On the relationships among queue lengths
at arrival, departure, and random epochs
in the discrete-time queue with D-BMAP arrivals

Nam K. Kim, Seok H. Chang, Kyung C. Chae*

Department of Industrial Engineering, Korea Advanced Institute of Science and
Technology, Daejon, 305-701, Korea

Abstract

We consider finite- and infinite-capacity queues with discrete-time batch Markovian
arrival processes (D-BMAP) under the assumption of the Late Arrival System with
Delayed Access as well as the Early Arrival System. Using simple arguments such as the
balance equation, “rate in = rate out,” we derive relationships among the stationary
queue lengths at arrival, at departure, and at random epochs. Such relationships hold for
a broad class of discrete-time queues with D-BMAP arrivals.

Keywords: Discrete-time queues; Markovian arrival process; Queue length

* Corresponding author. Tel: +82-42-869-2915; fax: +82-42-869-3110.
E-mail address: kcchae@kaist.ac.kr

Acknowledgements: The authors are grateful to Dr. Mohan L. Chaudhry and the anonymous reviewer for
their constructive comments that greatly improved the presentation of this paper.
1. Introduction and notation

In recent years, there has been a growing interest in the analysis of discrete-time queues due to their applications in slotted digital communication systems and other related areas. One of the reasons for this fact is that discrete-time queueing models fit better the discrete nature of computer and communication systems than the continuous-time counterparts, and therefore, they can give more accurate performance measures of these systems.

In this paper, we consider the stationary queue lengths of finite- and infinite-capacity queues with discrete-time batch Markovian arrival processes (D-BMAP). Using simple arguments such as “rate in = rate out,” we derive relationships among the stationary queue lengths at arrival, at departure, and at random epochs. Relationships of this kind are useful in the sense that once one obtains a solution for certain epochs, e.g., departure epochs, it is immediate to get solutions for the others through these relationships. Continuous-time counterparts of these relationships are available in Takine and Takahashi [13] and Kim et al. [8].

Remark 1. The D-BMAP is a versatile process that contains a large class of discrete-time point processes such as Bernoulli arrival process, Markov modulated Bernoulli process, and batch Bernoulli process with correlated batch arrivals. (see, e.g., Blondia and Casals [2]). Furthermore, we do not assume any particular service mechanism, e.g., the service discipline, the number of servers, the batch service, generalized vacations, etc. Thus the relationships presented in this paper hold for a broad class of discrete-time queues with D-BMAP arrivals.

In discrete-time queueing models, the time axis is divided into fixed-length intervals, called slots, and customer arrivals and departures are assumed to be taking place at slot boundaries. Different assumptions can be made on the order of an arrival and a departure simultaneously taking place at a slot boundary: either an arrival may have precedence over a departure or vice versa. The former case is referred to as the Late Arrival System (LAS or arrivals first policy) and the latter as the Early Arrival System (EAS or departures first policy). Until turning our attention to the EAS in Section 4, we
assume the \textit{LAS with Delayed Access} (LAS-DA), where delayed access means that arriving customers who find the server(s) available do not immediately get served but are delayed until the beginning of the next slot. For more details on these models, see Bruneel and Kim [4, p.1], Hunter [7, p.193], and Takagi [12, p.4].

In a queue with D-BMAP arrivals, groups of customers of size $k$ arrive at the queue according to a D-BMAP with representation $\{D_k, k \geq 0\}$, where $D_k$ is an $m \times m$ matrix. Note that $m$ denotes the number of phases in the underlying Markov chain (UMC) that governs the arrival process. Suppose that the UMC is in some phase $i$ in the middle of a slot. Then, under the LAS-DA, with respective probabilities $(D_k)_{ij}$ and $(D_0)_{ij}$, there is a phase transition to $j$ with a batch arrival of size $k \geq 1$ and without an arrival, just prior to the end of the slot. Just after the slot boundary, there may be a departure of customer(s), if any. Note that nothing could happen somewhere in the middle of a slot (for more details on the D-BMAP, see Blondia [1], [2] and Herrmann [6]).

Let $\lambda$ and $\lambda_g$, respectively, denote the average numbers of customer arrivals and batch arrivals per slot. They are then given by

$$\lambda = \pi \sum_{k=1}^{\infty} k D_k e,$$

$$\lambda_g = \pi \sum_{k=1}^{\infty} D_k e,$$

where $\pi$ is the stationary probability vector of the UMC with the transition probability matrix $D = \sum_{k=0}^{\infty} D_k$, and $e$ is a column vector of 1’s.

Let $N$ be the upper limit on the number of customers in the system. We assume the \textit{Partial Rejection Policy} (PRP), in which arriving customers in a batch fill empty positions in the system and, once all available positions are filled, the remaining customers in the batch are rejected (see, e.g., Takagi [11, p. 412]).

Now, we consider a discrete-time bivariate process $\{L_k, S_k; k \geq 1\}$, where $L_k$ denotes the number of customers in the system and $S_k$ denotes the phase of the UMC, both in the middle of the $k$th slot. We define the following probabilities in number-average sense for $0 \leq i \leq N$ and $1 \leq j \leq m$:

3
$y_{i,j}$: stationary probability that the process is in state $(i, j)$ in the middle of a slot

$Y_i = (y_{i,1}, \ldots, y_{i,m})$

(We interpret $y_{i,j}$ as the long-run fraction of slots during which the state is $(i, j)$.)

$y_{i,j}^b$: stationary probability that a batch (or group) finds $i$ customers in the system and that the UMC is in phase $j$ just after its arrival

$Y_i^b = (y_{i,1}^b, \ldots, y_{i,m}^b)$

(We interpret $y_{i,j}^b$ as the long-run fraction of batches that find $i$ customers in the system and make transitions of the UMC into $j$.)

$y_{i,j}^-$: stationary probability that an individual customer finds $i$ customers in the system (including accepted customers who precede her in her own batch – see Remark 2 below) and that the UMC is in phase $j$ just after the arrival of her batch

$Y_i^- = (y_{i,1}^-, \ldots, y_{i,m}^-)$

(We interpret $y_{i,j}^-$ as the long-run fraction of individual customers who find $i$ customers in the system with the phase of the UMC being $j$ just after the arrivals of their batches.)

$x_{i,j}$: stationary probability that an individual customer leaves behind $i$ customers in the system (including served customers, if any, who follow her to depart at the same time – see Remark 2 below) and that the UMC is in phase $j$ just after her departure

$X_i = (x_{i,1}, \ldots, x_{i,m})$

(We interpret $x_{i,j}$ as the long-run fraction of individual customers who leave behind $i$ customers in the system with the phase of the UMC being $j$ just after her departure.)
Remark 2. Note that we are adopting a well known convention (see e.g., Wolff [14, p.388]) that customers of an arriving/departing batch enter/leave the system not simultaneously but one at a time instantaneously. That is, we suppose that customers of an arriving batch form a line to enter the system one after another. Likewise, when we consider bulk departures in batch-service or multi-server queues, we suppose that served customers departing at the same epoch form a line to leave the system one after another. Particularly, individual customers in such an arriving line who find \( N \) customers in the system are assumed not to enter the system but to immediately depart from the system leaving those \( N \) customers behind. Note that we take such departures by rejected customers into account as well as those by accepted customers so that the average number of customer departures per slot is still \( \lambda \).

2. A relationship between the queue lengths at random and at departure epochs

In this section, we derive a relationship between the stationary queue lengths at random epochs in the middle of slots and just after departure epochs by using the balance equation, “rate in = rate out.” Hereafter, the average numbers of transitions into and out of some state \((i, j)\) per slot are called the transition rates into and out of \((i, j)\). Especially for such queues as bulk-arrival, bulk-service and multi-server queues in which customers may arrive and/or depart in group, we adopt the following convention to make the process \( \{L_k, S_k; k \geq 1\} \) skip-free with respect to \( L_k \):

Skip-free convention: When a batch arrival of size \( l > 1 \) makes a transition from \((i, j)\) to \((i+l, k)\), we suppose that this transition instantaneously goes through a sequence of in-between states, \((i+1, k), (i+2, k), \ldots, (i+l-1, k)\). That is, a giant transition from \((i, j)\) to \((i+l, k)\) is supposed to be made up of \( l \) instantaneous one-step transitions: \((i, j)\) to \((i+1, k)\), \((i+1, k)\) to \((i+2, k)\), \ldots, \((i+l-1, k)\) to \((i+l, k)\). Similarly, when a batch departure of size \( l > 1 \) makes a transition from \((i, j)\) to \((i-l, j)\), we suppose that this giant transition instantaneously goes through a
sequence of in-between states, \((i-1, j), (i-2, j), \ldots, (i-L+1, j)\).

We need this convention in order to represent transition rates into and out of some state \((i, j)\) in terms of such quantities as \(x_{i,j}\) (and \(y_{i,j}\)), which are defined in an individual-customer-average sense (note that their definitions correspond not to giant transitions but to one-step transitions). Now, we equate the transition rates into and out of state \((i, j)\), assuming that they exist. As a result, we have

**Theorem 1.** For a finite-capacity queue with D-BMAP arrivals under the assumptions of LAS-DA and PRP, \(y_i\) and \(x_i\) are related by

\[
\text{(i)} \quad \sum_{k=0}^{i} y_k D_{i-k} - y_i = \lambda x_{i-1} - \lambda x_i, \quad 0 \leq i \leq N-1, \tag{1}
\]

\[
\text{(ii)} \quad \sum_{k=0}^{N} \sum_{n=0}^{N} y_k D_{n-k} - y_N = \lambda x_{N-1}. \tag{2}
\]

Before presenting a proof of this theorem, we give an essential argument for calculating the rate of the transitions caused by customer departures. Note that \(\lambda\) is the average number of individual customer departures per slot (including departures by rejected customers) and that \(x_{i-1,j}, 1 \leq i \leq N\), is the fraction of individual customers who leave behind the system with state \((i-1, j)\). That is, \(\lambda x_{i-1,j}\) is the average number of individual customers per slot who depart to make transitions from \((i, j)\) to \((i-1, j)\) -- note that such a transition may be a part of a giant transition. Thus \(\lambda x_{i-1,j}\) is considered to be the transition rate out of \((i, j)\) as well as the transition rate into \((i-1, j)\) caused by individual customer departures. Note also that there are \(\lambda x_{N,j}\) rejected customers per slot who leave behind the system with state \((N, j)\). Their departures do not make any transitions at all because rejected customers can not make any effects on the state of the process except the first customers of rejected batches who can make phase-transitions of the UMC at their arrivals.
Proof. For the process \( \{L_k, S_k; k \geq 1\} \), we note that transitions into and out of \((i, j)\) during a slot are caused either by a phase transition of the UMC (with or without an arrival) or by a departure.

(i) The case \( 0 \leq i \leq N - 1 \)

We first consider the transition rate out of \((i, j)\). The out-rate caused by phase transitions of the UMC is given by

\[
\lambda_{i,j} \left( 1 - (D_{ij})_{ij} \right) + X, \quad \text{where } X \text{ is the rate of giant transitions instantaneously going through } (i, j). \quad \text{(That is, } X \text{ is the transition rate from } \{(i - n, l) | n > 0, 1 \leq l \leq m\} \text{ to } \{(i + n', j) | n' > 0\} \text{ caused by batch arrivals. Under skip-free convention, this rate of instantaneous transitions should be considered as a part of the transition rates into and out of } (i, j). \text{ However, we do not have to elaborate more on } X \text{ because it cancels out in the balance equation.) In addition, the rate out of } (i, j) \text{ caused by departures is given by } \lambda x_{i-1,j}, \text{ as discussed earlier. Thus we have, for } 10 \leq i \leq N - 1 \text{ and } 1 \leq j \leq m,
\]

\[
\text{transition rate out of } (i, j) = \left\{ \lambda_{i,j} \left( 1 - (D_{ij})_{ij} \right) + X \right\} + \lambda x_{i-1,j}, \quad (3)
\]

where \( x_{i-1,j} = 0 \).

Next, we consider the transition rate into \((i, j)\). Note that the rate into \((i, j)\) caused by phase transitions of the UMC is given by the total transition rate from \((k, l)\) to \((i, j)\) for all \((k, l) \neq (i, j)\) plus \(X\), the giant transition rate instantaneously going through \((i, j)\). In addition, the rate into \((i, j)\) caused by departures is given by \( \lambda x_{i,j} \).

Thus we have, for \( 0 \leq i \leq N - 1 \text{ and } 1 \leq j \leq m \),

\[
\text{transition rate into } (i, j) = \left\{ \sum_{k=0}^{l} \sum_{l=1}^{m} y_{k,l} (D_{k,l})_{ij} - y_{i,j} (D_{ij})_{ij} + X \right\} + \lambda x_{i,j}. \quad (4)
\]

Finally, (1) follows from equating (3) and (4).

(ii) The case \( i = N \)

In a similar manner, we first consider the transition rate out of \((N, j)\). The out-rate caused by phase transitions of the UMC is given by the total transition rate from \((N, l)\) to \((N, j)\) for all \( l \neq j \). Note that in this case there are no giant transitions having \((N, j)\) as an in-between state. In addition, there are \( \lambda x_{N-1,j} \) customer-departures per
slot that make transitions out of \((N, j)\). Thus we have

\[
\text{transition rate out of } (N, j) = y_{N,j} \left\{ 1 - \sum_{k=0}^{\infty} (D_k)_y \right\} + \lambda x_{N-1,j}.
\]  

(5)

Similarly, the rate into \((N, j)\) caused by phase transitions of the UMC is given by the total transition rate from \((k,l)\) to \((N, j)\) for all \((k,l) \neq (N, j)\). And customer departures cannot make a transition into \((N, j)\), as discussed earlier. Thus we have

\[
\text{transition rate into } (N, j) = \left\{ \sum_{k=0}^{N} \sum_{n=N}^{\infty} \sum_{l=1}^{m} y_{k,l} (D_{n-k})_{lj} - y_{N,N} \sum_{k=0}^{\infty} (D_k)_{lj} \right\}.
\]  

(6)

Finally, (2) follows from equating (5) and (6).

Note that \(\lambda x_i, 0 \leq i \leq N-1\), of (1) and (2) can be replaced by \(\lambda^* x_i^*\), where \(\lambda^* = \lambda(1-x_N e)\) is the effective arrival (or departure) rate and \(x_i^* = x_i/(1-x_N e)\) represents the probability of the queue length left behind by an accepted customer.

When the arrival process is a Bernoulli process with parameter \(\lambda\) (in this case, \(D_0 = 1-\lambda\), \(D_1 = \lambda\), and \(D_k = 0\), \(k \geq 2\)), (1) and (2) simplify to \(y_i = x_i, 0 \leq i \leq N\).

Now, multiplying both sides of (1) by \(z^i\) and summing up for all \(i \geq 0\), we have

**Corollary 1.** For a stable infinite-capacity queue with D-BMAP arrivals under the assumption of LAS-DA, \(Y(z)\) and \(X(z)\) are related by

\[
Y(z)(D(z) - I) = \lambda(z-1)X(z),
\]  

(7)

where \(Y(z) = \sum_{i=0}^{\infty} y_i z^i\), \(X(z) = \sum_{i=0}^{\infty} x_i z^i\), and \(D(z) = \sum_{k=0}^{\infty} D_k z^k\).

We remark that Theorem 1 and Corollary 1 are discrete-time analogues of the continuous-time BMAP results. For example, (7) is the discrete-time counterpart of the continuous-time result, \(Y(z)D(z) = \lambda(z-1)X(z)\) [8, 13].
3. Relationships between the queue lengths at random and at arrival epochs

In this section, we present relationships between the stationary queue lengths at random and at arrival epochs. We view an arrival in two respects: a batch arrival and an individual-customer-arrival. First, we present a relationship between the stationary queue length at random epochs in the middle of slots and that found by batch arrivals as follows:

**Theorem 2.** For a finite-capacity queue with D-BMAP arrivals under the assumptions of LAS-DA and PRP, \( y_i \) and \( y_i^{\epsilon^-} \) are related by

\[
\lambda_g y_i^{\epsilon^-} = y_i(D - D_0), \quad 0 \leq i \leq N. \tag{8}
\]

**Proof.** We note that \( \lambda_g \) is the average number of batch arrivals per slot and that \( y_i^{\epsilon^-} \) is the fraction of batches that find \( i \) customers in the system and make transitions of the UMC into \( j \). That is, \( \lambda_g y_i^{\epsilon^-} \) is the average number of batch arrivals of such kind per slot. This number obviously equals the transition rate out of \( \{i,l|1 \leq l \leq m\} \) caused by batch arrivals that make transitions of the UMC into \( j \). Thus we have, for \( 0 \leq i \leq N \) and \( 1 \leq j \leq m \),

\[
\lambda_g y_i^{\epsilon^-} = \sum_{k=1}^{\infty} \sum_{l=1}^{m} y_{i,l} (D_k)_{lj},
\]

from which the desired result follows.

We note that Theorem 2 is still valid for the Total Rejection Policy (TRP), in which all customers in an arriving batch are rejected if the number of available positions is less than the batch size (see, e.g., [11, p. 412]).

Now, multiplying both sides of (8) by \( z^i \) and summing up for all \( i \geq 0 \), we have

**Corollary 2.** For a stable infinite-capacity queue with D-BMAP arrivals under the assumption of LAS-DA, \( Y(z) \) and \( Y^{\epsilon^-}(z) \) are related by
\[
Y^g(z) = Y(z) \frac{D - D_0}{\lambda_g},
\]

where \( Y^g(z) = \sum_{i=0}^{\infty} y^g_i z^i \).

Now, we present a relationship between the stationary queue length at random epochs in the middle of slots and that found by individual customers (including accepted customers who precede them in their own batches).

**Theorem 3.** For a finite-capacity queue with D-BMAP arrivals under the assumptions of LAS-DA and PRP, \( y_i \) and \( y^-_i \) are related by

(i) \[
\lambda y^-_i = \sum_{k=0}^{j} \sum_{m=k+1}^{\infty} y_k D_{n-k}, \quad 0 \leq i \leq N - 1,
\]

(ii) \[
\lambda y^-_N = \sum_{k=0}^{N} \sum_{n=N+1}^{\infty} (n-N) \cdot y_k D_{n-k}.
\]

**Proof.** (i) The case \( 0 \leq i \leq N - 1 \)

We note that \( \lambda \) is the average number of individual customer arrivals per slot (including rejected customers) and that \( y^-_{i,j} \) is the fraction of individual customers who find \( i \) customers in the system with the phase of the UMC being \( j \) just after the arrivals of their batches. That is, \( \lambda y^-_{i,j} \) is the average number of individual customer arrivals of such kind per slot. This number obviously equals the average number of batch arrivals per slot that make transitions from \( \{k,l\} | k \leq i, 1 \leq l \leq m \} \) to \( \{(n,j) | n \geq i+1\} \). Thus we have, for \( 0 \leq i \leq N - 1 \) and \( 1 \leq j \leq m \),

\[
\lambda y^-_{i,j} = \sum_{k=0}^{i} \sum_{n=k+1}^{m} \sum_{l=1}^{m} y_{k,l} (D_{n-k})_l,
\]

from which (9) follows.

(ii) The case \( i = N \)

Under the PRP, all rejected customers find \( N \) customers in the system when they arrive (see Remark 2). Thus we have, in a similar manner,
\[
\lambda y_{N,j}^- = \sum_{k=0}^{N} \sum_{n=N+1}^{\infty} \sum_{j=1}^{\infty} (n-N) \cdot y_{k,l}^j (D_{n-k})_j,
\]

from which (10) follows.

Note that \( y_{N,e}^- ( = x_N e ) \) is the probability of an individual customer being rejected. Also note that when the arrival process is a Bernoulli process with parameter \( \lambda \), (9) and (10) simplify to \( y_i^- = y_i \), \( 0 \leq i \leq N \), which is a well known property called BASTA (Bernoulli arrivals see time averages, Boxma and Groenendijk [3]).

Now, multiplying both sides of (9) by \( z^i \) and summing up for all \( i \geq 0 \), we have

**Corollary 3.** For a stable infinite-capacity queue with D-BMAP arrivals under the assumption of LAS-DA, \( Y(z) \) and \( Y^-(z) \) are related by

\[
Y^-(z) = Y(z) \frac{D-D(z)}{\lambda(1-z)},
\]

where \( Y^-(z) = \sum_{i=0}^{\infty} y_i^- z^i \).

We remark that the relationships in Theorems 2 and 3 and Corollaries 2 and 3 are the same as the continuous-time BMAP counterparts [8].

4. Relationships in the Early Arrival System (EAS)

In this section, we consider the EAS (or departures first policy), where customers arrive early during a slot. Suppose that the UMC is in some phase \( i \) at an exact slot boundary. Then, under the EAS, with respective probabilities \( (D_i)_j \) and \( (D_0)_j \), there is a phase transition to \( j \) with a batch arrival of size \( k \geq 1 \) and without an arrival, just after the slot boundary. Just prior to the end of the slot, there may be a departure of customer(s), if any. Note again that nothing could happen somewhere in the middle of a slot. (For more details on the EAS, see [7, p.193] and [12, p.8].)
In addition to the probabilities defined in Section 1, we define for the EAS the following probabilities in number-average sense, for \( 0 \leq i \leq N \) and \( 1 \leq j \leq m \):

\( w_{i,j} \): stationary probability that the process is in state \((i, j)\) at an exact slot boundary

\[ w_i = (w_{i,1}, \ldots, w_{i,m}) \]

(We interpret \( w_{i,j} \) as the long-run fraction of slot boundaries at which the process is in state \((i, j)\).)

Note that the values of \( y_i, x_i, y_i^-, y_i^-, \) and \( w_i \) under the assumption of the EAS are, in general, different from those under the LAS-DA (see Remark 3 below). That is, they depend on which assumption they are under. Their relationships, however, are similar to each other, as discussed below.

We note first that \( w_i \) under the EAS corresponds to the queue length just before transition epochs of the UMC, just like \( y_i \) under the LAS-DA. Then we replace “LAS-DA”, \( y_{i,j}, x_i, y_i^-, y_i^- \), and \( Y(z) \) in Sections 2 and 3 with “EAS”, \( w_{i,j}, w_i \), and \( W(z) = \sum_{i=0}^{\infty} w_i z^i \), respectively. With this replacement, it is not difficult to see that all the theorems and corollaries stated for the LAS-DA still hold for the EAS as well. Finally, the following relationship between \( w_i \) and \( y_i \), both under the EAS, is obvious (we omit the proof):

**Theorem 4.** For a finite-capacity queue with D-BMAP arrivals under the assumptions of EAS and PRP, \( w_i \) and \( y_i \) are related by

(i) \[ y_i = \sum_{k=0}^{i} w_k D_{i-k}, \quad 0 \leq i \leq N-1, \]

(ii) \[ y_N = \sum_{k=0}^{N} \sum_{n=k}^{\infty} w_k D_{n-k}. \]

In case of a stable infinite-capacity queue, \( Y(z) \) and \( W(z) \) are related by

(iii) \[ Y(z) = W(z)D(z). \]
Remark 3. For a simple model such as an infinite-capacity queue without vacations, \( y_i \) under the EAS is the same as that under the LAS-DA. This is because the queue length in the middle of a slot is not affected by the order of the arrival and the departure at the preceding slot boundary [12, p.38]. That is, sample paths during a slot under the two assumptions are exactly the same [7, p.194]. This is not the case, however, when it comes to finite capacity queues or queues with vacations. In such queues, customers who can be accepted under the EAS may not be accepted under the LAS, or the server eligible for a vacation under the EAS may not be under the LAS. That is, sample paths may evolve differently according to the assumption of the order of the arrival and the departure simultaneously taking place at a slot boundary.

5. Concluding Remarks

For finite- and infinite-capacity queues with D-BMAP arrivals, we derived relationships among \( y_i, x_i, y_i^+, \) and \( y_i^- \), \( 0 \leq i \leq N \), under the LAS-DA (or EAS) and PRP. Because we did not assume any particular service mechanism, these relationships hold for a broad class of discrete-time queues with D-BMAP arrivals. Suppose one has the solution for \( x_i, 0 \leq i \leq N \), of a D-BMAP//GV queue with the threshold value of activating the server \( a \), the maximum size of a batch for service \( b \), the number of servers \( c \), and generalized vacations (GV). (Usually, \( x_i \) of a simple model can be obtained by considering the embedded Markov chain at departures (see, e.g., [1, 6])). Then, through the relationships, it is immediate to have \( y_i, y_i^+, \) and \( y_i^- \). Without these relationships, such as (1) for example, it may require lengthy calculations to obtain \( y_i \) from \( x_i \) (see, e.g., [1, 6]).

Another way to utilize these relationships is to combine them with other available equations. One can then solve these equations simultaneously to obtain the desired distributions. To show how it works, we consider a stable D-BMAP/G/1 queue under the LAS-DA as an example. Let \( S \) be an independent and identically distributed...
random variable of service time with \( \text{Pr}(S = k) = s_k, \ k \geq 1 \). Applying the so-called arrival time approach (Chae et al. [5]), one can obtain the following equations:

\[
\begin{align*}
Y(z) &= (1 - \rho)g + \rho X^s(z)A_E(z), \\
X(z) &= X^s(z)A(z)z^{-1},
\end{align*}
\]

where

\[
\rho = \lambda E(S) > 0,
\]

which represents the probability that the server is busy at an arbitrary slot,

\( g \) is the probability vector that represents the distribution of the phase of the UMC when the server is idle. (It can be shown that \( g \) is the invariant probability vector of \( G \), i.e., \( gG = g \) and \( gE = 1 \), where \( G \) represents the transition probability matrix of the phase change during the so-called fundamental period (Neuts [10, p.6]). This is a discrete-time analogue of the continuous-time BMAP result (Lucantoni [9]).)

\( X^s(z) \) is the generating function of the joint probability vector of the number of customers in the system and the phase of the UMC just after a service start epoch,

\[
A(z) = \sum_{k=0}^{\infty} s_k D(z)^k,
\]

which represents the matrix generating function of the joint probability of the number of customers arriving during a service time and the phase change of the UMC during the service time, and

\( A_E(z) \) represents the matrix generating function of the joint probability of the number of customers arriving during elapsed slots since the beginning of the ongoing service and the phase change of the UMC during these elapsed slots. It is easy to show that \( A_E(z) \) has the following relation to \( A(z) \):

\[
A_E(z)(I - D(z)) = E(S)^{-1}(I - A(z)).
\]

Note that (11) is obtained by conditioning on \( Y(z) \) by whether the server is busy or not: in particular, under the condition that the server is busy at an arbitrary slot, the number of customers in the system at this slot equals the number at the beginning of the current service plus the number of customers arriving during the elapsed slots since the
beginning of this service. Then considering the phase change during these elapsed slots, (11) follows. Also note that (12) is obtained from the obvious relation that the number of customers just after a service completion epoch equals the number at the beginning of the service plus the number of customers arriving during the service time minus one (for the departing customer). For details, see [5].

Now simultaneously solving (11), (12) and \( Y(z)(D(z) - I) = \lambda(z - 1)X(z) \) (see (7)), we finally have
\[
Y(z)(A(z) - zI) = (1 - \rho)(1 - z)gA(z),
\]
\[
X(z)(A(z) - zI) = \lambda^{-1}(1 - \rho)gA(z)(I - D(z)),
\]
\[
X^{\alpha}(z)(A(z) - zI) = \lambda^{-1}(1 - \rho)zg(I - D(z)).
\]
(See, e.g., [9] for the continuous counterparts of these results.) \( Y^<^()(z) \) and \( Y^<(z) \) then follow from Corollaries 2 and 3.

We hope that our intuitive derivations and the results obtained would help readers better understand the discrete-time queues with D-BMAP arrivals.
References