Batch Markovian Arrival Processes (BMAP)

James D. Cordeiro
Department of Mathematics and Statistics
Air Force Institute of Technology
2950 Hobson Way (AFIT/ENC)
Wright Patterson AFB, OH 45433
Email: cordeirj@yahoo.com

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 stochstic 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 distinct classes of batch arrivals, each of which are determined by the transition type of an external Markov process with m transient states and one absorbing state,

$m + 1$. One type of arrival is from a Markov-modulated Poisson process, and this type occurs during the sojourn of the exogenous process in any of the m transient states. Another type occurs when the Markov process in state i transitions to state j, $j \neq i$, which is an ordinary 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 uses the concept of arrival dependence upon an external Markov process but does not distinguish between classes of arrivals. The MAP 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 are typically classified as two main types: the $GI/M/1$-type and the $M/G/1$-type. A well-known third type, the quasi-birth-and-death (QBD) process, can be viewed 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 $D\text{-}BMAP/G/1$ 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 interested reader to gain a deeper understanding of 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 $\boldsymbol{Q}$. 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 ith diagonal element of $\boldsymbol{Q}$. 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

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

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

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

and

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

The matrix $\boldsymbol{D}_0$ contains the transition rates of J for which no arrivals occur, and the matrices $\{\boldsymbol{D}_k : k \geq 1\}$ contain the transition rates for which a batch size k occurs. Assuming $\boldsymbol{D}_0$ 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

$$\boldsymbol{Q} = \sum_{k=0}^{\infty} \boldsymbol{D}_k.$$

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

$$\boldsymbol{Q}^* = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 & \boldsymbol{D}_2 & \boldsymbol{D}_3 & \cdots \\ 0 & \boldsymbol{D}_0 & \boldsymbol{D}_1 & \boldsymbol{D}_2 & \cdots \\ 0 & 0 & \boldsymbol{D}_0 & \boldsymbol{D}_1 & \cdots \\ 0 & 0 & 0 & \boldsymbol{D}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

3

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{P}(n, t) = \sum_{r=0}^{n} \boldsymbol{P}(r, t) \boldsymbol{D}_{n-r}, \tag{4}$$

$$\boldsymbol{P}(0, 0) = \boldsymbol{I},$$

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

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

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

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{P}^*(z, t) = \boldsymbol{P}^*(z, t) \boldsymbol{D}(z), \ t \geq 0, \tag{6}$$

$$\boldsymbol{P}^*(z, 0) = \boldsymbol{I},$$

where

$$\boldsymbol{D}(z) \equiv \sum_{k=0}^{\infty} \boldsymbol{D}_k z^k, \quad |z| \leq 1. \tag{7}$$

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

$$\boldsymbol{P}^*(z, t) = \boldsymbol{P}^*(z, 0) \exp(\boldsymbol{D}(z) t) = \exp(\boldsymbol{D}(z) t), \quad |z| \leq 1, \ t \geq 0,$$

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 $\boldsymbol{D}(z)$ to account for batch sizes larger than unity.

Using the generating function of $\boldsymbol{P}(n,t)$, 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 $\mathbb{P}_i$, the probability law of (N,J) given $J(0)=i$ and $N(0)=0$. Then $\mathbb{E}_i(N(t)\mathbf{1}(J(t)=j))$ is the (i,j)th entry of the $m \times m$ matrix

$$\frac{\mathrm{d}}{\mathrm{d}z}\boldsymbol{P}^*(z,t)\bigg|_{z=1} = \boldsymbol{D}(1)\exp(\boldsymbol{D}(1)\,t) = \boldsymbol{Q}\exp(\boldsymbol{Q}\,t).$$

The conditional kth factorial moment can be obtained by taking the kth-order derivative $\boldsymbol{P}^*(z,t)$ and evaluating at $z=1$ in the usual way.

The limiting behavior of the continuous-time BMAP is discussed next. Let $\boldsymbol{\pi} = [\pi_1,\ldots,\pi_m]$ be the invariant probability vector of the CTMC, $\{J(t): t \geq 0\}$, with generator matrix $\boldsymbol{Q}$; that is, $\boldsymbol{\pi}$ is the unique positive solution to the system of equations

$$\boldsymbol{\pi}\boldsymbol{Q} = \mathbf{0} \text{ and } \boldsymbol{\pi}\boldsymbol{e} = 1$$

where $\mathbf{0}$ is the zero (row) vector and $\boldsymbol{e}$ is a (column) vector of ones. Then the *fundamental rate*, or the stationary rate of arrivals in a BMAP, is given by

$$\lambda = \boldsymbol{\pi}\left(\sum_{k=1}^{\infty} k\,\boldsymbol{D}_k\right)\boldsymbol{e}. \tag{8}$$

On the other hand, the arrival rate of *batches* is given by

$$\lambda_g = -\boldsymbol{\pi}\boldsymbol{D}_0\boldsymbol{e},$$

which is never zero since $\boldsymbol{D}_0$ 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 λ, 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 λ. In this case, the matrices $\{\boldsymbol{D}_k: k \geq 0\}$ 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, $\{N(t): t \geq 0\}$ is a BMAP

with generator matrix

$$
\boldsymbol{Q}^* =
\begin{bmatrix}
-\lambda & \lambda & 0 & 0 & 0 & 0 & \cdots \\
0 & -\lambda & \lambda & 0 & 0 & 0 & \cdots \\
0 & 0 & -\lambda & \lambda & 0 & 0 & \cdots \\
0 & 0 & 0 & -\lambda & \lambda & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
$$

2. **Batch Poisson Process**: If we allow a batch size greater than unity in the standard Poisson process with rate λ, 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 the batch Poisson Process is a BMAP with generator matrix

$$
\boldsymbol{Q}^* =
\begin{bmatrix}
-\lambda & p_1\lambda & p_2\lambda & p_3\lambda & p_4\lambda & p_5\lambda & \cdots \\
0 & -\lambda & p_1\lambda & p_2\lambda & p_3\lambda & p_4\lambda & \cdots \\
0 & 0 & -\lambda & p_1\lambda & p_2\lambda & p_3\lambda & \cdots \\
0 & 0 & 0 & -\lambda & p_1\lambda & p_2\lambda & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
$$

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 $\boldsymbol{Q}$. Whenever $J(t) = i$, arrivals are according to a Poisson process with rate λ_i ($\lambda_i > 0$). Define the vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and let $\Delta(\boldsymbol{\lambda}) = \text{diag}(\boldsymbol{\lambda})$. 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 $\boldsymbol{D}_0 = \boldsymbol{Q} - \Delta(\boldsymbol{\lambda})$, $\boldsymbol{D}_k = p_k\Delta(\boldsymbol{\lambda})$ 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 τ_n denotes the nth 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 $(\boldsymbol{\alpha}, \boldsymbol{T})$, where $\boldsymbol{T}$ 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 $\boldsymbol{D}_0 = \boldsymbol{T}$ and $\boldsymbol{D}_k = p_k\boldsymbol{T}^0\boldsymbol{\alpha}$, $k \geq 1$ where $\boldsymbol{T}^0 = -\boldsymbol{T}\boldsymbol{e}$.

5. **Superposition of Independent BMAPs**. The superposition of N independent BMAPs is again a BMAP. Let $\{\boldsymbol{D}_k(i) : k \geq 0\}$, $i = 1, 2, \ldots, N$, denote a collection of N independent

BMAPs such that the phase process of the ith 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 $\boldsymbol{D}_k$ by

$$\boldsymbol{D}_k = \boldsymbol{D}_k(1) \oplus \cdots \oplus \boldsymbol{D}_k(N),$$

where the operator $\oplus$ is the *Kronecker matrix sum* (see [8, 11]). Then $\{\boldsymbol{D}_k : k \geq 0\}$ is the representation of the superposition of the N independent BMAPs.

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 nth 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

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

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 $\boldsymbol{D}_k = [q_{ij}(k)]_{i,j \in E}$ and assume that $\boldsymbol{I} - \boldsymbol{D}_0$ is nonsingular to ensure that the arrival of one or more customers occurs with probability one. The transition probability matrix of J, denoted by $\boldsymbol{P}$, is given by

$$\boldsymbol{P} = \sum_{k=0}^{\infty} \boldsymbol{D}_k,$$

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

As for the continuous-time BMAP, it is possible to construct a bivariate Markov chain representation of the D-BMAP. For $r \geq 0$, let N_r be the total number of arrivals up to, and including, the rth 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 $\boldsymbol{P}^*$ given by

$$\boldsymbol{P}^* = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 & \boldsymbol{D}_2 & \boldsymbol{D}_3 & \ldots \\ 0 & \boldsymbol{D}_0 & \boldsymbol{D}_1 & \boldsymbol{D}_2 & \ldots \\ 0 & 0 & \boldsymbol{D}_0 & \boldsymbol{D}_1 & \ldots \\ 0 & 0 & 0 & \boldsymbol{D}_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Next, we elucidate an important property connected to the counting process as documented in Blondia [41]. Define the r-step transition matrix $\boldsymbol{P}(n,r)$ whose (i,j)th entry is defined as

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

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

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

It can be shown that, for $r \geq 1$,

$$\boldsymbol{P}^*(z,r) = [\boldsymbol{P}^*(z,1)]^r = [\boldsymbol{D}(z)]^r, \quad |z| \leq 1, \tag{9}$$

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

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

$$\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{P} \text{ and } \boldsymbol{\pi}\boldsymbol{e} = 1.$$

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

$$\lambda = \boldsymbol{\pi}\left(\sum_{k=1}^{\infty} k\,\boldsymbol{D}_k\right)\boldsymbol{e}.$$

Here, the stationary batch arrival rate may be computed as

$$\lambda_g = \boldsymbol{\pi}(\boldsymbol{I} - \boldsymbol{D}_0)\boldsymbol{e},$$

which is always nonzero due to the assumption that $\boldsymbol{I} - \boldsymbol{D}_0$ 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 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 Bernoulli 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 an m-state DTMC with transition probability matrix $\boldsymbol{P}$. 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

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m).$$

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 $\boldsymbol{D}_0 = \Delta(\boldsymbol{e} - \boldsymbol{\eta})\boldsymbol{P}$ and $\boldsymbol{D}_k = p_k \Delta(\boldsymbol{\eta})\boldsymbol{P}$ 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 $(\boldsymbol{\alpha}, \boldsymbol{T})$ with $\boldsymbol{T}$ of order m. The process is then a D-BMAP with $\boldsymbol{D}_0 = \boldsymbol{T}$ and $\boldsymbol{D}_k = p_k \boldsymbol{T}^0 \boldsymbol{\alpha}$, $k \geq 1$, where $\boldsymbol{T}^0 = -\boldsymbol{T}\boldsymbol{e}$ and $\boldsymbol{\alpha}$ is the vector of initial probabilities for the discrete PH-distribution.

Further Reading

The original formulation of the continuous-time BMAP was the versatile Markovian point process (VMPP) introduced by Neuts [1] in 1979. The current form of the MAP was developed in [3] as an arrival process to a single-server queueing system, and the extension to the BMAP is detailed in [5]. The D-BMAP first appeared in the works of Blondia [41, 40] 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 processes has received considerable attention in the stochastic modeling community. Specific examples of single-server queueing models with MAP input can be found in [12, 13, 14, 15, 16]. Machihara [17] examined single-server queues with batch arrivals and state-dependent service times, while Hofmann [18] considered state-dependent batch arrival rates. Krieger, et al. [19] 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 can be found in [20, 21, 22, 23, 24]. Chydzinsky [25] provided a transient analysis of the $MMPP/G/1/k$ loss system, and in [26], analyzed the first passage time to buffer overflow in the $BMAP/G/1/k$ queue. The processor-sharing queueing discipline in systems with MAP or BMAP input has been studied extensively in [27, 28, 29].

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 customers retry service after encountering a busy (or failed) server. For example, they are extremely useful for modeling the retransmission of data packets in computer and communications networks, or the call-back behavior of customers in a customer contact center. Some early examples of the BMAP in retrial queues include [30, 31, 32, 33]. Chakravarthy, et al. [34, 35] introduced single- and multi-server retrial models with group service and exponential retrials. Breuer, et al. [36] studied a $BMAP/PH/n$ multi-server retrial system, while Li et al. [37] introduced 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 [38] to maintenance models [39]. However, the preponderance of applications lies in computer and communications networking, with the bulk of these assuming D-BMAP input processes. The D-BMAP is often used to model specific characteristics of source signals, such as burstiness, in telecommunications systems. Blondia [40] introduced the D-BMAP model, and in [41], analyzed the steady-state system size distribution of a $D\text{-}BMAP/G/1/N$ queue. Van Houdt and Blondia [42] used a D-BMAP to model packet arrivals in a centralized wireless local area network. In [43], they examined contention resolution in a network among many users generating signals modeled as a D-BMAP. Zhao et al. [44] extended the work of [45] on the $D\text{-}BMAP/PH/1/N$ 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 these models in networks with heterogeneous data streams.

References

[1] Neuts, MF. A versatile Markovian point process. J. App. Prob. 1979; 16(4):764–779.

[2] Neuts, MF. Matrix-analytic methods in queueing theory. Eur. J. Oper. Res. 1984; 15:2–12.

[3] Lucantoni, DM, Meier-Hellstern, KS, Neuts, MF. A single-server queue with server vacations and a class of non-renewal arrival processes. Adv. App. Prob. 1990 Sep; 22(3):676–705.

[4] Lucantoni, DM. New results on the single server queue with a batch Markovian arrival process. Commun. Statist.-Stochastic Models. 1991; 7(1):1–46.

[5] Lucantoni, DM. The $BMAP/G/1$ Queue: A Tutorial. Pp 330–358 in Donatiello, L. and Nelson, R., eds. Models and Techniques for Performace Evaluation of Computer and Communications Systems. 1993. Springer-Verlag, London, U.K.

[6] Chakravarthy, SR. The Batch Markovian Arrival Process: A Review and Future Work. Pp 21–39 in Krishnamoorthy, A., Raju, N., and Ramaswami, V., eds. Advances in Probability Theory and Stochastic Processes. 2000. Notable Publications, New Jersey.

[7] Neuts, MF. Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. 1981. Dover Publications, Inc., New York, NY.

[8] Neuts, MF. Structured Stochastic Matrices of $M/G/1$ Type and Their Applications. 1989. Marcel Dekker, Inc., New York, NY.

[9] Ramaswami, V. The $N/G/1$ queue and its detailed analysis. J. App. Prob. 1980; 12:222–261.

[10] Latouche, G. and Ramaswami, V. Introduction to Matrix–Analytic Methods in Stochastic Modeling. 1999. American Stat. Assoc. and the Soc. for Indust. and Applied Mathematics, Alexandria, VA and Philadelphia, PA.

[11] Fischer, W., Meier-Hellstern, K. The Markov-modulated Poisson process (MMPP) cookbook. Perf. Eval. 1992; 18:149–171.

[12] Subramanian, V, Srikant, R. Tail probabilities of low-priority waiting times and queue lengths in $MAP/GI/1$ queues. Queue. Sys. 2000; 34(1-4):215–236.

[13] Shioda, S. Departure process of the $MAP/SM/1$ queue. Queue. Sys. 2003; 44(1):31–50.

[14] Adan, IJ, Kulkarni, VG. Single-server queue with Markov-dependent inter-arrival and service times. Queue. Sys. 2003; 45(2):113–134.

[15] Li, QL, Zhao, YQ. A $MAP/GI/1$ queue with negative customers. Queue. Sys. 2004; 47(1-2):5–43.

[16] Lee, HW, Cheon, SH, Lee, EY, Chae, KC. Workload and waiting time analyses of $MAP/GI/1$ queue under D-policy. Queue. Sys. 2004; 48:421–443.

[17] Machihara, F. A $BMAP/SM/1$ queue with service times depending on the arrival process. Queue. Sys. 1999; 33(4):277–291.

[18] Hofmann, J. The $BMAP/G/1$ queue with level-dependent arrivals - An overview. Telecomm. Sys. 2001; 16:(3-4)347–359.

[19] Krieger, U, Klimenok, VI, Kazimirsky, AV, Breuer, L, Dudin, AN. A $BMAP/PH/1$ queue with feedback operating in a random environment. Mathematical & Computer Modelling. 2005; 41(8-9):867–882.

[20] Shin, YW, Pearce, CEM. The $BMAP/G/1$ vacation queue with queue-length dependent vacation schedule. J. Austral. Math. Soc. Ser. B. 1998; 40:207–221.

[21] Ferng, HW. Departure processes of $BMAP/G/1$ queues. Queue. Sys. 2001; 39(2-3):109–135.

[22] Chang, SH, Takine, T, Chae, KC, Lee, HW. A unified queue length formula for $BMAP/G/1$ queue with generalized vacations. Stoch. Models. 2002; 18(3):369–386.

[23] Lee, HW, Park, NI. Using factorization for waiting times in $BMAP/G/1$ queues with N-policy and vacations. Stoch. Analy. & Appl. 2004; 22(3):755–773.

[24] Baek, JW, Lee, HW, Lee, SW, Ahn, S. A factorization property for $BMAP/G/1$ vacation queues under variable service speed. Ann. Oper. Res. 2008; 160:19–29.

[25] Chydzinski, A. Transient analysis of the $MMPP/G/1/K$ queue. Telecommun. Syst. 2006; 32:247–262. DOI: 10.1007/s11235-006-9001-5.

[26] Chydzinski, A. Time to reach buffer capacity in a BMAP queue. Stoch. Mod. 2006; 23:195–209. DOI: 10.1080/15326340701300746

[27] D'Apice, C, Manzo, R, Pechinkin, AV. A finite $MAP_k/G_k/1$ queueing system with generalized foreground-background processor-sharing discipline. Automation and Remote Control. 2004; 65(11):1793–1799.

[28] Li, QL, Lian, Z, Liu, L. An RG-factorization approach for a $BMAP/M/1$ generalized processor-sharing[3] queue. 2005; 21:507–530. DOI: 10.1081/STM-200056223.

[29] D'Apice, C, Manzo, R, Pechinkin, AV. A finite capacity $BMAP_k/G_k/1$ queueing system with generalized foreground-background processor-sharing discipline. Automation and Remote Control. 2006; 67(3):428–434.

[30] Choi, BD, Chang, Y. $MAP_1, MAP_2/M/c$ retrial queue with the retrial group of finite capacity and geometric loss. Mathematical and Computer Modelling. 1999; 30(3–4):99–113. DOI:10.1016/S0895-7177(99)00135-1

[31] Dudin, A, Klimenok, V. A retrial $BMAP/SM/1$ system with linear repeated requests. Queueing Sys. 2000; 34:47–66.

[32] Choi, BD, Chung, YH, Dudin, AN. The $BMAP/SM/1$ retrial queue with controllable operation modes. European J. of Operational Research. 2001; 131(1):16–30.

[33] Klimenok, VI. A multiserver retrial queueing system with batch Markov arrival process. Automation and Remote Control. 2001; 62(8): 1312–1322.

[34] Chakravarthy, SR, Dudin, AN. A Single Server Retrial Queuing Model with Batch Arrivals and Group Services. Pp 1–21 in Artalejo, J.R. and Krishnamoorthy, A., eds. Advances in Stochastic Modeling. 2002. Notable Publications, New Jersey.

[35] Chakravarthy, SR, Dudin, AN. A multi-server retrial queue with $BMAP$ arrivals and group services. Queueing Sys. 2002; 42: 5–31.

[36] Breuer, L, Dudin, A, Klimenok, V. A retrial $BMAP/PH/N$ system. Queueing Sys. 2002; 40:433–457.

[37] Li, Q, Ying, Y, Zhao, YQ. A $BMAP/G/1$ retrial queue with a server subject to breakdowns and repairs. Ann. Oper. Res. 2006; 141:233–270. DOI: 10.1007/s10479-006-5301-0.

[38] 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.

[39] Lee, HW, Park, NI. A maintenance model for manufacturing lead time in a production system with BMAP input and bilevel setup control. Asia-Pacific Journal of Operational Research. 2008; 25(6):807–825.

[40] Blondia, C. A discrete time Markovian arrival process. RACE Document PRLB-123-0015-CD-CC. August 1989.

[41] Blondia, C. A discrete time batch Markovian arrival process as B-ISDN traffic model. Belgian J. Oper. Res., Statistics and Computer Science. 1993; 32:3–23.

[42] Van Houdt, B, Blondia, C. Robustness properties of FS-ALOHA++: A contention resolution algorithm for dynamic bandwidth allocation. Mobile Networks and Applications. 2003; 8:237-253.

[43] Van Houdt, B, Blondia, C. Throughput of q-ary splitting algorithms for contention resolution in communication networks. Communications in Information and Systems. 2005; 4(2):135–164.

[44] Zhao, J, Li, B, Cao, X, Ahmad, I. A matrix-analytic solution for the $DBMAP/PH/1$ priority queue. Queuing Sys. 2006. 53:127–145. DOI: 10.1007/s11134-006-8306-0.

[45] Blondia, C and Casals, O. Statistical multiplexing of VBR sources: a matrix-analytic approach. Perf. Eval. 1992; 16(1-3):5–20.