stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BN 19/73 AUGUST

H.C. TIJMS
ON A CONTROL POLICY FOR A SINGLE SERVER SYSTEM

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 60K25, 90B05

ON A CONTROL POLICY FOR A SINGLE SERVER SYSTEM

HENK TIJMS

ABSTRACT

This paper considers for the M/G/1 queue a control policy which turns the server off when the system is empty and turns him on when the total work to be done exceeds a given value. The purpose of this paper is to derive an expression for the average queue length and to correct an expression recently found in the literature for the average amount of work to be done.

Introduction

Customers arrive at a single server service-station in accordance with a Poisson process with rate λ . The service times of the customers are independent random variables having a common probability distribution function $F(t)$ with finite first moment μ and finite second moment $\mu^{(2)}$. It is assumed that $F(0) = 0$ and $\rho < 1$, where $\rho = \lambda\mu$. The service time of each customer is known upon his arrival. The policy for controlling the system is to turn the server off when the system is empty and to turn him on when the cumulative service times of the customers in the system exceeds the value D, where D is a given non-negative number. Following Balachandran [1], we call this policy the D-policy. When turned on the server provides service, where the customers are served in order of arrival.

The purpose of this paper is to derive an expression for the average number of customers in the system and to correct an expression found in [1] for the average amount of work to be done. The formula (10) in [1] is incorrect (except for constant service times) due to the fact the relations (7) and (9) in [1] are false; the relation (7) is false because $M_D Y_i$ and S_i are dependent. The incorrectness of the relations (7) and (9) may be verified also directly by taking $D = 3$ and $P\{S_i=2\} = P\{S_i=4\} = 1/2$. Also, formula (20) in [1] is incorrect. Finally, it will appear that under the cost structure given in [1] the D-policy is superior to the control policy based on the queue size when the service time is exponential.

Preliminary results

For convenience we assume throughout this paper that the server is turned off at epoch 0, so the system is empty at epoch 0. Let T be the next epoch at which the server is turned off. For any $t \geq 0$, let $L(t)$ be the number of customers in the system at time t (including the customer being served at time t, if any). Observe that $\int_0^t L(s)ds$ represents the total time spent by customers in the system during $[0,t]$. For any $t \geq 0$, let $V(t)$ be the sum of the cumulative service times of the customers queueing at time t and the remaining service time of the customer being served at time t (if any).

The epoch T is a regeneration epoch for each of the processes $\{L(t)\}$ and $\{V(t)\}$. We shall see hereafter that T, $\int_0^T L(s)ds$ and $\int_0^T V(s)ds$ have finite expectations. Now, by the theory of regenerative processes (see p. 98 in Ross [3] and Theorem 1 of Stidham [4]),

$$L = \lim_{t\to\infty}(1/t)E\{\int_0^t L(s)ds\} \text{ and } V = \lim_{t\to\infty}(1/t)E\{\int_0^t V(s)ds\}$$

exist and are given by

$$(1) \qquad L = E\{\int_0^T L(s)ds\}/ET \text{ and } V = E\{\int_0^T V(s)ds\}/ET.$$

Let $M(t) = \sum_{n=1}^{\infty} F^n(t)$ for $t \geq 0$, where $F^n(t)$ denotes the n-fold convolution of $F(t)$ with itself. From $M(t) = F(t) + \int_0^t M(t-y)dF(y)$ the renewal function $M(t)$ may be computed (see also Jacquette [2]). Let $X_0 = 0$, let X_k be the service time of the kth customer , and let τ_k be the arrival epoch of the kth customer. For any $x \geq 0$, let

$$\nu(x) = \min\{n \mid \sum_{k=1}^{n} X_k > x\},\ T(x) = \tau_{\nu(x)},\ W(x) = \int_0^{T(x)} L(s)ds,$$
$$Y(x) = \sum_{k=1}^{\nu(x)} X_k \text{ and } Z(x) = \sum_{k=0}^{\nu(x)-1} X_k.$$

That is, for the x-policy, $T(x)$ is the first epoch at which the server is turned on, $\nu(x)$ is the number of customers present at epoch $T(x)$, $W(x)$ is the total time spent by customers in the system up to epoch $T(x)$, and $Y(x)$ ($Z(x)$) is the cumulative service times of the customers present at (just before) epoch $T(x)$. Also, let $K(x)$ be distributed as the number of customers to arrive during the time needed to serve the customers present at epoch $T(x)$. We will use repeatedly the following fact (e.g. p. 17 in [3]): Given that n arrivals have occurred in $(0,s)$, then each of the n arrival epochs has expectation $s/2$.

THEOREM 1. *Let* $a(x) = 2\int_0^x t[1+M(x-t)]dF(t)$, $x \geq 0$. *Then, for any* $x \geq 0$,

(a) $E\nu(x) = 1 + M(x)$, $E\{\nu(x)(\nu(x)-1)\} = 2M(x) + 2\int_0^x M(x-y)dM(y)$.

(b) $ET(x) = (1/\lambda)E\nu(x)$, $EW(x) = (1/2\lambda)E\{\nu(x)(\nu(x)-1)\}$.

(c) $EY(x) = \mu E\nu(x)$, $E[Y(x)]^2 = \mu^{(2)}[1+M(x)] + \mu[a(x)+\int_0^x a(x-y)dM(y)]$.

(d) $EK(x) = \lambda EY(x)$, $E\{K(x)(K(x)-1)\} = \lambda^2 E[Y(x)]^2$.

(e) $E\{\sum_{k=1}^{\nu(x)} (X_1+\ldots+X_k)\} = (\mu/2)[E\{\nu(x)(\nu(x)-1)\}+2E\nu(x)]$.

(f) $E\{\nu(x)Z(x)\} = a(x) + \int_0^x a(x-y)dM(y)$.

PROOF. (a) Part (a) is well known, e.g. p. 32 and p. 57 in [3].

(b) By $E\{T(x)|\nu(x)\} = n/\lambda$ and $E\{W(x)|\nu(x)=n,\ T(x)=t\} = (n-1)t/2$, we get (b).

(c) By Wald's equation, $EY(x) = \mu E\nu(x)$. Under the condition that $X_1 = t$ we have $Y(x) = t$ when $t > x$ and $Y(x) = t + Y(x-t)$ when $t \leq x$. Hence

$$E[Y(x)]^2 = \mu^{(2)} + 2\int_0^x tEY(x-t)dF(t) + \int_0^x E[Y(x-t)]^2 dF(t), \qquad x \geq 0.$$

From this renewal equation (cf. [3]) we get the desired result.

(d) The distribution of $K(x)$ given that $Y(x) = t$ is Poisson with mean λt.

(e) We have $E\{\sum_{k=1}^{\nu(x)} (X_1+\ldots+X_k)\} = \mu\sum_{n=1}^{\infty} nP\{\nu(x)\geq n\}$ as can be shown by copying the proof of Theorem 3.6 in [3]. Now, (e) easily follows.

(f) Conditioning on X_1 yields $EZ(x) = \int_0^x tdF(t) + \int_0^x EZ(x-t)dF(t)$. The solution of this renewal equation yields after partial integration,

$$EZ(x) = \int_0^x tdF(t) + \int_0^x tM(x-t)dF(t), \qquad x \geq 0.$$

From $E\{\nu(x)Z(x)\} = \int_0^x E[\{(1+\nu(x-t)\}\{t+Z(x-t)\}]dF(t)$, we get

$$E\{\nu(x)Z(x)\} = EZ(x) + \int_0^x t[1+M(x-t)]dF(t) + \int_0^x E\{\nu(x-t)Z(x-t)\}dF(t), \quad x\geq 0.$$

From this renewal equation we get the desired result. This ends the proof.

Finally, for the standard M/G/1 queue, let t_b be the expected length of a busy period and let w_b be the expected total time spent by customers in the system during one busy period. Then,

$$(2) \qquad t_b = (1-\rho)^{-1}\mu \text{ and } w_b = (1-\rho)^{-1}\mu + [2(1-\rho)^2]^{-1}\lambda\mu^{(2)}.$$

as follows from $t_b = \mu + \lambda\mu t_b$ and $w_b = \mu + \lambda\mu^{(2)}/2 + \lambda\mu w_b + \lambda^2\mu^{(2)}t_b/2$.

The averages L and V

It is readily seen that

$$(3) \quad ET = ET(D) + EY(D) + EK(D)t_b,$$

$$(4)\qquad E\{\int_0^T L(s)ds\} = EW(D) + E\{\sum_{k=1}^{\nu(D)}(X_1+\ldots+X_k)\} + (\lambda/2)E[Y(D)]^2 + \\ + EK(D)w_b + (1/2)E\{K(D)(K(D)-1)\}t_b.$$

By the formulae (2) and (4) in [1] (in the derivation of formula (4) in [1] it is helpful to read $j[\alpha(\nu/ES-1)ES + (\nu/ES)ES^2/2]$ for the left-hand side of the equation on the last line on p. 1014 in [1]),

$$(5)\qquad E\{\int_0^T V(s)ds\} = (2\lambda)^{-1}\, E\{\nu(D)Z(D)\} + [2(1-\rho)]^{-1}\, E[Y(D)]^2 + \\ + [2(1-\rho)^2]^{-1}\, \rho\mu^{(2)}\, E\nu(D).$$

Together (1) - (5) and Theorem 1 yield after some algebra the next theorem.

THEOREM 2. *Let* $b(x) = \int_0^x t[1+M(x-t)]dF(t)$, $x \geq 0$. *Then, for any D-policy,*

$$L = \rho + [2(1-\rho)]^{-1}\lambda^2\mu^{(2)} + [1+M(D)]^{-1}[(1-\rho^2)\psi_1(D)+\lambda\rho\psi_2(D)],$$
$$V = [2(1-\rho)]^{-1}\lambda\mu^{(2)} + [1+M(D)]^{-1}\psi_2(D),$$

where $\psi_1(D) = M(D) + \int_0^D M(D-y)dM(y)$ *and* $\psi_2(D) = b(D) + \int_0^D b(D-y)dM(y)$.

Consider the case where $F(t) = 1 - e^{-\alpha t}$. Then, by $M(t) = \alpha t$,

$$L = \rho + \rho^2(1-\rho)^{-1} + [2(1+\alpha D)]^{-1}[\alpha^2D^2+2\alpha(1-\rho^2)D],$$
$$V = [\alpha(1-\rho)]^{-1}\, \rho + [2(1+\alpha D)]^{-1}\alpha D^2.$$

Finally, for the cost structure considered in [1] straightforward calculations show that for exponential service the best D-policy has a lower average cost than the best policy in the class of policies which turn the server off when the system is empty and turn him on when the queue size reaches a certain number (see [1] for the average cost of the latter policy).

References

1. BALACHANDRAN, K.R., "Control Policies for a Single Server System," *Management Sciences,* Vol 19 (1973), pp. 1013 - 1018.

2. JACQUETTE, D.J., "Approximations for the Renewal Function m(t)", *Operations Research,* Vol. 20 (1972), pp. 722 - 727.

3. ROSS, S.M., *Applied Probability Models with Optimization Applications,* Holden-Day, Inc., California, 1969.

4. STIDHAM, S., JR., "Regenerative Processes in the Theory of Queues, with Applications to the Alternating Priority Queue", *Advances in Applied Probability,* Vol. 4 (1972), pp. 542 - 577.