Probability in the Engineering and Informational Sciences, **15**, 2001, 311–325. Printed in the U.S.A.

TWO QUEUES IN TANDEM WITH RETRIAL CUSTOMERS

EVANGELOS MOUTZOUKIS AND CHRISTOS LANGARIS

Department of Mathematics
University of Ioannina
Ioannina 45 110, Greece
E-mail: clagar@cc.uoi.gr

We consider two single-server queues in tandem. The customers receive service at the first station and then they proceed to the second. There are no intermediate waiting positions and, so, the blocking phenomenon is observed. The arriving customers who find the first station unavailable (busy or blocked) behave like retrial customers. We consider the case of "constant retrial policy." For such a system, we obtain the system state probabilities in steady state in terms of generating functions. We present also some interesting measures of the system performance and directly deduce results for the corresponding ordinary (without retrials) model. We consider, finally, the special case of exponential service times as an example.

1. INTRODUCTION

Queuing systems connected in tandem is a special class of queuing networks which has been extensively studied in the literature. When, in such a system, limitations are imposed on the lengths of the intermediate waiting rooms, the blocking phenomenon is observed. Thus, a customer whose service is completed at station i cannot enter station $i + 1$ if the latter is fully occupied. We then say that station i is blocked: The customer remains at station i, which cannot provide service to other customers. Queuing networks with blocking have been useful in modeling situations that arise in computer systems, telecommunication systems, and production lines.

The first work on two queues in tandem with a finite intermediate buffer is due to Hunt [8], who considered the case of exponential service times. Later, Avi-Itzhak and Yadin [2] investigated the same model assuming no intermediate waiting space and arbitrarily distributed service times. The authors derived steady-state results for

© 2001 Cambridge University Press 0269-9648/01 $12.50

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

the probability generating functions of the number of customers in queue for station 1, in front of station 2, and in the whole system. The same results have been derived by Suzuki [18], who, in addition, investigated formally the ergodicity of the system. Finally, Prabhu [16] considered the transient behavior of the same model. Variants and extensions of this model have been studied by several authors (e.g., Neuts [11]). For a complete bibliography on the subject, the reader is referred to Perros [13].

Retrial queuing systems are characterized by the fact that an arriving customer who finds the server busy leaves the system (joins the so-called retrial group) and repeats his demand after a random amount of time. Practical use of retrial queuing systems arises in telephone-switching systems and in telecommunication and computer networks. For more details, we refer to the surveys of Falin [5], Yang and Templeton [20], and Kulkarni and Liang [9]. Recently, Falin and Templeton [6] presented a comprehensive review of the retrial queues.

The majority of previous works in retrial queues assume that all customers in the retrial group try, independently of each other, to enter service after an exponentially distributed (with parameter μ) amount of time. However, as Artalejo [1] has pointed out, there exist situations in which the server is required to search for customers. Then, the retrial policy, called the "constant retrial policy," is as follows. If the retrial group is not empty, then the time until the next attempt is exponentially distributed with parameter μ. If the server is free at the time of this attempt, then the customer at the head of the group is chosen to commence service. A different interpretation of "constant retrial policy" is given by Farahmand [7]. According to this, the parameter of the exponential retrial time of each customer in the retrial group is μ/n, where n is the size of the retrial group. Thus, the total retrial intensity is μ. Applications of the constant retrial policy are found in computer networks and especially in the area of local area networks (LANs). Choi et al. [3] analyzed a retrial queue with "constant retrial policy" arising in computer networks operating under the CSMA|CD communication protocol. Other similar applications may be found in Martin and Artalejo [10] and in Artalejo [1]. The $M/G/1$ constant retrial queue has been analyzed by Farahmand [7]; Martin and Artalejo [10] investigated the same model but with two types of customer (queued up and retrial customers). Artalejo [1] studied the $M/M/2$ constant retrial queue, and recently Ramalhoto and Gomez-Corral [17] gave some decomposition formulas for the $M/M/r/r + d$ queue with constant retrial policy.

All of the retrial models mentioned are models consisting of a single station. In this work, we consider a model of two stations in tandem. The customers receive service at the first station and then they proceed to the second. There is no intermediate waiting room and so a customer whose service is completed in the first station cannot proceed to the second if the latter is busy. Instead, the customer remains in the first station and the station is blocked. The important feature and the main originality of our model is that the arriving customer who finds the first station unavailable (busy or blocked) does not join a queue, but he is placed, instead, in a retrial group and retries for service under the constant retrial policy.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

General queuing networks with retrial customers have been investigated by Pourbabai [14, 15] and Takahara [19]. The authors pointed out that the exact analysis of this kind of model seems rather impossible and obtained approximation results using approximation methods. We consider our analysis of the two-stage retrial model here as a first attempt to isolate a part of a general queuing network and to obtain some closed-form results on it.

A simple example of a system modeled as a two-stage queue with blocking and retrials is the telephone-switching system used for the intercommunication between, for example, a director and his secretary. The clients contact the secretary first, and if the line is busy, then they repeat their call later. If the line is idle, then the secretary asks the client's personal data and, in the sequel, either she connects him with the director, if the line is available, or she delays the client until the director becomes available.

The article is organized as follows. After the full description of the model in Section 2, the Markov chain that is imbedded in our system is investigated and the sufficient and necessary condition for statistical equilibrium is obtained in Section 3. These results are used, in Section 4, to investigate the general stochastic process governing the system states and to obtain the generating functions of the steady-state probabilities. Some measures of the system performance are obtained in Section 5, whereas in Section 6, the obtained formulas for the retrial model are used to deduce results directly for the ordinary (without retrials) model of two queues in tandem without intermediate space. Finally, in Section 7 we consider, as an example, the special case of exponential service times and give a table of numerical calculations.

2. THE MODEL

We consider two single-server stations in tandem. Customers arrive at the first station, one at a time, according to the Poisson distribution with parameter λ. Each customer receives service at station 1 and then proceeds to station 2 for an additional service. After the second service, the customer departs from the system. The service times at stations 1 and 2 are independent and arbitrarily distributed random variables with probability density functions (p.d.f.'s) $b_i(x)$, distribution functions $B_i(x)$, and finite mean values $\bar{u}_i$, for $i = 1$ and 2, respectively.

We assume further that there are no waiting positions between the two stations. Consequently, the blocking phenomenon is observed: A customer whose service is completed at station 1 cannot proceed into station 2 if the latter is busy. Instead, the customer remains at station 1, which is said to be blocked, until station 2 becomes empty. During blocking time, no customers may enter station 1.

The arriving customers who find the first station unavailable (busy or blocked) behave like retrial customers: They leave the system and join a (hypothetical) retrial queue of infinite capacity. The so-called constant retrial policy is used; that is, each customer in the retrial group retries for service according to the exponential distribution parameter μ/n, where n is the number of customers in the retrial box at the epoch of retrial. If the server of station 1 is free at the time of an attempt, then the customer receives service immediately. Otherwise, he repeats his demand later.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

3. THE IMBEDDED MARKOV CHAIN

Let $X(t)$ represent the number of customers in the retrial box at time t and let t_i be the instant of the ith departure from station 1. We assume, without loss of generality, that $t_0 = 0$ is the time of a departure from station 1. If we denote $X_n = X(t_n + 0)$, then it is clear that the stochastic process $\{X_n : n = 0,1,2,\ldots\}$ is an irreducible and aperiodic Markov chain with transition probability matrix $\mathcal{P} = \{p_{ij}\}$, where

$$
\begin{aligned}
p_{0n} &= \int_0^\infty \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} f_0(t)\, dt, && n = 0,1,2,\ldots,\\
p_{kn} &= \int_0^\infty \frac{(\lambda t)^{n-k}}{(n-k)!}\, e^{-\lambda t} f_1(t)\, dt \\
&\quad + \int_0^\infty \frac{(\lambda t)^{n-k+1}}{(n-k+1)!}\, e^{-\lambda t} f_2(t)\, dt, && n = 1,2,\ldots, \quad k = 1,2,\ldots,n \qquad \textbf{(1)}\\
p_{n+1n} &= \int_0^\infty e^{-\lambda t} f_2(t)\, dt, && n = 0,1,2,\ldots,
\end{aligned}
$$

and the functions $f_i(t)$ are defined by

$$
\begin{aligned}
f_0(t) &= \int_0^\infty \lambda e^{-\lambda w} \frac{d}{dt}\left(B_1(t)B_2(t+w)\right) dw,\\
f_1(t) &= \int_0^\infty \lambda e^{-(\lambda+\mu)w} \frac{d}{dt}\left(B_1(t)B_2(t+w)\right) dw,\\
f_2(t) &= \int_0^\infty \mu e^{-(\lambda+\mu)w} \frac{d}{dt}\left(B_1(t)B_2(t+w)\right) dw.
\end{aligned}
$$

Now define the function

$$
\psi_u(s) = \int_0^\infty u e^{-uw}\, dw \int_0^\infty e^{-sx}\, d_x\left(B_1(x)B_2(x+w)\right) \qquad \textbf{(2)}
$$

and denote

$$
v_u = -\left.\frac{d\psi_u(s)}{ds}\right|_{s=0}, \qquad \rho^* = \frac{\lambda}{\lambda+\mu} + \lambda v_{\lambda+\mu}. \qquad \textbf{(3)}
$$

Consider also the probabilities

$$
\pi_k = \lim_{n\to\infty} P[X_n = k], \qquad k = 0,1,2,\ldots,
$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

and the generating function

$$\Pi(z) = \sum_{n=0}^{\infty} \pi_n z^n, \qquad |z| < 1.$$

Then, the following theorem holds.

THEOREM 1: *The Markov chain X_n is positive recurrent if and only if $\rho^* < 1$. In this case, the generating function $\Pi(z)$ is given by*

$$\Pi(z) = \frac{z\psi_\lambda(\lambda - \lambda z) - \dfrac{\lambda z + \mu}{\lambda + \mu}\,\psi_{\lambda+\mu}(\lambda - \lambda z)}{z - \dfrac{\lambda z + \mu}{\lambda + \mu}\,\psi_{\lambda+\mu}(\lambda - \lambda z)}\,\pi_0, \tag{4}$$

where

$$\pi_0 = \frac{1 - \rho^*}{1 - \rho^* + \lambda v_\lambda}. \tag{5}$$

PROOF: *Sufficiency.* Suppose that $\rho^* < 1$. We will show the sufficiency by using the criterion of Pakes [12]:

> An irreducible and aperiodic Markov chain $(Y_n; n \geq 0)$, with state space the non-negative integers, is positive recurrent if $|\delta_k| < \infty$ for all $k = 0,1,2,\ldots$ and $\limsup_{k\to\infty} \delta_k < 0$, where
>
> $$\delta_k = E[Y_{n+1} - Y_n | Y_n = k].$$

In our model here, one can see after some manipulations that

$$\delta_0 = \lambda v_\lambda,$$

$$\delta_k = \rho^* - 1, \qquad k = 1,2,3,\ldots,$$

and so it is clear that for $\rho^* < 1$, the criterion holds.

Consider now the usual matrix equation

$$\underline{\pi} = \underline{\pi}\mathcal{P}, \tag{6}$$

where $\underline{\pi} = (\pi_0, \pi_1, \ldots)$. Then, using Eq. (1) in Eq. (6),

$$\Pi(z) = \pi_0\psi_\lambda(\lambda - \lambda z) + \frac{\mu}{\lambda + \mu}\,\frac{\Pi(z) - \pi_0}{z}\,\psi_{\lambda+\mu}(\lambda - \lambda z) + \frac{\lambda}{\lambda + \mu}\,(\Pi(z) - \pi_0)\psi_{\lambda+\mu}(\lambda - \lambda z),$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

and so we obtain relation (4). Moreover,

$$\lim_{z\to 1}\Pi(z)=\frac{1-\rho^*+\lambda v_\lambda}{1-\rho^*}\,\pi_0 \tag{7}$$

and so π_0 must be given from Eq. (5).

Necessity. Suppose now that X_n is positive recurrent and $\rho^*>1$. Consider the function

$$m(z)=\frac{\lambda z+\mu}{\lambda+\mu}\,\psi_{\lambda+\mu}(\lambda-\lambda z).$$

By defining, for ϵ sufficiently small, the curve $C=\{z:|z|=1-\epsilon\}$ and applying Rouche's theorem, one can show that the equation

$$z-m(z)=0$$

has one and only one root inside C if $\rho^*=-[dm(1-\epsilon)/d\epsilon]|_{\epsilon=0}>1$ and no roots inside C if $\rho^*\le 1$. Thus, for $\rho^*>1$ the denominator of $\Pi(z)$ in Eq. (4) has a zero inside the unit disk and it is easy to verify that this zero cannot be a zero of the numerator as well. This contradicts, of course, the hypothesis that X_n is positive recurrent.

For $\rho^*=1$ finally, it is clear from Eq. (7) that X_n cannot be positive recurrent and this completes the proof. ■

Remark 1: Let T_0 and T_1 represent the time intervals between two consecutive departures from station 1, given that at the first departure, the retrial box is empty and not empty, respectively. Also, let $r_0(t)$ and $r_1(t)$ represent the corresponding probability density functions. Then, it is easy to see that

$$r_0(t)=\int_0^t \lambda e^{-\lambda w}\,\frac{d}{dt}\left[B_1(t-w)B_2(t)\right]dw,$$

$$r_1(t)=\int_0^t (\lambda+\mu)e^{-(\lambda+\mu)w}\,\frac{d}{dt}\left[B_1(t-w)B_2(t)\right]dw,$$

and so using Eq. (2),

$$E(T_0)=\frac{1}{\lambda}+v_\lambda,\qquad E(T_1)=\frac{1}{\lambda+\mu}+v_{\lambda+\mu}. \tag{8}$$

Now it is clear from Eq. (3) that

$$\rho^*=\lambda E(T_1),$$

and so ρ^* can be considered as the traffic intensity of the system.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

4. THE GENERAL PROCESS

For a stochastic process $(A(t);\ t \geq 0)$ we will say that it is stable if its limiting probabilities as $t \to \infty$ exist, are positive, and sum to unity.

Now, let $X(t)$ represent the number of customers in the retrial box at time t and for $i = 1,2$,

$$\xi^i(t) = \begin{cases} 0 & \text{if the } i\text{th server is idle at } t \\ 1 & \text{if the } i\text{th server is working at } t \\ 2 & \text{if the } i\text{th server is blocked at } t. \end{cases}$$

Consider also the process

$$\mathbf{Y}(t) = \{X(t), \xi^1(t), \xi^2(t)\},$$

and define for $i = 0,1,2$, $j = 0,1$, and $k = 0,1,2,\ldots$,

$$p_{ij}(k) = \lim_{t\to\infty} \Pr[X(t) = k,\ \xi^1(t) = i,\ \xi^2(t) = j],$$

$$P_{ij}(z) = \sum_{k=0}^{\infty} p_{ij}(k) z^k, \qquad |z| < 1,$$

$$\phi_u(s) = \frac{u}{u-s}\left(\frac{b_2^*(s)}{s} - \frac{b_2^*(u)}{u}\right) - \frac{\psi_u(s)}{s},$$

$$\bar{\phi}_u(s) = \frac{b_1^*(s)}{s} - \frac{\psi_n(s)}{s},$$

where $b_i^*(\cdot)$ denotes the Laplace transform of $b_i(\cdot)$ for $i = 1,2$. Then, the following holds.

THEOREM 2: *For $\rho^* < 1$, the process $\mathbf{Y}(t)$ is stable. In this case, the generating functions $P_{ij}(z)$ are given for any $|z| \leq 1$ by*

$$P_{00}(z) = \frac{\lambda b_2^*(\lambda+\mu)}{\lambda+\mu}[\Pi(z) - \pi_0] + b_2^*(\lambda)\pi_0, \tag{9}$$

$$P_{01}(z) = \lambda\frac{1 - b_2^*(\lambda+\mu)}{\lambda+\mu}[\Pi(z) - \pi_0] + (1 - b_2^*(\lambda))\pi_0, \tag{10}$$

$$P_{10}(z) = \lambda\frac{\lambda z + \mu}{z(\lambda+\mu)}\phi_{\lambda+\mu}(\lambda - \lambda z)[\Pi(z) - \pi_0] + \lambda\phi_\lambda(\lambda - \lambda z)\pi_0, \tag{11}$$

$$P_{11}(z) = \lambda[(\lambda z + \mu)\Pi(z) - \mu(1-z)\pi_0]\frac{1 - b_1^*(\lambda - \lambda z)}{z(\lambda+\mu)(\lambda - \lambda z)} - P_{10}(z), \tag{12}$$

$$P_{21}(z) = \lambda\frac{\lambda z + \mu}{z(\lambda+\mu)}\bar{\phi}_{\lambda+\mu}(\lambda - \lambda z)[\Pi(z) - \pi_0] + \lambda\bar{\phi}_\lambda(\lambda - \lambda z)\pi_0, \tag{13}$$

where $\Pi(z)$ and π_0 are known from Eqs. (4) and (5), respectively.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

PROOF: Denote by T_n $(n = 0,1,2,\ldots)$ the epoch of the nth departure from station $1(T_0 \equiv 0)$. Then, it is clear that the process $(X,T) = \{X_n, T_n : n \geq 0\}$ is a Markov Renewal Process (MRP) and the process $\mathbf{Y}(t)$ is a semiregenerative process with imbedded MRP (X,T). Suppose now that $\rho^* < 1$. Then, the process (X,T) is an irreducible and aperiodic positive recurrent process and $\underline{\pi} = (\pi_0, \pi_1, \ldots)$ is an invariant measure (distribution) for X. Thus, if

$$\underline{\pi} \cdot \underline{m} \equiv \sum_{k=0}^{\infty} \pi_k E[T_1 | X_0 = k] < \infty,$$

then (see Cinlar [4, p. 347]) the limiting probabilities $p_{ij}(k)$ exist and are given by

$$p_{ij}(k) = \frac{1}{\underline{\pi} \cdot \underline{m}} \sum_{m=0}^{\infty} \pi_m \int_0^{\infty} K_m(k,i,j,t)\, dt, \tag{14}$$

where

$$K_m(k,i,j,t) = \Pr[X(t) = k,\ \xi^1(t) = i,\ \xi^2(t) = j,\ T_1 > t | X_0 = m].$$

Using the definition of the function $\psi_u(s)$, we obtain

$$E[T_1 | X_0 = k] = \begin{cases} \dfrac{1}{\lambda} + v_{\lambda}, & k = 0 \\ \dfrac{1}{\lambda + \mu} + v_{\lambda+\mu}, & k > 0, \end{cases}$$

and, finally, for $\rho^* < 1$,

$$\underline{\pi} \cdot \underline{m} = \frac{1}{\lambda}. \tag{15}$$

To finish the proof, we need the functions $K_m(k,i,j,t)$. For $i = 0$, it is easy to see that

$$K_m(k,0,j,t) = \begin{cases} (\delta_{\{m=k=0\}} e^{-\lambda t} + \delta_{\{m=k>0\}} e^{-(\lambda+\mu)t}) B_2(t), & j = 0 \\ (\delta_{\{m=k=0\}} e^{-\lambda t} + \delta_{\{m=k>0\}} e^{-(\lambda+\mu)t})(1 - B_2(t)), & j = 1, \end{cases} \tag{16}$$

where $\delta_{\{A\}} = 1$ if A holds and zero if not.

Now, define the function

$$a_{ij}(t,s) = \begin{cases} B_2(t)(1 - B_1(t-s)), & i = 1, j = 0 \\ (1 - B_2(t))(1 - B_1(t-s)), & i = 1, j = 1 \\ (1 - B_2(t)) B_1(t-s), & i = 2, j = 1. \end{cases}$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

Then, for $i \neq 0$,

$$K_m(k,i,j,t) = \delta_{\{m=0\}} \int_0^t \lambda e^{-\lambda s} a_{ij}(t,s) \frac{[\lambda(t-s)]^k}{k!} e^{-\lambda(t-s)}$$
$$+ \delta_{\{0<m\leq k\}} \int_0^t \lambda e^{-(\lambda+\mu)s} a_{ij}(t,s) \frac{[\lambda(t-s)]^{k-m}}{(k-m)!} e^{-\lambda(t-s)}$$
$$+ \delta_{\{0<m\leq k+1\}} \int_0^t \mu e^{-(\lambda+\mu)s} a_{ij}(t,s) \frac{[\lambda(t-s)]^{k-m+1}}{(k-m+1)!} e^{-\lambda(t-s)}. \quad \textbf{(17)}$$

Finally, using relations (15)–(17) in Eq. (14), we arrive at Eqs. (9)–(13) and the theorem has been proved. ■

5. MEASURES OF THE SYSTEM PERFORMANCE

Here, we will try to use the formulas derived in the previous section to obtain some important measures of the system performance:

1. Summing relations (11) and (12) and putting $z = 1$, we arrive at

$$\Pr[\text{the server in station 1 is busy}] = \lambda \bar{u}_1. \quad \textbf{(18)}$$

2. Summing relations (10), (12), and (13) and putting $z = 1$,

$$\Pr[\text{the server in station 2 is busy}] = \lambda \bar{u}_2.$$

3. Summing relations (9) and (10) and putting $z = 1$, we get

$$\Pr[\text{the server in station 1 is idle}]$$
$$= \frac{\lambda}{\lambda+\mu}(1-\pi_0) + \pi_0$$
$$= \frac{\lambda}{\lambda+\mu} + \frac{\dfrac{\mu}{\lambda+\mu} - \lambda v_{\lambda+\mu}}{1 + \dfrac{\lambda(\lambda+\mu)}{\mu}(v_\lambda - v_{\lambda+\mu})}, \quad \textbf{(19)}$$

and so from Eqs. (18) and (19),

$$\Pr[\text{the server in station 1 is blocked}]$$
$$= \frac{\mu}{\lambda+\mu} - \lambda \bar{u}_1 - \frac{\dfrac{\mu}{\lambda+\mu} - \lambda v_{\lambda+\mu}}{1 + \dfrac{\lambda(\lambda+\mu)}{\mu}(v_\lambda - v_{\lambda+\mu})}.$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

Finally, by differentiating the generating functions at the point $z = 1$, we arrive at the mean number of customers in our model:

$$L_r = E\begin{bmatrix}\text{number of customers}\\ \text{in the retrial box}\end{bmatrix}$$
$$= \frac{\lambda^2\left[\dfrac{\mu}{\lambda+\mu}\,v_\lambda^{(2)} + \dfrac{2\mu}{(\lambda+\mu)^2}\,v_\lambda + \lambda(v_\lambda v_{\lambda+\mu}^{(2)} - v_\lambda^{(2)} v_{\lambda+\mu})\right]}{2\left(\dfrac{\mu}{\lambda+\mu} + \lambda v_\lambda - \lambda v_{\lambda+\mu}\right)\left(\dfrac{\mu}{\lambda+\mu} - \lambda v_{\lambda+\mu}\right)},$$

$$L_1 = E\begin{bmatrix}\text{number of customers in}\\ \text{station 1 including the one}\\ \text{in service or in blocking}\end{bmatrix} = L_r + \frac{\dfrac{\mu\lambda}{\lambda+\mu}\,v_\lambda}{\dfrac{\mu}{\lambda+\mu} + \lambda(v_\lambda - v_{\lambda+\mu})}, \qquad \textbf{(20)}$$

$$L_s = E[\text{number of customers in the system}] = L_1 + \lambda\bar{u}_2,$$

where $v_u^{(2)} = -[d^2\psi_u(s)/ds^2]|_{s=0}$.

Remark 2: In our model, the customer who departs from station 1 always leaves behind an idle server and so the time between two consecutive departures from station 1 can be considered as a Busy Cycle. Now, from Eqs. (8) and (5), it is easy to see that

$$E(\text{Busy Cycle}) = E(T_1)(1-\pi_0) + E(T_0)\pi_0 = \frac{1}{\lambda}$$

and so

$$\Pr[\text{server in station 1 idle}] = \frac{\text{E(Idle Period)}}{E(\text{Busy Cycle})} = \frac{\dfrac{1}{\lambda+\mu}(1-\pi_0) + \dfrac{1}{\lambda}\pi_0}{\dfrac{1}{\lambda}}$$

and this is another way to obtain (and to explain) relation (19).

Remark 3: If we assume in our system that for all $x \geq 0$, $B_2(x) \equiv 1$, then the model becomes the usual $M/G/1$ queuing model with constant retrial policy. In this case, we obtain the results of Farahmand [7]. Thus, by substituting $B_2(x + w) = 1$ in Eq. (2), it is easy to see that $\psi_u(s)$, v_u, and $v_u^{(2)}$ become, for all u, the Laplace transform and the first two moments, respectively, of the service time. Moreover the condition $\rho^* < 1$ coincides with Farahmand's condition for ergodicity, and π_0 and L_r in Eqs. (5) and (20) agree with p_0 and $\bar{m}$, respectively, in Eqs. (2.14) and (3.3) of Farahmand [7].

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

6. THE ORDINARY MODEL

Let us assume now that in our retrial model, the mean retrial rate μ tends to infinity. Then, it is easy to understand that the customers in the retrial box try continuously to find a position for service and so they become, in fact, ordinary (queued up) customers. Thus, when $\mu \to \infty$, our model becomes the ordinary model of two queues in tandem without intermediate space. Define the function $\psi(s)$ as

$$\psi(s) = \int_0^\infty e^{-sx}\, d_x(B_1(x)B_2(x)),$$

and the quantities

$$\rho = -\lambda \left.\frac{d\psi(s)}{ds}\right|_{s=0} = \lambda \int_0^\infty x\, d_x(B_1(x)B_2(x)),$$

$$\rho^{(2)} = \lambda^2 \left.\frac{d^2\psi(s)}{ds^2}\right|_{s=0} = \lambda^2 \int_0^\infty x^2\, d_x(B_1(x)B_2(x)).$$

It is easy to see that

$$\lim_{\mu\to\infty} \psi_{\lambda+\mu}(s) = \psi(s), \qquad \lim_{\mu\to\infty} \rho^* = \rho, \qquad \lim_{\mu\to\infty} v_{\lambda+\mu} = \frac{\rho}{\lambda}, \qquad \lim_{\mu\to\infty} v^{(2)}_{\lambda+\mu} = \frac{\rho^{(2)}}{\lambda^2},$$

and from Eq. (19) or (5),

$$\bar\pi_0 = \lim_{\mu\to\infty} \Pr[\text{the server in station 1 is idle}] = \lim_{\mu\to\infty} \pi_0 = \frac{1-\rho}{1-\rho+\lambda v_\lambda},$$

and from Eq. (4),

$$\bar\Pi(z) = \lim_{\mu\to\infty} \Pi(z) = \frac{z\psi_\lambda(\lambda-\lambda z) - \psi(\lambda-\lambda z)}{z-\psi(\lambda-\lambda z)}\,\bar\pi_0.$$

Using the above relations in Eqs. (9)–(13), we arrive easily at

$$\pi_{00} = \Pr[\text{total system idle}] = \lim_{\mu\to\infty} P_{00}(z) = b_2^*(\lambda)\bar\pi_0,$$

$$\pi_{01} = \Pr\begin{bmatrix}\text{first server idle}\\ \text{second server busy}\end{bmatrix} = \lim_{\mu\to\infty} P_{01}(z) = (1-b_2^*(\lambda))\bar\pi_0,$$

$$\bar P_{10}(z) = \lim_{\mu\to\infty} P_{10}(z) = \frac{\lambda\bar\Pi(z)[b_2^*(\lambda-\lambda z) - \psi(\lambda-\lambda z)]}{z(\lambda-\lambda z)} + \frac{\bar\pi_0}{z}\left\{\frac{\lambda\psi(\lambda-\lambda z) - \lambda z\psi_\lambda(\lambda-\lambda z)}{\lambda-\lambda z} - b_2^*(\lambda)\right\},$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

$$\bar{P}_{11}(z) = \lim_{\mu\to\infty} P_{11}(z) = \frac{\lambda[\bar{\Pi}(z) - (1-z)\bar{\pi}_0]}{z} \frac{1 - b_1^*(\lambda - \lambda z)}{(\lambda - \lambda z)} - \bar{P}_{10}(z),$$

$$\bar{P}_{21}(z) = \lim_{\mu\to\infty} P_{21}(z) = \frac{\lambda\bar{\Pi}(z)[b_1^*(\lambda - \lambda z) - \psi(\lambda - \lambda z)]}{z(\lambda - \lambda z)} + \frac{\bar{\pi}_0}{z}\left\{\frac{\lambda\psi(\lambda - \lambda z) - \lambda z\psi_\lambda(\lambda - \lambda z)}{\lambda - \lambda z} - b_1^*(\lambda - \lambda z)\right\}.$$

Finally, from relations (20),

$$L_1^* = E[\text{number of customers in front of station 1}] = \lim_{\mu\to\infty} L_r = \frac{\lambda[\lambda v_\lambda^{(2)}(1-\rho) + v_\lambda \rho^{(2)}]}{2(1-\rho)(1-\rho+\lambda v_\lambda)},$$

$$L_2^* = E[\text{number of customers in front of station 2}] = L_1^* + \frac{\lambda v_\lambda}{1-\rho+\lambda v_\lambda},$$

$$L_s^* = E[\text{number of customers in the system}] = L_2^* + \lambda\bar{u}_2. \quad \textbf{(21)}$$

Remark 4: Note here that if we define $U_0(z)[U_1(z)]$, the probability generating function of the number of customers in station 1, including the one in service or in blocking, when the server in station 2 is idle (busy), then

$$U_0(z) = \pi_{00} + z\bar{P}_{10}(z) = \frac{z\psi_\lambda(\lambda - \lambda z) - \psi(\lambda - \lambda z)}{z - \psi(\lambda - \lambda z)} \frac{b_2^*(\lambda - \lambda z) - z}{1 - z}\bar{\pi}_0,$$

$$U_1(z) = \pi_{01} + z\bar{P}_{11}(z) + z\bar{P}_{21}(z) = \frac{1 - b_2^*(\lambda - \lambda z)}{1 - z} \frac{z\psi_\lambda(\lambda - \lambda z) - \psi(\lambda - \lambda z)}{z - \psi(\lambda - \lambda z)}\bar{\pi}_0. \quad \textbf{(22)}$$

Relations (22) are seen to agree with the corresponding results [relations (46) and (47)] of Prabhu [16], whereas relations (21) are, in fact, relations (19)–(21) of Avi-Itzhak and Yadin [2].

7. THE CASE OF EXPONENTIAL SERVERS

Let us suppose now that

$$B_1(t) = 1 - e^{-\mu_1 t}, \qquad B_2(t) = 1 - e^{-\mu_2 t}. \quad \textbf{(23)}$$

Then, from Eq. (2),

$$\psi_u(s) = \frac{\mu_1}{\mu_1 + s} - \frac{\mu_1 us}{(\mu_2 + u)(\mu_2 + s)(\mu_1 + \mu_2 + s)} \quad \textbf{(24)}$$

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

and so

$$
\begin{aligned}
v_u &= \frac{1}{\mu_1} + \frac{u}{\mu_2 + u}\,\frac{\mu_1}{\mu_2(\mu_1+\mu_2)} \\
v_u^{(2)} &= \frac{2}{\mu_1^2} + \frac{u}{\mu_2 + u}\,\frac{2\mu_1(\mu_1+2\mu_2)}{\mu_2^2(\mu_1+\mu_2)^2} \\
\rho^* &= \frac{\lambda}{\lambda+\mu} + \frac{\lambda}{\mu_1} + \frac{\lambda+\mu}{\lambda+\mu+\mu_2}\,\frac{\lambda\mu_1}{\mu_2(\mu_1+\mu_2)}.
\end{aligned}
\tag{25}
$$

A question that arises here is: For given μ, μ_1, and μ_2, what is the maximum permitted arrival rate λ^* so as to maintain statistical equilibrium? To give an answer, let $R(\lambda) \equiv \rho^* - 1$. Then, from Eq. (25),

$$
R(\lambda) = \frac{a_3\lambda^3 + a_2\lambda^2 + a_1\lambda - a_0}{\mu_1\mu_2(\lambda+\mu)(\lambda+\mu+\mu_2)(\mu_1+\mu_2)} \tag{26}
$$

with

$$
\begin{aligned}
a_3 &= \mu_1^2 + \mu_2(\mu_1+\mu_2), \\
a_2 &= 2\mu_1^2\mu + \mu_2(\mu_1+\mu_2)(2\mu+\mu_2), \\
a_1 &= \mu_1^2\mu^2 + \mu_2\mu(\mu_1+\mu_2)(\mu+\mu_2-\mu_1), \\
a_0 &= \mu\mu_1\mu_2(\mu_1+\mu_2)(\mu+\mu_2),
\end{aligned}
$$

and the question becomes, For what values of $\lambda > 0$, is $R(\lambda)$ negative?

The denominator of $R(\lambda)$ is, for all $\lambda \geq 0$, always positive, whereas for the numerator $Rn(\lambda)$ of $R(\lambda)$, we observe that

1. $Rn(0) = -a_0 < 0$.
2. $\lim_{\lambda\to+\infty} Rn(\lambda) = +\infty$ and $\lim_{\lambda\to-\infty} Rn(\lambda) = -\infty$.
3. $\lim_{\lambda\to+\infty}(d/d\lambda)Rn(\lambda) = \lim_{\lambda\to-\infty}(d/d\lambda)Rn(\lambda) = +\infty$.
4. If $[dRn(\lambda)/d\lambda]|_{\lambda=0} = a_1 \geq 0$, then $dRn(\lambda)/d\lambda$ is always positive for $\lambda > 0$ and, so, $Rn(\lambda)$ has no turning point in $\{\lambda > 0\}$.
5. If $a_1 < 0$, then $dRn(\lambda)/d\lambda$ has one negative and one positive zero and, so, $Rn(\lambda)$ has only one turning point in $\{\lambda > 0\}$.

Now, it is clear that $R(\lambda)$ has one and only one zero in $\{\lambda > 0\}$, λ^* say, and that $R(\lambda) < 0$ for all $\lambda \in [0, \lambda^*)$. This λ^* is the maximum permitted arrival rate to maintain statistical equilibrium in our system. Moreover, by substituting μ_1 instead of λ in Eq. (26), one can easily see that $R(\lambda = \mu_1)$ is always positive and, so, $\lambda^* < \mu_1$.

Using relations (25) in Eqs. (19) and (20), one can easily write down the closed-form relations for the performance measures of the system in the case of exponential service times. In Table 1, we observe the way in which the mean number of customers in the retrial box L_r varies when we change the values of the parameters λ and μ.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

TABLE 1. The Case of Exponential Service Times ($\mu_1 = \mu_2 = 5$); Values of the Mean Number of Customers in the Retrial Box (L_r)

	μ					
	10	**5**	**2**	**1**	**0.5**	**0.1**
λ	$\lambda^* = 2.859$	$\lambda^* = 2.547$	$\lambda^* = 2.029$	$\lambda^* = 1.625$	$\lambda^* = 1.258$	$\lambda^* = 0.637$
0.5	0.020	0.026	0.047	0.083	0.162	1.603
0.6	0.031	0.041	0.073	0.130	0.264	8.057
1.0	0.109	0.148	0.284	0.581	1.755	
1.2	0.181	0.250	0.509	1.219	10.935	
1.6	0.444	0.656	1.769	36.320		
1.8	0.683	1.069	4.216			
2.0	1.060	1.830	41.125			
2.5	4.151	34.630				

Thus, for $E[\text{retrial rate}] = 1/\mu = 0.2$, for example, L_r increases from 0.026 to 34.630 when we pass from $\lambda = 0.5$ to $\lambda = 2.5$, whereas for $\lambda = 1.6$, L_r increases from 0.444 to 36.32 when we pass from $E[\text{retrial rate}] = 0.1$ to $E[\text{retrial rate}] = 1$.

References

1. Artalejo, J.R. (1996). Stationary analysis of the characteristics of the $M/M/2$ queue with constant repeated attempts. *Opsearch* 33: 83–95.
2. Avi-Itzhak, B. & Yadin, M. (1965). A sequence of two servers with no intermediate queue. *Management Science* 11: 553–564.
3. Choi, B.D., Shin, Y.W., & Ahn, W.C. (1992). Retrial queues with collision arising from unslotted CSMA/CD protocol. *Queueing Systems* 11: 335–356.
4. Cinlar, E. (1975). *Introduction to stochastic processes.* New York: Prentice-Hall.
5. Falin, G. (1990). A survey of retrial queues. *Queueing Systems* 7: 127–168.
6. Falin, G. & Templeton, J.G.C. (1997). *Retrial queues.* New York: Chapman & Hall.
7. Farahmand, K. (1990). Single line queue with repeated demands. *Queueing Systems* 6: 223–228.
8. Hunt, G.C. (1956). Sequential arrays of waiting lines. *Operations Research* 4: 674–683.
9. Kulkarni, V.G. & Liang, H.M. (1997). Retrial queues revisited. In J.H. Dshalalow (ed.), *Frontiers in queueing.* Boca Raton, FL: CRC Press, pp. 19–34.
10. Martin, M. & Artalejo, J.R. (1995). Analysis of an $M/G/1$ queue with two types of impatient units. *Advances in Applied Probability* 27: 840–861.
11. Neuts, M.F. (1968). Two queues in series with a finite, intermediate waiting room. *Journal of Applied Probability* 5: 123–142.
12. Pakes, A.G. (1969). Some conditions for ergodicity and recurrence of Markov chains. *Operations Research* 17: 1058–1061.
13. Perros, H.G. (1989). A bibliography of papers on queueing networks with finite capacity queues. *Performance Evaluation* 10: 255–260.
14. Pourbabai, B. (1990). Tandem behavior of a telecommunication system with finite buffers and repeated calls. *Queueing Systems* 6: 89–108.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press

15. Pourbabai, B. (1993). Tandem behavior of a telecommunication system with repeated calls: A general case without buffers. *European Journal of Operations Research* 65: 247–258.
16. Prabhu, N.U. (1967). Transient behaviour of a tandem queue. *Management Science* 13: 631–639.
17. Ramalhoto, M.F. & Gomez-Corral, A. (1998). Some decomposition formulae for $M/M/r/r + d$ queues with constant retrial policy. *Communications in Statistics—Stochastic Models* 14: 123–145.
18. Suzuki, T. (1964). On a tandem queue with blocking. *Journal of the Operations Research Society of Japan* 6: 137–157.
19. Takahara, G.K. (1996). Fixed point approximations for retrial networks. *Probability in the Engineering and Informational Sciences* 10: 243–259.
20. Yang, T. & Templeton, J.G.C. (1987). A survey on retrial queues. *Queueing Systems* 2: 203–233.

https://doi.org/10.1017/S0269964801153027 Published online by Cambridge University Press