A tandem queueing model with coupled processors

A tandem queueing model with coupled processors

Jacques Resing
Department of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Lerzan Örmeci *
Department of Industrial Engineering
Koc University
Sariyer-Istanbul
Turkey

Abstract

We consider a tandem queueing model consisting of two stations. Special feature of the model is that the total service capacity of the stations together is constant. When both stations are nonempty, a given proportion of this capacity is allocated to the first station and the remaining part to the second station. However, if one of the stations becomes empty, the total capacity of the two stations together is allocated to the other station.

The model is motivated by a situation encountered in multi-access communication in cable TV networks. Before users are actually allowed to transmit data over a communication channel, they first have to obtain a kind of grant in order to avoid collisions. The total capacity of the communication channel is divided over the two different stages: allocation of the grants on one hand and transmission of actual data on the other hand.

We study the two-dimensional Markov process representing the numbers of jobs in the two stations. A functional equation for the generating function of the stationary distribution of this Markov process is derived and the solution of the functional equation is obtained. In the analysis we use the theory of Riemann-Hilbert boundary value problems.

*The research was done while the author was at EURANDOM, Eindhoven, The Netherlands

1 Introduction

In this paper we consider a tandem queueing model consisting of two stations. Jobs arrive at the first station according to a Poisson process. After receiving service at this station, they move to the second station, and upon completion of service at the second station they leave the system. The amount of work that a job requires at a station is an exponentially distributed random variable. The total service capacity of the two stations together is constant. When both stations are nonempty, a given proportion of the capacity is allocated to station 1, and the remaining proportion is allocated to station 2. However, if one of the stations is empty, the total service capacity of the stations is allocated to the other station. The model we consider is motivated by the following situation encountered in cable TV networks.

Cable TV networks are currently being upgraded to enable bidirectional communications between the network terminations (NTs) at the customer premises and a centrally located head end (HE). In order to coordinate upstream transmission (i.e., from NTs to HE) a medium access protocol is needed. This protocol can be a *request-grant* mechanism consisting of two stages. At the first stage, an NT which has data to transmit sends a *request* to the HE in a dedicated time slot to specify the number of data slots it needs. If only one NT sends a request in a certain time slot, then the HE receives the request successfully. If more NTs send a request simultaneously in a certain time slot, a collision occurs, upon which a collision resolution algorithm (CRA) is started for these NTs. The NTs involved in the collision have to retransmit their request. Hence, for a request to reach the HE successfully, a random number of time slots is needed, depending on the number of NTs involved in the collision and the CRA employed by the system. Upon receiving a request successfully, the HE starts the second stage of the mechanism, the actual data transmission, by sending a *grant* to the corresponding NT to transmit its data in specified data slots. Note that also the actual transmission of data from the NTs to the HE needs a random number of time slots since each NT has a different amount of data to transmit. Furthermore, the capacity of the upstream channel is divided between these two stages by the appropriate use of time slots. Some of the time slots are dedicated to data transmission of NTs already having a grant, and the rest is dedicated to requests of NTs not yet having a grant. In our model, service at station 1 represents the process of receiving the requests, whereas service at station 2 represents the transmission of the actual data corresponding to the successfully received requests. Hence, the total server capacity represents the total upstream bandwidth, and its allocation to the two stations corresponds to the time-sharing of the upstream channel by the two stages described above.

What is a clever way to divide the total service capacity over the two individual service stations? In [6], Klimov considers the minimization of the average holding costs in a time-sharing queueing system with a number of stations in series attended by a single server. In the case of two stations in series with the objective of minimizing the average sojourn time, Klimov's results imply that the optimal policy would be to allocate the whole capacity to the second station whenever this station is not empty. However, in the above mentioned application, the policy to first allocate time slots for data transmission to the NTs which requests have already been received, and allocate only the remaining time slots for receiving new requests, turns out to be not very sensible. The reason for this is that there is a *round trip delay* (RTD) on the collision feedback. Upon a collision, the HE announces that a collision occurred at a certain slot and all the NTs which tried to send a request in that slot have to try again according to the CRA. However, this announcement reaches the NTs only after some time due to the RTD. This feedback delay is very difficult to incorporate in a queueing model, but also may not be ignored completely due to its substantial effect on the whole process. Sala et al. [11] show, through simulations, that whenever the feedback delay is long, mean sojourn times at each stage can be shortened by allocating request slots on a more regular basis. That is why we study the model in which a fixed part of the total service capacity is always allocated to the first station, and only the remaining part to the second station when both stations have at least one job. Only when one of the two stations is empty, the total service capacity is allocated to the other station.

Systems in which the service rates of stations change at the moments that one of the stations becomes empty, are known in the literature as systems with coupled processors. In a pioneering paper, Fayolle and Iasnogorodski [4] were the first to consider such a system. They analyzed two coupled servers *in parallel* with exponential service times and derived a solution for the generating function of the stationary distribution of the Markov process describing the number of jobs in both queues, using the theory of Riemann-Hilbert boundary value problems. Konheim, Meilijson and Melkman [7] determined the generating function of the joint queue length distribution in the completely symmetric case (identical arrival and service rate at both servers) using a uniformization method. In Cohen and Boxma [3], the ordinary coupled processor model is analyzed for the case of generally distributed service times. Our model can be viewed as the *tandem version* of the model in [3, 4, 7]. Like the ordinary coupled processor model, our model will also be analyzed using the theory of boundary value problems.

Another way to divide the total service capacity over the individual service stations is to completely allocate the total service capacity to one of the two stations in an alternating order. This would lead to a polling system with two stations in tandem attended by a single server. For some tandem polling systems with different types of switching rules, such as gated and exhaustive service, Katayama [5] has given explicit expressions for the mean sojourn time of jobs in the system.

The rest of the paper is organized in the following way. In the next section, we describe in detail the model under consideration. In section 3, we derive a functional equation for the generating function of the stationary joint distribution of the number of jobs in both queues. This functional equation is analysed in section 4 for the two extreme cases in which the total capacity is allocated to one of the two stations, even if both stations are nonempty. For the intermediate cases, in which the stations really share the capacity when both stations are nonempty, the functional equation is studied in section 5. First, the kernel of the functional equation is analyzed and after that a boundary value problem is formulated and its solution is presented. In section 6, we briefly discuss a slightly more general model. As a special case of this more general model, we prove the well-known product form solution for the stationary distribution of the ordinary tandem queue with exponential interarrival and service times using the theory of boundary value problems. We conclude this paper with giving conclusions and mentioning some topics for further research in section 7.

2 Model description

We consider a tandem queueing model consisting of two stations. Jobs arrive at station 1 according to a Poisson process with rate λ, and they demand service from both stations before leaving the system. Each job requires an exponential amount of work with parameter ν_j at station j, $j = 1, 2$. The total service capacity *of the two service stations together* is fixed. Without loss of generality we assume that this total service capacity equals one unit of work per time unit. Whenever both stations are nonempty, a proportion p of the capacity is allocated to station 1, and the remaining part $(1-p)$ is allocated to station 2. Thus, when there is at least one job at each station, the departure rate of jobs at station 1 is $\nu_1 p$ and the departure rate of jobs at station 2 is $\nu_2(1-p)$. However, when one of the stations becomes empty, the total service capacity is allocated to the other station. Hence, the departure rate at that station, say station j, is temporarily increased to ν_j. In the sequel we will denote with $\rho_j = \lambda/\nu_j$ the average amount of work per time unit required at station j, $j = 1, 2$.

Clearly, the two-dimensional process $X(t) = (X_1(t), X_2(t))$, where $X_j(t)$, $j = 1, 2$, is the number of jobs at station j at time t, is a Markov process. The transition rate diagram of this process is given in Figure 1.

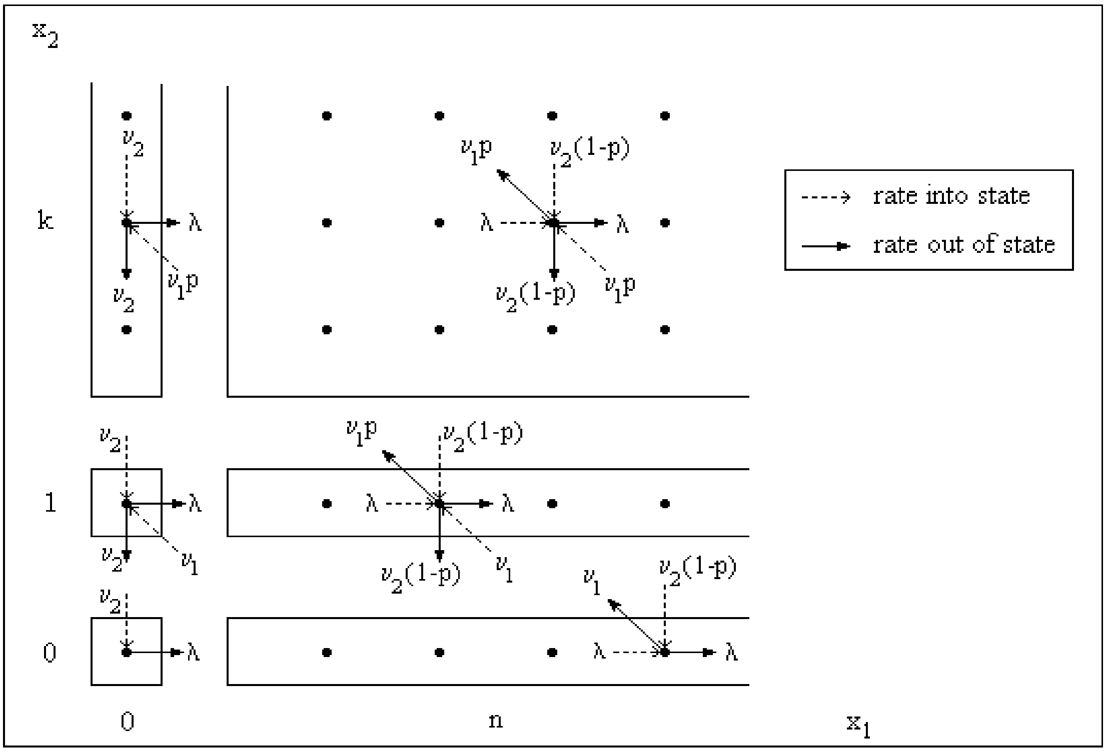

Figure 1: The transition rate diagram of the system

Under the ergodicity condition

$$\rho_1 + \rho_2 < 1, \tag{1}$$

the process $X(t)$ has a unique stationary distribution. In the sequel we are interested in determining this stationary distribution.

3 Functional equation

Let us denote with $\pi(n,k)$ the stationary probability of having n customers in station 1 and k customers in station 2. From the transition rate diagram of the model, we can derive the set of balance equations

$$\begin{aligned}
\lambda\pi(0,0) &= \nu_2\pi(0,1), \\
(\lambda+\nu_1)\,\pi(n,0) &= \lambda\pi(n-1,0)+(1-p)\,\nu_2\pi(n,1), \quad n\geq 1, \\
(\lambda+\nu_2)\,\pi(0,1) &= \nu_1\pi(1,0)+\nu_2\pi(0,2), \\
(\lambda+p\,\nu_1+(1-p)\,\nu_2)\,\pi(n,1) &= \lambda\pi(n-1,1)+\nu_1\pi(n+1,0)+(1-p)\,\nu_2\pi(n,2), \quad n\geq 1, \\
(\lambda+\nu_2)\,\pi(0,k) &= p\,\nu_1\pi(1,k-1)+\nu_2\pi(0,k+1), \quad k\geq 2, \\
(\lambda+p\,\nu_1+(1-p)\,\nu_2)\,\pi(n,k) &= \lambda\pi(n-1,k)+p\,\nu_1\pi(n+1,k-1)+(1-p)\,\nu_2\pi(n,k+1), \\
&\qquad n\geq 1, \quad k\geq 2.
\end{aligned}$$

Now we define, for $|x|\leq 1$, $|y|\leq 1$, the joint probability generating function

$$P(x,y) := \sum_{n\geq 0}\sum_{k\geq 0}\pi(n,k)x^n y^k.$$

From the balance equations it follows that $P(x,y)$ satisfies the following functional equation

$$\begin{aligned}
&\Big((\lambda+p\,\nu_1+(1-p)\,\nu_2)xy-\lambda x^2y-p\,\nu_1y^2-(1-p)\,\nu_2x\Big)\,P(x,y) \\
&\quad=\Big((1-p)\,[\nu_1\,y(y-x)+\nu_2\,x(y-1)]\Big)P(x,0) \\
&\quad+\Big(p\,[\nu_2\,x(1-y)+\nu_1\,y(x-y)]\Big)P(0,y) \\
&\quad+\Big(p\,\nu_2\,x(y-1)+(1-p)\,\nu_1\,y(x-y)\Big)P(0,0).
\end{aligned} \tag{2}$$

The constant $P(0,0)$ can be determined by substituting $x=(\nu_1y^2)/(\nu_1y-\nu_2(y-1))$ in (2). For this choice of x, both the factor in front of $P(x,0)$ and the factor in front of $P(0,y)$ are equal to zero, and hence equation (2) reduces to

$$P\Big(\frac{\nu_1y^2}{\nu_1y-\nu_2(y-1)},y\Big)=\frac{\nu_2(y-1)}{\nu_2(y-1)+\lambda y\big(1-\frac{\nu_1y^2}{\nu_1y-\nu_2(y-1)}\big)}P(0,0). \tag{3}$$

Now, letting $y\uparrow 1$ in (3), we obtain $P(0,0)=1-\rho_1-\rho_2$. This result can, of course, be explained by the fact that, independent of p, the two stations together always work at capacity 1 (if there is work in the system) and the fact that $\rho_1+\rho_2$ equals the amount of work brought into the system per time unit.

How can we find the solution $P(x,y)$ of the functional equation (2)? In the next section, we will give the explicit solution for $P(x,y)$ in the special cases $p=0$ and $p=1$. After that we show, in section 5, how for the case $0<p<1$ the solution of (2) can be obtained using the theory of boundary value problems.

4 The cases $p = 0$ and $p = 1$

In the case $p = 0$, resp. $p = 1$, the model that we consider can be alternatively viewed as a tandem queueing model with one single server for both stations together, in which the server gives preemptive priority to station 2, resp. station 1. It turns out that for these cases, the functional equation (2) can be solved relatively easily. This is mainly due to the fact that either the factor in front of $P(0, y)$ in equation (2), in case $p = 0$, or the factor in front of $P(x, 0)$, in case $p = 1$, is equal to zero. In fact, in case $p = 0$, the model we consider is well-known. However, as far as we know, the model is not studied before in case $p = 1$. Therefore, we will particularly pay attention to the latter case in this section.

4.1 The case $p = 0$

If $p = 0$, every time the server has completed a service of a job at station 1, he will immediately continue the service of this same job at station 2, due to the fact that service at station 2 has priority. Hence, the analysis of the model essentially reduces to the analysis of a single $M/C_2/1$ queue, in which the service time consists of two exponential phases with parameters ν_1 and ν_2 respectively. This model can, for example, also be analysed using the spectral expansion method (see [8]) or the matrix-geometric method (see [10]). Equation (2) reduces in this case to

$$\begin{aligned}\Big((\lambda + \nu_2)xy - \lambda x^2 y - \nu_2 x\Big)\, P(x, y) = \\ \Big(\nu_1\, y(y - x) + \nu_2\, x(y - 1)\Big)P(x, 0) + \Big(\nu_1\, y(x - y)\Big)P(0, 0).\end{aligned} \qquad (4)$$

Now, because for $y = \nu_2/(\lambda + \nu_2 - \lambda x)$ the factor in front of $P(x, y)$ in (4) is zero, also the righthandside of (4) should be equal to zero. Hence,

$$P(x, 0) = \frac{1 - \rho_1 - \rho_2}{1 - \frac{\rho_1 x(1+\rho_2-\rho_2 x)}{1-\rho_2 x}}. \qquad (5)$$

Substituting (5) in (4), we obtain after straightforward but lengthy calculations

$$P(x, y) = \frac{(1 - \rho_1 - \rho_2)(1 + \rho_2(y - x))}{1 - (\rho_1 + \rho_2 + \rho_1\rho_2)x + \rho_1\rho_2 x^2}. \qquad (6)$$

4.2 The case $p = 1$

If $p = 1$, the model is a tandem queue with a single server and preemptive priority for the first queue. Equation (2) reduces in this case to

$$\begin{aligned}\Big((\lambda + \nu_1)xy - \lambda x^2 y - \nu_1 y^2\Big)\, P(x, y) = \\ \Big(\nu_2\, x(1 - y) + \nu_1\, y(x - y)\Big)P(0, y) + \Big(\nu_2\, x(y - 1)\Big)P(0, 0).\end{aligned} \qquad (7)$$

Now, for $x = \xi(y)$, the unique root in the unit circle of the equation $\lambda x^2 - (\lambda + \nu_1)x + \nu_1 y = 0$, the righthandside of (7) should again be equal to zero. Hence, we obtain

$$P(0, 0) = \left(1 - \rho_2 y\frac{(1 - \xi(y))}{1 - y}\right) P(0, y), \qquad (8)$$

or, alternatively,

$$P(0,y) = \frac{1-\rho_1-\rho_2}{1-\rho_2 y \frac{(1-\xi(y))}{1-y}}. \tag{9}$$

Furthermore, substitution of (8) in (7) gives

$$P(x,y) = \frac{\rho_1 x(1-\xi(y)) + x - y}{(\rho_1+1)x - \rho_1 x^2 - y} P(0,y). \tag{10}$$

A nice probabilistic explanation for the results in equations (9) and (10) can be given. First remark that the root $\xi(y)$ can be interpreted as the generating function of the number of jobs served in a busy period of an $M/M/1$ queue with arrival rate λ and service rate ν_1 (see e.g., Cohen [2], page 190). Now, if we look at our model only during periods that the first queue is empty (i.e., we glue together idle periods of the first queue), the second queue behaves as an $M^X/M/1$ queue with arrival rate λ, batch size generating function $\xi(y)$ and service rate ν_2. Hence, the function $P(0,y)/P(0,1)$, i.e. the generating function of the conditional distribution of the number of jobs in the second queue given that the first queue is empty, is the same as the generating function of the number of jobs in the above mentioned $M^X/M/1$ queue. The latter one is well known (see e.g., [2], page 387) and immediately gives equation (9).

To explain equation (10) we introduce the random vector (Y_1, Y_2), denoting the stationary number of customers in the system (Y_1) and the stationary number of customers already served in the current busy period (Y_2), at a point in time in which the server is busy in an $M/M/1$ queue with arrival intensity λ and service intensity ν_1. Furthermore, let $Q(x,y)$ be the generating function of (Y_1, Y_2). The function $Q(x,y)$ can be straightforwardly obtained by studying the two-dimensional Markov process corresponding to (Y_1, Y_2). This process has almost the same transition rates as the ones in Figure 1 with $p=1$, only the rates near the vertical boundary differ. It turns out that $Q(x,y)$ is equal to

$$Q(x,y) = \frac{(1-\rho_1)x(x-\xi(y))}{(\rho_1+1)x - \rho_1 x^2 - y}. \tag{11}$$

Now, if we denote by (X_1, X_2) the stationary number of jobs at the two stations in our model at arbitrary points in time, and by $(0, X_2^{(i)})$ the same quantities during idle periods of the first station, then we have

$$(X_1, X_2) \stackrel{d}{=} \begin{cases} (0, X_2^{(i)}), & \text{with probability } 1-\rho_1, \\ (0, X_2^{(i)}) + (Y_1, Y_2), & \text{with probability } \rho_1. \end{cases} \tag{12}$$

Here, the random vectors $(0, X_2^{(i)})$ and (Y_1, Y_2) are furthermore independent. Hence, because the random vector $(0, X_2^{(i)})$ has generating function $P(0,y)/P(0,1)$, we have

$$P(x,y) = \frac{P(0,y)}{P(0,1)} \left(1-\rho_1+\rho_1 Q(x,y)\right). \tag{13}$$

Using $P(0,1) = 1-\rho_1$, and combination of (11) and (13), directly gives (10).

Remark: Decomposition result (12) also holds for generally distributed service times and hence may be starting point for the analysis of the model with arbitrary service times.

5 The case $0 < p < 1$

In this section we will derive the solution of functional equation (2) for $0 < p < 1$. A key role is played by the kernel

$$K(x,y) := (\lambda + p\,\nu_1 + (1-p)\,\nu_2)xy - \lambda x^2 y - p\,\nu_1 y^2 - (1-p)\,\nu_2 x.$$

5.1 Zeros of the kernel

Because the kernel $K(x,y)$ is, for each x, a polynomial of degree 2 in y, we have that for every value of x there are two possible values of y, say $y_1(x)$ and $y_2(x)$, such that $K(x,y_1(x)) = K(x,y_2(x)) = 0$.

Lemma 1 *The algebraic function $y(x)$ defined by $K(x,y(x)) = 0$ has four real branch points $0 = x_1 < x_2 \le 1 < x_3 < x_4$.*

Proof: Branch points are zeros of the discriminant, $D(x)$, of the equation $K(x,y) = 0$ as function of y, i.e.,

$$D(x) = \Big(\lambda x^2 - (\lambda + p\nu_1 + (1-p)\nu_2)x\Big)^2 - 4p(1-p)\nu_1\nu_2 x.$$

Clearly, $D(0) = 0$, $D(x) < 0$ for small positive x, $D(1) \ge 0$, $D((\lambda + p\nu_1 + (1-p)\nu_2)/\lambda) < 0$ and $\lim_{x\to\infty} D(x) = \infty$. Hence, the lemma follows. □

Lemma 2 *For each $x \in [x_1, x_2]$, the two roots $y_1(x)$ and $y_2(x)$ are complex conjugate. Hence, the interval $[x_1, x_2]$ is mapped by $x \mapsto y(x)$ onto a closed contour L, which is symmetric with respect to the real line.*

Proof: Follows directly from the fact that the discriminant $D(x)$ is zero for $x = x_1$ and $x = x_2$ and negative for $x \in (x_1, x_2)$. □

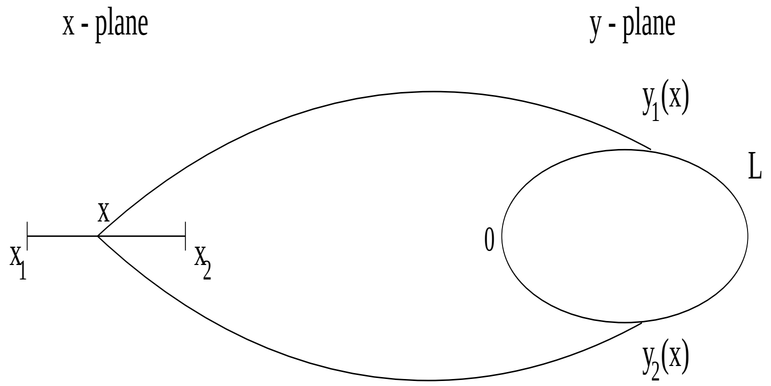

Figure 2: The contour L

In Figure 2, the result of Lemma 2 is illustrated. In the sequel we will denote the interior of the contour L by L^+. Finally, notice that for a point $y(x)$ on the contour L we have that

$$y(x)\overline{y}(x) = \frac{(1-p)\nu_2 x}{p\nu_1}, \tag{14}$$

where the notation $\overline{y}$ indicates the complex conjugate of y.

5.2 The boundary value problem

Next, we will formulate a boundary value problem for the function $P(0,y)$.

Lemma 3 *The function $P(0,y)$ is regular in the domain L^+ and satisfies for $y \in L$ the condition*

$$\operatorname{Im} P(0,y) = \operatorname{Im}\left(\frac{p^2\nu_2\overline{y}(y-1)+(1-p)y\,[p\nu_1\overline{y}-(1-p)\nu_2]}{p\,[y((1-p)\nu_2-p\nu_1\overline{y})+p\nu_2\overline{y}(y-1)]}P(0,0)\right). \tag{15}$$

Proof: For zeropairs (x,y) of the kernel for which $P(x,y)$ is finite, we have

$$\begin{aligned}&\Big((1-p)\,[\nu_1\,y(y-x)+\nu_2\,x(y-1)]\Big)P(x,0)\\&+\Big(p\,[\nu_2\,x(1-y)+\nu_1\,y(x-y)]\Big)P(0,y)\\&+\Big(p\,\nu_2\,x(y-1)+(1-p)\,\nu_1\,y(x-y)\Big)P(0,0)=0.\end{aligned} \tag{16}$$

We can rewrite this equation, by substituting $(1-p)\nu_2 x = p\nu_1 y\overline{y}$ (see (14)), in

$$P(0,y) = \frac{p^2\nu_2\overline{y}(y-1)+(1-p)y\,[p\nu_1\overline{y}-(1-p)\nu_2]}{p\,[y((1-p)\nu_2-p\nu_1\overline{y})+p\nu_2\overline{y}(y-1)]}P(0,0)+\frac{1-p}{p}P(x,0). \tag{17}$$

Now, if $\frac{(1-p)\nu_2}{p\nu_1}x_2 \le 1$, then L lies entirely within the unit circle ($y(x_2)$ is the point on L with largest absolute value). Hence, $P(0,y)$ is regular in L^+. Finally, (15) follows from (17) by taking $x \in [x_1,x_2]$ and using that $P(x,0)$ is real for those x.

If $\frac{(1-p)\nu_2}{p\nu_1}x_2 > 1$, then $P(0,y(x))$ can be continued analytically over the interval $[x_1,x_2]$ via equation (16), because $P(x,0)$ is regular on this interval. Hence, the analytic continuation of $P(0,y)$ is finite at $y=y(x_2)$. Because $P(0,y)$ has a power series expansion at $y=0$ with positive coefficients, this implies that $P(0,y)$ is regular for $|y|<y(x_2)$ and hence in L^+. □

Lemma 3 shows that the determination of $P(0,y)$ reduces to the determination of the solution of the following Riemann-Hilbert boundary value problem on the contour L:

Determine a function $P(0,y)$ such that

1. $P(0,y)$ is regular for $y\in L^+$ and continuous for $y \in L^+\cup L$.
2. Re $[iP(0,y)] = c(y)$, for $y\in L$,

where

$$c(y) = -\operatorname{Im}\left(\frac{p^2\nu_2\overline{y}(y-1)+(1-p)y\,[p\nu_1\overline{y}-(1-p)\nu_2]}{p\,[y((1-p)\nu_2-p\nu_1\overline{y})+p\nu_2\overline{y}(y-1)]}P(0,0)\right).$$

The standard way to solve this type of boundary value problem (see, e.g., Muskhelishvili [9]) is to transform the boundary condition (15), by using conformal mappings, to a condition on the unit circle. Let $z=f(y)$ be the conformal map of L^+ onto the unit circle $C^+=\{z : |z|<1\}$ and denote by $y=f_0(z)$ the inverse mapping, i.e., the conformal map of C^+ onto L^+.

Now, if the function $H(z)$ is the solution of the problem (P):

Determine a function $H(z)$ such that

1. $H(z)$ is regular for $z \in C^+$ and continuous for $z \in C^+ \cup C$.
2. Re $[iH(z)] = \tilde{c}(z)$, for $z \in C$, where $\tilde{c}(z) = c(f_0(z))$,

then $P(0, y) = H(f(y))$ is the solution of the original problem. The solution of problem (P), a so-called Dirichlet problem on the circle, is well-known (see [9]) and given by

$$H(z) = \frac{1}{2\pi} \int_C \tilde{c}(w) \frac{w+z}{w-z} \frac{dw}{w} + K,$$

where K is some constant.

In this way, $P(0, y)$ has been formally determined. Substitution, first in (16) to obtain $P(x, 0)$ and after that in (2), then yields $P(x, y)$, so that the generating function of the joint stationary distribution of the queue lengths in the tandem queue has been obtained. In a future study the details of this analysis will be provided. For example, the determination of the conformal map generally poses an interesting problem in the analysis of these boundary value problems.

Remark: In most problems, for the determination of the conformal map f and the inverse conformal map f_0, a numerical technique (e.g., Theodorsen's procedure, see [3]) has to be applied. However, for this specific problem, an explicit expression for the conformal mapping $f(y)$ can be found (see the paper of Blanc [1], in which the time-dependent behaviour of the ordinary tandem queue without coupled processors is studied).

6 Generalization

The model that we considered so far in the paper is a special case of the following model. The system has two stations in tandem, each station having its own server. Customers arrive at station 1 according to a Poisson process with rate λ, and they require an exponentially distributed service time from both stations before leaving the system. The service rate at station j is equal to rate μ_j whenever both stations have at least one customer. If one of the stations becomes empty, the service rate at the other station changes from μ_j to μ_j^*.

For this model the functional equation becomes

$$\begin{aligned}\Big((\lambda + \mu_1 + \mu_2)xy - \lambda x^2 y - \mu_1 y^2 - \mu_2 x\Big) P(x, y) \\ = \Big((\mu_1^* - \mu_1)y(y-x) + \mu_2 x(y-1)\Big) P(x, 0) \\ + \Big((\mu_2^* - \mu_2)x(1-y) + \mu_1 y(x-y)\Big) P(0, y) \\ + \Big((\mu_2^* - \mu_2)x(y-1) + (\mu_1^* - \mu_1)y(x-y)\Big) P(0, 0). \qquad (18)\end{aligned}$$

After calculations, similar to those done in the previous sections, we get for y on the contour L,

$$\begin{aligned}&\text{Im}\left(\frac{(\mu_2^* - \mu_2)\mu_1\overline{y}(1-y) + \mu_1 y\,(\mu_1\overline{y} - \mu_2)}{(\mu_1^* - \mu_1)y\,(\mu_1\overline{y} - \mu_2) + \mu_1\mu_2\overline{y}(1-y)} P(0, y)\right) \\ &= \text{Im}\left(\frac{(\mu_2^* - \mu_2)\mu_1\overline{y}(1-y) - (\mu_1^* - \mu_1)y\,(\mu_1\overline{y} - \mu_2)}{(\mu_1^* - \mu_1)y\,(\mu_1\overline{y} - \mu_2) + \mu_1\mu_2\overline{y}(1-y)} P(0, 0)\right).\end{aligned}$$

Again, we can now formulate a Riemann-Hilbert boundary value problem for the function $P(0, y)$ on the contour L of the following form:

Determine a function $P(0, y)$ such that

1. $P(0, y)$ is regular for $y \in L^+$ and continuous for $y \in L^+ \cup L$.
2. Re $[g(y)P(0, y)] = c(y)$, for $y \in L$.

The study of this more general Riemann-Hilbert boundary value problem will be a topic for further research. In the remaining part we restrict our attention to the solution of the problem for the special case of an ordinary tandem queue, i.e., $\mu_j^* = \mu_j$, $j = 1, 2$.

6.1 The ordinary tandem queue

In the case $\mu_j^* = \mu_j$, $j = 1, 2$, it is of course well-known that the stationary joint distribution of the number of jobs at the two stations has a product form. We now show how this result follows from (18). For zeropairs (x, y) of the kernel for which $P(x, y)$ is finite, we have, from (18),

$$\mu_1 y(x - y)P(0, y) = \mu_2 x(1 - y)P(x, 0). \qquad (19)$$

Multiplying both sides by $(\lambda(1 - \overline{y}))/(\mu_1\mu_2)$, we obtain

$$\rho_2 y(x - y)(1 - \overline{y})P(0, y) = \rho_1 x(1 - y)(1 - \overline{y})P(x, 0), \qquad (20)$$

where now $\rho_j = \lambda/\mu_j$. Clearly, for real x, the righthandside of (20) is real, and furthermore for y on the contour L, we have $\rho_2 y\overline{y} = \rho_1 x$. Using these two facts, we conclude that

$$\text{Im } ((x - y)(\rho_2 y - \rho_1 x)P(0, y)) = 0.$$

Finally, using again that (x, y) is a zeropair of the kernel, this reduces to

$$\text{Im } ((1 - \rho_2 y)P(0, y)) = 0.$$

The solution of this boundary value problem is given by

$$P(0, y) = \frac{K}{1 - \rho_2 y},$$

where K is a constant. Substituting this in (19) gives (again using $\rho_2 y\overline{y} = \rho_1 x$)

$$P(x, 0) = \frac{K}{1 - \rho_1 x}.$$

Finally, substituting the formulas for $P(0, y)$ and $P(x, 0)$ in (18) gives

$$P(x, y) = \frac{K}{(1 - \rho_1 x)(1 - \rho_2 y)}.$$

In this way, we find the product form solution for the tandem queueing system directly from the boundary value problem.

7 Conclusions and topics for further research

In this paper we analysed a tandem queueing model consisting of two stations in which the total service capacity of the two stations together is constant. The service capacity of the individual stations depends on whether or not one of the stations is empty. The stationary joint distribution of the number of jobs in the two stations is analysed, using the theory of boundary value problems.

The numerical evaluation of the solution is a topic for further research. Furthermore, the analysis of the more general tandem queueing model with coupled processors, briefly described in section 6, will also be part of a future study.

Acknowledgement The authors like to thank Onno Boxma (Eindhoven University of Technology) and Dee Denteneer (Philips Research, Eindhoven) for stimulating discussions.

References

[1] J.P.C. Blanc. The relaxation time of two queueing systems in series. *Stochastic Models*, 1:1-16, 1985.

[2] J.W. Cohen. *The Single Server Queue.* North-Holland, Amsterdam, 1982.

[3] J.W. Cohen and O.J. Boxma. *Boundary Value Problems in Queueing System Analysis.* North-Holland, Amsterdam, 1983.

[4] G. Fayolle and R. Iasnogorodski. Two coupled processors: The reduction to a Riemann-Hilbert problem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 47:325–351, 1979.

[5] T. Katayama. A cyclic-service tandem queueing model with semi-exhaustive service. In T. Hasegawa, H. Takagi, and Y. Takahashi, editors, *Performance of Distributed and Parallel Systems*, Elsevier Science Publishers B.V., Amsterdam, 1989, pages 183-196.

[6] G.P. Klimov. Time-sharing service systems I. *Theory of Probability and its Applications*, 12:532-551, 1974.

[7] A.G. Konheim, I. Meilijson and A. Melkman. Processor-sharing of two parallel lines. *Journal of Applied Probability*, 18:952–956, 1981.

[8] I. Mitrani and D. Mitra. A spectral expansion method for random walks on semi-infinite strips. In R. Beauwens and P. de Groen, editors, *Iterative Methods in Linear Algebra.* North-Holland, Amsterdam, 1992, pages 141–149.

[9] N.I. Muskhelishvili. *Singular Integral Equations.* Dover Publications, New York, 1992.

[10] M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models.* Johns Hopkins, Baltimore, 1981.

[11] D. Sala, J.O. Limb and S.U. Khaunte. Adaptive control mechanism for cable modem MAC protocols. In *Proceedings of IEEE Infocom 1998*, San Francisco, CA, USA, pages 1392-1399.