June 1985 LIDS-P-1481

THE BULK SERVICE QUEUE WITH A GENERAL CONTROL STRATEGY: THEORETICAL ANALYSIS AND A NEW COMPUTATIONAL PROCEDURE

Warren B. Powell*
Pierre Humblet**

EES-83-5

Revised May, 1985

*Department of Civil Engineering, Princeton University.
Research supported by NSF under grant No. CEE-8203476.

**Research was conducted at the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology with partial support provided by NSF under grant No. NSF-ECS-8310698.

This paper has been accepted by Operations Research.

Abstract

A general framework is developed for analyzing a wide class of vehicle dispatching strategies for bulk arrival, bulk service queues. A simple derivation of the queue length transform for the imbedded Markov chain is provided and a new computational procedure is developed for finding the moments of the queue length distribution. Extensive computational tests are reported which demonstrate that the new procedure is significantly faster and more stable than the standard method referred to in the literature, which requires solving a set of simultaneous linear equations. Formulas for the mean and variance of the length of the queue are provided for the general case of compound Poisson arrivals, random batch capacities, general service times and a general control strategy.

We consider the problem of analyzing bulk arrival, bulk service queues in steady state where customers arrive in groups at a point and wait for the next available vehicle. Under the simplest policy, termed here the bulk queue with no control, the vehicle, on arriving, will accept customers up to the capacity of the vehicle and then leave, independent of the number of customers waiting. The available capacity of the vehicle is assumed to be random but independent of the length of the queue or the capacity of all previous vehicles. The time from the departure of one vehicle to the earliest possible departure of the next vehicle is a service period, where successive service periods are assumed to be independently and identically distributed. The end of a service period is referred to as a dispatch instant, since a departure may or may not occur, depending on the control strategy in use and the length of the queue. If the vehicle is simply held until, for example, the length of the queue reaches a specified minimum, then the time during which the departure is delayed is an idle period.

Additional constraints are frequently placed on vehicle departures to avoid the possibility of having vehicles depart with uneconomically small loads. In this paper an approach is described for analyzing a very wide class of control strategies, where two of the strategies are of particular interest. The first of these, which has not been dealt with previously in the literature, is a cancellation strategy whereby if the queue is not sufficiently long, the departure is cancelled and any waiting customers must wait an extra service period before they may leave. The instant at which the departure is cancelled is still counted as a dispatch instant even though no departure actually occurred.

Vehicle cancellations are used frequently in freight transportation since it reduces the number of vehicle departures, thereby saving operating costs. The second control policy considered is a vehicle holding strategy which has been studied previously by other authors under the name of a general bulk service rule. This policy assumes that if a queue is not long enough to justify sending the vehicle, then the vehicle is simply held until the queue reaches a particular length. The classical problem where the server becomes idle when the system is empty and immediately serves the next arriving customer is simply a special case of a holding strategy.

A large number of contributions have been made to the bulk queueing literature since the original paper, Bailey [1954], which looked at the problem where vehicle departures occur independent of the number of customers waiting. The first paper to explicitly investigate vehicle control strategies is Neuts [1967], which introduced the general bulk service rule where if the queue is less than some minimum m when the vehicle is ready to leave, then the departure is delayed until the queue meets or exceeds m. Other authors have since expanded on this queueing system, including Borthakur [1971], Medhi [1975, 1979], Medhi and Borthakur [1972], and Sim and Templeton [1981a, 1981b], who look at queues with negative exponential service times and both single and multiple servers using the general bulk service rule. Teghem _et al_. [1969] and Borthakur and Medhi [1974] look at the general bulk service rule for bulk arrival, bulk service queues. Deb and Serfozo [1973] shows that this strategy can be used to minimize either the total discounted or average operating costs of the queue, and demonstrate how to find the optimal m. They assume a general cost function for holding x customers per unit of time, and assume the cost of serving a batch of y customers is A + By, where A and B are given constants. Powell [1983a] considers both the general bulk service rule and a vehicle cancellation strategy (defined below), drawing off the theory presented in this paper, and compares these strategies in terms of operating costs and level of service. Powell [1983b] investigates the same problems using an iterative numerical algorithm which avoids the use of classical transform techniques.

The remainder of this paper will use the term holding strategy in place of general bulk service rule as it is more accurate and is actually a special case of a broader set of vehicle control strategies. Section 1 of the paper outlines the basic approach for describing the queueing process, making use of several results in Keilson [1979]. Section 2 looks at the nontrivial problem of obtaining numerical results from classical analyses of bulk service queues using transform techniques. A new approach to solving the transform is presented which is shown, in Section 3, to be computationally very fast and numerically stable.

1. Theoretical Background

The problem considered in this paper is a situation common in transportation systems where customers arrive in groups of random size and wait for a vehicle to arrive. On arrival, the vehicle normally loads all waiting customers up to its capacity and departs. In many instances, however, the operator will wish to control the departure of the vehicles to achieve some objective, such as minimizing costs, by delaying or cancelling departures to avoid uneconomically small loads. The purpose of this section is to present a general theoretical framework for modeling a broad range of vehicle dispatching strategies. The objective is to derive the transform of the distribution of the length of the queue in steady state at the time of a service completion.

The system can be modeled by the following equation:

$$Q^{n+1} = Q^n - V^n + E^n + Y^{n+1} \tag{1}$$

where

Q^n = number of customers waiting at the n^{th} dispatch instant

V^n = capacity of the n^{th} vehicle

Y^n = number of customers arriving during the service period prior to the n^{th} dispatch instant

E^n = a control variable, incorporating the effect of the control strategy in use as well as ensuring the nonnegativity of $Q^n - V^n + E^n$, which can be interpreted as the number of customers left over immediately after the n^{th} vehicle departs (or is cancelled)

A dispatch instant occurs immediately at the end of each service period. Depending on the control strategy, a vehicle may be dispatched, held or cancelled. If the vehicle is held, it may be eventually sent or cancelled. The term dispatch instant is used since vehicle departures do not always occur between successive service periods. A new service period begins immediately following each departure or cancellation.

The assumption is made that $\{Y^n\}$ and $\{V^n\}$ are sequences of independently and identically distributed random variables that are independent of Q^1. The vehicle size V^n is assumed to be independent of Q^n. In addition, let $C = \max \{i:P(V^n=i) > 0\}$. The importance of treating V^n as a random variable arises in several contexts in transportation applications. Most frequently, a vehicle makes several stops, and hence the available capacity at a given stop is random due to the presence of other customers who boarded the vehicle at an earlier stop. Alternatively, variations in vehicle types may make the actual capacity random (a related example occurs in rail freight transportation, where uncertainty in the availability of locomotives creates uncertainty in the maximum number of cars a train may pull). A last example in freight transportation is that variability in the size of shipments to be loaded onto a truck may be treated as variability in the number of shipments that will fit on a truck.

The control variable E^n is determined by Q^n and V^n and is assumed to be conditionally independent of any past history, given Q^n. It is also assumed that $E^n = 0$ if $Q^n \geq V^n$ and that there exists an integer K (determined by the control strategy being used) such that $E^n = 0$ whenever K or more customers remain in the system at the beginning of the nth service period. Examples of how E^n is defined for different control strategies are the following:

No control - This policy, first introduced in Bailey [1954] represents a strategy where vehicle departures occur regardless of the length of the queue. For this case, $K = 1$ and

$$E^n = \max\{0, V^n - Q^n\} \tag{2}$$

Vehicle cancellations - A vehicle cancellation policy is defined as one where, if $Q^n < K$, the n^{th} departure is cancelled and any waiting customers, up to $K - 1$, must wait an additional service period for the next departure. In this context, K is referred to as the minimum load constraint (which must be specified) and

$$E^n = \begin{cases} V^n & \text{if } Q^n < K \\ \max\{0, V^n - Q^n\} & \text{if } Q^n \geq K \end{cases} \tag{3}$$

An interesting variant of this policy when V^n is random is where the cancellation decision is based on the number of empty spaces on a vehicle.

Vehicle holding strategies - The vehicle holding strategy was first introduced by Neuts [1967] and later extended by several authors (e.g. Teghem et al. [1969]). In its simplest form a vehicle will depart at the end of the $n - 1^{st}$ service period only if $Q^n \geq M$, where M is the minimum load constraint. If $Q^n < M$, then the vehicle is simply held until the number of customers is at least M. Let Y_I^n be the number of arrivals that occur while the vehicle is being held, where Y_I^n depends on both Q^n, which is the length of the queue at the beginning of the idle period, and M. After some manipulation it can be verified that

$$E^n = \begin{cases} Y_I^n & Q^n < M,\ Q^n + Y_I^n \geq V^n \\ V^n - Q^n & Q^n < M,\ Q^n + Y_I^n < V^n \\ V^n - Q^n & M \leq Q^n < V^n \\ 0 & Q^n \geq V^n \end{cases} \tag{4}$$

Note that if arrivals occur singly, then $Y_I^n = \max\{0, M-Q^n\}$, and the case reduces to that of the queue with no control (K = 1). Assume now that arrivals occur in groups of maximum size N, and let $U = \min\{i : P[V^n = i] > 0\}$. Then the largest number of customers that can be left over when $Q^n < M$ is $\max\{0, M + N - 1 - U\}$ and hence $K = \max\{1, M + N - U\}$. Note in this case we must require that $M \leq U$ to avoid the possibility of a vehicle cancellation when the number of customers actually meets or exceeds the capacity of the vehicle.

Combined vehicle holding and cancellation strategies - If at the end of a service period $Q^n < M$, the vehicle is held until one of two conditions is satisfied: a) if the number of new arrivals, Y_I^n, brings the length of the queue above M, and if these arrivals occur within time T, then the vehicle is dispatched; b) if after time T the total length of the queue is less than M, the vehicle is cancelled. The variable Y_I^n is defined to be the number of new arrivals up to the instant where $Q^n + Y_I^n \geq M$ or up to time T, whichever occurs first. Then:

$$E^n = \begin{cases} Y_I^n & Q^n < M,\ Q^n + Y_I^n \geq V^n \\ V^n - Q^n & Q^n < M,\ M \leq Q^n + Y_I^n < V^n \\ V^n + Y_I^n & Q^n < M,\ Q^n + Y_I^n < M \\ \max\{0,\ V^n - Q^n\} & Q^n \geq M \end{cases} \tag{5}$$

Defining U and N as above, the factor K is now given by $K = \max[M, M + N - U]$.

Using the assumptions on the sequences $\{Y^n\}$, $\{V^n\}$ and $\{E^n\}$ it is easily shown that $\{Q^n\}$ forms a Markov chain. The proof of this result follows simply from the assumptions that $\{Y^n\}$ and $\{V^n\}$ are i.i.d. sequences and that E^n, given Q^n, is conditionally independent of any previous history. In addition, the assumption is also made throughout the paper that the chain is irreducible, a condition that depends on the statistics of Y^n, V^n and E^n and must therefore be proved on a case by case basis. If the assumption is made, for example, that $P[Y^n = i] > 0$, $i = 0,1,2,\ldots$, then it is trivial to show that the chains for the cases described above are irreducible. Finally, to demonstrate that the process $\{Q^n\}$ is ergodic we state the following:

Theorem 1: If the Markov chain $\{Q^n\}$ is irreducible and aperiodic and if

$$E[Y^n] < E[V^n] \tag{6}$$

then the Markov chain is ergodic.

Proof: Let $\gamma_i = E[Q^{n+1} - Q^n | Q^n = i]$. According to theorem 2 in Pakes [1969] the Markov chain is ergodic if a) $|\gamma_i| < \infty$ for all i and b) $\limsup_{i\to\infty} \gamma_i < 0$. Conditioning on the event $Q^n = i$ and taking expectations of (1) gives

$$\gamma_i = E[Q^{n+1} - Q^n | Q^n = i] = E[Y^n] + E[E^n | Q^n = i] - E[V^n] \tag{7}$$

All the terms on the right hand side of (7) are bounded, and hence condition a is satisfied. To demonstrate condition b, we note that we require $E^n = 0$ if $Q^n \geq V^n$. Thus for all $i \geq C$, $\gamma_i = E[Y^n] - E[V^n]$, and, if (6) is satisfied, then $\gamma_i < 0$ for $i \geq C$. $\square$

Let $Q^n(z) = \sum_{i=0}^{\infty} q_i^n z^i$, where $q_i^n = \text{Prob}[Q^n = i]$, and define $Q(z) = \lim_{n\to\infty} Q^n(z)$, where the existence of $Q(z)$ is guaranteed by theorem 1. To find $Q(z)$ we use the following theorem adapted from Keilson [1979, p. 50]:

Theorem 2: Q(z) is given by

$$Q(z) = \frac{\Psi(z)\, Y(z)}{1 - Y(z)V(1/z)} \tag{8}$$

where $Y(z)$ and $V(z)$ are the z-transforms of the variables Y^n and V^n (the superscript n is dropped since the sequences $\{Y^n\}$ and $\{V^n\}$ are i.i.d.). $\Psi(z) = \sum_{i=-C}^{K-1} \Psi_i z^i$, where the elements Ψ_i, $i = -C,\ldots,K-1$, are described below.

Proof: Taking transforms of both sides of (10 gives

$$Q^{n+1}(z) = [Q^n(z)V^n(1/z) + \Psi^n(z)]Y^n(z) \tag{9}$$

where $\Psi^n(z)$ is the transform of the distribution of $Q^n - V^n + E^n$ minus the transform of the distribution of $Q^n - V^n$ (note as a consequence that $\Psi^n(1) = 0$). Taking the limit as $n \to \infty$, putting $\Psi(z) = \lim_{n\to\infty} \Psi_n(z)$, and solving for $Q(z)$ gives (8). $\square$

The function $\Psi(z)$ plays a special role in the analysis. Let $\underline{\Psi}$ be a column vector with elements $(\Psi_{-C}, \Psi_{-C+1}, \ldots, \Psi_{K-1})$ and let q be a column vector with elements $(q_0, q_1, \ldots, q_{C-1})$ where $q_i = P[Q = i]$. The vector $\underline{\Psi}$ can be calculated using

$$\underline{\Psi} = X \cdot \underline{q} \tag{10}$$

where the matrix X is given by

$$X = \begin{bmatrix} x_{-C,0} & \cdots & x_{-C,\, C-1} \\ \vdots & & \vdots \\ x_{K-1,0} & & x_{K-1,C-1} \end{bmatrix} \tag{11}$$

The entries of the matrix X are given by

$$x_{i,j} = [\text{Prob}(V^n - E^n = j-i, Q^n = j) - \text{Prob}(V^n = j-i, Q^n = j)]/\text{Prob}[Q^n = j] \quad \text{(12a)}$$

$$= \text{Prob}[V^n - E^n = j-i \mid Q^n = j] - \text{Prob}[V^n = j-i] \qquad \begin{matrix} 0 \le j \le C-1 \\ -C \le i \le K-1 \end{matrix} \quad \text{(12b)}$$

In Kielson's parlance $\underline{\Psi}$ is a compensating measure while X is the difference between the probability transition matrix induced by $Q \to Q - V + E$ and the one corresponding to the underlying random walk $Q \to Q - V$ without boundary or control. The dimensions of X result from the assumption that $E^n = 0$ whenever $Q^n - V^n + E^n \ge K$ and the constraint that $V^n \le C$.

The column sums of X are zero. Letting $v_i = P[V^n = i]$, the constraints on E^n imply that X may be written

$$X = \left[\begin{array}{cccc} -v_C & 0 & \cdots & \\ -v_{C-1} & -v_C & \cdots & \\ & \vdots & & \\ -v_1 & -v_2 & \cdots & -v_C \\ \hline x_{o,o} & x_{o,1} & \cdots & x_{o,C-1} \\ & \vdots & & \\ x_{K-1,o} & x_{K-1,1} & \cdots & x_{K-1,C-1} \end{array}\right] = \left[\begin{array}{c} X_1 \\ \\ X_2 \end{array}\right] \qquad (13)$$

where X_1 is a (C,C) matrix comprising the first C rows of X and X_2 is a (K,C) matrix comprising the last K rows. Note that X_1 is lower triangular, Toeplitz and, with the condition that $v_C > 0$, invertible.

The elements $x_{i,j}$ are determined by the control strategy being used. The vector $\underline{q}$, however, is unknown. The usual approach to finding $\underline{q}$ is by first observing that if Y(z) is bounded on the circle $|z| = 1 + \delta$, for some $\delta > 0$, and assuming (6) holds, then the denominator of (8) must have C zeroes inside and on the unit circle. The proof of this involves a classical application of Rouché's theorem. Using these zeroes, the vector $\underline{q}$ can be found by setting up a system of linear equations that forces $\Psi(z)$ to have zeroes matching those of the denominator of (8) that are in the unit circle.

This approach to eliminating the vector of unknown $\underline{q}$ is a standard one used by almost every paper on bulk service queues (see, for example, Ohno [1978]). While this is a satisfactory approach for small problems, it can be numerically hazardous for large problems (C greater than 50) as a result of the effects of computer roundoff error. The difficulties of obtaining numerical solutions are covered in greater detail in section 3.

The next section presents an alternative approach for solving for the remaining unknowns that is numerically more stable as well as being computationally much more efficient.

2. An Efficient Solution Procedure

A numerically stable and computationally efficient procedure for solving for the remaining unknowns can be developed by taking advantage of the fact that for many problems the parameter K is much smaller than C. We now present an approach that leads to solving only a system of K linear equations.

Let z_i, $i = 0,1,\ldots,C + K - 1$ be the zeroes of $\Psi(z)$ and assume that $z_0,\ldots,z_C$, are the zeroes of the denominator of (8). We first observe that we may write $\Psi(z) = A(z)B(z)$ where

$$A(z) = z^{-C}(E[V] - E[Y])\ (z - 1) \prod_{i=1}^{C-1} \left(\frac{z - z_i}{1 - z_i}\right) = \sum_{i=-C}^{0} a_i z^i \tag{14}$$

and

$$B(z) = \sum_{i=0}^{K-1} b_i z^i \tag{15}$$

Note that A(z) is defined to contain the C zeroes of the denominator within the unit circle, and hence is not affected by the specific control strategy being used. In fact, the coefficients $a_{-C}, a_{-C+1},\ldots, a_0$ can be obtained directly from the zeroes by simply expanding the polynomial on the right hand side of (14). The process of obtaining the coefficients is described in further detail in section 3. For the moment, however, we will assume that $A(z)$ is known. $B(z)$, on the other hand, is a polynomial of degree $K-1$ with zeroes not directly related to those of $A(z)$, and depends on the control strategy. Given the way $A(z)$ has been normalized, it is easy to verify $B(1) = 1$.

We now introduce the matrix A, made up from the coefficients of the polynomial $A(z)$, which is given by

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \tag{16}$$

where A_1 and A_2 are (C,K) and (K,K) matrices, respectively, given by

$$A_1 = \begin{bmatrix} a_{-C} & 0 & 0 & \cdots \\ a_{1-C} & a_{-C} & 0 & \\ & \vdots & & \\ a_{K-1-C} & a_{K-2-C} & \cdots & a_{-C} \\ & \vdots & & \\ a_{-1} & a_{-2} & & a_{-K} \end{bmatrix} \tag{16a}$$

$$A_2 = \begin{bmatrix} a_0 & a_{-1} & \cdots & & a_{-K+1} \\ 0 & a_0 & a_{-1} & \cdots & a_{-K+2} \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & a_0 \end{bmatrix} \tag{16b}$$

Since $\Psi(z) = A(z)B(z)$, we find that

$$\underline{\Psi} = A\underline{b} \tag{17}$$

where $\underline{b} = (b_0,b_1,\ldots,b_{K-1})^T$. Note that A_2 is triangular, Toeplitz and invertible.

Having structured the problem in this way, the principal result can be obtained directly. For ease of reference, it is summarized in the following theorem.

Theorem 3. The unknown vectors b and q can be determined by solving the following system of linear equations:

$$[A_2 - X_2 X_1^{-1} A_1] \underline{b} = 0 \tag{18}$$

combined with

$$\underline{e}^T \underline{b} = 1 \tag{19}$$

where $\underline{e} = (1, 1, \ldots, 1)^T$ is a vector of dimension K. Once $\underline{b}$ is determined, we may find $\underline{q}$ using

$$\underline{q} = X_1^{-1} A_1 \underline{b} \tag{20}$$

Proof: Equations (10) and (17) imply $X\underline{q} = A\underline{b}$, or $X_1\underline{q} = A_1 b$ and $X_2\underline{q} = A_2\underline{b}$, from which (18) and (20) follow directly. Equation 19 results from the observation that $\lim_{z\to 1} B(z) = 1$. Note that one of the equations in (18) is redundant and can be replaced by (19), leaving us a system of K linear equations for $\underline{b}$. $\square$

The importance of theorem 3 is that we can take advantage of the fact that K is typically much smaller than C, and hence we may solve a much smaller system of linear equations. The result is significantly faster execution times and reduced computer roundoff error. An important feature of the calculations in theorem 3 is that X_1 is lower triangular and Toeplitz and hence can be inverted extremely easily. In fact, X_1^{-1}, can often be computed in closed form, as it will also be lower triangular and Toeplitz.

In view of the definitions of A(z) and B(z) and equation 19, it seems plausible to deduce that B(z) is the transform of a probability distribution. Proof of this conjecture first requires the following result:

Proposition: The elements of the matrix A_2^{-1} are nonnegative.

Proof: The coefficients $\{a_i\}$ are determined by the zeroes of (9) and hence are not affected by the control strategy being used. For the queue with no control, $B(z) = 1$ and thus $\Psi(z) = A(z)$ or, equivalently, $\Psi_i = a_i$. Combining (10) and (13), with the observation that $q_i \geq 0$, implies that $a_i \leq 0$ for $i = -C, \ldots, -1$, since the first C rows of X are nonnegative. Finally, since $A(z) \neq 0$ for some z and $A(1) = 0$, a_o, which makes up the diagonal of A_2, must be strictly positive. Under these conditions, it is straightforward to show that A_2^{-1} is nonnegative. □

With this proposition the following theorem follows easily.

Theorem 4: If $E^n=0$ when $Q^n \geq V^n$, B(z) is the transform of a probability mass function

Proof: We already know that $B(1)=1$, hence we need only show that the vector of coefficients $\underline{b}$ is nonnegative. First define

$$\underline{\Psi}_2 = X_2\underline{q} \tag{21}$$

where $\underline{\Psi}_2 = (\Psi_o, \Psi_1, \ldots, \Psi_{K-1})$. We know that $\underline{b} = A_2^{-1}\,\underline{\Psi}_2$ and that A_2^{-1} is nonnegative, so it is sufficient to show that $\underline{\Psi}_2$ is nonnegative or, equivalently, that X_2 is nonnegative.

If $E^n=0$ when $Q^n \geq V^n$, then the event $[V^n=j-i,\ Q^n=j]$, $i \geq 0$ implies $E^n=0$ and thus $[V^n-E^n=j-i,\ Q^n=j]$. Going back to the definition of $x_{i,j}$ in (12a), we find that $x_{i,j} \geq 0$ for $i \geq 0$, which proves that $X_2 \geq 0$. □

Having determined the vector $\underline{b}$, it is possible to determine the moments of Q directly, without having to solve for $\underline{q}$. First let $Q^{(K)}(z)$ be the transform of the p.m.f. of Q when a control strategy is being used, and let $Q^{(1)}(z)$ be the corresponding transform when no control is being used. From the definition of A(z), we may write

$$Q^{(1)}(z) = \frac{A(z)Y(z)}{1 - Y(z)V(\frac{1}{z})} \tag{22}$$

We may then write

$$Q^{(K)}(z) = Q^{(1)}(z)\ B(z) \tag{23}$$

where B(z) is a function of K and reflects the control strategy being used. Let $Q^{(K)}$ and $Q^{(1)}$ be, respectively, the random variables described the length of the queue at dispatch instants with and without a control strategy, respectively. Also let B be the random variable whose transform is B(z). Equation 23 implies that $Q^{(K)} = Q^{(1)} + B$, where $Q^{(1)}$ and B are independent random variables, allowing us to write the moments of $Q^{(K)}$ as the sums of the moments of $Q^{(1)}$ and B. The moments of B are easily calculated once the vector $\underline{b}$ is known. The first two moments of $Q^{(1)}$, after a considerable amount of algebraic manipulation, are given by:

$$\overline{Q}^{(1)} = \frac{\overline{\overline{V}} + \overline{\overline{Y}} + (\overline{V} - \overline{Y})(1 + 2(\overline{V} - C)) - (\overline{V} - \overline{Y})^2}{2(\overline{V} - \overline{Y})} + \sum_{i=1}^{C-1} \frac{1}{1-z_i} \tag{24}$$

$$\overline{\overline{Q}}^{(1)} = \frac{-4(\overline{\overline{\overline{V}}} - \overline{\overline{\overline{Y}}})(\overline{V} - \overline{Y}) + 3(\overline{\overline{V}} + \overline{\overline{Y}})^2 - [6(\overline{\overline{V}} - \overline{\overline{Y}}) - 1](\overline{V} - \overline{Y})^2 - (\overline{V} - \overline{Y})^4}{12(\overline{V} - \overline{Y})^2}$$

$$- \sum_{i=1}^{C-1} \frac{z_i}{(1-z_i)^2} \tag{25}$$

where, for example, $\bar{V}$, $\bar{\bar{V}}$ and $\bar{\bar{\bar{V}}}$ denotes the mean, variance and third moment about the mean of the random variable V.

3. Numerical Experiments

Several experiments were conducted to test the efficiency and stability of the procedures described in the first two sections. The basic problem used for the experiments assumed compound Poisson arrivals, where λ is the rate of arrivals of groups and where the size of each group, G, is described using a shifted geometric distribution given by

$$g_i = \text{Prob}\ [G = i] = (1 - r)r^{i-L} \qquad i = L, L+1, L+2, \ldots \tag{26}$$

where L is the shift parameter. L = 1 and r = .6 were used throughout. Vehicle sizes were assumed to be deterministic or random, where in the latter case two distributions were tested. The first is the shifted binomial distribution given by

$$P[V = k + M_1] = \binom{M_2}{k} \sigma^k (1 - \sigma)^{M_2 - k} \qquad k = 0, 1, \ldots, M_2 \tag{27}$$

and the second is the shifted discrete uniform distribution, given by

$$P[V = k + M_1] = \frac{1}{M_2} \qquad k = 1, 2, \ldots, M_2 \tag{28}$$

In both cases, M_1 is a shift parameter, M_2 governs the spread of the distribution, and the maximum vehicle capacity is $C = M_1 + M_2$. In (27), a value of $\sigma = .5$ was used throughout. The arrival rate λ was always fixed to produce a desired value of ρ. Most of the experiments assumed $\rho = .6$, but other values ranging from .01 to .99 were also tested.

The first problem is to calculate the vector of coefficients {a}, after which the vector $\underline{q}$ can be found using (20). The standard approach is to solve a system of C linear equations. A much faster and more stable method is to use a simple procedure which is termed here the polynomial expansion algorithm. The steps of this procedure, which is described in the appendix, uses the known roots $z_0, z_1, \ldots, z_{C-1}$ to successively build up the polynomial $A(z)$.

A series of experiments were conducted to test the efficiency of the polynomial expansion algorithm relative to that of the old approach which involved solving a set of C simultaneous linear equations. The results, shown in table 1, demonstrate that the polynomial expansion algorithm is significantly faster than the old approach, an advantage that improves as C is increased. In addition, the old approach proved to be increasingly unstable for values of C greater than 50, whereas the new approach worked well for C = 250, which was the largest vehicle size tested.

Once the vector $\{a\}$ is found, the queue length transform $Q(z)$ can be found (and hence all the moments of the distribution) without calculating the vector $\underline{q}$. If, however, these probabilities are required, then equation (20) must be used. This step is simplified since X_1 is lower triangular and invertible, making X_1^{-1} easy to calculate. In fact, depending on the distribution of V, it may be possible to calculate X_1^{-1} in closed form (the simplest case occurs if V=C, in which case X_1 is the identity matrix). Let $\hat{x}_i$ denote the first element in the i^{th} row of X_1^{-1}. If V follows a shifted binomial distribution as given by (27), then

$$\begin{aligned} \hat{x}_1 &= -\sigma^{-M_2} \\ \hat{x}_{i+1} &= -\frac{(M_2 + i - 1)}{i} \; \frac{1-\sigma}{\sigma} \; \hat{x}_i \qquad i = 1,2,\ldots,C-1 \end{aligned} \tag{29}$$

Since X_1^{-1} is Toeplitz and lower triangular, the first column determines the entire matrix. If V is described by a shifted discrete unform distribution, then

$$\hat{x}_i = \begin{cases} M_2 & i = 1, 1+M_2, 1+2M_2, \ldots \\ -M_2 & i = 2, 2+M_2, 2+2M_2, \ldots \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

In addition, X_1 was also inverted numerically, a step which required a negligible amount of time and appeared to be very stable.

If vehicle sizes were deterministic or described by a discrete uniform distribution, equation (20) reliably yielded $\underline{q}$ in every instance. On the other hand, if the vehicle size follows a binomial distribution, the calculation of $\underline{q}$ proved unreliable for large values of M_2 and for small values of σ. This behavior is not surprising since small σ and large M_2 produces extremely large numbers in the matrix X_1^{-1}, thereby magnifying even small errors in the vector $\{a\}$.

If no control strategy is in use (K=1), then the vector $\{a\}$ determines $Q(z)$. If a control strategy is in use (K>1), then theorem 3 can be used to find the vector $\underline{b}$. To test this strategy, a simple cancellation strategy is used where the last K rows of X are given by:

$$x_{i,j} = \begin{cases} 1 & \text{if } i = j, \quad i = 1,2,\ldots,K-1 \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

The process requires just finding the vector $\{a\}$, and then using theorem 3 to find $\underline{b}$; knowledge of $\underline{q}$ is not needed.

As before, a series of experiments were run. Assume V is either deterministic or described by a discrete uniform or binomial distribution. Different runs were conducted varying ρ and K (recall that we assume that K must be less than or equal to M_1). Each problem was solved using both the old approach, which requires solving a system of C linear equations, and the new approach. Table 2 shows the results when V is deterministic or described by a discrete uniform distribution. In all cases, the new approach proved to be both stable and extremely fast. The old approach was much slower and frequently unstable for large values of C. Table 3 shows the results when the vehicle size is described by a binomial distribution. Here, the new approach did demonstrate some instability for certain combinations of M_2 and K, suggesting that the maximum value K could take diminished as M_2 was increased.

A final experiment was run to shed some light on the accuracy of the new procedure. It is possible to show that if $V \equiv C$ and $K = C$ (the go-when-filled cancellation policy) then the vector $\underline{b}$ is simply the discrete uniform distribution between 0 and C-1, regardless of the other parameters of the problem. For this special case, the calculation of the vector $\underline{b}$ proved to be accurate in every instance to ten decimal places. Since these calculations depend on the vector $\{a\}$, this result suggest these numbers are also accurate to at least ten places.

Acknowledgements

The authors would like to gratefully acknowledge the persistent and painstaking editorial assistance of Dr. Daniel P. Heyman and several anonymous referees.

APPENDIX

The Polynomial Expansion Algorithm

The polynomial expansion algorithm is a simple procedure for finding the coefficients of the polynomial A(z) using the known roots of A(z). The procedure works for problems where the roots fall along a smooth contour which crosses the negative real axis once. Let $M = [\frac{C-1}{2}]$, where [x] is the largest integer less than or equal to x. Let z_i, i = 0, 1,...,M be the subset of roots of A(z) such that $Im(z_i) \geq 0$ and $Re(z_i) < Re(z_{i-1})$, i = 1,2,...,M. Let $z_0 = 1$, and let $\bar{z}_i$ be the complex conjugate of z_i. Now define the polynomial

$$\begin{aligned} P_i(z) &= (z - z_i)(z - \bar{z}_i) \\ &= |z_i|^2 - 2\,Re(z_i)z + z^2, \qquad i = 1,\ldots,M \end{aligned} \tag{A.1}$$

Let ϕ be a constant given by

$$\phi = (\bar{V} - \bar{Y}) \prod_{i=1}^{C-1} \frac{1}{1 - z_i} \tag{A.2}$$

calculated in a straightforward manner. Finally, define a series of partial polynomials $A^{(i)}(z)$, each of which contains some subset of the zeroes. The polynomials $A^{(1)}(z)$, $A^{(2)}(z)$,..., are now calculated recursively according to the following scheme. For a given constant I (discussed below):

Step 0: If C is odd, set $A^{(0)} = 1$. If C is even, set $A^{(0)} = z - z_{C/2}$, where $z_{C/2}$ is the root located on the negative real axis.
Set i = 0, n = 0

Step 1: Set i = i + 1, k = i

Step 2: Set $n = n + 1$ and find

$$A^{(n)}(z) = A^{(n-1)}(z)\ P_{M-k+1}(z)$$

Step 3: Set $k = k + I$; if $k \leq M$, go to step 2

Step 4: If $i < I$, go to step 1

Step 5: Set $A(z) = \phi \cdot z^{-C}\ (z-1) \cdot A^{(n)}(z)$

If $I = 1$, then the algorithm brings each pair of conjugate roots into the polynomial starting with the root with the most negative real part. Experiments showed that $I = 1$ did not work well for larger vehicle sizes (over 50) but values of I equal to 7, 8 or 9 worked extremely well for vehicles with capacity up to 250, which was the largest vehicle examined.

References

Bailey, N. T. J. 1954. On Queueing Processes with Bulk Service. J. Royal Stat. Soc. B. Vol. 16, pp. 80-87.

Borthakur, A. 1971. A Poisson Queue with a General Bulk Service Rule. J. Assam. Sc. Soc. Vol. XIV, No. 2, pp. 162-167.

Borthakur, A. and J. Medhi. 1974. A Queueing System with Arrival and Service in Batches of Variable Size. Cahiers du Centre d'Etudes de Recherche Opérationelle. Vol. 16, pp. 117-126.

Deb, R. K. and R. F. Serfozo. 1973. Optimal Control of Batch Service Queues. Adv. Appl. Prob. Vol. 5, pp. 340-361.

Keilson, J. 1979. Markov Chain Models - Rarity and Exponentiality. Springer-Verlag, New York.

Medhi, J. 1975. Waiting Time Distribution in a Poisson Queue with a General Bulk Service Rule. Mgmt. Sci. Vol. 21, No. 7, pp. 777-782.

Medhi, J. 1979. Further Results on Waiting Time Distribution in a Poisson Queue Under a General Bulk Service Rule. Cahiers du Centre d'Etudes de Recherche Opérationelle. Vol. 21, No. 2, pp. 183-189.

Medhi, J. and A. Borthakur. 1972. On a Two Server Markovian Queue with a General Bulk Service Rule. Cahiers du Centre d'Etudes de Recherche Opèrationelle. Vol. 14, pp. 151-158.

Neuts, M. F. 1967. A General Class of Bulk Queues with Poisson Input. Ann. Math. Stat. Vol. 38, pp. 759-770.

Ohno, K. 1978. Computational Algorithm for a Fixed Cycle Traffic Light and New Approximate Expressions for Average Delay. Trans. Sci. Vol. 12, No. 2, pp. 29-47.

Pakes, A. G. 1969. Some Conditions for Ergodicity and Recurrence of Markov Chains. Opns. Res., 17, pp. 1058-1061.

Powell, W. B. 1983a. Analysis of Vehicle Holding and Cancellation Strategies in Bulk Arrival, Bulk Service Queues. Trans. Sci. (to appear).

Powell, W. B. 1983b. Iterative Algorithms for Bulk Arrival, Bulk Service Queues with Poisson and Non-Poisson Arrivals. Trans. Sci. (to appear).

Sim, S. H. and J. G. C. Templeton. 1981. A Multi-Server Queueing System with General Bulk Service Rules. University of Toronto, Department of Industrial Engineering. Working Paper 81-012.

Sim. S. H. and J. G. C. Templeton. 1981. Further Results for the M/M(a,b)/C Queueing System. University of Toronto, Department of Industrial Engineering. Working Paper 81-011.

Teghem, J., J. Lois-Teghem and J. P. Lambotte. 1969. Modèles d'attente M/G/1 et GI/M/1 à arrivées et services en groups. Lecture Notes in Operations Research and Mathematical Economics, 8. Springer-Verlag, New York.

Table 1

Comparison of Simultaneous Equations and
Polynomial Expansion Methods for
Finding $\underline{q}$*

Deterministic vehicle size
No control policy

Vehicle Capacity	To find Roots	To find $\underline{q}$: Simultaneous Equations	To find $\underline{q}$: Polynomial Expansion
50	.024 secs.	.293 secs.	.006 secs.
100	.044	3.479#	.017
150	.067	11.996#	.039
250	.086	DNR**	.068

* All times are in CPU seconds measured using an internal clock which excluded any I/O. The program was compiled using the IBM Fortran H compiler and run on an IBM 3081. The system of equations was solved using the IMSL routine LEQT2F. All calculations were performed using double precision arithmetic.

\# Method might or might not produce a valid solution depending on the problem.

** DNR-did not run. Method produced negative probabilities.

Table 2

Comparison of Methods for Finding
the Moments of the Length
of the Queue

Vehicles operated under a cancellation policy with minimum load m

Vehicle size C	Vehicle size M_2	Minimum load K	Execution Times* OLD**	Execution Times* NEW**
Deterministic batch sizes				
100	0	50	6.672#	.122
100	0	100	DNR***	.764
150	0	20	11.283#	.049
150	0	50	15.239#	.392
Shifted discrete uniform distribution for batch sizes				
150	100	20	9.122#	.050
150	100	50	13.511#	.146
150	50	50	16.340#	.415
150	50	100	DNR	.774

* Execution times exclude time required to find roots

** OLD uses system of simultaneous equations to find $\underline{q}$; NEW uses polynomial expansion method to find $\{a_i\}$, and then applies theorem 1 to find $\underline{b}$.

*** DNR-did not run. Method produced negative probabilities.

For some problems (e.g. different values of ρ) method would not produce valid results.

Table 3

Comparison of Methods for Finding the Moments of the Length of the Queue

Vehicles operated under a cancellation policy with minimum load K

Vehicle Size C	M_2	Minimum load K	Execution times OLD	Execution times NEW
Shifted binomial distribution for batch sizes				
Group 1				
30	10	20	.023	.013
40	10	20	.401	.016
50	10	20	.617	.016
60	10	20	DNR	.023
100	10	20	DNR	.035
200	10	20	DNR	.090
250	10	20	DNR	.126
Group 2				
50	20	20	DNR	.021
100	20	20	DNR	.038
150	20	20	DNR	.061
200	20	20	DNR	.089
250	20	20	DNR	.119
Group 3				
100	10	20	DNR	.038
100	10	50	DNR	.145
100	10	90	DNR	.615
100	20	20	DNR	.039
100	20	30	DNR	.060
100	20	40	DNR	.089
100	20	50	DNR	DNR
100	30	10	DNR	.027
100	30	20	DNR	.036
100	30	30	DNR	.058
100	30	40	DNR	DNR
100	40	10	DNR	.024
100	40	20	DNR	.035
100	40	30	DNR	DNR