

On the Preemptive Priority Queues

OLDŘICH VAŠIČEK

This article deals with the preemptive priority queues with single server in which the preempted items do not return to service and are lost. In the case of Poisson input and exponential service times the stationary distributions of the number of items of two priority classes are given. The Laplace-Stieltjes transform of the waiting time distribution has been derived.

1. INTRODUCTION

Recently, the priority queueing problems have been of increasing interest in the theory of queues. Before studying our special case, a classification of these problems will be given according to R. G. Miller Jr. [1].

Units (or items) with m types of priorities (1), (2), ..., (m) arrive at a service mechanism. A type (i) item will be selected for service in preference to a type (j) one if $i < j$, and the first-come, first-served policy is kept within each class. When a type (j) item is in service and a type (i) item arrives, $i < j$, two primary disciplines are distinguished. In the case of head-of-the-line discipline the unit of type (j) completes service but the type (i) unit is served ahead of any other lower priority item. This problem has been studied by A. Cobham [2], P. M. Morse [3], R. G. Miller [1] and others. The preemptive discipline withdraws the type (j) unit from service and replaces it by the type (i) item. When the preemptive scheme in which displaced lower priority items return to service is considered, there are two cases. The preemptive resume policy allows the preempted item to resume service at the point at which it was preempted so that its service time upon reentry has been reduced by the amount of time the item has already spent in service. The preemptive repeat policy requires the preempted item to commence service again at the beginning. H. White and L. S. Christie [4] and others have investigated this type of priority queues. In this paper another case of preemptive priority, where items of lower type are lost if displaced, will be studied.

Let us have a single channel service system with infinite waiting line. Let there be m types of priorities, (1), (2), ..., (m) (the smaller the number, the higher the priority). Suppose that the input process of type (i) items is Poisson with arrival rate λ_i , $i = 1, 2, \dots, m$, and that the input processes are independent. Units wait to be served according to preemptive discipline; when an item of type (j) is in service and a type (i) one arrives, $i < j$, the lower priority unit is displaced from service and lost without return. Let the service time distribution for a type (i) unit be exponential with parameter μ_i , $i = 1, 2, \dots, m$.

The queueing process for items of types (1), (2), ..., (i), $i < m$ is independent on the input process of types ($i + 1$), ($i + 2$), ..., (m) items and it is the same as in a system where only the first i classes of priorities are considered. Especially, the queueing process of type (1) units is described by the characteristics corresponding to the system M/M/1.

The time t_i a type (i) item spent in service has the distribution function

$$\begin{aligned} P[t_i \leq t] &= 1 - P[t_i > t, \tau_1 > t, \tau_2 > t, \dots, \tau_{i-1} > t] = \\ &= 1 - P[t_i > t] \cdot P[\tau_1 > t] \cdot P[\tau_2 > t] \dots P[\tau_{i-1} > t] = \\ &= 1 - \exp\left[-\left(\mu_i + \sum_{j=1}^{i-1} \lambda_j\right) t\right], \end{aligned}$$

where t_i is the service time of a type (i) item if no interruption of the service has occurred and τ_j , $j = 1, 2, \dots, i - 1$ is the length of the time interval to the arrival of a type (j) item. The mean value of the time spent by an item of type (i) in service is then

$$\frac{1}{\mu_i + \sum_{j=1}^{i-1} \lambda_j},$$

and the necessary and sufficient condition for the existence of the steady-state solution is

$$(1) \quad \sum_{i=1}^m \frac{\lambda_i}{\mu_i + \sum_{j=1}^{i-1} \lambda_j} < 1.$$

If the condition

$$\sum_{i=1}^i \frac{\lambda_i}{\mu_i + \sum_{j=1}^{i-1} \lambda_j} < 1$$

holds and if, at the same time

$$\sum_{i=1}^{i+1} \frac{\lambda_i}{\mu_i + \sum_{j=1}^{i-1} \lambda_j} > 1$$

for some l , $1 \leq l < m$, the queue will become saturated with type $(l + 1)$ items but the number of first l classes of priority will be in statistical equilibrium. 149

We shall denote

$$\rho_i = \frac{\lambda_i}{\mu_i + \sum_{j=1}^{i-1} \lambda_j}.$$

When the condition (1)

$$\rho_1 + \rho_2 + \dots + \rho_m < 1$$

is satisfied, the probability

$$(2) \quad p_0 = 1 - \rho_1 - \rho_2 - \dots - \rho_m$$

is the stationary probability that there is no item of any class in system, and

$$(3) \quad p_i = \rho_i$$

is the stationary probability that a type (i) unit is in service.

The probability that a type (i) item will not finish its service is given by

$$\begin{aligned} P\left[\min_{1 \leq j \leq i-1} j\tau < i b\right] &= 1 - P[1\tau > i b, 2\tau > i b, \dots, i-1\tau > i b] = \\ &= 1 - \prod_{j=1}^{i-1} P[j\tau > i b] = \\ &= 1 - \prod_{j=1}^{i-1} \int_0^{\infty} (1 - e^{-\mu t}) \lambda_j e^{-\lambda_j t} dt = \\ &= 1 - \prod_{j=1}^{i-1} \mu_j (\mu_i + \lambda_j). \end{aligned}$$

2. GENERATING FUNCTION FOR THE STATIONARY PROBABILITIES ($m = 2$)

We shall consider in more details the case of two classes of priority (1), (2). All our results will hold true for queuing process of the first two categories of priority even in the system with $m > 2$ classes.

Let

$$\rho_1 + \rho_2 = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2 + \lambda_1} < 1.$$

Let p_{nk} , $n, k = 0, 1, 2, \dots$ be the steady-state probabilities that there are n type (1) items and k type (2) items in the system. Then the steady-state equations may be

150 written

$$\begin{aligned}
 (4) \quad & \mu_1 p_{10} + \mu_2 p_{01} - (\lambda_1 + \lambda_2) p_{00} = 0, \\
 & \lambda_1 p_{00} + \lambda_1 p_{01} + \mu_1 p_{20} - (\lambda_1 + \lambda_2 + \mu_1) p_{10} = 0, \\
 & \lambda_2 p_{0k-1} + \mu_1 p_{1k} + \mu_2 p_{0k+1} - (\lambda_1 + \lambda_2 + \mu_2) p_{0k} = 0, \quad k > 0, \\
 & \lambda_1 p_{n-10} + \mu_1 p_{n+10} - (\lambda_1 + \lambda_2 + \mu_1) p_{n0} = 0, \quad n > 1, \\
 & \lambda_1 p_{0k+1} + \lambda_2 p_{1k-1} + \mu_1 p_{2k} - (\lambda_1 + \lambda_2 + \mu_1) p_{1k} = 0, \quad k > 0, \\
 & \lambda_1 p_{n-1k} + \lambda_2 p_{nk-1} + \mu_1 p_{n+1k} - (\lambda_1 + \lambda_2 + \mu_1) p_{nk} = 0, \quad n > 1, k > 0,
 \end{aligned}$$

which together with the condition

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} p_{nk} = 1$$

determine the probabilities p_{nk} .

We shall define the generating functions

$$\begin{aligned}
 (5) \quad Q(x, y) &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} x^n y^k p_{nk}, \\
 A_n(y) &= \sum_{k=0}^{\infty} y^k p_{nk}, \quad n = 0, 1, 2, \dots, \\
 H(y) &= \sum_{k=0}^{\infty} y^k \sum_{n=0}^k p_{nk-n} = Q(y, y).
 \end{aligned}$$

Multiplying the equations (4) by the appropriate powers of x and y and adding them, we obtain

$$\begin{aligned}
 & Q(x, y) \cdot \left(\lambda_1 x + \lambda_2 y + \frac{\mu_1}{x} - \lambda_1 - \lambda_2 - \mu_1 \right) = \\
 & = A_0(y) \left(\lambda_1 x - \lambda_1 \frac{x}{y} + \frac{\mu_1}{x} - \frac{\mu_2}{y} - \mu_1 - \mu_2 \right) - p_0 (\lambda_1 x + \mu_2) \left(1 - \frac{1}{y} \right),
 \end{aligned}$$

or

$$(6) \quad Q(x, y) = \frac{A_0(y) \left[(\lambda_1 x + \mu_2) \left(1 - \frac{1}{y} \right) - \mu_1 \left(1 - \frac{1}{x} \right) \right] - p_0 (\lambda_1 x + \mu_2) \left(1 - \frac{1}{y} \right)}{\lambda_1 (x-1) + \lambda_2 (y-1) - \mu_1 \left(1 - \frac{1}{x} \right)},$$

where $p_0 = p_{00} = 1 - \varrho_1 - \varrho_2$ is the probability that there is no item in the system. Then

$$(7) \quad H(y) = \frac{A_0(y) (\lambda_1 y - \mu_1 - \mu_2) - p_0 (\lambda_1 y + \mu_2)}{(\lambda_1 + \lambda_2) y - \mu_1}.$$

Multiplying the equations (4) by y^k and summing them on k yields the following set of equations in $A_n(y)$ 151

$$(8) \quad \begin{aligned} & - \left(\lambda_1 + \lambda_2 - \mu_2 - \lambda_2 y - \frac{\mu_2}{y} \right) A_0(y) + \mu_1 A_1(y) = \mu_2 p_0 \left(\frac{1}{y} - 1 \right) \\ & \frac{\lambda_1}{y} A_0(y) - (\lambda_1 + \lambda_2 + \mu_1 - \lambda_2 y) A_1(y) + \mu_1 A_2(y) = \lambda_1 p_0 \left(\frac{1}{y} - 1 \right) \\ & \lambda_1 A_{n-1}(y) - (\lambda_1 + \lambda_2 + \mu_1 - \lambda_2 y) A_n(y) + \mu_1 A_{n+1}(y) = 0, \quad n > 1. \end{aligned}$$

Introducing another set of functions $\{A_n^*(y)\}$, $n = -1, 0, 1, 2, \dots$ by the relations

$$(9) \quad \begin{aligned} A_{-1}^*(y) &= \frac{1}{\lambda_1} A_0(y) \left[(\lambda_1 + \lambda_2 + \mu_2 - \lambda_1 y) \left(\frac{1}{y} - 1 \right) + \frac{\mu_1}{y} \right] - \\ & \quad - \frac{1}{\lambda_1} p_0 \left(\frac{1}{y} - 1 \right) (\lambda_1 + \lambda_2 + \mu_1 + \mu_2 - \lambda_2 y), \\ A_0^*(y) &= \frac{1}{y} A_0(y) - p_0 \left(\frac{1}{y} - 1 \right), \\ A_n^*(y) &= A_n(y), \quad n > 0 \end{aligned}$$

gives an equivalent form of the equations (8) as follows

$$(10) \quad \lambda_1 A_{n-1}^*(y) - (\lambda_1 + \lambda_2 + \mu_1 - \lambda_2 y) A_n^*(y) + \mu_1 A_{n+1}^*(y) = 0, \quad n = 0, 1, 2, \dots$$

The solution of this set of equations is given by

$$A_n^*(y) = C(y) \alpha^n(y),$$

where

$$(11) \quad \alpha(y) = \frac{1}{2\mu_1} \{ \lambda_1 + \lambda_2(1-y) + \mu_1 - \sqrt{[\lambda_1 + \lambda_2(1-y) + \mu_1]^2 - 4\mu_1\lambda_1} \}$$

is that of the roots of the quadratic equation

$$(12) \quad \lambda_1 - (\lambda_1 + \lambda_2 + \mu_1 - \lambda_2 y) \alpha(y) + \mu_1 \alpha^2(y) = 0,$$

which is smaller than unity. Therefore

$$(13) \quad A_n^*(y) = \alpha(y) A_{n-1}^*(y), \quad n = 0, 1, 2, \dots$$

152 Substituting $A_{-1}^*(y)$, $A_0^*(y)$ from (9) into (13), we get the equation, enabling us to determine $A_0(y)$

$$(14) \quad A_0(y) = \frac{p_0[\mu_1 \alpha(y) + \mu_2](y-1)}{[\lambda_1 + \lambda_2(1-y) + \mu_2]y - (\mu_1 \alpha(y) + \mu_2)}.$$

This expression can then be substituted back in (6) and (7) to obtain a complete expression for the generating functions $Q(x, y)$ and $H(y)$.

The expected number of items in the system is

$$L = H'(1) = \frac{\rho_1}{1 - \rho_1} + \frac{\rho_1 \frac{\lambda_2}{\mu_1 - \lambda_1} + \rho_2(1 - \rho_1)}{1 - \rho_1 - \rho_2},$$

and the mean numbers of type (1) items and of type (2) items are

$${}_1L = \frac{\rho_1}{1 - \rho_1},$$

$${}_2L = \frac{\rho_1 \frac{\lambda_2}{\mu_1 - \lambda_1} + \rho_2(1 - \rho_1)}{1 - \rho_1 - \rho_2},$$

respectively.

3. WAITING TIME DISTRIBUTION ($m = 2$)

We shall denote $W(t)$ the waiting time distribution and ${}_1W(t)$, ${}_2W(t)$ the waiting time distributions for a type (1) items and for a type (2) items, respectively. Let $\gamma(s)$, ${}_1\gamma(s)$, ${}_2\gamma(s)$ be the respective Laplace-Stieltjes transforms. Then

$$(15) \quad \gamma(s) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot {}_1\gamma(s) + \frac{\lambda_2}{\lambda_1 + \lambda_2} {}_2\gamma(s),$$

where ${}_1\gamma(s)$ is given by

$${}_1\gamma(s) = (1 - \rho_1) \frac{\mu_1 + s}{\mu_1 - \lambda_1 + s}.$$

Now we shall find ${}_2\gamma(s)$. Let ${}_2w_{nk}$, $n, k = 0, 1, 2, \dots$ be the waiting time of a type (2) item before which there are n type (1) units and k type (2) units in the system. Let

$${}_2\gamma_{nk}(s) = \int_0^\infty e^{-st} dP[{}_2w_{nk} < t].$$

The type (2) items may be served only in the intervals between the busy periods of the system with items of the class (1). Let there be n type (1) units in the system in definite instant and let g_n be the length of the time period from this instant to the nearest moment when no type (1) item is in the system. Let further g denote the busy period of the system with the type (1) items and v the length of the time interval in which no type (1) item is in the system. Then

$$(16) \quad \begin{aligned} {}_2W_{n0} &= g_n \\ {}_2W_{nk} &= g_n + z_k, \quad k > 0, \end{aligned}$$

where z_k is the waiting time of a type (2) item before which there are k type (2) items and no type (1) item in the system.

Let v items of type (2) finish service within the period v and let $b^{(1)}, b^{(2)}, \dots, b^{(v)}$ be their service times. Besides, one type (2) unit, the service of which has been interrupted, leaves the system at the termination of the period v . Therefore

$$(17) \quad \begin{aligned} z_k &= v + g + z_{k-v-1}, & \text{if } v = 0, 1, \dots, k-2, \\ &= v + g, & \text{if } v = k-1, \\ &= b^{(1)} + b^{(2)} + \dots + b^{(k)}, & \text{if } v \geq k. \end{aligned}$$

Let $G_n(t), G(t), V(t), Z_k(t), B(t)$ be the distribution functions of the random variables $g_n, g, v, z_k, b^{(i)}$, respectively. We have

$$\begin{aligned} V(t) &= 1 - e^{-\lambda_1 t}, \\ B(t) &= 1 - e^{-\mu_2 t}. \end{aligned}$$

The Laplace-Stieltjes transform $\Gamma(s)$ of the distribution function $G(t)$ can be obtained from the equation (Takács)

$$\Gamma(s) = \frac{\mu_1}{\mu_1 + s + \lambda_1 - \lambda_1 \Gamma(s)};$$

it follows

$$(18) \quad \Gamma(s) = \frac{1}{2\lambda_1} \{ \mu_1 + \lambda_1 + s - \sqrt{[(\mu_1 + \lambda_1 + s)^2 - 4\lambda_1\mu_1]} \}.$$

Now we shall determine $G_n(t)$. Random variable g_n is a sum of service times of n type (1) units in the system and service times of all the type (1) items which will be arriving during these and all subsequent service periods. It is independent on the order in which these items are served; we can assume that the $n-1$ items waiting in the queue at the beginning will be served at last. Therefore, g_n equals to the sum

154 of two independent random variables with distribution function $G(t)$, $G_{n-1}(t)$ and hence

$$G_n(t) = \int_0^t G_{n-1}(t-x) dG(x).$$

If $\Gamma_n(s)$ is the Laplace-Stieltjes transform of $G_n(t)$, we have

$$\Gamma_n(s) = \Gamma_{n-1}(s) \Gamma(s), \quad n = 2, 3, \dots$$

and

$$(19) \quad \Gamma_n(s) = \Gamma^n(s), \quad n = 1, 2, 3, \dots,$$

because $\Gamma_1(s) = \Gamma(s)$.

On the right-hand side of equations (17) the summands are independent random variables. The random variable v is Poisson distributed with parameter μ_2 . Then the distribution function of z_k can be expressed in the form

$$\begin{aligned} Z_k(t) = & \int_0^t \sum_{j=0}^{k-2} e^{-\mu_2 y} \frac{(\mu_2 y)^j}{j!} \int_0^{t-y} Z_{k-j-1}(t-y-x) dG(x) dV(y) + \\ & + \int_0^t e^{-\mu_2 y} \frac{(\mu_2 y)^{k-1}}{(k-1)!} G(t-y) dV(y) + \\ & + \int_0^t B_k(y) dV(y) + B_k(t) \int_t^\infty dV(y), \quad k > 0, \end{aligned}$$

where

$$B_k(t) = 1 - \sum_{j=0}^{k-1} e^{-\mu_2 t} \frac{(\mu_2 t)^j}{j!}$$

is the k -fold convolution of $B(t)$. Substituting for $V(t)$, $B_k(t)$, we have

$$\begin{aligned} (20) \quad Z_k(t) = & \int_0^t dy \sum_{j=0}^{k-2} \lambda_1 e^{-(\lambda_1 + \mu_2)y} \frac{(\mu_2 y)^j}{j!} \int_0^{t-y} Z_{k-j-1}(t-y-x) dG(x) + \\ & + \int_0^t \lambda_1 e^{-(\lambda_1 + \mu_2)y} \frac{(\mu_2 y)^{k-1}}{(k-1)!} G(t-y) dy - \\ & - \int_0^t \sum_{j=0}^{k-1} \lambda_1 e^{-(\lambda_1 + \mu_2)y} \frac{(\mu_2 y)^j}{j!} dy + \\ & + 1 - e^{-\lambda_1 t} \sum_{j=0}^{k-1} e^{-\mu_2 t} \frac{(\mu_2 t)^j}{j!}, \quad k > 0. \end{aligned}$$

Let

155

$$\zeta_k(s) = \int_0^{\infty} e^{-st} dZ_k(t).$$

Applying the Laplace-Stieltjes transform to (20) yields

$$\begin{aligned} \zeta_k(s) &= \sum_{j=0}^{k-2} \frac{\lambda_1}{\mu_2} \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^{j+1} \Gamma(s) \zeta_{k-j-1}(s) + \\ &+ \frac{\lambda_1}{\mu_2} \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^k \Gamma(s) - \sum_{j=0}^{k-1} \frac{\lambda_1}{\mu_2} \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^{j+1} + \\ &+ 1 - \frac{s}{s + \lambda_1} \left[1 - \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^k \right]. \end{aligned}$$

After some computation

$$\begin{aligned} \zeta_k(s) &= \frac{\lambda_1}{\mu_2} \Gamma(s) \sum_{j=1}^{k-1} \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^j \zeta_{k-j}(s) + \\ &+ \left(1 + \frac{\lambda_1}{\mu_2} \Gamma(s) \right) \left(\frac{\mu_2}{s + \mu_2 + \lambda_1} \right)^k, \quad k > 0. \end{aligned}$$

This recurrent relation is solved to give

$$(21) \quad \zeta_k(s) = \zeta^k(s),$$

where

$$(22) \quad \zeta(s) = \frac{\mu_2 + \lambda_1 \Gamma(s)}{\mu_2 + \lambda_1 + s}.$$

Now

$${}_2\gamma_{nk}(s) = \Gamma^n(s) \zeta^k(s),$$

and the Laplace-Stieltjes transform ${}_2\gamma(s)$ of ${}_2W(t)$ is

$${}_2\gamma(s) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} {}_2\gamma_{nk}(s) \cdot p_{nk} = Q(\Gamma(s), \zeta(s)).$$

Finally, the Laplace-Stieltjes transform $\gamma(s)$ of the waiting time distribution is given by

$$(23) \quad \gamma(s) = \frac{\lambda_1}{\lambda_1 + \lambda_2} (1 - q_1) \frac{\mu_1 + s}{\mu_1 - \lambda_1 + s} + \frac{\lambda_2}{\lambda_1 + \lambda_2} Q(\Gamma(s), \zeta(s)).$$

156 The mean value \bar{W} of the waiting time is obtained by differentiating (23) and setting $s = 0$. This gives

$$\bar{W} = \frac{1}{\lambda_1 + \lambda_2} \left(\frac{\rho_1^2}{1 - \rho_1} + \frac{\rho_1 \frac{\lambda_2}{\mu_1 - \lambda_1} + \rho_2^2}{1 - \rho_1 - \rho_2} \right),$$

which corresponds to the relation $\bar{W} = L_q / (\lambda_1 + \lambda_2)$, where $L_q = L - \rho_1 - \rho_2$ is the expected number of items in queue.

(Received July 19th, 1966.)

REFERENCES

- [1] R. G. Miller, Jr.: Priority Queues. *Ann. Math. Stat.* 31 (1960), 86–103.
- [2] A. Cobham: Priority Assignment in Waiting Line Problems. *J. Opns Res. Soc. Am.* 2 (1954), 70–76. Priority Assignment — A Correction. *J. Opns Res. Soc. Am.* 3 (1955), 547.
- [3] P. M. Morse: Queues, Inventories, and Maintenance. J. Wiley and Sons, New York 1958.
- [4] H. White, L. S. Christie: Queueing With Preemptive Priorities or With Breakdown. *J. Opns Res. Soc. Am.* 6 (1958), 79–95.
- [5] Марьянович Т. П.: Однолинейная система массового обслуживания с ненадежным прибором. *Укр. мат. ж.* 14 (1962), 4, 417–422.

VÝTAH

O silných přednostech v systémech hromadné obsluhy

OLDŘICH VAŠIČEK

V práci je vyšetřován systém hromadné obsluhy s jednou obsluhovací linkou a jednou nekonečnou linkou pro čekání, do něhož vstupují dva typy zákazníků. První typ je silně nadřazen druhému, tzn. zákazník první kategorie, který v okamžiku příchodu nalezne linku obsazenu jednotkou druhého typu, ji vyřazuje z obsluhy. Zákazník nižší kategorie, který byl takto vyřazen, se nevrací do fronty a ztrácí se. Uvnitř obou kategorií je dodržována obsluha v pořadí příchodů. Oba vstupní proudy jsou Poissonovy, doba obsluhy má rozložení exponenciální s různými parametry pro oba typy. Za těchto předpokladů jsou vyvozena rozložení pravděpodobnosti délky fronty pro oba typy ve formě vytvořující funkce a podán výraz pro rozložení doby čekání. Některé výsledky (podmínka pro existenci stacionárního stavu, pravděpodobnosti dokončení obsluhy) platí i pro n typů vstupů.

Ing. Oldřich Vašíček, Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.