

SUPPLEMENTARY VARIABLE METHOD APPLIED TO THE MAP/G/1 QUEUEING SYSTEM

BONG DAE CHOI¹, GANG UK HWANG¹ and DONG HWAN HAN²

(Received 19 August 1996)

Abstract

In this paper we consider the MAP/G/1 queueing system with infinite capacity. In analysis, we use the supplementary variable method to derive the double transform of the queue length and the remaining service time of the customer in service (if any) in the steady state. As will be shown in this paper, our method is very simple and elegant. As a one-dimensional marginal transform of the double transform, we obtain the generating function of the queue length in the system for the MAP/G/1 queue, which is consistent with the known result.

1. Introduction

In B-ISDNs, the major features of input traffics are high burstiness and strong correlation between cell arrivals, and to support such characteristics the Markovian Arrival Process (MAP) and the Batch Markovian Arrival Process (BMAP) have been proposed. The BMAP is a convenient representation of the versatile Markovian point process (N-process) and a class of which includes IPP, MMPP, Phase-type Renewal Process and superpositions of such processes. Matrix-Geometric methods and Markov renewal theory are usually employed for the analysis of the MAP/G/1 queueing system. Ramaswami [12] analyzed the N/G/1 queue and Lucantoni [5] restated the results of the N/G/1 queue into BMAP notation. Lucantoni, Meier-Hellstern and Neuts [4] analyzed the MAP/G/1 queue with server vacations. Lucantoni, Choudhury and Whitt [6] considered the transient BMAP/G/1 queue and derived the two-dimensional transforms of the transient workload and queue-length distribution for the single-server queue with general service times.

Recently it has been shown that the solution of the nonlinear matrix equation arising in the MAP/G/1 queue satisfies a Matrix-Exponential form. Ramaswami [13]

¹Center for Applied Mathematics and Department of Mathematics, Korea Advanced Institute of Science and Technology, 373-1 Kusong-Dong, Yusong-Gu, Taejon, 305-701, Korea

²Department of Mathematics, Sun Moon University, Asan-Kun, Chungnam, 337-840, Korea

© Australian Mathematical Society, 1998, Serial-fee code 0334-2700/98

considered the Matrix-Exponential form for the N/G/1 and GI/N/1 queue. Lucantoni [5] considered the Matrix-Exponential form for the BMAP/G/1 queue. Takine and Hasegawa [17] analyzed the workload in the MAP/G/1 queue with state-dependent services by using Matrix-Exponential form. Lucantoni and Neuts [7] obtained the Matrix-Exponential form for the MAP/G/1 and the MAP/SM/1 queues with less calculation effort than in earlier proofs.

However to the best of authors' knowledge, there is no study of applying the supplementary variable method to the analysis of the MAP/G/1 queue. The supplementary variable method is known as a simple and convenient way of deriving the double transform of the queue length and the supplementary variable used. We have two kinds of supplementary variable, in general, one of which is the elapsed service time and the other is the remaining service time, and for both cases the approaches of deriving the double transforms are different. When we use the elapsed service time as supplementary variable, the hazard rate function and the boundary conditions are needed. However, when we use the remaining service time as supplementary variable, all of these are not needed and we have much simpler calculations. In this paper we use the remaining service time as our supplementary variable.

The main work of our paper is to apply the supplementary variable method to the derivation of the double transform for the MAP/G/1 queueing system. As one dimensional marginal transform of the double transform, we obtain the generating function of the number of customers in the system for the MAP/G/1 queue, which is in accord with a known result. It is shown that, when the arrival process is a Poisson process, our result is reduced to the double transform of the queue length and the remaining service time of the customers in service for the M/G/1 queue. As will be shown later, our method is very simple and elegant. So we expect that the supplementary variable method can be widely used in the analysis of variants of the MAP/G/1 queue.

This paper is organized as follows. In Section 2, we review MAP. In Section 3, we derive the Chapman-Kolmogorov matrix differential equation for the queue length and the remaining service time, and find the double transform of the queue length and the remaining service time of the customer in service (if any) in the steady state. As a corollary of our main result, we derive the generating function of the number of customers in the steady state.

2. Markovian Arrival Process

To introduce the Markovian Arrival Process, first consider a Poisson arrival process with rate λ . Let A_t denote the number of customers arriving during $[0, t]$. The Poisson process $\{A_t\}$ is a Markov process with state space $\{i | i \geq 0\}$ and infinitesimal generator

of the form

$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The MAP is a natural generalization of the above Markov process where arrivals are governed by an underlying m -state Markov chain, and diagonal elements and upper diagonal elements of the above infinitesimal generator are replaced by $m \times m$ matrices $C = [C_{ij}]$ and $D = [D_{ij}]$, respectively. The matrix C has negative diagonal elements and nonnegative off-diagonal elements and D has nonnegative elements. Here C_{ij} , $i \neq j$ is the state transition rate from state i to state j in the underlying Markov chain without an arrival and D_{ij} is the state transition rate from state i to state j in the underlying Markov chain with an arrival. Let A_t be the number of customers arriving during $[0, t]$, and J_t indicate the state of the underlying Markov chain at time t with state space $\{1, 2, \dots, m\}$. Then $\{A_t, J_t\}$ is a two-dimensional Markov process on state space $\{(i, j) | i \geq 0, 1 \leq j \leq m\}$ with infinitesimal generator

$$\begin{pmatrix} C & D & 0 & 0 & \cdots \\ 0 & C & D & 0 & \cdots \\ 0 & 0 & D & D & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Then $\{A_t, J_t\}$ is called the Markovian Arrival Process (MAP). Since the above matrix is the infinitesimal generator of the MAP, we have

$$(C + D)e = 0,$$

where e is an $m \times 1$ column vector whose elements are all equal to 1. Furthermore, $C + D$ is the infinitesimal generator of the underlying Markov chain $\{J_t\}$. We assume the underlying Markov chain is irreducible, so that there exists a stationary probability vector π such that

$$\pi(C + D) = 0, \quad \pi e = 1.$$

We assume that the service times of customers are i. i. d. with distribution function $G(x)$ and density function $g(x)$. Let ρ be the traffic intensity, that is,

$$\rho = \pi D e \int_0^{\infty} x g(x) dx.$$

Throughout this paper we assume $\rho < 1$ to guarantee the stability of our system.

3. MAP/G/1 Queueing system

In this section we consider the MAP/G/1 queueing system. The capacity of the queue is assumed to be infinite. Let N_t and J_t be the queue length and the state of the underlying Markov chain of the MAP at time t . Let X_t be the remaining service time of customer in service (if any) at time t . Let ξ_t be the number of busy servers at time t , that is,

$$\xi_t = \begin{cases} 1, & \text{if the server is busy,} \\ 0, & \text{if the server is idle.} \end{cases}$$

Define

$$\pi_i(n, x; t)\Delta x = P\{N_t = n, x < X_t < x + \Delta x, J_t = i, \xi_t = 1\}, \quad n \geq 0, x \geq 0$$

and

$$p_i(t) = P\{J_t = i, \xi_t = 0\}.$$

Then we have the following Chapman-Kolmogorov equations. When $n = 0$ and $1 \leq i \leq m$,

$$\begin{aligned} \pi_i(0, x - \Delta t; t + \Delta t) &= \pi_i(0, x; t) [1 + C_{ii}\Delta t] + \sum_{j \neq i} \pi_j(0, x; t)C_{ji}\Delta t \\ &\quad + \pi_i(1, 0; t)g(x)\Delta t + \sum_j p_j(t)D_{ji}g(x)\Delta t + o(\Delta t). \end{aligned} \tag{1}$$

When $n \geq 1$ and $1 \leq i \leq m$,

$$\begin{aligned} \pi_i(n, x - \Delta t; t + \Delta t) &= \pi_i(n, x; t)[1 + C_{ii}\Delta t] + \sum_{j \neq i} \pi_j(n, x; t)C_{ji}\Delta t \\ &\quad + \pi_i(n + 1, 0; t)g(x)\Delta t + \sum_j \pi_j(n - 1, x; t)D_{ji}\Delta t + o(\Delta t). \end{aligned} \tag{2}$$

For $p_i(t)$ we get

$$p_i(t + \Delta t) = p_i(t) [1 + C_{ii}\Delta t] + \sum_{j \neq i} p_j(t)C_{ji}\Delta t + \pi_i(0, 0; t)\Delta t + o(\Delta t). \tag{3}$$

From equations (1), (2) and (3) we have the following system of partial differential equations. For $n \geq 0$ and $1 \leq i \leq m$,

$$\begin{aligned} -\frac{\partial}{\partial x}\pi_i(n, x; t) + \frac{\partial}{\partial t}\pi_i(n, x; t) &= \sum_j \pi_j(n, x; t)C_{ji} + \pi_i(n + 1, 0; t)g(x) + \sum_j \pi_j(n - 1, x; t)D_{ji} \end{aligned} \tag{4}$$

and

$$\frac{d}{dt} p_i(t) = \sum_j p_j(t) C_{ji} + \pi_i(0, 0; t), \tag{5}$$

where $\pi_j(-1, x; t) = p_j(t)g(x)$.

Let $\pi(n, x; t)$, $\pi(n, 0; t)$ and $p(t)$ be row vectors whose i -th components are $\pi_i(n, x; t)$, $\pi_i(n, 0; t)$ and $p_i(t)$, respectively. The above equations (4) and (5) can be rewritten in matrix notation as

$$\begin{aligned} -\frac{\partial}{\partial x} \pi(n, x; t) + \frac{\partial}{\partial t} \pi(n, x; t) \\ = \pi(n, x; t)C + \pi(n + 1, 0; t)g(x) + \pi(n - 1, x; t)D \end{aligned} \tag{6}$$

and

$$\frac{d}{dt} p(t) = p(t)C + \pi(0, 0; t). \tag{7}$$

Since we assume $\rho < 1$, the limits of $\pi_i(n, x; t)$ and $p_i(t)$ exist. Define, for $n \geq 0$, $1 \leq i \leq m$ and $x \geq 0$,

$$\begin{aligned} \pi_i(n, x) &= \lim_{t \rightarrow \infty} \pi_i(n, x; t), \\ \pi(n, x) &= [\pi_1(n, x), \dots, \pi_m(n, x)] \end{aligned}$$

and

$$p_i = \lim_{t \rightarrow \infty} p_i(t), \quad p = [p_1, \dots, p_m].$$

From (6) and (7) we have, for $n \geq 0$,

$$-\frac{\partial}{\partial x} \pi(n, x) = \pi(n, x)C + \pi(n + 1, 0)g(x) + \pi(n - 1, x)D \tag{8}$$

and

$$\pi(0, 0) = -pC, \tag{9}$$

where $\pi(-1, x) = pg(x)$. Define the generating function $\Pi(z, x)$ of $\pi(n, x)$ by

$$\Pi(z, x) = \sum_{n=0}^{\infty} z^n \pi(n, x),$$

for $0 < z < 1$ and $x \geq 0$. By multiplying z^n on both sides of (8) and summing over n , we get

$$\begin{aligned} -\frac{\partial}{\partial x} \Pi(z, x) &= \Pi(z, x)[C + zD] + \frac{1}{z} [\Pi(z, 0) - \pi(0, 0)] g(x) \\ &\quad + pDg(x), \quad 0 < z < 1. \end{aligned} \tag{10}$$

Let $\Pi^*(z, s)$ be the Laplace Transform of $\Pi(z, x)$ for $s > 0$, that is,

$$\Pi^*(z, s) = \int_0^{\infty} \Pi(z, x)e^{-sx} dx \quad \text{for } s > 0.$$

By taking the Laplace transform in (10) we get

$$-\Pi^*(z, s)s + \Pi(z, 0) = \Pi^*(z, s)[C + zD] + \frac{1}{z} [\Pi(z, 0) - \pi(0, 0)] G^*(s) + pDG^*(s),$$

that is,

$$\Pi^*(z, s)[sI + C + zD] = \Pi(z, 0) \left[I - \frac{1}{z} G^*(s)I \right] + \frac{1}{z} \pi(0, 0) G^*(s) - pDG^*(s). \quad (11)$$

In above, $G^*(s)$ denotes the Laplace transform of $g(x)$. Since $\pi(0, 0) = -pC$,

$$\Pi^*(z, s)[sI + C + zD] = \Pi(z, 0) \left[I - \frac{1}{z} G^*(s)I \right] - p \left[\frac{1}{z} C + D \right] G^*(s). \quad (12)$$

Next, to find $\Pi(z, 0)$ we need to extend the domain of the Laplace transform to the region

$$\mathfrak{S} = \left\{ S \mid S = [S_{ij}] \text{ is an irreducible real matrix such that } -S_{ij} \geq 0 \text{ for } i \neq j, \text{ and } \sum_j (-S_{ij}) \leq 0 \text{ with strict inequality for some } i \right\}.$$

Let $G^*(S)$ and $\Pi^*(z, S)$ be the Laplace transforms of $g(x)$ and $\pi(z, x)$ defined on \mathfrak{S} , respectively, that is,

$$G^*(S) = \int_0^{\infty} g(x)e^{-Sx} dx,$$

$$\Pi^*(z, S) = \int_0^{\infty} \Pi(z, x)e^{-Sx} dx, \quad \text{for } S \in \mathfrak{S}.$$

Since all elements of e^{-Sx} are dominated by 1 [3, 9], $G^*(S)$ and $\Pi^*(z, S)$ are well-defined. Further, $G^*(S)$ and $\Pi^*(z, S)$ are natural generalizations of $G^*(s)$ and $\Pi^*(z, s)$ because any positive real s can be identified with sI . Note that the notion of the Laplace transform $G^*(s)$ of matrix argument was used in Lucantoni, Choudhury and Whitt [6].

We further need the following two lemmas for later use.

LEMMA 1. Let $\Pi^*(z, S)$ be the Laplace transform defined on \mathfrak{S} . Then

$$\Pi^*(z, S)S - \Pi(z, 0) = \int_0^\infty \frac{d}{dx} \Pi^*(z, x)e^{-Sx} dx.$$

PROOF. Note that, for any $S \in \mathfrak{S}$, S is invertible. So, from integration by parts, we can easily derive the equation.

LEMMA 2. The matrix $-C - zD$ is an element of \mathfrak{S} for $0 < z < 1$.

PROOF. Since $C + D$ is irreducible, $C + zD$ is also irreducible for $0 < z < 1$. Further, since $C + D$ is an infinitesimal generator of the underlying Markov chain and D is a nonnegative matrix, $-C - zD$ is an element in \mathfrak{S} .

By multiplying both sides of (10) by $e^{(C+zD)x}$, and integrating both sides with respect to x , and using Lemma 1, we get

$$\begin{aligned} & -\Pi^*(z, -C - zD)[-C - zD] + \Pi(z, 0) \\ &= \Pi^*(z, -C - zD)[C + zD] + \frac{1}{z}[\Pi(z, 0) - \pi(0, 0)]G^*(-C - zD) \\ & \quad + pDG^*(-C - zD). \end{aligned}$$

Simplifying the above equation yields

$$\begin{aligned} \Pi(z, 0) \left[I - \frac{1}{z}G^*(-C - zD) \right] &= \left[-\frac{1}{z}\pi(0, 0) + pD \right] G^*(-C - zD) \\ &= p \left[\frac{1}{z}C + D \right] G^*(-C - zD). \end{aligned} \tag{13}$$

Combining (12) and (13), we can eliminate $\Pi(z, 0)$ to give

$$\begin{aligned} & \Pi^*(z, s)[sI + C + zD] \left[I - \frac{1}{z}G^*(-C - zD) \right] \\ &= p \left[\frac{1}{z}C + D \right] [G^*(-C - zD) - G^*(s)I]. \end{aligned}$$

Therefore we have

$$\begin{aligned} & \Pi^*(z, s)[sI + C + zD][zI - G^*(-C - zD)] \\ &= p[C + zD][G^*(-C - zD) - G^*(s)I], \end{aligned} \tag{14}$$

for $0 < z < 1$ and $s > 0$.

It remains to determine the vector p . Note that p is the stationary vector during idle periods. Let Q be the infinitesimal generator of the underlying Markov chain which is obtained by excising busy periods. By Takine and Hasegawa [17], Q satisfies the equation

$$Q = C + \int_0^\infty g(x)De^{Qx} dx. \tag{15}$$

Note that Q can be calculated by the limit of the recursive formulas

$$\begin{aligned} Q_{n+1} &= C + \int_0^\infty g(x)De^{Q_n x} dx, \quad n \geq 1, \\ Q_1 &= C + \int_0^\infty g(x)De^{Cx} dx, \\ Q &= \lim_{n \rightarrow \infty} Q_n. \end{aligned}$$

(See Takine and Hasegawa [17] for details.) Hence p must satisfy

$$pQ = 0, \quad pe = \alpha, \tag{16}$$

where α is a constant to be determined later. It will turn out to be $1 - \rho$ as we expect.

For $0 < z < 1$ and $s > 0$, let $U_i^*(z, s)$ be the double transform of the queue length and the remaining service time of the customer in service (if any) when the underlying Markov chain is in state i in the steady state, that is,

$$U_i^*(z, s) = \lim_{t \rightarrow \infty} P \{J_t = i, \xi_t = 0\} + \lim_{t \rightarrow \infty} E [z^{N_t} e^{-sX_t}; J_t = i, \xi_t = 1].$$

Let $U^*(z, s)$ be the vector whose i -th component is $U_i^*(z, s)$. Then by definition, we have

$$U^*(z, s) = p + \Pi^*(z, s).$$

Therefore, from (14), we have

$$\begin{aligned} U^*(z, s)[sI + C + zD][zI - G^*(-C - zD)] \\ = p\{s[zI - G^*(-C - zD)] + [C + zD][zI - G^*(s)I]\}. \end{aligned} \tag{17}$$

Now it remains to calculate the constant α . Letting $s \rightarrow 0^+$ in (17), we get

$$U^*(z, 0^+)[C + zD][zI - G^*(-C - zD)] = p[C + zD](z - 1).$$

Since $C + zD$ is invertible, we get

$$U^*(z, 0^+)[zI - G^*(-C - zD)] = p(z - 1).$$

By differentiating the above equation with respect to z , we get

$$\frac{d}{dz}U^*(z, 0^+)[zI - G^*(-C - zD)] + U^*(z, 0^+) \left[I - \frac{d}{dz}G^*(-C - zD) \right] = p.$$

By multiplying e from the right and letting $z \rightarrow 1^-$, we get

$$1 - \rho = \alpha.$$

Thus we obtain our main result.

THEOREM 1. *The vector $U^*(z, s)$ of the double transform of the queue length and the remaining service time of the customer in service in the steady state for $0 < z < 1$ and $s > 0$ is given by*

$$U^*(z, s)[sI + C + zD][zI - G^*(-C - zD)] = p \{s[zI - G^*(-C - zD)] + [C + zD][zI - G^*(s)I]\}.$$

The vector p satisfies the equations

$$pQ = 0, \quad pe = 1 - \rho,$$

where

$$Q = C + \int_0^\infty g(x)De^{Qx} dx.$$

From the above theorem we get the following corollary.

COROLLARY 1. *Let $\phi_i(z)$ be the generating function of the number of customers in the system (that is, the number of customers in the queue and the service facility) when the underlying Markov chain is in state i . Let $\phi(z)$ be the vector of which i -th component is $\phi_i(z)$. Then the generating function $\phi(z)$ of the number of customers in the system at an arbitrary time is given by*

$$\phi(z)[zI - G^*(-C - zD)] = p(z - 1)G^*(-C - zD).$$

PROOF. Note that

$$\phi(z) = p + z\Pi^*(z, 0).$$

From (14) we get, for $0 < z < 1$

$$\begin{aligned} \phi(z)[C + zD][zI - G^*(-C - zD)] &= p[C + zD][zI - G^*(-C - zD)] + zp[C + zD][G^*(-C - zD) - I] \\ &= p[C + zD](z - 1)G^*(-C - zD). \end{aligned}$$

Since $C + zD$ is invertible,

$$\phi(z)[zI - G^*(-C - zD)] = p(z - 1)G^*(-C - zD).$$

Note that our corollary is consistent with the equation (35) in Lucantoni [5], if we restrict batch arrivals to a single arrival as in his model.

As a special case, for the $M/G/1$ queueing system with arrival rate λ , we have $C = -\lambda$ and $D = \lambda$. Hence

$$\begin{aligned}\Pi^*(z, s) &= (1 - \rho)[\lambda - \lambda z][z - G^*(\lambda - \lambda z)]^{-1}[G^*(s) - G^*(\lambda - \lambda z)][s - \lambda + \lambda z] \\ &= \frac{(1 - \rho)\lambda[1 - z][G^*(s) - G^*(\lambda - \lambda z)]}{[z - G^*(\lambda - \lambda z)][s - \lambda + \lambda z]}\end{aligned}$$

which is a well-known result [16].

Acknowledgement

This work is partially supported by ROSET, the Center of Applied Mathematics at KAIST.

References

- [1] R. Bellman, *Introduction to matrix analysis* (McGraw-Hill, New York, 1970).
- [2] B. T. Hokstad, "A supplementary variable technique applied to the M/G/1 queue", *Scand. J. Stat.* **2** (1976) 95–103.
- [3] S. Karlin and H. M. Taylor, *A first course in stochastic processes* (Academic Press, Inc., New York, 1975).
- [4] D. M. Lucantoni, K. S. Meier-Hellstern and M. F. Neuts, "A single-server queue with server vacations and a class of non-renewal arrival processes", *Adv. Appl. Prob.* **22** (1990) 676–705.
- [5] D. M. Lucantoni, "New results on the single server queue with a batch Markovian arrival process", *Stoch. Mod.* **7** (1991) 1–46.
- [6] D. M. Lucantoni, G. L. Choudhury and W. Whitt, "The transient BMAP queue", *Stoch. Mod.* **10** (1994) 145–182.
- [7] D. M. Lucantoni and M. F. Neuts, "Simpler proofs of some properties of the fundamental period of the MAP/G/1 queue", *J. Appl. Prob.* **31** (1994) 235–243.
- [8] H. Minc, *Nonnegative Matrices* (John Wiley & Sons, Inc., New York, 1988).
- [9] M. F. Neuts, "A versatile Markovian point process", *J. Appl. Prob.* **16** (1979) 764–779.
- [10] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models* (The John Hopkins University Press, Baltimore, MD, 1981).
- [11] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications* (Marcel Dekker, New York, 1989).
- [12] V. Ramaswami, "The N/G/1 queue and its detailed analysis", *Adv. Appl. Prob.* **12** (1980) 222–261.
- [13] V. Ramaswami, "From the matrix-geometric to the matrix-exponential", *Queueing Systems* **6** (1990) 229–260.
- [14] E. Seneta, *Nonnegative Matrices: An Introduction to Theory and Applications* (John Wiley & Sons, New York, 1980).
- [15] B. Sengupta, "Markov processes whose steady state distribution is Matrix-exponential with an application to the GI/PH/1 queue", *Adv. Appl. Prob.* **21** (1989) 159–180.

- [16] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation Volume 1: Vacation and Priority Systems, Part I* (Elsevier Science Publisher B. B., (North Holland), Amsterdam, 1991).
- [17] T. Takine and T. Hasegawa, "The workload in the MAP/G/1 queue with state-dependent services: Its application to a queue with preemptive resume priority", *Stoch. Mod.* **10** (1994) 183–204.