Batch Markovian Arrival Processes (BMAP)

James D. Cordeiro
Directorate of Force Management Policy
Headquarters, United States Air Force
1235 South Clark Street, Suite 301
Arlington, VA 22202 USA
Ph: (703) 604-0729; Fax: (703) 604-1657
Email: james.cordeiro@pentagon.af.mil

Jeffrey P. Kharoufeh
Department of Industrial Engineering
University of Pittsburgh
1048 Benedum Hall
3700 O’Hara Street
Pittsburgh, PA 15261 USA
Ph: (412) 624-9832; Fax: (412) 624-9831
Email: jkharouf@pitt.edu

Abstract

This article describes the batch Markovian arrival process (BMAP), a point process that is characterized by Markov-modulated batch arrivals of random size. The BMAP is a generalization of many well-known processes including the Markovian arrival process (MAP), the Poisson process, and the Markov-modulated Poisson process. It provides a common framework for modeling arrival processes in a variety of applications. We formally define the continuous- and discrete-time BMAP, review a few basic results for each, and show how these processes generalize many common point processes. Additionally, we provide suggestions for further reading on the subject.

Keywords: Batch Markovian arrival process, Markovian arrival process, batch arrivals

Introduction

The batch Markovian arrival process (BMAP) is a stochastic point process that generalizes the standard Poisson process (and other point processes) by allowing for “batches” of arrivals, dependent inter-arrival times, nonexponential inter-arrival time distributions, and correlated batch sizes. The Markovian arrival process (MAP) is a special case of the BMAP in which the batch size is restricted to unity. For a detailed description of the MAP, please see the article "Markovian Arrival Processes (MAP)."

The origins of the BMAP can be traced to the development of the versatile Markovian point process (VMPP) by Marcel Neuts [1] whose primary objective was to extend the standard Poisson process to account for more complex customer arrival processes in queueing models. The VMPP is characterized by three different classes of batch arrivals, each of which are determined according to the type of transition that takes place in an external Markov process with $m$ transient states and a single absorbing state $m + 1$. One type of arrival is a Markov-modulated Poisson process, which occurs during the sojourn of the Markov process in any of the $m$ transient states. Another occurs when the Markov process in state $i$ transitions to state $j$, $j \neq i$, which is a normal transition between two transient states. The third type of arrival occurs when the process in transient state $i$
transitions to the absorbing state $m + 1$ and then restarts in state $j$; this type of transition is called an $(i, j)$-renewal transition, and, by virtue of restarting the Markov process, admits the possibility of a “self-transition” from a transient state $i$ to itself. From this description, it is clear that the VMPP is founded upon the notion of a phase-type (PH) distribution, or the distribution of the time to absorption of an absorbing Markov process. Neuts [2] played a major role in advancing the use of the PH-distribution in queueing theory, culminating ultimately in the development of the VMPP.

Lucantoni, et al., [3] sought to extend the original definition of the VMPP while simultaneously easing its notational burden by defining the Markovian arrival process (MAP). The MAP also leverages the concept of arrival dependence upon an external Markov process but does not distinguish between classes of arrivals. This was generalized to the BMAP in [4] by permitting batch arrivals. Originally it was thought that the VMPP was a special case of the BMAP, and it was only later that Lucantoni and others [5, 6] asserted the equivalence of the VMPP and BMAP. The term “BMAP” has persisted due to its widespread acceptance in the stochastic modeling community.

The analysis of queueing systems is assisted by the fact that they may often be modeled, either directly or via embedding, as structured Markov chains. Structured Markov chains classified as two main types: the $GI/M/1$- and $M/G/1$-types, with a well-known third type, the quasi-birth-and-death (QBD) process, formed as the juxtaposition of the other two. Structured Markov chains help facilitate the use of matrix-analytic methods in the steady-state analysis of queueing systems with MAP and BMAP input. The use of matrix-analytic methods in the analysis of queueing systems is detailed in Neuts’ two classic texts [7, 8], which describe the theory and method underlying the derivation of the stationary distributions of the structured Markov chains. The first queueing model to be considered is the single-server model with infinite capacity. Ramaswami [9] incorporated the BMAP (or VMPP), which he called the $N$-process in honor of Neuts, as an arrival process to a single-server queue with generally-distributed service times. From this work, a generalization of the Polleczek-Khinchin formula to the $N/G/1$ queue was derived. Basic results for the steady-state analysis of the $MAP/G/1$ queue are provided in [3]. The $BMAP/G/1$ is subsequently considered in [4] while the first known transient analysis of the $BMAP/G/1$ queue is presented in [5]. Various aspects of the $BMAP/G/1$ continue to be studied, as are queueing variants such as the $D-BMAP/G/1$ and the BMAP retrial queue. See “Further Reading” for references that pertain to these subjects.

In the sections that follow, we formally define the continuous-time BMAP and provide some basic results including the generating function of its counting process and its fundamental rate. We likewise define the discrete-time BMAP (D-BMAP) and describe a variety of arrival processes that are special cases of the BMAP and D-BMAP. Finally, we will provide suggestions for further reading on the subject for the interest reader to gain a deeper understanding, and appreciation for, these versatile arrival processes.
The Continuous-Time BMAP

Let $J \equiv \{J(t) : t \geq 0\}$ be an irreducible, continuous-time Markov chain (CTMC) with state space $E = \{1, 2, \ldots, m\}$, where $m$ is a finite, positive integer. The infinitesimal generator matrix of this CTMC is denoted by $Q$. Suppose $J$ has just entered state $i \in E$. The process spends an exponentially distributed amount of time in state $i$ with rate $\lambda_i = -q_{ii}$ where $q_{ii}$ is the $i$th diagonal element of $Q$. The transition that follows this sojourn can be one of two types. For the first type, an “arrival” of batch size $k$ ($k \geq 1$) occurs and the process transitions to state $j \in E$ with probability $p_{ij}(k)$, where $j$ may be equal to $i$. For the second type, the batch size is 0 and the process transitions to state $j \neq i$ with probability $p_{ij}(0)$. For each $i \in E$, the probabilities $p_{ij}(k)$ satisfy

$$\sum_{k=1}^{\infty} \sum_{j=1}^{m} p_{ij}(k) + \sum_{j \in E \setminus \{i\}} p_{ij}(0) = 1. \quad (1)$$

Next, for $k \geq 0$, define the matrices $D_k = [d_{ij}(k)]_{i,j \in E}$, where

$$d_{ij}(0) = \begin{cases} -\lambda_i, & j=i, \\ \lambda_i p_{ij}(0), & j \neq i, \end{cases} \quad (2)$$

and

$$d_{ij}(k) = \lambda_i p_{ij}(k), \quad i,j \in E, \ k \geq 1. \quad (3)$$

The matrix $D_0$ contains the transition rates of $J$ for which no arrivals occur, and the matrices $\{D_k : k \geq 1\}$ contain the transition rates for which a batch size $k$ occurs. Assuming $D_0$ is a stable matrix (i.e., it is nonsingular), then the interarrival times will be finite almost surely, which is equivalent to stating that the BMAP will not terminate. From condition (1) and eqs. (2) and (3), it is not hard to see that

$$Q = \sum_{k=0}^{\infty} D_k.$$ 

Now, let $N(t)$ denote the total number of arrivals up to time $t$. The joint process, $(N, J) \equiv \{(N(t), J(t)) : t \geq 0\}$, is called a Batch Markovian Arrival Process (BMAP). Obviously, it is a Markov process with state space $\{(n, j) : n \geq 0, j \in E\}$ and infinitesimal generator matrix

$$Q^* = \begin{bmatrix} D_0 & D_1 & D_2 & D_3 & \ldots \\ 0 & D_0 & D_1 & D_2 & \ldots \\ 0 & 0 & D_0 & D_1 & \ldots \\ 0 & 0 & 0 & D_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$ 

In the context of a BMAP, $\{J(t) : t \geq 0\}$ is normally called the phase process and $\{N(t) : t \geq 0\}$ is the counting process. The matrices, $\{D_k : k \geq 0\}$, are said to form a representation of the BMAP, i.e., the BMAP is completely specified by these matrices.
Let us now consider the joint probability distribution of \((N(t), J(t))\) via its (probability) generating function. Adopting the notation of Lucantoni [5], denote the transition functions of \((N, J)\) by

\[ P_{ij}(n, t) = \mathbb{P}(N(t) = n, J(t) = j \mid N(0) = 0, J(0) = i), \]

and define the matrix \( P(n, t) = [P_{ij}(n, t)]_{i,j \in E} \). Then, for each \( n \geq 0 \) and \( t \geq 0 \), \( P(n, t) \) satisfies the Chapman-Kolmogorov equations

\[ \frac{d}{dt} P(n, t) = \sum_{r=0}^{n} P(r, t) D_{n-r}, \quad (4) \]

\[ P(0, 0) = I, \]

where \( I \) is the identity matrix of order \( m \). Define the matrix generating function of \( P(n, t) \) by

\[ P^*(z, t) = \sum_{n=0}^{\infty} P(n, t) z^n, \quad |z| \leq 1, \quad t \geq 0. \quad (5) \]

Differentiating both sides of (5) with respect too \( t \), substituting (4) and summing shows that

\[ \frac{d}{dt} P^*(z, t) = P^*(z, t) D(z), \quad t \geq 0, \quad (6) \]

\[ P^*(z, 0) = I, \]

where

\[ D(z) \equiv \sum_{k=0}^{\infty} D_k z^k, \quad |z| \leq 1. \quad (7) \]

The (ordinary) matrix differential equation (6) has the obvious solution

\[ P^*(z, t) = P^*(z, 0) \exp(D(z) t) = \exp(D(z) t), \quad |z| \leq 1, \quad t \geq 0, \]

where \( \exp(A) \) is the matrix exponential of a square matrix \( A \) defined by

\[ \exp(A) = \sum_{i=0}^{\infty} \frac{A^i}{i!}. \]

We pause here to note the similarity between the generating function of the BMAP and that of a standard Poisson process which is given by the scalar function

\[ P^*(z, t) = \exp(-\lambda + \lambda z)t. \]

For the BMAP, the exponential term \(-\lambda + \lambda z\) is replaced by the matrix \( D(z) \) to account for batch sizes larger than unity.

Using the generating function of \( P(n, t) \), one can obtain the (conditional) expectation of the number of arrivals in the interval \((0, t]\). Define this conditional expectation by \( \mathbb{E}_i(N(t) \mathbf{1}(J(t) = j)) \) where \( \mathbf{1}(B) \) denotes the indicator variable of event \( B \) and \( \mathbb{E}_i \) denotes expectation with respect to
the probability law of \((N, J)\) given \(J(0) = i\) and \(N(0) = 0\). Then \(E_i(N(t)1(J(t) = j))\) is the \((i,j)\)th entry of the \(m \times m\) matrix
\[
\left. \frac{d}{dz} P^*(z, t) \right|_{z=1} = D(1) \exp(D(1) t) = Q \exp(Q t).
\]
The conditional \(k\)th factorial moment can be obtained by taking the \(k\)th-order derivative \(P^*(z, t)\) and evaluating at \(z = 1\) in the usual way.

The limiting behavior of the continuous-time BMAP is discussed next. Let \(\pi = [\pi_1, \ldots, \pi_m]\) be the invariant probability vector of the CTMC, \(\{J(t) : t \geq 0\}\), with generator matrix \(Q\); that is, \(\pi\) is the unique positive solution to the system of equations
\[
\pi Q = 0 \quad \text{and} \quad \pi e = 1
\]
where 0 is the zero (row) vector and \(e\) is a (column) vector of ones. Then the fundamental rate, or the stationary rate of arrivals in a BMAP, is given by
\[
\lambda = \pi \left( \sum_{k=1}^{\infty} k D_k \right) e. \tag{8}
\]
On the other hand, the arrival rate of batches is given by
\[
\lambda_g = -\pi D_0 e,
\]
which is never zero since \(D_0\) is assumed to be nonsingular. If all the batch sizes are equal to unity, then the process is a Markovian arrival process (MAP), and \(\lambda = \lambda_g\).

In the next section, we describe several arrival processes that are special cases of the BMAP. A working knowledge of phase-type (PH) distributions is assumed for this discussion. For a thorough treatment of continuous- and discrete-time PH-distributions, the reader should consult [1, 6, 7, 10]. A cogent summary of PH-distributions is also provided in the article *Phase-Type (PH) Distributions*.

**Common Continuous-Time BMAPs**

1. **Poisson Process**: If the state space \(E\) consists of only a single state (i.e., \(m = 1\)), the time between “transitions” is exponentially distributed with rate \(\lambda\), and an arrival of batch size 1 occurs at each transition, then the counting process, \(\{N(t) : t \geq 0\}\), is a Poisson process with rate \(\lambda\). In this case, the matrices \(\{D_k : k \geq 0\}\) are replaced by scalars \(\{D_k : k \geq 0\}\).

Specifically, \(D_0 = -\lambda\), \(D_1 = \lambda\), and \(D_k = 0\) for all \(k \geq 2\). Then, the counting process \(\{N(t) : t \geq 0\}\) is a BMAP with generator matrix
\[
Q^* = \begin{bmatrix}
-\lambda & \lambda & 0 & 0 & 0 & 0 & \ldots \\
0 & -\lambda & \lambda & 0 & 0 & 0 & \ldots \\
0 & 0 & -\lambda & \lambda & 0 & 0 & \ldots \\
0 & 0 & 0 & -\lambda & \lambda & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
\]
2. **Batch Poisson Process**: If we allow a batch size greater than unity in the standard Poisson process with rate \( \lambda \), the resulting *batch Poisson process* is a BMAP. Let \( p_k \) denote the probability that an arrival is of batch size \( k \), \( k \geq 1 \), and note that \( \sum_{k \geq 1} p_k = 1 \). For this process, \( m = 1 \), \( D_0 = -\lambda \), and \( D_k = \lambda p_k \) for each \( k \geq 1 \). Then batch Poisson Process is a BMAP with generator matrix

\[
Q^* = \begin{bmatrix}
-\lambda & p_1 \lambda & p_2 \lambda & p_3 \lambda & p_4 \lambda & p_5 \lambda & \ldots \\
0 & -\lambda & p_1 \lambda & p_2 \lambda & p_3 \lambda & p_4 \lambda & \ldots \\
0 & 0 & -\lambda & p_1 \lambda & p_2 \lambda & p_3 \lambda & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
\]

Moreover, as noted by Lucantoni [5], if \( g(z) \) is the generating function of \( \{p_k : k \geq 1\} \), then \( D(z) = -\lambda + \lambda g(z) \), \( |z| \leq 1 \).

3. **Batch Markov-Modulated Poisson Process (MMPP)**. Consider a Poisson process whose rate is modulated by an exogenous, irreducible Markov process, \( \{J(t) : t \geq 0\} \), with state space \( \{1, 2, \ldots, m\} \) and generator matrix \( Q \). Whenever \( J(t) = i \), arrivals are according to a Poisson process with rate \( \lambda_i \) (\( \lambda_i > 0 \)). Define the vector \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m) \) and let \( \Delta(\lambda) = \text{diag}(\lambda) \). Arrivals occur in batches of size \( k \) with probability \( p_k \), \( k \geq 1 \). If \( N(t) \) denotes the number of arrivals up to time \( t \), then \( \{(N(t), J(t)) : t \geq 0\} \) is a BMAP with \( D_0 = Q - \Delta(\lambda) \), \( D_k = p_k \Delta(\lambda) \) for \( k \geq 1 \). (An excellent summary of the standard Markov-Modulated Poisson Process is provided by Fischer, et al. [11].)

4. **Batch PH-Renewal Process**. Suppose that arrivals are according to a renewal process, \( \{\tau_n : n \geq 0\} \), where \( \tau_n \) denotes the \( n \)th arrival epoch. The PH-renewal process is a renewal process for which the inter-renewal times \( S_n \equiv \tau_{n+1} - \tau_n \), \( n \geq 0 \), form an i.i.d. sequence of PH-distributed random variables with representation \((\alpha, T)\), where \( T \) is of order \( m \). Again, let \( p_k \) denote the probability that the batch size is \( k \), \( k \geq 1 \). The batch PH-renewal process is then a BMAP with \( D_0 = T \) and \( D_k = p_k T^0 \alpha \), \( k \geq 1 \) where \( T^0 = -Te \).

5. **Superposition of Independent BMAsps**. The superposition of \( N \) independent BMAsps is again a BMAP. Let \( \{D_k(i) : k \geq 0\}, i = 1, 2, \ldots, N \), denote a collection of \( N \) independent BMAsps such that the phase process of the \( i \)th BMAP is of order \( m(i) \). Let

\[
M = \prod_{i=1}^{N} m(i),
\]

and define for each \( k \geq 0 \) the \( M \times M \) matrix \( D_k \) by

\[
D_k = D_k(1) \oplus \cdots \oplus D_k(N),
\]

where the operator \( \oplus \) is the *Kronecker matrix sum* (see [8, 11]). Then \( \{D_k : k \geq 0\} \) is the representation of the superposition of the \( N \) independent BMAsps.
The Discrete-Time BMAP (D-BMAP)

Consider an irreducible discrete-time Markov chain (DTMC) \( J \equiv \{ J_r : r \geq 0 \} \) on the state space \( E = \{1, 2, \ldots, m\} \) which allows self-transitions. Suppose that the \( n \)th transition of \( J \) triggers the arrival of a batch of customers of size \( Y_n \) (\( Y_n \geq 0 \)), with \( Y_0 = 0 \). Define the conditional probabilities

\[
q_{ij}(k) = P(X_{n+1} = j, Y_n = k \mid X_n = i), \quad i, j \in E, \ n \geq 0,
\]

which are the joint probabilities of a transition of the discrete-time chain \( J \) from \( i \) to \( j \) and an arrival of batch size \( k \geq 0 \). Next, define the (sub-stochastic) matrices \( D_k = [q_{ij}(k)]_{i,j \in E} \) and assume that \( I - D_0 \) is nonsingular to ensure that the arrival of one or more customers occurs with probability one. The transition probability matrix of \( J \), denoted by \( P \), is given by

\[
P = \sum_{k=0}^{\infty} D_k,
\]

whose entries are necessarily finite. The matrices \( \{ D_k : k \geq 0 \} \) completely specify a discrete-time batch Markovian arrival process (D-BMAP).

As with the continuous-time BMAP, it is possible to construct a Markov chain representation for the D-BMAP. For \( r \geq 0 \), let \( N_r \) be the total number of arrivals up to, and including, the \( r \)th transition of \( J \). The process \( \{ N_r : r \geq 0 \} \) is the counting process of the D-BMAP, which, together with the phase process \( J \), allows us to define the bivariate process

\[
\{(N_r, J_r) : r \geq 0\},
\]

which is a two-dimensional DTMC with transition probability matrix \( P^* \) given by

\[
P^* = \begin{bmatrix}
D_0 & D_1 & D_2 & D_3 & \ldots \\
0 & D_0 & D_1 & D_2 & \ldots \\
0 & 0 & D_0 & D_1 & \ldots \\
0 & 0 & 0 & D_0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

Next we elucidate an important property connected to the counting process as documented in Blondia [12]. Define the \( r \)-step transition matrix \( P(n, r) \) whose \((i, j)\)th entry is defined as

\[
[P(n, r)]_{ij} = P(N_r = n, J_r = j \mid J_0 = i), \quad r \geq 1, \ n \geq 0.
\]

The matrix generating function of \( P(n, r) \), denoted by \( P^*(z, r) \), is given by

\[
P^*(z, r) = \sum_{n=0}^{\infty} P(n, r)z^n, \quad |z| \leq 1.
\]

It can be shown that, for \( r \geq 1 \),

\[
P^*(z, r) = [P^*(z, 1)]^r = [D(z)]^r, \quad |z| \leq 1,
\]  

(9)
where \( D(z) = \sum_{k=0}^{\infty} D_k z^k \). If \( D(z) \) is known explicitly, then the \( k \)-th-factorial moments of the number of arrivals in \( r \) \((r \geq 1)\) transitions can be obtained by computing the \( k \)-th-order derivative of \( P^*(z,r) \) and evaluating at \( z = 1 \).

Next we consider the fundamental rate of the D-BMAP. Let \( \pi = [\pi_1, \ldots, \pi_m] \) be the invariant probability vector of the phase process \( J \); that is, \( \pi \) is the unique positive solution to the system of equations

\[
\pi P = P \quad \text{and} \quad \pi e = 1.
\]

Then the fundamental, or stationary arrival, rate of the D-BMAP is given by

\[
\lambda = \pi \left( \sum_{k=1}^{\infty} k D_k \right) e.
\]

This time, the stationary batch arrival rate may be computed as

\[
\lambda_g = \pi(I - D_0)e,
\]

which is always nonzero due to the assumption that \( I - D_0 \) is nonsingular. As before, we have that \( \lambda = \lambda_g \) if the maximum possible batch size is one, as in a discrete-time MAP.

In the next section, we discuss a few common D-BMAPs and point the reader to a few more extensive references on the subject.

**Common Discrete-Time BMAPs**

1. **Batch Geometric Process**: Arrivals here are considered to be a sequence of independent trials for which the “success” probability \( p_0 \) \((0 < p_0 < 1)\) corresponds to a batch size of zero. This process is a D-BMAP with \( m = 1 \), \( D_0 = p_0 \), and \( D_k = p_k(1 - p_0) \), where \( \{p_k : k \geq 1\} \) are the batch-size probabilities conditioned upon the arrival of a batch of size \( k \geq 1 \). For the single-arrival process, we note that \( p_1 = 1 \) and \( p_k = 0 \) for \( k \geq 2 \), thus giving \( D_0 = p_0 \), \( D_1 = 1 - p_0 \), and \( D_k = 0 \) for \( k \geq 2 \).

2. **Batch Markov-Modulated Bernoulli Process**: The Markov-Modulated Bernoullie Process (MMBP) is the discrete-time analogue of the MMPP. For both the single-arrival and batch versions of the MMBP, arrivals are triggered by the transitions of a \( m \)-state DTMC with transition probability matrix \( P \). If the process ends up in state \( j \in \{1, \ldots, m\} \), then the probability of an arrival is given by \( \eta_j \in (0,1] \), while the probability of a null arrival is given by \( 1 - \eta_j \). For notational convenience, define the vector

\[
\eta = (\eta_1, \ldots, \eta_m).
\]

As with the batch geometric process, the conditional probabilities of batch size are given by the sequence \( \{p_k : k \geq 1\} \), with the usual adjustments made for the single-arrival version. The batch MMBP may be expressed as a D-BMAP with elements \( D_0 = \Delta(e - \eta)P \) and \( D_k = p_k \Delta(\eta)P \) for \( k \geq 1 \).
3. **Batch PH-Renewal Process:** As before, we consider a renewal process whose renewal epochs are the set of points \( \{ \tau_n : n \geq 0 \} \) with inter-renewal times \( S_{n+1} \equiv \tau_{n+1} - \tau_n, \ n \geq 0 \). Here, the i.i.d. sequence of random variables, \( \{ S_n : n \geq 1 \} \), has a discrete phase-type (PH) distribution with representation \((\alpha, T)\) with \( T \) of order \( m \). The process is then a D-BMAP with \( D_0 = T \) and \( D_k = p_k T^0 \alpha, \ k \geq 1 \), where \( T^0 = -Te \) and \( \alpha \) is the vector of initial probabilities for the discrete PH-distribution.

**Further Reading**

The prototypical formulation of the continuous-time BMAP is the versatile Markovian point process (VMPP) introduced by Neuts [1] in 1979. The current form of the MAP was developed in [3] in conjunction with its application as an arrival process to a single-server queueing system, and the extension to the BMAP is detailed in [5]. The D-BMAP first appears in the works of Blondia [12, 13] and has since pervaded the queueing and computer and communications networking literature, just as its continuous-time predecessor. An excellent summary of the BMAP and D-BMAP, along with examples and selected applications, can be found in Chakravarthy [6].

The analysis of queueing systems with BMAP (or related) input has received considerable attention in the stochastic modeling community. For specific examples of single-server queueing models with MAP input, the reader is referred to [14, 15, 16, 17, 18]. Machihara [19] examined single-server queues with batch arrivals and state-dependent service times, while Hofmann [20] considered state-dependent batch arrival rates. Krieger, et al. [21] studied a Markov-modulated BMAP\( /G/1 \) queue. Queues with BMAP input and server vacations have received much attention, beginning with [3]. Lucantoni [5] provided a nice summary of a number of important results for the BMAP\( /G/1 \) system. A sampling of the ensuing literature available on the subject may be found in [22, 23, 24, 25, 26]. Chydzinski pursues the transient analysis, first of the MMPP\( /G/1/k \) loss system in [27] and in [28] goes on to consider the time of first passage to buffer overflow in the BMAP\( /G/1/k \) queue. The processor-sharing queueing discipline has been studied extensively in [29, 30, 31].

Another fruitful area of research has considered queueing systems with BMAP input and retrials. Retrial queueing systems are extensively used to model systems in which jobs retry service after encountering a busy or failed server. For example, they are extremely useful for modeling the retransmission of data packets in telecommunications networks or the call-back behavior of customers in a customer contact center. Some early examples of the BMAP in retrial queues include [32, 33, 34, 35]. Chakravarthy, et al. [36, 37] introduce single- and multi-server retrial models with group services and exponential retrials. Breuer, et al. [38] consider a BMAP\( /PH/n \) multi-server retrial system while Li, et al. [39] consider the complication of an unreliable server in the BMAP retrial model.

The BMAP has been applied extensively in a number of areas from inventory management [40] to maintenance models [41]. However the preponderance of applications lie in computer and
communications networking, with the bulk of these utilizing stochastic processes with D-BMAP input. The D-BMAP is often used to model specific characteristics of source signals, such as burstiness, in telecommunications systems. Blondia [13] introduces the D-BMAP model and in [12] analyzes the long-run system size distribution of a $D-BMAP/G/1/N$ queue. Van Houdt and Blondia [42] used a D-BMAP to model packet arrivals in a centralized wireless local area network. In [43], they study contention resolution in a network among many users generating signals modeled as a D-BMAP. Zhao, et al. [44] extended the work in [45] on the $D-BMAP/PH/1/N$ queue to the case of prioritized service. Queues with PH-distributed service times have proven useful in modeling video streaming over networks, and the addition of a prioritization scheme enhances the usefulness of this model in networks with heterogeneous data streams.

References


[40] Yadavalli, VSS, Sivakumar, B, Arivarignan, G. Stochastic inventory management at a service facility with a set of reorder levels. ORiON (Journal of the Operations Research Society of South Africa ORSSA)). 2007; 23(2):137-149.


