ON THE PERFORMANCE OF THE $M_1,M_2/G_1,G_2/1$ RETRIAL QUEUE WITH PRE-EMPTIVE RESUME POLICY

Leila BOUTARFA
University of 08 May 1945, Guelma, Algeria
Leila.bout@gmail.com

Natalia DJELLAB
Laboratory LaPS, University Badji Mokhtar, Annaba, Algeria
djellab@yahoo.fr

Received: February 2013 / Accepted: December 2013

Abstract: Priority mechanism is an invaluable scheduling method that allows customers to receive different quality of a service. Service priority is clearly today a main feature of the operation of any manufacturing system. We are interested in an $M_1,M_2/G_1,G_2/1$ priority retrial queue with pre-emptive resume policy. For the model in question, we discuss the problem of ergodicity, and by using the method of supplementary variables, we find the partial generating functions of the steady state system state distribution. Moreover, some pertinent performance measures are obtained and numerical study is also performed.

Keywords: Retrial Queue, Pre-emptive Resume Policy, Steady-state Distribution, Performance Measure, Priority Customer.

MSC: 60K25.

1. INTRODUCTION

Retrial queuing systems or systems with repeated attempts are characterized by the requirement that customers finding the service space busy must join the retrial group and reply for service at random intervals. A review of the main results on this topic can be found in [7], [12]. A comparison between retrial queues and their standard counter parts with classical waiting line is released in [3]. Retrial queues have been widely used as mathematical models of different communication systems: shared bus local area
networks operating under transmission protocols like CSMA/CD (Carrier Sense Multiple Access with Collision Detection), cellular mobile networks, IP networks [9], [1], [6].

In this paper, we are interested in single server retrial queues with priority phenomenon. Priority mechanism is an invaluable scheduling method that allows customers to receive different quality of service. Service priority is clearly today a main feature of the operation of any manufacturing system. For this reason, the priority queues, in particular the retrial priority ones, have received considerable attention in the literature [2], [8], [11]. A review of main results of such retrial models can be found in the survey paper of Choi and Chang (1999) [4].

In this work, we deal with an $M_1,M_2/G_1,G_2/1$ priority retrial queue with pre-emptive resume policy, which can be used to model a situation (frequently observed in some information desks) where a single agent answers the telephone calls and serves the present customer. In this context, the telephone calls have pre-emptive priority over the present customer. For the model in question, we discuss the problem of ergodicity, and by using the method of supplementary variables, we find the partial generating functions of the steady state system state distribution. Moreover, some pertinent performance measures are obtained, and numerical study is also performed.

The rest of the paper is organized as follows. The next section contains the model description. In section 3, we describe the structure of the embedded Markov chain and obtain ergodicity condition. The steady state distribution of the system state, as well as, some performance measures are obtained in section 4. We conclude this work by numerical illustrations in section 5.

### 2. MATHEMATICAL MODEL

We consider a single server queuing system at which two different types of primary customers arrive, according to independent Poisson processes with rates $\lambda_1$ and $\lambda_2$, respectively. Customers from the first flow of rate $\lambda_1 > 0$ have a pre-emptive priority over customers from the second flow of rate $\lambda_2 > 0$. Thus, the following rules govern the dynamic of the customers:

- Any arriving primary customer, finding the server idle, immediately occupies the server and leaves the system after service completion.
- Any arriving priority customer, finding the server busy by another priority customer, joins the retrial group (orbit). The retrial times follow an exponential law with rate $\theta > 0$.
- Any arriving priority customer (primary or orbiting), finding the server busy by a non-priority customer, goes directly into the server. Non-priority customer, whose service was interrupted, persists in the service area until the completion of priority customer service so to start his service again from where it was interrupted (pre-emptive resume policy).
- Any arriving primary non-priority customer finding the server busy leaves the system without service.

The service times of both types of customers follow a general distribution with distribution function $B_i(x)$, $i \in \{1,2\}$, and Laplace-Stieltjes transform
\[
\tilde{B}_i(s) = \int_0^\infty \exp(-sx)dB_i(x), \quad i \in \{1, 2\} \quad \text{and} \quad \Re(s) > 0. \quad \text{Let} \quad \beta_{i,k} = (-1)^k \tilde{B}_i^{(k)}(0) \quad \text{be the} \ k\text{-th moment of the service time about the origin;}
\]

\[
b_i(x) = \frac{B'_i(x)}{1-B_i(x)}, \quad i \in \{1, 2\}, \quad \text{be the instantaneous service intensity of the customer type} \ i \quad \text{given that the elapsed service time is equal to} \ x; \quad K_i(z_1, z_2) = \tilde{B}_i(\lambda_1 - \lambda_2 z_1 + \lambda_2 - \lambda_1 z_2) \quad \text{be the generating function of the number of primary customers of both types that arrive during the service time of an} \ i\text{-th type customer} \ (i \in \{1, 2\}). \quad \text{Finally, we admit the hypothesis of mutual independence between all random variables defined above.}
\]

The state of the system at time \( t \) can be described by means of the process

\[
\left\{ C(t), N_0(t), \xi_1(t), \xi_2(t), t \geq 0 \right\},
\]

where \( N_0(t) \) is the number of priority customers in the orbit, and \( C(t) \) represents the state of the service station at time \( t \). We define \( C(t) \) as equal to 0, 1, 2 or 3 depending on whether the server is idle, a priority customer is served and there is no interrupted non-priority customer in the service station, a non-priority customer is served, or a priority customer is served and there is an interrupted non-priority customer in the service station. If \( C(t) \in \{1, 3\} \) (If \( C(t) = 2 \)), \( \xi_1(t) \) (\( \xi_2(t) \)) represents the elapsed service time of the priority customer (the non-priority customer) in service at time \( t \).

The transitions among states are defined as follows:

- \( C(t) = 0, t \geq 0 \)
- \( C(t) = 1, t \geq 0 \)
- \( C(t) = 2, t \geq 0 \)
- \( C(t) = 3, t \geq 0 \)

\[
\begin{align*}
& \lambda_1, \\
& \lambda_2, \\
& b_1(x), \\
& b_2(x), \\
& \sigma \\
\end{align*}
\]

\[ \text{Figure 1} \]
3. ERGODICITY CONDITION

Let $t_d$ be the time of the $d$-th departure, $N_{1,d}$ ($N_{2,d}$) be the number of priority customers (non-priority customers) in the orbit (the service station) just before the time $t_d$. We have the following fundamental equations:

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d};$$  
(2)

$$N_{2,d} = N_{2,d-1} \times B_d \times Y_{2,d} = 0;$$  
(3)
where \( V_d \) is equal to 0 or to 1, according to the type of \( d \)-th served customer as primary priority customer or a repeated priority one; \( Y_{1,d} \) (\( Y_{2,d} \)) is the number of priority/non-priority customers arriving at the system during the service time of the \( d \)-th customer; \( B_d \) is equal to 0 or to 1 depending on whether the \( d \)-th served customer is a non-priority customer or a priority one. Let \( U_d \) be the type of the \( d \)-th served customer (which can be 1 or 2 depending on whether the customer in question is a priority customer or a non-priority one).

The random vector \( (U_d, V_d) \) depends on the history of the system before time \( I_d \) only through the vector \( (N_{1,d-1}, N_{2,d-1}) \). Its conditional distribution is defined by

\[
P\left((U_d = 1, V_d = 1) \mid (N_{1,d-1}, N_{2,d-1}) = (m, n)\right) = \frac{m \theta}{\lambda_1 + \lambda_2 + m \theta};
\]

\[
P\left((U_d = 1, V_d = 0) \mid (N_{1,d-1}, N_{2,d-1}) = (m, n)\right) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m \theta};
\]

\[
P\left((U_d = 2, V_d = 1) \mid (N_{1,d-1}, N_{2,d-1}) = (m, n)\right) = 0;
\]

\[
P\left((U_d = 2, V_d = 0) \mid (N_{1,d-1}, N_{2,d-1}) = (m, n)\right) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m \theta}.
\]

The random vector \( (Y_{1,d}, Y_{2,d}) \) depends on the events that have occurred during the service time of the \( d \)-th customer only through \( U_d \), and has the following conditional distribution:

\[
P\left((Y_{1,d}, Y_{2,d}) = (m, n) \mid (U_d = l)\right)
= \int_0^{\infty} \exp\left(-\lambda_1 x\right) \left(\frac{\lambda_2}{m!}\right)^n x^n \exp\left(-\lambda_2 x\right) \left(\frac{\lambda_2}{n!}\right)^m x^m dB_l (x) = k_{l,m,n},
\]

with \( l \in \{1, 2\} \), \( m \geq 0 \) and \( n \in \{0, 1\} \).

The sequence \( \{X_d = (U_d, N_{1,d}, N_{2,d}), d \geq 1\} \) forms a Markov chain with state space \( S = \{1, 2\} \times \mathbb{Z}_+ \times \{0, 1\} \), which is the embedded Markov chain for our queuing system. Its one-step transition probabilities \( \eta((i,m,n),(c,i,j)) = P(X_d = (c,i,j) \mid X_{d-1} = (l,m,n)) \), where \( c \in \{1, 2\} \), are given by

\[
\eta((i,m,n),(2,i,j)) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + m \theta} k_{2,j-m,n} \cdot n \in \{0, 1\} \ \text{and} \ m = i;
\]  

(4)
The next question to be investigated is the ergodicity of our chain. Since the fundamental equations (2)-(3) have a recursive structure, we will use Foster’s criterion [5]. To this end, consider the following Lyapunov function on the state space $S$ defined above:

$$f((l,m,n)) = m + (1 - \rho_1)n,$$

where $\rho_1 = \lambda_1 \beta_{1,1}$. The mean drift $x_{(l,m,n)} = E\left[\frac{(f(X_d) - f(X_{d-1}))}{X_{d-1} = (l,m,n)}\right]$ can be obtained in the following manner:

i) If $n = 0$, the equations (2)-(3) become

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}; \quad N_{2,d} = 0.$$

Under these circumstances, $f(X_{d-1}) = m$ and $f(X_d) = m - V_d + Y_{1,d}$, which implies that

$$x_{(l,m,n)}^{(1)} = E\left[\frac{(-V_d + Y_{1,d})}{((N_{1,d-1}, N_{2,d-1}) = (m,0))}\right] = \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \frac{\lambda_1 \beta_{1,1}}{\lambda_1 + \lambda_2 + m\theta}.$$

ii) If $n = 1$, from (2)-(3) we have

$$N_{1,d} = N_{1,d-1} - V_d + Y_{1,d} = m - V_d + Y_{1,d}; \quad N_{2,d} = B_d$$

and $f(X_d) - f(X_{d-1}) = -V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1)$.

Thus

$$x_{(l,m,n)}^{(2)} = E\left[\frac{(-V_d + Y_{1,d} + (1 - \rho_1)(B_d - 1))}{((N_{1,d-1}, N_{2,d-1}) = (m,1))}\right] = -\frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} + \frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} \frac{\lambda_1 \beta_{1,1} + (1 - \rho_1)\left(\frac{\lambda_1 + m\theta}{\lambda_1 + \lambda_2 + m\theta} - 1\right)}{\lambda_1 + \lambda_2 + m\theta}.$$

At present, consider $x_{(l,m,n)}^{(1)} = \lim_{m \to \infty} x_{(l,m,n)}^{(1)} = \rho_1 - 1$ and $x_{(l,m,n)}^{(2)} = \lim_{m \to \infty} x_{(l,m,n)}^{(2)} = \rho_1 - 1$. Then $x^{(1)}_1 < x^{(2)}_1 < 0$ if $\rho_1 < 1$. Therefore, the sufficient condition is $\rho_1 < 1$. Since $r_{(l,m,n),(i,j)} = r_{(l,m,n),(i,j)}^\prime = 0$ for $i < m-1$, $\rho_1 < 1$ is also a necessary condition for ergodicity (according to Kaplan’s condition [10], $\rho_1 \geq 1$ gives non ergodicity of our embedded Markov chain). Finally, $\{X_d, d \geq 1\}$ is ergodic if and only if $\rho_1 < 1$. 

$$\eta_{(l,m,n),(i,j)} = \frac{\lambda_1}{\lambda_1 + \lambda_2 + m\theta} k_{i,m-1,0} + \frac{m\theta}{\lambda_1 + \lambda_2 + m\theta} k_{i,m+1,0}. \tag{5}$$
4. STEADY STATE DISTRIBUTION OF THE SYSTEM STATE

Now, we investigate the steady state distribution of the process (1) by using the method of supplementary variables. To this end, we assume that $\rho_1 < 1$ and introduce

$$R_{0,i} = \lim_{t \to \infty} P(C(t) = 0, N_x(t) = i), \quad i \geq 0;$$

$$R_{1,i}(x) = \lim_{t \to \infty} \frac{d}{dx} P(C(t) = 1, x \leq x, N_x(t) = i), \quad i \geq 0 \text{ and } x \geq 0;$$

$$R_{2,i}(y) = \lim_{t \to \infty} \frac{d}{dy} P(C(t) = 2, y \leq y, N_y(t) = i), \quad i \geq 0 \text{ and } y \geq 0;$$

$$R_{3,i}(x, y) = \lim_{t \to \infty} \frac{d}{dx} P(C(t) = 3, x, y \leq y, N_x(t) = i), \quad i \geq 0, x \geq 0 \text{ and } y \geq 0.$$

Then, following the method of supplementary variables, we find that the considered probabilities satisfy the equations of statistical equilibrium

$$\left(\lambda_1 + \lambda_2 + i\theta\right) R_{0,i} = \int_0^\infty R_{1,i}(y) b_2(y) dy + \int_0^\infty R_{1,i}(x) b_1(x) dx;$$

$$R_{1,i}(x) = -\left(h_1(x) + \lambda_1\right) R_{1,i}(x) + \lambda_1 (1 - \delta_0) R_{2,i-1}(x);$$

$$R_{2,i}(0) = \lambda_2 R_{1,i} + (i+1) \theta R_{2,i+1};$$

$$R_{2,i}(y) = -\left(h_2(y) + \lambda_1 + i\theta\right) R_{2,i}(y) + \int_0^\infty R_{2,i}(x, y) b_1(x) dx;$$

$$R_{3,i}(0, y) = \lambda_2 R_{2,i}(y) + (i+1) \theta R_{3,i+1}(y).$$

With the help of the generating functions $P_0(z) = \sum_{i=0}^\infty z^i P_{0,i}$, $P_1(z, x) = \sum_{i=0}^\infty z^i P_{1,i}(x)$, $P_2(z, y) = \sum_{i=0}^\infty z^i P_{2,i}(y)$ and $P_3(z, x, y) = \sum_{i=0}^\infty z^i P_{3,i}(x, y)$, system (6) becomes
\[
\begin{align*}
    z_i \theta \frac{dP_0(z_i)}{dz_1} &= -(\lambda_1 + \lambda_2)P_0(z_i) + \int_0^{\infty} P_2(z_1, y) b_2(y) dy + \int_0^{\infty} P_1(z_1, x) b_1(x) dx; \\
    \frac{\partial P_1(z_1, x)}{\partial x} &= (-b_1(x) - \lambda_1 + \lambda_2 z_i) P_1(z_1, x); \\
    P_1(z_1, 0) &= \lambda_1 P_0(z_i) + \theta \frac{dP_0(z_i)}{dz_1}; \\
    \frac{\partial P_2(z_1, y)}{\partial y} &= -(b_2(y) + \lambda_1) P_2(z_1, y) - \theta z_1 \frac{\partial P_2(z_1, y)}{\partial z_1} + \int_0^{\infty} P_1(z_1, x, y) b_1(x) dx; \\
    P_2(z_1, 0, y) &= \lambda_2 P_0(z_i); \\
    \frac{\partial P(z_1, x, y)}{\partial x} &= (-b_1(x) - \lambda_1 + \lambda_2 z_1) P_1(z_1, x, y); P_3(z_1, 0, y) = \theta \frac{\partial P_3(z_1, y)}{\partial z_1} + \lambda_2 P_2(z_1, y).
\end{align*}
\]

Solving the system (7) in the usual way (see for example [7]) requires fastidious computations, and gives the following partial generating functions:

\[
\begin{align*}
P_0(z_i) &= \frac{1}{2 \lambda_2 (\lambda_1 + \lambda_2 + 1)} \exp \left[ \int_1^{\infty} \frac{\lambda_1 - \lambda_2 K_1(z_1, 1)}{\theta (K_1(z_1, 1) - z_1)} dz_1 \right]; \\
P_1(z_i, x) &= \frac{\lambda_1}{2 \lambda_2 (1 + \lambda_1 + \lambda_2)} \left(1 - B_1(x)\right) e^{(\lambda_1 - \lambda_2) x} \left( \frac{1 - z_1}{K_1(z_1, 1) - z_1} \right) \\
&\quad \times \exp \left[ \int_1^{z_1} \frac{\lambda_1 - \lambda_2 K_1(z_1, 1)}{\theta (K_1(z_1, 1) - z_1)} dz_1 \right]; \\
P_1(z_1) &= \int_0^{\infty} P_1(z_1, x) dx \\
&= \frac{1}{2 \lambda_2 (1 + \lambda_1 + \lambda_2)} \left(1 - K_1(z_1, 1) \right) \exp \left[ \int_1^{\infty} \frac{\lambda_1 - \lambda_2 K_1(z_1, 1)}{\theta (K_1(z_1, 1) - z_1)} dz_1 \right]; \\
P_2(z_1, y) &= \frac{1}{2 (1 + \lambda_1 + \lambda_2)} \left(1 - B_2(y)\right) \exp \left[ \int_1^{\infty} \frac{\lambda_1 - \lambda_2 K_1(z_1, 1)}{\theta (K_1(z_1, 1) - z_1)} dz_1 \right]; \\
P_2(z_1) &= \int_0^{\infty} P_2(z_1, y) dy \\
&= \frac{1}{2 (1 + \lambda_1 + \lambda_2)} \left(1 - B_2(y)\right) \exp \left[ \int_1^{\infty} \frac{\lambda_1 - \lambda_2 K_1(z_1, 1)}{\theta (K_1(z_1, 1) - z_1)} dz_1 \right].
\end{align*}
\]
At present, we can find the generating function of the number of customers in the orbit
\[ Q(z_1) = P_2(z_1) + P_1(z_1) + P_2(z_1) + P_3(z_1) \]
\[ = \frac{1}{2(1 + \lambda_1 + \lambda_2)} \int_0^\infty (1 - B_2(y)) dy \left( \frac{1 - K_1(z_1,1)}{K_1(z_1,1) - z_1} \right) \]
\[ \times \exp \left( \int \frac{\lambda_1 - \lambda_2 K_1(z_1,1)}{\theta(K_1(z_1,1) - z_1)} dz_1 \right). \]

as well as the generating function of the number of priority customers in the system
\[ H(z_1) = P_2(z_1) + z_1 P_1(z_1) + P_2(z_1) + z_1 P_3(z_1) \]
\[ = \frac{K_1(z_1,1)(1 - z_1)}{2(1 + \lambda_1 + \lambda_2) \lambda_2 (K_1(z_1,1) - z_1)} \]
\[ \times \exp \left( \int \frac{\lambda_1 - \lambda_2 K_1(z_1,1)}{\theta(K_1(z_1,1) - z_1)} dz_1 \right). \]

With the help of the obtained generating functions, we can get various performance characteristics of our system, say

- Probability, \( p_1 \), that the server is occupied by a priority customer and there is no interrupted non-priority customer in the service station
\[ p_1 = P_1(1) = \frac{\rho_1}{2\lambda_2 (1 - \rho_1)(1 + \lambda_1 + \lambda_2)}. \]
- Probability, $P_2$, that the server is occupied by a non-priority customer

$$p_2 = P_2(1) = \frac{1}{2(1+\lambda_1 + \lambda_2)} \int_0^\infty (1 - B_2(y)) \, dy.$$  \hspace{1cm} (9)

- Probability, $P_3$, that the server is occupied by a priority customer and there is an interrupted non-priority customer in the service station

$$p_3 = P_3(1) = \frac{\rho_1}{2(1+\lambda_1 + \lambda_2)(1-\rho_1)} \int_0^\infty (1 - B_2(y)) \, dy.$$  \hspace{1cm} (10)

- Mean number of priority customers in the orbit, $n_o$

$$n_o = Q'(1) = \frac{(2\rho_1\lambda_1 + \theta(1-\rho_1)\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2)}{4\theta(1-\rho_1)(1+\lambda_1 + \lambda_2)(1+\beta_{1,1})^2} \times \left[1 + \lambda_2 \int_0^\infty (1 - B_2(y)) \, dy \right].$$  \hspace{1cm} (11)

- Mean number of priority customers in the system, $\bar{n}$

$$\bar{n} = H'(1) = \frac{(2\rho_1\lambda_1 + \theta(1-\rho_1)\beta_{1,2} + 4\rho_1\lambda_1\beta_{1,1} + 2\lambda_1\beta_{1,1}^2)}{4\theta(1-\rho_1)(1+\lambda_1 + \lambda_2)(1+\beta_{1,1})^2} \times \left[1 + \lambda_2 \int_0^\infty (1 - B_2(y)) \, dy \right] + \frac{\rho_1}{2(1-\rho_1)} \frac{1 + \lambda_2}{\lambda_2} \int_0^\infty (1 - B_2(y)) \, dy.$$  \hspace{1cm} (12)

5. NUMERICAL ILLUSTRATIONS

In this section, we present some numerical results in order to illustrate the effect of priority customer arrival rate $\lambda_1$ (because this rate defines the ergodicity condition) and retrial intensity $\theta$ on the obtained performance measures. To this end, consider an $M_1,M_2/G_1,G_2/1$ retrial queue with priority customers and resume priority discipline where the service times follow a two-stage Erlang distribution (E₂). So,
Throughout this section, we suppose that the mean service times of both types of customers are $\beta_{1,1} = \beta_{2,1} = 1$ and that the arrival rate of non priority customers $\lambda_2 = 0.3$.

In the figure below, we present the behaviour of the probabilities $p_1$, $p_2$ and $p_3$ (given by (8)-(10)) with respect to $\lambda_1$. It must be noted that the choice of numerical values of the system parameters is performed in the way to ensure the steady state of our system.

As is expected (intuitively), increasing the arrival rate of priority customers $\lambda_1$ results in a significant increase of the probabilities $p_1$ and $p_3$ (related to the priority customers) and in a low decrease of the probability $p_2$ that the server is occupied by a non-priority customer.

Now, we show how the retrial intensity $\theta$ influences the mean number of priority customers in the orbit $\bar{\pi}_n$ and also in the system $\bar{\pi}$ (given by (11)-(12)).
In Figure 2 (where \( \lambda_1 = \rho_1 = 0.6 \)), we observe that the increase of \( \theta \) gives a sensitive improvement of the measures in question.

**REFERENCES**


