Research Article

Sample-Path Analysis of Single-Server Queue with Multiple Vacations

Muhammad El-Taha

Department of Mathematics and Statistics, University of Southern Maine, 96 Falmouth Street, Portland, ME 04104-9300, USA

Correspondence should be addressed to Muhammad El-Taha, eltaha@usm.maine.edu

Received 14 March 2011; Accepted 14 May 2011

Academic Editors: F. Jauberteau and M. S. Noorani

Copyright © 2011 Muhammad El-Taha. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We give a deterministic (sample path) proof of a result that extends the Pollaczek-Khintchine formula for a multiple vacation single-server queueing model. We also give a conservation law for the same system with multiple classes. Our results are completely rigorous and hold under weaker assumptions than those given in the literature. We do not make stochastic assumptions, so the results hold almost surely on every sample path of the stochastic process that describes the system evolution. The article is self contained in that it gives a brief review of necessary background material.

1. Introduction

Consider a single-server queue with multiple vacations general arrival process and general-service times. The server works until all customers in the queue are served then takes a vacation; the server takes a second vacation if when he is back, there are no customers waiting, and so on, until he finds one or more waiting customers at which point he resumes service until all customers, including new arrivals, are served. A vacation is not initiated when there are customers in the system. Under the stochastic assumptions of stationarity and *i.i.d.* inter-arrival and service times, it is known that the mean customer delay in the queue is the sum of two components: mean queue delay and mean vacation time.

This model has applications in communication systems and repair systems among others. Choudhury [1] considers a batch arrival queue with a single vacation where a server takes exactly one vacation after the end of each busy period. This model has applications in manufacturing systems of job-shop type. The paper derives the steady-state distribution of queue length, busy period, and unfinished work using Laplace-Stieltjes transformation approach. Boxma and Groenendijk [2] give (pseudo) conservation laws for multiqueue single

server systems with cyclic service. Their work is based on well-known stochastic decomposition results. Green and Stidham Jr. [3] use a sample-path approach (similar to ours) to address conservation laws and their applications to scheduling and fluid systems. They use the achievable region method to solve stochastic problems utilizing mathematical programming formulations.

References also include Bisdikian [4] who presents a simple method for obtaining conservation laws for single-server queues with batch arrivals and multiple classes using the ASTA (El-Taha [5]) property. Relevant references also include Green and Stidham [3], Boxma and Groenendijk [2], and Choudhury [1]. There is a vast literature on queues with vacations, and conservation laws. The reader may consult Takagi [6] and references therein.

Most analyses give the transform of the waiting time and queue length distribution and then invert these transforms to give explicit expressions. The reader is left with the task of understanding the details of the mathematical analysis. The contribution of this article is to give completely rigorous intuitive proofs that avoid transform methods. We accomplish this by using sample-path analysis. In other words, our proofs are intuitive, completely rigorous and hold pathwise in the sense that they are true on every realization of the stochastic process of interest. Our results turn out to be valid under weaker assumptions than those required in the literature.

This paper is organized as follows: in Section 2, we give preliminary background results that are used in proving main results. In particular, we review $H = \lambda G$ relation and give sample-path proofs for the residual service time and Pollaczek-Khintchine formula for $M/G/1$ queues. In Section 3, we focus on systems with multiple vacations and give the main result which relates virtual delay at any given time with systems actual delay and vacation times. In Section 4, we give a sample path proof of the conservation law and consider a few special cases.

2. Preliminary and Background Results

In this section, we review a few preliminary results that are used in the proof of the main result. Our proof uses the sample path relation $H = \lambda G$ which is a generalization of the well-known Little's formula.

We are given a deterministic sequence of time points, $\{T_k, k \geq 1\}$, with $0 \leq T_k \leq T_{k+1} < \infty$, $k \geq 1$, and we define $N(t) := \max\{k \geq 1 : T_k \leq t\}$, $t \geq 0$, so that $N(t)$ is the number of points in $[0, t]$. We assume that $T_k \rightarrow \infty$ as $k \rightarrow \infty$, so that there are only a finite number of events in any finite time interval ($N(t) < \infty$ for all $t \geq 0$), and we note that $N(t) \rightarrow \infty$ as $t \rightarrow \infty$, since $T_k < \infty$ for all $k \geq 1$. Associated with each time point T_k , there is a function $f_k : [0, \infty) \rightarrow [0, \infty)$. The bivariate sequence $\{(T_k, f_k(\cdot)), k \geq 1\}$ constitutes the basic data, in terms of which the behavior of the system is described. We assume that $f_k(t)$ is Lebesgue integrable on $t \in [0, \infty)$, for each $k \geq 1$.

With $H(t)$ and G_k defined by $H(t) = \sum_{k=1}^{\infty} f_k(t)$ and $G_k = \int_0^{\infty} f_k(t) dt$, respectively, define the following limiting averages, when they exist:

$$\begin{aligned} \lambda &:= \lim_{t \rightarrow \infty} t^{-1} N(t), \\ H &:= \lim_{t \rightarrow \infty} t^{-1} \int_0^t H(s) ds, \\ G &:= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n G_k. \end{aligned} \tag{2.1}$$

Following Stidham Jr. [7], Heyman and Stidham Jr. [8] and El-Taha and Stidham Jr. [9] suppose that the bivariate sequence $\{(T_k, f_k(\cdot)), k \geq 1\}$ satisfies the following condition.

Condition A. There exists a sequence $\{W_k, k \geq 1\}$ such that

- (i) $W_k/T_k \rightarrow 0$ as $k \rightarrow \infty$; and
- (ii) $f_k(t) = 0$ for $t \notin [T_k, T_k + W_k)$.

In economic terms, Condition A says that all the cost associated with the k th point (e.g., the k th customer) is incurred in a finite time interval beginning at the point (e.g., the arrival of the customer), and that the lengths of these intervals cannot grow at the same rate as the points themselves, as $k \rightarrow \infty$. This is a stronger-than-necessary condition for $H = \lambda G$, but it is satisfied in most applications to queueing systems, in which the time points T_k and $T_k + W_k$ correspond to customer arrivals and departures, respectively, and it is natural to assume that customers can only incur cost while they are physically present in the system. The following theorem is given by El-Taha and Stidham Jr. [9, Chapter 6].

Theorem 2.1. Suppose $t^{-1}N(t) \rightarrow \lambda$ as $t \rightarrow \infty$, where $0 \leq \lambda \leq \infty$, and Condition A holds. Then if $n^{-1} \sum_{k=1}^n G_k \rightarrow G$ as $n \rightarrow \infty$, where $0 \leq G \leq \infty$, then $t^{-1} \int_0^t H(s)ds \rightarrow H$ as $t \rightarrow \infty$, and $H = \lambda G$, provided λG is well defined.

We next consider two applications of $H = \lambda G$. The first deals with residual service times, and the second application is the well-known Pollaczek-Khintchine formula for $M/G/1$ queues.

Residual Service Times

Let T_k with $T_0 = 0, T_k \leq T_{k+1}, k = 0, \dots$ be a deterministic point process, and $A_k = T_{k+1} - T_k$ be the k th interevent. Let $R(s)$ be the residual time, that is, time until next event, that is,

$$R(s) = \sum_{k=1}^{\infty} (T_{k+1} - s) \mathbf{1}_{\{T_k \leq s < T_{k+1}\}}. \quad (2.2)$$

Define the following limits when they exist:

$$\begin{aligned} EA &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n A_k; \\ EA^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n A_k^2; \\ R &= \lim_{t \rightarrow \infty} t^{-1} \int_0^t R(s)ds. \end{aligned} \quad (2.3)$$

We interpret EA as the asymptotic average time between events, EA^2 as the asymptotic second moment of the time between events, and R as the asymptotic (long-run) time-average residual time of $R(s)$. In a queueing setting, it is sometimes useful to think of A_k as the service

time of k th arrival, and $R(s)$ as the residual service time of the customer in service at time s , but we shall see other examples. Now, we state the following preliminary result.

Lemma 2.2. *The asymptotic average residual time is*

$$R = \frac{EA^2}{2EA}. \quad (2.4)$$

Proof. This proof uses $H = \lambda G$. Let $f_k(t) = (T_{k+1} - t)\mathbf{1}\{T_k \leq t < T_{k+1}\}$. Now, $H(t) = \sum_{k=1}^{\infty} f_k(t)$ is the residual time at a randomly given time, and H is the asymptotic residual time. Moreover, $G_k = \int_0^{\infty} f_k(t)dt = A_k^2/2$; thus, $G = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n A_k^2/2 = EA^2/2$. Therefore (with $\lambda = 1/EA$)

$$H = \frac{\lambda EA^2}{2} = \frac{EA^2}{2EA}, \quad (2.5)$$

which completes the proof. $\square$

An alternative proof of this result uses the simpler relation, $Y = \lambda X$, of (El-Taha and Stidham Jr. [9]). Now, consider an $M/G/1$ queue with multiple vacations as defined in Section 3. Let S_k and D_k be the service requirement and queueing delay of k th arrival, respectively, and let $f_k(t) = (S_k - (t - T_k - D_k))\mathbf{1}\{T_k + D_k \leq t < T_k + D_k + S_k\}$. Then we immediately see that the mean residual service time of the customer in service is given by $R = \lambda ES^2/2$, noting that $\lambda = 1/EA$ in Lemma 2.2. Note that R may also be written as ($\rho = \lambda ES$)

$$R = \frac{\rho ES^2}{2ES} \quad (2.6)$$

which is the product of *the probability that the server is busy* and *the mean residual service time*. Similarly, thinking of T_k as time instants when a server takes a vacation, the asymptotic mean residual vacation time, V_R , is also given by

$$V_R = \frac{EV^2}{2EV}. \quad (2.7)$$

Pollaczek-Khintchine Formula

Consider any $G/G/1$ -FIFO queue. Let T_k , $k \geq 1$ be the time instant of k th arrival; S_k and D_k be the k th arrival service requirement and delay (time in queue) respectively. Also let $N(t) = \max\{k : T_k \leq t\}$ be the total arrivals during $[0, t]$. Define the following limits when they exist:

$$\lambda = \lim_{t \rightarrow \infty} \frac{N(t)}{t},$$

$$ES = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n S_k,$$

$$\begin{aligned}
ES^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n S_k^2, \\
ED &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n D_k, \\
ED^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n D_k^2, \\
ESD &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n S_k D_k.
\end{aligned} \tag{2.8}$$

Note that we use the suggestive expectation notation even though the quantities are defined pathwise. Let

$$f_k(t) = \begin{cases} S_k, & \text{if } T_k \leq t < T_k + D_k \\ S_k - (t - T_k - D_k), & \text{if } T_k + D_k \leq t < T_k + D_k + S_k \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Now,

$$H(t) = \sum_{k=1}^{\infty} f_k(t) \tag{2.10}$$

is the total amount of work in the system at a randomly given time, and H is the asymptotic average amount of work in the system. Moreover,

$$G_k = \int_0^{\infty} f_k(t) dt = S_k D_k + \frac{S_k^2}{2}, \tag{2.11}$$

so that

$$G = \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \left[S_k D_k + \frac{S_k^2}{2} \right] = ESD + \frac{ES^2}{2}. \tag{2.12}$$

Therefore,

$$H = \lambda \left[ESD + \frac{ES^2}{2} \right], \tag{2.13}$$

The first term of the r.h.s. is the total amount of work in the system associated with customers waiting (excluding the one in service), and the second term is the residual service time of the customer in service.

M/G/1 Model

For an $M/G/1$ -queue, we make the additional assumptions that S and D are independent, and arrivals are Poisson. We also assume that $\rho = \lambda ES < 1$. Then by *PASTA*, we have $ED = H$ which implies $ED = \lambda(ES)(ED) + \lambda ES^2/2$. Therefore,

$$ED = \frac{\lambda ES^2}{2(1-\rho)}. \quad (2.14)$$

In the next section, we extend Pollaczek-Khintchine formula to systems with multiple vacations using $H = \lambda G$ and maintaining a pure sample path approach.

3. Systems with Multiple Vacations

Consider any $G/G/1$ -FIFO stable queue with multiple vacations of length $\{V_k, k \geq 1\}$. The server takes a vacation as soon as the number of customers in the system drops to 0 and continues to take successive vacations until the state N , where N is the number of customers in the system, is ≥ 1 when the server returns from vacation. The process of taking vacations is repeated at the end of the following busy period. A vacation is not initiated when there are customers in the system. Let $\{v_k, k = 1, 2, \dots\}$ be the start time of k th vacation. Let $M(t)$ be the number of vacations up to time t so that $M(t) = \max\{k : v_k \leq t\}$. Define the following limits when they exist:

$$\begin{aligned} \gamma &= \lim_{t \rightarrow \infty} \frac{M(t)}{t}, \\ EV &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n V_k, \\ EV^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n V_k^2. \end{aligned} \quad (3.1)$$

Let $\rho := \lim_{t \rightarrow \infty} t^{-1} \sum_k^{N(t)} S_k$, then the system is stable if $\rho = \lambda ES < 1$. Note that the ρ represents the long-run fraction of time the server is busy and $1 - \rho$ is the long-run fraction of time the server is idle (i.e., on vacation).

Definition 3.1. The two sequences $\{S_k, k \geq 1\}$ and $\{D_k, k \geq 1\}$ are said to be asymptotically pathwise uncorrelated if $ESD = ESED$.

Note that the requirement that $ESD = ESED$ is weaker than the corresponding stochastic assumption that the random variables S and D , representing customer service times and delays, respectively, are independent.

Theorem 3.2. Consider the $G/G/1$ -FIFO multivacation model. Suppose that $ED < \infty$ and $EV < \infty$, and that $\{S_k, k \geq 1\}$ and $\{D_k, k \geq 1\}$ are asymptotically pathwise uncorrelated. Then virtual delay in the system is given by

$$H = \rho ED + \rho \frac{ES^2}{2ES} + (1 - \rho) \frac{EV^2}{2EV}. \quad (3.2)$$

Proof. Let T_k and v_k be time instants of customer arrivals, and vacation starts, respectively. Let $A = \{T_k, k = 1, 2, \dots\}$, $B = \{v_k, k = 1, 2, \dots\}$. Note that $A \cap B = \emptyset$. Let $f_k^1(t)$ and $f_k^2(t)$ be defined such that

$$\begin{aligned} f_k^1(t) &= S_k \mathbf{1}\{T_k \leq t < T_k + D_k\} + (S_k - (t - T_k - D_k)) \mathbf{1}\{T_k + D_k \leq t < T_k + D_k + S_k\}; \\ f_k^2(t) &= (V_k - (t - v_k)) \mathbf{1}\{v_k \leq t < v_k + V_k\}. \end{aligned} \quad (3.3)$$

Note that $f_k^1(t)$ is the same as the $f_k(t)$ given by (2.9). Here, $f_k^1(t)$ is the work remaining to be done for the k th arrival at time t ; $f_k^2(t)$ is the vacation time remaining for the k th vacation at time t . Consider the bivariate sequence $\{T_k, f_k^1(t)\}$ and let $H^1(t) := \sum_{k=1}^{\infty} f_k^1(t)$, so that $H^1(t)$ is the total amount of work in the system at time t . Using $H = \lambda G$, we obtain (see (2.13))

$$H^1 := \lim_{t \rightarrow \infty} t^{-1} \int_0^t H^1(t) dt = \lambda \left[ESD + \frac{ES^2}{2} \right]. \quad (3.4)$$

Similarly, consider the second bivariate sequence $\{v_k, f_k^2(t)\}$ and let $H^2(t) := \sum_{k=1}^{\infty} f_k^2(t)$, $G_k^2 = \int_0^{\infty} f_k^2(t) dt = V_k^2/2$. Here, $H^2(t)$ is the residual vacation time in the system at time t , and G_k^2 is the remaining vacation time encountered by the k th arrival. Applying $H = \lambda G$, we obtain

$$\begin{aligned} G^2 &= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \left(\frac{V_k^2}{2} \right) = \frac{EV^2}{2}, \\ H^2 &:= \lim_{t \rightarrow \infty} t^{-1} \int_0^t H^2(t) dt = \gamma \left[\frac{EV^2}{2} \right], \end{aligned} \quad (3.5)$$

where γ is the unconditional vacation rate. We need to compute γ .

Recall that $M(t) = \max\{k : v_k \leq t\}$ is the number of vacations up to time t . Let $I(t)$ be the status of the server at time t , that is, $I(t) = 0$ if the server is idle at time t and 1 otherwise. Note that the server can be idle (on vacation) while customers are waiting for a vacation to end. It follows that

$$\begin{aligned} EV &:= \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n V_k = \lim_{t \rightarrow \infty} M(t)^{-1} \sum_{k=1}^{M(t)} V_k \\ &= \lim_{t \rightarrow \infty} M(t)^{-1} \int_0^t \mathbf{1}\{I(t) = 0\} dt \\ &= \lim_{t \rightarrow \infty} t M(t)^{-1} (1 - \rho), \end{aligned} \quad (3.6)$$

so that

$$\gamma := \lim_{t \rightarrow \infty} t^{-1} M(t) = \frac{1 - \rho}{EV}. \quad (3.7)$$

Therefore,

$$H^2 = \frac{(1-\rho)EV^2}{2EV}. \quad (3.8)$$

Let $H(t) := H^1(t) + H^2(t)$ be the virtual delay, that is, total amount of work and remaining vacation time in the system at time t , and $H = \lim_{t \rightarrow \infty} t^{-1} \int_0^t H(s)ds$ is the asymptotic (long-run) average amount of work and residual vacation in the system, in other words, H is the asymptotic average virtual delay for a randomly arriving customer. Therefore,

$$H := H^1 + H^2 = \lambda ESD + \frac{\lambda ES^2}{2} + \frac{(1-\rho)EV^2}{2EV}. \quad (3.9)$$

Now using the condition that S_k and D_k are asymptotically pathwise uncorrelated and simplifying, we obtain the result. $\square$

Equation (3.9) gives virtual delay H under conditions weaker than those given in Theorem 3.2. Specifically, (3.9) holds without the assumption that S_k and D_k are asymptotically pathwise uncorrelated.

M/G/1 Queue with Vacations

Suppose that S and D are independent, arrivals are Poisson, and vacations V_k are *i.i.d.* having a general distribution function with mean EV . Also assume that the system is stable, that is, $\rho = \lambda ES < 1$. Then, by PASTA, we have $ED = H$, so Theorem 3.2 implies

$$ED = \rho ED + \frac{\rho ES^2}{2ES} + \frac{(1-\rho)EV^2}{2EV}. \quad (3.10)$$

Therefore,

$$ED = \frac{\lambda ES^2}{2(1-\rho)} + \frac{EV^2}{2EV}. \quad (3.11)$$

For systems with exponential vacations

$$ED = \frac{\lambda ES^2}{2(1-\rho)} + EV. \quad (3.12)$$

Additionally, if service times are exponential with $ES = 1/\mu$, we obtain

$$ED = \frac{\rho}{\mu(1-\rho)} + EV. \quad (3.13)$$

4. Conservation Laws

Now we consider multiclass $G/G/1$ queue with multiple vacations, that is, a single-server system with general (not necessarily *i.i.d.*) interarrival and service times. Conservation laws hold for a wide variety of service disciplines like *FIFO*, *LCFS*, service in random order, and other priority rules. For nonvacation models, any scheduling rule has to be work-conserving, and the server is never idle when there is work in the system, but for vacation models this is not possible due to server vacation so the rule is adapted so that the server is never idle except when on vacation. We define a work-conserving scheduling rule similar to El-Taha and Stidham Jr. [9, pages 204–211], but adapted for a server that takes vacations.

Work-Conserving Scheduling Rules

Consider a bivariate sequence, $\{(T_n, S_n), n = 1, 2, \dots\}$, where T_n and S_n are the arrival instant and work requirement, respectively, of n th arrival. A scheduling rule is said to be *nonanticipative* if the decision about which job to process at time t depends only on $\{(T_n, S_n), n = 1, 2, \dots, N(t)\}$, where $N(t) = \max\{n : T_n \leq t\}$, and possibly on decisions taken before time t . It is *nonidling* if the server does not take a vacation when there is at least one job in the system.

Definition 4.1. A single-server model with nonpreemptive *work-conserving scheduling* (WCS) rules consists of

- (i) a single-server working at unit rate,
- (ii) a set of *non-anticipative* and *non-idling* scheduling rules.

We also assume that the scheduling rules are *service time independent*, and *nonregenerative* in the sense that the decision to schedule a job does not use any information from previous busy cycles. It is immediate from the definition of a work-conserving scheduling system that $U(t) = \sum_{n=1}^{\infty} f_k^1(t)$, the total work in the system at time t is invariant with respect to WCS rules. It follows that the limiting average total work in the system,

$$U = \lim_{t \rightarrow \infty} t^{-1} \int_0^t U(s) ds, \quad (4.1)$$

is also invariant. The residual vacation at time t is defined by $V(t) = \sum_{k=1}^{\infty} f_k^2(t)$, and the asymptotic (long-run) time-average residual vacation is given by

$$V_R = \lim_{t \rightarrow \infty} t^{-1} \int_0^t V(s) ds. \quad (4.2)$$

So V_R is also invariant. Now let H be the limiting average virtual delay in the system as defined in Theorem 3.2. Note that H and U are different due to the fact that the server takes multiple vacations. It is straight forward to see that $H(t) = U(t) + V(t)$, therefore,

$$H = U + V_R, \quad (4.3)$$

so that H is invariant. Therefore, we have the following.

Corollary 4.2. *Consider the G/G/1 multivacation model with WCS rules that are nonpreemptive, regenerative and service time independent. Suppose that $ED < \infty$ and $EV < \infty$, and that $\{S_k, k \geq 1\}$ and $\{D_k, k \geq 1\}$ are asymptotically pathwise uncorrelated. Then virtual delay in the system is given by*

$$H = \rho ED + \rho \frac{ES^2}{2ES} + (1 - \rho) \frac{EV^2}{2EV}. \quad (4.4)$$

Basically, this Corollary extends Theorem 3.2 from FIFO to any nonpreemptive WCS rule. Next, we focus attention on multiclass systems.

Multiclass Multiple Vacation Model

Consider the multi-class GI/GI/1 multiple vacation model as given in Section 3. Suppose that the discipline is work conserving when the server is not on vacation, when the server starts service, it continues until all work is cleared before taking vacation. The scheduling rule is nonanticipative and independent of previous arrival and service times, and within each class $j, j = 1, \dots, J$. That is, assume WCS rules. We give the following conservation law for a multiclass single-server system with multiple vacations.

Theorem 4.3. *Consider the G/G/1 stable multi-class multiple vacation model with WCS rules that are non-preemptive, regenerative, and within each class, they are service time independent. Also, suppose $ED < \infty$ and $EV < \infty$, and that for each $j \{S_{kj}, k \geq 1\}$ and $\{D_{kj}, k \geq 1\}$ are asymptotically pathwise uncorrelated. The vector $(ED_1, \dots, ED_J)$ of expected actual queue delays per customer satisfies the following conservation law:*

$$\sum_{j=1}^J \rho_j ED_j = H - \sum_{j=1}^J \rho_j \frac{ES_j^2}{2ES_j} - (1 - \rho) \frac{EV^2}{2EV}. \quad (4.5)$$

Proof. Let H_j and $U_j, j = 1, \dots, J$ be the class j virtual delay and work in the system respectively, where $H = \sum_{j=1}^J H_j$ and $U = \sum_{j=1}^J U_j$. Now (4.3) becomes

$$\sum_{j=1}^J H_j = \sum_{j=1}^J U_j + V_R, \quad (4.6)$$

Recall that H, U , and V_R are invariant with respect to WCS rules. For each $j, j = 1, \dots, J$, let λ_j, S_j , and D_j be the class j mean arrival rate, service time, and queue delay, respectively. Also let $\rho_j = \lambda_j ES_j$, and suppose that $\sum_{j=1}^J \rho_j < 1$. Then it follows from (3.4)

$$EU_j = \lambda_j \left[ES_j D_j + \frac{ES_j^2}{2} \right]. \quad (4.7)$$

Using the fact that the S_j and D_j are asymptotically pathwise uncorrelated and $\rho_i = \lambda_j ES_j$, we have

$$EU_j = \rho_j ED_j + \frac{\rho_j ES_j^2}{2ES_j}. \quad (4.8)$$

Let $B_j = ES_j^2/2ES_j$. It follows by (4.6) and (4.8) that

$$H = \sum_{j=1}^J \rho_j ED_j + \sum_{j=1}^J \rho_j B_j + \frac{(1-\rho)EV^2}{2EV} \quad (4.9)$$

is invariant over all WCS rules, since we have working conserving scheduling rules. Now let $\lambda = \sum_{j=1}^J \lambda_j$, and let S be a random variable with distribution function $F(t) = \sum_{j=1}^J (\lambda_j / \lambda) P(S_j \leq t)$. Then

$$\sum_{j=1}^J \rho_j B_j = \frac{\lambda ES^2}{2} \quad (4.10)$$

is also invariant over all WCS rules. Thus $\sum_{j=1}^J \rho_j ED_j$ is also invariant, and we have the following conservation law satisfied by the vector $(ED_1, \dots, ED_J)$ of expected queue delays

$$\sum_{j=1}^J \rho_j ED_j = H - \sum_{j=1}^J \rho_j B_j - \frac{(1-\rho)EV^2}{2EV} \quad (4.11)$$

which leads to (4.5), thus proving the theorem. $\square$

In the $M/G/1$ multi-class multiple vacation model, we derive an explicit expression for $H = ED$, using the fact that H is invariant and, therefore, is equal to the delay in the $M/G/1$ -FIFO discipline which is given by (3.11). Hence, the conservation equation (4.5) becomes, by substituting (3.11) and (4.10) in (4.5),

$$\sum_{j=1}^J \rho_j ED_j = \frac{\lambda ES^2}{2(1-\rho)} + \frac{EV^2}{2EV} - \frac{\lambda ES^2}{2} - \frac{(1-\rho)EV^2}{2EV}, \quad (4.12)$$

which simplifies to

$$\sum_{j=1}^J \rho_j ED_j = \frac{\lambda \rho ES^2}{2(1-\rho)} + \frac{\rho EV^2}{2EV}. \quad (4.13)$$

Theorem 4.3 can be used to construct conservation laws for waiting time in the system W_j , number of customers in the system L_j , and number of customers in the queue L_j^q using the fact

that $EW_j = ED_j + ES_j$ and Little's formula. In this paper, we give two primary results. The first result presented in Theorem 3.2 gives a relation between virtual delay and actual delay for single-server multiple vacation model under conditions that are weaker than those given in the literature. This result is extended to multi-class models where a conservation law is given in Theorem 4.3 which is our second result. We use sample path analysis which allows us to give rigorous arguments by focusing on one realization of the stochastic process that describes the system evolution.

References

- [1] G. Choudhury, "A batch arrival queue with a vacation time under single vacation policy," *Computers & Operations Research*, vol. 29, no. 14, pp. 1941–1955, 2002.
- [2] O. J. Boxma and W. P. Groenendijk, "Pseudo-conservation laws in cyclic-service systems," *Journal of Applied Probability*, vol. 24, no. 4, pp. 949–964, 1987.
- [3] T. C. Green and S. Stidham, Jr., "Sample-path conservation laws, with applications to scheduling queues and fluid systems," *Queueing Systems*, vol. 36, no. 1–3, pp. 175–199, 2000.
- [4] C. Bisdikian, "A note on the conservation law for queues with batch arrivals," *IEEE Transactions on Communications*, vol. 41, no. 6, pp. 832–835, 1993.
- [5] M. El-Taha, *PASTA and Related Results*, Encyclopedia of Operations Research and Management Science, John Wiley & Sons, New York, NY, USA, 2010.
- [6] H. Takagi, *Queueing Analysis, Volume 1: Vacation and Priority Systems, Part 1*, North-Holland Publishing, Amsterdam, The Netherlands, 1991.
- [7] S. Stidham, Jr., "A last word on $L = \lambda W$," *Operations Research*, vol. 22, no. 2, pp. 417–421, 1974.
- [8] D. P. Heyman and S. Stidham, Jr., "The relation between customer and time averages in queues," *Operations Research*, vol. 28, no. 4, pp. 983–994, 1980.
- [9] M. El-Taha and S. Stidham, Jr., *Sample-Path Analysis of Queueing Systems*, International Series in Operations Research & Management Science, Kluwer Academic Publishing, Boston, Mass, USA, 1999.