Probability in the Engineering and Informational Sciences, **22**, 2008, 537–555. Printed in the U.S.A.
doi:10.1017/S0269964808000326

ON A QUEUING MODEL WITH SERVICE INTERRUPTIONS

ONNO BOXMA
EURANDOM and Department of Mathematics and Computer Science
Eindhoven University of Technology
5600 MB Eindhoven, The Netherlands
E-mail: boxma@win.tue.nl

MICHEL MANDJES
Korteweg-de Vries Institute for Mathematics
The University of Amsterdam
1018 TV Amsterdam, The Netherlands
and CWI
1090 GB Amsterdam, The Netherlands
E-mail: mmandjes@science.uva.nl

OFFER KELLA
Department of Statistics
The Hebrew University of Jerusalem
Mount Scopus, Jerusalem 91905, Israel
E-mail: Offer.Kella@huji.ac.il

Single-server queues in which the server takes vacations arise naturally as models for a wide range of computer, communication, and production systems. In almost all studies on vacation models, the vacation lengths are assumed to be independent of the arrival, service, workload, and queue length processes. In the present study, we allow the length of a vacation to depend on the length of the previous active period (viz. the period since the previous vacation). Under rather general assumptions regarding the offered work during active periods and vacations, we determine the steady-state workload distribution, both for single and multiple vacations. We conclude by discussing several special cases, including polling models, and relate our findings to results obtained earlier.

© 2008 Cambridge University Press 0269-9648/08 $25.00
https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

1. INTRODUCTION

In traditional queuing models a stream of customers arrives at a service station according to some stochastic process. These customers enter a waiting line when they find the server busy, where they wait until they can be processed. Although it is commonly assumed that the server is always available, an interesting alternative model is a system in which the server takes *vacations*, that is, it alternates between active and inactive modes. During the active times, it is working at full speed, whereas it is not processing any work during the vacations.

Systems with server vacations arise naturally as models for an important class of computer, communication, and production systems. The vacations might for example, represent server breakdowns or periods in which the server processes work generated by another class of customers. Static priority queues and polling models constitute important classes of models for the latter situation. Performance measures like queue lengths, waiting times, and workloads have been intensively studied for many types of vacation models. For extensive surveys on vacation queues, we refer to [5], its update [7], the article [21] that focuses on control aspects, and the book [18]. See also [19,20] for surveys on polling models.

In almost all studies on vacation and polling models, the vacation lengths (or switchover times) are assumed to be independent of other vacations (switchover times) and of the arrival, service, workload, and queue length processes. A notable exception is the work of Harris and Marchal [12], in which the probability of the server taking a vacation after a service completion and the length of the vacation depend on the number of customers present when a service ends. In polling, Altman [1], Groenevelt and Altman [11], and Eliazar [9] studied models in which there is interdependence between switchover times. The latter work also considers a generalization from the common compound Poisson input processes to input according to general Lévy subordinator processes.

1.1. Contribution and Methodology

The main goal of the present article is to extend earlier vacation results to the situation that allows the length of a vacation to depend on the length of the previous *active* period (viz. the period since the previous vacation). It is noted that, in many applications, such a dependence is very natural. For example, in polling models, a relatively long visit of the server to a tagged queue probably leads to a substantial accumulation of work in subsequent queues and hence to a relatively long *intervisit* time of the tagged queue. Indeed, we will see that particular model choices give rise to polling systems. We remark that our findings might be viewed as queuing-theoretic results. We have chosen, however, to put them into the more general framework of Lévy processes, and they can be interpreted as results on storage processes (cf. Prabhu [14]). Let us now give a brief model description.

- During *active periods*, work is generated according to a Lévy process $X_{\mathrm{D}}(\cdot)$ with negative drift, until the workload reaches zero (i.e., the storage reservoir is empty).

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

- From then on, the storage level behaves according to a second Lévy process $X_U(\cdot)$, which we assume to be nondecreasing. As during this period work accumulates in the queue, we may interpret the period as a vacation; it lasts $I + V$, where I is the amount of work generated by some other Lévy process $X_E(\cdot)$ during the length of the preceding active period and V is an independent vacation time.

  The case in which the workload is still zero after $I + V$ has to be treated separately. This case can only occur when $X_U(\cdot)$ is a compound Poisson process. Then there are two possibilities. The first is that the vacation period is extended until work is generated by $X_U(\cdot)$ (the *single vacation* case). The second is when the vacation period is extended by an independent sequence of random variables distributed like V until a positive content level is found at the end of a vacation (the *multiple vacation* case).
- Subsequently, a new active period starts. etc. Note that the multiple-vacation case is equivalent to a model in which the assumption is that a new active period starts immediately after each vacation, because when the content level is zero, the length of the active period is zero.

In this way the stochastic storage process alternatingly experiences active and passive (vacation) periods. Observe that the classical $M/G/1$ queue with single and multiple vacations is a special case of our model; it is obtained by taking $X_E \equiv 0, X_U(\cdot)$, a compound Poisson process, and $X_D(\cdot)$, the same compound Poisson process minus a linear slope with rate one. By taking $X_E(t) = at$, (i.e., the Lévy process is a linear drift process), we could make the length of the interruption period depend linearly on the length of the preceding active period. By taking $X_E(\cdot)$, a compound Poisson process, we could let the interruption periods represent periods in which some other $M/G/1$-type queue is being served (see Section 5).

We concentrate on the derivation of the steady-state workload distribution for the above-described model. We do so by first considering the distribution at the embedded epochs in which the server switches between a vacation and an active period. We express the state of the system at such an embedded epoch in terms of the system at the previous embedded epoch and then find the Laplace-Stieltjes transform of the workload by iteration.

Then we use the result for the embedded epochs to characterize the distribution of the workload at *arbitrary* epochs, relying on martingale techniques and Palm-theoretic arguments. For the classical $M/G/1$ queue with server vacations, Doshi [6] presented a related analysis, specializing to the distributions during active periods and vacations.

1.2. Organization

The article is organized as follows. Section 2 contains a detailed model description and some preliminary results. In Section 3 we determine the distribution of the workload at epochs in which the process switches from *passive* to *active*. In Section 4 this result is exploited to obtain the steady-state workload distribution during active periods, during passive periods, and overall. Section 5 considers special cases and ramifications.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

2. MODEL DESCRIPTION AND SOME PRELIMINARIES

In this section we formally introduce the model of our storage system, which can alternatively be interpreted as a queue with service interruptions. We also present some preliminaries that we frequently use in our analysis.

The dynamic behavior of the storage system consists (alternatingly) of service periods (or active periods) and interruptions (or passive periods, vacations), as follows:

- Suppose there is $z \geq 0$ present in the storage system at the beginning of a service period. The storage level evolves according to a Lévy process $X_{\mathrm{D}}(\cdot)$ until the storage level reaches 0. Let $\varrho_{\mathrm{D}} := -\mathbb{E}[X_{\mathrm{D}}(1)]$ be positive and finite. Throughout, it is assumed that $X_{\mathrm{D}}(\cdot)$ has no negative jumps; as an immediate consequence, level 0 is actually attained, say at time $\tau(z)$:

$$\tau(z) := \inf\big\{t \geq 0 : z + X_{\mathrm{D}}(t) = 0\big\}. \tag{1}$$

  We define the *Laplace exponent* of $X_{\mathrm{D}}(\cdot)$ by $\varphi_{\mathrm{D}}(\alpha) := \log \mathbb{E}[\exp(-\alpha X_{\mathrm{D}}(1))]$ and hence $\mathbb{E}[\exp(-\alpha X_{\mathrm{D}}(t))] = e^{\varphi_{\mathrm{D}}(\alpha)t}$ and $\varrho_{\mathrm{D}} = \varphi'_{\mathrm{D}}(0)$. Also, $\psi_{\mathrm{D}}(\cdot)$ denotes the inverse of $\varphi_{\mathrm{D}}(\cdot)$. It is well known that for any Lévy process that has no negative jumps, $\tau(\cdot)$ is a Lévy process itself, with Laplace exponent $-\psi_{\mathrm{D}}(\alpha)$; that is,

$$\mathbb{E}[e^{-\alpha\tau(z)}] = e^{-\psi_{\mathrm{D}}(\alpha)z}; \tag{2}$$

  see, for instance, Theorem 46.3 in [15]. Note that, for $\alpha \geq 0$, $\psi_{\mathrm{D}}(\alpha)$ is uniquely defined as the inverse of $\varphi_{\mathrm{D}}(\cdot)$, as $\varphi_{\mathrm{D}}(\cdot)$ increases on $[0, \infty)$.

- At the moment that the storage system hits zero, a service interruption starts. From then on, the storage level evolves according to a second Lévy process $X_{\mathrm{U}}(\cdot)$, which is assumed to be nondecreasing; in other words: $X_{\mathrm{U}}(\cdot)$ is a *subordinator*. This service interruption lasts for a time $\sigma(z)$ that equals $I + V$, where I denotes the amount of work generated by an *external* Lévy process (also a subordinator) $X_{\mathrm{E}}(\cdot)$ during $\tau(z)$ and V an independent vacation time (the vacations constitute a sequence of independent and identically distributed (i.i.d.) nonnegative random variables). Let $-\eta_{\mathrm{U}}(\cdot)$ be the Laplace exponent of $X_{\mathrm{U}}(\cdot)$ and let $-\eta_{\mathrm{E}}(\cdot)$ the one of $X_{\mathrm{E}}(\cdot)$. Finally, we let $\varrho_{\mathrm{U}} := \eta'_{\mathrm{U}}(0) = \mathbb{E}[X_{\mathrm{U}}(1)] < \infty$ and $\varrho_{\mathrm{E}} := \eta'_{\mathrm{E}}(0) = \mathbb{E}[X_{\mathrm{E}}(1)] < \infty$ be the corresponding (positive) drifts.

  Equation (2) immediately yields the transform of the duration of the service interruption, if the preceding service period started off at storage level z:

$$\mathbb{E}[e^{-\alpha\sigma(z)}] = \mathbb{E}[e^{-\eta_{\mathrm{E}}(\alpha)\tau(z)}]\mathbb{E}[e^{-\alpha V}] = e^{-\psi_{\mathrm{D}}(\eta_{\mathrm{E}}(\alpha))z}\mathbb{E}[e^{-\alpha V}]. \tag{3}$$

  The case that the queue is still empty after $I + V$ (i.e., at time $\tau(z) + \sigma(z)$) should be handled separately. This case can only occur when $X_{\mathrm{U}}(\cdot)$ is a

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

compound Poisson process with arrival rate λ_U and jumps distributed like some random variable B; that is, when $\eta_U(\alpha) = \lambda_U(1 - \mathbb{E}[e^{-\alpha B}])$, so that $\lambda_U = \eta_U(\infty)$ and

$$\mathbb{E}[e^{-\alpha B}] = 1 - \frac{\eta_U(\alpha)}{\lambda_U} . \tag{4}$$

We note that if $X_U(\cdot)$ is not compound Poisson, then $\eta_U(\alpha) \to \infty$ as $\alpha \to \infty$, and for each $t > 0$,

$$\mathbb{P}[X_U(t) = 0] = \lim_{\alpha\to\infty} \mathbb{E}[e^{-\alpha X_U(t)}] = \lim_{\alpha\to\infty} e^{-\eta_U(\alpha)t} = 0 . \tag{5}$$

In the situation of an empty queue after $I + V$, we either extend the vacation period until work is generated by $X_U(\cdot)$ or immediately take a sequence of independent vacations distributed like V until the first time a positive content level is found at the end of a vacation. For the first case, we denote this additional period by a random variable L, which is exponentially distributed with rate λ_U, whereas the storage level as soon as it becomes positive is distributed like B.

- Then a new service period starts again (i.e., the Lévy process $X_D(\cdot)$ becomes active again).

3. EQUILIBRIUM DISTRIBUTION OF THE EMBEDDED PROCESS

In this section we concentrate on the distribution of the storage level at embedded epochs (viz. the epochs right before an active period starts) it is clear that these storage levels constitute a discrete-time Markov process on $[0, \infty)$. We also note that because a Lévy process is continuous in probability, considering the process right before or right after makes a difference only for the single-vacation case when $X_U(\cdot)$ is compound Poisson and only at epochs when there is an arrival to an empty system.

Let Z_n denote the storage level right *before* the nth active/passive cycle. For the multiple-vacation case, it is convenient to assume that there is such a cycle after every vacation even when the content level is zero, since then the length of the active period is zero anyway. We recall that there is a difference between the single- and multiple-vacation cases only when $X_U(\cdot)$ is compound Poisson. We also recall that, for the latter, the arrival rate is $\lambda_U = \eta_U(\infty)$ and the jump size distribution has the Laplace-Stieltjes transform (LST) $\beta(\alpha) := \mathbb{E}[e^{-\alpha B}] = 1 - \eta_U(\alpha)/\lambda_U$ (see (4)). Applying standard Markov process terminology, we denote by $\mathbb{E}_z$ and $\mathbb{P}_z$ the expected value and probability when this embedded chain is initiated at some level $z \geq 0$.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

Starting from an initial position $z > 0$, $Z_1 = X_{\mathrm{U}}(X_{\mathrm{E}}(\tau(z)) + V)$, so that

$$\begin{aligned}\mathbb{E}_z[e^{-\alpha Z_1}] &= \mathbb{E}[e^{-\alpha X_{\mathrm{U}}(X_{\mathrm{E}}(\tau(z))+V)}] \\ &= \mathbb{E}[e^{-\eta_{\mathrm{U}}(\alpha)(X_{\mathrm{E}}(\tau(z))+V)}] \\ &= \mathbb{E}[e^{-\eta_{\mathrm{U}}(\alpha)V}]\mathbb{E}[e^{-\eta_{\mathrm{E}}(\eta_{\mathrm{U}}(\alpha))\tau(z)}] \\ &= \mathbb{E}[e^{-\eta_{\mathrm{U}}(\alpha)V}]e^{-\psi_{\mathrm{D}}(\eta_{\mathrm{E}}(\eta_{\mathrm{U}}(\alpha)))z} \\ &= g(\alpha)e^{-h(\alpha)z}, \end{aligned} \tag{6}$$

where $g(\alpha) := \mathbb{E}[e^{-\eta_{\mathrm{U}}(\alpha)V}]$ and $h(\alpha) := \psi_{\mathrm{D}}(\eta_{\mathrm{E}}(\eta_{\mathrm{U}}(\alpha)))$. When $X_{\mathrm{U}}(\cdot)$ is not compound Poisson, then we only need to consider $z > 0$, as at the end of a vacation there is never an empty system. In this case, there is no distinction between the single- and multiple-vacation cases.

When $X_{\mathrm{U}}(\cdot)$ is compound Poisson, for the multiple-vacation case it is clear that (6) is valid also for $z = 0$ (i.e., $\mathbb{E}_0[e^{-\alpha Z_1}] = g(\alpha)$). For the single-vacation case, starting from zero just before time 0 is the same as starting from a level distributed like B right after time 0. Thus,

$$\begin{aligned}\mathbb{E}_0[e^{-\alpha Z_1}] &= \mathbb{E}_B[e^{-\alpha Z_1}] \\ &= g(\alpha)\mathbb{E}[e^{-h(\alpha)B}] \qquad (7) \\ &= g(\alpha)\beta(h(\alpha)) \\ &= g(\alpha)\left(1 - \frac{\eta_{\mathrm{U}}(h(\alpha))}{\lambda_{\mathrm{U}}}\right). \qquad (8)\end{aligned}$$

Denoting

$$\gamma := h'(0) = \psi'_{\mathrm{D}}(0)\eta'_{\mathrm{E}}(0)\eta'_{\mathrm{U}}(0) = \frac{\varrho_{\mathrm{U}}\varrho_{\mathrm{E}}}{\varrho_{\mathrm{D}}}, \tag{9}$$

it follows either by differentiation or directly that for $z > 0$,

$$\mathbb{E}_z[Z_1] = \gamma z + \varrho_{\mathrm{U}}\mathbb{E}[V] \tag{10}$$

so that it is reasonable to guess that the stability condition is $\gamma < 1$ or equivalently $\varrho_{\mathrm{U}}\varrho_{\mathrm{E}} < \varrho_{\mathrm{D}}$. Since h is a composition of increasing concave functions, it is also increasing and concave with $h(0) = 0$. Thus, $h(\alpha)/\alpha$ is decreasing so that in particular, $h(\alpha) \le \gamma\alpha$. Thus, if we let $h^{(0)}(\alpha) = \alpha$ and $h^{(n)}(\alpha) = h(h^{(n-1)}(\alpha))$, then it is clear that $h^{(n)}(\alpha) \le \gamma^n\alpha$, and when $\gamma < 1$, for each $\alpha > 0$, $h^{(n)}(\alpha) \to 0$ geometrically fast as $n \to \infty$. From here on, we indeed assume that $\gamma < 1$ and we will see that this is indeed the stability condition.

Note that, being an LST, g is log-convex and decreasing, so that $f(\alpha) = -\log(g(\alpha))$ is increasing and concave. Thus, $f(\alpha) \le f'(0)\alpha$, where $f'(0) = -g'(0) =$

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

$\varrho_U \mathbb{E}[V]$, as $g(0) = 1$. This also implies that $f(h^{(i)}(\alpha)) \le f'(0)h^{(i)}(\alpha) \le f'(0)\gamma^i\alpha$, so that

$$\sum_{i=0}^{\infty} f(h^{(i)}(\alpha)) \le \frac{f'(0)}{1-\gamma}\alpha. \tag{11}$$

For the noncompound Poisson case with $z > 0$ and for the compound Poisson multiple-vacation case with $z \ge 0$, we saw that

$$\mathbb{E}_z[e^{-\alpha Z_1}] = g(\alpha)e^{-h(\alpha)z} \tag{12}$$

and, hence, by induction

$$\mathbb{E}_z[e^{-\alpha Z_n}] = e^{-h^{(n)}(\alpha)z} \prod_{i=0}^{n-1} g(h^{(i)}(\alpha)). \tag{13}$$

Since $h^{(n)}(\alpha) \le \gamma^n\alpha \to 0$, it follows that for every $z > 0$,

$$\lim_{n\to\infty} \mathbb{E}_z[e^{-\alpha Z_n}] = \prod_{i=0}^{\infty} g(h^{(i)}(\alpha)), \tag{14}$$

where the right-hand side is independent of z. It is easy to check that if we replace z by any positive (finite) random variable, the result remains valid.

To show that the right-hand side is indeed an LST, it suffices to show that as $\alpha \downarrow 0$, it converges to 1 (no mass escapes to infinity). Using (11), this is obvious from

$$\prod_{i=0}^{\infty} g(h^{(i)}(\alpha)) = \exp\left(-\sum_{i=0}^{\infty} f(h^{(i)}(\alpha))\right) \ge e^{-[f'(0)/(1-\gamma)]\alpha}. \tag{15}$$

Clearly $\zeta(\alpha) := \prod_{i=0}^{\infty} g(h^{(i)}(\alpha))$ is the transform of a stationary distribution π, as

$$\mathbb{E}_\pi[e^{-\alpha Z_1}] = g(\alpha)\mathbb{E}_\pi[e^{-h(\alpha)Z_0}] = g(\alpha)\zeta(h(\alpha)) = \zeta(\alpha). \tag{16}$$

The fact that there is a unique stationary distribution follows from the fact that the (unique) limit is independent of initial conditions.

Now that we have analyzed the noncompound Poisson case and the compound Poisson multiple-vacation case, we turn to the compound Poisson single-vacation case. This case turns out to be considerably harder, because of the possibility of the queue still being empty after $I + V$. We first focus on the question of the existence of a unique stationary distribution of the process Z_n, which we establish in Theorem 1. To this end, we prove two useful lemmas.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

Define $N := \inf\{n \in \mathbb{N} |\ Z_n = 0\}$. From

$$\mathbb{P}_z[N > n] = \mathbb{E}_z[1_{\{Z_1>0\}}\mathbb{P}_{Z_1}[N > n-1]] \tag{17}$$

and

$$\begin{aligned}\mathbb{E}_z[1_{\{Z_1>0\}}e^{-\alpha Z_1}] &= \mathbb{E}_z[e^{-\alpha Z_1}] - \mathbb{P}_z[Z_1 = 0] \\ &= g(\alpha)e^{-h(\alpha)z} - g(\infty)e^{-h(\infty)z}\end{aligned} \tag{18}$$

for $z > 0$, the following may be shown by induction.

LEMMA 1: *For every $z > 0$ and every $n \geq 2$,*

$$\mathbb{P}_z[N > n] = 1 - \sum_{i=1}^{n} a_i^n e^{-h^{(i)}(\infty)z}, \tag{19}$$

where $a_1^1 := g(\infty)$,

$$a_1^n := g(\infty)\left(1 - \sum_{i=1}^{n-1} a_i^{n-1}\right), \tag{20}$$

and $a_i^n := a_{i-1}^{n-1} g(h^{(i-1)}(\infty))$ for $i = 2, \ldots, n$.

Letting $z \downarrow 0$, (19) implies that

$$\mathbb{P}_{0+}[N > n] = 1 - \sum_{i=1}^{n} a_i^n, \tag{21}$$

so that from (20) we have that $a_1^n = \mathbb{P}_{0+}[N > n-1]g(\infty)$ and, consequently, that

$$\begin{aligned}a_i^n &= a_1^{n-i+1} g(h(\infty)) \cdot \cdots \cdot g(h^{(i-1)}(\infty)) \\ &= \mathbb{P}_{0+}[N > n-i] g(\infty) g(h(\infty)) \cdot \cdots \cdot g(h^{(i-1)}(\infty))\end{aligned} \tag{22}$$

for all $1 \leq i \leq n$. Hence, $a_i^n > 0$ for all $1 \leq i \leq n$ and $\sum_{i=1}^{n} a_i^n < 1$. We summarize with the following.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

LEMMA 2: *Let*

$$G_i := \begin{cases} 1, & i = 0 \\ g(\infty), & i = 1 \\ g(\infty)g(h(\infty)) \cdot \cdots \cdot g(h^{i-1}(\infty)), & i \geq 2; \end{cases} \tag{23}$$

then

$$\mathbb{P}_z[N > n] = 1 - \sum_{i=1}^{n} \mathbb{P}_{0+}[N > n - i] G_i e^{-h^{(i)}(\infty)z} \tag{24}$$

and for $z \downarrow 0$,

$$\sum_{i=0}^{n} \mathbb{P}_{0+}[N > n - i] G_i = 1. \tag{25}$$

We recall from (15) that $\prod_{i=0}^{\infty} g(h^{(i)}(\alpha)) > 0$ for all $\alpha > 0$, and, in particular, when inserting $\alpha = h(\infty)$, it implies that $G_i \geq G_\infty := g(\infty) \prod_{i=1}^{\infty} g(h^{(i)}(\infty)) > 0$. Therefore, (25) is bounded below by $G_\infty \sum_{i=0}^{n} \mathbb{P}_{0+}[N > i]$ so that

$$\mathbb{E}_{0+}[N] = \sum_{n=0}^{\infty} \mathbb{P}_{0+}[N > n] \leq G_\infty^{-1} < \infty. \tag{26}$$

We are now ready to show the following.

THEOREM 1: *Consider the single-vacation model. When* $X_{\mathrm{U}}(\cdot)$ *is compound Poisson, for every initial distribution of* Z_0 *having a finite mean, the expected value of* N *is finite. Hence, there is a unique stationary distribution (being equal to the limiting distribution) for the process* $(Z_n)_{n\in\mathbb{N}}$*, which is independent of initial conditions.*

PROOF: Due to (24) and (25), we have that for $n \geq 1$,

$$\begin{aligned} \mathbb{P}_z[N > n] &= \mathbb{P}_{0+}[N > n] + \sum_{i=1}^{n} \mathbb{P}_{0+}[N > n - i] G_i (1 - e^{-h^{(i)}(\infty)z}) \\ &\leq \mathbb{P}_{0+}[N > n] + \sum_{i=1}^{n} \mathbb{P}_{0+}[N > n - i] \gamma^{i-1} h(\infty) z, \end{aligned} \tag{27}$$

where the second inequality follows from $1 - e^{-h^{(i)}(\infty)z} \leq h^{(i)}(\infty)z \leq \gamma^{i-1} h(\infty) z$ as well as $G_i \leq 1$. Noting that $\mathbb{P}_z[N > 0] = \mathbb{P}_{0+}[N > 0] = 1$, we have, upon summation, that

$$\mathbb{E}_z[N] \leq \mathbb{E}_{0+}[N] \left(1 + \frac{h(\infty)}{1 - \gamma} z\right). \tag{28}$$

In particular, if we replace z by any arbitrary random variable having a finite mean, the expected value remains finite. As zero is an accessible state for $(Z_n)_{n\in\mathbb{N}}$, the proof is complete. ∎

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

We remark that $\mathbb{P}_{0+}[N > n]$ and $\mathbb{E}_{0+}[N]$ are *not* zero, as for any $z > 0$, the length of the interruption period is at least V. Also, we note that what we have shown until now is that regardless of whether $X_{\mathrm{U}}(\cdot)$ is compound Poisson, and when it is regardless of whether we are dealing with the single- or multiple-vacation case, the embedded Markov process $(Z_n)_{n\in\mathbb{N}}$ has a unique stationary distribution (being equal to the limiting distribution). For the cases where $X_{\mathrm{U}}(\cdot)$ is not compound Poisson or the multiple-vacation case, we already observed that the LST of this distribution is given by the right-hand side of (14).

For the single-vacation case where $X_{\mathrm{U}}(\cdot)$ is compound Poisson, we recall that for $z > 0$,

$$\mathbb{E}_z[e^{-\alpha Z_1}] = g(\alpha)e^{-h(\alpha)z}, \tag{29}$$

and for $z = 0$,

$$\mathbb{E}_0[e^{-\alpha Z_1}] = g(\alpha)\beta(h(\alpha)) = g(\alpha)\left(1 - \frac{\eta_{\mathrm{U}}(h(\alpha))}{\lambda_{\mathrm{U}}}\right). \tag{30}$$

Therefore, for $z > 0$,

$$\mathbb{P}_z[Z_1 = 0] = \lim_{\alpha\to\infty} \mathbb{E}_z[e^{-\alpha Z_1}] = g(\infty)e^{-h(\infty)z}, \tag{31}$$

where $g(\infty) = \mathbb{E}[e^{-\lambda_{\mathrm{U}} V}]$ and $h(\infty) = \psi_{\mathrm{D}}(\eta_{\mathrm{E}}(\lambda_{\mathrm{U}}))$. Similarly,

$$\mathbb{P}_0[Z_1 = 0] = g(\infty)\left(1 - \frac{\eta_{\mathrm{U}}(h(\infty))}{\lambda_{\mathrm{U}}}\right). \tag{32}$$

Therefore, if we let Z_0 have the desired stationary distribution $\pi(\cdot)$ with $p = \pi(\{0\})$ and LST $\zeta(\cdot)$, then we must have

$$\begin{aligned}\zeta(\alpha) &= \mathbb{E}_\pi[e^{-\alpha Z_1}]\\ &= g(\alpha)\left[\mathbb{E}_\pi[1_{\{Z_0>0\}}e^{-h(\alpha)Z_0}] + p\left(1 - \frac{\eta_{\mathrm{U}}(h(\alpha))}{\lambda_{\mathrm{U}}}\right)\right]\\ &= g(\alpha)[\zeta(h(\alpha)) - p\lambda_{\mathrm{U}}^{-1}\eta_{\mathrm{U}}(h(\alpha))],\end{aligned} \tag{33}$$

which implies that

$$\zeta(h^{(i)}(\alpha)) = g(h^{(i)}(\alpha))[\zeta(h^{(i+1)}(\alpha)) - p\lambda_{\mathrm{U}}^{-1}\eta_{\mathrm{U}}(h^{(i+1)}(\alpha))] \tag{34}$$

and, thus, for $i \geq 0$,

$$\begin{aligned}&p\lambda_{\mathrm{U}}^{-1}\left(\prod_{j=0}^{i} g(h^{(j)}(\alpha))\right)\eta_{\mathrm{U}}(h^{(i+1)}(\alpha))\\ &\quad= \left(\prod_{j=0}^{i} g(h^{(j)}(\alpha))\right)\zeta(h^{(i+1)}(\alpha)) - \left(\prod_{j=0}^{i-1} g(h^{(j)}(\alpha))\right)\zeta(h^{(i)}(\alpha)),\end{aligned} \tag{35}$$

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

where an empty product is 1. Since $h^{(n)}(\alpha) \to 0$ and thus $\zeta(h^{(n)}(\alpha)) \to 1$ as $n \to \infty$, this implies, upon summation, that

$$p\lambda_{\mathrm{U}}^{-1} \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i} g(h^{(j)}(\alpha)) \right) \eta_{\mathrm{U}}(h^{(i+1)}(\alpha)) = \prod_{j=0}^{\infty} g(h^{(j)}(\alpha)) - \zeta(\alpha). \tag{36}$$

In particular, noting that $\zeta(\infty) = p$, we can let $\alpha \to \infty$ and evaluate p. Therefore, we finally have the following result.

THEOREM 2: *When $\gamma = \varrho_{\mathrm{U}}\varrho_{\mathrm{E}}/\varrho_{\mathrm{D}} < 1$, the LST of the stationary distribution of the embedded chain is given by*

$$\zeta(\alpha) = \prod_{j=0}^{\infty} g(h^{(j)}(\alpha)) - c \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i} g(h^{(j)}(\alpha)) \right) \eta_{\mathrm{U}}(h^{(i+1)}(\alpha)), \tag{37}$$

where $c = 0$ except for the single-vacation case with $X_{\mathrm{U}}(\cdot)$ compound Poisson, in which case $c := p\lambda_{\mathrm{U}}^{-1}$, where

$$p := \frac{\prod_{j=0}^{\infty} g(h^{(j)}(\infty))}{1 + \lambda_{\mathrm{U}}^{-1} \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i} g(h^{(j)}(\infty)) \right) \eta_{\mathrm{U}}(h^{(i+1)}(\infty))}. \tag{38}$$

Remark 1: Just as we can define Z as the stationary version of the Z_n, we can also define Z_+ as the stationary version of the $Z_{n,+}$, where $Z_{n,+}$ denotes the buffer level right *after* the start of the nth active/passive cycle. Recalling that $\zeta(\alpha) = \mathbb{E}[e^{-\alpha Z}]$, it is readily verified that

$$\mathbb{E}[e^{-\alpha Z_+}] = \zeta(\alpha) - (1 - \beta(\alpha))\zeta(\infty). \tag{39}$$

The rest of this section is devoted to computations regarding the mean workload and correlation structure. From (15), it follows that

$$0 \le 1 - \prod_{i=0}^{\infty} g(h^{(i)}(\alpha)) \le 1 - e^{-[f'(0)/(1-\gamma)]\alpha} \le \frac{f'(0)}{1-\gamma}\alpha. \tag{40}$$

Moreover, from the concavity of $\eta_{\mathrm{U}}(\cdot)$, it follows that $\eta_{\mathrm{U}}(h^{(i)}(\alpha)) \le \gamma^i \varrho_{\mathrm{U}}\alpha$ and, thus,

$$0 \le \sum_{i=0}^{\infty} \left(\prod_{j=0}^{i} g(h^{(j)}(\alpha)) \right) \eta_{\mathrm{U}}(h^{(i+1)}(\alpha)) \le \frac{\gamma\varrho_{\mathrm{U}}}{1-\gamma}\alpha \tag{41}$$

so that $(1 - \zeta(\alpha))/\alpha$ is bounded and thus $\mathbb{E}_\pi[Z_0]$ is finite. To compute it, we note from (10) for the single-vacation case when $X_{\mathrm{U}}(\cdot)$ is compound Poisson, via conditioning

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

on Z_0, that

$$\begin{aligned}\mathbb{E}_\pi[Z_0] &= \mathbb{E}_\pi[Z_1] \\ &= \mathbb{E}_\pi[\gamma(Z_0 1_{\{Z_0>0\}} + B1_{\{Z_0=0\}}) + \varrho_{\mathrm{U}}\mathbb{E}[V]] \\ &= \gamma\mathbb{E}_\pi[Z_0] + \gamma p\mathbb{E}[B] + \varrho_{\mathrm{U}}\mathbb{E}[V] \\ &= \gamma\mathbb{E}_\pi[Z_0] + \varrho_{\mathrm{U}}(\gamma c + \mathbb{E}[V]) \end{aligned} \tag{42}$$

recalling that $c = p\lambda_{\mathrm{U}}^{-1}$ and $\varrho_{\mathrm{U}} = \lambda_{\mathrm{U}}\mathbb{E}[B]$. Therefore,

$$\mathbb{E}_\pi[Z_0] = \frac{\varrho_{\mathrm{U}}(\gamma c + \mathbb{E}[V])}{1-\gamma}. \tag{43}$$

For all other cases, the same holds only with $c = 0$. Using similar ideas, it is not hard to show that if $\mathbb{E}[V^2]$, $\sigma_{\mathrm{D}}^2 := \varphi_{\mathrm{D}}''(0) = \mathrm{Var}(X_{\mathrm{D}}(1))$, $\sigma_{\mathrm{E}}^2 := -\eta_{\mathrm{E}}''(0) = \mathrm{Var}(X_{\mathrm{E}}(1))$, and $\sigma_{\mathrm{U}}^2 := -\eta_{\mathrm{U}}''(0) = \mathrm{Var}(X_{\mathrm{U}}(1))$, are all finite, then $\mathbb{E}_\pi[Z_0^2]$ is finite as well and can be computed from (37) via differentiation or by conditioning from

$$\mathrm{Var}_\pi(Z_0) = \mathrm{Var}_\pi(Z_1) = \mathrm{Var}_\pi(X_{\mathrm{U}}(X_{\mathrm{E}}(\tau(Z_0)) + 1_{\{Z_0=0\}}X_{\mathrm{E}}(\tau(B)) + V)). \tag{44}$$

As the formulas are somewhat cumbersome and will not be needed in the sequel, we omit the details. For all cases except the single-vacation case with $X_{\mathrm{U}}(\cdot)$ compound Poisson, the result is

$$\mathrm{Var}_\pi(Z_0) = \frac{\varrho_{\mathrm{U}}^2}{1-\gamma^2}\left(\frac{\mathbb{E}[V]}{1-\gamma}\left[\frac{\sigma_{\mathrm{U}}^2}{\varrho_{\mathrm{U}}^2} + \varrho_{\mathrm{E}}\gamma\left(\frac{\sigma_{\mathrm{E}}^2}{\varrho_{\mathrm{E}}^2} + \frac{\sigma_{\mathrm{D}}^2}{\varrho_{\mathrm{D}}^2}\right)\right] + \mathrm{Var}(V)\right). \tag{45}$$

Finally, we note that

$$\begin{aligned}\mathbb{E}_\pi[Z_0Z_1] &= \mathbb{E}_\pi[Z_0(\gamma Z_0 + \varrho_{\mathrm{U}}(c\gamma + \mathbb{E}[V]))] \\ &= \gamma\mathbb{E}_\pi[Z_0^2] + \varrho_{\mathrm{U}}(c\gamma + \mathbb{E}[V])\mathbb{E}_\pi[Z_0] \\ &= \gamma\mathbb{E}_\pi[Z_0^2] + (1-\gamma)(\mathbb{E}_\pi[Z_0])^2 \end{aligned} \tag{46}$$

and thus

$$\mathrm{Cov}_\pi(Z_0, Z_1) = \gamma\mathrm{Var}_\pi(Z_0). \tag{47}$$

Interestingly, the load γ is apparently equal to the correlation coefficient between two subsequent values of the embedded process, say Z_n and Z_{n+1}, given that Z_n has the equilibrium distribution.

4. EQUILIBRIUM DISTRIBUTION OF THE WORKLOAD

In the previous section we have analyzed the transform of the equilibrium distribution of the storage level right before the start of the active period. The present section

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

translates this into the transform of the equilibrium distribution at an *arbitrary* instant in time using standard point process methodology. In particular, initiating the process with the stationary distribution of the embedded process, one then obtains a stationary marked point process where the points are the epochs at the beginning of the active cycles and the marks are the evolution of the process between these epochs, noting that the stationary independent increments property of all the building blocks involved ensures that the structure fits the stationary marked point process framework. Thus, in order to show that the continuous-time process $\{Z(s)|\ s \geq 0\}$ has a stationary version, it suffices to show that the times between consecutive points have a finite expected value and then the computation of the one-dimensional stationary distribution is done in the same way as for regenerative processes; that is, one integrates a function of the process (in this case, the function $u(z) = e^{-\alpha z}$) over the interval until the first point, takes the expected value and then divides by the expected value of the time until the first point. Starting from the stationary distribution of the embedded chain, we note that the time until the first point is given by

$$\tau(Z_0) + X_{\mathrm{E}}(\tau(Z_0)) + V + 1_{\{X_{\mathrm{U}}(X_{\mathrm{E}}(\tau(Z_0))+V)=0\}} L, \tag{48}$$

where L is exponential with rate λ_{U} for the single-vacation case when $X_{\mathrm{U}}(\cdot)$ is compound Poisson and zero otherwise. Thus, the expected value under the stationary distribution π is

$$\frac{1 + \varrho_{\mathrm{E}}}{\varrho_{\mathrm{D}}} \mathbb{E}_\pi[Z_0] + \mathbb{E}[V] + c < \infty \tag{49}$$

(as $\mathbb{E}_\pi[Z_0]$ was shown to be finite in Section 3), where we recall that $c = \mathbb{P}_\pi[Z_1 = 0]\mathbb{E}[L] = p\lambda_{\mathrm{U}}^{-1}$ for the single-vacation compound Poisson case and zero otherwise.

In our case, the procedure followed uses decoupling of the active periods and the interruptions; the desired transform follows by weighing these in an appropriate way.

4.1. Active Periods

First concentrate on the active periods. Consider the martingale

$$e^{-\alpha X_{\mathrm{D}}(t)} - e^{-\alpha X_{\mathrm{D}}(0)} - \varphi_{\mathrm{D}}(\alpha) \int_0^t e^{-\alpha X_{\mathrm{D}}(s)}\, ds, \tag{50}$$

with stopping time $\tau(z)$. Using the fact that $X_{\mathrm{D}}(\cdot)$ has no negative jumps, we derive the identity, by applying "optional stopping,"

$$\mathbb{E}\left[\int_0^{\tau(z)} e^{-\alpha X_{\mathrm{D}}(s)}\, ds\right] = \frac{e^{\alpha z} - 1}{\varphi_{\mathrm{D}}(\alpha)}. \tag{51}$$

Recalling that, on $[0, \tau(z)]$, it holds that $z + X_{\mathrm{D}}(s) = Z(s)$, this immediately yields

$$\mathcal{L}(\alpha) := \mathbb{E}\left[\int_0^{\tau(Z_+)} e^{-\alpha Z(s)}\, ds\right] = \frac{1 - \mathbb{E}[e^{-\alpha Z_+}]}{\varphi_{\mathrm{D}}(\alpha)}, \tag{52}$$

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

where, as earlier, the random variable Z_+ denotes the storage level at (i.e., right after) the beginning of an active period (such that $\mathbb{E}[e^{-\alpha Z_+}]$ can be computed as in section 3; see Remark 1). This expression can be interpreted by using the integrated tail $Z_+^{\rm res}$ of Z_+, characterized by the transform

$$\mathbb{E}[e^{-\alpha Z_+^{\rm res}}] = \frac{1}{\alpha \mathbb{E}[Z_+]}(1 - \mathbb{E}[e^{-\alpha Z_+}]). \tag{53}$$

Now, $\mathcal{L}(\alpha)$ can be expressed in terms of the distribution of $Z_+^{\rm res}$:

$$\mathcal{L}(\alpha) = \frac{\alpha \mathbb{E}[e^{-\alpha Z_+^{\rm res}}]}{\varphi_{\rm D}(\alpha)} \mathbb{E}[Z_+]. \tag{54}$$

Division by $\mathbb{E}[\tau(Z_+)] = \mathbb{E}[Z_+]/\varrho_{\rm D}$ yields an expression for the steady-state workload during active periods:

$$\frac{\varrho_{\rm D} \alpha \mathbb{E}[e^{-\alpha Z_+^{\rm res}}]}{\varphi_{\rm D}(\alpha)}; \tag{55}$$

notice the similarity of transform (55) with the celebrated Pollaczek–Khinchine formula $\varphi_{\rm D}'(0)\alpha/\varphi_{\rm D}(\alpha)$ [2, Cor. IX.3.4] describing the steady state of a Lévy process (with negative drift) reflected at zero. Formula (55) is a decomposition that is related to formula (5.6) of [6].

4.2. Interruptions

Now, concentrate on the service interruptions. The time-average distribution during these intervals is characterized by the ratio of

$$\mathbb{E}\left[\int_0^{\sigma(Z_+)} e^{-\alpha X_{\rm U}(s)}\, ds\, 1\{X_{\rm U}(\sigma(Z_+)) > 0\}\right] + \mathbb{E}\left[\int_0^{\sigma(Z_+)+L} e^{-\alpha X_{\rm U}(s)}\, ds\, 1\{X_{\rm U}(\sigma(Z_+)) = 0\}\right] \tag{56}$$

and

$$\mathbb{E}\left[\sigma(Z_+)\, 1\{X_{\rm U}(\sigma(Z_+)) > 0\}\right] + \mathbb{E}\left[(\sigma(Z_+) + L)\, 1\{X_{\rm U}(\sigma(Z_+)) = 0\}\right]. \tag{57}$$

Let us start by considering the numerator (56), which can be rewritten as

$$\mathcal{N}(\alpha) := \mathbb{E}\left[\int_0^{\sigma(Z_+)} e^{-\alpha X_{\rm U}(s)}\, ds\right] + \mathbb{P}(X_{\rm U}(\sigma(Z_+)) = 0)\mathbb{E}[L]; \tag{58}$$

use the fact that $X_{\rm U}(s) = 0$ until the end of the period L. One readily verifies that

$$\mathbb{E}\left[\int_0^{\sigma(z)} e^{-\alpha X_{\rm U}(s)}\, ds\right] = \int_0^\infty \int_0^x e^{-\eta_{\rm U}(\alpha)s}\, ds\, d\mathbb{P}(\sigma(z) \le x) = \frac{\mathbb{E}\left[e^{-\eta_{\rm U}(\alpha)\sigma(z)}\right] - 1}{-\eta_{\rm U}(\alpha)}; \tag{59}$$

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

recall that both numerator and denominator of the last expression are negative (for positive α). We conclude that (56) reduces to

$$\mathcal{N}(\alpha) = \frac{\mathbb{E}[e^{-h(\alpha)Z_+}]g(\alpha) - 1}{-\eta_{\mathrm{U}}(\alpha)} + \mathbb{E}[e^{-h(\infty)Z_+}]g(\infty)\mathbb{E}[L]. \tag{60}$$

Likewise, the denominator (57) equals

$$\begin{aligned}\mathcal{D} &:= \mathbb{E}[\sigma(Z_+)] + \mathbb{E}[L\,1\{X_{\mathrm{U}}(\sigma(Z_+)) = 0\}] \\ &= \mathbb{E}[Z_+]\frac{\varrho_{\mathrm{E}}}{\varrho_{\mathrm{D}}} + \mathbb{E}[V] + \mathbb{E}[e^{-h(\infty)Z_+}]g(\infty)\mathbb{E}[L]. \end{aligned} \tag{61}$$

Denoting by W the steady-state storage level at an arbitrary instant in time, we obtain the following result.

THEOREM 3: *The LST of the steady-state storage level W is given by*

$$\mathbb{E}[e^{-\alpha W}] = \frac{\mathcal{L}(\alpha) + \mathcal{N}(\alpha)}{\mathcal{C} + \mathcal{D}}, \tag{62}$$

where $\mathcal{C} := \mathbb{E}[\tau(Z_+)] = \mathbb{E}[Z_+]/\varrho_{\mathrm{D}}$ and where $\mathcal{L}(\alpha)$, $\mathcal{N}(\alpha)$, and $\mathcal{D}$ are given by (54), (60)*, and* (61)*, respectively.*

Remark 2: Notice that, except for the single-vacation compound Poisson case, one can simplify (60) to $\mathcal{N}(\alpha) = [1 - \mathbb{E}[e^{-\alpha Z_+}]]/\eta_{\mathrm{U}}(\alpha)$ and thus (62) to

$$\mathbb{E}[\mathrm{e}^{-\alpha W}] = \frac{1 - \mathbb{E}[e^{-\alpha Z_+}]}{\alpha \mathbb{E}Z_+}\left(\frac{\alpha}{\eta_{\mathrm{U}}(\alpha)} + \frac{\alpha}{\varphi_{\mathrm{D}}(\alpha)}\right)\Big/\left(\frac{1}{\varrho_{\mathrm{U}}} + \frac{1}{\varrho_{\mathrm{D}}}\right) \tag{63}$$

$$= \mathbb{E}[e^{-\alpha Z_+^{\mathrm{res}}}]\frac{1 + (\varphi_{\mathrm{D}}(\alpha)/\eta_{\mathrm{U}}(\alpha))}{1 + \varrho_{\mathrm{D}}/\varrho_{\mathrm{U}}}\frac{\alpha\varphi_{\mathrm{D}}'(0)}{\varphi_{\mathrm{D}}(\alpha)}. \tag{64}$$

This result can be further interpreted with Remark 4.4 of Kella [13].

5. SPECIAL CASES AND RAMIFICATIONS

In this section we consider three special cases of the general model considered so far. Subsequently we discuss two model variants that can also be analyzed in detail.

Example 1: The main feature of the general model under consideration is the dependence between the length of a passive period and the length of the preceding active period. This dependence is eliminated by taking $X_{\mathrm{E}} \equiv 0$, yielding $h(\alpha) \equiv 0$. Consider the classical $M/G/1$ vacation queue "with single vacations", an $M/G/1$ queue in which the server goes on vacation when the system has become idle, and when

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

finding the system empty upon returning, the server waits until the arrival of the first customer.

The following notation is used. The customers arrive according to a Poisson process with rate λ. The amounts of work brought along by the customers are an i.i.d. sequence of random variables (where the size of such a job has transform $\beta(\cdot)$). It is assumed that $\varrho := -\lambda\beta'(0) < 1$.

It is easily seen that this model is a special case of our model; take

$$X_{\mathrm{E}} \equiv 0, \qquad \eta_{\mathrm{U}}(\alpha) = \lambda(1-\beta(\alpha)), \qquad \varphi_{\mathrm{D}}(\alpha) = -\lambda(1-\beta(\alpha)) + \alpha. \tag{65}$$

Notice that we could have taken other functions than $\beta(\alpha)$, thus allowing for different service time distributions of customers who arrive during an active period, a passive period and at the end of L. Also, by taking $\lambda^\star$ rather than λ in the definition of $-\eta_{\mathrm{U}}(\cdot)$, we could have allowed for a different arrival rate during passive periods. Shanthikumar [16] and Doshi [8] study vacation models in which the arrival rate changes per period (active/vacation).

We now specify the LST of Z_+, from Remark 1:

$$\mathbb{E}[e^{-\alpha Z_+}] = \mathbb{E}[e^{-\lambda(1-\beta(\alpha))V}] - (1-\beta(\alpha))\mathbb{E}[e^{-\lambda V}]; \tag{66}$$

notice that in the first product in the right-hand side of (37) only the $(j=0)$ factor is not equal to 1 and that in the sum, only the $(i=0)$ term is not equal to 0. The four elements of the expression for the LST of the storage level W in Theorem 3 are (observe that $\varrho_{\mathrm{D}} = 1-\lambda\mathbb{E}[B] = 1-\varrho$)

$$\mathcal{C} = \mathbb{E}[\tau(Z_+)] = \frac{\mathbb{E}[Z_+]}{1-\varrho}, \qquad \mathcal{D} = \mathbb{E}[V] + \mathbb{E}[e^{-\lambda V}]\frac{1}{\lambda}, \tag{67}$$

$$\mathcal{L}(\alpha) = \frac{(1-\varrho)\alpha}{\alpha - \lambda(1-\beta(\alpha))}\mathbb{E}\left[e^{-\alpha Z_+^{\mathrm{res}}}\right]\mathbb{E}[\tau(Z_+)], \tag{68}$$

$$\mathcal{N}(\alpha) = \frac{1-\mathbb{E}[e^{-\lambda(1-\beta(\alpha))V}]}{\lambda(1-\beta(\alpha))} + \mathbb{E}[e^{-\lambda V}]\frac{1}{\lambda}. \tag{69}$$

We refer to Takagi [19] for an extensive discussion of the $M/G/1$ queue with single vacations. The above expression for $\mathcal{L}(\alpha)$ can be found in Eq. (2.28a) of [19] (Takagi presents the waiting time transform of a customer who arrives during an active period; PASTA implies that this is also the conditional workload transform).

Remark 3: It should be observed that in this model, Z can be thought of as the workload that has accumulated during the passive period. The structure of $\mathcal{L}(\alpha)$ reveals a decomposition of the workload during active periods into an $M/G/1$ workload and an independent additional term. Such decomposition results play a central role in the literature on vacation models. Indeed, Fuhrmann and Cooper [10] prove that, for a large

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

class of $M/G/1$-type vacation queues, a decomposition of the following type holds: The steady-state queue length in the vacation model equals, in distribution, the sum of two independent quantities (viz. the queue length in the corresponding model without vacations and a term representing the effect of the vacations). Similar decompositions have been obtained for waiting times and workloads. See [3,4] for such a workload decomposition in single-server queues with multiple customer classes, like polling models.

Example 2: Let us consider a two-queue polling model, with exhaustive service at both queues Q_1 and Q_2 and with independent Poisson arrival processes with rates λ_1 and λ_2 and service time LSTs $\beta_1(\alpha)$ and $\beta_2(\alpha)$ (with means μ_1 and μ_2, respectively). Compared to the classical two-queue polling model, cf. [17], we introduce one slight adaptation: When the server leaves a queue and *both* queues are empty, the server waits for the first arrival at Q_1 (instead of waiting for the first arrival at *any* of the queues). One can adapt the definition of the period L from Section 2 to retrieve the classical polling model.

Using the results of the previous two sections, the workload distribution in Q_1 is obtained by making the following choices:

$$V = 0, \qquad \varphi_{\mathrm{D}}(\alpha) = -\lambda_1(1 - \beta_1(\alpha)) + \alpha,$$
$$\eta_{\mathrm{U}}(\alpha) = \lambda_1(1 - \beta_1(\alpha)), \qquad \eta_{\mathrm{E}}(\alpha) = \lambda_2(1 - p_2(\alpha)).$$

Here, $p_2(\alpha)$ is the transform of the busy period distribution in an $M/G/1$ queue with arrival rate λ_2 and service time transform $\beta_2(\alpha)$. This choice of $\eta_{\mathrm{E}}(\cdot)$ accomplishes the following: During an active period of Q_1, customers arrive at Q_2 according to a Poisson process with rate λ_2. During the subsequent passive period of Q_1, each of those, say, N_2 customers at Q_2 is served, along with all those arriving in Q_2 during their service time and so on. This amounts to N_2 busy periods at Q_2, reflecting exhaustive service at Q_2.

It is readily checked that the stability constraint $\varrho_{\mathrm{U}}\varrho_{\mathrm{E}} < \varrho_{\mathrm{D}}$ of this model translates to

$$\lambda_1\mu_1\left(\frac{\lambda_2\mu_2}{1 - \lambda_2\mu_2}\right) < 1 - \lambda_1\mu_1. \tag{70}$$

Defining $\varrho_i := \lambda_i\mu_i$, this reduces to the familiar $\varrho_1 + \varrho_2 < 1$.

A crucial feature of the analysis of the previous sections was that $\mathbb{E}[e^{-\alpha\tau(z)}] = e^{-f(\alpha)z}$ for some function $f(\cdot)$. This is an important property of Lévy processes but can also be enforced by appropriate different choices of $\tau(z)$. We consider two such choices.

Ramification 1: Assume that an active period lasts until the workload has been reduced to a fraction F of its value at the beginning of the active period (the situation described

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

in the previous sections corresponds to $F = 0$). So,

$$\tau(z) := \inf\{t \geq 0 : z + X_{\mathrm{D}}(t) = Fz\}, \qquad 0 \leq F < 1, \tag{71}$$

which is in distribution equal to $\inf\{t \geq 0 : (1-F)z + X_{\mathrm{D}}(t) = 0\}$. Hence,

$$\mathbb{E}[e^{-\alpha\tau(z)}] = e^{-\psi_{\mathrm{D}}(\alpha)(1-F)z}. \tag{72}$$

The case in which $F \in (0,1)$ is relatively easy, since the system never empties. Then (33) is modified into the following form:

$$\zeta(\alpha) = \zeta(k(\alpha))g(\alpha), \tag{73}$$

with $k(\alpha) := \psi_{\mathrm{D}}(\eta_{\mathrm{E}}(\eta_{\mathrm{U}}(\alpha)))(1-F) = (1-F)h(\alpha)$ and, as before, $g(\alpha) = \mathbb{E}[e^{-\eta_{\mathrm{U}}(\alpha)V}]$. The stability condition is easily seen to be $\varrho_{\mathrm{U}}\varrho_{\mathrm{E}}(1-F) < \varrho_{\mathrm{D}}(1-F)$, which reduces to $\varrho_{\mathrm{U}}\varrho_{\mathrm{E}} < \varrho_{\mathrm{D}}$.

It is not surprising that this criterion is independent of F. During the active periods, the storage level decreases by an amount $z(1-F)$, if the initial storage level is z. During the vacations, the storage level increases on average by $z(1-F)\varrho_{\mathrm{E}}\varrho_{\mathrm{U}}/\varrho_{\mathrm{D}}$. To ensure stability we should have that the average increase is smaller than the decrease $z(1-F)$. We see that the factor $1-F$ cancels.

If the stability condition (which is equivalent with $k'(0) < 1$) holds, then (73) yields the following expression for the steady-state workload Z at the embedded points of beginnings of activity periods:

$$\mathbb{E}[e^{-\alpha Z}] = \prod_{m=0}^{\infty} g(k^{(m)}(\alpha)). \tag{74}$$

Ramification 2: Let us consider a two-queue polling model with gated service at queue Q_1 and exhaustive service at queue Q_2 and with independent Poisson arrival processes with rates λ_1 and λ_2 and service time transform $\beta_1(\alpha)$ and $\beta_2(\alpha)$. The gated service policy amounts to the following: When the server visits a queue, it serves exactly all of the work (customers) present upon arrival and then moves on to the next queue. As in Example 2, we introduce one slight adaptation to the classical two-queue gated/exhaustive polling model: When the server leaves Q_2 and Q_1 is empty, the server waits for the first arrival at Q_1. If we replace the definition of $\tau(z)$ in Section 2 by $\tau(z) := z$ and we choose $\varphi_{\mathrm{D}}(\alpha) := -\lambda_1(1-\beta_1(\alpha)) + \alpha$, then an activity period can be viewed as the visit period of an $M/G/1$ queue Q_1 with arrival rate λ_1 and service time LST $\beta_1(\alpha)$, operating under the gated service policy. Using the results of the previous two sections, the workload distribution in Q_1 is obtained by making the following choices:

$$V = 0, \qquad \varphi_{\mathrm{D}}(\alpha) = -\lambda_1(1-\beta_1(\alpha)) + \alpha,$$
$$\eta_{\mathrm{U}}(\alpha) = \lambda_1(1-\beta_1(\alpha)), \qquad \eta_{\mathrm{E}}(\alpha) = \lambda_2(1-p_2(\alpha)).$$

Here, $p_2(\alpha)$ is defined as in Example 2.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press

Remark 4: It should now also be clear how to model polling models with exhaustive service at one queue and gated at another. Furthermore, choosing $V > 0$ allows for switch-over times between queues.

Acknowledgments

The authors are grateful to the referee for several interesting questions and useful suggestions. Part of this work was done while the third author visited EURANDOM and CWI as a Stieltjes visiting professor. The third author is also supported by grant 964/06 from the Israel Science Foundation and the Vigevani Chair in Statistics. The research of the first and second author was done within the framework of the BRICKS project and the European Network of Excellence Euro-NGI.

References

1. Altman, E. (2002). Stochastic recursive equations with applications to queues with dependent vacations. *Annals of Operations Research* 112: 43–61.
2. Asmussen, S. (2003). *Applied probability and queues.* New York: Springer-Verlag.
3. Boxma, O.J. (1989). Workloads and waiting times in single-server systems with multiple customer classes. *Queueing Systems* 5: 185–214.
4. Boxma, O.J. & Groenendijk, W.P. (1987). Pseudo-conservation laws in cyclic-service systems. *Journal of Applied Probability* 24: 949–964.
5. Doshi, B.T. (1986). Queueing systems with vacations: A survey. *Queueing Systems* 1: 29–66.
6. Doshi, B.T. (1989). Conditional and unconditional distributions for $M/G/1$ type queues with server vacations. *Queueing Systems* 7: 229–251.
7. Doshi, B.T. (1990). Single server queues with vacations. In: H. Takagi (ed.) *Stochastic analysis of computer and communication systems.* Amsterdam: North-Holland pp. 217–265.
8. Doshi, B.T. (1990). Generalizations of the stochastic decomposition results for single server queues with vacations. *Stochastic Models* 6: 307–333.
9. Eliazar, I. (2005). Gated polling systems with Lévy inflow and inter-dependent switchover times: A dynamical-systems approach. *Queueing Systems* 49: 49–72.
10. Fuhrmann, S.W. & Cooper, R.B. (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. *Oper. Res.* 33, 1117–1129.
11. Groenevelt, R. & Altman, E. (2005). Analysis of alternating-priority queueing models with (cross) correlated switchover times. *Queueing Systems* 51: 199–247.
12. Harris, C.M. & Marchal, W.G. (1988). State dependence in $M/G/1$ server-vacation models. *Operations Research* 36: 560–565.
13. Kella, O. (1998). An exhaustive Lévy storage process with intermittent output. *Stochastic Models* 14: 979–992.
14. Prabhu, N.U. (1998). *Stochastic storage processes.* New York: Springer-Verlag.
15. Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge: Cambridge University Press.
16. Shanthikumar, J.G. (1988). On stochastic decomposition in $M/G/1$ type queues with generalized server vacations. *Operations Research* 36: 566–569.
17. Takács, L. (1968). Two queues attended by a single server. *Operations Research* 16: 639–650.
18. Takagi, H. (1990). Queueing analysis of polling models: An update. In: H. Takagi (ed.), *Stochastic analysis of computer and communication systems.* Amsterdam: North-Holland, pp. 267–318.
19. Takagi, H. (1991). *Queueing analysis.* Vol. 1: *Vacation and priority systems, Part 1.* Amsterdam: North-Holland.
20. Takagi, H. (1997). Queueing analysis of polling models: Progress in 1990–1994. In: J.H. Dshalalow (ed.), *Frontiers in queueing.* Boca Raton, FL: CRC Press, pp. 119–146.
21. Teghem, J. Jr. (1986). Control of the service process in a queueing system. *European Journal of Operations Research* 23: 141–158.

https://doi.org/10.1017/S0269964808000326 Published online by Cambridge University Press