Mathematical Problems in Engineering
Volume 3, pp. 243–253
Reprints available directly from the publisher
Photocopying permitted by license only

M/G/1 Vacation Model with Limited Service Discipline and Hybrid Switching-on Policy

ALEXANDER V. BABITSKY

Department of Applied Mathematics, Belarusian State University, Minsk, Belarus

(Received April 1994; Revised February 1996; In final form February 1996)

The author studies an M/G/1 queueing system with multiple vacations. The server is turned off in accordance with the K-limited discipline, and is turned on in accordance with the T-N-hybrid policy. This is to say that the server will begin a vacation from the system if either the queue is empty or K customers were served during a busy period. The server idles until it finds at least N waiting units upon return from a vacation.

Formulas for the distribution generating function and some characteristics of the queueing process are derived. An optimization problem is discussed.

Keywords: Queueing process; vacation model; limited discipline; embedded Markov chain; optimal control

1 INTRODUCTION

Today, queueing systems with vacations are a popular object of research. These models are applied to the modelling of local area networks, data communication networks, and flexible manufacturing systems.

The large number and variety of systems which have been studied prompted researchers to classify them. As such, queueing systems are divided into systems with T-, N-, and D-policies of switching on (see [4] and [9]). As to switching off, systems are divided into exhaustive, semi-exhaustive, gated, T-, and K-limited (see [10]). There are papers on systems with the T-N-hybrid policy ([3],[5],[7],[9]) and papers on systems with K-limited service discipline ([2],[6],[8]) which study various modifications of the input stream, service distribution, and so on. The present paper is the first that combines this service discipline and this switching-on policy. The server under the K-limited discipline turns off if either the queue becomes empty or K customers have been served from the beginning of a busy period. Then the server goes on vacation (idle time begins). Each vacation lasts a random time. The queue length is examined after each completion of a vacation. If this length is more than or equal to a preassigned value N, the server turns on; otherwise, the idle period continues.

This paper is a generalization of [3] and [8]. The present system and the system in [3] have the same policy of switching on and results of [3] may be obtained for $K \to \infty$. The system in [8] and the one we study have the service discipline in common. The results of article [8] and those of this paper coincide for $N = 1$.

The embedded Markov chain approach is used to obtain the steady state distribution of queue length. The embedded moments are not only moments of service completion, but also moments of vacation completion.

A sufficient condition of the steady state is given. The Laplace-Stieltjes transform of the sojourn time, the generating function of the distribution, the mean queue length, the mean busy period length, and other characteristics are obtained. The search for optimal values of K and N is discussed.

2 THE MODEL

Customers arrive at the system in accordance with a Poisson process with intensity λ. Service times and vacation times are independent random variables with distribution functions $B(t)$ and $H(t)$, their respective Laplace-Stieltjes transforms (LST) $\beta(s)$ and $h(s)$, and finite moments $b_m = \int_0^\infty t^m dB(t)$, $h_m = \int_0^\infty t^m dH(t)$, $m = 1, 2$. The server begins a vacation in either of the two cases: 1) there are no waiting customers; 2) K units have been served from the beginning of a busy period. The server turnes on if the queue length is equal to or greater than N upon return from a vacation.

Let t_k be the k-th moment of service (vacation) completion, i_{t_k} be the number of customers at moment t_k, and l_{t_k} be the number (from the beginning of a busy period) of a customer whose service was finished at the moment t_k. Obviously, $\{i_{t_k}, l_{t_k}\}$ is a Markov chain.

Suppose that the system is in the steady state. The following theorem provides a sufficient condition for the existence of this state.

THEOREM 1. *The inequality*

$$\lambda < \frac{K}{Kb_1 + h_1} \tag{1}$$

is the sufficient condition for the stationary queue length distribution existence.

Proof. This theorem will be proved with the help of the ergodic Moustafa's theorem. For the studied system it assumes the following form. The irreducible and aperiodic Markov chain is ergodic and the ergodic distribution coincides with the stationary one if $\varepsilon > 0$, natural i_0, $x_i^{(l)} \geq 0$ $(l = 0,\ldots, K, i = 0, 1,\ldots)$ exist, and the following inequalities hold true:

$$\sum_{0 \leq n \leq K, j \geq 0} p_{ij}^{(l,n)} x_j^{(n)} \leq x_i^{(l)} - \varepsilon \text{ for } i > i_0, l = 0,\ldots, K,$$

$$\sum_{0 \leq n \leq K, j \geq 0} p_{ij}^{(l,n)} x_j^{(n)} < +\infty \text{ for } i \leq i_0, l = 0,\ldots, K, \tag{2}$$

where $p_{ij}^{(l,n)}$ is the probability of the transition from a state with i customers after finishing the l-th service (vacation) to a state with j customers after completion of the n-th service (vacation).

The probabilities $p_{ij}^{(l,n)}$ are defined by

$$p_{ij}^{(l,n)} = \begin{cases} a_{j-i+1}, & \text{if } n = 2,\ldots, K, l = n - 1, i \geq 1, j \geq i - 1 \\ & \text{or } n = 1, l = 0, i \geq N, j \geq i - 1; \\ d_{j-i}, & \text{if } n = 0, \text{ and } l = 0, i < N, j \geq i \\ & \qquad \text{or } l = 1,\ldots, K - 1, i = 0, j \geq 0 \\ & \qquad \text{or } l = K, i \geq 0, j \geq i \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $a_n = \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} dB(t)$, $d_n = \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} dH(t)$.

Let $i_0 = N - 1$. Substitute i_0 and (3) into (2) to arrive at

$$\left.\begin{aligned} \sum_{j=i-1}^{\infty} a_{j-i+1} x_j^{(l+1)} \le x_i^{(l)} - \varepsilon, \quad l = 0,\dots, K-1 \\ \sum_{j=i}^{\infty} d_{j-i} x_j^{(0)} \le x_i^{(K)} - \varepsilon, \end{aligned}\right\} \text{for } i > i_0, \tag{4}$$

$$\left.\begin{aligned} \sum_{j=i-1}^{\infty} a_{j-i+1} x_j^{(l)} < +\infty, \quad l = 2,\dots, K\,(i > 0) \\ \sum_{j=i}^{\infty} d_{j-i} x_j^{(0)} < +\infty, \end{aligned}\right\} \text{for } i \le i_0. \tag{5}$$

Let $x_i^{(l)} = i + q_l$, where q_l are some real values, formulas for which will be provided later. By substituting the expression for $x_i^{(l)}$ and (3) into (4) and replacing addition and integration there, expression (4) assumes the following form:

$$\lambda b_1 + q_{l+1} - q_l - 1 \le -\varepsilon, \quad l = 0,\dots, K-1$$

$$\lambda h_1 + q_0 - q_K \le -\varepsilon.$$

By summing up these inequalities we get $K(1 - \lambda b_1) - \lambda h_1 \ge (K + 1)\varepsilon$. Obviously, this is equivalent to (1) and it allows us to assign parameters of Moustafa's theorem:

$$\varepsilon = \frac{K(1 - \lambda b_1) - \lambda h_1}{K + 1}, \qquad q_l = l(1 - \lambda b_1 - \varepsilon), \quad l = 0,\dots, K.$$

It is easy to see that $\varepsilon > 0$ and $q_l \ge 0$ due to (1), and consequently, $x_i^{(l)} \ge 0$. The validity of (4) and (5) is verified by substituting the expressions of ε and $x_i^{(l)}$. Likewise, the conditions of Moustafa's theorem are satisfied and the steady state exists. The theorem is proved.

Suppose that the condition of theorem 1 is satisfied. The steady state distribution will be targeted. Let $\pi_i^{(l)}$ be a stationary probability of a state of the process $\{i_{t_k}, l_{t_k}\}$, where i is the number of customers in the system at the embedded moment t_k (return from a vacation ($l = 0$) or from a service ($l = 1,\dots, K$)) and l is the number (from a busy period beginning) of customers, whose service was finished at the embedded moment.

Obviously,

$$\pi_i^{(l)} > 0 \text{ for } l + i \geq N, \ (l = 1,\dots, K) \text{ and } i \geq \max\{0, N - K\}, \ (l = 0). \tag{6}$$

Suppose, for convenience, that $N \leq K$; though all formulas are correct for $N > K$, too (see Remarks, Section 3).

Define the generating functions as

$$\pi^{(l)}(z) = \sum_{i=0}^{\infty} \pi_i^{(l)} z^i, \quad l = 0,\dots, K, \quad \pi(z) = \sum_{l=1}^{K} \pi^{(l)}(z) \tag{7}$$

THEOREM 2.

$$\pi(z) = \frac{\beta(\lambda - \lambda z)(h(\lambda - \lambda z) - 1)}{(z - \beta(\lambda - \lambda z))(z^K - h(\lambda - \lambda z)\beta^K(\lambda - \lambda z))}$$
$$\times \left[\sum_{j=0}^{N-1} \pi_j^{(0)} z^j (z^K - \beta^K(\lambda - \lambda z)) + \sum_{j=N}^{K-1} \pi_0^{(j)} z^j (z^{K-j} - \beta^{K-j}(\lambda - \lambda z)) \right] \tag{8}$$

where $\pi_j^{(0)}, j = 0, \dots, N - 1, \pi_0^{(j)}, j = N, \dots, K - 1$ is a unique solution of the linear equation system:

$$\begin{cases} \dfrac{d^{l_s}}{dz^{l_s}} \left[\displaystyle\sum_{j=0}^{N-1} \pi_j^{(0)} z^j (z^K - \beta^K(\lambda - \lambda z)) + \right. \\ \quad \left. + \displaystyle\sum_{j=N}^{K-1} \pi_0^{(j)} z^j (z^{K-j} - \beta^{K-j}(\lambda - \lambda z)) \right] \Big|_{z=z_s} = 0, \\ \quad s = 1, \dots, t, \ l_s = 0,\dots, i_s - 1, \\ \displaystyle\sum_{j=0}^{N-1} \pi_j^{(0)} K + \sum_{j=N}^{K-1} \pi_0^{(j)} (K - j) = \dfrac{K - \lambda(h_1 + Kb_1)}{1 - \lambda b_1 + \lambda h_1} \end{cases} . \tag{9}$$

where $z_1,\dots, z_t$ are roots of $z^K - h(\lambda - \lambda z)\beta^K(\lambda - \lambda z) = 0$ in $|z| < 1$ with multiplicities $i_1,\dots, i_t$.

Proof. For the probabilities $\pi_i^{(l)}$ the following Chapman-Kolmogorov equations take place:

$$\pi_i^{(l)} = \sum_{j=K-l+1}^{i+1} a_{i-j+1}\, \pi_j^{(l-1)}, \quad i = N-l, N-l+1,\ldots, l = 1,\ldots, K,$$

$$\pi_i^{(0)} = \sum_{j=0}^{min\{i,N-1\}} \pi_j^{(0)}\, d_{i-j} + \sum_{j=0}^{i} \pi_j^{(K)}\, d_{i-j} + d_i \sum_{j=N}^{K-1} \pi_0^{(j)}, \quad i = 0, 1,\ldots, \tag{10}$$

where a_n and d_n were defined in theorem 1.

By multiplying each equality of (10) by z^i respectively, and summing them up, we have:

$$\pi^{(0)}(z) = h(\lambda - \lambda z)\left(\pi^{(K)}(z) + \sum_{j=N}^{K-1} \pi_0^{(j)} + \sum_{j=0}^{N-1} \pi_j^{(0)} z^j\right),$$

$$\pi^{(l)}(z) = \frac{\beta(\lambda - \lambda z)}{z}\, \pi^{(l-1)}(z), \quad l = 1,\ldots, N, \tag{11}$$

$$\pi^{(l)}(z) = \frac{\beta(\lambda - \lambda z)}{z}\, (\pi^{(l-1)}(z) - \pi_0^{(l-1)}), \quad l = N+1,\ldots, K.$$

Formula (8) is obtained from (11) by using formulas for the generating functions and (7).

Determine the probabilities $\pi_j^{(0)}, j = 0,\ldots, N-1$, $\pi_0^{(j)}, j = N,\ldots, K-1$ which are used in (8). Rewrite the first part of (8):

$$\frac{\beta(\lambda - \lambda z)(h(\lambda - \lambda z) - 1)}{(z - \beta(\lambda - \lambda z))} \cdot \frac{1}{(z^K - h(\lambda - \lambda z)\beta^K(\lambda - \lambda z))}.$$

The first factor of this expression is bounded in the region $|z| \leq 1$. Due to theorem 4.1 in statement A of [1], the denominator of the second factor has K roots (counting multiplicities) in the region $\{|z| \leq 1\}$ under condition $(h(\lambda - \lambda z)\beta^K(\lambda - \lambda z))'|_{z=1} < K$ (this is the assumed condition of steady distribution existence). It is easy to see, that only the simple root $z = 1$ lies on the boundary of the region, and therefore other $K - 1$ roots, say $z_1,\ldots,$

z_t, with multiplicities $i_1, \ldots, i_t$ ($i_1 + \ldots + i_t = K - 1$), lie inside. As long as $\pi(z)$ is regular in the region $\{|z| < 1\}$, then the expression in the square brackets in (8) has the same roots $z_1, \ldots, z_t$ with the corresponding multiplicities. Equate this expression and its derivations $(d / dz, \ldots, d^{i_t - 1} / dz^{i_t - 1})$ to zero in points $z_1, \ldots, z_t$. Therefore, equations in (9) for probabilities $\pi_j^{(0)}$ and $\pi_0^{(j)}$ are given. The last equation in (9) is a normalization condition $\pi(1) + \pi^{(0)}(1) = 1$. The theorem is proved.

3 SOME CHARACTERISTICS

Having determined the generating function $\pi(z)$, we can derive some characteristics of the system.

The LST and the average value of the sojourn time.

Let $V(t)$ and $v(s)$ be the distribution function and LST of the sojourn time (the waiting and the service time). The probability that a customer, who has already been served, leaves i units in the system (that is, i units came into the system during the sojourn time) is equal to the probability that i customers are present in the system at the embedded moment under the condition that the moment is the end of a service. This can be rewritten as

$$\int_0^\infty \frac{(\lambda t)^i}{i!} e^{-\lambda t} dV(t) = \frac{\sum_{l=1}^{K} \pi_i^{(l)}}{\pi(1)}, \quad i = 0, 1, \ldots.$$

By multiplying these equalities by z^i, respectively, and summing them up we get:

$$v(s) = \frac{\pi(1 - s/\lambda)}{\pi(1)}.$$

Consequently, the mean sojourn time, say V_1, is

$$V_1 = -v'(0) = \frac{\pi'(1)}{\lambda \pi(1)}. \tag{12}$$

Values of $\pi(1)$ and $\pi'(1)$ are determined by (8) and (11) as

$$\pi(1) = \frac{\lambda h_1}{1 - \lambda b_1 + \lambda h_1} \text{ and}$$

$$\pi'(1) = \sum_{j=0}^{N-1} \pi_j^{(0)} \left(\tilde{D} I_K + D \tilde{I}_K^j\right) + \sum_{j=N}^{K-1} \pi_0^{(j)} \left(\tilde{D} I_{K-j} + D \tilde{I}_{K-j}^j\right),$$

where

$$D = \frac{\lambda h_1}{(1 - \lambda b_1)A}, \quad A = K - \lambda(h_1 + K b_1),$$

$$\tilde{D} = \frac{\lambda^2(2b_1 h_1 + h_2)}{2(1 - \lambda b_1)A} + \frac{\lambda h_1 C}{(1 - \lambda b_1)A^2} + \frac{\lambda^3 h_1 b_2}{2(1 - \lambda b_1)^2 A},$$

$$C = -\frac{K(K-1)}{2} + \frac{\lambda^2}{2}\left(h_2 + K b_2 + K(K-1) b_1^2 + 2K b_1 h_1\right),$$

$$I_l = l(1 - \lambda b_1), \text{ and } \tilde{I}_l^j = jl(1 - \lambda b_1) + \frac{l}{2}\left(l - 1 - \lambda^2\left(b_2 + (l-1) b_1^2\right)\right).$$

The average number of customers in the system.

Denote this characteristic by L and determine it by Little's formula:

$$L = \lambda V_1 = \frac{\pi'(1)}{\pi(1)}.$$

The mean busy period.

The number of units that have been served between vacation periods is a random variable, say ξ valued $l = N, \ldots, K$ with probabilities $\tilde{\pi}^{(l)}$:

$$\tilde{\pi}^{(l)} = \frac{\pi_0^{(l)}}{\pi^*}, \quad l = N, \ldots, K-1, \quad \tilde{\pi}^{(K)} = \frac{\sum_{i=0}^{\infty} \pi_i^{(K)}}{\pi^*} = \frac{\pi^{(K)}(1)}{\pi^*},$$

where $\pi^* = \sum_{l=N}^{K-1} \tilde{\pi}_0^{(l)} + \tilde{\pi}^{(K)}(1)$ is a probability that an embedded moment is the completion of a busy period. Rewrite the expression for π^* with the aid of (11):

$$\pi^* = \frac{\lambda h_1 \sum_{j=0}^{N-1} \pi_j^{(0)} + (1 - \lambda b_1) \sum_{j=N}^{K-1} \pi_0^{(j)} (K - j)}{K - \lambda(h_1 + K b_1)}.$$

Denote the mean of ξ as L_{bp}, $L_{bp} = \sum_{l=N}^{K} l\tilde{\pi}^{(l)}$. After some elementary algebra it can be rewritten as $L_{bp} = \pi(1) / \pi^*$.

Let T_{bp} be the mean busy period length (the average time from the moment the server finished a vacation to the moment when the server finished the last service before a vacation). Obviously, $T_{bp} = b_1 L_{bp}$. Therefore,

$$T_{bp} = b_1 \frac{\pi(1)}{\pi^*}. \tag{13}$$

The average number of switches from service to vacation per unit time. Denote this characteristic by Q_s. It is easy to see that Q_s is equal to the average number of busy period completions per unit time. As long as λ is the average number of customers served per unit time and $\pi^* / \pi(1)$ is the probability that a current service finishes a busy period, the expression for Q_s can be written as

$$Q_s = \frac{\lambda \pi^*}{\pi(1)}. \tag{14}$$

Remarks. Formulas for generating functions and characteristics can be rewritten for $N > K$ in the following way. The sum $\sum_{j=N}^{K-1} \pi_0^{(j)}$ will be dropped in them not only due to the rule that the sum is equal to zero if the lower limit is greater than the upper one, but also because of the zero value of probabilities $\pi_0^{(j)}$; (see (6)). The lower limit of the sum $\sum_{j=0}^{N-1} \pi_j^{(0)}$ is replaced by $j = N - K$, because $\pi_j^{(0)} = 0, j = 0,\ldots, N - K - 1$ due to (6).

4 THE SEARCH OF OPTIMAL VALUES OF K AND N

The problem consists of determining the optimal values of K and N to minimize the total cost function.

In any optimization problem utilizing queueing theory it is necessary to minimize the sojourn time. Besides, it may be desirable to reduce the number of switches from service to vacation. For systems, which describe databases with dynamic information, it is desirable to impose a limitation to the mean and maximal values of the busy period length, because large intervals of time between reorganizations of data (between vacations in the present system) can destroy the data base. The following problem takes into account all these requirements.

$$c_v V_1(N, K) + c_q Q_s(N, K) \to \min_{N,K} \tag{15}$$

$$T_{bp}(N, K) \leq T_{max} \tag{16}$$

$$K \leq K_{max}$$

where c_v, c_q are linear holding costs; T_{max}, K_{max} are constraints on the busy period length; V_1, Q_s, T_{bp} are defined by (12), (13), (14). Due to the complexity of these formulas it is difficult to obtain explicit expressions for N and K. However, the following remark helps searching for them: optimal values of K and N should satisfy the inequality $N_{opt} \leq K_{opt}$, because for $N > K$ $V(N, K) > V(K, K)$, $Q_s(N, K) = Q_s(K, K)$ and $T_{bp}(N, K) = T_{bp}(K, K)$.

Numerical example. Problem (15) with the limitations (16) was numerically solved for $c_v = 0.05$, $c_q = 10$, $T_{max} = 1.87$, $K_{max} = 8$, $B(t) = 1 - e^{-3t}$, $G(t) = 1 - e^{-t}$, $\lambda = 0.3$. The optimal values K_{opt} and N_{opt} are equal to 8 and 4, respectively.

References

[1] Abolnikov, L. and Dukhovny, A. (1991). Markov chains with transition delta-matrix: ergodicity conditions, invariant probability measures and applications, *JAMSA*, **4**, No. 4, 335–355.

[2] Bocharov, P. P. and Vornaritkh Ch. (1991). On finite capacity queueing system under limited service discipline, *Communication and computing networks as queueing models*. Minsk. 26–27.

[3] Dudin, A. N. (1989). Analysis of model of data bases dynamic reorganizations in computer networks, *Automatic control and Computer sciences*, **3**, 70–75.

[4] Heyman, D. P. (1968). Optimal operating policies for M/G/1 queueing systems, *Oper. Res.*, **16**, 362–382.

[5] Jacob, M. J. and Madhusoodanatan, I. P. (1987). Transient solution for a finite capacity $M|G^{a,b}|1$ queueing system with vacation to the server, *Queueing systems* **2**, 381–386.

[6] Kramer, M. (1989). Stationary distributions in a queueing system with vacation times and limited service, *Queueing systems* **4**, 57–68.

[7] Lee, H. S. and Srinivasan, M. M. (1989). Control policies for the $M^x|G|1$ queueing system, *Management science*, V.35, No. 6, 708–721.

[8] Lee, T. T. M/G/1/N queue with vacation time and limited service discipline, *Performance Evaluation*, **9**, 181–190, 1988/89.

[9] Muh, D. (1993). A bulk queueing system under N-policy with bilevel service delay discipline and start-up time, *JAMSA*, vol. 6, No. 4.

[10] Takagi, H. (1988). A survey of queueing analysis of polling models, *Data communications systems and their performance—Elseveier science publishers*, 263–281.