Journal of the Korean Statistical Society (2004), **33**: 2, pp 159–167

A SIMPLE APPROACH TO THE WORKLOAD ANALYSIS OF $M/G/1$ VACATION QUEUES†

NAM KI KIM[1], YON IL PARK[2] AND KYUNG CHUL CHAE[3]

ABSTRACT

We present a simple approach to finding the stationary workload of $M/G/1$ queues having generalized vacations and exhaustive service discipline. The approach is based on the level crossing technique. According to the approach, all that we need is the workload at the beginning of a busy period. An example system to which we apply the approach is the $M/G/1$ queue with both multiple vacations and D-policy.

AMS 2000 subject classifications. Primary 60K25; Secondary 90B05.
Keywords. Infinite dam, compound Poisson input, unfinished work, multiple vacations, D-policy, decomposition property.

1. INTRODUCTION

In this paper, we present a simple and instructive approach to finding the stationary workload for a class of $M/G/1$ queueing systems under the following assumptions.

(A.1) During a busy period customers are served continuously until there are no customers in the system (exhaustive service discipline).

(A.2) A predetermined stopping rule, which is independent of future arrivals, governs when to stop an idle period and start the next busy period (generalized vacation).

Received October 2003; accepted December 2003.
†This work was supported by Korea Research Foundation Grant (KRF-2002-070-C00021).
[1]Department of Industrial Engineering, Chonnam National University, Gwangju 500-757, Korea (e-mail : freedom@chonnam.ac.kr)
[2]National Library of Medicine, Bethesda, MD 20894, U.S.A. (e-mail : park@ncbi.nlm.nih.gov)
[3]Department of Industrial Engineering, KAIST, Daejeon 305-701, Korea (e-mail : kcchae@kaist.ac.kr)

The workload of such a class of queues corresponds to the water level of an infinite dam with a compound Poisson input and with the following output policy. During a release period, the water is released at a constant rate until the reservoir becomes empty (A.1). Then we stop releasing the water until the next release period allowing the water level increase jumpwise. When to begin to release the water is governed by a predetermined rule which is independent of the future input (A.2).

A typical rule is the D-policy, under which we begin to release the water at the instant the water level exceeds a predetermined constant D or the single server begins to serve the customers at the instant the total workload brought in by the customers who arrived during an idled period exceeds D (see Park and Chae, 1999 and 2003). Note that the D-policy corresponds to the special case $M = 1$ and $\lambda = D$ in the P_λ^M-policy of Lee and Ahn's (1998) dam model.

Suppose the water level is monitored not continuously but at discrete random points in time. The inter-monitoring times are assumed to be *iid* exponential by Lee and Lee (1993) and *iid* general random variables by Park *et al.* (2000). These models correspond to the $M/G/1$ queue with both D-policy and multiple vacations: *iid* exponential vacations in the former and *iid* general vacations in the latter.

In the references cited thus far, authors solve Chapman-Kolmogorov equations to obtain desired results. Our approach, on the other hand, is based on the level crossing technique (LCT), which is simpler and intuitively appealing (see, *e.g.*, Kim and Lee, 2002).

By using LCT, we show that desired results can be obtained *via* the workload at the beginning of a busy period for the $M/G/1$ queues satisfying (A.1) and (A.2). For such queues, we also prove the so-called decomposition property.

Then as a demonstration of our approach, we find the stationary workload of the $M/G/1$ queue with both D-policy and multiple vacations. It should be noted that the model of Park *et al.* (2000) corresponds only to the idle period of the $M/G/1$ queue with D-policy and multiple vacations.

Readers who are interested in the queue length distribution of the $M/G/1$ queue with D-policy are referred to Chae and Park (2001) and Dshalalow (1998).

2. LCT BASED APPROACH

2.1. Preliminaries

We consider $M/G/1$ queues satisfying (A.1) and (A.2). In an $M/G/1$ queue, customers arrive according to a Poisson process with a rate λ. The service times (or the workloads brought in by the customers) are *iid* general random variables, denoted by S, and independent of the arrival process. We assume that $\rho = \lambda E(S) < 1$ to guarantee the stationary workload process.

Let W and $W^*(\theta)$ denote the stationary workload and its Laplace-Stieltjes transform (LST), respectively. We can express $W^*(\theta)$ as

$$W^*(\theta) = P(I)W^*(\theta|I) + P(B)W^*(\theta|B) \tag{2.1}$$

where $P(I)$ ($P(B)$) denotes the probability that the server is idling (busy) and $W^*(\theta|I)$ ($W^*(\theta|B)$) denotes the corresponding conditional LST. It is well known that

$$P(B) = \rho = \lambda E(S) \text{ and } P(I) = 1 - \rho. \tag{2.2}$$

2.2. Applying LCT to W

The rationale of LCT is as follows. The rate of transitions from '$W \leq x$' to '$W > x$', $x \geq 0$, should be balanced by the rate of transitions from '$W > x$' to '$W \leq x$', where the former is called the upcrossing rate and the latter the downcrossing rate.

During a busy period both upcrossings and downcrossings are possible but during an idle period only upcrossings are possible. The downcrossing rate is known to be $P(B)dF_B(x)/dx$, $x > 0$, where $F_B(x)$ denotes the conditional (cumulative) distribution function of W given that the server is busy (see Cohen, 1977; Lee, 1998, p. 419). Note that

$$W^*(\theta|B) = \int_0^\infty e^{-\theta x} dF_B(x). \tag{2.3}$$

On the other hand, $W^*(\theta|I)$ is given by

$$W^*(\theta|I) = P_I(0) + \int_0^\infty e^{-\theta x} dF_I(x), \tag{2.4}$$

where $P_I(0)$ denotes the conditional probability that $W = 0$ and $F_I(x)$ denotes the conditional distribution function of W both given that the server is idling.

Upcrossings occur only when customers arrive. We can express the upcrossing rate as follows.

$$\lambda P(B)\int_0^x P(S>x-y)dF_B(y)$$
$$+\lambda P(I)\left\{P_I(0)\cdot P(S>x)+\int_0^x P(S>x-y)dF_I(y)\right\}. \qquad (2.5)$$

Note that the two integrals in (2.5) are the so-called convolutions. Thus, the LST of (2.5) can be expressed in terms of (2.3) and (2.4) as

$$\lambda P(B)W^*(\theta|B)\{1-S^*(\theta)\}\theta^{-1}+\lambda P(I)W^*(\theta|I)\{1-S^*(\theta)\}\theta^{-1}, \qquad (2.6)$$

where $\{1-S^*(\theta)\}/\theta=\int_0^\infty e^{-\theta x}P(S>x)dx$ and $S^*(\theta)$ is the LST of S.

Finally, we equate (2.6) with $P(B)W^*(\theta|B)$ which is the LST of the down-crossing rate. Then substituting (2.2) and solving for $W^*(\theta|B)$, we have that

$$W^*(\theta|B)=W^*_{M/G/1}(\theta)\cdot\frac{1-S^*(\theta)}{\theta E(S)}\cdot W^*(\theta|I), \qquad (2.7)$$

where

$$W^*_{M/G/1}(\theta)=\frac{\theta(1-\rho)}{\theta-\lambda+\lambda S^*(\theta)} \qquad (2.8)$$

is the workload LST of the standard $M/G/1$ queue.

2.3. The decomposition property

Substituting (2.2), (2.7), (2.8) into (2.1) and simplifying the result, we have that

$$W^*(\theta)=W^*_{M/G/1}(\theta)W^*(\theta|I). \qquad (2.9)$$

(2.9) is called the decomposition property. Due to (2.9), finding $W^*(\theta)$ reduces to finding $W^*(\theta|I)$ since $W^*_{M/G/1}(\theta)$ is already known. Sometimes, however, finding $W^*(\theta|B)$ is easier than finding $W^*(\theta|I)$. In such cases, we can make use of the following relation which is obtained by substituting (2.9) into (2.7):

$$W^*(\theta)=W^*(\theta|B)\cdot\frac{\theta E(S)}{1-S^*(\theta)}. \qquad (2.10)$$

One way to find $W^*(\theta|B)$ is as follows. Let Ω and $\Omega^*(\theta)$ denote the workload and its LST at the beginning of a busy period, respectively. Then, based on the delay cycle analysis (see, *e.g.*, Takagi, 1991, p. 27), we have that

$$W^*(\theta|B)=W^*_{M/G/1}(\theta)\cdot\frac{1-\Omega^*(\theta)}{\theta E(\Omega)}. \qquad (2.11)$$

3. $W^*(\theta)$ FOR THE $M/G/1$ QUEUE WITH D-POLICY AND MULTIPLE VACATIONS

3.1. D-policy and multiple vacations

Let $V_1, V_2, \ldots$ be *iid* random variables which are independent of both the arrival process and the service times. As soon as the system becomes empty, the server leaves for a vacation of length V_1. If the workload accumulated during V_1 exceeds D when he returns, he immediately begins to serve the customers. If not, he leaves for another vacation of length V_2, and so on until the total workload exceeds D when he returns from a vacation.

A useful notion introduced by Lee *et al.* (1994) is the grand vacation. The length of the i^{th} grand vacation, denoted by G_i, is defined as the following stopped random sum:

$$G_i = \sum_{k=1}^{K_i} V_{ik}, \; i = 1, 2, \ldots, \tag{3.1}$$

where the stopping time K_i is defined as follows: $K_i = 1$ if some customers arrive during V_{i1}; and $K_i = k \geq 2$ if no customers arrive during $V_{i1}, \ldots, V_{i\,k-1}$ but some arrive during V_{ik}.

Note that the customers arriving during G_i are those arriving during V_{iK_i}. Thus the probability generating function (PGF) of the number of arrivals during G_i equals the conditional PGF of the number of arrivals during a vacation given that the number of arrivals during this vacation is at least one. Therefore the PGF of the number of arrivals during G_i is

$$\frac{V^*(\lambda - \lambda z) - V^*(\lambda)}{1 - V^*(\lambda)}, \tag{3.2}$$

where $V^*(\theta)$ is the common (and unconditional) LST of *iid* V's. Note that $V^*(\lambda - \lambda z)$ is the unconditional PGF of the number of Poisson arrivals (with a rate λ) during a vacation and that $V^*(\lambda)$ is the probability of no arrivals during a vacation, thereby (3.2) is the conditional PGF.

Let U_i denote the workload accumulated during G_i. Since by definition, $G_1, G_2, \ldots$ are *iid* random variables, so are $U_1, U_2, \ldots$ Let $U^*(\theta)$ denote the common LST of *iid* U's. Recall that the service times are independent of both the arrival process and the vacation times. Thus we have that

$$U^*(\theta) = \frac{V^*(\lambda - \lambda S^*(\theta)) - V^*(\lambda)}{1 - V^*(\lambda)}, \tag{3.3}$$

which is obtained by substituting $S^*(\theta)$ for z in (3.2), and that

$$E(U) = \frac{\rho E(V)}{1 - V^*(\lambda)}. \tag{3.4}$$

3.2. Main results

Now we consider a renewal process whose interrenewal times consist of *iid* U's. Let $m(\cdot)$ denote the renewal function. Then we can express Ω as a stopped random sum as

$$\Omega = \sum_{i=1}^{N} U_i, \tag{3.5}$$

where the stopping time N has the following distribution and the expectation:

$$P(N = n) = P\left(\sum_{i=1}^{n-1} U_i \leq D, \sum_{i=1}^{n} U_i > D\right), \tag{3.6}$$

$$E(N) = m(D) + 1. \tag{3.7}$$

Note that $m(D)$ in (3.7) is the expected number of renewals occurred within $(0, D]$ and 1 represents the first renewal within (D, ∞). By the Wald's equation, we have that

$$E(\Omega) = E(N)E(U). \tag{3.8}$$

In addition, according to Park (2000), we have that

$$P(\Omega > x) = P(U > x) + \int_0^D P(U > x - y)dm(y), \;\; x \geq D. \tag{3.9}$$

Let $F_\Omega(x)$ and $F_U(x)$ denote the distribution functions of Ω and U, respectively. Since $F_\Omega(x) = 0$ for $x \leq D$, $\Omega^*(\theta)$ is defined as $\int_D^\infty e^{-\theta x} dF_\Omega(x)$. Thus from (3.9), we have that

$$\begin{aligned}
&\Omega^*(\theta) \\
&= \int_D^\infty e^{-\theta x} dF_U(x) + \int_{x=D}^\infty e^{-\theta x} \int_{y=0}^D dF_U(x - y)dm(y) \\
&= \int_0^\infty e^{-\theta x} dF_U(x) - \int_0^D e^{-\theta x} dF_U(x) \\
&\quad + \int_{y=0}^D e^{-\theta y} \int_{x=D}^\infty e^{-\theta(x-y)} dF_U(x - y)dm(y)
\end{aligned}$$

$$= U^*(\theta) - \int_0^D e^{-\theta x} dF_U(x) + \int_{y=0}^D e^{-\theta y} \int_{x=y}^\infty e^{-\theta(x-y)} dF_U(x-y) dm(y)$$

$$- \int_{y=0}^D e^{-\theta y} \int_{x=y}^D e^{-\theta(x-y)} dF_U(x-y) dm(y)$$

$$= U^*(\theta) - \left\{ \int_0^D e^{-\theta x} dF_U(x) + \int_{y=0}^D \int_{x=y}^D e^{-\theta x} dF_U(x-y) dm(y) \right\}$$

$$+ U^*(\theta) \int_0^D e^{-\theta y} dm(y)$$

$$= U^*(\theta) - \int_0^D e^{-\theta x} dm(x) + U^*(\theta) \int_0^D e^{-\theta y} dm(y)$$

$$= U^*(\theta) - \{1 - U^*(\theta)\} \int_0^D e^{-\theta x} dm(x)$$

$$= U^*(\theta) \left\{ 1 + \int_0^D e^{-\theta x} dm(x) \right\} - \int_0^D e^{-\theta x} dm(x). \tag{3.10}$$

From (2.11), (3.3), (3.4), (3.8), (3.10), we have that

$$W^*(\theta|B)$$

$$= W^*_{M/G/1}(\theta) \left\{ \frac{1 - U^*(\theta)}{\theta E(U)} \right\} \left\{ \frac{1 + \int_0^D e^{-\theta x} dm(x)}{E(N)} \right\}$$

$$= W^*_{M/G/1}(\theta) \left[\frac{1 - V^*(\lambda - \lambda S^*(\theta))}{\{\lambda - \lambda S^*(\theta)\} E(V)} \right] \left\{ \frac{1 - S^*(\theta)}{\theta E(S)} \right\} \left\{ \frac{1 + \int_0^D e^{-\theta x} dm(x)}{E(N)} \right\}. \tag{3.11}$$

Then, from (2.7), (2.9), (3.11), we have that

$$W^*(\theta|I) = \left[\frac{1 - V^*(\lambda - \lambda S^*(\theta))}{\{\lambda - \lambda S^*(\theta)\} E(V)} \right] \left\{ \frac{1 + \int_0^D e^{-\theta x} dm(x)}{E(N)} \right\}, \tag{3.12}$$

$$W^*(\theta) = W^*_{M/G/1}(\theta) \left[\frac{1 - V^*(\lambda - \lambda S^*(\theta))}{\{\lambda - \lambda S^*(\theta)\} E(V)} \right] \left\{ \frac{1 + \int_0^D e^{-\theta x} dm(x)}{E(N)} \right\}. \tag{3.13}$$

3.3. Remarks on main results

We interpret (3.12) as follows. The workload observed at a random point during an idle period is the sum of two independent random variables: one is the workload at the beginning of the ongoing grand vacation and the other is the workload brought in by those (if any) who arrived during the elapsed grand vacation. Note that no customers arrived during the preceding vacations (if any) belonging to the ongoing grand vacation. Thus the customers who arrived during the elapsed grand vacation must be those who arrived during the elapsed time of the ongoing vacation. The LST of the elapsed vacation is $\{1-V^*(\theta)\}/\{\theta E(V)\}$. Then substituting '$\lambda - \lambda z$' for θ, we have the PGF of the number of arrivals during the elapsed vacation. Substituting again $S^*(\theta)$ for z, we have the LST of the workload brought in by these arrivals. This explains the first term at the right hand side of (3.12).

The second term can be explained by the renewal reward theorem. The denominator, $E(N)$, is the expected number of grand vacations in an idle period. The numerator is of the form $E(\sum_{i=1}^{N} e^{-\theta X_i})$, where X_i denotes the workload at the beginning of the i^{th} grand vacation. Since $P(X_1 = 0) = 1$, we have that $e^{-\theta \cdot 0} = 1$. Then we have that

$$\begin{aligned} E\left(\sum_{i=2}^{N} e^{-\theta X_i}\right) &= \sum_{n=2}^{\infty} P(N=n) E\left(\sum_{i=2}^{n} e^{-\theta X_i} \Big| N=n\right) \\ &= \sum_{n=2}^{\infty} P(N=n) \int_0^D \sum_{i=2}^{n} e^{-\theta x} P(x < X_i \le x + dx | N = n) \\ &= \int_0^D e^{-\theta x} \sum_{n=2}^{\infty} \sum_{i=2}^{n} P(N=n, x < X_i \le x + dx) \\ &= \int_0^D e^{-\theta x} dm(x). \end{aligned}$$

We consider two special cases as follows. For the $M/G/1$ queue with multiple vacations (and without D-policy), the last term of (3.13) (and of (3.11), (3.12)) disappears since $D = 0$ and $E(N) = 1$.

For the $M/G/1$ queue with D-policy (and without multiple vacations), on the other hand, two modifications are necessary. First, the middle term at the right hand side of (3.13) disappears since no customers can arrive during a vacation of length zero. Secondly, in the last term of (3.13), $E(N)$ becomes the expected number of arrivals during an idle period and $m(\cdot)$ becomes the renewal function of the renewal process whose interrenewal times consist of *iid* service times.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for their careful reading and valuable suggestions.

REFERENCES

CHAE, K. C. AND PARK, Y. I. (2001). "The queue length distribution for the $M/G/1$ queue under the D-policy", *Journal of Applied Probability*, **38**, 278–279.

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

DSHALALOW, J. H. (1998). "Queueing processes in bulk systems under the D-policy", *Journal of Applied Probability*, **35**, 976–989.

KIM, S. AND LEE, E. Y. (2002). "A level crossing approach to the analysis of finite dam", *Journal of the Korean Statistical Society*, **31**, 405–413.

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, E. Y. AND LEE, J. (1993). "A model for a system subject to random shocks", *Journal of Applied Probability*, **30**, 979–984.

LEE, H. W. (1998). *Queueing Theory*, Sigma Press, Seoul.

LEE, H. W., LEE, S. S., PARK, J. O. AND CHAE, K. C. (1994). "Analysis of the $M^X/G/1$ queue with N-policy and multiple vacations", *Journal of Applied Probability*, **31**, 476–496.

PARK, Y. I. (2000). *Analysis of Queueing Models under the D-policy*, Ph.D. Thesis, Department of Industrial Engineering, Korea Advanced Institute of Science and Technology.

PARK, Y. I. AND CHAE, K. C. (1999). "Analysis of unfinished work and queue waiting time for the $M/G/1$ queue with D-policy", *Journal of the Korean Statistical Society*, **28**, 523–533.

PARK, Y. I. AND CHAE, K. C. (2003). "Correction and addendum : Analysis of unfinished work and queue waiting time for the $M/G/1$ queue with D-policy", *Journal of the Korean Statistical Society*, **32**, 311.

PARK, Y. I., CHAE, K. C. AND LEE, H. S. (2000). "A random review replacement model for a system subject to compound Poisson shocks", *Stochastic Analysis and Applications*, **18**, 145–157.

TAKAGI, H. (1991). *Queueing Analysis*, Vol. 1, North-Holland, Amsterdam.