FINITE CAPACITY ASSEMBLY-LIKE QUEUES

U. NARAYAN BHAT

Department of Statistics, Southern Methodist University, Dallas, Texas 75275, USA

Received 3 September 1985
(Revised 27 January 1986)

Abstract

In assembly lines, service involves assembling units coming from more than one source. In queue terminology, we may consider this situation as one in which service is rendered only to groups of customers – one from each class. In this paper we give procedures to determine response time characteristics of such a system under Markovian assumptions when a finite capacity restriction is imposed. This restriction is imposed to reflect reality as well as to make analysis tractable. In the course of this study, we also give a recursive technique to determine the distribution of the time taken for a specific number of departures in a Poisson queue from an arbitrary initial state. We demonstrate that this distribution is related to the response time distribution of the assembly-like queue. We believe that this procedure will also be of independent interest.

Keywords

Assembly-like queues, queues with paired customers, response time, Markovian systems, first-passage time for departures.

1. Introduction

At an assembly point on a production line, units coming from various sources are assembled to make a product. In this process, typically one unit from each source is needed. Such queueing systems in which service can be rendered only to groups of customers – one from each source – have come to be known as assembly-like queues. Harrison [4] investigated the equilibrium properties of the underlying waiting time processes and came to the conclusion that, when there is no restriction on system capacity, the vector waiting time process (with waiting time of the nth arriving customer in each class as elements) does not converge in distribution to a nondefective random vector as $n \to \infty$. This result is only to be expected, since the slowest traffic dominates the waiting time process. Harrison also showed that a modified waiting time process which is the minimum of waiting times of the nth customers in the various classes did converge in distribution as $n \to \infty$. Although useful in understanding the properties of the underlying processes, this result is not useful in practice if one is interested in the waiting times of the various classes of customers.

In its generality, the structure of an assembly-like queue may be described as follows:

(i) Customers from m sources arrive for service at a single facility.

(ii) Service is given only to groups of customers with groups made up of one customer from each source.

(iii) The waiting room for customers from each source has a specified capacity.

(iv) For the study of waiting times, one assumes that the customer groups are served in the order they are formed.

(v) The service facility might also be equipped with the capability to serve more than one type of customer group, each having its own set of sources.

The system analyzed here is a simplified version with assumptions for arrival and service processes that make the system Markovian. We consider the case $m = 2$ in detail and outline the procedure by which it can be extended to $m > 2$, using the case $m = 3$. We assume exponential interarrival and service times and consider only one set of sources, thus resulting in service for only one type of customer groups. We also assume that the waiting rooms of customers are finite, not necessarily being the same for each source. This assumption of finiteness overcomes the problems identified in the Harrison study. However, it should be noted that the capacity of the waiting room is limited only by our computational capabilities.

In order to avoid the problems posed by the unlimited waiting room, Latouche [5] bounded the excess of one class of customers over the other in an assembly-like queue with two customer classes. He used the algorithmic technique introduced by Neuts [7] to determine the equilibrium distribution of the number of customers in the system. However, in many situations bounding the number of customers in each class rather then bounding the excess seems more practical. This is the approach we take in our study.

Our study of assembly-like queues was motivated by the dataflow model of a computer system. In the dataflow architecture of computer systems (Dennis [2]), execution of programs is data driven in the sense that each instruction is enabled for execution just when each required operand has been supplied by the completion of the predecessor instruction. Thus, operands are processed only when one operand from each of the required classes is available, exactly in an assembly-like manner. Thus, in its generality a processor node of a dataflow model may exhibit all properties of assembly-like queues described above, and the system we consider is a simplified model for such a node. Even though the primary justification for assumptions in our model is the ease of analysis, the finite waiting room assumption can be justified also by the architectural restrictions imposed by the initial attempts in system design (see Dennis and Misunas [3]). As can be seen later, only approximate solution techniques seem feasible if some of the major model assumptions are relaxed.

The objective of our study is the determination of the distribution and the moments of the response time (= waiting time + service time) using known results for the steady state distribution of the number of customers in the system from each source. If one is interested only in the unconditional mean response time, Little's law can be used to derive it if the queue length distribution is known. The value of our study is in providing expressions for response time characteristics conditional on the state of the system on the arrival of the customer and the ability to determine the unconditional distribution when the state space is relatively small.

In an assembly-like queue, when an arriving customer finds more customers of its class than the others, the waiting time has the characteristics of the time taken for a specified number of departures in a regular queue. Therefore, in the next section we give a procedure to determine the distribution characteristics of a specified departure time in an M/M/1 (Poisson arrival, exponential service, single server) queue. This result may also be of independent interest since we are not aware of the availability of a similar result in the literature for the finite waiting room case.

Section 3 describes the procedure for the analysis of assembly-like queues with two customer classes. An outline of the extension to three customer classes is given in sect. 4. The paper concludes with some remarks on the feasibility of similar analysis in the general case in the last section.

2. Preliminary results: A specified departure time in an M/M/1 queue

Consider a single-server queue with Poisson arrivals and exponential service. Let λ be the arrival rate and μ the service rate. Let $A(t)$ and $D(t)$ be the number of arrivals and the number of departures during $(0, t)$, and $Q(t)$ the number of customers in the system at time t. Let

$$T_d^{(i)} = \inf\{t \,|\, D(t) = d,\ Q(0) = i\} .$$

The random variable $T_d^{(i)}$ is the first passage time for the counting process $D(t)$ and is the time taken for d departures, having started initially with i customers in the system.

For $i = 0$ and an infinite system capacity, the distribution $T_d^{(0)}$ can be determined from the joint distribution of $A(t)$ and $D(t)$ given by Pegden and Rosenshine [8]. For an arbitrary i, the distribution of $T_d^{(i)}$ can be obtained from similar results derived by Boxma [1]. To use Boxma's results, let

$$F_{nr}^{(i)}(t) = P(A(t) = n,\ D(t) = r \mid Q(0) = i).$$

Now one may write

$$P(T_d^{(i)} \leqslant t) = \sum_{r=d}^{\infty} \sum_{n=0}^{\infty} F_{nr}^{(i)}(t). \tag{2.1}$$

Besides the fact that these results are for the infinite capacity queue, the explicit expressions are quite cumbersome for getting $E(T_d^{(i)})$. The method we propose here is quite simple and is based on the Kolmogorov equations for the process. We consider the queue with a finite capacity K. Even though the expressions for the distribution become cumbersome for larger values of K, one should be able to obtain the mean of the departure time directly for specific parameter values even for large K.

Define

$$P_{nr}^{(i)}(t) \equiv P_{nr}(t) = P(Q(t) = n,\ D(t) = r \mid Q(0) = i), \tag{2.2}$$

$$n = 0, 1, 2, \ldots, K$$

$$r = 0, 1, 2, \ldots, d-1.$$

Also

$$P_{nr}^{(i)}(0) = 1 \quad \text{if } n = i,\ r = 0$$

$$= \text{otherwise}.$$

Note that in order to consider the first passage time of the process $\{Q(t), D(t)\}$ to state $D(t) = d$, we make the state (n, d), $n = 0, 1, 2, \ldots, K$ absorbing. Defining $g_d^{(i)}(t)$ as the density function of $T_d^{(i)}$, we have

$$g_d^{(i)}(t) = \mu \sum_{n=1}^{K} P_{n,d-1}^{(i)}(t), \tag{2.3}$$

which remains true even when $K \to \infty$. The transition probabilities $P_{nr}(t)$ satisfy the following forward Kolmogorov equations.

$$
\begin{aligned}
P'_{00}(t) &= -\lambda P_{00}(t), && i = 0 \\
P'_{i0}(t) &= -(\lambda+\mu)P_{i0}(t), && i > 0 \\
P'_{n0}(t) &= -(\lambda+\mu)P_{n0}(t) + \lambda P_{n-1,0}(t), && n = i+1,\dots,K \\
P'_{i-r,r}(t) &= -(\lambda+\mu)P_{i-r,r}(t) + \mu P_{i-r+1,r-1}(t), && r = 1,2,\dots,i \\
P'_{0r}(t) &= -\lambda P_{0r}(t) + \mu P_{1,r-1}(t), && r = i+1,\dots,d-1 \\
P'_{nr}(t) &= -(\lambda+\mu)P_{nr}(t) + \lambda P_{n-1,r}(t) + \mu P_{n+1,r-1}(t), \\
&\qquad \{(n,r) = (i+j-r+1,r) \quad \text{for } r = 1,2,\dots,i \\
&\qquad\qquad\;\; = (1+j,r) \quad \text{for } r = i+1,\dots,d-1, \\
&\qquad j = 0,1,2,\dots,K-i-1\} \\
P'_{Kr}(t) &= -\mu P_{Kr}(t) + \lambda P_{K-1,r}(t), && r = 1,2,\dots,d-1.
\end{aligned}
\tag{2.4}
$$

Define the Laplace transform

$$
\phi_{nr}(\theta) = \int_0^\infty e^{-\theta t} P_{nr}(t)\,dt \qquad \mathrm{Re}(\theta) > 0 .
$$

With the initial conditions defined in (2.2), we get

$$
\int_0^\infty e^{-\theta t} P'_{nr}(t)\,dt = \begin{cases} \theta\phi_{i0}(\theta) - 1 & \text{if } n=i,\ r=0 \\ \theta\phi_{nr}(\theta) & \text{otherwise} . \end{cases}
$$

Taking transforms, corresponding to the sets of equations (2.4), we have, respectively,

$$
\phi_{00}(\theta) = (\theta + \lambda)^{-1}
$$

$$
\phi_{i0}(\theta) = (\theta + \lambda + \mu)^{-1}
$$

$$\phi_{n0}(\theta) = \lambda(\theta + \lambda + \mu)^{-1} \phi_{n-1,0}(\theta)$$

$$\phi_{i-r,r}(\theta) = \mu(\theta + \lambda + \mu)^{-1} \phi_{i-r+1,r-1}(\theta)$$

$$\phi_{0r}(\theta) = \mu(\theta + \lambda)^{-1} \phi_{1,r-1}(\theta)$$

$$\phi_{nr}(\theta) = (\theta + \lambda + \mu)^{-1} [\lambda\phi_{n-1,r}(\theta) + \mu\phi_{n+1,r-1}(\theta)]$$

$$\phi_{Kr}(\theta) = \lambda(\theta + \mu)^{-1} \phi_{K-1,r}(\theta). \tag{2.5}$$

When K and d are small, these equations can be solved explicitly. For larger values of K and d, an algorithmic procedure which can be implemented using a computer may be suggested. A diagrammatic representation (fig. 1) of the transition structure helps in describing this procedure.

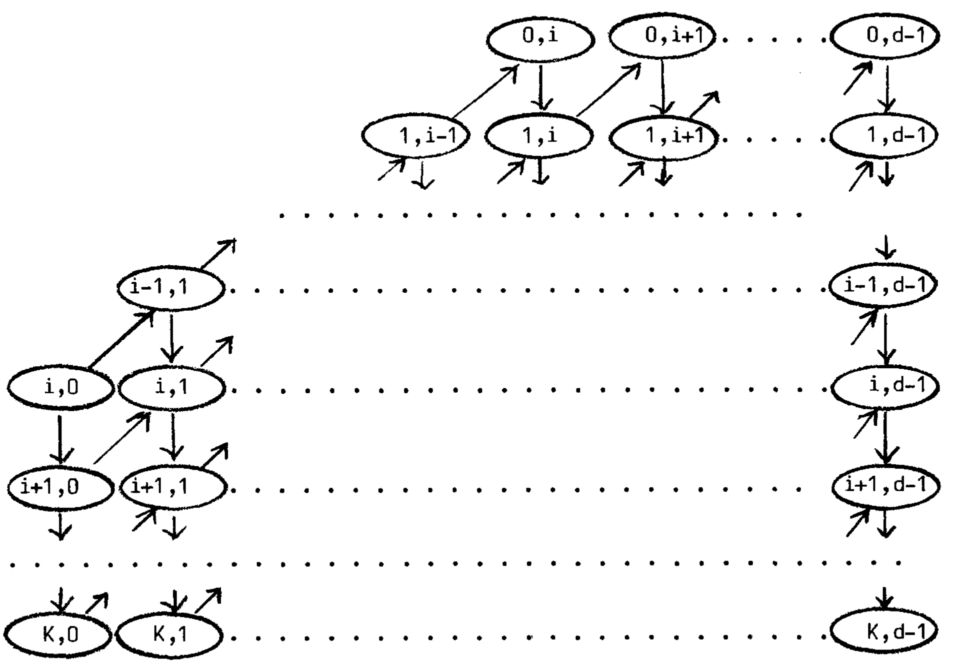

Figure 1.

The following observations may be noted.

(1) When the process is in any one of the states $(0,r)$, $r = i, \ldots, d-1$, only arrivals affect transition. The residence time in such a state is exponential with rate λ

and the probability is 1 that the next state will be $(0, r + 1)$. The corresponding Laplace transform for the transition time is $\lambda/(\lambda + \theta)$.

(2) When the process is any one of the states (K, r), $r = 0, 1, \ldots, d-1$, only departures affect transition. The residence time in such a state is exponential with rate μ and the probability is 1 that the next state is $(K-1, r+1)$. The corresponding transform for the transition time is given by $\mu/(\mu + \theta)$.

(3) From all other states, both arrivals and departures affect transition and the probability of arrival is $\lambda/(\lambda + \mu)$ and the probability of departure is $\mu/(\lambda + \mu)$. Residence times in these states are exponential with mean $(\lambda + \mu)^{-1}$ and therefore corresponding to an $(n, r) \to (n + 1, r)$ transition we have a Laplace transform $(\lambda/(\lambda + \mu))$ $((\lambda + \mu)/(\lambda + \mu + \theta))$ and corresponding to an $(n, r) \to (n - 1, r + 1)$ transition we have a Laplace transform $(\mu/(\lambda + \mu))$ $((\lambda + \mu)/(\lambda + \mu + \theta))$. Thus, because of the Markovian properties of the system, obtaining the Laplace transform $\phi_{n,d-1}(\theta)$ is simply a problem of tracing the various paths from $(i, 0)$ to $(n, d - 1)$, counting the number of transitions of the four types mentioned above, attaching the appropriate Laplace transform with each transition and getting their product. Thus, suppose a path from $(i, 0)$ to $(n, d - 1)$ includes N_1 of type $1, N_2$ of type 2, and N_3^u of type 3 with an arrival and N_3^d of type 3 with a departure, then the corresponding transform is simply

$$\tau(N_1, N_2, N_3^u, N_3^d) = \left(\frac{\lambda}{\lambda + \theta}\right)^{N_1} \left(\frac{\mu}{\mu + \theta}\right)^{N_2} \frac{\lambda^{N_3^u} \mu^{N_3^d}}{(\lambda + \mu + \theta)^{N_3}}, \tag{2.6}$$

where we have written $N_3^u + N_3^d = N_3$. If there are $H(N_1, N_2, N_3^u, N_3^d)$ paths of this type, we get

$$\begin{aligned}\phi_{n,d-1}(\theta) &= (\theta + \lambda + \mu)^{-1} \sum H(N_1, N_2, N_3^u, N_3^d)\, \tau(N_1, N_2, N_3^u, N_3^d) \\ &= \sum \Gamma(N_1, N_2, N_3^u, N_3^d) \left(\frac{\lambda}{\theta + \lambda}\right)^{N_1} \left(\frac{\mu}{\mu + \theta}\right)^{N_2} \left(\frac{\lambda + \mu}{\lambda + \mu + \theta}\right)^{N_3+1},\end{aligned} \tag{2.7}$$

where the term $(\theta + \lambda + \mu)^{-1}$ of the first expression indicates that the process is still in state $(n, d - 1)$, and that the summation is over possible sets of values of N_1, N_2, N_3^u and N_3^d. The number of paths $H(N_1, N_2, N_3^u, N_3^d)$ can be determined by using a computer algorithm.

Now the Laplace transform of the distribution of $T_d^{(i)}$ is obtained from (2.3) as

$$\int_0^\infty e^{-\theta t} g_d^{(i)}(t)\,dt = \mu \sum_{n=1}^{K} \phi_{n,d-1}(\theta)\,. \tag{2.8}$$

The inversion of this transform can be accomplished by noting from (2.7) that it consists of three factors which are transforms of Erlangian (Gamma) distributions. Let $f_N(t/\lambda)$ denote an Erlangian density such that

$$\int_0^\infty e^{-\theta t} f_N(t\,|\,\lambda)\,dt = \left(\frac{\lambda}{\theta+\lambda}\right)^N .$$

Using this notation, after inverting (2.8) one gets

$$g_d^{(i)}(t) = \mu \sum_{n=1}^{K} \sum \Gamma\left(N_1, N_2, N_3^u, N_3^d\right)$$

$$f_{N_1}(t\,|\,\lambda) \star f_{N_2}(t\,|\,\mu) \star f_{N_3+1}(t\,|\,\lambda+\mu)\,, \tag{2.9}$$

where the second summation is over different values of N_1, N_2, N_3^u and N_3^d resulting from different paths from $(i, 0)$ to $(n, d-1)$. Also, we have used $\star$ to denote convolution. The transform (2.8) can also be inverted using partial fraction expansions. In this case, one would be able to express $g_d^{(i)}(t)$ as

$$g_d^{(i)}(t) = \mu\left[\sum_{i=1}^{N_1} c_{1i} f_i(t\,|\,\lambda) + \sum_{i=1}^{N_2} c_{2i} f_i(t\,|\,\mu) + \sum_{i=1}^{N_3+1} c_{3i} f_i(t\,|\,\lambda+\mu)\right] \tag{2.10}$$

where c_{ji} $(j = 1, 2, 3)$ are appropriate coefficients appearing in the partial fraction expansion.

It may be noted here that Boxma's [1] procedure is basically similar to this solution technique. Without the finite upper limit on $Q(t)$, he is able to use random walk methods to give the number of each type of path. Also, for the same reason, there are only two types of transitions [(1) and (3)], thus simplifying the inversion.

If we are interested only in the mean first passage time to state $D(t) = d$, the procedure given above can be used to determine this quantity in a recursive manner.

Using a well-known property of non-negative random variables, we have

$$E\left[T_d^{(i)}\right] = \int_0^\infty P(T_d^{(i)} > t)\,\mathrm{d}t\,. \tag{2.11}$$

Combining this result with (2.3) we get

$$\begin{aligned} E\left[T_d^{(i)}\right] &= \int_0^\infty \sum_{n=0}^{K} \sum_{r=0}^{d-1} P_{nr}^{(i)}(t)\,\mathrm{d}t \\ &= \lim_{\theta \to 0} \sum_{n=0}^{K} \sum_{r=0}^{d-1} \phi_{nr}(\theta) \\ &= \sum_{n=0}^{K} \sum_{r=0}^{d-1} \phi_{nr}, \end{aligned} \tag{2.12}$$

where we have written $\lim_{\theta \to 0} \phi_{nr}(\theta) = \phi_{nr}$. These quantities ϕ_{nr} can be determined from the set of equations (2.5) recursively as follows. Let $\theta \to 0$ in (2.5). We now have the set of equations

$$\begin{aligned} \phi_{00} &= \lambda^{-1}, && i = 0 \\ \phi_{i0} &= (\lambda + \mu)^{-1}, && i > 0 \\ \phi_{n0} &= \lambda(\lambda + \mu)^{-1} \phi_{n-1,0}, && n = i+1, \ldots, K \\ \phi_{i-r,r} &= \mu(\lambda + \mu)^{-1} \phi_{i-r+1,r-1}, && r = 1, 2, \ldots, i \\ \phi_{0r} &= \mu\lambda^{-1} \phi_{1,r-1}, && r = i+1, \ldots, d-1 \end{aligned}$$

$$\phi_{nr} = (\lambda + \mu)^{-1} [\lambda \phi_{n-1,r} + \mu \phi_{n+1,r-1}] ,$$

$$(n, r) = (i + j - r + 1, r) \qquad \text{for } r = 1, 2, \ldots, i$$

$$= (1 + j, r) \qquad \text{for } r = i + 1, \ldots, d - 1$$

$$j = 0, 1, 2, \ldots, K - i - 1$$

$$\phi_{Kr} = \lambda \mu^{-1} \phi_{K-1,r} , \qquad r = 1, 2, \ldots, d - 1. \tag{2.13}$$

For specific values of λ and μ, these can be recursively solved. When K and d are small, explicit expressions can be easily obtained, but when K and d are large, a computer can eliminate the drudgery of algebraic simplifications. The type of results one obtains from this is illustrated in the following example.

EXAMPLE: $i = 3, K = 6, d = 5$

The recursion is illustrated in fig. 2.

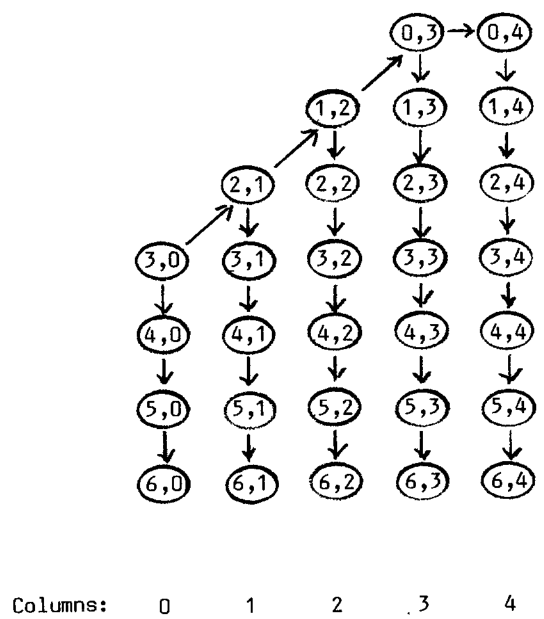

Figure 2.

The first four equations of the set (2.13) give the ϕ values for (3, 0) to (6, 0), (2, 1) to (0, 3), and (0, 4), which are all on the upper part of the boundary. The fifth and sixth equations are used to determine the remaining ϕ values. After some algebraic simplification, we get

$$\phi_{3+r,0} = \frac{\lambda^r}{(\lambda+\mu)^{r+1}}, \qquad r = 0, 1, 2, 3$$

$$\phi_{3+r,1} = \frac{\binom{r+2}{1}\lambda^{r+1}\mu}{(\lambda+\mu)^{r+3}}, \qquad r = -1, 0, 1, 2, 3$$

$$\phi_{3+r,2} = \frac{\binom{r+4}{2}\lambda^{r+2}\mu^2}{(\lambda+\mu)^{r+5}}, \qquad r = -2, -1, 0, 1, 2, 3$$

$$\phi_{3+r,3} = \frac{\binom{r+6}{3}\lambda^{r+3}\mu^3}{(\lambda+\mu)^{r+7}}, \qquad r = -3, -2, -1, 0, 1, 2, 3$$

$$\phi_{3+r,4} = \frac{\lambda^{r+4}\mu^4}{(\lambda+\mu)^{r+9}}\left[\binom{4}{3}\frac{\lambda+\mu}{\lambda} + \binom{5}{3} + \binom{6}{3} + \ldots + \binom{r+7}{3}\right],$$

$$r = -3, -2, -1, 0, 1, 2, 3 . \qquad (2.14)$$

It should be noted that the expressions are easy to obtain for columns 0, 1, 2, . . . , i. For columns beyond i, because of the contribution of the numerous paths, the expressions become quite cumbersome, as evidenced by the last result in (2.14). In view of this, computer implementation of the recursion is recommended.

The recursive scheme can also be used as $K \to \infty$. Clearly, the terms decrease as one moves down along the columns and appropriate cutoff values can be prescribed to terminate calculations in each column.

3. Assembly-like queues with two customer classes

We assume the following queue characteristics:

(i) Two classes of customers arrive in Poisson streams with rates λ_1 and λ_2.

(ii) Service requires pairs of customers – one of each class.

(iii) Service times are independent and identically distributed exponential random variables with rate μ.

(iv) Waiting space is provided for only a maximum of K customers in the system for each class including the one in service, if any. It should be noted that the analysis remains unchanged even when queue capacities are different for different classes of customers. We make the assumption of equal capacity for notational ease.

(v) Service is rendered in the order the pairs are formed.

Let p_{n_1,n_2} be the steady state probability that there are n_1 of class 1 and n_2 of class 2 customers in the system. With these assumptions we can easily write down the state balance equations and solve them in the usual manner. When the capacity constraint is K, there are $(K+1)^2$ probabilities to be determined. For the remainder of this discussion we shall assume that K is relatively small and our computational capabilities are such that these probabilities can be determined without significant error.

Let $f^{(i)}_{n_1,n_2}(x)$ be the probability density of the response time (= waiting time + service time) of a class i $(i = 1, 2)$ customer when it finds n_1 of class 1 and n_2 of class 2 customers on its arrival $(n_1, n_2 = 0, 1, 2, \ldots, K)$. Also, let $R^{(i)}_{n_1 n_2}$ be the corresponding mean response time. In the determination of $f^{(i)}_{n_1 n_2}(x)$ and $R^{(i)}_{n_1 n_2}$ we have to delineate different cases.

Case (i): $n_1 = n_2 = n \geqslant 0$.

When a customer arrives, if there are only pairs (and no excess in any class) in the system, then the response time will include the amount of time needed to serve the pairs and its own service time if a customer of the other class has already arrived by that time, or the amount of time for a customer of the other class to arrive plus its service time. We have

$$f^{(i)}_{nn}(x) = \int_0^x e^{-\mu y} \frac{\mu^n y^{n-1}}{(n-1)!} \left(1 - e^{-\lambda_j y}\right) \mu e^{-\mu(x-y)}\, dy$$

$$+ \int_0^x e^{-\mu y} \frac{\mu^n y^{n-1}}{(n-1)!} e^{-\lambda_j y} f^{(i)}_{00}(x-y)\, dy$$

$$K \geqslant n \geqslant 0; \quad i,j = 1,2, \quad i \neq j, \tag{3.1}$$

giving

$$R_{nn}^{(i)} = \frac{n+1}{\mu} + \frac{\mu^n}{\lambda_j(\mu+\lambda_j)^n},$$

$$i,j = 1,2, \quad i \neq j; \quad 0 \leqslant n \leqslant K. \tag{3.2}$$

Case (ii): $n_i < n_j \quad (i,j = 1,2; \ i \neq j)$.

When a customer of class i arrives if the excess of customers (after forming pairs) is in class j $(j \neq i)$, then its response time is the total service time of the pairs ahead of it plus one service time. We have

$$f_{n_1 n_2}^{(1)}(x) = \mathrm{e}^{-\mu x} \frac{\mu^{n_1+1} x^{n_1}}{n_1!}, \quad n_1 < n_2, \quad n_1, n_2 = 0,1,2,\ldots,K, \tag{3.3}$$

giving

$$R_{n_1 n_2}^{(1)} = \frac{n_1+1}{\mu}, \qquad n_1 < n_2, \quad n_1, n_2 = 0,1,2,\ldots,K. \tag{3.4}$$

Similar expressions follow when $n_2 < n_1$.

Case (iii): $n_i > n_j \quad (i,j = 1,2; \ i \neq j)$.

When a customer of class i arrives if the excess of customers is in its own class, then the response time is the service times of pairs in the system plus the amount of time needed to deplete the $(n_i - n_j + 1)$ customers to zero. Since no more arrival of class i is needed, this time is simply the time taken for $n_i + 1$ departures in an M/M/1 queue with arrival rate λ_j and an initial number of n_j customers. As defined in sect. 2, this is the distribution of the random variable $T_{n_i+1}^{(n_j)}$ for that system. For convenience, we shall denote its probability density as $g_{n_i+1}^{(n_j)}(t \mid \lambda_j)$. Thus we have

$$f_{n_1 n_2}^{(1)}(x) = g_{n_1+1}^{(n_2)}(x \mid \lambda_2), \qquad n_1 > n_2,$$

$$n_1, n_2 = 0,1,2,\ldots,K \tag{3.5}$$

giving

$$R^{(1)}_{n_1 n_2} = E\left[T^{(n_2)}_{n_1+1} \mid \lambda_2\right], \qquad n_1 > n_2 . \tag{3.6}$$

Similar expressions follow when $n_2 > n_1$.

Combining the distributions of the number of customers in the system and their response time we have, for $f^{(i)}(x)$ the response time density of an arriving customer of class i ($i = 1, 2$)

$$f^{(i)}(x) = q_i^{-1} \sum_{n_1=0}^{K-1} \sum_{n_2=0}^{K} p_{n_1 n_2} f^{(i)}_{n_1 n_2}(x) , \tag{3.7}$$

giving $R^{(i)}$, the corresponding mean response time, as

$$R^{(i)} = q_i^{-1} \sum_{n_1=0}^{K-1} \sum_{n_2=0}^{K} p_{n_1 n_2} R^{(i)}_{n_1 n_2} = \frac{1}{q_i \lambda_i} \sum_{n_1} n_1 \sum_{n_2} p_{n_1 n_2} , \tag{3.8}$$

where we have used q_i ($i = 1, 2$) to denote the probability that a class i customer will not be blocked from entering the system. For instance, we have

$$q_1 = \sum_{n_1=0}^{K-1} \sum_{n_2=0}^{K} p_{n_1 n_2} .$$

The last expression in (3.8) is obtained using Little's law.

In order to extend these results for an assembly-like queue with three classes of customers, we need the distribution and the mean of the time taken to serve a specified number of pairs in a two-class system starting with some initial state (i, j). This can be obtained exactly the same way as for a specified departure time in a one-class (M/M/1) system, described in sect. 2.

4. Assembly-like queues with three customer classes

We make the following modifications to the system characteristics assumed at the beginning of sect. 3.

(i) Three classes of customers arrive in Poisson streams with rates λ_1, λ_2 and λ_3.

(ii) Service requires triplets of customers – one coming from each class.

(iii) Service is rendered in the order of formation of triplets. In order to simplify the discussion, we denote the above system by $(\mathrm{ALQ})_{123}$. In a similar manner, if we are considering an assembly-like queue with two Poisson arrival streams with rates λ_1 and λ_2, we denote it by $(\mathrm{ALQ})_{12}$, etc.

Let $p_{n_1 n_2 n_3}$ be the steady state probability that there are n_1 of class 1, n_2 of class 2, and n_3 of class 3 customers in the system. With a finite capacity constraint of K for the number of customers in each class, the steady state probability $p_{n_1 n_2 n_3}$ can be determined by solving the $(K + 1)^3$ state balance equations in the usual manner.

Let $f^{(i)}_{n_1 n_2 n_3}(x)$ be the probability density of the response time (= waiting time + service time) of a class i customer when it finds n_1 of class 1, n_2 of class 2, and n_3 of class 3 customers on its arrival ($i = 1, 2, 3$; $n_1, n_2, n_3 = 0, 1, 2, \ldots, K$). This conditional density can be determined in different cases as indicated below.

Case (i): $(n_1 = n_2 = n_3 = 0)$.

If an arriving customer finds no customers in any of the classes, the waiting time is the maximum of the arrival times of customers from the other two classes.

Case (ii): (Customer of one class arrives; only one of the other two classes has customers in the system.)

Suppose a class 2 customer arrives when $n_1 > 0$, $n_2 = n_3 = 0$. Then the waiting time for service is the time until a customer of class 3 arrives.

Case (iii): (After the arrival, the arriving customer class has the least or one of the least number of customers.)

Suppose a customer of class 1 arrives to find n_1 customers. Let $n_1 \geqslant 0$, $n_2, n_3 > 0$, and $n_1 < n_2, n_3$. Now the response time is equivalent to the service time of $n_1 + 1$ customers.

Case (iv): (After the arrival, the number of customers in that class is larger than the number in one class, but smaller than the other.)

Suppose a customer of class 1 arrives. Let $n_1 + 1 > n_2$ and $n_1 + 1 \leqslant n_3$. Now the class 3 process does not influence the response time. The response time is the time needed to serve $n_1 + 1$ customers in an M/M/1 queue with arrival rate λ_2 and starting with an initial number n_2.

Case (v): (After the arrival, the arriving class has the largest number of customers in the system.)

Suppose a customer of class 1 arrives. Let $n_1 + 1 > n_2, n_3$. Now the response time is the time needed for $n_1 + 1$ departures in an $(\mathrm{ALQ})_{23}$ system starting with n_2 and n_3 customers in the two classes, respectively. The distribution characteristics of this quantity can be obtained as indicated at the end of sect. 3.

The conditional response time distributions from these different cases can be combined to give the unconditional distribution as described in sect. 3 for the two-class system.

5. Remarks

As demonstrated in the case of the assembly-like queue with three classes, for the determination of the response time characteristics of an n-class system we need specified departure time characteristics of r-tuplets ($r = 1, 2, \ldots, n - 1$) from the corresponding ALQ systems. Clearly, as n gets larger, this becomes quite cumbersome and untractable. It should also be noted that the determination of the steady state distribution of the number of customers in different classes is not a simple problem either in the general case. Furthermore, in the dataflow computer system model described in the introduction, the processor is likely to be required to handle more than one type of customer sets. For instance, the same processor might be used to add/subtract as well as multiply/divide different sets of numbers. One conclusion that can be drawn from the analysis given in sects. 3 and 4 is that, in extensions such as these, exact analyses are likely to be untractable. Then approximation techniques seem to be the only answer to determine general properties of underlying models. An example of such an attempt is the article by Lipper and Sengupta [6] appearing elsewhere in this issue.

Acknowledgements

The author gratefully acknowledges the comments made by the referees, which have improved the presentation.

References

[1] O.J. Boxma, The joint arrival and departure process for the M/M/1 queue, Preprint No. 307, Dept. of Mathematics, Utrecht University (1983).

[2] J.B. Dennis, Data flow super computers, IEEE Computer 13, 11(1980)48.

[3] J.B. Dennis and D.P. Misunas, A preliminary architecture for a basic data flow processor, in: *Proc. 2nd Symp. Comp. Arch.*, Houston (1975) p. 126.

[4] J.M. Harrison, Assembly-like queues, J. Appl. Prob. 10(1973)354.

[5] G. Latouche, Queues with paired customers, J. Appl. Prob. 18(1981)684.

[6] E.H. Lipper and B. Sengupta, Assembly-like queues with finite capacity: Bounds, asymptotics and approximations, Queueing Systems, Theor. and Appl. 1(1986)67.

[7] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models* (The John Hopkins University Press, Baltimore, 1981).

[8] C.D. Pegden and M. Rosenshine, Some new results for the M/M/1 queue, Management Science 28(1982)821.