MANAGEMENT SCIENCE
Vol. 22, No.2, October 1975
Printed in U.S.A.

UTILIZATION OF IDLE TIME IN AN $M/G/1$ QUEUEING SYSTEM*

YONATAN LEVY AND URI YECHIALI

Tel-Aviv University

This paper studies an $M/G/1$ queue where the idle time of the server is utilized for additional work in a secondary system. As usual, the server is busy as long as there are units in the main system. However, as soon as the server becomes idle he leaves for a "vacation." The duration of a vacation is a random variable with a known distribution function. Two models are considered. In the first, upon termination of a vacation the server returns to the main queue and begins to serve those units, if any, that have arrived during the vacation. If no units have arrived the server waits for the first arrival when an ordinary $M/G/1$ busy period is initiated. In the second model, if the server finds the system empty at the end of a vacation, he immediately takes another vacation, etc. For both models Laplace-Stieltjes transforms of the occupation period, vacation period and waiting time are derived and generating functions of the number of units in the system are calculated. The two models are then compared to each other, and for some special cases the optimal mean vacation times are found.

1. Introduction

We consider an $M/G/1$ queueing system where, as usual, the server serves the queue continuously as long as there is at least one unit in the system. When the server finishes serving a unit and finds the system empty he goes away for a length of time called a vacation. The vacation time is utilized for some additional work so that the idle time of the server is not completely lost. At the end of the vacation the server returns to the main system.

We study two models. In the first one, upon termination of a single vacation, the server returns to the main queue and begins to serve those units, if any, that have arrived during the vacation. If no units have arrived the server waits for the first one to arrive when an ordinary $M/G/1$ busy period is initiated. In the second model, if the server finds the system empty at the end of a vacation, he immediately takes another vacation, and continues in this manner until he finds at least one waiting unit upon return from a vacation.

Although it seems that (because of the Poissonian properties of the arrival stream) Model 2 is superior to Model 1, there may be cases where Model 1 is the more reasonable one. For example, suppose that the server is a machine that is being checked each time it becomes idle. Clearly, there is no point in checking it again before it resumes operating again.

The analysis of the above two models, besides being interesting by itself as a possible solution for the problem of utilization of the server's idle time, may be used as an intermediate step for the analysis of other queueing models such as priority queues, cyclic queues, etc. For example, the second model has been partially used by Cooper [2] to analyze a system of queues served in cyclic order. In that study, for any given queue, say i, the "vacation time" is the length of time the server spends idle or working

* Processed by Professor Marcel F. Neuts, Departmental Editor for Queueing Theory and Stochastic Processes and by John P. Lehoczky, Associate Editor; received May 28, 1974, revised December 8, 1974. This paper has been with the authors 2 months for revision.

on other queues before returning and beginning service on queue i. However, Cooper was not interested in the length of time the server spends outside queue i, but only in the number of customers the server finds upon return. In our study we analyze this model completely (using a somewhat different technique) and obtain detailed and explicit results. Yadin and Naor [5] and Heyman [3] analyzed an $M/G/1$ system where the service facility is turned off when no customers are present and is turned on only when the rth unit has arrived. Yadin and Naor's model differs from ours in that the server returns to the main queue immediately when the rth unit arrives, while in our model the server reutrns to the main queue only upon termination of its vacation—independent of the number of customers present there.

In §2 Model 1 is analyzed. We obtain the Laplace-Stieltjes (LS) transforms for the occupation period and cycle time. We then consider an embedded Markov chain defined on an extended state space, derive the generating functions of the limiting probabilities and calculate the mean queue size.

Model 2 is studied in §3. LS transforms, generating functions and mean queue size are calculated. In addition, the average waiting time of a customer is derived as a limiting case of Cobham's [1] nonpreemptive priority model. In §4 we compare the two models and find optimal vacation lengths for some special cases.

2. Model 1

We consider an $M/G/1$ queueing system where the stream of arrivals is a homogeneous Poisson process with rate λ. The service times $V_1, V_2, \ldots$ are independent random variables having common distribution $H(v)$ and finite mean $E(V)$. When a service is completed and no customers are present in the system the server leaves for a "vacation" (which may be utilized for some additional and different work) whose duration U is a random variable with distribution $F(u)$ and finite mean $E(U)$. After finishing his "vacation" the server returns to the main system. If, on returning, the server finds customers waiting for him (customers who arrived during the "vacation" time U) he starts service immediately and keeps busy until the system becomes idle again and he leaves for another vacation. If no customers have arrived during the vacation time the server waits for the first customer to arrive when an ordinary $M/G/1$ busy period starts. At the termination of the busy period the server takes another vacation, etc.

2.1. *Occupation Period and Cycle Time*

Let the occupation period T_s be the total time elapsing from the moment the server returns from a vacation until he leaves for another one, and let $T_1, T_2, \ldots, T_i, \ldots$ represent a sequence of ordinary busy periods in an $M/G/1$ queue. Also, let N be the number of customers present at the end of a vacation, then

$$(1) \qquad \begin{aligned} T_s &= X + T_1 && \text{if } N = 0, \\ &= T_1 + T_2 + \ldots + T_N && \text{if } N \geqq 1. \end{aligned}$$

The cycle time, T, is thus given by $T = T_s + U$.

For any random variable, say Y, we denote its Laplace-Stieltjes transform (LST) by $\Gamma_Y(z) = E\{e^{-zY}\}$.

The LST's of T_s and T are now readily obtained. Let

$$(2) \qquad b_j = P(N = j) = \int_0^\infty e^{-\lambda t} (\lambda t)^j/j! \, dF(t) \qquad (j = 0, 1, 2, \ldots).$$

By conditioning on N and recalling $\Gamma_X(z) = \lambda/(\lambda + z)$, we obtain

$$\Gamma_{T_s}(z) = \frac{\lambda}{\lambda + z} \Gamma_{T_1}(z)\Gamma_U(\lambda) + \Gamma_U(\lambda - \lambda\Gamma_{T_1}(z)) - \Gamma_U(\lambda). \tag{3}$$

Since U and T_s are not independent the LST of T is not the product of the LST's of U and T_s. However, $\Gamma_T(z)$ is derived by conditioning on U and N

$$\Gamma_T(z) = \frac{\lambda}{\lambda - z} \Gamma_{T_1}(z)\Gamma_U(\lambda+z) + \Gamma_U(\lambda+z-\lambda\Gamma_{T_1}(z)) - \Gamma_U(\lambda+z). \tag{4}$$

Now, the mean cycle time is

$$E(T) = (1/\lambda + E(T_1))(\Gamma_t(\lambda) + \lambda E(U)). \tag{5}$$

Using the well-known result [4] that $E(T_1) = E(V)/(1 - \lambda E(V))$ we finally have

$$E(T) = \left[\frac{1}{(1 - \lambda E(V))}\right]\left[\frac{\Gamma_U(\lambda)}{\lambda} + E(U)\right] \tag{6}$$

and

$$E(T_s) = E(T) - E(U) = \left[\frac{1}{(1 - \lambda E(V))}\right]\left[\frac{\Gamma_U(\lambda)}{\lambda} + \lambda E(U)E(V)\right]. \tag{7}$$

Let $P_{0\bullet}$ denote the fraction of time the server spends on vacation. Clearly

$$P_{0\bullet} = \frac{E(U)}{E(T)} = [1 - \lambda E(V)] \frac{\lambda E(U)}{[\lambda E(U) + \Gamma_U(\lambda)]}. \tag{8}$$

The proportion of time that the server is not busy serving customers in the main system (i.e., when he is either on vacation or idle) is

$$\frac{E(U) + (1/\lambda)P(N = 0)}{E(T)} = 1 - \lambda E(V). \tag{9}$$

Result (9) indicates, as is intuitively clear, that the condition for the system to be in a steady state regime is $1 > \lambda E(V)$, as is the case in the ordinary $M/G/1$ queue. The difference between the two models is that in our case a single busy period is (stochastically) longer than a single busy period in the ordinary $M/G/1$ queue.

2.2. *An Extended-Markov-Chain Representation*

The common approach now would be to consider the system at epochs of service completion or vacation termination, and to define a Markov chain with transitions occurring at these instants. However, if we then want to find the mean number of customers in the system we cannot apply the standard argument that each departing unit leaves behind it precisely those units that arrived during its sojourn time. The problem is that if we define the state space to be $\{0, 1, 2, 3, \dots\}$ then for each state j there is no distinction between the epoch of departure and the epoch of vacation termination. To distinguish between these two instants we define an extended state space $\{(i, j): i = 0, 1; j = 0, 1, 2, \dots\}$ such that if $i = 0$ then j counts the number of customers at epochs of vacation termination, and if $i = 1$ then j denotes the number of customers immediately after a service completion.

If $t_1, t_2, \dots, t_n$ are transition moments and $\tau(t)$ denotes the state of the system at time t, then the sequence of random variables $(i_n, j_n) = \tau(t_n + 0)$ determines a semi-Markov chain with a law of transition given by

$$(10)\qquad \begin{aligned} (i_{n+1}, j_{n+1}) &= (1, j_n + \xi - 1), && j_n \geqq 1, \\ &= (1, \xi), && (i_n, j_n) = (0, 0), \\ &= (0, N), && (i_n, j_n) = (1, 0), \end{aligned}$$

where ξ denotes the number of arrivals during a service time.

Let $a_k = P(\xi = k) = \int_0^\infty e^{-\lambda V} (\lambda V)^k/k!\, dH(V)$ $(k = 0, 1, 2, \dots)$, then, when in a steady-state regime, there exists a unique distribution $\pi_{i,j} = \lim_{n\to\infty} P(i_n = i, j_n = j)$, $i = 0, 1; j = 0, 1, 2, \dots$, that satisfies

$$(11a)\qquad \pi_{0,j} = \pi_{1,0} b_j, \qquad j = 0, 1, 2, \dots,$$

$$(11b)\qquad \pi_{1,j} = \pi_{0,0} a_j + \textstyle\sum_{k=1}^{j+1} \pi_{\cdot k}\, a_{j-k+1}, \qquad j = 0, 1, 2, \dots,$$

$$(11c)\qquad \textstyle\sum_{(i,j)} \pi_{i,j} = 1,$$

where

$$(12)\qquad \pi_{\cdot j} = \pi_{0,j} + \pi_{1,j}.$$

Define the generating functions

$$(13a)\qquad A(z) = \textstyle\sum_{j=0}^\infty a_j z^j, \quad B(z) = \sum_{j=0}^\infty b_j z^j, \quad \text{and}$$

$$(13b)\qquad \begin{aligned} \pi_i(z) &= \textstyle\sum_{j=0}^\infty \pi_{i,j} z^j, \qquad (i = 0, 1); \\ \pi(z) &= \pi_0(z) + \pi_1(z) = \textstyle\sum_{j=0}^\infty \pi_{\cdot j} z^j. \end{aligned}$$

Multiplying each of the equations (11a) and (11b) by z^j and summing over j we arrive at

$$(14a)\qquad \pi_0(z) = \pi_{1,0} B(z),$$

$$(14b)\qquad \pi_1(z) = \pi_{0,0} A(z) + [\pi(z)A(z) - \pi_{\cdot 0} A(z)]/z.$$

Using (11a), (12) and (14) we obtain

$$(15a)\qquad \pi_1(z) = \frac{b_0 A(z)(z-1) + A(z)[B(z) - 1]}{z - A(z)}\, \pi_{1,0},$$

$$(15b)\qquad \pi(z) = \frac{zB(z) - A(z) + b_0 A(z)(z-1)}{z - A(z)}\, \pi_{1,0}.$$

It follows that a solution exists if $A(z) = z$ does not have a solution in the interval (0, 1), which is true if $\sum_{k=0}^\infty k a_k = \lambda E(V) < 1$.

Applying l'Hôpital's rule on (15b) and (15a) one gets

$$(16a)\qquad 1 = \pi(1) = \frac{1 + \lambda E(U) - \lambda E(V) + b_0}{1 - \lambda E(V)}\, \pi_{1,0},$$

$$(16b)\qquad \begin{aligned} \pi_{1\cdot} = \textstyle\sum_{j=0}^\infty \pi_{1,j} = \pi_1(1) &= \frac{b_0 + \lambda E(U)}{1 - \lambda E(V)}\, \pi_{1,0} \\ &= \frac{b_0 + \lambda E(U)'}{1 - \lambda E(V) + b_0 + \lambda E(U)}. \end{aligned}$$

$\pi_{1\cdot}$ is the probability that at a transition instant a service completion has occurred.

We proceed now to find the LST of the waiting time distribution of an arbitrary arrival. Let $C(\cdot)$ be the distribution function of an arbitrary unit's sojourn time, W, defined as the elapsed time between the unit's arrival and departure epochs. This customer leaves behind him j customers with probability $\pi_{1,j}/\pi_{1\cdot}$. Hence

$$\pi_{1,j}/\pi_{1\cdot} = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^j}{j!} dC(t), \qquad j = 0, 1, 2, \ldots . \tag{17}$$

Substituting (17) in (13) yields (analogous to the ordinary $M/G/1$ queue)

$$\pi_1(z) = \Gamma_W[\lambda(1-z)]\pi_{1\cdot}. \tag{18}$$

Differentiating (18) with respect to z we obtain

$$\pi_1'(1)/\pi_{1\cdot} = -\lambda\Gamma_W'(0) = \lambda E(W). \tag{19}$$

Hence, from Little's law, the mean number of customers in the system is

$$L = \pi_1'(1)/\pi_{1\cdot}. \tag{20}$$

The specific calculations for $\pi_1'(1)$ result in

$$L = \lambda E(V) + \frac{\lambda^2 E(V^2)}{2(1-\lambda E(V))} + \frac{\lambda^2 E(U^2)}{2(b_0 + \lambda E(U))}. \tag{21}$$

We observe that the first two terms in (21) give the well-known Khintchine-Pollaczek formula for the ordinary $M/G/1$ queue, while the third term in (21) is the result of the server's vacation periods and, indeed, is independent of the service time.

The LS transform of W is calculated using (15), (16b), and (18). Letting $\alpha = \lambda(1-z)$ and, since $A(z) = \Gamma_V(\lambda(1-z))$ and $B(z) = \Gamma_U(\lambda(1-z))$, one obtains

$$\Gamma_W(\alpha) = \left(\frac{1-\lambda E(V)}{b_0 + \lambda E(U)}\right)\left(\frac{1-\Gamma_U(\alpha) + b_0\alpha/\lambda}{\Gamma_V(\alpha) - (1-\alpha/\lambda)}\right)\Gamma_V(\alpha). \tag{22}$$

3. Model 2

We consider now a variation of the model studied previously. In this variation the underlying structure is, as before, an $M/G/1$ queue with server's vacations. However, in this case if the server finds the system empty at the end of a vacation, he immediately takes another vacation, and continues in this manner until he finds at least one waiting unit upon return from vacation.

We distinguish between a single vacation, U, having distribution $F(\cdot)$, and a 'vacation period,' T_R, defined as the time elapsed between the moment the server leaves the main system (after a service completion) and the moment he starts serving again.

Again, T_s is the occupation period in which, contrary to the first model, the server is always busy. In the sequel we use the same notations as for the first model.

A vacation period, T_R, is the sum of geometric number (with parameter $b_0 = \Gamma_U(\lambda)$) of independent vacations U. Thus

$$\Gamma_{T_R}(z) = \sum_{k=0}^\infty [\Gamma_U(z)]^{k+1} b_0^k (1-b_0) = \frac{(1-b_0)\Gamma_U(z)}{1-b_0\Gamma_U(z)}. \tag{23}$$

An occupation period is composed of j ordinary $M/G/1$ busy periods if the server finds j units waiting for service upon return from vacation. Similar to Model 1 we get

(24) $$\Gamma_{T_S}(z) = (\Gamma_U[\lambda - \lambda\Gamma_{T_1}(z)] - b_0)/(1 - b_0).$$

The respective means of T_R and T_S are

(25) $$E(T_R) = E(U)/(1-b_0)$$

which is intuitively clear because of the geometric distribution of the number of vacations in a vacation period, and

(26) $$E(T_S) = \frac{\lambda E(U)E(T_1)}{1 - b_0} = \left(\frac{\lambda E(U)}{1 - b_0}\right)\left(\frac{E(V)}{1 - \lambda E(V)}\right).$$

The expected cycle time is given by

(27) $$E(T) = E(T_R) + E(T_S) = \frac{E(U)}{(1 - b_0)(1 - \lambda E(V))},$$

hence, the fraction of time that the server is vacationing is

(28) $$P_{0\bullet} = E(T_R)/E(T) = 1 - \lambda E(V)$$

which is the proportion of time the server is idle in an ordinary $M/G/1$ queue. However, the occupation period is longer than an ordinary busy period in an $M/G/1$ queue since the former starts with $j \geqq 1$ customers while the latter always starts with a single customer. This is also seen from (26) since $\lambda E(U) > 1 - b_0$.

In order to be able to derive the LST of the waiting time, W, and to calculate the mean number of customers in the system we define an embedded Markov chain with a state space identical to the one introduced in § 2.2 except for the (0, 0) state which does not exist in this model. We have

(29) $$\begin{aligned} (i_{n+1}, j_{n+1}) &= (1, j_n + \xi - 1), & j_n &\geqq 1, \\ &= (0, N^*), & (i_n, j_n) &= (1, 0), \end{aligned}$$

where N^* is the number of customers present at the end of a vacation period. Clearly, $P(N^* = j) = b_j/(1 - b_0)$, $j = 1, 2, \ldots$. The limiting probabilities $[\pi_{i,j}]$, satisfy the following equations

(30a) $$\pi_{0,j} = \frac{b_j}{1 - b_0}\pi_{1,0}, \qquad j = 1, 2, \ldots,$$

(30b) $$\pi_{1,j} = \sum_{k=1}^{j+1}\pi_{\bullet k}a_{j-k+1}, \qquad j = 0, 1, 2, \ldots.$$

Using standard procedures we arrive at

(31) $$\pi_0(z) = \frac{B(z) - b_0}{1 - b_0}\pi_{1,0},$$

(31) $$\pi_1(z) = \frac{A(z)(B(z) - 1)/(1 - b_0)}{z - A(z)}\pi_{1,0}.$$

Again, it is seen that the condition for the existence of positive limiting probabilities is the nonexistence of a solution for $A(z) = z$ in the interval (0, 1).

Now,

(33) $$\pi_1(1) = \frac{\lambda E(U)}{(1 - b_0)(1 - \lambda E(V))}\pi_{1,0}.$$

Since $\pi_1(1) + \pi_{1,0} = 1$ we arrive at

$$\pi_{1,0} = (1 - \lambda E(V))[1 - \lambda E(V) + \lambda E(U)/(1 - b_0)]^{-1}. \tag{34}$$

Substituting (34) in (31) and (32) yields the complete expressions for $\pi_0(z)$ and $\pi_1(z)$.

Result (32) as well as the expression for $\pi_{1,0}/\pi_1(1)$, given by (33), were obtained by Cooper ([2], equations (12) and (13), respectively) while observing the system only at epochs of departure.

The mean number of customers in the system is derived in a way similar to the derivation of (21) and is

$$L = \lambda E(V) + \frac{\lambda^2 E(V^2)}{2(1 - \lambda E(V))} + \frac{\lambda E(U^2)}{2E(U)}. \tag{35}$$

Note that the mean waiting time of an arbitrary arrival until the end of a vacation is given by the, so-called, Random Modification and is $E(U^2)/(2E(U))$. Hence the average increase in queue length due to vacations is given by the third term of (35).

Equation (18) holds for this model as well, and hence

$$\Gamma_W(\alpha) = \left(\frac{1 - \lambda E(V)}{\lambda E(U)}\right) \left(\frac{1 - \Gamma_U(\alpha)}{\Gamma_V(\alpha) - (1 - \alpha/\lambda)}\right) \Gamma_V(\alpha). \tag{36}$$

Representation as a Limiting Case of a Nonpreemptive

The average waiting time $E(W_q) = L/\lambda - E(V)$ may be derived as a limiting case of the nonpreemptive priority model developed by Cobham [1]. The queue in the main system is considered to have a higher priority over a dummy saturated lower-priority queue. Whenever the server completes serving all higher-priority customers he turns to the lower-priority queue where each single vacation is considered as a service time. However, Cobham's results are true for an unsaturated system. His well-known formula for the waiting time of an arbitrary arrival in the higher-priority queue is

$$E(W_w) = \frac{\lambda_1 E(V_1^2) + \lambda_2 E(V_2^2)}{2(1 - \lambda_1 E(V_1))} \tag{37}$$

where λ_i and $E(V_i)$, $i = 1, 2$, are the mean arrival rates and service times, respectively.

Thus, when $\lambda_1 E(V_1) + \lambda_2 E(V_2) \to 1$,

$$E(W_1) \to \frac{\lambda_1 E(V_1^2)}{2(1 - \lambda_1 E(V_1))} + \frac{E(V_2^2)}{2E(V_2)}.$$

Letting $\lambda_1 = \lambda$; $E(V_1) = E(V)$, $E(V_2) = E(U)$ we have $\lim E(W_1) = E(W_q)$, i.e.,

$$E(W_q) = \frac{\lambda E(V^2)}{2(1 - \lambda E(V))} + \frac{E(U^2)}{2E(U)} \tag{38}$$

which may be obtained directly from (35).

4. Comparison Between the Models and Optimization

The two models analyzed previously differ from each other only by their vacation policy. While in the second model the entire idle time of the server is utilized, only part of it is being used in the first one. This fact is seen quantitatively if we compare the mean queue lengths and the fraction of time the server spends vacationing in each of the two models. Considering (21) and (35) and denoting by L'_i the third terms of L in model $i (i = 1, 2)$ we get $L_1'/L_2' = \lambda E(U)/(b_0 + \lambda E(U)) \equiv f < 1$, i.e., for a given arrival rate λ and vacation distribution $F(\cdot)$ the mean queue in Model 1 is always smaller than its correspondence in Model 2. A similar result is true for the fraction of time the server spends on vacation as is readily seen from (8) and (28). This is why the occupation periods and the customers' waiting times are longer in Model 2.

We suppose now that a vacation period's set-up cost is K, the waiting cost of a customer is c per unit time, and the reward for the work done during the vacation is r per unit time. The revenue per unit time is given by

(39) $$R = rP_{0\bullet} - K/E(T) - cL'$$

where, for each of the two models, L' is the excess queue due to vacations. Denote by R_i the revenue for model i $(i = 1, 2)$ and let $\rho = \lambda E(V)$, then

(40a) $$R_1 = r(1-\rho)f - K\,\frac{(1-\rho)}{E(U)}\,f - c\,\frac{\lambda E(U^2)}{2E(U)}\,f,$$

(40b) $$R_2 = r(1-\rho) - K\,\frac{(1-\rho)}{E(U)}\,(1-b_0) - c\,\frac{\lambda E(U^2)}{2E(U)},$$

from which

(41) $$R_1 = \left(R_2 - K\,\frac{(1-\rho)}{E(U)}\,b_0\right)f\,.$$

From (41) it follows that if $R_1 > 0$ then, since $f < 1$, $R_2 > R_1$. Moreover, $R_2 > 0$ implies that $R_2 > R_1$. That is, if, for a given vacation distribution $F(\cdot)$, it is profitable to operate a system that utilizes the idle time of the server, then it is always better to operate it under the policy of Model 2. However, there may be cases where $R_2 \leqq 0$. Clearly, $R_2 \leqq 0 \Rightarrow R_1 < 0$. In such cases it is not at all clear that Model 2 is superior to using (41), it follows that $R_2 < (=)\, R_1$ if and only if $R_2 < (=) - \lambda(1-\rho)K$. That is, external considerations, we are obliged to operate the system using vacations.) Again, using (41), it follows that $R_2 < (=)\, R_1$ if and only if $R_2 < (=) - \lambda(1-\rho)K$. That is, for a fixed distribution $F(\cdot)$ Model 2 is superior to Model 1 if and only if

(42) $$R_2 > -\lambda(1-\rho)K\,.$$

We now turn to find the optimal vacation lengths. From (40) it is readily seen that R_i is dependent on $F(\cdot)$ through $E(U)$, $E(U^2)$ and b_0. Thus, it is impossible to find the optimal R_i explicitly. Nevertheless, we will consider the exponential and deterministic distributions as special cases.

We start with Model 1. Let $E(U) = y$. Hence, for the exponential case, $E(U^2) = 2y^2$ and $b_0 = 1/(1 + \lambda y)$. Substituting in (40a) yields

(43) $$R_1(y) = \frac{\lambda(\lambda y + 1)}{\lambda^2 y^2 + \lambda y + 1}\,[(ry - K)(1-\rho) - c\lambda y^2]\,.$$

To find the optimal vacation mean we differentiate with respect to y, equate to zero and obtain a fourth-degree polynomial in y from which the optimal value, y^*, may be calculated numerically. We just show that $R_1(y)$ assumes its maximum in the interval $(0, \infty)$. The function $R_1(y)$ satisfies

(44a) $$R'_1(0) = \lambda r(1-\rho) > 0\,,$$

(44b) $$R_1(y) \to_{y\to 0} -\lambda(1-\rho)K,$$

(44c) $$R_1(y) \to_{y\to\infty} -\infty.$$

That is, the revenue function is finite at the origin, increasing at the neighborhood of zero and goes to $-\infty$ as $y \to \infty$. Hence, there must exist at least one point $y^* \in (0, \infty)$ for which $R_1(y^*)$ is maximum.

For the deterministic case (again, $E(U) = y$)

(45) $$R_1(y) = \frac{\lambda}{e^{-\lambda y} + \lambda y}\left[(ry - K)(1-\rho) - \tfrac{1}{2}c\lambda y^2\right].$$

It is easy to verify that equations (44)(a), (b), (c) hold as well and the consequences are the same.

For Model 2, as is intuitively clear, the profit from the vacation, $r(1-\rho)$, is independent of the vacation length, and therefore, the objective is to minimize the cost caused by the vacation. For example, if no set-up cost is incurred ($K = 0$) then R_2 is dependent on the ratio $E(U^2)/E(U)$. For many distributions this ratio increases with $E(U)$ (e.g., deterministic and exponential) and we wish to have the smallest vacation possible. For the normal distribution with fixed varience σ^2 this ratio is a convex function of $E(U)$ and has a minimum at $E(U) = \sigma$. When $K \neq 0$ one would look for an optimal vacation length. Considering the *exponential* case

(46) $$R_2(y) = r(1-\rho) - c\lambda y - K\lambda(1-\rho)/(1+\lambda y).$$

Since $-(1+\lambda y)^{-1}$ is concave for $y > 0$, then so is $R_2(y)$. Thus, if $R_2(y)$ has an extremum in $(0, \infty)$ this is a maximum point. If an extremum does not exist then the maximum is at zero, since $R_2(y) \to -\infty$ as $y \to \infty$ and $R_2(y) \to (r - \lambda K)(1-\rho)$ as $y \to 0$. To derive the condition for $y^* > 0$ we let $R'_2(y) = 0$ which yields

(47) $$c\lambda^2 y^2 + 2c\lambda y + c - \lambda(1-\rho)K = 0.$$

The solution of (47) is

(48) $$y = \lambda^{-1}\left[\pm(\lambda(1-\rho)K/c)^{1/2} - 1\right].$$

The condition for $y^* > 0$ is

(49) $$\lambda(1-\rho)K > c.$$

Thus, the optimal vacation mean is

(50) $$\begin{aligned} y^* &= \lambda^{-1}[\lambda(1-\rho)K/c)^{1/2} - 1] && \text{if } \lambda(1-\rho)K > c, \\ &= 0 && \text{otherwise.} \end{aligned}$$

For a *deterministic* vacation we get

(51) $$R_2(y) = r(1-\rho) - \tfrac{1}{2}c\lambda y - K(1-\rho)(1-e^{-\lambda y})/y.$$

One can show that for $y > 0$, $(1-e^{-\lambda y})/y$ is strictly convex and hence $R_2(y)$ is strictly concave. Thus, if there is an extremum in $(0, \infty)$ it is a unique maximum. Since $R_2(y)$ has a continuous derivative there is a maximum in $(0, \infty)$ if and only if $R_2'(0) > 0$. Using l'Hôpital's rule it is readily derived that

$$\lim_{y\to 0} R_2'(y) = -\tfrac{1}{2}c\lambda + \lambda^2(1-\rho)K/2,$$

i.e., $R_2'(0) > 0 \Leftrightarrow \lambda(1-\rho)K > c$, which is the same condition as for the exponential case. Also, if $R_2'(0) \leqq 0$, it follows, using the same arguments, that $y^* = 0$, i.e., 'no vacations' is the best strategy.

For the special case where $K = 0$ (42) simply implies that Model 2 is superior to Model 1 if and only if $R_2 > 0$, and if $F(\cdot)$ is exponential (deterministic) then by (46) (by (51)) $R_2(y) < 0$ iff $y > r(1 - \rho)/c\lambda$ ($y > 2r(1 - \rho)/c\lambda$).

References

1. Cobham, A., "Priority Assignment in Waiting Line Problems," *Operations Research,* Vol. 2 (1954), pp. 70–76.
2. Cooper, R.B., "Queues Served in Cyclic Order: Waiting Times," *Bell Systems Tech. J.,* Vol. 49 (1970), pp. 399–413.
3. Heyman, D. P., "Optimal Operating Policies for $M/G/1$ Queueing Systems," *Operations Research,* Vol. 16 (1968), pp. 362–382.
4. Karlin, S., *A First Course in Stochastic Processes,* Academic Press, New York, 1966.
5. Yadin, M. and Naor, P., "Queueing Systems with a Removable Service Station," *Operational Research Quart.,* Vol. 14 (1963), pp. 393–405.