ON THE CHARACTERISTIC EQUATION OF GENERAL QUEUEING SYSTEMS

Dimitris Bertsimas and Daisuke Nakazato

OR 246-91 September 1990

On the characteristic equation of general queueing systems

Dimitris J. Bertsimas * Daisuke Nakazato †

September 1990

Abstract

There are numerous papers in the literature, which analyze the behavior of specific queueing systems in terms of certain roots of a nonlinear equation. In this paper we show that a very general class of queueing systems, including queues with heterogeneous servers, multiple arrival streams, bulk queues and feedback, is described by a characteristic equation. We prove that the number in queue and the waiting time distribution is a sum of geometric and exponential terms respectively involving the roots of the characteristic equation. The sum is finite if the service time distributions belong to the class R of distributions with rational Laplace transform. We find explicitly the characteristic equation for a wide variety of queueing systems and furthermore, we provide an easy method to generate the characteristic equation for even more general systems. Our results are structural for the case of arbitrary distributions and aim to unify the field, but they can potentially lead to an algorithmic solution for queues with distributions in the class R.

*Dimitris Bertsimas, Sloan School of Management and Operations Research Center, MIT, Cambridge, Ma 02139. The research of the author was partially supported by grants from the Leaders for Manufacturing program at MIT and from Draper Laboratory.

†Daisuke Nakazato, Operations Research Center, MIT, Cambridge, MA 02139.

1 Introduction

One of the classical problems in queueing theory is the steady state and transient analysis of the $GI/G/m$ queue. Ramaswami and Lucantoni [7] and de Smit [8] analyze the steady state behavior of the $GI/R/m$ queue through the matrix geometric method and complex variable methods respectively. Bertsimas [1] analyzes the $R/R/m$ queue, where R is the class of distributions with rational Laplace transforms using the method of stages. All these investigations address the case where the servers are homogeneous and there is only one arrival stream which forms a renewal process. Yu [10] proposes an approach for the case of heterogeneous servers and Erlang interarrival times. There are also numerous papers in the literature, which analyze the behavior of specific queueing systems in terms of certain roots of a nonlinear equation.

In this paper we aim to understand the structure of the solution of an arbitrary $G/G/m$ queue with heterogeneous servers and multiple arrival streams, as well as more complicated systems involving bulk queues and queues with feedback. Moreover, we want to demonstrate that a very general class of queueing systems, including queues with heterogeneous servers, multiple arrival streams, bulk queues and feedback, is described by a characteristic equation, which we characterize explicitly.

Although our results are only structural in the case of systems with arbitrary distributions, they can potentially lead to an algorithmic solution in the case where all the distributions have rational Laplace transform. Moreover, we believe that they lead to a certain unification of the field.

The methodology we used is to extend the method of stages, which has been used for queueing systems only for distributions with rational Laplace transform (Cox [5]), to arbitrary distributions using supplementary variables.

In section 2 we introduce our techniques with respect to the $GI/G/1$ queue. We then proceed in section 3 to the $\{G\}^L/\{G\}^m/m$ queue with L arriving streams and m heterogeneous servers. In section 4 we describe the general methodology to derive the characteristic equation of arbitrary systems and apply it in the context of bulk queues and queues with feedback, while the final section contains some concluding remarks.

2 The generalized method of stages in the GI/G/1 queue

Since we are addressing the transient queue length distribution we assume the system is initially idle. Let $a(x), b(x)$ be the interarrival and service time distribution respectively and let $\alpha(s), \beta(s)$ be the corresponding Laplace transforms. Let

$$\eta_a(x) = \frac{a(x)}{\int_x^\infty a(y)\,dy}, \quad \eta_b(x) = \frac{b(x)}{\int_x^\infty b(y)\,dy}$$

be the corresponding hazard rate functions.

Let $P_n(t, x_a, x_b)$ be the probability that at time t there are n customers in the system, x_a is the elapsed time since the last arrival and x_b is the elapsed time since the current service initiation. We can then write down the Kolmogorov equation that describes the dynamics of the system. For $n \geq 1$,

$$P_n(t+\Delta t, x_a+\Delta t, x_b+\Delta t) = (1 - \eta_a(x_a)\Delta t - \eta_b(x_b)\Delta t)P_n(t, x_a, x_b) +$$

$$\delta(x_a)\int_0^\infty P_{n-1}(t, u, x_b)\eta_a(u)du\ \Delta t + \delta(x_b)\int_0^\infty P_{n+1}(t, x_a, u)\eta_b(u)du\ \Delta t,$$

where $\delta(x)$ is the usual delta function. Taking the limit as $\Delta t \to 0$ we find that for $n \geq 1$

$$\frac{\partial}{\partial t}P_n(t, x_a, x_b) = \delta(x_a)\int_0^\infty P_{n-1}(t, u, x_b)\eta_a(u)du + \delta(x_b)\int_0^\infty P_{n+1}(t, x_a, u)\eta_b(u)du$$

$$-[\frac{\partial}{\partial x_a}P_n(t, x_a, x_b) + \eta_a(x_a)P_n(t, x_a, x_b)] - [\frac{\partial}{\partial x_b}P_n(t, x_a, x_b) + \eta_b(x_b)P_n(t, x_a, x_b)].$$

We now define the following operators that act on a function $f(.)$ from the left:

$f(x_a)\Delta_{0a} = \frac{\partial}{\partial x_a}f(x_a) + \eta_a(x_a)f(x_a)$

$f(x_b)\Delta_{0b} = \frac{\partial}{\partial x_b}f(x_b) + \eta_b(x_b)f(x_b)$

$f(x_a)\Delta_{1a} = -\delta(x_a)\int_0^\infty f(u)\eta_a(u)du$

$f(x_b)\Delta_{1b} = -\delta(x_b)\int_0^\infty f(u)\eta_b(u)du.$

With these definitions the Kolmogorov equations become:

$$\frac{\partial}{\partial t}P_n(t,x_a,x_b) = -P_{n-1}(t,x_a,x_b)\Delta_{1a} - P_n(t,x_a,x_b)\{\Delta_{0a}+\Delta_{0b}\} - P_{n+1}(t,x_a,x_b)\Delta_{1b}.$$

Taking the Laplace transform in terms of t we obtain that for $n \geq 1$:

$$s\pi_n(s,x_a,x_b)+\pi_{n-1}(s,x_a,x_b)\Delta_{1a}+\pi_n(s,x_a,x_b)\{\Delta_{0a}+\Delta_{0b}\}+\pi_{n+1}(s,x_a,x_b)\Delta_{1b} = 0, \tag{1}$$

where $\pi_n(s,x_a,x_b) = \mathcal{L}[P_n(t,x_a,x_b)]$.

Equation (1) is a difference equation in terms of n with constant coefficients. In terms of x_a, x_b it is a linear partial differential equation. In order to solve (1) we use the separation of variables method and we assume that a general solution is $\tilde{\pi}_n(s,x_a,x_b) = \omega(s)^n\phi_a(x_a)\phi_b(x_b)$. Substituting to (1) we obtain:

$$\phi_a(x_a)\phi_b(x_b)\{s + [\Delta_{0a} + \frac{1}{\omega(s)}\Delta_{1a}] + [\Delta_{0b} + \omega(s)\Delta_{1b}]\} = 0,$$

which can be rewritten as

$$\phi_a(x_a)\phi_b(x_b)\{s + \Delta_a(\frac{1}{\omega(s)}) + \Delta_b(\omega(s))\} = 0,$$

if we define the operators $\Delta_i(z) = \Delta_{0i} + z\Delta_{1i},\ i = a, b$.

As a result, $\phi_a(x_a)$, $\phi_b(x_b)$ should be eigenfunctions of $\Delta_a(\frac{1}{\omega(s)})$, $\Delta_b(\omega(s))$ respectively with corresponding spectrums $-\theta_a(s)$, $-\theta_b(s)$. We are thus naturally led in the next lemma to the investigation of the eigenfunction and eigenvalue structure of the operators $\Delta_i(z),\ i = a, b$.

Lemma 1 *Let $\phi_a(x)$ denote an eigenfunction of $\Delta_a(z)$ with a corresponding spectrum $-\theta_a$. Then*

$$\phi_a(x) = e^{-\theta_a x}\bar{A}(x),$$

where

$$\bar{A}(x) = \int_x^\infty a(y)\,dy$$

and the characteristic equation of the spectrum is

$$z\alpha(\theta_a) = 1.$$

Proof

The eigenvalue equation is $\phi_a(x)\Delta_a(z) = -\theta_a\phi_a(x)$, which reduces to the ordinary differential equation (ODE)

$$\frac{d\phi_a(x)}{dx} + \phi_a(x)\eta_a(x) - z\delta(x)\int_0^\infty \phi_a(y)\eta_a(y)\,dy = -\theta_a\phi_a(x)$$

$$\phi_a(x) = 0 \quad (x < 0)$$

Since this is a first order linear ODE it can be solved by direct integration as follows:

$$\phi_a(x) = zKe^{-\theta_a x - \int_0^x \eta_a(y)\,dy} \quad (x \geq 0), \tag{2}$$

where

$$K = \int_0^\infty \phi_a(y)\eta_a(y)\,dy.$$

Substituting (2) to the expression for K we obtain

$$1 = z\int_0^\infty e^{-\theta_a x - \int_0^x \eta_a(y)\,dy}\eta_a(x)\,dx = z\int_0^\infty e^{-\theta_a x}\int_x^\infty a(y)\,dy\eta_a(x)\,dx =$$

$$z\int_0^\infty e^{-\theta_a x}a(x)dx = z\alpha(\theta_a),$$

i.e. the characteristic equation for the spectrum is

$$z\alpha(\theta_a) = 1.$$

Moreover, up to a constant factor the eigenfunction $\phi_a(x)$ is

$$\phi_a(x) = e^{-\theta_a x - \int_0^x \eta_a(y)\,dy} = e^{-\theta_a x}\int_x^\infty a(y)\,dy \quad (x \geq 0).\square$$

Applying now the previous lemma we obtain that

$$\tilde{\pi}_n(s, x_a, x_b) = \omega(s)^n e^{-\theta_a(s)x_a - \theta_b(s)x_b}\bar{A}(x_a)\bar{B}(x_b)$$

and $\omega(s)$ satisfies the following system of equations

$$\begin{cases} \theta_a(s) + \theta_b(s) = s \\ \omega(s) = \alpha(\theta_a(s)) \\ \omega(s)\beta(\theta_b(s)) = 1 \\ |\omega(s)| < 1 \end{cases} \tag{3}$$

The above system of equations has several solutions for $\omega(s)$. If $\beta(s)$ is a rational function where the denominator is a polynomial of degree k, i.e for the model $GI/R/1$, an easy application of Rouche theorem establishes that there are exactly k roots $\omega_1(s), ..., \omega_k(s)$ in the unit circle. In this case the solution would be a linear combination of the form:

$$\pi_n(s, x_a, x_b) = \sum_{i=1}^{k} C_i(s)\omega_i(s)^n e^{-\theta_{a,i}(s)x_a - \theta_{b,i}(s)x_b} \bar{A}(x_a)\bar{B}(x_b), \ n \geq 1.$$

For the $GI/G/1$ queue, in which the service time distribution is completely arbitrary, there is a continuum of roots $\omega_u(s)$ characterized by a continuous parameter u. As a result, we conclude that the transform of the queue length distribution[1] $\pi_n(s)$ is characterized as follows:

$$\pi_n(s) = \sum_u C_u(s)\omega_u(s)^n$$

where $C_u(s)$ is a function independent of n. We have thus established the following result:

Theorem 1 *In a $GI/G/1$ queue, which is initially empty, the transform of the queue length distribution has the following structure for $n \geq 1$:*

$$\pi_n(s) = \sum_u C_u(s)\omega_u(s)^n,$$

where the roots $\omega_u(s)$ are found from (3) The steady-state solution is achieved by letting $s = 0$.

[1]Note that

$$\tilde{\pi}_n(s) = \int_0^\infty \int_0^\infty \tilde{\pi}_n(s, x_a, x_b)\, dx_a dx_b = \frac{-1}{\theta_a(s)\theta_b(s)}(1 - \omega(s))^2 \omega(s)^{n-1}$$

For the steady-state waiting time W under FCFS we let $s = 0$ and we apply the distributional Little's law of Bertsimas and Nakazato [2] to find that the density function of the waiting time is:

$$f_W(t) = \sum_u \hat{C}_u(0)\, e^{-\theta_{a,u}(0)t},$$

where $\hat{C}_u(0) = -\frac{C_u(0)\theta_{a,u}^2(0)}{\lambda(1-w_u)(1-\theta_{a,u}(0))}$ i.e the waiting time distribution is a linear combination of exponential terms. As before, there are finitely many terms in the summation for the $GI/R/1$ queue.

3 The $\{G\}^L/\{G\}^m/m$ queue

In this section we generalize and extend the techniques of the previous section to the multiserver $\{G\}^L/\{G\}^m/m$ queue with L arrival streams and m heterogeneous servers. This queue is notorious for its difficulty. Our main result is that the tail of the transform of the queue length distribution is also a linear combination of geometric terms which we explicitly characterize. Moreover, the steady state distribution of the queue length and the waiting time under FCFS is also a linear combination of geometric and exponential terms respectively. The number of terms in the linear combination is finite only in the case in which all the service time distributions have rational Laplace transform.

In this queue there are L arrival streams. The interarrival times in stream i form a renewal process. Let $\alpha_i(s)$ be the Laplace transform of the interarrival time of stream i, $i = 1, ...L$. Customers from all the streams form a single waiting line. We assume that the service discipline is FCFS. There are m servers in this queue. Server j has a service time distribution with Laplace transform $\beta_j(s)$. For the transient queue length distribution we assume that initially the system is empty.

Let $P_n(t, x_{a_1}, \ldots, x_{a_L}, x_{b_1}, \ldots, x_{b_m})$ be the probability that there are n customers in the system at time t, the elapsed time since the last arrival for each stream i ($i = 1, \ldots, L$) is x_{a_i} and the elapsed time since the current service initiation for server j ($j = 1,\ldots,m$) is x_{b_j}. Let $\pi_n(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m})$ be the Laplace transform of $P_n(t, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m})$.

As in the previous section we write the Kolmogorov equation that describes the dynamics of the system and take transforms with respect to t. Using the operator notation of the previous section we obtain that for $n \geq m$:

$$
\begin{aligned}
&s\pi_n(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m}) + \pi_{n-1}(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m})\sum_{i=1}^{L}\Delta_{1a_i} \\
&+\pi_n(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m})\{\sum_{i=1}^{L}\Delta_{0a_i} + \sum_{j=1}^{m}\Delta_{0b_j}\} \\
&+\pi_{n+1}(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m})\sum_{j=1}^{m}\Delta_{1b_j} = 0,
\end{aligned} \tag{4}
$$

where the operators $\Delta_{0a_i}, \Delta_{1a_i}, \Delta_{0b_j}, \Delta_{1a_i}$ are defined as in the previous section. Our analysis now follows exactly the same path. We use the separation of variables technique to solve (4). We introduce a general solution

$$\tilde{\pi}_n(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m}) = \omega(s)^n \prod_{i=1}^{L}\phi_{a_i}(x_{a_i})\prod_{j=1}^{m}\phi_{b_j}(x_{b_j}),$$

and we obtain that

$$\prod_{i=1}^{L}\phi_{a_i}(x_{a_i})\prod_{j=1}^{m}\phi_{b_j}(x_{b_j})\{s + \sum_{i=1}^{L}\Delta_{a_i}(\frac{1}{\omega(s)}) + \sum_{j=1}^{m}\Delta_{b_j}(\omega(s)) = 0.$$

Arguing as before we require that the $\phi_{a_i}(x_{a_i})$, $\phi_{b_j}(x_{b_j})$, $i = 1,\ldots,L$, $j = 1,\ldots,m$ are eigenfunctions of $\Delta_{a_i}(\frac{1}{\omega(s)})$, $\Delta_{b_j}(\omega(s))$ respectively with corresponding spectrums $-\theta_{a_i}(s)$, $-\theta_{b_j}(s)$. Applying lemma 1 we find that

$$\tilde{\pi}_n(s, x_{a_1},\ldots,x_{a_L},x_{b_1},\ldots,x_{b_m}) = \omega(s)^n e^{-\sum_{i=1}^{L}\theta_{a_i}(s)x_{a_i} - \sum_{j=1}^{m}\theta_{b_j}(s)x_{b_j}}\prod_{i=1}^{L}\bar{A}_i(x_{a_i})\prod_{j=1}^{m}\bar{B}_j(x_{b_j}),$$

and the characteristic equation for $\omega(s)$ is:

$$
\left\{
\begin{array}{lll}
\sum_{i=1}^{L}\theta_{a_i}(s) + \sum_{j=1}^{m}\theta_{b_j}(s) = s & & (a) \\
\omega(s) = \alpha_i(\theta_{a_i}(s)) & (i = 1\ldots L) & (b) \\
\omega(s)\beta_j(\theta_{b_j}(s)) = 1 & (j = 1\ldots m) & (c) \\
|\omega(s)| < 1 & & (d)
\end{array}
\right. \tag{5}
$$

Therefore, we can summarize the previous discussion in the following theorem.

Theorem 2 *In a $\{G\}^L/\{G\}^m/m$ queue, which is initially empty, the transform of the queue length distribution has the following structure for $n \geq m$:*

$$\pi_n(s) = \sum_u C_u(s)\, \omega_u(s)^n,$$

where the roots $\omega_u(s)$ are found from (5) and $C_u(s)$ is independent of n. The steady-state solution is achieved by letting $s = 0$.

The previous theorem is in agreement with the results of Bertsimas [1] who considers the $R/R/m$ queue with homogeneous servers. In the $G/R/m$ case, where the degree of the denominator of the transform of the service time distribution is k, there are $\binom{m+k-1}{m}$ roots inside the unit circle. More generally, in the $\{G\}^L/\{R\}^m/m$ case there are finitely many roots in the linear combination. Theorem 2 generalizes results of Takahashi [9] about the asymptotic behavior as $n \to \infty$ of the steady-state probability that there are n customers in a $Ph/Ph/m$ system.

Moreover, in the case $L = 1$ we can also find the structure of steady state waiting time distribution under FCFS using the distributional Little's law of Bertsimas and Nakazato [2]. Then the density function of the waiting time is

$$f_W(t) = \sum_u \hat{C}_u(0)\, e^{-\theta_{a,u}(0)t},$$

where $\hat{C}_u(0) = -\frac{C_u(0)\theta_{a,u}^2(0)}{\lambda(1-w_u)(1-\theta_{a,u}(0))}$ i.e the waiting time distribution is a linear combination of exponential terms. For the case with L arrival streams using a similar approach to that of Bertsimas and Nakazato [3] we can find that

$$f_W(t) = \sum_u \hat{B}_u(0)\, e^{-\sum_{i=1}^{L} \theta_{a_i,u}(0)t},$$

i.e. the waiting time has exponential tails in this case as well.

3.1 Algorithmic issues

In the case where all distributions belong to the class R of distributions with rational Laplace transform then the variables x_{a_i}, x_{b_j} are discrete and therefore there are only finitely many unknowns $\pi_n(s, x_{a_1}, \ldots, x_{a_L}, x_{b_1}, \ldots, x_{b_m})$ for $n \leq m$. Moreover, there are also only finite terms $C_u(s)$. Using the Kolmogorov equations for $n \leq m$ one would have to solve a (large) linear system in order to find these unknowns. Therefore, a conceptual algorithm for the solution of such systems would be to solve the system of equations (5) to find the roots and then solve a linear system to find the unsaturated probabilities and the coefficients $C_u(s)$. Bertsimas [1] implemented an algorithm of this type for the $R/R/m$ queue. The main difficulty with such an algorithm is not the solution for the roots, but rather that the linear system becomes very large very quickly. Our experience at least with the $R/R/m$ is that finding the roots is computationally quite easy, which is in agreement with the comments of Chaudhry, Harris and Marchal [4]. The algorithm spends most of its time in the solution of the linear system.

4 On the characteristic equation of an arbitrary queue

Our initial goal in this section is to understand the nature and character of the characteristic equation (5) and to generalize it to even more general queueing systems. The final goal is to be able to write the characteristic equation directly for an arbitrary queueing system. We then apply this technique to find the characteristic equation of several systems with bulk arrivals, bulk service and feedback.

4.1 A physical interpretation of the queueing system

To generalize our analysis further, we introduce a direct method to obtain the characteristic equations (5) from the dynamics of the system. The method is one of the most commonly used tools in modern physics. We believe it can also be useful in queueing theory.

The key player of our method is the operator:

$$T(z) = s + \sum_{i=1}^{L} \Delta_{a_i}(z) + \sum_{j=1}^{m} \Delta_{b_j}(\frac{1}{z}) \tag{6}$$

We interpret the definition (6) as follows.

The first term s indicates the transient behavior. The second term is due to arrival processes with the argument z representing an increase in the number in the system by one. The last term is due to service processes, and the argument $1/z$ means a decrease of the number in the system by one. Adding these terms means that all the processes in the system are independent. Therefore, an addition of a new operator $\Delta(u(z))$ to the operator $T(z)$ is equivalent to introduce a new independent process to the system and the argument in $u(z)$ describes how the number in the system changes at each renewal epoch.

The characteristic system of equations (5) is simply the condition that the spectrum of $T(\frac{1}{w(s)})$ equals to 0. An alternative approach to see this is as follows. If $\Pi(z)$ is the generating function

$$\Pi(z) = \sum_n \pi_n(s) z^n,$$

then the problem can be formulated using the compensation method (Keilson [6]) as

$$\Pi(z)T(z) = \chi(z), \tag{7}$$

where $T(z)$ is the operator that describes the homogeneous dynamics and $\chi(z)$ is the compensation part which is regular.

From the structure of the solution

$$\Pi(z) = \sum_n \pi_n(s) z^n = \sum_u \frac{C_u(s)}{1 - w_u(s) z},$$

and therefore $z = \frac{1}{w(s)}$ is a singular point of $\Pi(z)$. From (7) however $\Pi(z) = \chi(z)T^{-1}(z)$. Since $\chi(z)$ is regular it means that $z = \frac{1}{w(s)}$ must be a singular point of $T^{-1}(z)$, i.e. the spectrum of $T^{-1}(\frac{1}{w(s)})$ should be infinity and thus the spectrum of $T(\frac{1}{w(s)})$ should be 0.

In addition, equation (5a) means that the spectrum of $T(\frac{1}{w(s)})$ is the sum of the individual spectrums θ of the operators $\Delta(.)$. Equations (5b,5c) are the individual characteristic equations for the spectrums θ of each of the operators $\Delta(.)$ from lemma 1, while the final equation (5d) is the ergodicity condition.

4.2 The characteristic equation in bulk queues

In this subsection we apply the interpretation of the previous subsection to a general bulk queueing system. Consider a combined bulk service and bulk arrival queue with m heterogeneous servers and L arrival streams, in which server j serves r customers with probability q_{jr} and in the arrival stream i at each arrival epoch there are k customers arriving with probability p_{ik}. Let

$$\nu_j(z) = \sum_r q_{jr} z^r$$

be the generating function of the number of customers server j serves and

$$\mu_i(z) = \sum_k p_{ik} z^k$$

be the generating function of the number of customers arriving in the arrival stream i . Our goal is to characterize again the structure of the system.

We saw in the previous subsection that the argument of the operator $\Delta(u(z))$ is a description of how the number in the system changes at every renewal epoch. Therefore, the operator that describes the system is

$$T(z) = s + \sum_{i=1}^{L} \Delta_{a_i}(\mu_i(z)) + \sum_{j=1}^{m} \Delta_{b_j}(\nu_j(\frac{1}{z})),$$

since the number in the system increases according to $\mu_i(z)$ and decreases according to $\nu_i(z)$. The characteristic equation for this system is thus

$$\begin{cases} \sum_{i=1}^{L} \theta_{a_i}(s) + \sum_{j=1}^{m} \theta_{b_j}(s) = s \\ \mu_i(\frac{1}{\omega(s)})\alpha_i(\theta_{a_i}(s)) = 1 & (i = 1 \ldots L) \\ \nu_j(\omega(s))\beta_j(\theta_{b_j}(s)) = 1 & (j = 1 \ldots m) \\ |\omega(s)| < 1 \end{cases}$$

and the transform of the queue length distribution for $n \geq m$ is

$$\pi_n(s) = \sum_u C_u(s)\, \omega(s)^n.$$

4.3 Queues with feedback

Consider again a heterogeneous system with one arriving stream, with the modification that after each service completion the customer is fed back into the system with probability q, while with probability $1 - q$ leaves the system. In this case

$$\nu(z) = q + (1 - q)z$$

is the generating function of how the number in the system changes after a service completion. In this case the characteristic equation is as follows:

$$\begin{cases} \theta_a(s) + \sum_{j=1}^{m} \theta_{b_j}(s) = s \\ \frac{1}{\omega(s)}\alpha(\theta_a(s)) = 1 \\ \nu(\omega(s))\beta_j(\theta_{b_j}(s)) = 1 & (j = 1 \ldots m) \\ |\omega(s)| < 1 \end{cases}$$

and the transform of the queue length distribution for $n \geq m$ is

$$\pi_n(s) = \sum_u C_u(s)\, \omega(s)^n.$$

Obviously one can consider quite complicated situations, for example combinations of bulking and feedback. These can also be analyzed with the technique we introduced.

5 Concluding Remarks

We have generalized the method of stages to an arbitrary queueing system using operators rather than matrices. This generalization enabled us to derive a rather general procedure to describe the characteristic equation that an arbitrary queue satisfies. Although our method can be only seen as a structural result when the interarrival and service time distributions are arbitrary, it can potentially lead to a finite algorithm in the case in which all the distributions have rational Laplace transform. Moreover, our methods prove that a quite general class of queueing systems (queues with heterogeneous servers, multiple arrival streams, with bulking and feedback) are characterized by geometric tails for the queue length distribution and exponential tail for the waiting time distribution.

References

[1] Bertsimas, D. (1990). "An analytic approach to a general class of $G/G/s$ queueing systems", *Operations Research*, 1, 139-155.

[2] Bertsimas D. and Nakazato D. (1990). "The general distributional Little's law", Operations Research Center technical report, MIT.

[3] Bertsimas, D. and Nakazato, D. (1989). "Transient and busy period analysis of the $GI/G/1$ queue; the method of stages", Operations Research Center technical report, MIT.

[4] Chaudhry M., Harris C. and Marchal W. (1990). "Robustness of rootfinding in single-server queueing models", *ORSA Journal on Computing*, 2, 273-286.

[5] Cox, D.R. (1955). "A use of complex probabilities in the theory of stochastic processes", *Proc. Camb. Phil. Soc.*, 51, 313-319.

[6] Keilson, J. (1965). *Green's function method in probability theory*, Hafner, New York.

[7] Ramaswami, V. and Lucantoni, D. M. (1985). "Stationary waiting time distribution in queues with phase type service and in quasi-birth-death-process", *Stochastic Models*, 1, 125-136.

[8] de SMIT, J.H.A, (1983). "The queue $GI/M/s$ with customers of different types or the queue $GI/H_m/s$", *Advances in Applied Probability*, 15, 392-419.

[9] Takahashi Y., (1981). "Asymptotic exponentiallity of the tail of the waiting time distribution in a Ph/Ph/c queue", *Adv. Appl. Prob.*, 13, 619-630.

[10] Yu O., (1977). "The steady-state solution of a heterogeneous server queue with Erlang service time", *TIMS studies in Management Science*, 7, 199-213.