

A $MAP_K/G_K/1/\infty$ Queueing System with Generalized Foreground-Background Processor Sharing Discipline¹

C. D'Apice*, M. L. Cristofano*, and A. V. Pechinkin**

*University of Salerno, Italy

**Institute of Information Problems, Russian Academy of Sciences, Moscow, Russia

Received December 8, 2003

Abstract—A queueing system with Markov arrival process, several customer types, generalized foreground-background processor sharing discipline with minimal served length, and an infinite buffer for all types of customers is studied. The joint stationary distribution of the number of customers of all types and the stationary distribution of time of sojourn of customers of every type are determined in terms of generating functions and Laplace–Stieltjes transforms.

1. FORMULATION OF THE PROBLEM

One possible variant of the processor sharing discipline is the foreground-background processor sharing (FBPS) discipline with minimal served length. Queueing systems with Markov arrival process, several customer types, and generalized foreground-background processor sharing discipline with either separate finite buffers for customers of different types or a common finite buffer for customers of all types are investigated in [1]. Mathematical relations for computing the joint stationary distributions of the number of customers of all types in these systems are derived. It also gives a long list of papers devoted to systems under processor sharing discipline.

In this paper, we study a similar system, but with an infinite buffer for customers of all types. As will be clear from what follows, for systems with infinite buffers we can find the joint stationary distribution of the number of customers of all types in terms of generating functions, but also the stationary distribution of the time of sojourn of customers of every type in the system in terms of Laplace–Stieltjes transform.

We shall study a single-server system with a flow of customers of K types.

The input is a Markov process with a finite state set $\{1, \dots, I\}$ and defined by matrices Λ_k , $k = \overline{1, K}$, corresponding to the change in the generation phase with the arrival of a customer of the k th type, and matrix M corresponding to the change in the generation phase without any arrival (see [1]).

The total number of customers of any type is unbounded (system with an infinite buffer for customers of all types).

The service time (called the length in the sequel) of a customer of the k th type has a distribution function $B_k(x)$ with mean $\bar{b}_k = \int_0^\infty (1 - B_k(x)) dx < \infty$. For the sake of simplicity of presentation, we assume that the distribution densities $b_k(x) = B'_k(x)$ exist for deriving differential equations.

We also assume that the necessary and sufficient condition $\rho < 1$ for the existence of a stationary operation mode for the system holds. Here $\rho = \sum_{k=1}^K \lambda_k \bar{b}_k$ is the traffic intensity, $\lambda_k = \pi_a \Lambda_k \mathbf{1}$ is

¹ This work was supported in part by the Russian Foundation for Basic Research, project no. 02-07-90147.

the intensity of the input flow of customers of the k th type, and π_a is the vector of stationary probabilities of the Markov customer generation process.

The generalized foreground-background processor sharing (FBPS) discipline with minimal served length for our system is described as follows. A k th-type customer with served length x has a priority index $r_k(x)$, and the condition $r_k(0) = 0$ holds for every k . Moreover, $r_k(x)$ is a strictly increasing continuous function. The inverse $r_k^{(-1)}(y)$ of the function $r_k(x)$ is assumed to have a bounded derivative $\gamma_k(y) = dr_k^{(-1)}(y)/dy$. Then a customer with minimal priority index is served at any instant. If l customers have the same minimal priority index y , they are served concurrently. If the i th of these customers is of the k_i th type, his service rate is $\gamma_{k_i}(y) / \sum_{j=1}^l \gamma_{k_j}(y)$.

In Kendall's classification, our system is denoted by $MAP_K/G_K/1/\infty$ /FBPS.

In what follows, we use the following notation (see also [1]).

The intensity of completion of service of a customer of the k th type, which depends on his priority index y , is denoted by $\beta_k(y) = \gamma_k(y)b_k(r_k^{(-1)}(y))/(1 - B_k(r_k^{(-1)}(y)))$.

Vectors whose coordinates are numbered by indexes $i, i = \overline{1, I}$, are denoted by boldface letters, and vectors whose coordinates are numbered by indexes $k, k = \overline{1, K}$, are denoted by an over-arrow; for example, $\mathbf{p} = (p_1, \dots, p_I)$ and $\vec{n} = (n_1, \dots, n_K)$. Moreover, we shall not make any distinction between column vectors and row vectors in expressions since this is always clear from the context. Indeed, if a matrix is left multiplied by a vector, the result is a column vector; for the converse case, we obtain a row vector.

We say that $\vec{n} \leq \vec{m}$ if all coordinates the vector \vec{n} are not greater than the corresponding coordinates of the vector \vec{m} , and $\vec{n} < \vec{m}$ if all coordinates of the vector \vec{n} are not greater than the corresponding coordinates of the vector \vec{m} and $n_k < m_k$ for at least one k .

The scalar product of two vectors \vec{n} and \vec{m} is denoted by $(\vec{n}, \vec{m}) = \sum_{k=1}^K n_k m_k$. A vector all whose coordinates, except for the k th coordinate, are zero and the k th coordinate is 1 is denoted by \vec{e}_k . The vector $\vec{m} + \vec{e}_k$ is denoted by \vec{m}_k^+ , a vector of dimension I all whose coordinates are 1 is denoted by $\mathbf{1}$, and a vector all whose coordinates are 0 is denoted by $\mathbf{0}$. We also use the notation $|\vec{m}| = \sum_{i=1}^K m_i$ and $\vec{z}^{\vec{m}} = \prod_{k=1}^K z_k^{m_k}$.

2. STATIONARY DISTRIBUTION OF THE NUMBER OF CUSTOMERS

Let us consider a busy period. Let $A_k(\vec{m} | y), \vec{m} \geq \vec{0}$, denote the matrix, whose element $a_{kij}(\vec{m} | y)$ is the probability that at the instant of a busy period when the priority index takes the maximal value y or at the instant of completion of the busy period if the priority index does not take this value y in this busy period, then the system contains \vec{m} customers and the generation phase is j , provided the busy period began with generation phase i and service of a k th-type customer.

The functions $A_k(\vec{m} | y)$ satisfy the infinite system of first-order ordinary differential equations

$$\begin{aligned} \frac{d}{dy} A_k(\vec{m} | y) &= (\vec{m}, \vec{\gamma}(y)) A_k(\vec{m} | y) M \\ &+ \sum_{l=1}^K \left(\sum_{\vec{0} \leq \vec{i} \leq \vec{m}} (\vec{i}, \vec{\gamma}(y)) A_k(\vec{i} | y) \Lambda_l A_l(\vec{m} - \vec{i} | y) + (m_l + 1) \beta_l(y) A_k(\vec{m}_l^+ | y) - m_l \beta_l(y) A_k(\vec{m} | y) \right) \end{aligned} \quad (1)$$

under the initial condition

$$A_k(\vec{m} | 0) = \begin{cases} E, & \vec{m} = \vec{e}_k \\ 0, & \vec{m} \neq \vec{e}_k. \end{cases}$$

Using the generating function $\tilde{A}_k(\vec{z} | y) = \sum_{\vec{m} \geq \vec{0}} \vec{z}^{\vec{m}} A_k(\vec{m} | y)$, we can reduce system (1) to the finite system of first-order quasilinear partial differential equations

$$\frac{\partial}{\partial y} \tilde{A}_k(\vec{z} | y) = \sum_{l=1}^K \frac{\partial}{\partial z_l} \tilde{A}_k(\vec{z} | y) \left(z_l \gamma_l(y) \left[M + \sum_{n=1}^K \Lambda_n \tilde{A}_n(\vec{z} | y) \right] + \beta_l(y) [1 - z_l] \right) \tag{2}$$

under the boundary conditions

$$\tilde{A}_k(\vec{z} | 0) = z_k E. \tag{3}$$

Let us consider the busy period that began with generation phase i and service of a k th-type customer. Let $Q_k(\vec{m} | y)$, $\vec{m} > \vec{0}$, denote the matrix whose element $q_{kij}(\vec{m} | y)$ is the mean time of residence of the system in a state in which the system contains \vec{m} customers and the generation phase is j until the instant when the priority index attains the maximal value y or until the completion of the busy period if the priority index does not take this value y in this busy period.

The functions $Q_k(\vec{m} | y)$ satisfy the infinite system of first-order ordinary differential equations

$$\frac{d}{dy} Q_k(\vec{m} | y) = (\vec{m}, \vec{\gamma}(y)) A_k(\vec{m} | y) + \sum_{l=1}^K \sum_{\vec{0} < \vec{i} < \vec{m}} (\vec{i}, \vec{\gamma}(y)) A_k(\vec{i} | y) \Lambda_l Q_l(\vec{m} - \vec{i} | y)$$

under the initial condition $Q_k(\vec{m} | 0) = 0$, or the finite systems of first-order ordinary differential equations

$$\frac{d}{dy} \tilde{Q}_k(\vec{z} | y) = \sum_{l=1}^K z_l \gamma_l(y) \frac{\partial}{\partial z_l} \tilde{A}_k(\vec{z} | y) \left(E + \sum_{n=1}^K \Lambda_n \tilde{Q}_n(\vec{z} | y) \right)$$

in terms of the generating function $\tilde{Q}_k(\vec{z} | y) = \sum_{\vec{m} > \vec{0}} \vec{z}^{\vec{m}} Q_k(\vec{m} | y)$ under the initial condition $\tilde{Q}_k(\vec{z} | 0) = 0$.

Now, let

$A_k = A_k(\vec{0} | \infty)$ be the matrix whose elements are the probabilities a_{kij} that the generation phase at the end of a busy period is j , provided at the beginning of the busy period it was i and a k th-type customer was taken for service,

$Q_k(\vec{m}) = Q_k(\vec{m} | \infty)$ be the matrix whose elements are the mean times $q_{kij}(\vec{m})$ of sojourn of the system in a busy period in state (j, \vec{m}) , provided the generation phase at the beginning of the busy period was i and a k th-type customer was taken up for service,

$\tilde{Q}_k(\vec{z}) = \sum_{\vec{m} > \vec{0}} \vec{z}^{\vec{m}} Q_k(\vec{m}) = \tilde{Q}_k(\vec{z} | \infty)$ be the generating function of the matrix $Q_k(\vec{m})$ and,

$Q_k = \sum_{\vec{m} > \vec{0}} Q_k(\vec{m}) = \tilde{Q}_k(\mathbf{1})$ be the matrix whose elements are the mean times q_{kij} of sojourn of the generation process in phase j in a busy period, provided it was in phase i at the beginning of the busy period and a k th-type customer was taken up for service.

Now we can find the stationary distribution of the number of customers in the system. For this purpose, let us consider the imbedded Markov chain generated by the numbers of the states of the control process at the instants of completion of busy periods.

The probability that the busy period following an idle period begins with the service of a k th-type customer and state j of the control process, provided it was in state i at the beginning of the idle period is given by the formula in matrix form

$$\int_0^\infty e^{Mt} \Lambda_k dt = -M^{-1} \Lambda_k.$$

Therefore, the transition probability matrix P of the imbedded Markov chain is of the form

$$P = -M^{-1} \sum_{k=1}^K \Lambda_k A_k.$$

The vector π of stationary probabilities $\pi_i, i = \overline{1, I}$, of the imbedded Markov chain is determined from the system of equilibrium equations

$$\pi = \pi P$$

under the normalization condition

$$\pi \mathbf{1} = 1.$$

The matrix Q_0 whose elements are the mean sojourn times q_{0ij} of the control process in state j in an idle period, provided at the beginning of the idle period it was in state i and a k th-type customer was taken up for service is defined by the expression

$$Q_0 = \int_0^\infty e^{Mt} dt = -M^{-1}.$$

The mean time T between the instants of change in the state of the imbedded Markov chain for a system in stationary operation mode is defined by the expression

$$T = \pi Q_0 \mathbf{1} - \pi M^{-1} \sum_{k=1}^K \Lambda_k Q_k \mathbf{1} = -\pi M^{-1} \left(\mathbf{1} + \sum_{k=1}^K \Lambda_k Q_k \mathbf{1} \right).$$

Now we can find the vectors $\mathbf{p}(\vec{m}), \vec{m} \geq \vec{0}$, of time-stationary probabilities $p_i(\vec{m}), i = \overline{1, I}$, that the system contains \vec{m} customers of different types and control process is in state i :

$$\begin{aligned} \mathbf{p}(\vec{0}) &= \frac{1}{T} \pi Q_0 = -\frac{1}{T} \pi M^{-1}, \\ \mathbf{p}(\vec{m}) &= -\frac{1}{T} \pi M^{-1} \sum_{k=1}^K \Lambda_k Q_k(\vec{m}), \quad \vec{m} > \vec{0}. \end{aligned}$$

Using the generating function $\tilde{\mathbf{p}}(\vec{z}) = \sum_{\vec{m} \geq \vec{0}} z^{\vec{m}} \mathbf{p}(\vec{m})$, we obtain the formula

$$\tilde{\mathbf{p}}(\vec{z}) = -\frac{1}{T} \pi M^{-1} \left(E + \sum_{k=1}^K \Lambda_k Q_k(\vec{z}) \right).$$

The vector $\mathbf{p}_k^*(\vec{m}), \vec{m} \geq \vec{0}$, of stationary probabilities $p_{ki}^*(\vec{m}), i = \overline{1, I}, k = \overline{1, K}$, that a k th-type customer upon arrival finds \vec{m} other customers of different types in the system and the control process is in state i is given by the formula

$$\mathbf{p}_k^*(\vec{m}) = \frac{1}{\lambda_k} \mathbf{p}(\vec{m}) \Lambda_k, \quad \vec{m} \geq \vec{0},$$

or in terms of the generating function $\tilde{\mathbf{p}}_k^*(\vec{z}) = \sum_{\vec{m} \geq \vec{0}} z^{\vec{m}} \mathbf{p}_k^*(\vec{m})$, by the formula

$$\tilde{\mathbf{p}}_k^*(\vec{z}) = \frac{1}{\lambda_k} \tilde{\mathbf{p}}(\vec{z}) \Lambda_k.$$

In numerical computations, it is not easy to find the solution of system (2) under the boundary conditions (3). Therefore, we now state the functional relationships for $A_k(\vec{z} | y)$ derived from special probabilistic considerations.

Let us consider, along with the initial system, an y -system, which is defined as follows. The input flow, customer length, and service discipline for the y -system are the same as for the initial system. The difference is that the service of a k th-type customer with maximal priority index greater than y (i.e., length greater than $r_k^{(-1)}(y)$) is terminated in the y -system at the instant when the priority index attains the value y and he quits the y -system.

It is easily seen that the busy periods of the initial and y -systems coincide up to the instant τ when the maximal value of the served priority indexes of the customers in these systems attains the value y (at this instant both systems contain only customers of priority index y). Thereafter the busy period of the initial system continues, whereas it ends in the y -system. All customers of initial priority index greater than y that had arrived prior to the instant τ are assumed to have been served and, therefore, quit the system.

Therefore, the numbers of customers of different types in the initial system at the instant in a busy period when the maximal value of the priority index attains the value y , coincide with the numbers of customers of maximal priority index greater than y served in a busy period of the y -system, and the generating function $\tilde{A}_k(\vec{z} | y)$ coincides with the generating function of the number of customers of priority index greater than y served in the busy period of the y -system.

Note that for conservative service disciplines (such is the generalized FBPS discipline), the length of a busy period and numbers of served customers of different priority indexes and types in a busy period are invariant characteristics of the service discipline. Therefore, the generating function of the number of customers of length greater than y served in a busy period of the y -system can be computed using any conservative discipline. In our opinion, the LCFs discipline (see [2]) is best suited for this purpose.

Thus, let us consider a $MAP_K/G_K/1/\infty$ queueing system whose input is the same Markov flow of customers of K types as that of the initial system. This system is also referred to as the y -system. But, unlike the initial system, in the y -system

customers are served according to the LCFs discipline (with no regard for the type, length, priority index, etc.) and

customers of the k th type are subdivided into two classes: customers of priority index less than y (i.e., of length less than $r_k^{(-1)}(y)$) and customers of priority index greater than y (i.e., of length greater than $r_k^{(-1)}(y)$ in the initial system). The length of a class I customer is the same as before, whereas the length of a class II customer in the y -system is $r_k^{(-1)}(y)$.

Now the generating function $\tilde{A}_k(\vec{z} | y)$ coincides with the generating function of the number of class II customers of different types served in a busy period of the y -system.

Let $\tilde{A}_x^*(\vec{z} | y)$ denote the generating function of the number of class II customers of different types served in a busy period of the y -system opened by a class I customer of length x (the customer that opens a busy period may have any nonnegative length). The matrix expression takes account of the generation phase at the beginning and end of a busy period. Reasoning as in [2], we obtain

$$\tilde{A}_x^*(\vec{z} | y) = e^{\left(M + \sum_{k=1}^K \Lambda_k \tilde{A}_k(\vec{z} | y) \right) x} \tag{4}$$

and

$$\tilde{A}_k(\vec{z} | y) = \int_0^{r_k^{(-1)}(y)} \tilde{A}_x^*(\vec{z} | y) b_k(x) dx + z \tilde{A}_{r_k^{(-1)}(y)}^*(\vec{z} | y) \left[1 - B_k \left(r_k^{(-1)}(y) \right) \right]. \tag{5}$$

The system of functional Eqs. (4) and (5) can be solved numerically by the methods of [2].

3. STATIONARY DISTRIBUTION OF THE SOJOURN TIME OF A CUSTOMER

Let us examine the y -system introduced at the end of the previous section. Let $W_x^*(u|y)$ denote the matrix whose element $w_{xij}^*(u|y)$ is the probability that the busy period of the y -system is less than u and the generation phase at the end of this busy period is j , provided the total length of customers at the beginning of the busy period was x and generation phase was i . Then the Laplace–Stieltjes transform $\widetilde{W}_x^*(s|y)$ of the matrix $W_x^*(u|y)$ is given by the formula

$$\widetilde{W}_x^*(s|y) = e^{\left[M + \sum_{k=1}^K \Lambda_k \widetilde{W}_k(s|y) - sE \right] x}, \tag{6}$$

where the matrix $\widetilde{W}_k(s|y)$ is determined from Eq. (6) and equations

$$\widetilde{W}_k(s|y) = \int_0^{r_k^{(-1)}(y)} \widetilde{W}_x^*(s|y) b_k(x) dx + \widetilde{W}_{r_k^{(-1)}(y)}^*(s|y) \left[1 - B_k \left(r_k^{(-1)}(y) \right) \right].$$

Let $\mathbf{w}(x|y)$ denote the vector of probabilities $w_i(x|y)$ that the virtual waiting time in the y -system in stationary mode is less than x and the generation phase is i . Note that the vector $\mathbf{w}(x|y)$, being a function of x , has at the point 0 a discontinuity $\mathbf{w}(0+|y)$ equal to the vector of stationary probabilities $w_i(0+|y)$ that there are no customers in the y -system and the generation phase is i .

Let $\widetilde{\mathbf{w}}(s|y)$ denote the Laplace–Stieltjes transform of the vector $\mathbf{w}(x|y)$ and let

$$\beta_k(s|y) = \int_0^{r_k^{(-1)}(y)} e^{-sx} b_k(x) dx + e^{-sr_k^{(-1)}(y)} \left[1 - B_k \left(r_k^{(-1)}(y) \right) \right]$$

denote the Laplace–transform of the length of a k th-type customer in the y -system. Then $\widetilde{\mathbf{w}}(s|y)$ is given by the formula (see [2])

$$\widetilde{\mathbf{w}}(s|y) = s\mathbf{w}(0+|y) \left[sE + M + \sum_{k=1}^K \Lambda_k \beta_k(s|y) \right]^{-1},$$

where $\mathbf{w}(0+|y)$ can be found from the equation

$$\mathbf{w}(0+|y) \left[M + \sum_{k=1}^K \Lambda_k \widetilde{W}_k(0|y) \right] = \mathbf{0}$$

under the normalization condition

$$\mathbf{w}(0+|y) \mathbf{1} = 1 - \rho_y,$$

where

$$\rho_y = \sum_{k=1}^K \lambda_k \left(\int_0^{r_k^{(-1)}(y)} x b_k(x) dx + r_k^{(-1)}(y) \left[1 - B_k \left(r_k^{(-1)}(y) \right) \right] \right)$$

is the traffic intensity of the y -system.

Let us turn back to the initial system.

Let $\mathbf{v}_k(u|x)$ denote the vector of stationary probabilities $v_{ki}(u|x)$ that the sojourn time of a k th-type customer of length x is less than u and the generation phase at the instant of completion of his service is i .

To find $\mathbf{v}_k(u|x)$, let us take a k th-type customer of length x , i.e., of maximal priority index $y = r_k(x)$. Note that customers of priority index greater than y in the system have no influence on

the service of the customer we have chosen. Furthermore, customers of maximal priority index less than y in the system at the instant of arrival of the chosen customer and the customers that arrive in the course of service of the chosen customer necessarily quit the system earlier than the chosen customer. Finally, customers of maximal priority index greater than y whose priority index at the instant of arrival of the chosen customer was less than y or customers that arrive in the course of service of the chosen customer will have a priority index equal to y at the instant when the chosen customer quits the system. Therefore, the time of sojourn of the chosen customer in the system is equal to the time between the instant of his arrival and the instant when the system is free of customers of priority index less than y ; in other words, is equal to the time between the instant of arrival of the chosen customer at the y -system and the instant of completion of the busy period of the y -system. Hence, since the stationary distribution of the virtual waiting time in the y -system is defined by the vector $\mathbf{w}(v|y)$, we obtain

$$\mathbf{v}_k(u|x) = \frac{1}{\lambda_k} \int_0^\infty \mathbf{w}(dv|y) \Lambda_k W_{x+v}^*(u|y),$$

or, by virtue of (6),

$$\tilde{\mathbf{v}}_k(s|x) = \frac{1}{\lambda} \int_0^\infty \mathbf{w}(dv|y) \Lambda_k \tilde{W}_{x+v}^*(s|y) = \frac{1}{\lambda} \int_0^\infty \mathbf{w}(dv|y) \Lambda_k e^{\left[M + \sum_{k=1}^K \Lambda_k \tilde{W}_k(s|y) - sE \right] (x+v)} \quad (7)$$

in terms of the generating function.

The vector $\mathbf{v}_k(u)$ of stationary probabilities $v_{ki}(u)$ that the time of sojourn of a k th-type customer is less than u and generation exists in phase i at the instant of completion of his service is given by the formula

$$\mathbf{v}_k(u) = \int_0^\infty \mathbf{v}_k(u|x) dB_k(x). \quad (8)$$

Finally, the stationary distribution $V_k(u|x)$ of the sojourn time of a k th-type customer of length x and the stationary distribution $V_k(u)$ of a k th-type customer are

$$V_k(u|x) = \mathbf{v}_k(u|x) \mathbf{1} \quad (9)$$

and

$$V_k(u) = \mathbf{v}_k(u) \mathbf{1}. \quad (10)$$

Formulas (7)–(10) define the stationary characteristics related to the time of sojourn of a customer in the system.

REFERENCES

1. D'Apice, C., Manzo, R., and Pechinkin, A.V., A $MAP_K/G_K/1$ Queueing System of Finite Capacity with Generalized Foreground-Background Processor Sharing Discipline, *Avtom. Telemekh.*, 2004, no. 11, pp. 114–121.
2. Bocharov, P.P., D'Apice, C., Pechinkin, A.V., and Salerno, S., *Queueing Theory*, Boston: VSP, 2004.

This paper was recommended for publication by V.M. Vishnevskii, a member of the Editorial Board