ANALYSIS OF A CYCLIC-SERVICE TANDEM QUEUE WITH MULTIPLE SERVER VACATIONS AND EXHAUSTIVE SERVICES

Tsuyoshi Katayama
Toyama Prefectural University

Abstract   This paper presents an analysis of a cyclic-service tandem queue ($M/G/1$ type queue) with exhaustive services and multiple server vacations, in which a single server takes repeated vacations until it finds at least one message in the first queue after emptying the queueing system. It is assumed that the vacation time is generally distributed, and switch-over times are zero. This paper analyzes queue-length distributions and message sojourn time distributions. By using the results of the two-stage tandem-queue analysis and the decomposition property in vacation models, some explicit formulas are derived for the mean total sojourn time in the multi-stage tandem queue and the mean waiting time in the first stage. These results are applicable to the performance analysis for message processing in packet switching systems.

1. Introduction

Cyclic-service tandem queueing models attended by a single server may find a number of examples in the performance evaluation of computer operating systems, packet switching systems in telecommunication networks and production systems, Refs. [4, 8, 9, 14]. This paper considers such a cyclic-service tandem queueing model with multiple server vacations, in which the server takes repeated vacations until it find at least one message in the first queue after emptying the queueing system (*multiple vacations* [2, 13]). For example, in packet switching systems, the vacation time can be utilized for secondary priority tasks such as traffic data processing and aperiodic processing for systems maintenance and testing, while after the vacation period, the server returns to the main system to execute the primary message processing with higher priority [3].

There have been some analytical studies on cyclic-service tandem queueing models, the literature of which is given in the reference lists of [4, 5, 10], but the subject is not as well-investigated as polling models [7, 12]. Vacation models have been analyzed for various queueing system, e.g. $M^{[X]}/G/1$, $GI/G/1$, priority queueing systems and so on, an important result for which is the well-known decomposition property [1, 2, 13]. Doshi [2] presented an overview of the state of the art of vacation model analysis, as well as an extensive list of references. However, there are few studies in the literature treating the cyclic-service tandem queueing models with vacations. From such a point of view, we analyze a basic cyclic-service tandem queueing model with multiple server vacations which is a generalized model of the ordinary single-stage ($M/G/1$-type) vacation model. This queueing model is derived from message processing in packet switching systems. The analysis results are also summarized by taking account of the tandem queueing model without vacation times. It is useful to clarify the influence of vacation time on some performance measures in such switching systems.

The rest of this paper is organized as follows: In Section 2, we describe a cyclic-service tandem queueing model with multiple server vacations. In Sections 3 and 4, we derive generating functions of joint queue-length distributions, Laplace-Stieltjes transforms (LSTs)

for waiting times in each stage and the total sojourn time in the tandem queueing system, and formulas on the mean performance measures. In Section 5, we give the explicit expression for the mean total sojourn time in an $N(\geq 2)$-stage tandem queue with vacations. We conclude the paper in Section 6.

2. Cyclic-Service Tandem Queueing Model with Vacations

This section presents a cyclic-service tandem queueing model with server vacations and introduces some notations. The queueing system is composed of two service stages connected in series. The n-th stage, $n = 1, 2$ has a service counter S_n and a waiting room with infinite capacity Q_n. Messages arrive at Q_1 according to a Poisson process with rate λ. Each message requires exactly two services before leaving the queueing system, i.e. after completion of the service (*pre-processing*) in S_1, the message goes to Q_2 to receive the service (*primary processing*) in S_2, and leaves the system after service-completion in S_2. Service time τ_n at each counter $S_n, n = 1, 2$ are random variables with a general distribution function $H_n(t)$, with finite first and second moments h_n and $h_n^{(2)}$. The LST of $H_n(t)$ is denoted by $H_n^*(s), n = 1, 2$.

Messages in Q_1 and Q_2 are served by a single server in accordance with *the exhaustive service*, i.e. two queues are served by the server in cyclic order, $S_1 \to S_2 \to S_1 \to S_2 \cdots$, and the server will continue to serve at $S_n, n = 1, 2$ until Q_n becomes empty. However, if there is no waiting message in Q_1 just after switching over to S_1, the server then begins a vacation. The vacation may be repeated if the server finds that Q_1 is still empty upon his return (called *multiple vacations*). If any, the server continues serving messages in Q_1 in accordance with the exhaustive service. Messages in each queue are served in the order of their arrivals (FIFO). The vacation time V is a random variable with a general distribution function $V(t)$, with finite first and second moments v and $v^{(2)}$. The LST of $V(t)$ is denoted by $V^*(s)$. Service times τ_1 and τ_2 and vacation time V are mutually independent random variables.

For simplicity, the following notations are introduced:

$$\rho_n := \lambda h_n \quad n = 1, 2 \quad \rho := \rho_1 + \rho_2$$
$$\eta := \lambda v \tag{1}$$

and

$$H^*(s) := H_1^*(s) \cdot H_2^*(s)$$
$$h := h_1 + h_2 \qquad h^{(2)} := h_1^{(2)} + h_2^{(2)} + 2h_1 h_2. \tag{2}$$

Server utilization $\rho = \lambda h$ is assumed to be less than unity for stability (see Remark 2.1).

Here, let $q_n(k)$ denote the probability that k messages arrive at Q_1 during a service time $\tau_n, n = 1, 2$ and denote by $Q_n(x)(:= \sum_{k=0}^{\infty} q_n(k)x^k)$ the generating function for $q_n(k)$. Then, we have

$$Q_n(x) = H_n^*\{\lambda(1 - x)\} \quad n = 1, 2. \tag{3.a}$$

Similarly, we denote by $r(k)$ the probability that k messages arrive at Q_1 during a vacation time V, and denote by $R(x)(:= \sum_{k=0}^{\infty} r(k)x^k)$ the generating function for $r(k)$, which yields

$$R(x) = V^*\{\lambda(1 - x)\}. \tag{3.b}$$

Remark 2.1: The present model and the standard $M/G/1$ vacation model with arrival rate λ, and the LSTs of service time distribution and vacation time distribution, $H^*(s)$ and $V^*(s)$, have the same distribution of the number of messages served in a busy period and the same vacation times, i.e. the same starting/teminating epochs for the vacations [2, 13].

□

3. Queueing Analysis

By the imbedded Markov-chain approach, we first derive a set of generating functions for joint queue-length distributions. Let $t_r, r = 1, 2, 3, \cdots$ denote the successive epochs just after departure of messages from S_1 or S_2. Let $L_1(r)$ and $L_2(r)$ denote the number of waiting messages in Q_1 and Q_2 at t_r, respectively, and let $\sigma(r - 0) = n$ denote the server location $S_n (n = 1, 2)$ at the point just before t_r. Then, a vector chain $\{Y_r\}$ defined by

$$Y_r := (\sigma(r - 0), L_1(r), L_2(r))$$

becomes an imbedded Markov chain [4]. Let $\{\pi_n(i, j)\}$ denote the stationary distribution of the Markov chain $\{Y_r\}$, i.e.

$$\pi_n(i, j) := \lim_{r \to \infty} P_r\{Y_r = (n, i, j)\}. \tag{4.a}$$

Further, let us introduce the generating functions,

$$\Pi_n(x, y) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \pi_n(i, j) x^i y^j \quad |x|, |y| \le 1 \quad n = 1, 2. \tag{4.b}$$

Then, considering the events that occur during two successive service completion epochs t_r and $t_{r+1}, r = 1, 2, 3, \cdots$, we have the following relationships:

$$Pr\{Y_{r+1} = (1, i, 1)\} = Pr\{Y_r = (2, 0, 0)\} \sum_{m=1}^{\infty} r(0)^{m-1} \sum_{k=0}^{i} r(i - k + 1) q_1(k)$$

$$+ \sum_{k=0}^{i} Pr\{Y_r = (2, i - k + 1, 0)\} q_1(k) \quad \text{for } i \ge 0 \tag{5.a}$$

$$Pr\{Y_{r+1} = (1, i, j)\} = \sum_{k=0}^{i} Pr\{Y_r = (1, i - k + 1, j - 1)\} q_1(k) \quad \text{for } i \ge 0, j \ge 2 \tag{5.b}$$

$$Pr\{Y_{r+1} = (2, i, j)\} = Pr\{Y_r = (1, 0, j + 1)\} q_2(i)$$

$$+ \sum_{k=0}^{i} Pr\{Y_r = (2, i - k, j + 1)\} q_2(k) \quad \text{for } i, j \ge 0 \tag{5.c}$$

where it should be noted that $Pr\{Y_r = (1, i, 0)\} = 0$ for any $i \ge 0$. From (4.a), (4.b) and (5.a)~(5.c), we get the following functional relationships for $\Pi_n(x, y), n = 1, 2$:

$$\Pi_1(x, y) = \Pi_2(0, 0) \sum_{m=1}^{\infty} r(0)^{m-1} (R(x) - R(0)) Q_1(x) \frac{y}{x}$$

$$+ \{\Pi_2(x, 0) - \Pi_2(0, 0)\} Q_1(x) \frac{y}{x}$$

$$+ \{\Pi_1(x, y) - \Pi_1(0, y)\} Q_1(x) \frac{y}{x} \tag{6.a}$$

and

$$\Pi_2(x,y) = \{\Pi_2(x,y) - \Pi_2(x,0)\}Q_2(x)\frac{1}{y}$$
$$+ \Pi_1(0,y)Q_2(x)\frac{1}{y}. \tag{6.b}$$

By using (3.a), (3.b), Eqs. (6.a) and (6.b) can be rewritten as

$$\Pi_1(x,y) = \frac{yH_1^*\{\lambda(1-x)\}}{x - yH_1^*\{\lambda(1-x)\}}\left[\psi(x) - \varphi(y) - \frac{\pi_2(0,0)}{1 - r(0)}[1 - V^*\{\lambda(1-x)\}]\right] \tag{7.a}$$

$$\Pi_2(x,y) = \frac{H_2^*\{\lambda(1-x)\}}{y - H_2^*\{\lambda(1-x)\}}\left[\varphi(y) - \psi(x)\right] \tag{7.b}$$

where

$$\psi(x) := \Pi_2(x,0) \quad \varphi(y) := \Pi_1(0,y). \tag{7.c}$$

It is necessary to determine the unknown probability $\pi_2(0,0)$ and unknown functions $\psi(x)$ and $\varphi(y)$ on the right-hand sides of (7.a) and (7.b).

(1) Determination of $\pi_2(0,0)$

The unknown probability $\pi_2(0,0)$ should be determined by the normalization condition,

$$\Pi_1(1,1) + \Pi_2(1,1) = 1. \tag{8}$$

From (8) and the relationships derived by letting $x, y \to 1$ in (7.a) and (7.b), we get simultaneous linear equations with five unknowns [4], the solution of which is given by:

$$\frac{\pi_2(0,0)}{1 - r(0)} = \frac{1 - \rho}{2\eta} \tag{9.a}$$

and

$$\Pi_n(1,1) = \frac{1}{2} \quad n = 1, 2$$
$$\Pi_{1y}'(0,1) = \varphi'(1) = \frac{1}{2}, \quad \Pi_{2x}'(1,0) = \psi'(1) = \frac{1}{2}\rho_2 \tag{9.b}$$

where

$$\Pi_{ny}'(a,b) := \left[\frac{\partial}{\partial y}\Pi_n(x,y)\right]_{x=a, y=b}. \tag{9.c}$$

(We use the same notation below.)

A probabilistic interpretation for (9.a) is given by using the decomposition theorem on workload in the vacation system [1,2]: The probability p_0 that there is no message at an arbitrary time in the ordinary $M/G/1$ vacation system with arrival rate λ and the LSTs of service time and vacation time, $H^*(s)$ and $V^*(s)$, respectively, is given by

$$p_0 = (1 - \rho) \cdot \frac{1/\lambda}{\displaystyle\sum_{m=1}^{\infty} mv \cdot \{r(0)^{m-1}(1 - r(0))\}} \tag{10.a}$$
$$= (1 - \rho) \cdot \frac{1/\lambda}{v/(1 - r(0))}.$$

The first factor on the right-hand side of (10.a), $1 - \rho$, stands for the probability that the workload is zero at an arbitrary point in the $M/G/1$ queue without vacations. The second factor represents the probability that the workload is zero at an arbitrary point in a vacation period in the $M/G/1$ vacation system, where the numerator, $1/\lambda$, represents the mean time interval that there is no message in the system and the denominator, $v/(1 - r(0))$, represents the mean vacation period. Further, it is derived from the property of the $M/G/1$ type queue (Burke's theorem and *PASTA* property [13]) that $p_0 = \pi_0$, where π_0 denotes the probability that there is no message in the system at the message departure epoch. We also get

$$\pi_0 = \frac{\pi_2(0,0)}{\Pi_2(1,1)}. \tag{10.b}$$

Hence from (9.b), (10.a) and (10.b), we get (9.a).

(2) Determination of unknown functions $\psi(x)$ and $\varphi(y)$
   On the right-hand side of (7.a),

$$x - yH_1^*\{\lambda(1-x)\} = 0 \tag{11}$$

has exactly one root, say $x = \delta(y)$, in the unit circle $\mid x \mid \leq 1$ under the condition: $\rho_1 \leq 1$ and $\mid y \mid \leq 1$. An explicit expression for $\delta(y)$ is given by (17.b) known as Takács' lemma [11]. From the regularity of $\Pi_1(x, y)$ and $\Pi_2(x, y)$, the numerators on the right-hand sides of (7.a) and (7.b) should be equal to zero for $x = \delta(y)$ and $y = H_2^*\{\lambda(1-x)\}$, respectively. Thus, we obtain a simultaneous functional equation for two unknowns:

$$\psi(\delta(y)) - \varphi(y) = \frac{1-\rho}{2\eta}[1 - V^*\{\lambda(1-\delta(y))\}] \tag{12.a}$$

$$\varphi[H_2^*\{\lambda(1-x)\}] - \psi(x) = 0. \tag{12.b}$$

Here, eliminating $\psi(\delta(y))$ from (12.a) and (12.b) after setting $x = \delta(y)$ in (12.b), we obtain a non-homogeneous linear functional equation for $\varphi(y)$,

$$\varphi[f(y)] - \varphi(y) = g(y) \tag{13.a}$$

where

$$f(y) := H_2^*\{\lambda(1-\delta(y))\}$$
$$g(y) := \frac{1-\rho}{2\eta}[1 - V^*\{\lambda(1-\delta(y))\}]. \tag{13.b}$$

Using an iterative scheme, e.g. [4, 6], $\varphi(y)$ can be determined as follows: First, let us introduce a sequence of $\{y_i\}$ and a function defined by

$$y_0 = y \quad 0 \leq y \leq 1 \tag{14.a}$$
$$y_i = f(y_{i-1}) \qquad i = 1, 2, 3, \cdots$$

and

$$g\{y_i \mid y_0 = y\} := g(y_i) \qquad i = 0, 1, 2, \cdots. \tag{14.b}$$

Then, it follows from (13.a) that

$$\varphi(y_{i+1}) - \varphi(y_i) = g(y_i) \quad i = 0, 1, 2, \cdots. \tag{15.a}$$

Using this relationship repeatedly, we have

$$\varphi(y) = \kappa - \sum_{i=0}^{\infty} g\{y_i \mid y_0 = y\} \tag{15.b}$$

where κ is a constant which is independent of the sequence of $\{y_i\}$. The convergence of the infinite sum in (15.b) is ensured under the condition $\rho < 1$(i.e. $\rho_1 < 1$ and $\pi_2(0,0) \neq 0$ as also shown in Appendix in [6]). From a boundary condition, $\varphi(0) = \Pi_1(0,0) = 0$, the constant κ can be determined as

$$\kappa = \sum_{i=0}^{\infty} g\{y_i \mid y_0 = 0\}. \tag{15.c}$$

Thus, the unknown function $\varphi(y)$ has been determined, and $\psi(x)$ is obtained from (12.b). In this way, the generating functions $\Pi_n(x,y), n = 1,2$ have been completely determined. We thus obtain the following theorem.

Theorem 1. The generating functions $\Pi_n(x,y), n = 1,2$ for the joint queue-length distribution $\{\pi_n(i,j)\}$ are given by (7.a), (7.b) and

$$\varphi(y) = G(0) - G(y) \tag{16.a}$$
$$\psi(x) = G(0) - G[H_2^*\{\lambda(1-x)\}] \tag{16.b}$$

where

$$G(y) := \frac{1-\rho}{2\eta} \sum_{i=0}^{\infty} \left[1 - V^*\{\lambda(1 - \delta(y_i))\} \mid y_0 = y \right] \tag{17.a}$$

$$\delta(y) := \sum_{j=1}^{\infty} \frac{y^j}{j!} \int_0^{\infty} \exp\{-\lambda t\}(\lambda t)^{j-1} dH_1^{(j)}(t) \tag{17.b}$$

where $H_1^{(j)}(t)$ is the j-th iterated convolution of $H_1(t)$ with itself. $\qquad\square$

4. Sojourn Time Analysis

Let $\theta_n, n = 1,2$ denote the sojourn time in the n-th stage, i.e. the time elapsed from the arrival of a message at Q_n to the departure from S_n, and denote by $\Theta_n(t), n = 1,2$ the distribution function of θ_n. Similarly, let $w_n, n = 1,2$ denote the waiting time in the n-th queue and denote by $W_n(t), n = 1,2$ the distribution function of w_n. In addition, let θ denote the total sojourn time, i.e. the time elapsed from the arrival of a message at Q_1 to the departure from S_2, and denote by $\Theta(t)$ the distribution function of θ. Let $\Theta_n^*(s), W_n^*(s), n = 1,2$ and $\Theta^*(s)$ denote the LSTs of $\Theta_n(t), W_n(t), n = 1,2$ and $\Theta(t)$, respectively.

From the usual argument that a message departing from S_1 will leave behind messages in Q_1 which arrived during its sojourn time θ_1 because of the *FIFO* queueing discipline [4, 6], it follows that

$$\Theta_1^*\{\lambda(1-x)\} = \frac{\Pi_1(x,1)}{\Pi_1(1,1)}. \tag{18.a}$$

A similar equation holds for $\Theta^*(s)$:

$$\Theta^*\{\lambda(1-x)\} = \frac{\Pi_2(x,x)}{\Pi_2(1,1)}. \tag{18.b}$$

Thus, we get

$$\Theta_1^*(s) = 2\Pi_1(1 - s/\lambda, 1) \tag{19.a}$$

$$\Theta^*(s) = 2\Pi_2(1 - s/\lambda, 1 - s/\lambda). \tag{19.b}$$

Next, we shall derive $\Theta_2^*(s)$. First, we denote by $\Theta_2(t \mid (1, i, j))$ the conditional probability that the sojourn time of a message that has arrived at Q_2 shall be less than or equal to t, i.e. $\theta_2 \leq t$, under the condition of the state $(1, i, j)$. Further, we denote by $\Theta_2^*(s \mid (1, i, j))$ the LST of $\Theta_2(t \mid (1, i, j))$. From the *FIFO* discipline at S_2, it then follows that

$$\Theta_2^*(s \mid (1, i, j)) = B^*(s)^i H_2^*(s)^j \tag{20}$$

where $B^*(s)$ is the LST of a one-busy period at S_1 and satisfies the equation,

$$B^*(s) = H_1^*[s + \lambda - \lambda B^*(s)]. \tag{21}$$

(From Takács' lemma [11], the explicit expression for $B^*(s)$ is given by (25)). By unconditioning of $\Theta_2^*(s \mid (1, i, j))$, if follows that

$$\Theta_2^*(s) = \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \frac{\pi_1(i, j)}{\Pi_1(1, 1)} \Theta_2^*(s \mid (1, i, j))$$
$$= 2\Pi_1[B^*(s), H_2^*(s)]. \tag{22}$$

Hence, we obtain the following results by using the relation $\Theta_n^*(s) = W_n^*(s)H_n^*(s), n = 1, 2$.

Theorem 2. The LSTs of waiting time distributions $W_n(t), n = 1, 2$ and the total sojourn time distribution $\Theta(t)$ are given by:

$$W_1^*(s) = \frac{2\lambda}{s - \lambda(1 - H_1^*(s))} \left[\varphi(1) - \varphi(H_2^*(s)) + \frac{1 - \rho}{2\eta}(1 - V^*(s)) \right] \tag{23.a}$$

$$W_2^*(s) = \frac{2H_1^*\{\lambda(1 - B^*(s))\}}{B^*(s) - H_2^*(s)H_1^*\{\lambda(1 - B^*(s))\}}$$

$$\cdot \left[\varphi[H_2^*\{\lambda(1 - B^*(s))\}] - \varphi(H_2^*(s)) - \frac{1 - \rho}{2\eta}[1 - V^*\{\lambda(1 - B^*(s))\}] \right] \tag{23.b}$$

$$\Theta^*(s) = \frac{2\lambda H_2^*(s)}{s - \lambda(1 - H_2^*(s))} \left[\varphi(H_2^*(s)) - \varphi(1 - s/\lambda) \right] \tag{24}$$

where $\varphi(y)$ is given by (16.a), and $B^*(s)$ is explicitly expressed by

$$B^*(s) := \sum_{j=1}^{\infty} \frac{1}{j!} \int_0^{\infty} \exp\{-(s + \lambda)t\}(\lambda t)^{j-1} dH_1^{(j)}(t). \tag{25}$$

$\square$

Remark 4.1: Eqs. (23.a) and (24) yield a generalization of the previous results in Refs. [8, 14] to the case with vacation time. The LST $W_2^*(s)$ may be a new result. It is obvious that $W_2^*(s)$ is not of a form of $\Theta^*(s)/(\Theta_1^*(s)H_2^*(s))$ because of the dependence of w_1 and w_2. From Theorem 2, we can derive the covariance of the sojourn time in the first stage and that in the second stage, $Cov(\theta_1, \theta_2)$, by using a relationship $Var(\theta) = Var(\theta_1) + Var(\theta_2) + 2Cov(\theta_1, \theta_2)$.

$\square$

Here, we consider some mean performance measures derived from Theorem 2. We need to calculate $\varphi'(1)$ which is already given by (9.b) and $\varphi''(1)$ to obtain the mean waiting times and so on [4, 6]. The expression for $\varphi''(1)$ which can be also obtained by putting $y = 1$ after differentiating twice both sides of (13.a) with respect to y is given by

$$\varphi''(1) = \frac{1}{1 - \rho_1 + \rho_2}\left[\frac{\rho_1(1 - \rho_1)}{1 - \rho} + \frac{\lambda v^{(2)}}{2v} + \frac{\lambda^2}{2(1 - \rho)}(h_1^{(2)} + h_2^{(2)})\right] \qquad (26)$$

where we have used a relationship, $\delta(y) = yH_1^*\{\lambda(1 - \delta(y))\}$, in order to get $\delta(1) = 1, \delta'(1)$ and $\delta''(1)$. Thus, the results are summarized in the following corollaries.

Corollary 1. The mean waiting time in the n-th queue, $E(w_n), n = 1, 2$, and the mean total sojourn time $E(\theta)$ are given by:

$$E(w_1) = E(w_1)_0 + \frac{1 - \rho_1}{1 - \rho_1 + \rho_2}\frac{v^{(2)}}{2v} \qquad (27.a)$$

$$E(w_2) = E(w_2)_0 + \frac{\rho}{1 - \rho_1 + \rho_2}\frac{v^{(2)}}{2v} \qquad (27.b)$$

$$E(\theta) = E(\theta)_0 + \frac{1 + \rho_2}{1 - \rho_1 + \rho_2}\frac{v^{(2)}}{2v} \qquad (27.c)$$

where $E(w_n)_0, n = 1, 2$ and $E(\theta)_0$ denote the mean waiting times and the mean total sojourn time in the corresponding queueing system without vacations, respectively and are given by:

$$E(w_1)_0 := \frac{\rho_1\rho_2}{(1 - \rho)(1 - \rho_1 + \rho_2)}h_2 + \frac{\lambda(1 - \rho_1)}{2(1 - \rho)(1 - \rho_1 + \rho_2)}(h_1^{(2)} + h_2^{(2)}) \qquad (28.a)$$

$$E(w_2)_0 := \frac{\rho(1 - \rho_1)}{(1 - \rho)(1 - \rho_1 + \rho_2)}h_1 + \frac{\lambda\rho}{2(1 - \rho)(1 - \rho_1 + \rho_2)}(h_1^{(2)} + h_2^{(2)}) \qquad (28.b)$$

$$E(\theta)_0 := \frac{(1 - \rho_1)(1 + \rho_2)}{(1 - \rho)(1 - \rho_1 + \rho_2)}h_1 + h_2 + \frac{\lambda(1 + \rho_2)}{2(1 - \rho)(1 - \rho_1 + \rho_2)}(h_1^{(2)} + h_2^{(2)}). \qquad (28.c)$$

$\square$

Remark 4.2: Substituting $h_2 = h_2^{(2)} = 0$ into (27.a), it can be seen that (27.a) is consistent with the results of the ordinary $M/G/1$ vacation system [13], which is given by (32.a) and (32.b) below. It seems not easy to give a probabilistic interpretation for the second terms on the right-hand sides of (27.a) to (27.c). $\square$

From Corollary 1, we can derive a relationship between $E(w_1)$ and $E(w_2)$ called the pseudo-conservation law [1, 2].

Corollary 2. The mean waiting times in each queue satisfy

$$\{\rho E(w_1) + \rho_2 E(w_2)\}/\rho = \frac{\lambda h^{(2)}}{2(1 - \rho)} + \frac{v^{(2)}}{2v}. \qquad (29)$$

$\square$

It should be noted that the value of $\sum_{n=1}^{2} \rho_n E(w_n)/\rho$ is variant for different service disciplines, like in the case that $v^{(2)}/2v = 0$.

We will present a simpler but more general explanation of (29) by using the well-known decomposition property for the queue with vacations [1, 2, 13]. Here, we use the following notations:

$E(U)$: the expected workload at an arbitrary time in the queuing system with vacations introduced in Section 2.

$E(X)$: the expected workload at an arbitrary time in the queueing system without vacations, which is equivalent to the mean waiting time in the ordinary $M/G/1$ queue with arrival rate λ and the LST of service time distribution, $H^*(s)$.

$E(Y)$: the expected workload at arbitrary time during a vacation period.

$E(Q_n)$: the expected number of messages in $Q_n, n = 1, 2$ excluding the one being served at an arbitrary time.

$E(R)$: the expected residual workload of a message in service at an arbitrary time.

Recalling that two queues are connected in series, $E(U)$ and $E(R)$ are obtained by

$$E(U) = E(Q_1)(h_1 + h_2) + E(Q_2)h_2 + E(R)$$
$$= \rho E(w_1) + \rho_2 E(w_2) + E(R) \tag{30.a}$$

and

$$E(R) = \rho_1 \left(\frac{h_1^{(2)}}{2h_1} + h_2 \right) + \rho_2 \frac{h_2^{(2)}}{2h_2} \tag{30.b}$$

where $h_n^{(2)}/2h_n, n = 1, 2$ represents the mean residual service time at $S_n, n = 1, 2$. The first term on the right-hand side of (30.b), $h_1^{(2)}/2h_1 + h_2$, represents the expected workload of a message in the first stage in service at an arbitrary time [1]. The decomposition property leads to

$$E(U) = E(X) + E(Y) \tag{31.a}$$

where

$$E(X) = \frac{\lambda h^{(2)}}{2(1 - \rho)}. \tag{31.b}$$

From the assumption in Section 2 that the workload is zero in the queueing system with vacations when the server begins a vacation, it follows by using the mean elapsed vacation time, $v^{(2)}/2v$, [1, 13] that

$$E(Y) = \lambda \frac{v^{(2)}}{2v} \cdot (h_1 + h_2) = \rho \frac{v^{(2)}}{2v}. \tag{31.c}$$

Hence, from (30.a) to (31.c), we obtain (29).

Next, we consider a structure of the mean total sojourn time $E(\theta)$. First, let us denote by $E(w)_v$ the mean waiting time in the ordinary $M/G/1$ vacation system with arrival rate λ and the LSTs of service time and vacation time, $H^*(s)$ and $V^*(s)$, respectively. From the results of the $M/G/1$ vacation system analysis, e.g. see [2, 13], we have

$$E(w)_v = E(w)_{M/G/1} + \frac{E(Y)}{\rho} \tag{32.a}$$

where

$$E(w)_{M/G/1} := \frac{\lambda h^{(2)}}{2(1-\rho)}$$

$$E(Y) := \rho \frac{v^{(2)}}{2v}. \tag{32.b}$$

Then, we obtain the following result, where we have used (7.b), (9.b) and (26) in order to derive (33.c) below.

Corollary 3. The mean total sojourn time $E(\theta)$ is decomposed into a summation form as follows:

$$E(\theta) = E(w)_v + h + h_1 E(Q_2^*) \tag{33.a}$$

$$= \frac{\lambda h^{(2)}}{2(1-\rho)} + \frac{v^{(2)}}{2v} + h + h_1 E(Q_2^*) \tag{33.b}$$

where

$$E(Q_2^*) := \frac{1}{\Pi_2(1,1)} \left[\frac{\partial}{\partial y} \Pi_2(1,y) \right]_{y=1}$$

$$= \frac{1}{(1-\rho)(1-\rho_1+\rho_2)} \left[\rho_1(1-\rho_1) + \frac{\lambda^2}{2}\{h_1^{(2)} + h_2^{(2)}\} \right] + \frac{\lambda}{1-\rho_1+\rho_2} \frac{v^{(2)}}{2v}. \tag{33.c}$$

$$\square$$

In the same way as in [5], we will give a probabilistic interpretation for the decomposition of $E(\theta)$ as expressed in the right-hand side of (33.a). Since services in each queue are performed according to the *FIFO* discipline, all messages in front of an arbitrary tagged message, C^*, must first be served before C^* leaves the system. Therefore, the mean total sojourn time $E(\theta)$ of an arbitrary message C^* is given by the following three terms as also shown in (33.a): The first term, $E(w)_v$, represents the mean time interval in which the server needs to serve all messages in the queueing system on arrival of C^* plus the mean residual vacation time on the arrival of C^* since the tandem queueing system is equivalent to the standard $M/G/1$ vacation model as shown in Remark 2. 1. The second term, h, denotes the mean total service time of C^* at S_1 and S_2. The third term, $h_1 E(Q_2^*)$, represents the mean total service time for messages that have arrived at the system behind C^*, to receive their services at S_1 during the total sojourn time of C^*, where $E(Q_2^*)$ denotes the mean queue length in Q_2 when message C^* leaves the system.

Remark 4.3: It is obvious that $E(\theta)$ given by (33.b) agrees with (27.c) derived by the different procedure. Putting $v^{(2)}/2v = 0$ in (33.c), it is also confirmed that $E(Q_2^*)$ agrees with the previous one for the exhaustive service by using Eq. (A1.3) in [5]. It seems not easy to give a probabilistic interpretation for the coefficient of $v^{(2)}/2v$ on the right-hand side of (33.c). $\square$

5. Extension to Multi-Stage Tandem Queue with Vacations

This section considers an $N(\geq 2)$-stage tandem queue with vacations which is an extension of the two-satge tandem queue with exhaustive services described in Section 2. In this section, we use the similar notations to those in Sections 2 and 3 such as $H_n^*(s), h_n, h_n^{(2)}$ and

$\rho_n, n = 1, 2, \cdots, N$ and so on. We also use the same notation by adding abbreviation "N", i.e., $E(w_1)_N, E(\theta)_N$ and $W_1^*(s)_N$.

Corollary 2 can be easily extended to the case of the N-stage tandem queueing system with non-preemptive work-conserving service disciplines in the same way as in previous section. Thus, we summarize only the result in theorem.

Theorem 3. The mean waiting times in the n-th queue, $E(w_n)_N, n = 1, 2, \cdots, N$ satisfy the following relationship:

$$\sum_{n=1}^{N} \Big(\sum_{k=n}^{N} \rho_k \Big) E(w_n)_N / \rho = \frac{\lambda h^{(2)}}{2(1-\rho)} + \frac{v^{(2)}}{2v} \tag{34.a}$$

where

$$\rho := \lambda h \qquad h := \sum_{n=1}^{N} h_n$$

$$h^{(2)} := \sum_{n=1}^{N} h_n^{(2)} + 2 \sum_{i=1}^{N} h_i \sum_{j=i+1}^{N} h_j. \tag{34.b}$$

$\square$

Next, by using the results obtained in previous sections, we derive explicit expressions for the mean waiting time in the first stage, $E(w_1)_N$, and the mean total sojourn time, $E(\theta)_N$. We can apply the same replacement method as used in [5] to this queueing model with vacations. First of all, we introduce a modified tandem queue with two stages defined as follows: Service time $T_1 := \tau_1$ at the first stage has the same distribution as the original tandem queue with N-stages. However, service time T_2 at the second stage is equal to the sum of $\tau_2, \tau_3, \cdots$ and τ_N, i.e. $T_2 := \tau_2 + \tau_3 + \cdots + \tau_N$. Everything else, e.g. arrival process, vacation process, server's switching rule is exactly the same as in the original tandem queue with N stage.

Considering that the modified tandem queue with two stages is equivalent to the tandem queue treated in previous section by substituting $H_2^*(s) \cdot H_3^*(s) \cdots H_N^*(s)$ for $H_2^*(s)$, we thus have the following relationships:

$$W_1^*(s)_N = T^{\#}[W_1^*(s)] \tag{35.a}$$

$$E(w_1)_N = T^{\#}[E(w_1)] \tag{35.b}$$

and

$$E(Q_N^*) = T^{\#}[E(Q_2^*)] \tag{35.c}$$

where the capital letter $T^{\#}$ denotes an operator resulting from substituting $H_2^*(s) \cdot H_3^*(s) \cdots H_N^*(s)$ for $H_2^*(s)$, and which represents the corresponding replacement defined by

$$\Big\{ \sum_{n=2}^{N} h_n \to h_2, \sum_{n=2}^{N} h_n^{(2)} + 2 \sum_{i=2}^{N} h_i \sum_{j=i+1}^{N} h_j \to h_2^{(2)}, \sum_{n=2}^{N} \rho_n \to \rho_2 \Big\}. \tag{36}$$

By operating $T^{\#}$, i.e. by replacing according to (36), for (27.a) in Corollary 1 and (33.b) in Corollary 3, we obtain the following formulae, where we have used the same way as in previous section in order to get (38) below.

Theorem 4. The mean waiting time in the first stage, $E(w_1)_N$, in the N-stage tandem queueing system with exhaustive services is given by

$$E(w_1)_N = \frac{\rho_1 \sum_{n=2}^{N} \rho_n}{(1-\rho)(1-\rho_1 + \sum_{n=2}^{N} \rho_n)} \sum_{n=2}^{N} h_n$$

$$+ \frac{\lambda(1-\rho_1)}{2(1-\rho)(1-\rho_1 + \sum_{n=2}^{N} \rho_n)} \left[\sum_{n=1}^{N} h_n^{(2)} + 2\sum_{i=2}^{N} h_i \sum_{j=i+1}^{N} h_j \right] + \frac{1-\rho_1}{1-\rho_1 + \sum_{n=2}^{N} \rho_n} \frac{v^{(2)}}{2v}. \quad (37)$$

The mean total sojourn time $E(\theta)_N$ in the N-stage tandem queueing system with exhaustive services is decomposed into the summation form as follows:

$$E(\theta)_N = \frac{\lambda h^{(2)}}{2(1-\rho)} + \frac{v^{(2)}}{2v} + h + \sum_{n=1}^{N-1} h_n E(Q_N^*) \quad (38)$$

where

$$E(Q_N^*) := \frac{1}{1-\rho_1 + \sum_{n=2}^{N} \rho_n} \left[\frac{\rho_1(1-\rho_1)}{1-\rho} + \frac{\lambda^2 [\sum_{n=1}^{N} h_n^{(2)} + 2\sum_{i=2}^{N} h_i \sum_{j=i+1}^{N} h_j]}{2(1-\rho)} + \frac{\lambda v^{(2)}}{2v} \right] \quad (39.\text{a})$$

and

$$\rho := \lambda h \quad h := \sum_{n=1}^{N} h_n$$

$$h^{(2)} := \sum_{n=1}^{N} h_n^{(2)} + 2\sum_{i=1}^{N} h_j \sum_{j=i+1}^{N} h_j. \quad (39.\text{b})$$

$\square$

Remark 5.1:

(1) Putting $v^{(2)}/2v = 0$ in (38), it is confirmed that $E(\theta)_N$ agrees with the previous one given by Lemma 2 in [5]. (The derivation of (38) in the case without vacation time is also given in the proof of Lemma 2 in [5]). It should be noted that the relationship (38) holds also for *limited service* and *gated service* as shown in [5]. Hence, to obtain $E(\theta)_N$ for the limited service or the gated service in the first stage, we need at least to derive $E(Q_2^*)$ in the corresponding two-stage tandem queueing system with vacations.

(2) Eqs. (37) and (38) yield a generalization of the previous results, $E(w_N)_{ZS}$ and $E(\theta_N)_{ZS}$ given by Theorem 1 in [5] to the case with vacation time. $\square$

Remark 5.2:

From (19.b), we have

$$E(\theta) = \frac{1/\lambda}{\Pi_2(1,1)} \left[\Pi_{2x}'(1,1) + \Pi_{2y}'(1,1) \right] \quad (40)$$

which leads to (27.c). Here, let $\Pi_N(x, 0, \cdots, 0, y)$ denote the generating function in the N-stage tandem queue corresponding to $\Pi_2(x, y)$ in the two-stage tandem queue. Then, we have

$$\frac{\Pi_N(x, 0, \cdots, 0, 1)}{\Pi_N(1, 0, \cdots, 0, 1)} \neq T^\# \left[\frac{\Pi_2(x, 1)}{\Pi_2(1, 1)} \right] \tag{41.a}$$

and

$$\frac{\Pi_N(1, 0, \cdots, 0, y)}{\Pi_N(1, 0, \cdots, 0, 1)} = T^\# \left[\frac{\Pi_2(1, y)}{\Pi_2(1, 1)} \right] \tag{41.b}$$

Therefore, it follows from (40), (41.a) and (41.b) that

$$E(\theta)_N = \frac{1/\lambda}{\Pi_N(1, 0, \cdots, 0, 1)} \left[\Pi'_{Nx}(1, 0, \cdots, 0, 1) + \Pi'_{Ny}(1, 0, \cdots, 0, 1) \right]$$
$$\neq T^\# [E(\theta)]. \tag{42}$$

This is a reason why we have used the relationship (38), i.e. (33.b) instead of (27.c) in order to get an expression for $E(\theta)_N$. Furthermore, from (20), (21) and (22), we get

$$W_2^*(s)_N = \frac{1}{H_2^*(s)} T^\# \left[\frac{\Pi_1(B^*(s), y)}{\Pi_1(1, 1)} \right]_{y = H_2^*(s)}. \tag{43}$$

Hence, we can not use the simple replacement of (36) for $E(w_2)$ in (27.b) in order to obtain $E(w_2)_N$.   □

6. Concluding Remarks

In this paper, we have analyzed a cyclic-service tandem queueing model with exhaustive services and multiple server vacations. We have derived LSTs for waiting time distributions and sojourn time distributions, and mean performance measures, e.g. waiting times in each stage and total sojourn time in the queueing system. By using the results of two-stage, tandem-queue analysis, some explicit expressions are also derived for the mean total sojourn time in the multi-stage tandem queueing system and the mean waiting time in the first stage. Throughout the paper, we intend to give a probabilistic interpretation for the formulas obtained here by using the decomposition method. However, it remains an open problem to give an intuitive interpretation for $E(w_n), n = 1, 2$ in Corollary 1 and $E(Q_2^*)$ in Corollary 3.

Although the exhaustive service has been assumed throughout the paper, other policies such as gated and limited are often considered in vacation and polling models. It will be challenging to extend the approach of this paper to those queueing systems with limited or gated service disciplines.

Acknowledgment

References

[1] O. J. Boxma: "Workloads and waiting times in single-server systems with multiple customer classes," *Queueing Systems*, **5,** 1, pp. 185–214 (1989).

[2] B. T. Doshi: "Single server queues with vacations," *Stochastic Analysis of Computer and Communication Systems*, Ed. by H. Takagi, pp. 217–265, North-Holland (1990).

[3] F. Ishino, S. Tonami and M. Monden: "High-throughput packet switching system," *Review of the Electrical Communications Laboratories, NTT*, **34,** 6, pp. 689–696 (1986).

[4] T. Katayama: "Analysis of a tandem queueing system with gate attended by a moving server," *Review of the Electrical communications Laboratories, NTT*, **29**, 3–4, pp. 254–267 (1981).

[5] T. Katayama: "Mean sojourn times in a multi-stage tandem queue served by a single server," *J. Oper. Res. Soc. Japan*, **31**, 2, pp. 233–250 (1988).

[6] T. Katayama and Y. Takahashi: "Analysis of a two-class priority queue with Bernoulli schedules," *J. Oper. Res. Soc. Japan*, **35**, 3, pp. 236–249 (1992).

[7] H. Levy and M. Sidi: "Polling Systems: Applications, Modeling, and Optimization," *IEEE Trans. on Commun.*, **38**, 10, pp. 1750–1760 (1990).

[8] S. S. Nair: "Semi-Markov analysis of two queues in series attended by a single server," *Bull. Soc. Math. Belgique*, **22**, pp. 355–367 (1970).

[9] R. T. Nelson: "Dual-resource constrained series systems," *Oper. Res.*, **16**, 2, pp. 324–341 (1977).

[10] M. Sidi, H. Levy and S. W. Fuhrmann: "A queueing network with a single cyclically roving server," *Queueing Systems*, **11**, 1–2, pp. 121–144 (1991).

[11] L. Takács: "*Introduction to the Theory of Queues*," Oxford University Press, New York (1962).

[12] H. Takagi: "Queueing analysis of polling models: An update," *Stochastic Analysis of Computer and Communication Systems*, Ed. by H. Takagi, pp. 267–318, North-Holland (1990).

[13] H. Takagi: "*Queueing Analysis: A Foundation of Performance Evaluation Vol. 1, Vacation and Priority Systems, Part 1*," Elsevier Science Publisher B. V., North-Holland (1991).

[14] M. Taube-Netto: "Two queues in tandem attended by a single server," *Oper. Res.*, **25**, 1, pp. 140–147 (1977).

Tsuyoshi Katayama
Toyama Prefectural University
Toyama 939-03, Japan