Queue length and waiting time of the M/G/1 queue under the $D$-policy and multiple vacations

Ho Woo Lee · Sahng Hoon Cheon · Won Joo Seo

Received: 31 August 2005 / Revised: 24 July 2006
© Springer Science + Business Media, LLC 2006

Abstract We study the steady-state queue length and waiting time of the M/G/1 queue under the $D$-policy and multiple server vacations. We derive the queue length PGF and the LSTs of the workload and waiting time. Then, the mean performance measures are derived. Finally, a numerical example is presented and the effects of employing the $D$-policy are discussed.

Keywords M/G/1 · $D$-policy · Multiple vacations · Queue length · Workload · Waiting time

AMS Subject Classifications 60K25

1 Introduction

This paper studies the queue length and waiting time of the M/G/1 queue under the mixed control of the $D$-policy and $T$-policy (multiple server vacations). In contrast with the well-known $N$-policy of Yadin and Naor [33] and the $T$-policy of Heyman [17], the $D$-policy controls the queueing system by the workload of the waiting customers. The $D$-policy can be employed in actual manufacturing settings to control the number of start-ups per unit time and thereby reduce the overall average cost per unit time.

In our system under study, the server leaves for repeated vacations as soon as the system becomes empty. It resumes its service only when the cumulative workload is found to be greater than the predetermined threshold $D$ at the end of a vacation. Figure 1 shows both the queue length and workload processes on the synchronized time scale.

The behavioral complexity of our system is in the relationship between the service times of the customers who arrive during the idle period. For example, assume that the threshold $D$ is crossed by the third customer and there are four customers at the start of the busy period (see the lower part of Fig. 1). Obviously the service times $\{S_{1(4)}, S_{2(4)}, S_{3(4)}\}$ of the first three customers are not independent. Moreover, they are stochastically different from the ordinary service time random variable $S$ because $S_{1(4)}$ and $S_{2(4)}$ are smaller than $D$. The server spends $U_D = S_{1(4)} + S_{2(4)} + S_{3(4)} + S_{4(4)}$ amount of time on serving the four ‘special’ customers, during each of which, it is necessary to keep track of the arrivals of the ‘ordinary’ customers who have iid ordinary service times.

Studies on the $D$-policy queueing systems were pioneered by Balachandran [4], Balachandran and Tijms [5], Boxma [6], and Tijms [30] for the M/G/1 queue. Their primary concern was in the optimal control of $D$ under a linear cost structure. While these authors used the mean workload, Chae and Park [10] used the mean queue length to determine the optimal $D$.

Boxma [6] showed that the optimal $D$-policy is superior over the optimal $N$-policy for all service time distributions if the cost function consists of the startup cost and linear workload holding cost. But, Artalejo [2, 3] showed that the same is no longer true if the mean queue length is used (instead of the mean workload) in the cost function.

Gakis et al. [15] derived the distributions of the idle and busy periods under simple and dyadic policies. Sivazlian [28] provided an approximate formula for optimal $D$ in terms of
the first three moments of the service time. His formula was exact under exponential service times. Rhee [24] developed a new methodology to find the expected busy periods for controllable M/G/1 queueing models.

Li and Niu [21] considered the GI/G/1/D-policy queue and derived the waiting time distributions in transform-free style. Lillo and Martin [22] claimed the superiority of the \(N\)-policy over the \(D\)-policy if the mean queue length is used, but their argument was based on the erroneously derived mean queue length. Feinberg and Kella [14] considered switching costs, running costs and holding costs per unit time and proved the optimality of the \(D\)-policies. Readers are referenced to Tijms [31] for approximate numerical results and related analysis.

Lee and Song [18] and Lee et al. [19] studied the queue length and the waiting time of the MAP/G/1/D-policy queue.

Due to the behavioral and analytical complexities inherent in the \(D\)-policy queueing systems, the study on the queue length could not be found until Rubin and Zhang [26] studied the switch-on policies for communications and queueing systems. Readers could find the mean queue length as part of their formula which may be considered as the first success on the derivation of the mean queue length of the M/G/1/D-policy queue. Dshalalow [13] carried out the first extensive study on the queue length process of the batch arrival \(D\)-policy queues with vacations. But as pointed out in Artalejo [2] and Chae and Park [9, 10], his \(D\)-policy did not agree with the classical \(D\)-policy in the true sense because he implicitly assumed that the customers who have arrived during the idle period are assigned totally new iid service times when the busy period begins. But as noted in Artalejo [2], Dshalalow’s approach can be considered as a practical alternative to reduce the complexity of the \(D\)-policy. The queue length analysis of the same system under the genuine \(D\)-policy was carried out later by Agarwal and Dshalalow [1]. They applied the fluctuations of the multivariate delayed marked renewal process with mutually dependent components developed in [13]. The main difference between this paper and the work of [1] is that we are taking more rudimentary approach by just using the information of the workload at an arbitrary point of time during the idle period, and the workload and the queue length at the start of the busy period. For this purpose we invent the Workload Grand Vacation Process (WGVP) which allows us to handle the workload and the number of customers during the vacations collectively. Another difference is that we are also deriving the workload and waiting time distributions in this paper. Chae and Park [9] derived the probability generating function (PGF) of the queue length of the M/G/1/D-policy queue. Artalejo [2] derived the complete queue length distribution. Lee et al. [20] analyzed the queue length of the \(M^X/G/1\) queue under the \(D\)-policy.