Journal of the Korean Statistical Society (2007), **36***: 2, pp 285–297*

AN $M/G/1$ VACATION QUEUE UNDER THE P_λ^M-SERVICE POLICY†

JIYEON LEE[1]

ABSTRACT

We consider the P_λ^M-service policy for an $M/G/1$ queueing system in which the workload is monitored randomly at discrete points in time. If the level of the workload exceeds a threshold λ when it is monitored, then the service rate is increased from 1 to M instantaneously and is kept as M until the workload reaches zero. By using level-crossing arguments, we obtain explicit expressions for the stationary distribution of the workload in the system.

AMS 2000 subject classifications. Primary 60K25; Secondary 60K20.
Keywords. $M/G/1$ queues, multiple vacations, P_λ^M-service policy, stationary distributions.

1. INTRODUCTION

The P_λ^M-policy was originally introduced by Faddy (1974) as a releasing policy for water in a finite dam with Wiener inputs: it starts to release water at a rate of M per unit time as soon as the level of water reaches a threshold $\lambda > 0$ and keeps the release rate constant until the reservoir is empty. Lee and Ahn (1998) applied this policy to an infinite dam with inputs formed by a compound Poisson process. In the specific case of $M = 1$ and $\lambda = D$, the situation is the same as the D-policy applied to an $M/G/1$ queueing system. Bae *et al.* (2002) modified the P_λ^M-policy and introduced the P_λ^M-service policy for an $M/G/1$ queueing system: a server is initially idle, but when a customer arrives it starts to work with service rate 1, meaning that it is getting through its workload at a rate of 1 per unit time. As soon as the workload exceeds a threshold $\lambda > 0$, the server increases its service rate from 1 to $M > 1$, and continues to serve at rate M until its workload

Received September 2006; accepted January 2007.
†This work was supported by the Korea Research Foundation Grant (KRF-2004-002-C00056).
[1]Department of Statistics, Yeungnam University, Kyungpook 712-749, Korea (e-mail: leejy@yu.ac.kr)

is zero again. Bae *et al.* (2002) obtained the stationary workload distribution for this policy by using the decomposition technique introduced by Lee and Ahn (1998) and the level-crossing arguments of Brill and Posner (1977) and Cohen (1977). Recently Kim *et al.* (2006) have shown that, if costs are assigned to an $M/G/1$ queueing system, there exists an optimal service rate which minimizes the long-run average cost per unit time under the P_λ^M-service policy.

In this paper, we consider a system in which the workload is monitored not continuously but randomly at discrete points in time. The inter-monitoring times, denoted by $V_1, V_2, \ldots$, are assumed to be independent and identically distributed *(i.i.d.)* exponential random variables of rate ξ. If the workload of the system exceeds the level λ when it is monitored, then the server increases its service rate instantaneously from 1 to M, at which it remains until the server becomes idle, otherwise the service rate remains at 1. It is assumed that the arrival of customers follows a Poisson process of rate $\nu > 0$, and that the service times of customers are also *i.i.d.* random variables, with distribution function G and mean m.

Notice that, in the special case of $\xi = \infty$, our model corresponds to that introduced by Bae *et al.* (2002). Randomly monitored queueing systems are close to queueing systems with multiple vacations and have been studied by many researchers (Tagaki, 1991). More recently, Kim *et al.* (2004) presented a simple approach to finding the stationary workload distribution of $M/G/1$ queues with both multiple vacations and D-policy, and Lee and Kim (2007) derived the stationary distribution of $M/G/1$ queues under the P_λ^M-service policy with a single vacation.

We will now obtain an explicit formula for the stationary distribution of the workload by using level-crossing arguments. We will also show that results of Bae *et al.* (2002) follow our analysis.

2. Analysis of the Workload Process

Let $\mathbf{X} = \{X(t), t \geq 0\}$ be the workload process under the service policy described in the previous section. The process $\mathbf{X}$ is regenerated each time that the server starts to work. The length of a cycle C is the interval between two successive regeneration points. To analyze $\mathbf{X}$, we first decompose it into four Makov processes and then apply the level-crossing arguments. Let $\mathbf{X}_1$ be a process obtained from the original process $\mathbf{X}$ by connecting the periods during which the service rate is 1 which start at the beginning of the busy period and end at the time of the first exit from $(0, \lambda]$. Let a second process $\mathbf{X}_2$ be formed by separating out and connecting together the remainder of the periods of service rate 1 and similarly a third process $\mathbf{X}_3$ is made up of the periods of service rate M. Finally $\mathbf{X}_4$ is formed by connecting the idle periods of the original process $\mathbf{X}$, that is, $\mathbf{X}_4 \equiv 0$. Clearly, all these new processes are regenerative Markov processes. We will call each separated segment a cycle of each process and use C_i to denote the length of the cycle in $\mathbf{X}_i$, $i = 1, 2, 3, 4$.

Let F_i be the stationary distribution function of $\mathbf{X}_i$ for $i = 1, 2, 3, 4$ and let F be the stationary distribution function of $\mathbf{X}$. Since $E[C_4] = 1/\nu$ and $F_4(x) = 1$ for all $x \geq 0$, where ν is the arriving rate of customers, by applying the renewal reward theorem (Ross, 1996, p. 133), we can show that, for $x \geq 0$,

$$F(x) = \alpha \frac{E[C_1]}{E[C]} F_1(x) + \beta \frac{E[C_2]}{E[C]} F_2(x) + \gamma \frac{E[C_3]}{E[C]} F_3(x) + \frac{1/\nu}{E[C]},$$

where α, β and γ are the respective probabilities that the processes $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ exist in the cycle of $\mathbf{X}$. Note that $E[C] = \alpha E[C_1] + \beta E[C_2] + \gamma E[C_3] + 1/\nu$. We can immediately see that

$$\alpha = G(\lambda),$$

since there exists a process $\mathbf{X}_1$ in the cycle of $\mathbf{X}$ if and only if the workload brought by the first customer after the idle period is less than or equals to λ. We also observe that the probability β is the same as the probability that the workload process $\mathbf{X}$ crosses over the level λ during the cycle C. Let $\mathbf{W} = \{W(t), t \geq 0\}$ be the workload (virtual waiting time) process of an ordinary $M/G/1$ queue with customers arriving at a rate ν and with a distribution function of service times G of which the mean is m. Several results about the process $\mathbf{W}$ are summarized in the Appendix. Because the process $\mathbf{X}$ coincides with $\mathbf{W}$ until $\mathbf{X}$ upcrosses λ it follows that the probability β can be expressed as

$$\beta = 1 - \Pr\{D_\lambda = 0\},$$

where D_λ denotes the number of downcrossings of λ that occur during a cycle of $\mathbf{W}$. Substituting the distribution of D_λ from (2.6) in Cohen (1978) gives

$$\beta = \frac{H'_\rho(\lambda)}{\nu H_\rho(\lambda)},$$

for which the definition of H_ρ appears in (A.1). The probability γ will be calculated in (2.4) of Section 2.2.

In the following three subsections, we will successively evaluate the stationary distributions F_1, F_2 and F_3 and the corresponding expected values $E[C_1]$, $E[C_2]$ and $E[C_3]$.

2.1. Stationary distribution of $\mathbf{X}_1$

Let f_1 be the probability density function of the stationary distribution of $\mathbf{X}_1$ and let $D_x^{(1)}$ denote the number of downcrossings of x during a cycle of $\mathbf{X}_1$. From the level-crossing arguments in Brill and Posner (1977) and in Cohen (1977), it follows that

$$f_1(x) = \frac{E[D_x^{(1)}]}{E[C_1]}, \quad 0 < x \le \lambda.$$

Observe that the process $\mathbf{X}_1$ coincides with $\mathbf{W}$ until the point in the cycle at which the process $\mathbf{W}$ either crosses over λ or reaches 0, provided that both $\mathbf{X}_1$ and $\mathbf{W}$ start at the same level. Therefore, if we use $D_{0\lambda;yx}$ to denote the number of downcrossings of x until the process $\mathbf{W}$, starting at y, $0 < y \le \lambda$, either crosses over λ or reaches 0, then $E[D_x^{(1)}]$ can be calculated in terms of the starting level of the cycle C_1, as follows:

$$E[D_x^{(1)}] = \int_0^\lambda E[D_{0\lambda;yx}]\frac{dG(y)}{G(\lambda)}, \quad 0 < x \le \lambda.$$

Substituting the expression for $E[D_{0\lambda;yx}]$ from (A.2), Bae *et al.* (2002) obtained

$$E[D_x^{(1)}] = \frac{H_\rho(x)}{\nu G(\lambda)}\left(\frac{H'_\rho(x)}{H_\rho(x)} - \frac{H'_\rho(\lambda)}{H_\rho(\lambda)}\right), \quad 0 < x \le \lambda.$$

Conditioned on the starting level of the cycle C_1, the expected value of $E[C_1]$ can be expressed as follows:

$$E[C_1] = \int_0^\lambda E[C_1(y)]\frac{dG(y)}{G(\lambda)},$$

where $C_1(y)$ denotes the length of a cycle of $\mathbf{X}_1$, starting at level y. For an ordinary $M/G/1$ workload process $\mathbf{W}$, we define

$$T_{0\lambda;y} \equiv \inf\{t \ge 0 | W(t) \notin (0, \lambda], W(0) = y\}$$

to represent the first exit time from $(0, \lambda]$ when the process starts from $W(0) = y$.

It follows from (A.2) that

$$\begin{aligned} E[T_{0\lambda;y}] &= \int_0^\lambda E[D_{0\lambda;yx}]dx \\ &= \frac{H_\rho(\lambda - y)}{H_\rho(\lambda)} \int_0^\lambda H_\rho(x)dx - \int_0^{\lambda - y} H_\rho(x)dx, \quad 0 < y \le \lambda, \end{aligned} \tag{2.1}$$

which implies that

$$E[C_1] = \frac{1}{\nu G(\lambda)} \left(H_\rho(\lambda) - 1 - \frac{H'_\rho(\lambda)}{H_\rho(\lambda)} \int_0^\lambda H_\rho(z)dz \right),$$

because $C_1(y) \stackrel{\mathcal{D}}{=} T_{0\lambda;y}$, where $\stackrel{\mathcal{D}}{=}$ denotes the equality of distribution.

We can now express the stationary density $f_1(x)$ of the process $\mathbf{X}_1$ as

$$f_1(x) = \frac{H_\rho(\lambda)H'_\rho(x) - H'_\rho(\lambda)H_\rho(x)}{H_\rho(\lambda)(H_\rho(\lambda) - 1) - H'_\rho(\lambda) \displaystyle\int_0^\lambda H_\rho(y)dy}, \quad 0 < x \le \lambda,$$

which agrees with the result of Bae *et al.* (2002).

2.2. Stationary distribution of $\mathbf{X}_2$

We will use Y_2 to denote a starting level above λ in a cycle of the process $\mathbf{X}_2$. Using the Markov property of $\mathbf{X}_1$, Bae *et al.* (2002) obtained the distribution function $Q_2(y)$ of Y_2 in the form

$$Q_2(y) = 1 - \frac{1 - G(y) + \displaystyle\int_0^\lambda P(y - \lambda, z)dG(z)}{H'_\rho(\lambda)/\nu H_\rho(\lambda)}, \quad y > \lambda,$$

where

$$\begin{aligned} P(w, z) &\equiv \Pr\{W(T_{0\lambda;z}) > \lambda + w\} \\ &= c(w)H_\rho(\lambda - z) - \rho \int_{0-}^{\lambda - z} J_w(\lambda - z - u)dH_\rho(u), \quad w \ge 0, \quad 0 < z \le \lambda, \end{aligned}$$

where $c(w) = \rho(H_\rho * J_w)(\lambda)/H_\rho(\lambda)$, $J_w(x) = G_e(x + w) - G_e(w)$ and $G_e(x) = (1/m)\int_0^x (1 - G(u))du$, the equilibrium distribution function of G. We note that the starting levels of each cycle in the process $\mathbf{X}_2$ are independent and have the same distribution as the random variable Y_2.

Conditioning on the starting level, we now have

$$E[C_2] = \int_{\lambda}^{\infty} E[C_2(y)]dQ_2(y), \tag{2.2}$$

where $C_2(y)$ denotes the length of a cycle of $\mathbf{X}_2$ when the starting level is $y > \lambda$. For $0 \le x < y$, let $T_{x;y}$ be the time that an ordinary $M/G/1$ workload process $\mathbf{W}$ takes to reach a level x, starting from y, and let V be a generic random variable denoting the interval between successive times at which the system is monitored. Because of the memoryless property of the exponential random variable, $C_2(y)$ can now be described as follows:

$$C_2(y) \stackrel{\mathcal{D}}{=} \begin{cases} V & \text{if} \quad V \le T_{\lambda;y}, \\ T_{\lambda;y} + C_2' & \text{if} \quad V > T_{\lambda;y}, \end{cases}$$

where C_2' denotes the length of the period $C_2(y)$ that remains after the process $\mathbf{X}_2$ downcrosses the level λ. We note that C_2' is independent of the starting level y because $\mathbf{X}_2$ has already downcrossed λ. Again invoking the memoryless property of the exponential random variable V, we can express the expected value of $C_2(y)$ in the following terms:

$$\begin{aligned} E[C_2(y)] &= E[\min(V, T_{\lambda;y})] + E[C_2' \mid V > T_{\lambda;y}]\Pr\{V > T_{\lambda;y}\} \\ &= \frac{1}{\xi} + \tilde{T}(y-\lambda, \xi)\left(E[C_2'] - \frac{1}{\xi}\right), \end{aligned} \tag{2.3}$$

where the second equality follows from the facts that $E[\min(V, T_{\lambda;y})] = \{1 - \tilde{T}(y-\lambda,\xi)\}/\xi$ and $\Pr\{V > T_{\lambda;y}\} = \tilde{T}(y-\lambda,\xi)$, in which $\tilde{T}(y,\xi) \equiv E[e^{-\xi T_{0;y}}]$ is the Laplace-Stieltjes transform of $T_{0;y}$, and is derived in (A.4).

Let Y be a random variable which represents the sum of λ and the amount by which the level exceeds λ when the process $\mathbf{W}$, having started at λ, crosses over λ without returning to 0. Then the distribution function of Y is

$$\begin{aligned} \Pr\{Y \le y\} &= \Pr\{W(T_{0\lambda;\lambda}) \le y | W(T_{0\lambda;\lambda}) > \lambda\} \\ &= 1 - \frac{P(y-\lambda, \lambda)}{P(0,\lambda)}, \quad y > \lambda. \end{aligned}$$

Note that $P(0,\lambda) = 1 - 1/H_\rho(\lambda)$.

The Markov property of the process $\mathbf{X}_2$ also allows us to observe that

$$C_2' \stackrel{\mathcal{D}}{=} T_{0\lambda;\lambda} + 1_{\{W(T_{0\lambda;\lambda})>\lambda\}}\left(\min(V, T_{\lambda;Y}) + 1_{\{T_{\lambda;Y}<V\}}C_2'\right).$$

We can determine the expected value of C_2' from this equation by substituting $E[T_{0\lambda;\lambda}] = \int_0^\lambda H_\rho(x)dx/H_\rho(\lambda)$ from (2.1), and it then follows that

$$E[C_2'] = \frac{1}{\xi} + \frac{\displaystyle\int_0^\lambda H_\rho(x)dx - \frac{1}{\xi}}{H_\rho(\lambda)\left(1 + \displaystyle\int_\lambda^\infty \tilde{T}(y-\lambda,\xi)d_yP(y-\lambda,\lambda)\right)}.$$

Referring back to (2.2) and (2.3), we are now at last able to formulate the expectation

$$E[C_2] = (1-\gamma)\int_0^\lambda H_\rho(x)dx + \frac{\gamma}{\xi},$$

where γ is the probability that the process $\mathbf{X}_3$ exists in the cycle of $\mathbf{X}$, which is given by

$$\gamma = 1 - \frac{\displaystyle\int_\lambda^\infty \tilde{T}(y-\lambda,\xi)dQ_2(y)}{H_\rho(\lambda)\left(1 + \displaystyle\int_\lambda^\infty \tilde{T}(y-\lambda,\xi)d_yP(y-\lambda,\lambda)\right)}. \tag{2.4}$$

To apply the level-crossing arguments, we now need to know the expected number of downcrossings of x for the process $\mathbf{X}_2$ during its cycle, denoted by $E[D_x^{(2)}]$. Conditioning on the starting level y, we now have

$$E[D_x^{(2)}] = \int_\lambda^\infty E[D_{yx}^{(2)}]dQ_2(y), \tag{2.5}$$

where $D_{yx}^{(2)}$ is the number of downcrossings of x that occur in the process $\mathbf{X}_2$ during a cycle starting at $y > \lambda$. If $D_{yx}(t)$ is the number of times that an ordinary $M/G/1$ workload process $\mathbf{W}$ downcrosses x during an interval of length t, and $D_{\lambda;yx}$ is the number of downcrossings of x that occur before $\mathbf{W}$ hits λ, having started at $y > \lambda$, then

$$D_{yx}^{(2)} \stackrel{\mathcal{D}}{=} \begin{cases} D_{yx}(V) & \text{if } V \le T_{\lambda;y}, \\ D_{\lambda;yx} + D_{\lambda x}^{(2)} & \text{if } V > T_{\lambda;y} \end{cases} \tag{2.6}$$

and

$$D_{\lambda x}^{(2)} \stackrel{\mathcal{D}}{=} D_{0\lambda;\lambda x} + D_{Yx}^{(2)} 1_{\{W(T_{0\lambda;\lambda})>\lambda\}}. \tag{2.7}$$

We also observe that

$$\begin{aligned} E[D_{\lambda;yx}] &= E[D_{\lambda;yx}1_{\{V>T_{\lambda;y}\}}] + E[D_{\lambda;yx}1_{\{V\le T_{\lambda;y}\}}] \\ &= E[D_{\lambda;yx}1_{\{V>T_{\lambda;y}\}}] + E[D_{yx}(V)1_{\{V\le T_{\lambda;y}\}}] \\ &\quad + E[D_{\lambda;W_{y-\lambda}+\lambda x}]\Pr\{V \le T_{\lambda;y}\}, \end{aligned} \tag{2.8}$$

where

$$W_y \equiv W(V)|W(0) = y,\ V \le T_{0;y}$$

represents the workload of $\mathbf{W}$ after the exponential time V has elapsed, given that $\mathbf{W}$ starts at y and does not reach zero before the exponential time. From Lee and Kim (2007), W_y has the Laplace-Stieltjes transform

$$E[e^{-\theta W_y}] = \frac{\xi\left(\theta_0(\xi)e^{-\theta y} - \theta e^{-\theta_0(\xi)y} - \tilde{T}_y(\xi)(\theta_0(\xi) - \theta)\right)}{\theta_0(\xi)(\xi - \varphi(\theta))(1 - \tilde{T}_y(\xi))}, \quad \theta \ge 0,$$

where

$$\varphi(\theta) = \theta - \nu + \nu\tilde{G}(\theta). \tag{2.9}$$

Here $\tilde{G}(\theta) = \int_0^\infty e^{-\theta x}dG(x)$ is the Laplace-Stieltjes transform of G and $\theta_0(\xi)$ is the solution to the equation

$$\varphi(\theta) = \xi. \tag{2.10}$$

Taking the expectations in (2.6) and using the relation (2.8), we have

$$E[D^{(2)}_{yx}] = E[D_{\lambda;yx}] - (1 - \tilde{T}(y-\lambda,\xi))E[D_{\lambda;W_{y-\lambda}+\lambda x}] + \tilde{T}(y-\lambda,\xi)E[D^{(2)}_{\lambda x}].$$

And substituting the expectations of (2.7) into the above equation, we also have

$$\begin{aligned} E[D^{(2)}_{\lambda x}] = E[D_{0\lambda;\lambda x}] - \int_\lambda^\infty &\Big(E[D_{\lambda;yx}] - (1 - \tilde{T}(y-\lambda,\xi))E[D_{\lambda;W_{y-\lambda}+\lambda x}] \\ &+\tilde{T}(y-\lambda,\xi)E[D^{(2)}_{\lambda x}]\Big)\, d_yP(y-\lambda,\lambda). \end{aligned}$$

Solving for $E[D^{(2)}_{\lambda x}]$ gives

$$E[D^{(2)}_{\lambda x}] = \frac{E[D_{0\lambda;\lambda x}] - \int_\lambda^\infty \Big(E[D_{\lambda;yx}] - (1 - \tilde{T}(y-\lambda,\xi))E[D_{\lambda;W_{y-\lambda}+\lambda x}]\Big)\, d_yP(y-\lambda,\lambda)}{1 + \int_\lambda^\infty \tilde{T}(y-\lambda,\xi)d_yP(y-\lambda,\lambda)},$$

which finally allows us to evaluate (2.5).

2.3. *Stationary distribution of* $\mathbf{X}_3$

Let Y_3 be the starting level in a cycle of $\mathbf{X}_3$, where Y_3 depends on the starting level Y_2 in the preceding cycle C_2 of the process $\mathbf{X}_2$. We will write $Y_3(y)$ for the starting level of $\mathbf{X}_3$ if the starting level of C_2 is y. If $X_2(0) = y \geq \lambda$, it follows that

$$Y_3(y) \stackrel{\mathcal{D}}{=} \begin{cases} W(V) & \text{if} \quad V \leq T_{\lambda;y}, \\ Y_3(\lambda) & \text{if} \quad V > T_{\lambda;y}. \end{cases}$$

For the specific case of $y = \lambda$, we have

$$Y_3(\lambda) \stackrel{\mathcal{D}}{=} \begin{cases} Y_3 & \text{if} \quad W(T_{0\lambda;\lambda}) = 0, \\ Y_3(Y) & \text{if} \quad W(T_{0\lambda;\lambda}) > \lambda. \end{cases}$$

Hence, the Laplace-Stieltjes transform of $Y_3(y)$ is given by

$$\begin{aligned} &E[e^{-\theta Y_3(y)}] \\ &= E[e^{-\theta W(V)} 1_{\{V \leq T_{\lambda;y}\}} | W(0) = y] + \Pr\{V > T_{\lambda;y}\} E[e^{-\theta Y_3(y)} | V > T_{\lambda;y}] \\ &= E[e^{-\theta W(V)} | W(0) = y] - \Pr\{V > T_{\lambda;y}\} E[e^{-\theta W(V)} | W(0) = y, V > T_{\lambda;y}] \\ &\quad + \Pr\{V > T_{\lambda;y}\} E[e^{-\theta Y_3(\lambda)}] \\ &= E[e^{-\theta W(V)} | W(0) = y] - \tilde{T}(y - \lambda, \xi) E[e^{-\theta W(V)} | W(0) = \lambda] \\ &\quad + \tilde{T}(y - \lambda, \xi) E[e^{-\theta Y_3(\lambda)}], \quad y > \lambda, \end{aligned} \tag{2.11}$$

in which the last equality follows from the memoryless property of the exponential random variables. Using the results of Boxma *et al.* (2001) we find that for $y \geq 0$,

$$\psi(\theta, y) \equiv E[e^{-\theta W(V)} \mid W(0) = y] = \frac{\xi}{\xi - \varphi(\theta)} \left(e^{-\theta y} - \frac{\theta e^{-\theta_0(\xi) y}}{\theta_0(\xi)} \right), \quad \theta \geq 0,$$

where $\varphi(\theta)$ and $\theta_0(\xi)$ were defined in (2.9) and (2.10) respectively. Therefore $E[e^{-\theta Y_3(y)}]$ can be rewritten as

$$E[e^{-\theta Y_3(y)}] = \psi(\theta, y) - \tilde{T}(y - \lambda, \xi) \psi(\theta, \lambda) + \tilde{T}(y - \lambda, \xi) E[e^{-\theta Y_3(\lambda)}]. \tag{2.12}$$

Using a similar method, we can also calculate $E[e^{-\theta Y_3(\lambda)}]$ as follows:

$$\begin{aligned} E[e^{-\theta Y_3(\lambda)}] &= \Pr\{W(T_{0\lambda;\lambda}) = 0\} E[e^{-\theta Y_3} | W(T_{0\lambda;\lambda}) = 0] \\ &\quad + \Pr\{W(T_{0\lambda;\lambda}) > \lambda\} E[e^{-\theta Y_3(Y)} | W(T_{0\lambda;\lambda}) > \lambda] \\ &= \frac{1}{H_\rho(\lambda)} E[e^{-\theta Y_3}] + \left(1 - \frac{1}{H_\rho(\lambda)}\right) E[e^{-\theta Y_3(Y)} | W(T_{0\lambda;\lambda}) > \lambda] \end{aligned}$$

$$= \frac{1}{H_\rho(\lambda)} E[e^{-\theta Y_3}] - \int_\lambda^\infty E[e^{-\theta Y_3(y)} | X_2(0) = y] d_y P(y - \lambda, \lambda)$$
$$= \frac{1}{H_\rho(\lambda)} E[e^{-\theta Y_3}] - \int_\lambda^\infty \Big\{ \psi(\theta, y) - \tilde{T}(y - \lambda, \xi)\psi(\theta, \lambda)$$
$$+ \tilde{T}(y - \lambda, \xi) E[e^{-\theta Y_3(\lambda)}] \Big\} d_y P(y - \lambda, \lambda).$$

Solving the above equation yields

$$E[e^{-\theta Y_3(\lambda)}] = \frac{\frac{1}{H_\rho(\lambda)} E[e^{-\theta Y_3}] - \int_\lambda^\infty \left[\psi(\theta, y) - \tilde{T}(y - \lambda)\psi(\theta, \lambda) \right] d_y P(y - \lambda, \lambda)}{1 + \int_\lambda^\infty \tilde{T}(y - \lambda, \xi) d_y P(y - \lambda, \lambda)}.$$

Thus, conditioning on the initial level $X_2(0) = y$, and substituting the above formula into (2.12), we finally obtain

$$E[e^{-\theta Y_3}]$$
$$= \int_\lambda^\infty E[e^{-\theta Y_3(y)}] dQ_2(y)$$
$$= \frac{1}{\gamma} \left(\int_\lambda^\infty \psi(\theta, y) dQ_2(y) - (1 - \gamma) H_\rho(\lambda) \left[\psi(\theta, \lambda) + \int_\lambda^\infty \psi(\theta, y) d_y P(y - \lambda, \lambda) \right] \right),$$

from which we can determine the expected value of Y_3 and the distribution function $Q_3(y)$ of Y_3.

Notice that if we change the scale of time by making $1/M$ the unit of time, then the arrival rate of the process $\mathbf{X}_3$ during the period C_3 becomes ν/M, the service speed becomes 1, and the traffic intensity $\rho' \equiv \nu m/M$. Since $\rho' < 1$, the well-known fact in Wolff (1989, p. 393) about the expected busy period of $M/G/1$ queues with exceptional first service yields

$$E[C_3] = \frac{1}{M} \frac{E[Y_3]}{1 - \rho'} = \frac{E[Y_3]}{M - \rho},$$

in which the time scale is restored by multiplying by $1/M$.

Let $D'_{0;yx}$ denote the number of downcrossings of x that occur before the workload process $\mathbf{W}'$ with a traffic intensity of ρ', reaches 0, given that the workload starts at y and $D_x^{(3)}$ is the number of downcrossings of x that occur during the cycle of process $\mathbf{X}_3$. Conditioning on the starting level y, we have

$$E[D_x^{(3)}] = \int_\lambda^\infty E[D'_{0;yx}] dQ_3(y)$$
$$= H_{\rho'}(x)$$

for $0 < x \le \lambda$ and for $x > \lambda$ we have

$$\begin{aligned}E[D_x^{(3)}] &= \int_\lambda^x E[D'_{0;yx}]dQ_3(y) + \int_x^\infty E[D'_{0;yx}]dQ_3(y)\\ &= \int_\lambda^x \left(H_{\rho'}(x) - H_{\rho'}(x-y)\right) dQ_3(y) + \int_x^\infty H_{\rho'}(x)dQ_3(y)\\ &= H_{\rho'}(x) - H_{\rho'}Q_3(x).\end{aligned}$$

Let $f_3(x)$ be the probability density function of the stationary distribution of the process $\mathbf{X}_3$. Deploying the level-crossing arguments again, we can obtain

$$f_3(x) = \frac{E[D_x^{(3)}]}{M \cdot E[C_3]}, \quad 0 < x < \infty.$$

REMARK 2.1. We will now check our result for the special case of $\xi = \infty$, which was treated in Bae *et al.* (2002). It follows from (2.11) that

$$\begin{aligned}\lim_{\xi\to\infty} E[e^{-\theta Y_3(y)}] &= \lim_{\xi\to\infty} \psi(\theta, y)\\ &= e^{-\theta y},\end{aligned}$$

because

$$\begin{aligned}\lim_{\xi\to\infty} \frac{\theta_0(\xi)}{\xi} &= \lim_{\xi\to\infty} \frac{\nu - \nu\tilde{G}(\theta_0(\xi)) + \xi}{\xi}\\ &= 1.\end{aligned}$$

Since $\gamma \to 1$ we can deduce that

$$E[e^{-\theta Y_3}] = E[e^{-\theta Y_2}],$$

when $\xi = \infty$, which means that there is no period of process $\mathbf{X}_2$ in this case. Hence we can conclude that the stationary distributions are the same as those in Bae *et al.* (2002) when $\xi = \infty$.

APPENDIX : THE WORKLOAD PROCESS OF THE $M/G/1$ QUEUE

It is well-known in Cohen (1982, p. 255) that, under the assumption that $\rho \equiv \nu m < 1$, an ordinary $M/G/1$ workload process $\mathbf{W} = \{W(t), t \ge 0\}$ has a unique stationary distribution V given by

$$V(x) = (1-\rho)H_\rho(x)$$

with

$$H_\rho(x) = \sum_{n=0}^{\infty} \rho^n G_e^{*n}(x), \tag{A.1}$$

where G_e^{*n} is the n-fold recursive Stieltjes convolution of G_e with the Heaviside function G_e^{*0}.

Let us define $D_{0\lambda;yx}$ as the number of downcrossings of level x that occur before the process $\mathbf{W}$, having started from y, crosses over λ or reaches 0. Bae *et al.* (2002) showed that

$$E[D_{0\lambda;yx}] = \begin{cases} \dfrac{H_\rho(x)H_\rho(\lambda - y)}{H_\rho(\lambda)} & \text{if} \quad 0 < x < y \leq \lambda, \\ \dfrac{H_\rho(x)H_\rho(\lambda - y)}{H_\rho(\lambda)} - H_\rho(x-y) & \text{if} \quad 0 < y \leq x \leq \lambda. \end{cases} \tag{A.2}$$

If we let $D_{0;yx} \equiv \lim_{\lambda\to\infty} D_{0\lambda;yx}$, then it expresses the number of downcrossings of x that occur before the process $\mathbf{W}$, having started from y, returns to 0. Since $\lim_{x\to\infty} H_\rho(x) = 1/(1-\rho)$, from the monotone convergence theorem it follows that

$$\begin{aligned} E[D_{0;yx}] &= \lim_{\lambda\to\infty} E[D_{0\lambda;yx}] \\ &= \begin{cases} H_\rho(x) & \text{if} \quad 0 < x < y, \\ H_\rho(x) - H_\rho(x-y) & \text{if} \quad 0 < y \leq x, \end{cases} \end{aligned}$$

which coincides with the result in Bae *et al.* (2002).

Now we need the distribution of $T_{0;y}$, which is the time for the process $\mathbf{W}$ to reach 0 when it starts from y. From the Markovian property of $\mathbf{W}$, we can see that

$$T_{0;y} \stackrel{\mathcal{D}}{=} \begin{cases} y & \text{if} \quad N(y) = 0, \\ y + \sum_{i=1}^{N(y)} B_i & \text{if} \quad N(y) \geq 1, \end{cases} \tag{A.3}$$

where $N(y)$ is the number of customers who arrive during the time y, which is the Poisson random variable with parameter νy, and where B_i denotes the busy period of the $M/G/1$ queue. It is well known (Wolff, 1989, p. 390) that the Laplace-Stieltjes transform of B_i, denoted by $\tilde{B}(\theta)$, is the solution to the following equation:

$$\tilde{B}(\theta) = \tilde{G}(\theta + \nu - \nu\tilde{B}(\theta)).$$

Using Wald's equation (Ross, 1996, p. 105), it follows from (A.3) that the Laplace-Stieltjes transform of $T_{0;y}$ is given by

$$\tilde{T}_{0;y}(\theta) = \exp\left\{-(\theta + \nu - \nu\tilde{B}(\theta))y\right\} \quad (cf.\ \text{Wolff, 1989}). \tag{A.4}$$

REFERENCES

BAE, J., KIM, S. AND LEE, E. Y. (2002). "A P_λ^M-policy for an $M/G/1$ queueing system", *Applied Mathematical Modelling*, **26**, 929–939.

BOXMA, O. J., PERRY, D. AND STADJE, W. (2001). "Clearing models for $M/G/1$ queues", *Queueing Systems*, **38**, 287–306.

BRILL, P. H. AND POSNER, M. J. M. (1977). "Level crossings in point processes applied to queues: single-server case", *Operations Research*, **25**, 662–674.

COHEN, J. W. (1977). "On up- and downcrossings", *Journal of Applied Probability*, **14**, 405–410.

COHEN, J. W. (1978). "Properties of the process of level crossings during a busy cycle of the M/G/1 queueing system", *Mathematics of Operations Research*, **3**, 133–144.

COHEN, J. W. (1982). *The Single Server Queue*, revised ed., North-Holland, Amsterdam.

FADDY, M. J. (1974). "Optimal control of finite dams: discrete (2-stage) output procedure", *Journal of Applied Probability*, **11**, 111–121.

KIM, J., BAE, J. AND LEE, E. Y. (2006). "An optimal P_λ^M-service policy for an $M/G/1$ queueing system", *Applied Mathematical Modelling*, **30**, 38–48.

KIM, N. K., PARK, Y. I. AND CHAE, K. C. (2004). "A simple approach to the workload analysis of $M/G/1$ vacation queues", *Journal of the Korean Statistical Society*, **33**, 159–167.

LEE, E. Y. AND AHN, S. K. (1998). "P_λ^M-policy for a dam with input formed by a compound Poisson process", *Journal of Applied Probability*, **35**, 482–488.

LEE, J. AND KIM, J. (2007). "Workload analysis of an $M/G/1$ queue under the P_λ^M policy with a set-up time", *Applied Mathematical Modelling*, **31**, 236–244.

ROSS, S. M. (1996). *Stochastic Processes*, 2nd ed., John Wiley & Sons, New York.

TAKAGI, H. (1991). *Queueing Analysis, Volume 1: Vacation and Priority Systems*, North-Holland, Amsterdam.

WOLFF, R. W. (1989). *Stochastic Modeling and the Theory of Queues*, Prentice Hall, New Jersey.