Mathematical Problems in Engineering
Volume 5, pp. 255–273
Reprints available directly from the publisher
Photocopying permitted by license only

Optimal Control for a BMAP/G/1 Queue with Two Service Modes

ALEXANDER N. DUDIN [a,*] and SHOICHI NISHIMURA [b]

[a] *Department of Applied Mathematics and Informatics, Belarus State University, 4, F. Skorina Ave., Minsk-50, Belarus;* [b] *Department of Applied Mathematics, Science University of Tokyo, 1-3, Kagurazaka, Shinjuku-ku, Tokyo, 162 Japan*

(Received 23 September 1998)

Queueing models with controllable service rate play an important role in telecommunication systems. This paper deals with a single-server model with a batch Markovian arrival process (BMAP) and two service modes, where switch-over times are involved when changing the service mode. The embedded stationary queue length distribution and the explicit dependence of operation criteria on switch-over levels and derived.

Keywords: Hysteresis control; Batch Markovian arrival process

1 INTRODUCTION

Queueing models with controllable service rate have many promising applications for telecommunication systems. For e.g. they can be efficiently used when optimizing the transmission protocols in Integrated Service Digital Networks (ISDN). In ISDN, information of different type and value is transmitted simultaneously. So there is an opportunity to change dynamically the service rate of the priority flows by using controlled queueing models.

Another example of an effective application of such queues is the model of satellite business systems, see e.g. [10]. The part of the satellite bandwidth is strictly shared between the network subscribers.

* Corresponding author. Fax: 375-0172 265548. E-mail: Dudin@fpm.bsu.unibel.by.

The second part can be redistributed dynamically between the subscribers according to their demand. Controlled queues can be used for the optimal control by the bandwidth sharing and tariff policy optimization.

Also controlled queueing systems can be applied for the dynamic restriction of conversation durations in mobile cellular systems when congestion arises.

There is also a lot of different examples of potential applications of queueing models with controllable service rate to real-life systems. It explains the attention that these models received in the literature.

A brief review of papers dealing with the single-server systems with controllable service rate and Poisson or batch Poisson input was recently presented by Nobel [19].

But it is known that the real flows in modern communication networks are essentially non-Poisson because they have a "bursty" behaviour and dependent interarrival times. Combe [1] asserts that the batch Markovian arrival Process (BMAP) is a good mathematical model for these flows.

The great practical importance of queues with controlled service rate and the necessity to take into account the real nature of flows in communication networks motivates the interest of studying the BMAP/G/1 type models with controllable service rate. In Dudin [7], the model with N service modes and multithreshold strategy of control is investigated. But such strategies are effective only in situations where the switching of the modes is instantaneous and free of charge. In the present paper, we consider the system with two service modes but at the same time we take into consideration switching times and penalty for switching. We consider a hysteresis strategy of control instead of a threshold strategy. The hysteresis strategy reacts by increasing the service rate when the queue length increases (as well as the threshold strategy), but it is more flexible and does not cause frequent switching.

The rest of the paper is organized as follows. In Section 2, we describe the model. In Section 3, we derive an expression for the vector generating function of the queue steady state distribution at service completion epochs under fixed values of the thresholds (switch-over levels). In Section 4, we construct the algorithm for calculating the vector of the embedded queue being empty. In Section 5, we derive the precise result for the case when the lower switch-over level is equal to zero. In Section 6, an expression for the objective function in the threshold values is given.

2 THE MODEL

Consider a single-server queue whose arrival process is given by a BMAP. The definition of a BMAP was given by Lucantoni [14,15]. A BMAP is the natural generalization of a Poisson arrival process, which is commonly used by the telecommunication engineers.

In the BMAP, the interarrival times of customers are directed by a continuous time Markov process ν_t, $t \geq 0$ (the underlying or directing process) with a state space $\{0, 1, \ldots, W\}$. A transition from the state ν into the state r may induce the arrival of a batch of customers. The size of the batch depends on ν and r. Let the sojourn time in the state ν be exponentially distributed with the parameter $\lambda_\nu > 0$, $\nu = \overline{0, W}$. Given that an arrival takes place at state ν, $p_m(\nu, r)$ is the probability that the size of the batch arrival is m and the state transition is from ν to r. We suppose $p_m(\nu, r) \geq 0$ for $m > 0$, $\nu, r = \overline{0, W}$, $p_0(\nu, r) \geq 0$ for $r \neq \nu$,

$$\sum_{\substack{r=0,\\ r\neq\nu}}^{W} p_0(\nu, r) + \sum_{m=1}^{\infty} \sum_{r=0}^{W} p_m(\nu, r) = 1, \quad \nu = \overline{0, W}.$$

Introduce matrices D_m, $m \geq 0$, in the following way:

$$(D_0)_{\nu,\nu} = -\lambda_\nu, \ \nu = \overline{0, W}, \quad (D_0)_{\nu,r} = \lambda_\nu p_0(\nu, r), \ \nu, r = \overline{0, W}, \ \nu \neq r;$$
$$(D_m)_{\nu,r} = \lambda_\nu p_m(\nu, r), \quad m \geq 1, \ \nu, r = \overline{0, W}.$$

So a BMAP is defined by the sequence D_m, $m \geq 0$, of $(W+1) \times (W+1)$ matrices. The matrix D_0 governs transitions that correspond to no arrivals, and D_m governs transitions that correspond to arrivals of batches of size m. The matrix D_0 is a stable matrix, so the matrix $-D_0^{-1}$ exists and is nonnegative [14].

Define

$$D(z) = \sum_{m=0}^{\infty} D_m z^m, \quad |z| \leq 1, \ \ D = \sum_{m=0}^{\infty} D_m,$$

$$\bar{D} = \sum_{m=0}^{\infty} m D_m, \quad \bar{\bar{D}} = \sum_{m=0}^{\infty} m(m-1) D_m.$$

Let $\vec{X}$ be the stationary probability row-vector of the Markov chain ν_t. It satisfies the equations $\vec{X}D = 0$, $\vec{X}\mathbf{1}^{\mathrm{T}} = 1$, where $\mathbf{1}$ is a row-vector consisting of $W+1$ $1's$; $^{\mathrm{T}}$ is the symbol of transposition.

The intensity of the input λ is defined by the formula

$$\lambda = \vec{X}\bar{D}\mathbf{1}^{\mathrm{T}}, \tag{1}$$

where λ^{-1} is the mean interarrival time.

The system has unlimited waiting space and two service modes. These modes are alternately employed according to the number of customers presented at the service completion epochs. When the rth mode is employed, the service time has a distribution function $B_r(t)$ with a Laplace–Stieltjes transform $\beta_r(s) = \int_0^\infty \exp(-st)\mathrm{d}B_r(t)$ and finite initial moments $b_n^{(r)} = \int_0^\infty t^n \, \mathrm{d}B_r(t)$, $n = 1, 2$, $r = 1, 2$. The system can switch from one mode into another only at the service completion epochs. A switch-over time is involved when the system switches from the rth mode into the other. This time has a distribution function $G_r(t)$ with a Laplace–Stieltjes transform $g_r(s)$, $r = \overline{1,2}$. During the switching time, the service of customers is suspended.

The following cost criteria is imposed on the model:

$$I = a\lambda L + c_1 P_1 + c_2 P_2 + dM, \tag{2}$$

where L is an average queue length at service completion epochs, a is a holding cost, P_r is the average fraction of time, when the rth mode is used, c_r is the cost of rth mode per time unit, $r = \overline{1,2}$, M is the average number of mode switches per time unit, d is the fine per switch.

We suppose that $a > 0$, $d > 0$, $c_1 < c_2$, $\rho_1 > \rho_2$, where $\rho_r = \lambda b_1^{(r)}$, $r = \overline{1,2}$. Under such conditions the problem to determine the optimal switching strategy is not trivial. We find the optimal strategy in the class of so-called hysteresis strategies. This class is defined as follows.

Two nonnegative integers j and k, $j \leq k$, are fixed. They are called switch-over levels or thresholds. Let the queue length at the given service completion epoch be equal to i. If the following inequality: $i \leq j$ holds, the system will operate in the first mode. If $i > k$, then the system will operate in the second mode. And the system keeps the current mode in the case $j < i \leq k$.

The optimality of hysteresis strategies in the class of Markov homogeneous strategies is proved only in some basic cases, see e.g. Lu and Serfozo [13], Rykov [21]. But nevertheless, it is reasonable to use a hysteresis strategy of switching when the fine d is positive. So we will restrict ourselves by this class of strategies. Thus we have to indicate the optimal values (j^*, k^*) of the thresholds, which provide the minimal value cost criteria (2). To do this, we elaborate the algorithm for calculating the stationary queue length distribution under the arbitrary set (j, k) of thresholds and give the formulas for calculating the value of the cost criteria.

The analogous problem was solved for the $M/M/n$ system by Dudin [2], for the $M/G/1$ system by Dudin [3], Yamada and Nishimura [23], Nishimura and Jiang [17], for the $M^X/G/1$ system by Dudin [5], Nobel [19], Nobel and Tijms [20].

3 STATIONARY QUEUE LENGTH DISTRIBUTION

Let the thresholds (j, k), $0 \leq j \leq k$, be fixed. Let t_n be the nth customer's service completion epoch, i_n be the number of customers in the system at the moment $t_n + 0$, ω_n be the mode, which was in force at the moment $t_n - 0$, and ν_n be the state of the arrival directing process ν_t at the moment t_n.

The three-dimensional process $\{i_n, \omega_n, \nu_n\}$ is a Markov chain. Using the result of Dudin and Klimenok [4,6] and the result of Gail *et al.* [8,9], it can be shown that the necessary and sufficient condition for the existence of the stationary state distribution of this Markov chain is

$$\rho_2 < 1. \tag{3}$$

Let this inequality be fulfilled. Then the following limits exist:

$$\pi(i,\nu) = \lim_{n\to\infty} P\{i_n = i,\ \nu_n = \nu,\ \omega_n = 1\}, \quad i \geq 0,$$

$$\chi(i,\nu) = \lim_{n\to\infty} P\{i_n = i,\ \nu_n = \nu,\ \omega_n = 2\}, \quad i \geq j,$$

$$P\{(i,\nu,\omega) \to (l,r,\bar{\omega})\}$$
$$= \lim_{n\to\infty} P\{i_{n+1} = l,\ \nu_{n+1} = r,\ \omega_{n+1} = \bar{\omega} \mid i_n = i,\ \nu_n = \nu,\ \omega_n = \omega\},$$
$$\nu, r = \overline{0, W}, \quad l \geq \max\{0, i-1\}, \quad i \geq 0, \quad \omega, \bar{\omega} = 1, 2.$$

Define

$$\beta_m(-D(z)) = \int_0^\infty e^{D(z)t}\, dB_m(t),$$
$$g_m(-D(z)) = \int_0^\infty e^{D(z)t}\, dG_m(t).$$

It is easy to verify, that the matrices $W_l^{(m)}, \Gamma_l^{(m)}$ which are defined as follows:

$$\sum_{l=0}^{\infty} W_l^{(m)} z^l = \beta_m(-D(z)),$$

$$\sum_{l=0}^{\infty} \Gamma_l^{(m)} z^l = g_m(-D(z))$$

have the following probabilistic meaning: the (ν, r)th entry of matrix $W_l^{(m)}(\Gamma_l^{(m)})$ is the probability of the following event. The process ν_t transits from the state ν into the state r and l customers arrive into the system during the customer processing, which is performed in the mth mode (during the switching time from mth mode into the other). Because the matrix exponent does not possess certain nice properties of a scalar exponent, the problem of determination of the matrices $W_l^{(m)}, \Gamma_l^{(m)}$ appears to be rather difficult. The examples of solving this problem for some partial cases can be found in Dudin [7]. Here we consider these matrices to be known.

Taking into account the above mentioned reasons, the formula of composite probability, and the fixed strategy of control, the following lemma can be proved.

LEMMA 1 *The transition probabilities* $P\{(i,\nu,\omega)\rightarrow(l,r,\bar{\omega})\}$ *of Markov chain* $\{i_n, \nu_n, \omega_n\}$ *in the case* $j>0$ *are defined as follows:*

- $P\{(i,\nu,1)\rightarrow(l,r,1)\}$, $l\geq i-1$, $i>0$, *is the* (ν,r)*th entry of the matrix* $W^{(1)}_{l-i+1}$;
- $P\{(0,\nu,1)\rightarrow(l,r,1)\}$, $l\geq 0$, *is the* (ν,r)*th entry of the matrix*

$$-D_0^{-1}\sum_{i=0}^{l} D_{i+1} W^{(1)}_{l-i};$$

- $P\{(i,\nu,2)\rightarrow(l,r,2)\}$, $l\geq i-1$, $i\geq j+1$, *is the* (ν,r)*th entry of the matrix* $W^{(2)}_{l-i+1}$;
- $P\{(i,\nu,1)\rightarrow(l,r,2)\}$, $l\geq i-1$, $i\geq k+1$, *is the* (ν,r)*th entry of the matrix*

$$\sum_{m=i}^{l+1}\Gamma^{(1)}_{m-i}W^{(2)}_{l+1-m};$$

- $P\{(j,\nu,2)\rightarrow(l,r,1)\}$, $l\geq j-1$, *is the* (ν,r)*th entry of the matrix*

$$\sum_{m=j}^{l+1}\Gamma^{(2)}_{m-j}W^{(1)}_{l+1-m}.$$

In the case $j=0$ only the last transition probability is changed. Here $P\{(0,\nu,2)\rightarrow(l,r,1)\}$, $l\geq 0$, is the (ν,r)th entry of the matrix

$$-\Gamma^{(2)}_0 D_0^{-1}\sum_{i=0}^{l} D_{i+1}W^{(1)}_{l-i}+\sum_{i=1}^{l+1}\Gamma^{(2)}_i W^{(1)}_{l-i+1}.$$

Introduce into consideration the following row-vectors:

$$\vec{\pi}_l=\{\pi(l,0),\pi(l,1),\ldots,\pi(l,W)\},\quad l\geq 0,$$
$$\vec{\chi}_l=\{\chi(l,0),\chi(l,1),\ldots,\chi(l,W)\},\quad l\geq j.$$

Here we consider only the case $j>0$. The results for the case $j=0$ will be given in Section 5.

It is easy to see that the vectors $\vec{\pi}_l, \vec{\chi}_l$ satisfy the following system of equations:

$$\vec{\pi}_l = \vec{\pi}_0\left(-D_0^{-1}\sum_{i=0}^{l} D_{i+1}W_{l-i}^{(1)}\right) + \sum_{i=1}^{l+1}\vec{\pi}_i W_{l-i+1}^{(1)}, \quad l = \overline{0, j-2}, \quad (4)$$

$$\vec{\pi}_l = \vec{\pi}_0\left(-D_0^{-1}\sum_{i=0}^{l} D_{i+1}W_{l-i}^{(1)}\right) + \sum_{i=1}^{l+1}\vec{\pi}_i W_{l-i+1}^{(1)}$$

$$+ \vec{\chi}_j \sum_{m=j}^{l+1} \Gamma_{m-j}^{(2)} W_{l+1-m}^{(1)}, \quad l = \overline{j-1, k-1}, \quad (5)$$

$$\vec{\pi}_l = \vec{\pi}_0\left(-D_0^{-1}\sum_{i=0}^{l} D_{i+1}W_{l-i}^{(1)}\right) + \sum_{i=1}^{k}\vec{\pi}_i W_{l-i+1}^{(1)}$$

$$+ \vec{\chi}_j \sum_{m=j}^{l+1} \Gamma_{m-j}^{(2)} W_{l+1-m}^{(1)}, \quad l \geq k, \quad (6)$$

$$\vec{\chi}_l = \sum_{i=j+1}^{l+1} \vec{\chi}_i W_{l-i+1}^{(2)}, \quad l = \overline{j, k-1},$$

$$\vec{\chi}_l = \sum_{i=j+1}^{l+1} \vec{\chi}_i W_{l-i+1}^{(2)} + \sum_{i=k+1}^{l+1} \vec{\pi}_i \sum_{m=i}^{l+1} \Gamma_{m-i}^{(1)} W_{l+1-m}^{(2)}, \quad l \geq k. \quad (7)$$

Introduce the partial generating functions

$$\vec{\Pi}_1(z) = \sum_{i=0}^{k} \vec{\pi}_i z^i, \quad \vec{\Pi}_2(z) = \sum_{i=k+1}^{\infty} \vec{\pi}_i z^i, \vec{K}(z) = \sum_{i=j}^{\infty} \vec{\chi}_i z^i, \qquad |z| \leq 1.$$

THEOREM 1 *The partial generating functions* $\vec{\Pi}_1(z)$, $\vec{\Pi}_2(z)$, $\vec{K}(z)$ *satisfy the following equalities*:

$$\vec{\Pi}_1(z) = \vec{\pi}_0\left(\sum_{i=0}^{k} F_i z^i - v(k,j)\sum_{l=0}^{k-j} \Omega_l z^{l+j}\right), \quad (8)$$

$$\vec{\Pi}_2(z) = \vec{\Pi}_1(z)(\beta_1(-D(z)) - Ez)z^{-1}$$
$$+ \vec{\pi}_0(-D_0^{-1}D(z) + z^j v(k,j)g_2(-D(z)))\beta_1(-D(z))z^{-1}, \quad (9)$$
$$\vec{K}(z) = (\vec{\pi}_0 v(k,j)z^j - \vec{\Pi}_2(z)g_1(-D(z)))\beta_2(-D(z))$$
$$\times (\beta_2(-D(z)) - zE)^{-1}. \quad (10)$$

Here

- *matrices* F_l, $l = \overline{0,k}$, *are defined by the recurrent relations*

$$F_0 = E,$$
$$F_{l+1} = \left(F_l + D_0^{-1}\sum_{i=0}^{l} D_{i+1}W_{l-i}^{(1)} - \sum_{i=1}^{l} F_i W_{l-i+1}^{(1)}\right)\left(W_0^{(1)}\right)^{-1}, \quad l \geq 0; \quad (11)$$

- E *is* $(W+1)\times(W+1)$ *identity matrix;*
- *matrices* Ω_l *are defined as follows:*

$$\Omega_l = \sum_{r=0}^{l} \Gamma_r^{(2)}\tilde{\Omega}_{l-r}, \quad (12)$$

where matrices $\tilde{\Omega}_l$ *are defined by the recurrent relations*

$$\tilde{\Omega}_0 = E, \quad \tilde{\Omega}_l = F_l - D_0^{-1}\sum_{r=0}^{l-1} D_{l-r}\tilde{\Omega}_r, \ l \geq 1; \quad (13)$$

- *the matrix* $v(k,j)$ *is defined by the formula*

$$v(k,j) = F_{k+1}\Omega_{k-j+1}^{-1}. \quad (14)$$

Proof We use the matrix analogue of the principle of disregarding, which was formulated for the scalar case in Dudin [5].

Principle of disregarding. To find the form of the vectors $\vec{\pi}_l$, $l = \overline{1,i}$, up to the value of vector $\vec{\pi}_0$, we can take into consideration only the first i equations of the system (4), $i = \overline{1,j-1}$. To find the form of the vectors $\vec{\pi}_l$, $l = \overline{j,m}$, up to the values of vectors $\vec{\pi}_0, \vec{\chi}_j$, we can take into consideration only system (1) and Eq. (5) for $i = \overline{j-1,m-1}$. In both cases we can disregard the tails of the systems (4), (5) and even we can change the tails temporarily.

Apply this principle to the system (4). Set temporarily $j=\infty$ and introduce the temporary generating function

$$\hat{\vec{\Pi}}(z) = \sum_{l=0}^{\infty} \vec{\pi}_l z^l.$$

Multiplying Eq. (4) by the corresponding degree of z and summing up, we derive the following expression for the function $\hat{\vec{\Pi}}(z)$:

$$\hat{\vec{\Pi}}(z) = \vec{\pi}_0 D_0^{-1} D(z)\beta_1(-D(z))(\beta_1(-D(z)) - zE)^{-1}. \tag{15}$$

The matrices F_l are the coefficients of the expansion in series of the matrix $D_0^{-1}D(z)\beta_1(-D(z))(\beta_1(-D(z)) - zE)^{-1}$. They satisfy the system (11). Formula (15) makes clear the probabilistic sense of the matrices F_l: they coincide with the corresponding matrices in the following representation:

$$\vec{\pi}_l = \vec{\pi}_0 F_l, \quad l \geq 0,$$

of a stationary state probabilities of a classic BMAP/G/1 system operating in the first mode, see Lucantoni [14,15].

Now we recall, that system (4) actually holds only for $l = \overline{0, j-2}$, so the representation

$$\vec{\pi}_l = \vec{\pi}_0 F_l \tag{16}$$

for our system is valid only for $l = \overline{0, j-1}$.

Further, consider systems (4) and (5). Set temporarily $k=\infty$ and introduce a temporary generating function

$$\tilde{\vec{\Pi}}(z) = \sum_{l=0}^{\infty} \vec{\pi}_l z^l.$$

Multiplying Eqs. (4) and (5) by the corresponding degree of z and summing up, we obtain the following expression for the function $\tilde{\vec{\Pi}}(z)$:

$$\begin{aligned}\tilde{\vec{\Pi}}(z) = {} & \vec{\pi}_0 D_0^{-1} D(z)\beta_1(-D(z))(\beta_1(-D(z)) - zE)^{-1} \\ & - \vec{\chi}_j z^j g_2(-D(z))\beta_1(-D(z))(\beta_1(-D(z)) - zE)^{-1}. \end{aligned} \tag{17}$$

Expanding (17) in series, we have the following expression:

$$\vec{\pi}_{l+j} = \vec{\pi}_0 F_{l+j} - \vec{\chi}_j \Omega_l, \tag{18}$$

where the matrices Ω_l are defined by formulas (12) and (13). Recall that system (5) is actually valid only for $l = \overline{j-1, k-1}$. We see that (18) is valid for $l = \overline{0, k-j}$.

Thus, from (16) and (18) we deduce

$$\vec{\Pi}_1(z) = \vec{\pi}_0 \sum_{i=0}^{k} F_i z^i - \vec{\chi}_j \sum_{i=0}^{k-j} \Omega_l z^{l+j}. \tag{19}$$

To find the relationship between $\vec{\pi}_0$ and $\vec{\chi}_j$, we use Eq. (6) for $l = k$. By substituting (16) and (18) into this equation and taking into account the recurrent relations (11) and (13), after some algebra we have the relation

$$\vec{\chi}_j = \vec{\pi}_0 \nu(k,j), \tag{20}$$

where the matrix $\nu(k,j)$ is defined by formula (14). Substituting (20) into (19), we have proved formula (8).

Multiplying systems (4)–(6) by the corresponding degree of z and summing up, we prove formula (9). By analogy we have (10) from system (7).

So, Theorem 1 is proved.

Introduce also the following stationary probabilities:

$$R(i, \nu) = \lim_{n \to \infty} P\{i_n = i,\ \nu_n = \nu\},$$

the vectors $\vec{R}_i = (R(i,0), R(i,1), \ldots, R(i,W))$ and the vector generating function

$$\vec{R}(z) = \sum_{i=0}^{\infty} \vec{R}_i z^i.$$

It is easy to see, that $\vec{R}_i = \vec{\pi}_i + \vec{\chi}_i$ and hence

$$\vec{R}(z) = \vec{\Pi}_1(z) + \vec{\Pi}_2(z) + \vec{K}(z).$$

So, it is very easy to verify the validity of the following statement:

COROLLARY *The vector generating function $\vec{R}(z)$ is defined as follows*:

$$\vec{R}(z) = \vec{\pi}_0 \Bigg\{ D_0^{-1} D(z)\beta_1(-D(z))U(z) + z^j v(k,j)[\beta_2(-D(z)) \\ \times (\beta_2(-D(z)) - zE)^{-1} - g_2(-D(z))\beta_1(-D(z))U(z)] \\ + \left[\sum_{i=0}^{k} F_i z^i - v(k,j) \sum_{l=0}^{k-j} \Omega_l z^{l+j} \right] \\ \times (E - (\beta_1(-D(z)) - Ez)U(z)) \Bigg\}, \tag{21}$$

where $U(z) = (zE - (E - g_1(-D(z)))\beta_2(-D(z)))(\beta_2(-D(z)) - zE)^{-1} z^{-1}$.

The generating function $\tilde{R}(z)$ of the queue length stationary distribution:

$$\tilde{R}(z) = \sum_{i=0}^{\infty} \tilde{R}_i z^i, \quad \tilde{R}_i = \lim_{n \to \infty} P\{i_n = i\}$$

is defined as follows:

$$\tilde{R}(z) = \vec{R}(z)\mathbf{1}^{\mathrm{T}}.$$

Formulas (8)–(10), (21) in the scalar case coincide with the corresponding formulas in Dudin [5]. In that case they almost completely define stationary state probabilities because the probability π_0 of the system being empty is defined from the normalization condition rather simply. In our case the problem is not so simple because we have the vector $\vec{\pi}_0$ with $W+1$, still unknown, entries and we have to find its value.

As mentioned in Dudin [7], there are two possible ways to determine the vector π_0: the approach of Neuts [16] and the approach, which exploits the analyticity of corresponding generating function in the unit disc. As in Dudin [7], we follow the second way.

4 DETERMINATION OF THE VECTOR π_0

To develop the algorithm for finding the vector π_0, we need some preliminary considerations.

Using (8) and (9) in (10), we derive the formula

$$\begin{aligned}
&\vec{K}(z)(\beta_2(-D(z)) - zE)\\
&\quad= \vec{\pi}_0\left\{ v(k,j)z^j - \left[\left(\sum_{i=0}^{k} F_i z^i - v(k,j)\sum_{l=0}^{k-j}\Omega_l z^{l+j}\right)(\beta_1(-D(z)) - zE)z^{-1}\right.\right.\\
&\qquad\left. + (-D_0^{-1}D(z) + z^j v(k,j) g_2(-D(z)))\beta_1(-D(z))z^{-1}\right]\\
&\qquad\left. \times g_1(-D(z))\right\}\beta_2(-D(z)). \qquad (22)
\end{aligned}$$

It follows from (22) that

$$\begin{aligned}
\vec{K}(z) &= \vec{\pi}_0\left\{ v(k,j)z^j - \left[\left(\sum_{i=0}^{k} F_i z^i - v(k,j)\sum_{l=0}^{k-j}\Omega_l z^{l+j}\right)(\beta_1(-D(z)) - zE)z^{-1}\right.\right.\\
&\quad\left. + (-D_0^{-1}D(z) + z^j v(k,j) g_2(-D(z)))\beta_1(-D(z))z^{-1}\right]\\
&\quad\left. \times g_1(-D(z))\right\}\beta_2(-D(z))(\beta_2(-D(z)) - zE)^{-1}. \qquad (23)
\end{aligned}$$

Define $Q(z)$ the adjoint matrix to the matrix $\beta_2(-D(z)) - zE$, that is,

$$Q(z) = (\beta_2(-D(z)) - zE)^{-1} \cdot \det(\beta_2(-D(z)) - zE).$$

The equation

$$\det(\beta_2(-D(z)) - zE) = 0 \qquad (24)$$

has a simple root $z = 1$ and W roots (taking into account the multiplicity) inside the region $|z| < 1$, if condition (3) is fulfilled. This has been proved by Klimenok [11] and Gail *et al.* [8,9]. We denote these roots in the region $|z| < 1$ as z_r, $r = \overline{1,m}$ and their multiplicities by k_r, $r = \overline{1,m}$, $k_1 + \cdots + k_m = W$.

Because of the function $\vec{K}(z)$ being analytical in the region $|z|<1$, we derive from (23) the following system of linear algebraic equations for the components of the unknown vector $\vec{\pi}_0$:

$$\vec{\pi}_0\frac{\mathrm{d}^n}{\mathrm{d}z^n}\left\{\left\{v(k,j)z_r^j-\left[\left(\sum_{i=0}^{k}F_iz_r^i-v(k,j)\sum_{l=0}^{k-j}\Omega_l z_r^{l+j}\right)(\beta_1(-D(z_r))-z_rE)z_r^{-1}\right.\right.\right.$$
$$\left.+(-D_0^{-1}D(z_r)+z_r^jv(k,j)g_2(-D(z_r)))\beta_1(-D(z_r))z_r^{-1}\right]$$
$$\left.\left.\times g_1(-D(z_r))\right\}\beta_2(-D(z_r))Q(z_r)\right\}\mathbf{e}_1^{\mathrm{T}}=0,$$
$$n=\overline{0,k_r-1},\ r=\overline{1,m}, \tag{25}$$

where $\mathbf{e}_1$ is $(W+1)$-vector $(1,0,\ldots,0)$.

According to Gail *et al.* [9], it can be shown that Eq. (25) are linearly independent. So we have system (25) of W linear independent equations for $W+1$ entries of the vector $\vec{\pi}_0$. Now we need only one inhomogeneous equation. To obtain this equation we perform the following.

Introduce into consideration the matrices $B_m^{(n)}, G_m^{(n)}$, $m=1,2$, $n=1,2$, which are defined as follows:

$$\beta_m(-D(z))-zE=\beta_m(-D)-E+(z-1)B_m^{(1)}$$
$$+(z-1)^2B_m^{(2)}+o(z-1)^2,$$
$$g_m(-D(z))-zE=g_m(-D)-E+(z-1)G_m^{(1)}$$
$$+(z-1)^2G_m^{(2)}+o(z-1)^2.$$

The problem of finding the matrices $B_m^{(n)}, G_m^{(n)}$ can be solved by using the technique of eigenvalues as in Nishimura and Sato [18] under the conditions that the eigenvalues of D are simple and $D(z)$ is analytic in a neighbourhood at $z=1$. This assumption is not crucial in applications. Note that $\beta_m(-D)\mathbf{1}^{\mathrm{T}}=\mathbf{1}^{\mathrm{T}}$, $g_m(-D)\mathbf{1}^{\mathrm{T}}=\mathbf{1}^{\mathrm{T}}$. Formulas for calculating the vectors $B_m^{(n)}\mathbf{1}^{\mathrm{T}}$, $G_m^{(n)}\mathbf{1}^{\mathrm{T}}$ are given in Dudin [7] in the same partial cases.

Expanding (22) in powers of $(z-1)$, multiplying it by $\mathbf{1}^{\mathrm{T}}$, and equating the coefficients under corresponding degrees of $(z-1)$, we have the following relations:

$$\vec{K}(1)B_2^{(1)}\mathbf{1}^{\mathrm{T}} = \vec{\pi}_0\Bigg\{D_0^{-1}(\bar{D} + DB_1^{(1)} + D\beta_1(-D)N_1) - v(k,j) \times (N_2 - B_2^{(1)} + 2E + g_2(-D)\beta_1(-D)N_1) - \left(\sum_{i=0}^{k} F_i - v(k,j)\sum_{l=0}^{k-j}\Omega_l\right)((\beta_1(-D) - E)N_1 + B_1^{(1)})\Bigg\}\mathbf{1}^{\mathrm{T}}, \tag{26}$$

where

$$N_1 = G_1^{(1)} + g_1(-D)B_2^{(1)}, \quad N_2 = G_2^{(1)} + g_2(-D)B_1^{(1)},$$

and

$$\vec{K}'(1)B_2^{(1)}\mathbf{1}^{\mathrm{T}} = -\vec{K}(1)B_2^{(2)}\mathbf{1}^{\mathrm{T}} + \vec{\pi}_0 v(k,j)\left(B_2^{(2)} + jB_2^{(1)} + \frac{j(j+1)}{2}E\right)\mathbf{1}^{\mathrm{T}} - \Big\{\vec{\Pi}_2(1)\Big[g_1(-D)B_2^{(2)} + \Big(G_1^{(1)} + E\Big)\Big(E + B_2^{(1)}\Big) + G_1^{(2)}\Big] + \vec{\Pi}_2'(1)\Big[g_1(-D)\Big(E + B_2^{(1)}\Big) + G_1^{(1)} + E\Big] + \tfrac{1}{2}\vec{\Pi}_2''(1)\Big\}\mathbf{1}^{\mathrm{T}}. \tag{27}$$

Here $'$ is the symbol of derivative.

We can calculate the vectors $\vec{\Pi}_2(1)$, $\vec{\Pi}_2'(1)$ and the number $\vec{\Pi}_2''(1)\mathbf{1}^{\mathrm{T}}$ in (27) using the following formulas:

$$\vec{\Pi}_2(1) = \vec{\Pi}_1(1)(\beta_1(-D) - E) + \vec{\pi}_0(-D_0^{-1}D + v(k,j)g_2(-D))\beta_1(-D), \tag{28}$$

$$\vec{\Pi}_2(1)\mathbf{1}^{\mathrm{T}} = \vec{\pi}_0 v(k,j)\mathbf{1}^{\mathrm{T}},$$

where

$$\vec{\Pi}_1(1) = \vec{\pi}_0\left[\sum_{i=0}^{k} F_i - v(k,j)\sum_{l=0}^{k-j}\Omega_l\right], \tag{29}$$

$$\vec{\Pi}_2'(1) = \vec{\Pi}_1'(1)(\beta_1(-D) - E) + \vec{\Pi}_1(1)(B_1^{(1)} + E - \beta_1(-D))$$
$$+ \vec{\pi}_0[(-D_0^{-1}\bar{D} + jv(k,j) + v(k,j)(G_2^{(1)} + E))\beta_1(-D)$$
$$+ (-D_0^{-1}D + v(k,j)g_2(-D))(B_1^{(1)} + E - \beta_1(-D))], \qquad (30)$$

where

$$\vec{\Pi}_1'(1) = \vec{\pi}_0\left[\sum_{i=0}^{k} iF_i - v(k,j)\sum_{l=0}^{k-j}(l+j)\Omega_l\right],$$

$$\vec{\Pi}_2''(1)\mathbf{1}^{\mathrm{T}} = -2\Big\{\vec{\Pi}_2'(1)\mathbf{1}^{\mathrm{T}} - \vec{\Pi}_1'(1)B_1^{(1)}\mathbf{1}^{\mathrm{T}} - \vec{\Pi}_1(1)B_1^{(2)}\mathbf{1}^{\mathrm{T}}$$
$$- \vec{\pi}_0\Big\{-\tfrac{1}{2}D_0^{-1}\Big(\bar{\bar{D}} + 2\bar{D} + \Big(B_1^{(1)} + E\Big) + 2DB_1^{(2)}\Big)$$
$$+ v(k,j)\Big[j\Big(E + G_2^{(2)}\Big) + G_2^{(2)} + jg_2(-D)\Big(E + B_1^{(1)}\Big)$$
$$+\Big(G_2^{(1)} + E\Big)\Big(E + B_1^{(1)}\Big) + \frac{j(j-1)}{2}E + g_2(-D)B_1^{(2)}\Big]\Big\}\mathbf{1}^{\mathrm{T}}\Big\}. \qquad (31)$$

Now we are ready to describe the algorithm for determination of the vectors $\vec{\pi}_0, \vec{\pi}_l$, $l \geq 1$, $\vec{\Pi}_1(1)$, $\vec{\Pi}_2(1)$, $\vec{K}(1)$, $\vec{\Pi}_1'(1)$, $\vec{\Pi}_2'(1)$, $\vec{K}'(1)$.

The Algorithm

Step 1 Set $z = 1$ in (22) and consider (22) as the system of linear algebraic equations for the entries of vector $\vec{K}(1)$. The matrix $\beta_2(-D) - E$ of this system is singular. Replace one equation of this system by Eq. (26). Inverting the matrix of such modified system, we obtain the following relation:

$$\vec{K}(1) = \vec{\pi}_0 H, \qquad (32)$$

where H is a known matrix.

Step 2 Taking into account (28), (29), (32) and the normalization condition, we have the following equation for the entries of the unknown vector $\vec{\pi}_0$:

$$\vec{\pi}_0\left(H + v(k,j) + \left(\sum_{i=0}^{k} F_i - v(k,j)\sum_{l=0}^{k-j}\Omega_l\right)\right)\mathbf{1}^{\mathrm{T}} = 1. \qquad (33)$$

Solve systems (25) and (33) and obtain the value of vector $\vec{\pi}_0$.

Step 3 Using (16) and the known vector $\vec{\pi}_0$, obtain the values of the vectors $\vec{\pi}_l$ for any l, $l \geq 1$.

Step 4 From (32), (28) and (29) obtain the values of the vectors $\vec{\Pi}_1(1), \vec{\Pi}_2(1), \vec{K}(1)$.

Step 5 Differentiate Eq. (22) and set $z = 1$. Consider this equation as the system of linear algebraic equations for the entries of the vector $\vec{K}'(1)$. Replace one equation of this system by Eq. (27). The vectors $\vec{\Pi}_1'(1)$, $\vec{\Pi}_2'(1)$ and the number $\vec{\Pi}_2''(1)\mathbf{1}^{\mathrm{T}}$ are calculated from (30) and (31). Solve the modified system and obtain the value of the vector $\vec{K}'(1)$.

The End

Having formulas (8)–(10) and the value of the vector $\vec{\pi}_0$, which was calculated by means of our algorithm, we solved the problem of determining the stationary state probabilities of the embedded Markov chain $\{i_n, \nu_n, \omega_n\}$ for any fixed set of the thresholds (j, k), $0 < j \leq k$.

5 THE CASE $j = 0$

Above we calculated the stationary state distribution in the case $j > 0$. The case $j = 0$ requires special consideration because the probabilities $P\{(0, \nu, 2) \to (l, r, 1)\}$ have now another form (see Lemma 1). But notwithstanding these special considerations, we showed that the result coincides with the result in the case above up to some small modifications. Namely, in the case $j = 0$ we have to write

$$g_1(-D(z)) - \Gamma_0^{(2)} D_0^{-1} D(z) \text{ instead of } g_2(-D(z)),$$

$$g_2(-D) - \Gamma_0^{(2)} D_0^{-1} D \text{ instead of } g_2(-D),$$

$$\Omega_l - \Gamma_0^{(2)} F_l \text{ instead of } \Omega_l,$$

$$G_2^{(1)} - \Gamma_0^{(2)} D_0^{-1} \bar{D} \text{ instead of } G_2^{(1)},$$

$$G_2^{(2)} - \Gamma_0^{(2)} D_0^{-1} \bar{\bar{D}}/2 \text{ instead of } G_2^{(2)}.$$

By making these modifications, we obtain the stationary state distribution for $j = 0$ from the general formulas above.

6 CALCULATING THE VALUE OF THE COST CRITERIA

Having available the values of vectors $\vec{\pi}_0$, $\vec{\Pi}_1(1)$, $\vec{\Pi}_2(1)$, $\vec{K}(1)$, $\vec{\Pi}'_1(1)$, $\vec{\Pi}'_2(1)$, $\vec{K}'(1)$ under the fixed values of thresholds (j,k), we can calculate the value of the cost criteria (2), which corresponds to this set of thresholds. The formula for the calculation of the mean queue length L is evident:

$$L = (\vec{\Pi}'_1(1) + \vec{\Pi}'_2(1) + \vec{K}'(1))\mathbf{1}^{\mathrm{T}}. \tag{34}$$

By exploiting the ergodic theorems for Markov chains, see e.g. Skorokhod [22], and the strong law of large numbers, see Kolmogrow [12], it can be shown that

$$P_2 = \lambda \vec{K}(1)\mathbf{1}^{\mathrm{T}} b_1^{(2)}, \quad P_1 = 1 - P_2. \tag{35}$$

It is easy to see that $M = 2\lambda_{\vec{X}_j}\mathbf{1}^{\mathrm{T}}$ or

$$M = 2\lambda\vec{\pi}_0 v(k,j)\mathbf{1}^{\mathrm{T}}. \tag{36}$$

Substituting (1), (34)–(36) into (2), we get the value of the cost criteria (2). Having the possibility to calculate the value of criteria (2) for any fixed (j,k), practically we have an opportunity to find the optimal set (j^*,k^*) of thresholds.

Note, that the most costly part of the algorithm – calculating the roots of Eq. (24) – is implemented only one time, because of (j,k) not being involved in (24).

7 CONCLUSION

Throughout this paper, we have studied the controlled queueing model with two service modes. This model can be widely used by practical engineers to optimize the processing of the real-life systems when it is possible to handle the demands by means of fast and expensive or slow and cheap tools. We offer an algorithm to calculate the optimal strategy for dynamically changing the tool for processing the demands in accordance with the queue length. It allows to handle the trade-off between the mean queue length and the cost of lending the tools. The model of the input flow is more flexible than the stationary Poisson input which is used by the engineers traditionally.

References

[1] M.B. Combé, Queueing models with dependence structures, *Ph. D. dissertation.* Amsterdam, 1995.
[2] A.N. Dudin, Optimal assignment of the rate for service of customers in a multilinear two-rate service system, *Automation and Remote Control* **42** (1982), 74–82.
[3] A.N. Dudin, Queueing systems with varying modes of operation and their optimization, Doctoral dissertation, Minsk – Tomsk, 1992.
[4] A.N. Dudin and V.I. Klimenok, Characteristics calculation for the single server queueing system, which operates in the synchronized Markov random environment, *Automation and Remote Control* **58** (1997), 74–84.
[5] A.N. Dudin, Optimal control for a $M^X/\mathbf{G}/1$ queue with two operations modes, *Probability in the Engineering and Informational Sciences* **11** (1997), 255–265.
[6] A.N. Dudin and V.I. Klimenok, About the stationary state probabilities of two-dimensional Markov chains, *Automatic Control and Computer Sciences* **32** (1998), 15–22.
[7] A.N. Dudin, Optimal multithreshold control for a BMAP/G/1 queue with N service modes, *Queueing Systems* **30** (1998), 273–287.
[8] H.R. Gail, S.L. Hantler, M. Sidi and B.A. Taylor, Linear independence of root equations for $M/G/1$ type of Markov chains, *Queueing Sytems* **20** (1995), 321–339.
[9] H.R. Gail, S.L. Hantler and B.A. Taylor, Spectral analysis of $M/G/1$ and $/G/M/1$ type Markov chains, *Adv. Appl. Prob.* **28** (1996), 114–165.
[10] B. Goode, SBS TDMA-DA System with VAC and DAC, *Proceedings of the IEEE* **72** (1994), 1594–1610.
[11] V.I. Klimenok, The Rouche's theorem in the problem of the two-dimensional quasitoeplitz Markov chain stationary distribution determination, *Automatic Control and Computer Sciences* **32** (1998), 23–29.
[12] A.N. Kolmogorov, *Basic Notions of the Probability Theory*, ONTI, Moscow, 1936.
[13] F.V. Lu and R.F. Serfozo, $M/M/1$ queueing decision process with monotone hysteretic optimal policies, *Operations Research* **32** (1984), 1116–1129.
[14] D.M. Lucantoni, New results on the single server queue with a Batch Markovian Arrival Process, *Stochastic Models* **7** (1991), 1–46.
[15] D.M. Lucantoni, The BMAP/G/1 queue: A tutorial. *Models and Techniques for Performance Evaluation of Computer and Communication Systems*, Springer, Berlin, 1993.
[16] M.F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker Inc., New York, 1989.
[17] S. Nishimura and S. Jiang, An $M/G/1$ vacation model with two service modes, *Probability in the Engineering and Informational Sciences* **9** (1995), 355–374.
[18] S. Nishimura and H. Sato, Eigenvalue expression for mean queue length of BMAP/G/1 queue, *J. Oper. Res. Soc. Japan* **40** (1997), 122–132.
[19] R.D. Nobel, A regenerative approach for an $M^X/G/1$ queue with two service modes, *Automatic Control and Computer Sciences* **32** (1998), 3–14.
[20] R.D. Nobel and H.C. Tijms, Optimal control for a $M^X/G/1$ queue with two service modes, *European Journal of the Operations Research* (to appear).
[21] V.V. Rykov, Hysteretic phenomena in controllable queueing systems, *Vestnik of Russian University of People Friendship* **3** (1995), 101–111.
[22] A.V. Skorokhod, *Probability Theory and Random processes*, High School, Kiev, 1980.
[23] K. Yamada and S. Nishimura, A queueing system with a setup time for switching of the service distribution, *J. Oper. Res. Soc. Japan* **31** (1994), 271–285.