International Journal of Management Science
Vol 9, No 2, November 2003

A Note on the Relationships among the Queue Lengths at Various Epochs of a Queue with BMAP Arrivals*

Nam K. Kim**

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

Kyung C. Chae***

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

Ho W. Lee****

School of Systems Management Engineering,
Sung Kyun Kwan University, Suwon, 440-746, Korea

(Received Oct. 2001, Revised July 2003, Accepted Aug. 2003)

ABSTRACT

For a stationary queue with BMAP arrivals, Takine and Takahashi [8] present a relationship between the queue length distributions at a random epoch and at a departure epoch by using the rate conservation law of Miyazawa [6]. In this note, we derive the same relationship by using the elementary balance equation, "rate-in = rate-out." Along the same lines, we additionally derive relationships between the queue length distributions at a random epoch and at an arrival epoch. All these relationships hold for a broad class of finite- as well as infinite-capacity queues with BMAP arrivals.

* The authors would like to thank Dr. Yoshitaka Takahashi of Waseda University and Dr. Tetsuya Takine of Kyoto University for their encouraging comments

** Corresponding Author, Email: freedom@kaist.ac.kr

*** Email: kcchae@kaist.ac.kr

**** This research was supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University. Email: hwlee@skku.edu

1. INTRODUCTION AND NOTATION

In this note, we consider the stationary queue lengths at various epochs of a queue with a *batch Markovian arrival process* (BMAP). For this queue, Takine and Takahashi [9] give a rigorous proof of a relationship between the queue length distributions at a random epoch and at a departure epoch by using the *rate conservation law* of Miyazawa [7]. In this paper, we derive the same relationship, using the elementary balance equation, "rate in = rate out." Along the same lines, we additionally derive relationships between the queue length distributions at a random epoch and at an arrival epoch. Relationships of this kind are useful in the sense that once one obtains a solution for a certain epoch, e.g., for a departure epoch, it is immediate to get solutions for the other epochs through these relationships.

Remark 1: The BMAP is a versatile process that contains a large class of point processes such as a Poisson process, a Markov modulated Poisson process, and a batch Poisson process with correlated batch arrivals (see, e.g., Lucantoni [5, 6]). Furthermore, we (as well as Takine and Takahashi) do not assume any particular service mechanism, e.g., service discipline, number of servers, batch service, generalized vacations, etc. Thus the relationships presented in this paper hold for a broad class of queues with BMAP arrivals.

In a queue with BMAP arrivals, groups of customers of size k arrive according to a BMAP with representation $\{\mathbf{D}_k, k \geq 0\}$, where $\mathbf{D}_k$ 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. Specifically, suppose that the UMC is in some phase $1 \leq i \leq m$ at a certain epoch. Then, with transition rate $(\mathbf{D}_k)_{i,j}$, there is a phase transition to j with a batch arrival of size $k \geq 1$; with transition rate $(\mathbf{D_0})_{i,j}$, there is a phase transition to $j \neq i$ without an arrival; $(-\mathbf{D_0})_{i,i}$, in particular, is defined to be $(\sum_{k=1}^{\infty}\sum_{j=1}^{m}(\mathbf{D}_k)_{i,j} + \sum_{j \neq i}(\mathbf{D_0})_{i,j})$, so that $\mathbf{D} = \sum_{k=0}^{\infty}\mathbf{D}_k$ can be the *generator* of the UMC. (For details on BMAP, see Lucantoni [5, 6].) Let λ and λ_g denote the arrival rates of individual customers and batches (or groups), respectively. Then they are given by

$$\lambda = \boldsymbol{\pi}\sum_{k=1}^{\infty} k\mathbf{D}_k\mathbf{e},$$

$$\lambda_g = \boldsymbol{\pi}\sum_{k=1}^{\infty}\mathbf{D}_k\mathbf{e} ,$$

where $\boldsymbol{\pi}$ is the stationary probability vector of the UMC, and $\mathbf{e}$ is a column vector of 1's.

Now, we consider a bivariate process $\{L(t),S(t),t \geq 0\}$, where $L(t)$ denotes the number of customers in the system and $S(t)$ the phase of the UMC at $t \geq 0$. We define the following probabilities for $i \geq 0$ and $1 \leq j \leq m$:

$y_{i,j}$: stationary probability that the process is in state (i,j) *at a random epoch* in time

$\mathbf{y}_i = (y_{i,1}, \cdots, y_{i,m})$

(We interpret $y_{i,j}$ as the average amount of time the process spends in (i,j) per unit time.)

$y_{i,j}^{g-}$: stationary probability that a *batch* finds i customers in the system when it arrives and the UMC is in phase j just after its arrival

$\mathbf{y}_i^{g-} = (y_{i,1}^{g-}, \cdots, y_{i,m}^{g-})$

(We interpret $y_{i,j}^{g-}$ 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 those who precede her in her own batch – see Remark 2 below) when she arrives and the UMC is in phase j just after the arrival of her batch

$\mathbf{y}_i^{-} = (y_{i,1}^{-}, \cdots, y_{i,m}^{-})$

(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 those, if any, who follow her to depart at the same time – see Remark 2 below) when she departs from the system and the UMC is in phase j just after her departure

$\mathbf{x}_i = (x_{i,1}, \cdots, x_{i,m})$

(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 [11, p.388]) that customers arriving (departing) at the same time 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 batch departures, such as in a batch-service queue, we suppose that customers departing at the same time form a line to leave the system one after another.

Throughout the note, we assume a stable system so that the rates and the probabilities defined above are all well defined either in time-average or customer-average sense. It is also assumed that the phase of the UMC just before a departure is identical to that just after the departure. In cases where a phase transition of the UMC (with and without an arrival) and departure(s) happens at the same time, we assume that the phase transition occurs just before the departure(s).

2. A RELATIONSHIP BETWEEN THE QUEUE LENGTHS AT A RANDOM EPOCH AND AT A DEPARTURE EPOCH

In this section, we derive the relationship between the queue length distributions at a random epoch and at a departure epoch, using the elementary balance equation, "rate-in = rate-out." Especially for those queues in which customers may arrive and/or depart at the same time, we adopt the following convention, along the same lines as those discussed in Remark 2:

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, $(i+1,k)$, $(i+2,k)$, $\cdots$, $(i+l-1,k)$. 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 $(i+2,k),\cdots,(i+l-1,k)$ 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, $(i-1,j)$, $(i-2,j),\cdots$, $(i-l+1,j)$.

This convention makes the process $\{L(t),S(t),t\geq 0\}$ *skip-free* with respect to $L(t)$. We need this skip-free convention in order to represent the transition rates into and out of 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 instantaneous one-step transitions). Now, we equate the transition rates into and out of state (i,j). As a result, we have

Theorem 1. (Takine and Takahashi [9]) For a queue with BMAP arrivals, $\mathbf{Y}(z)$ and $\mathbf{X}(z)$ are related by

$$\mathbf{Y}(z)\mathbf{D}(z)=\lambda(z-1)\mathbf{X}(z), \tag{1}$$

where $\mathbf{Y}(z)=\sum_{i=0}^{\infty}\mathbf{y}_i z^i$, $\mathbf{X}(z)=\sum_{i=0}^{\infty}\mathbf{x}_i z^i$, and $\mathbf{D}(z)=\sum_{i=0}^{\infty}\mathbf{D}_k z^k$.

Before presenting a proof of this theorem, we give an explanation of the way we calculate the rate of the transitions caused by customer departures. Note that λ is the average number of individual customers departing per unit time and that $x_{i-1,j}$ is the fraction of those who leave behind the system with state $(i-1,j)$. Thus, $\lambda x_{i-1,j}$ can be interpreted as the average number of individual customers per unit time 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. That is, $\lambda x_{i-1,j}$ is the transition rate from (i,j) to $(i-1,j)$ caused by individual customer departures.

Proof. For the process $\{L(t),S(t),t\geq 0\}$, it is obvious that the transition rates into and out of (i,j) are equal for all $i\geq 0$ and $1\leq j\leq m$. We note that these transitions are caused either by a phase transition of the UMC (with or without an arrival) or by a departure.

We first consider the transition rate out of (i,j). The out-rate caused by phase transitions of the UMC is given by $y_{i,j}\cdot(-\mathbf{D}_0)_{j,j}+X$, where X is the rate of giant transitions instantaneously going through (i,j). (That is, X is the total transition rate from $\{(i-n,l)\mid n>0, 1\le l\le m\}$ to $\{(i-n',j)\mid n'>0\}$ caused by batch arrivals. Under the skip-free convention, this rate of instantaneous transitions should be considered as a part of the transition rates into and out of (i,j). However, we do not have to elaborate more on X because, whatever it turns out to be, it cancels out in the balance equation.) In addition, the rate out of (i,j) caused by departures is given by $\lambda x_{i-1,j}$, as discussed earlier. Thus we have, for $i\ge 0$ and $1\le j\le m$,

$$\text{transition rate out of } (i,j) = \{y_{i,j}\cdot(-\mathbf{D}_0)_{j,j} + X\} + \lambda\, x_{i-1,j}, \tag{2}$$

where $x_{-1,j}=0$.

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 $(k,l)\ne(i,j)$ 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 $i\ge 0$ and $1\le j\le m$,

$$\text{transition rate into } (i,j) = \left\{\sum_{k=0}^{i}\sum_{l=1}^{m} y_{k,l}(\mathbf{D}_{i-k})_{l,j} - y_{i,j}(\mathbf{D}_0)_{j,j} + X\right\} + \lambda x_{i,j}. \tag{3}$$

Now equating (2) and (3), we have

$$\sum_{k=0}^{i}\mathbf{y}_k\mathbf{D}_{i-k} = \lambda\,\mathbf{x}_{i-1} - \lambda\,\mathbf{x}_i. \tag{4}$$

Finally, multiplying both sides of (4) by z^i and summing up for all $i\ge 0$, we have the desired result. **Q.E.D.**

When the arrivals are Poisson with rate λ (in this case, $\mathbf{D}_0=-\lambda$, $\mathbf{D}_1=-\lambda$, and $\mathbf{D}_k=\mathbf{0}$, $k\ge 2$), (1) simplifies to $\mathbf{y}_i=\mathbf{x}_i$, $i\ge 0$.

3. RELATIONSHIPS BETWEEN THE QUEUE LENGTHS AT A RANDOM EPOCH AND AT AN ARRIVAL EPOCH

In this section, we present relationships between the stationary queue lengths at a random epoch and at an arrival epoch. 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 a random epoch and the queue length found by a batch arrival:

Theorem 2. For a queue with BMAP arrivals, $\mathbf{Y}(z)$ and $\mathbf{Y}^{g-}(z)$ are related by

$$\mathbf{Y}^{g-}(z) = \mathbf{Y}(z)\frac{\mathbf{D}-\mathbf{D}_0}{\lambda_g}, \tag{5}$$

where $\mathbf{Y}^{g-}(z) = \sum_{i=0}^{\infty} \mathbf{y}_i^{g-} z^i$.

Proof. We note that λ_g is the average number of batch arrivals per unit time and that $y_{i,j}^{g-}$ is the fraction of batches that find i customers in the system and make transitions of the UMC into j. Thus, $\lambda_g y_{i,j}^{g-}$ can be interpreted as the average number of batch arrivals of such kind per unit time. This number obviously equals the total transition rate out of $\{(i,l) \mid 1 \le l \le m\}$ caused by batch arrivals that make transitions of the UMC into j. Thus we have, for $i \ge 0$ and $1 \le j \le m$,

$$\lambda_g y_{i,j}^{g-} = \sum_{k=1}^{\infty}\sum_{l=1}^{m} y_{i,l}(\mathbf{D}_k)_{l,j},$$

which can be rewritten in a compact form as follows:

$$\lambda_g \mathbf{y}_i^{g-} = \mathbf{y}_i(\mathbf{D}-\mathbf{D}_0). \tag{6}$$

Now, multiplying both sides of equation (6) by z^i and summing up for all $i \ge 0$, we have the desired result. **Q.E.D.**

Next, we present a relationship between the stationary queue length at a random epoch and the queue length found by an individual customer arrival (including those who precede her in her own batch).

Theorem 3. For a queue with BMAP arrivals, $\mathbf{Y}(z)$ and $\mathbf{Y}^{-}(z)$ are related by

$$\mathbf{Y}^{-}(z) = \mathbf{Y}(z)\frac{\mathbf{D}-\mathbf{D}(z)}{\lambda(1-z)}, \tag{7}$$

where $\mathbf{Y}^{-}(z) = \sum_{i=0}^{\infty} \mathbf{y}_i^{g} z^i$.

Proof. We note that λ is the average number of individual customers arriving per unit time and that $y_{i,j}^-$ is the fraction of those who find i customers in the system with the phase of the UMC being j just after the arrivals of their batches. Thus, $\lambda y_{i,j}^-$ can be interpreted as the average number of individual customer arrivals of such kind per unit time. This number obviously equals the average number of *batch arrivals* per unit time that make transitions from $\{(k,l) \mid k \le i, 1 \le l \le m\}$ to $\{(n,j) \mid n \ge i+1\}$. Thus we have, for $i \ge 0$ and $1 \le j \le m$,

$$\lambda\, y_{i,j}^- = \sum_{k=0}^{i} \sum_{n=i+1}^{\infty} \sum_{l=1}^{m} y_{k,l} (\mathbf{D}_{n-k})_{l,j} ,$$

which can be rewritten in a compact form as follows:

$$\lambda\, \mathbf{y}_i^- = \sum_{k=0}^{i} \sum_{n=i+1}^{\infty} \mathbf{y}_k \mathbf{D}_{n-k} . \tag{8}$$

Now, multiplying both sides of (8) by z^i and summing up for all $i \ge 0$, we have the desired result. **Q.E.D.**

Note that when arrivals are Poisson with parameter λ, (7) simplifies to $\mathbf{y}_i^- = \mathbf{y}_i$, $i \ge 0$, which is a result of the well-known property called *PASTA* (*Poisson arrivals see time averages* [10]). We note that Lucantoni [6] establishes relationships, which are similar to those presented in Theorems 2 and 3, between the virtual waiting time and the actual waiting time of the BMAP/G/1 queue.

4. A NOTE ON A FINITE-CAPACITY QUEUE WITH BMAP ARRIVALS

In this section, we consider a finite-capacity queue with BMAP arrivals. Let N be the system capacity, such that at most N customers can be accommodated in the system. We assume the *Partial Rejection Policy* (PRP; see, e.g., Takagi [8, p.412]), in which only those individual customers who find the system full are rejected. Specifically, suppose that a batch of size k arrives when there are i customers in the system. In the case of $i+k > N$, the first $k-(N-i)$ individual customers of the batch are accepted, while the last $(i+k)-N$ individual customers are rejected. These rejected customers are supposed to find more than or equal to N customers in the system (including their preceding customers in their own batch) and depart from the system immediately after their arrivals. They are also supposed to leave behind those customers who follow them in the same arriving batch, *in addition to* N customers who have already been accommodated in the system (see Remark 2).

As discussed in Takine and Takahashi [9], a finite-capacity queue can be considered as a queue having a service mechanism in which an individual arriving customer who finds more than or equal to N customers in the system (including those who precede her in her batch) is expelled immediately after her arrival. Thus the relationships derived in Theorems 1 through 3 have already included the finite-capacity queue as well because we have not assumed any particular service mechanism to derive them. Among others, we present a finite-capacity version of Theorem 1 as follows:

Corollary 1. For a finite-capacity queue with BMAP arrivals under the PRP, $\mathbf{y}_i$ and $\mathbf{x}_i$ are related by

$$\sum_{k=0}^{i}\mathbf{y}_k\mathbf{D}_{i-k} = \lambda\mathbf{x}_{i-1} - \lambda\mathbf{x}_i\ , \quad 0 \le i \le N-1, \text{ and} \tag{9}$$

$$\sum_{k=0}^{N}\sum_{i=N}^{\infty}\mathbf{y}_k\mathbf{D}_{i-k} = \lambda\mathbf{x}_{N-1}\ . \tag{10}$$

Proof. As discussed earlier, (4) still holds for all $i \ge 0$ even in the case of a finite-capacity queue. Note, however, that $\mathbf{y}_k = \mathbf{0}$ for $k \ge N+1$, and that $\mathbf{x}_i$ for $i \ge N$ corresponds to the departures by rejected customers. While (9) is the same as (4), (10) is obtained by summing up both sides of (4) for all $i \ge N$. (This proof is based on the comment made by an anonymous referee of the earlier version of this paper.) **Q.E.D.**

Remark 3: Corollary 1 cannot be said to be new because it is a special case of the general relationship (Theorem 1). However, Corollary 1 may deserve an explicit presentation as above for its potential usefulness in the analysis of finite-capacity queues with BMAP arrivals. Specifically, note that $\lambda\mathbf{x}_i$, $0 \le i \le N-1$, of (9) and (10) can be replaced by $\lambda^*\mathbf{x}_i^*$, where $\lambda^* = \lambda\sum_{i=0}^{N-1}\mathbf{x}_i\mathbf{e}$ is the so-called effective arrival (or departure) rate and $\mathbf{x}_i^* = \mathbf{x}_i \big/ \sum_{i=0}^{N-1}\mathbf{x}_i\mathbf{e}$ is the probability vector of the queue length left behind by an accepted customer. This replacement is useful because $\mathbf{x}_i^*$ is easier to obtain through the analysis of the Markov chain embedded at departures than $\mathbf{x}_i$. Also, note that λ^* can be represented in terms of $\mathbf{x}_i^*$ (in the BMAP/G/1/N queue, for example, the mean inter-departure time of accepted customers is given by $(\lambda^*)^{-1} = E(S) + \mathbf{x}_0^*(-\mathbf{D}_0^{-1})\mathbf{e}$ (Gupta and Laxmi [1])). Once λ^* and $\mathbf{x}_i^*$ are obtained, $\mathbf{y}_i$ follows from the relationships (9) and (10) that are rewritten in terms of λ^* and $\mathbf{x}_i^*$ (in place of (10), one may want to use $\sum_{i=0}^{N} \mathbf{y}_i = \pi$).

5. CONCLUDING REMARKS

For a queue with BMAP arrivals, we have derived relationships among $\mathbf{y}_i$, $\mathbf{x}_i$, $\mathbf{y}_i^{g-}$, and $\mathbf{y}_i^-$, using a simple and intuitive argument. These relationships hold for a broad class of finite- as well as infinite-capacity queues with BMAP arrivals because we do not assume any particular service mechanism. Suppose one has a solution for $\mathbf{x}_i$ of a certain queueing system with BMAP arrivals (usually, $\mathbf{x}_i$ or $\mathbf{x}_i^*$ can be obtained by considering the embedded Markov chain at departures (see, e.g., Kasahara et al. [2], and Lucantoni [5])). Then $\mathbf{y}_i$, $\mathbf{y}_i^{g-}$, and $\mathbf{y}_i^-$ immediately follow from the relationships. Without these relationships, such as (1) for example, one may have to go through lengthy calculations to obtain $\mathbf{y}_i$ from $\mathbf{x}_i$ (see, e.g., Lucantoni et al. [4]).

Remark 4: The same relationships with respect to the number of customers in the *queue* (excluding those in service) at various epochs can be establisheds as well. This is done by replacing all the probabilities with respect to the number in the system with the corresponding probabilities with respect to the number in the queue. For example, we define $\mathbf{Y}_Q(z)$ and $\mathbf{X}_Q(z)$ as generating functions corresponding to $\mathbf{Y}(z)$ and $\mathbf{X}(z)$, such that $\mathbf{Y}_Q(z)$ and $\mathbf{X}_Q(z)$ represent the generating functions of the probability vectors of the numbers in the *queue* at a random epoch and at the epoch of a departure from the *queue*, respectively. Then, by following the same arguments (that is, applying the balance equations to the queue rather than to the system), it is easy to have

$$\mathbf{Y}_Q(z)\mathbf{D}(z) = \lambda(z-1)\mathbf{X}_Q(z).$$

The other relationships with respect to the number in the queue can be obtained similarly. We omit the details.

We finally remark that the discrete-time analogues of the results presented in this paper have been derived by using the similar arguments for queues with D-BMAP (discrete-time BMAP) arrivals (Kim et al. [3]). We hope that our intuitive derivations and the results obtained would help readers better understand the queues with BMAP arrivals.

REFERENCES

[1] Gupta, U. C. and P. V. Laxmi, "Analysis of the MAP/G^{a,b}/1/N Queue," *Queueing systems* 38 (2001), 109–124.

[2] Kasahara, S., T. Takine, Y. Takahashi, and T. Hasegawa, "MAP/G/1 queues under N-policy with and without vacations," *Journal of the Operations Research Society of Japan* 39 (1996), 188–212.

[3] Kim, N. K., S. H. Chang and K. C. Chae, "On the relationships among queue lengths at arrival, departure, and random epochs in the discrete-time queue with D-BMAP arrivals," *Operations Research Letters* 30 (2002), 25–32.

[4] Lucantoni, D. M., K. S. Meier-Hellstern, and M. F. Neuts, "A single-server queue with server vacations and a class of non-renewal arrival processes," *Advances in Applied Probability* 22 (1990), 676–705.

[5] Lucantoni, D. M., "New results on the single server queue with a batch Markovian arrival process," *Stochastic Models* 7 (1991), 1–46.

[6] Lucantoni, D. M., "The BMAP/G/1 queue: A tutorial," in L. Donatiello and R. Nelson (eds.): *Performance Evaluation of Computer and Communications Systems*, Springer-Verlag, Berlin 1993, 330–358.

[7] Miyazawa, M., "Rate conservation laws: A survey," *Queueing Systems* 15 (1994), 1–58.

[8] Takagi, H., *Queueing Analysis, Volume 2: Finite Systems*, North-Holland, Amsterdam 1993.

[9] Takine, T. and Y. Takahashi, "On the relationship between queue lengths at a random instant and at a departure in the stationary queue with BMAP arrivals," *Stochastic Models* 14 (1998), 601–610.

[10] Wolff, R.W., "Poisson arrivals see time averages," *Operations Research* 30 (1982), 223–231.

[11] Wolff, R.W., *Stochastic Modeling and the Theory of Queues*, Prentice Hall, Englewood Cliffs, New Jersey 1989.