Journal of the Korean Statistical Society (1999), 28: 1, pp 125–133

A Queueing System with Work-Modulated Arrival and Service Rates†

Jiyeon Lee[1]

ABSTRACT

We consider a FIFO single-server queueing model in which both the arrival and service processes are modulated by the amount of work in the system. The arrival process is a non-homogeneous Poisson process(NHPP) modulated by work, that is, with an intensity that depends on the work in the system. Each customer brings a job consisting of an exponentially distributed amount of work to be processed. The server processes the work at various service rates which also depend on the work in the system. Under the stability conditions obtained by Browne and Sigman(1992) we derive the exact stationary distribution of the work $W(t)$ and the first exit probability that the work level b is exceeded before the work level a is reached, starting from $x \in (a, b]$.

Keywords: Work-modulated Queue; Non-homogeneous Poisson Process; Stationary Distribution; First Exit Probability

1. INTRODUCTION

We consider a single-server queue with an unlimited waiting space and the first-in first-out discipline. The nth customer arrives at the system at time $t_n (t_0 \equiv 0)$ and brings a job consisting of an amount of work to be processed, $S_n \geq 0$, that is exponentially distributed random variable with mean $1/\nu$. The total work in system, $W(t)$, is defined as the sum of all the unprocessed work in system at time t. It is assumed that the arrival process $\{t_n, n \geq 0\}$ forms a non-homogeneous Poisson process(NHPP) with the intensity $\lambda(W(t))$ where λ is a non-increasing positive real-valued function on $[0, \infty)$ and $\lambda(0) > 0$ to avoid triviality. It is further assumed that the service rate at which the server processes the work per unit time at time t is given by $\mu(W(t))$ for a non-decreasing function μ on $[0, \infty)$

†The author wishes to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.

[1]Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea.

with $\mu(0) = 0$ meaning that the server stops processing whenever the workload is empty. Then the sample paths of the workload process $\{W(t), t \geq 0\}$ satisfy the following storage equation;

$$W(t) = W(0) + A(t) - \int_0^t \mu(W(s))ds,$$

where $A(t)$ denotes the total amount of work that arrived during $(0, t]$. In our model, $A(t) = \sum_{n=1}^{N(t)} S_n$, where $N(t)$ is the number of arrivals during $(0, t]$, the counting process for $\{t_n, n \geq 1\}$. Thus, at any time t, the amount of work in system is precisely the initial amount, plus how much has arrived during $(0, t]$, minus how much has been serviced during $[0, t]$. In Brockwell et al.(1982), $A(t)$ is assumed a pure positive jump Levy process whereas in Harrison et al.(1976) and Asmussen(1987), $A(t)$ assumes a compound Poisson process. In these classic storage models the input $A(t)$ is assumed to be independent of the amount of work(or storage) in system. In our model, however, we allow $A(t)$ to be a work-modulated nonhomogeneous compound Poisson process i.e. we allow the intensity of arrivals to be modulated by the work level.

Browne and Sigman(1992) established the sufficient conditions for the existence of the stationary distribution of the work $W(t)$. Under these stability conditions, in section 2 we obtain the stationary distribution function of $W(t)$ using the integro-differential equation for $W(t)$. In section 3 we compute the first exit probability that the work level b is exceeded before the work level a is reached, starting from $x \in (a, b]$.

A few researchers studied the state-dependent queueing model in which both the arrival process and the service process depend on the unfinished work or the virtual waiting time instead of the number of customers in the system. Knessl et al.(1987) considered the queue with NHPP arrivals and the exponential service times that depend on the work. It was concerned with asymptotic approximations for the busy period distribution. Whitt(1990) derived stability conditions and then approximations for the steady-state waiting time distribution for the queueing model where the service times and the interarrival times depend linearly and randomly on the waiting time. Perry and Asmussen(1995) considered a M/G/1 queue modified such that an arriving customer may be totally or partially rejected depending on the virtual waiting time, which can be the special case of our model.

2. THE STATIONARY DISTRIBUTION OF $W(t)$

We assume that the function μ satisfies

$$\lim_{x\to\infty}\mu(x)=\infty \quad \text{and} \quad \theta(x)=\int_0^x \frac{1}{\mu(y)}dy<\infty, \quad x\geq 0, \tag{2.1}$$

where $\theta(x)$ means the amount of time required to bring the work level x down to 0 in the absence of any new arrivals.(Asmussen(1987)) Under these conditions Browne and Sigman(1992) showed that the workload process $\{W(t), t\geq 0\}$ is positive recurrent with consecutive visits to the origin serving as regeneration points so that $\{W(t), t\geq 0\}$ has a unique stationary distribution.

Theorem 2.1. *Under the above assumptions (2.1), the stationary distribution of the workload process $\{W(t), t\geq 0\}$ is given by a probability p_0 at 0 and a density $f(x)$ on $(0,\infty)$ such that*

$$\begin{aligned} p_0 &= [1+\int_0^\infty \frac{\lambda(0)}{\mu(y)}\exp\{\omega(y)-\nu y\}dy]^{-1} \quad \textit{and} \\ f(x) &= p_0\frac{\lambda(0)}{\mu(x)}\exp\{\omega(x)-\nu x\}, \end{aligned} \tag{2.2}$$

where $\omega(x)=\int_0^x \frac{\lambda(y)}{\mu(y)}dy$.

Proof: Let $F(x,t)=P\{W(t)\leq x\}$ denote the distribution function of $W(t)$. Then notice that $F(x,t)$ consists of a discrete probability $F(0,t)$ and a density $f(x,t)$ for $x>0$. To derive the integro-differential equation for $F(x,t)$ we define

$$\phi(x,t)=\theta^{-1}(\theta(x)+t), \quad \text{for } t\geq 0, x\geq 0$$

for the strictly increasing function $\theta(x)$. Then $\phi(x,t)$ is the work level from which the workload process reaches at the level x after time t in the absence of arrivals. Observe that for $x\geq 0$,

$$\begin{aligned} &F(x,t+\delta t)= \\ &\quad P\{W(t)=0, W(t+\delta t)\leq x\}+P\{0<W(t)\leq \phi(x,\delta t), W(t+\delta t)\leq x\}. \end{aligned} \tag{2.3}$$

Conditioning on whether the customer arrives or not during the time interval $(t,t+\delta t)$, we obtain that

$$\begin{aligned} P\{W(t)=0, W(t+\delta t)\leq x\} &= F(0,t)-F(0,t)\lambda(0)\exp\{-\nu\phi(x,\delta t)\}\delta t \\ &\quad +o(\delta t) \end{aligned} \tag{2.4}$$

and

$$P\{0 < W(t) \le \phi(x,\delta t), W(t+\delta t) \le x\}$$
$$= \int_0^{\phi(x,\delta t)} f(y,t)dy - \int_0^{\phi(x,\delta t)} f(y,t)\lambda(y)\exp\{-\nu(\phi(x,\delta t)-y)\}dy\delta t$$
$$+o(\delta t). \tag{2.5}$$

From equation (2.3), (2.4) and (2.5) we have that

$$F(x,t+\delta t) = F(\phi(x,\delta t),t) - F(0,t)\lambda(0)\exp\{-\nu\phi(x,\delta t)\}\delta t$$
$$- \int_0^{\phi(x,\delta t)} f(y,t)\lambda(y)\exp\{-\nu(\phi(x,\delta t)-y)\}dy\delta t + o(\delta t).$$

Performing a Taylor series expansion on $F(\phi(x,\delta t),t)$, rearranging the above equation and letting $\delta t \to 0$, from the fact that $\lim_{\delta t\to 0}\phi(x,\delta t) = x$ and $\lim_{\delta t\to 0}\{\phi(x,\delta t)-x\}/\delta t = \mu(x)$ we can derive the following integro-differential equation

$$\frac{\partial F(x,t)}{\partial t} = \mu(x)f(x,t) - F(0,t)\lambda(0)\exp\{-\nu x\}$$
$$- \int_0^x f(y,t)\lambda(y)\exp\{-\nu(x-y)\}dy.$$

Notice that the stationary distribution is the same as the limiting distribution $\lim_{t\to\infty}F(x,t)$, since the workload process $\{W(t), t \ge 0\}$ is positive recurrent (Harrison and Resnick(1976)). Putting $\frac{\partial F(x,t)}{\partial t} = 0$ and letting $\lim_{t\to\infty}f(x,t) = f(x)$ and $\lim_{t\to\infty}F(0,t) = p_0$ the following integral equation for $f(x)$ is obtained;

$$f(x) = p_0K(x,0) + \int_0^x K(x,y)f(y)dy, \tag{2.6}$$

where $K(x,y) = \frac{\lambda(y)}{\mu(x)}\exp\{-\nu(x-y)\}$ for $0 \le y < x < \infty$.

To solve the above equation (2.6) we use the iteration method in Asmussen(1987). Define recursively

$$K_1(x,y) = K(x,y)$$
$$K_{n+1}(x,y) = \int_y^x K(x,z)K_n(z,y)dz, \quad n \ge 1.$$

It follows easily by induction that

$$K_n(x,y) = \frac{\lambda(y)}{\mu(x)}\exp\{-\nu(x-y)\}\frac{\{\omega(x)-\omega(y)\}^{n-1}}{(n-1)!}, \quad n \ge 1.$$

Since $\omega(x) \leq \lambda(0)\theta(x) < \infty$ for all $x \geq 0$, $K^*(x,y) = \sum_{n=1}^{\infty} K_n(x,y)$ is well-defined and finite. Furthermore

$$K^*(x,y) = \frac{\lambda(y)}{\mu(x)} \exp\{-\nu(x-y) + \omega(x) - \omega(y)\}.$$

Iterating the equation (2.6) $N-1$ times, this yields

$$f(x) = p_0 \sum_{n=1}^{N} K_n(x,0) + \int_0^x K_N(x,y) f(y) dy.$$

Letting $N \to \infty$ and using the bounded convergence theorem we get

$$f(x) = p_0 K^*(x,0)$$

which implies the density $f(x)$ and the probability p_0 in equation (2.2). □

Example 2.1. Consider a M/M/1 queue which has an arrival rate λ and an exponential service time with parameter ν. In this example the service time is considered as the amount of the work to be processed. The server processes the work at the rate $\mu(x) = 1 + x$ ($\mu(0) = 0$) when the total work in system $W(t)$ is equal to x. Let $V(t)$ denote the virtual waiting time at time t. Then

$$V(t) = \int_0^{W(t)} \frac{1}{1+x} dx = \ln(1 + W(t)).$$

We assume that the arriving customer is rejected if the virtual waiting time in front of the customer exceeds a specified time V_0, that is, the arriving customer who finds his waiting time exceeds his impatience V_0 is not admitted. Then it follows that the arrival process is the nonhomogeneous Poisson process with the intensity

$$\lambda(x) = \begin{cases} \lambda, & 0 \leq x \leq \exp\{V_0\} - 1 \\ 0, & x > \exp\{V_0\} - 1. \end{cases}$$

Since the service rate $\mu(x)$ satisfies the stability conditions (2.1) the stationary distribution of $W(t)$ is given by

$$\begin{aligned} p_0 &= [1 + \lambda \int_0^{\infty} g(x) dx]^{-1} \\ f(x) &= \lambda p_0 g(x), \quad x > 0, \end{aligned}$$

where

$$g(x) = \begin{cases} \exp\{-\nu x\}(1+x)^{\lambda - 1}, & 0 < x \leq \exp\{V_0\} - 1 \\ \exp\{\lambda V_0 - \nu x\}(1+x)^{-1}, & x > \exp\{V_0\} - 1. \end{cases}$$

3. FIRST EXIT PROBABILITY

Harrison and Resnick(1976) obtained the first exit probability for the storage process of which the arrival times are work-independent. Using the time transformation we can extend their result to the case of work-modulated NHPP arrivals. We assume that $0 \leq a < b < \infty$. Let $t^*(a) = \inf\{t > 0 | W(t) \leq a\}$ and $t^*(b) = \inf\{t > 0 | W(t) > b\}$ and we define

$$U(x) = P\{t^*(b) < t^*(a) | W(0) = x\} \quad \text{for } x \geq 0.$$

Then $U(x)$ is the probability that the work level b is exceed before the level a is reached, starting from x. It follows immediately that $U(x) = 0$ for $x \in [0, a]$ and $U(x) = 1$ for $x \in (b, \infty)$ from the fact that $W(t)$ almost surely converges to $W(0)$ as $t \to 0$. To obtain $U(x)$ for $x \in (a, b]$ consider the storage process $\{V(t), t \geq 0\}$ with an arrival intensity 1, release rate $r(x) = \mu(x)/\lambda(x)$ and the exponential size-jump with mean $1/\nu$.

Proposition 3.1. *The storage process* $\{V(t), t \geq 0\}$ *and the workload process* $\{W(t), t \geq 0\}$ *can be coupled as random time transformation of each other, that is*

$$W(t) = V(\Lambda(t)),$$

where $\Lambda(t) = \int_0^t \lambda(W(s))ds$.

Proof: Define

$$W_1(t) = V(\Lambda(t)),$$

where $\Lambda(t) = \int_0^t \lambda(W_1(s))ds$. Then the probability that $W_1(t)$ has an arrival in $[t, t + \delta t]$ is

$$[\Lambda(t + \delta t) - \Lambda(t)] \cdot 1 = \Lambda'(t)\delta t = \lambda(W_1(t))\delta t.$$

Also, in between jumps we have

$$\begin{aligned} \frac{dW_1(t)}{dt} &= \frac{dV(\Lambda(t))}{d\Lambda(t)} \cdot \frac{d\Lambda(t)}{dt} \\ &= -r(V(\Lambda(t)))\lambda(W_1(t)) \\ &= -\mu(W_1(t)). \end{aligned}$$

Hence $\{W_1(t), t \geq 0\}$ has the arrival intensity $\lambda(W_1(t))$ and the service rate $\mu(W_1(t))$ so that this process is our workload process $\{W(t), t \geq 0\}$. □

Theorem 3.1. *Let* $\alpha = 1 - U(b)$. *Then under the assumptions (2.1) for* $x \in (a, b]$,

$$U(x) = \int_a^x u(y)dy,$$

where

$$u(x) = \alpha \frac{\lambda(x)}{\mu(x)} \exp\{-\nu(b-x) + \omega(b) - \omega(x)\} \tag{3.1}$$

and

$$\alpha = [1 + \int_a^b \frac{\lambda(x)}{\mu(x)} \exp\{-\nu(b-x) + \omega(b) - \omega(x)\}dx]^{-1}.$$

Proof: Let

$$\begin{aligned} T^*(a) &= \inf\{T > 0 | V(T) \leq a\} \\ T^*(b) &= \inf\{T > 0 | V(T) \geq b\}. \end{aligned}$$

Then it easily follows that

$$\begin{aligned} t^*(a) &= \Lambda^{-1}(T^*(a)) \\ t^*(b) &= \Lambda^{-1}(T^*(b)). \end{aligned}$$

Since both Λ and Λ^{-1} are strictly increasing we have that $t^*(b) < t^*(a)$ if and only if $T^*(b) < T^*(a)$. Therefore we get that

$$U(x) = P\{T^*(b) < T^*(a) | V(0) = x\}.$$

Let T_1 denote the first input time of the storage process $\{V(T), T \geq 0\}$ and S_1 its jump size. Notice that $\omega(x) - \omega(a) = \int_a^x [r(y)]^{-1}dy$ is the time required for $\{V(T), T \geq 0\}$ to move from the storage level x down to level a. By the strong Markov property we have that

$$\begin{aligned} U(x) &= E[U(V(T_1)), T_1 < \omega(x) - \omega(a) | V(0) = x] \\ &= \int_0^{\omega(x)-\omega(a)} \exp\{-t\} E[U(\omega^{-1}(\omega(x) - t) + S_1)]dt \\ &= \int_a^x \frac{1}{r(y)} \exp\{-[\omega(x) - \omega(y)]\} \int_0^\infty U(y+z)\nu \exp\{-\nu z\}dzdy. \end{aligned} \tag{3.2}$$

Taking the factor $\exp\{-\omega(x)\}$ outside the integral, we can differentiate (3.2) to get

$$u(x) = \frac{1}{r(x)} \int_0^\infty [U(x+z) - U(x)]\nu \exp\{-\nu z\} dz. \tag{3.3}$$

Substituting

$$U(x+z) - U(x) = \begin{cases} \int_x^{x+z} u(y)dy & \text{if } x+z \le b \\ \int_x^b u(y)dy + \alpha & \text{if } x+z > b. \end{cases}$$

into equation (3.3) yields

$$u(x) = \frac{1}{r(x)}[\alpha \exp\{-\nu(b-x)\} + \int_x^b u(y) \exp\{-\nu(y-x)\} dy]. \tag{3.4}$$

Recalling $K(x,y) = \lambda(y) \exp\{-\nu(x-y)\}/\mu(x)$ and $r(x) = \mu(x)/\lambda(x)$, we rewrite (3.4) as

$$\frac{\mu(x)}{\mu(b)} u(x) = \alpha K(b,x) + \int_x^b K(y,x) \frac{\mu(y)}{\mu(b)} u(y) dy$$

and iterate this relationaship $N-1$ times to obtain

$$\frac{\mu(x)}{\mu(b)} u(x) = \alpha \sum_{n=1}^{N} K_n(b,x) + \int_x^b K_N(y,x) \frac{\mu(y)}{\mu(b)} u(y) dy.$$

Letting $N \to \infty$ under the conditions (2.1) in section 2 we obtain

$$\frac{\mu(x)}{\mu(b)} u(x) = \alpha K^*(b,x)$$

which implies (3.1). □

REFERENCES

Asmussen, S. (1987). *Applied Probability and Queues.* John Wiley & Sons.

Brockwell, P. J., Resnick, S. I. and Tweedie, R. L. (1982). "Storage processes with general release rule and additive inputs," *Advances in Applied Probability,* **14**, 392-433.

Browne, S. and Sigman, K. (1992). "Work-modulated queues with applications to storage processes," *Journal of Applied Probability,* **29**, 699–712.

Harrison, J. M. and Resnick, S. I. (1976). "The stationary distribution and first exit probabilities of a storage process with general release rule," *Mathematics of Operations Research*, **1**, 347–358.

Knessl, D., Matkowsky, B. J., Schuss, Z. and Tier, C. (1987). "Busy period distribution in state-dependent queues," *Queueing Systems*, **2**, 285–305.

Perry, D. and Asmussen, S. (1995). "Rejection rules in the M/G/1 queue," *Queueing Systems*, **19**, 105–130.

Whitt, W. (1990). "Queues with service times and interarrival times depending linearly and randomly upon waiting times," *Queueing Systems*, **6**, 335–352.