{21}}.$$

The numerator above is nonnegative by the negation of (ii), thus $\dot{f}(t)>0$.

5.2 Proof of Theorem 2

Suppose that $\bar{\mathbb{X}}$ is a fluid model solution with $\bar{Z}(0)=0$, $t\geq 0$. Let $f(t)=|\bar{Z}(t)|$. By Lemma 2, it suffices to show that $\dot{f}(t)>0$ for all $t>0$. We show this by considering three cases separately.

Case 1 Suppose (3) holds.
Since $\dot{\bar{T}}_1(t)\leq 1$ and $\dot{\bar{T}}_2(t)\leq 1$, by (39),

$$\dot{f}(t)\geq \lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1+(p_2^*-1)\mu_2>0,$$

for all $t>0$.

Case 2 Suppose (3) does not hold and (4) holds.
If $p_2^*=1$, then $\dot{f}(t)\geq\lambda_1+\lambda_2+\lambda+(p_1^*-1)\mu_1>0$ by (39) and (4). Now suppose that $p_2^*<1$. First we show that

$$\bar{Z}_2(t)=0, \quad t\geq 0. \qquad (50)$$

To prove (50), it suffices to show that $\dot{\bar{Z}}_2(t)\leq 0$ if $\bar{Z}_2(t)>0$. Suppose $\bar{Z}_2(t)>0$. Then by (32) and (33), $\dot{\bar{T}}_2(t)=1$. By (29),

$$\dot{\bar{A}}_2(t)=\dot{\bar{Z}}_2(t)+\mu_2. \qquad (51)$$

By subtracting (34) evaluated at $C=\{1\}$ from (35) evaluated at $C=\{1,2\}$, we have

$$\dot{\bar{A}}_2(t)\leq\lambda_2+\lambda+(p_1^*-p_{11})\mu_1\dot{\bar{T}}_1(t)+(p_2^*-p_{21})\mu_2. \qquad (52)$$

Substituting (51) into (52) and applying $\dot{\bar{T}}_1(t)\leq 1$ lead to

$$\dot{\bar{Z}}_2(t)\leq\lambda_2+\lambda+(p_1^*-p_{11})\mu_1+(p_2^*-p_{21}-1)\mu_2. \qquad (53)$$

Since by assumption, (3) does not hold,

$$\mu_2\geq\frac{\lambda_1+\lambda_2+\lambda+\mu_1(p_1^*-1)}{1-p_2^*}. \qquad (54)$$

Finally, by (53) and (54),

$$\dot{\bar{Z}}_2(t)\leq -\frac{p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1 p_1^*-\mu_1)}{1-p_2^*} - \frac{(1-p_2^*)(\lambda_1+\mu_1 p_{11}-\mu_1)}{1-p_2^*}.$$

Hence, $\dot{\bar{Z}}_2(t)<0$ by (4). Thus (50) holds.
Next, subtracting (34) evaluated at $C=\{1\}$ from (35) evaluated at $C=\{1,2\}$, we obtain

$$\begin{aligned}\dot{\bar{A}}_2(t) &\leq \lambda_2+\lambda+(p_1^*-p_{11})\mu_1\dot{\bar{T}}_1(t)\\ &\quad+(p_2^*-p_{21})\mu_2\dot{\bar{T}}_2(t)\\ &\leq\lambda_2+\lambda+(p_1^*-p_{11})\mu_1\\ &\quad+(p_2^*-p_{21})\mu_2\dot{\bar{T}}_2(t).\end{aligned} \qquad (55)$$

By (50), $\dot{\bar{Z}}_2(t)=0$ and so $\dot{\bar{A}}_2(t)=\mu_2\dot{\bar{T}}_2(t)$ by (29). Hence, employing (55), we have

$$\dot{\bar{T}}_2(t)\leq\frac{\lambda_2+\lambda+(p_1^*-p_{11})\mu_1}{\mu_2(1-p_2^*+p_{21})}. \qquad (56)$$

Substituting (56) into (39) and applying $\dot{\bar{T}}_1(t)\leq 1$ lead to

$$\dot{f}(t)\geq\frac{p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1 p_1^*-\mu_1)}{1-p_2^*+p_{21}} + \frac{(1-p_2^*)(\lambda_1+\mu_1 p_{11}-\mu_1)}{1-p_2^*+p_{21}}.$$

Thus (4) now implies $\dot{f}(t)>0$.

Case 3 Suppose (3) does not hold and (5) holds. By symmetry this case is completely analogous to Case 2.

5.3 Proof of Theorem 3

We first need to prove the following lemma.

Lemma 4 *Let $\bar{\mathbb{X}}$ be a fluid model solution. Consider a fixed regular $t>0$ and let $C\equiv C(t)=\{i\in\mathcal{J}:\bar{Z}_i(t)>0\}$. Then*

$$\sum_{i\in C}\dot{\bar{Z}}_i(t)=\frac{1-p^*}{1-p^*+P_C}\left(\Lambda_C+\frac{\lambda^*}{1-p^*}P_C-\mu_C\right). \qquad (57)$$

Proof Using (35), we have

$$\sum_{i\in\mathcal{J}}\dot{\bar{A}}_i(t)=\lambda^*+p^*\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t) \qquad (58)$$

and

$$\sum_{i\in C}\dot{\bar{A}}_i(t)=\Lambda_C+P_C\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t). \qquad (59)$$

Subtracting (59) from (58), yields

$$\sum_{i\in\mathcal{J}-C}\dot{\bar{A}}_i(t)=\lambda^*-\Lambda_C+(p^*-P_C)\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t). \tag{60}$$

Since $\bar{Z}_i(t)=0$ for $i\in\mathcal{J}-C$, $\dot{\bar{Z}}_i(t)=0$ for $i\in\mathcal{J}-C$. Hence, by (29),

$$\dot{\bar{A}}_i(t)=\mu_i\dot{\bar{T}}_i(t),\quad i\in\mathcal{J}-C. \tag{61}$$

Then (60) and (61) give

$$\sum_{i\in\mathcal{J}-C}\mu_i\dot{\bar{T}}_i(t)=\frac{\lambda^*-\Lambda_C+(p^*-P_C)\sum_{i\in C}\mu_i\dot{\bar{T}}_i(t)}{1-p^*+P_C}. \tag{62}$$

Since $\bar{Z}_i(t)>0$, (32) and (33) imply

$$\dot{\bar{T}}_i(t)=1,\quad i\in C. \tag{63}$$

Substituting (62) and (63) into (59) leads to

$$\sum_{i\in C}\dot{\bar{A}}_i(t)=\frac{(1-p^*)\Lambda_C+\lambda^*P_C+\mu_CP_C}{1-p^*+P_C}. \tag{64}$$

Next, (29) and (63) give us

$$\sum_{i\in C}\dot{\bar{Z}}_i(t)=\sum_{i\in C}\dot{\bar{A}}_i(t)-\mu_C. \tag{65}$$

Finally, substituting (64) into (65) yields (57). □

Proof of Sufficiency of Theorem 3 Suppose $\bar{\mathbb{X}}$ is a fluid model solution. Let $f(t)=|\bar{Z}(t)|$. Consider a fixed regular $t>0$ with $f(t)>0$ and again let $C=\{i\in\mathcal{J}:\bar{Z}_i(t)>0\}$, $t\geq 0$. Since $\dot{\bar{Z}}_i(t)=0$ for $i\in\mathcal{J}-C$, $\dot{f}(t)=\sum_{i\in C}\dot{\bar{Z}}_i(t)$. Hence, by Lemma 4,

$$\dot{f}(t)\leq-\epsilon, \tag{66}$$

for any such t, where

$$\epsilon=\min_{B\in\mathcal{P}}\frac{1-p^*}{1-p^*+P_B}\left(\mu_B-\Lambda_B-\frac{\lambda^*}{1-p^*}P_B\right)>0.$$

From (66), it is readily seen that the fluid model is stable. The proof is now completed by applying Lemma 1. □

Proof of Necessity of Theorem 3 By Lemma 3, it suffices to show that the augmented fluid model is weakly unstable if (6) does not hold for some $C\in\mathcal{P}$. Suppose then that (6) does not hold for some $C\in\mathcal{P}$. Let $\bar{\mathbb{X}}$ be an augmented fluid model solution with $\bar{Z}(0)=0$. In light of (29) and (34) we have

$$\sum_{i\in C}\dot{\bar{Z}}_i(t)\geq\Lambda_C+P_C\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)-\sum_{i\in C}\mu_i\dot{\bar{T}}_i(t). \tag{67}$$

Next, using (29) and (35), $\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t)=\lambda^*+(p^*-1)\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)$, which can be rewritten as

$$\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)=\frac{\lambda^*}{1-p^*}-\frac{1}{1-p^*}\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t). \tag{68}$$

Substituting (68) into (67) yields

$$\sum_{i\in C}\dot{\bar{Z}}_i(t)\geq\Lambda_C+\frac{\lambda^*}{1-p^*}P_C-\frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t)-\sum_{i\in C}\mu_i\dot{\bar{T}}_i(t).$$

Equation (37) then implies that

$$\frac{1-p^*+P_C}{1-p^*}\sum_{i\in C}\dot{\bar{Z}}_i(t)+\frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}-C}\dot{\bar{Z}}_i(t)>\Lambda_C+\frac{\lambda^*}{1-p^*}P_C-\mu_C. \tag{69}$$

Now, let

$$f(t)=\frac{1-p^*+P_C}{1-p^*}\sum_{i\in C}\bar{Z}_i(t)+\frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}-C}\bar{Z}_i(t).$$

Then by (69) and the negation of (6), $\dot{f}(t)>0$ for all $t>0$, which proves that the augmented fluid model is weakly unstable. □

5.4 Proof of Theorem 4

We first need the following lemma.

Lemma 5 *Let $\bar{\mathbb{X}}$ be a fluid model solution. Then, for any $C\in\mathcal{P}$,*

$$\frac{1-p^*+P_C}{1-p^*}\sum_{i\in C}\dot{\bar{Z}}_i(t)+\frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}-C}\dot{\bar{Z}}_i(t)\geq\Lambda_C+\frac{\lambda^*}{1-p^*}P_C-\mu_C. \tag{70}$$

Proof Equations (29) and (34) imply,

$$\sum_{i\in C}\dot{\bar{Z}}_i(t)\geq\Lambda_C+P_C\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)-\sum_{i\in C}\mu_i\dot{\bar{T}}_i(t). \tag{71}$$

Now (29) and (35) give $\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t)=\lambda^*+(p^*-1)\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)$, which can be rewritten as

$$\sum_{i\in\mathcal{J}}\mu_i\dot{\bar{T}}_i(t)=\frac{\lambda^*}{1-p^*}-\frac{1}{1-p^*}\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t). \tag{72}$$

By substituting (72) into (71), we get

$$\sum_{i\in C}\dot{\bar{Z}}_i(t) \ge \Lambda_C + \frac{\lambda^*}{1-p^*}P_C - \frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}}\dot{\bar{Z}}_i(t) - \sum_{i\in C}\mu_i\dot{\bar{T}}_i(t).$$

Since $\dot{\bar{T}}_i(t) \le 1$, $i \in C$, (70) is obtained. □

Proof of Theorem 4 Suppose $\bar{\mathbb{X}}$ is a fluid model solution with $\bar{Z}(0) = 0$, and let $C \in \mathcal{P}$ be such that it satisfies (7). Let

$$f(t) = \frac{1-p^*+P_C}{1-p^*}\sum_{i\in C}\bar{Z}_i(t) + \frac{P_C}{1-p^*}\sum_{i\in\mathcal{J}-C}\bar{Z}_i(t).$$

By Lemma 5, $\dot{f}(t) > 0$ for all $t > 0$. Thus $f(t) > 0$ and so $|\bar{Z}(t)| > 0$ for all $t > 0$. Hence the fluid model is weakly unstable and the proof is completed by applying Lemma 2. □

Appendix 1: Proofs of propositions

To prove Proposition 2, we first state and prove the following lemma, as applied to the network in Fig. 1. Clearly, the lemma can be extended to a general setting like multiclass queueing networks with general distributions as in Dai [2] or stochastic processing networks as in Dai and Lin [5].

Lemma 6 *Assume that the continuous time Markov chain* $X = \{(Z_1(t), Z_2(t), Z_3(t), Y(t)) : t \ge 0\}$ *is positive recurrent with stationary distribution* $\pi = \{\pi_{i_1,i_2,i_3,i_4} : (i_1,i_2,i_3,i_4) \in \mathbb{Z}_+^4\}$. *Let the initial state* $X(0) = x$ *be fixed. Then,* $\mathbb{P}_x$*-a.s., for each fluid limit* $(\bar{X}, \bar{T})$,

$$\bar{T}_j(t) = \left(1 - \sum_{(i_1,i_2,i_3,i_4)\in\mathbb{B}_j}\pi_{(i_1,i_2,i_3,i_4)}\right)t \quad \text{for each } j = 1,2,3 \text{ and each } t \ge 0, \tag{73}$$

where $\mathbb{B}_j = \{(i_1,i_2,i_3,i_4) \in \mathbb{Z}_+^4 : i_j = 0\}$.

Proof For notational convenience, we prove the case for $j = 1$. The proofs for other cases are identical.

Since a nonidling service policy is assumed, we have for each $s \ge 0$

$$\frac{T_1(s)}{s} = \frac{1}{s}\int_0^s \mathbf{1}_{\{Z_1(u)>0\}}\,du = 1 - \frac{1}{s}\int_0^s \mathbf{1}_{\{Z_1(u)=0\}}\,du.$$

By the positive recurrence of the Markov chain, we have

$$\mathbb{P}_x\left\{\lim_{s\to\infty}\frac{T_1(s)}{s} = 1 - \lim_{s\to\infty}\frac{1}{s}\int_0^s \mathbf{1}_{\{Z_1(u)=0\}}\,du = 1 - \sum_{(i_1,i_2,i_3,i_4)\in\mathbb{B}_1}\pi_{(i_1,i_2,i_3,i_4)}\right\} = 1. \tag{74}$$

For each sample path in the event set of (74) and for each $t \ge 0$,

$$\bar{T}_1(t) = \lim_{n\to\infty}\frac{T_1(nt)}{n} = t\lim_{s\to\infty}\frac{T_1(s)}{s} = t\left(1 - \sum_{(i_1,i_2,i_3,i_4)\in\mathbb{B}_1}\pi_{(i_1,i_2,i_3,i_4)}\right),$$

thus proving the lemma. □

Proof of Proposition 2 Let $X = \{(Z_1(t), Z_2(t), Z_3(t), Y(t)) : t \ge 0\}$ and let $r > 0$ be fixed. Recall that the 3-dimensional process $\{(Z_1(t), Z_2(t), Y(t)) : t \ge 0\}$ is an irreducible CTMC. If $\lambda_1 + \lambda_2 + \lambda_4 \ge \mu_1 + \mu_2$, then by Theorem 1, the 3-dimensional CTMC is not positive recurrent, and so neither is X. This establishes the necessity of the first condition in Proposition 2.

Thus, we assume that $\lambda_1 + \lambda_2 + \lambda_4 < \mu_1 + \mu_2$ throughout the remainder of this proof. Let $\{\kappa_{ijk}(r) : i, j, k\}$ be the stationary distribution of the 3-dimensional Markov chain $\{(Z_1(t), Z_2(t), Y(t)) : t \ge 0\}$. We now show that X is positive recurrent if and only if

$$\lambda_3 + \mu_1\left(1 - \sum_{j=0}^{\infty}\sum_{k=0}^{2}\kappa_{0jk}(r)\right) < \mu_3. \tag{75}$$

Fix an initial state $X(0)$, say, $X(0) = (0,0,0,0)$. Let $(\bar{X}, \bar{T})$ be a fluid limit. It follows that it satisfies the following fluid model equation (see, e.g., Dai [2])

$$\bar{Z}_3(t) = \lambda_3 t + \mu_1\bar{T}_1(t) - \mu_3\bar{T}_3(t), \quad t \ge 0,$$

Applying Lemma 6 to the 3-dimensional Markov chain, we have

$$\bar{Z}_3(t) = \left[\lambda_3 + \mu_1\left(1 - \sum_{j,k}\kappa_{0jk}(r)\right)\right]t - \mu_3\bar{T}_3(t), \quad t \ge 0. \tag{76}$$

Assume that X is positive recurrent with stationary distribution $\pi = \{\pi_{(i_1,i_2,i_3,i_4)}\}$, but that condition (75) does not hold. Since $\sum_{(i_1,i_2,i_4)\in\mathbb{Z}_+^3}\pi_{(i_1,i_2,0,i_4)} > 0$, it follows from Lemma 6 and (76) that $\bar{Z}_3(t) > 0$ for each fluid limit and each time $t > 0$. Therefore, the fluid limit model is weakly unstable as defined in [3]. It follows from Theorem 4.2 of [3] that X is transient, and hence not positive recurrent, contradicting the assumption that X is positive recurrent. Thus we have proved that X is positive recurrent only if (75) holds.

Now suppose that (75) holds. For each fluid limit $(\bar{X}, \bar{T})$, $\bar{Z}_3(t) \geq 0$ for each $t \geq 0$. Thus, (76) implies that

$$\bar{T}_3(t) \leq \frac{\lambda_3 + \mu_1(1 - \sum_{j,k} \kappa_{0jk}(r))}{\mu_3} \cdot t = (1 - \epsilon)t, \tag{77}$$

where

$$\epsilon = 1 - \frac{1}{\mu_3}\left(\lambda_3 + \mu_1\left(1 - \sum_{j,k} \kappa_{0jk}(r)\right)\right) > 0.$$

Since (77) holds for every fluid limit, we have, $\mathbb{P}_x$-a.s.,

$$\limsup_{t\to\infty} \frac{1}{t} T_3(t) \leq 1 - \epsilon.$$

Therefore, $\mathbb{P}_x$-a.s.,

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{Z_3(u)=0\}}\, du \geq \epsilon. \tag{78}$$

Let B be a finite set such that

$$\sum_{(i,j,k)\notin B} \kappa_{ijk}(r) < \epsilon. \tag{79}$$

Now, define $\tilde{B} \equiv \{(i, j, 0, k) : (i, j, k) \in B\}$. By (78) and (79), $\mathbb{P}_x$-a.s.,

$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{(Z_1(u),Z_2(u),Z_3(u),Y(u))\in\tilde{B}\}}\, du \\
&\quad \geq \liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{Z_3(u)=0\}}\, du \\
&\qquad - \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{(Z_1(u),Z_2(u),Y(u))\notin B\}}\, du \\
&\quad = \liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{Z_3(u)=0\}}\, du - \sum_{(i,j,k)\notin B} \kappa_{ijk}(r) \\
&\quad \geq \epsilon - \sum_{(i,j,k)\notin B} \kappa_{ijk}(r) > 0.
\end{aligned}$$

By Fatou's lemma and Fubini's theorem,

$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{P}_x\left\{(Z_1(u), Z_2(u), Z_3(u), Y(u)) \in \tilde{B}\right\} du \\
&\quad \geq \epsilon - \sum_{(i,j,k)\notin B} \kappa_{ijk}(r) > 0.
\end{aligned} \tag{80}$$

Since $\tilde{B}$ is a finite set, (80) implies that the Markov chain $\{(Z_1(t), Z_2(t), Z_3(t), Y(t)) : t \geq 0\}$ is positive recurrent. □

Appendix 2: An equivalent form of the stability condition in Theorem 1

In order to facilitate a comparison of our results with Kurkova's results [11] for a special case, we now show that "(i)–(iii) in Theorem 1" is equivalent to the following:

$$\begin{cases} \lambda_1 + \mu_2 p_{21} + \mu_1 p_{11} < \mu_1, \\ \lambda_2 + \mu_1 p_{12} + \mu_2 p_{22} < \mu_2, \\ \lambda_1 + \lambda_2 + \lambda + \mu_1 p_1^* + \mu_2 p_2^* < \mu_1 + \mu_2, \end{cases} \tag{81}$$

or

$$\begin{cases} \lambda_1 + \mu_2 p_{21} + \mu_1 p_{11} \geq \mu_1, \\ p_{21}(\lambda_1 + \lambda_2 + \lambda + \mu_1 p_1^* - \mu_1) \\ \quad + (1 - p_2^*)(\lambda_1 + \mu_1 p_{11} - \mu_1) < 0, \end{cases} \tag{82}$$

or

$$\begin{cases} \lambda_2 + \mu_1 p_{12} + \mu_2 p_{22} \geq \mu_2, \\ p_{12}(\lambda_1 + \lambda_2 + \lambda + \mu_2 p_2^* - \mu_2) \\ \quad + (1 - p_1^*)(\lambda_2 + \mu_2 p_{22} - \mu_2) < 0. \end{cases} \tag{83}$$

It is not difficult to see that "(i)–(iii) in Theorem 1" implies "(81) or (82) or (83)". Hence we show the converse of this only. By symmetry, it suffices to show that

(a) (81) implies (ii) in Theorem 1,
(b) (82) implies (i) in Theorem 1, and
(c) (82) implies (iii) in Theorem 1.

Proof of (a) Suppose that (81) holds and $p_2^* < 1$. If

$$\frac{\lambda_2 + \lambda + (p_1^* - p_{11})\mu_1}{1 - p_2^* + p_{21}} \leq \mu_2,$$

then a little algebra shows that this inequality plus the first inequality in (81) implies (ii) in Theorem 1. If

$$\frac{\lambda_2 + \lambda + (p_1^* - p_{11})\mu_1}{1 - p_2^* + p_{21}} > \mu_2,$$

then this plus the third inequality in (81), implies (ii) in Theorem 1. □

Proof of (b) Suppose that (82) holds. The second inequality in (82) implies that either $p_{21} > 0$ or $p_2^* < 1$. So, $1 - p_2^* + p_{21}$ is positive and we can then use, respectively, the first and second inequalities in (82) to obtain

$$\begin{aligned}
&\lambda_1 + \mu_2 p_{21} + \mu_1(p_{11} - 1) \\
&\quad \geq 0 \\
&\quad > \frac{p_{21}(\lambda_1 + \lambda_2 + \lambda + \mu_1 p_1^* - \mu_1)}{1 - p_2^* + p_{21}} \\
&\qquad + \frac{(1 - p_2^*)(\lambda_1 + \mu_1 p_{11} - \mu_1)}{1 - p_2^* + p_{21}}.
\end{aligned}$$

A reduction of this yields

$$\lambda_2+\lambda+(p_1^*-p_{11})\mu_1-p_{21}\mu_2-(1-p_2^*)\mu_2<0. \tag{84}$$

Next, (84) and the second inequality in (82) give

$$\begin{aligned}&[\lambda_2+\lambda+(p_1^*-p_{11})\mu_1-p_{21}\mu_2-(1-p_2^*)\mu_2](1-p_2^*)\\&\quad+[p_{21}(\lambda_1+\lambda_2+\lambda+\mu_1p_1^*-\mu_1)\\&\quad+(1-p_2^*)(\lambda_1+\mu_1p_{11}-\mu_1)]<0,\end{aligned}$$

and a final reduction produces

$$\begin{aligned}&(1-p_2^*+p_{21})(\lambda_1+\lambda_2+\lambda+\mu_1p_1^*-\mu_1+\mu_2p_2^*-\mu_2)\\&\quad<0.\end{aligned}$$

Thus, (i) in Theorem 1 is obtained. □

Proof of (c) Suppose that (82) holds and $p_1^*<1$. The first inequality in (82) can be used to obtain

$$\lambda_1+\mu_2p_{21}+\mu_1p_{11}+\lambda+\mu_1p_{1\{1,2\}}+\mu_2p_{2\{1,2\}}\ge\mu_1,$$

which can be rewritten as

$$\lambda_1+\lambda+(p_2^*-p_{22})\mu_2+(p_1^*-p_{12})\mu_1\ge\mu_1.$$

Some algebra yields

$$\frac{\lambda_1+\lambda+(p_2^*-p_{22})\mu_2}{1-p_1^*+p_{12}}\ge\mu_1. \tag{85}$$

Note that (i) in Theorem 1 holds by (b). Then combining (85) with (i) in Theorem 1, we have (iii) in Theorem 1. □

Appendix 3: A network with two queues

We consider the network depicted in Fig. 1. Customers arrive according to a Poisson process with rate λ. These arrivals join the shorter of the two queues 1 and 2. If the queue lengths are equal at the time of an arrival, the customer breaks the tie using an independent Bernoulli(r) random variable, with a success indicating that the customer joins queue 1. Therefore, following the notation in Sect. 1, we have

$$\gamma_{\{1,2\},1}=r \quad\text{and}\quad \gamma_{\{1,2\},2}=1-r.$$

The service times are all assumed to be independent and exponentially distributed with mean $1/\mu$. A customer who finishes a service by station 1 or station 2 departs from the system permanently.

Since all distributions are exponential, $Z=\{(Z_1(t),Z_2(t)):t\ge0\}$ is a Markov chain, where, as before, $Z_i(t)$ is the number of customers that are either waiting in queue i or being served by station i at time t. By Theorem 1, when $\lambda<2\mu$, the process Z is positive recurrent. For a given value r, let $\kappa_{ij}(r)$ be the probability that Z is in state (i,j) under the stationary distribution of Z.

Lemma 7 *For all r_1,r_2 such that $0\le r_1<r_2\le1$,*

(i) *If $i>j$, then $\kappa_{ij}(r_1)<\kappa_{ij}(r_2)$.*
(ii) *If $i<j$, then $\kappa_{ij}(r_1)>\kappa_{ij}(r_2)$.*
(iii) *If $i=j$, then $\kappa_{ij}(r_1)=\kappa_{ij}(r_2)$.*
(iv) *$\kappa_{ij}(r_1)+\kappa_{ji}(r_1)=\kappa_{ij}(r_2)+\kappa_{ji}(r_2)$.*

Proof Let $Q^{(r)}=(q^{(r)}_{(i,j)(i'j')})$ be the rate matrix governing Z, i.e.,

$$q^{(r)}_{(i,j)(i',j')}=\begin{cases}\mu & \text{if } (i',j')=(i-1,j)\\ & \text{or } (i',j')=(i,j-1),\\ \lambda & \text{if } (i',j')=(i+1,j), i<j\\ & \text{or } (i'j')=(i,j+1), i>j,\\ r\lambda & \text{if } (i',j')=(i+1,j), i=j,\\ (1-r)\lambda & \text{if } (i',j')=(i,j+1), i=j,\\ -(\lambda+2\mu) & \text{if } (i',j')=(i,j), i>0, j>0,\\ -(\lambda+\mu) & \text{if } (i',j')=(i,j), i+j=1,\\ -\lambda & \text{if } (i',j')=(i,j)=(0,0),\\ 0 & \text{otherwise.}\end{cases}$$

Next, let $p^{(r)}_{(i,j)(i'j')}$, $(i,j),(i',j')\in\{0,1,\dots\}^2$ be the one-step transition probabilities of the embedded discrete Markov chain (in this case we include virtual transitions needed by the standard uniformization technique). Then we have

$$p^{(r)}_{(i,j)(i'j')}=\begin{cases}q^{(r)}_{(i,j)(i',j')}/(1+\lambda+2\mu)\\ \quad\text{if } (i',j')\neq(i,j),\\ 1+q^{(r)}_{(i,j)(i',j')}/(1+\lambda+2\mu)\\ \quad\text{if}(i',j')=(i,j).\end{cases}$$

Define a sequence of probability distributions $\{\kappa^{(n)}_{ij}:i,j=0,1,2,\dots\}$, $n=0,1,2,\dots$, as follows

$$\kappa^{(0)}_{ij}=\kappa_{ij}(r_1), \tag{86}$$

$$\kappa^{(n)}_{ij}=\sum_{i'j'}\kappa^{(n-1)}_{i'j'}\,p^{(r_2)}_{(i',j')(i,j)},\quad n=1,2,\dots. \tag{87}$$

Since $p^{(r_2)}_{(i,j)(i'j')}$, $(i,j),(i',j')\in\{0,1,\dots\}^2$, are the transition probabilities of the embedded chain, when $r\equiv r_2$, we conclude

$$\kappa_{ij}(r_2)=\lim_{n\to\infty}\kappa^{(n)}_{ij}. \tag{88}$$

The lemma can now be proved using (86), (88) and the following claim: for $n\ge1$,

(a) If $i>j$, then $\kappa^{(n-1)}_{ij}=\kappa^{(n)}_{ij}$ for $n<i-j$ and $\kappa^{(n-1)}_{ij}<\kappa^{(n)}_{ij}$ for $n\ge i-j$;

(b) If $i < j$, then $\kappa_{ij}^{(n-1)} = \kappa_{ij}^{(n)}$ for $n < j-i$ and $\kappa_{ij}^{(n-1)} > \kappa_{ij}^{(n)}$ for $n \geq j-i$;

(c) If $i = j$, then $\kappa_{ij}^{(n-1)} = \kappa_{ij}^{(n)}$;

(d) $\kappa_{ij}^{(n-1)} + \kappa_{ji}^{(n-1)} = \kappa_{ij}^{(n)} + \kappa_{ji}^{(n)}$.

We show the claim by induction on n. Since $\kappa_{ij}^{(0)}$ is a stationary probability of Z when $r \equiv r_1$, we have

$$\kappa_{ij}^{(0)} = \sum_{i'j'} \kappa_{i'j'}^{(0)} p_{(i',j')(i,j)}^{(r_1)},$$

which can be written explicitly as follows

$$\kappa_{ij}^{(0)} \equiv \begin{cases} \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(0)} + \mu\kappa_{i+1,j}^{(0)} + \mu\kappa_{i,j+1}^{(0)} + \kappa_{ij}^{(0)}), & 1 \leq i < j-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i,j-1}^{(0)} + \mu\kappa_{i+1,j}^{(0)} + \mu\kappa_{i,j+1}^{(0)} + \kappa_{ij}^{(0)}), & 1 \leq j < i-1, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{1,j}^{(0)} + \mu\kappa_{0,j+1}^{(0)} + (\mu+1)\kappa_{0j}^{(0)}), & i=0,\ j \geq 2, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{i+1,0}^{(0)} + \mu\kappa_{i1}^{(0)} + (\mu+1)\kappa_{i0}^{(0)}), & i \geq 2,\ j=0, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(0)} + \lambda(1-r_1)\kappa_{i,j-1}^{(0)} + \mu\kappa_{i+1,j}^{(0)} + \mu\kappa_{i,j+1}^{(0)} + \kappa_{ij}^{(0)}), & 1 \leq i = j-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i,j-1}^{(0)} + \lambda r_1\kappa_{i-1,j}^{(0)} + \mu\kappa_{i+1,j}^{(0)} + \mu\kappa_{i,j+1}^{(0)} + \kappa_{ij}^{(0)}), & 1 \leq j = i-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda(1-r_1)\kappa_{00}^{(0)} + \mu\kappa_{11}^{(0)} + \mu\kappa_{02}^{(0)} + (\mu+1)\kappa_{01}^{(0)}), & i=0,\ j=1, \\ \frac{1}{\lambda+2\mu+1}(\lambda r_1\kappa_{00}^{(0)} + \mu\kappa_{20}^{(0)} + \mu\kappa_{11}^{(0)} + (\mu+1)\kappa_{10}^{(0)}), & i=1,\ j=0, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(0)} + \lambda\kappa_{i,j-1}^{(0)} + \mu\kappa_{i+1,j}^{(0)} + \mu\kappa_{i,j+1}^{(0)} + \kappa_{ij}^{(0)}), & 1 \leq i = j, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{10}^{(0)} + \mu\kappa_{01}^{(0)} + (2\mu+1)\kappa_{00}^{(0)}), & i=j=0. \end{cases} \tag{89}$$

Next, the recursion (87) can be written out as follows

$$\kappa_{ij}^{(n)} \equiv \begin{cases} \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(n-1)} + \mu\kappa_{i+1,j}^{(n-1)} + \mu\kappa_{i,j+1}^{(n-1)} + \kappa_{ij}^{(n-1)}), & 1 \leq i < j-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i,j-1}^{(n-1)} + \mu\kappa_{i+1,j}^{(n-1)} + \mu\kappa_{i,j+1}^{(n-1)} + \kappa_{ij}^{(n-1)}), & 1 \leq j < i-1, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{1,j}^{(n-1)} + \mu\kappa_{0,j+1}^{(n-1)} + (\mu+1)\kappa_{0j}^{(n-1)}), & i=0,\ j \geq 2, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{i+1,0}^{(n-1)} + \mu\kappa_{i1}^{(n-1)} + (\mu+1)\kappa_{i0}^{(n-1)}), & i \geq 2,\ j=0, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(n-1)} + \lambda(1-r_2)\kappa_{i,j-1}^{(n-1)} + \mu\kappa_{i+1,j}^{(n-1)} + \mu\kappa_{i,j+1}^{(n-1)} + \kappa_{ij}^{(n-1)}), & 1 \leq i = j-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i,j-1}^{(n-1)} + \lambda r_2\kappa_{i-1,j}^{(n-1)} + \mu\kappa_{i+1,j}^{(n-1)} + \mu\kappa_{i,j+1}^{(n-1)} + \kappa_{ij}^{(n-1)}), & 1 \leq j = i-1, \\ \frac{1}{\lambda+2\mu+1}(\lambda(1-r_2)\kappa_{00}^{(n-1)} + \mu\kappa_{11}^{(n-1)} + \mu\kappa_{02}^{(n-1)} + (\mu+1)\kappa_{01}^{(n-1)}), & i=0,\ j=1, \\ \frac{1}{\lambda+2\mu+1}(\lambda r_2\kappa_{00}^{(n-1)} + \mu\kappa_{20}^{(n-1)} + \mu\kappa_{11}^{(n-1)} + (\mu+1)\kappa_{10}^{(n-1)}), & i=1,\ j=0, \\ \frac{1}{\lambda+2\mu+1}(\lambda\kappa_{i-1,j}^{(n-1)} + \lambda\kappa_{i,j-1}^{(n-1)} + \mu\kappa_{i+1,j}^{(n-1)} + \mu\kappa_{i,j+1}^{(n-1)} + \kappa_{ij}^{(n-1)}), & 1 \leq i = j, \\ \frac{1}{\lambda+2\mu+1}(\mu\kappa_{10}^{(n-1)} + \mu\kappa_{01}^{(n-1)} + (2\mu+1)\kappa_{00}^{(n-1)}), & i=j=0. \end{cases} \tag{90}$$

For $n = 1$, (89) and (90) yield,

(a′) If $i > j+1$, then $\kappa_{ij}^{(0)} = \kappa_{ij}^{(1)}$; if $i = j+1$, then $\kappa_{ij}^{(0)} < \kappa_{ij}^{(1)}$;

(b′) If $i < j-1$, then $\kappa_{ij}^{(0)} = \kappa_{ij}^{(1)}$; if $i = j-1$, then $\kappa_{ij}^{(0)} > \kappa_{ij}^{(1)}$;

(c′) If $i = j$, then $\kappa_{ij}^{(0)} = \kappa_{ij}^{(1)}$;

(d′) $\kappa_{ij}^{(0)} + \kappa_{ji}^{(0)} = \kappa_{ij}^{(1)} + \kappa_{ji}^{(1)}$.

Hence the claim holds when $n = 1$. Now, suppose that the claim holds for $n = k$ ($k \geq 1$). Then, we have

(a″) If $i > j+k$, then $\kappa_{ij}^{(k-1)} = \kappa_{ij}^{(k)}$; if $j < i \leq j+k$, then $\kappa_{ij}^{(k-1)} < \kappa_{ij}^{(k)}$;

(b″) If $i < j-k$, then $\kappa_{ij}^{(k-1)} = \kappa_{ij}^{(k)}$; if $j-k \leq i < j$, then $\kappa_{ij}^{(k-1)} > \kappa_{ij}^{(k)}$;

(c″) If $i = j$, then $\kappa_{ij}^{(k-1)} = \kappa_{ij}^{(k)}$;

(d″) $\kappa_{ij}^{(k-1)} + \kappa_{ji}^{(k-1)} = \kappa_{ij}^{(k)} + \kappa_{ji}^{(k)}$.

Using the above and (90) with $n = k$ and $n = k + 1$, we obtain

(a‴) If $i > j + k + 1$, then $\kappa_{ij}^{(k)} = \kappa_{ij}^{(k+1)}$; if $j < i \le j + k + 1$, then $\kappa_{ij}^{(k)} < \kappa_{ij}^{(k+1)}$;

(b‴) If $i < j - k - 1$, then $\kappa_{ij}^{(k)} = \kappa_{ij}^{(k+1)}$; if $j - k - 1 \le i < j$, then $\kappa_{ij}^{(k)} > \kappa_{ij}^{(k+1)}$;

(c‴) If $i = j$, then $\kappa_{ij}^{(k)} = \kappa_{ij}^{(k+1)}$;

(d‴) $\kappa_{ij}^{(k)} + \kappa_{ji}^{(k)} = \kappa_{ij}^{(k+1)} + \kappa_{ji}^{(k+1)}$.

Hence the claim holds for $n = k + 1$. Therefore the claim holds for all $n \ge 1$ by induction. □

References

1. Bramson, M.: Stability of two families of queueing networks and a discussion of fluid limits. Queueing Syst. Theory Appl. **28**, 7–31 (1998)
2. Dai, J.G.: On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. Ann. Appl. Probab. **5**, 49–77 (1995)
3. Dai, J.G.: A fluid-limit model criterion for instability of multiclass queueing networks. Ann. Appl. Probab. **6**, 751–757 (1996)
4. Dai, J.G.: Stability of fluid and stochastic processing networks. In: MaPhySto Miscellanea Publication, No. 9, Centre for Mathematical Physics and Stochastics (1999)
5. Dai, J.G., Lin, W.: Maximum pressure policies in stochastic processing networks. Oper. Res. **53**, 197–218 (2005)
6. Foley, R.D., McDonald, D.R.: Bridges and networks: exact asymptotics. Ann. Appl. Probab. **15**, 542–586 (2005)
7. Foley, R.D., McDonald, D.R.: Large deviations of a modified Jackson network: stability and rough asymptotics. Ann. Appl. Probab. **15**, 519–541 (2005)
8. Foss, S.G., Chernova, N.: On the stability of a partially accessible multi-station queue with state-dependent routing. Queueing Syst. Theory Appl. **29**, 55–73 (1998)
9. Jackson, J.R.: Networks of waiting lines. Oper. Res. **5**, 518–521 (1957)
10. Jackson, J.R.: Jobshop-like queueing systems. Manag. Sci. **10**, 131–142 (1963)
11. Kurkova, I.A.: A load-balanced network with two servers. Queueing Syst. Theory Appl. **37**, 379–389 (2001)
12. Medhi, J.: Stochastic Models in Queueing Theory, 2nd edn. Academic, San Diego (2002)
13. Resnick, S.I.: Adventures in Stochastic Processes. Birkhäuser, Boston (1992)
14. Suhov, Y.M., Vvedenskaya, N.D.: Fast Jackson networks with dynamic routing. Probl. Inf. Transm. **38**, 136–153 (2002)
15. Wolff, R.W.: Stochastic Modeling and the Theory of Queues. Prentice Hall, Englewood Cliffs (1989)