Communication

On the Time to Buffer Overflow in a Queueing Model with a General Independent Input Stream and Power-Saving Mechanism Based on Working Vacations

Martyna Kobielnik * and Wojciech Kempa

Department of Mathematics Applications and Methods for Artificial Intelligence, Faculty of Applied Mathematics, Silesian University of Technology, ul. Kaszubska 23, 44-100 Gliwice, Poland; wojciech.kempa@polsl.pl
* Correspondence: martyna.kobielnik@polsl.pl

Abstract: A single server $GI/M/1$ queue with a limited buffer and an energy-saving mechanism based on a single working vacation policy is analyzed. The general independent input stream and exponential service times are considered. When the queue is empty after a service completion epoch, the server lowers the service speed for a random amount of time following an exponential distribution. Packets that arrive while the buffer is saturated are rejected. The analysis is focused on the duration of the time period with no packet losses. A system of equations for the transient time to the first buffer overflow cumulative distribution functions conditioned by the initial state and working mode of the service unit is stated using the idea of an embedded Markov chain and the continuous version of the law of total probability. The explicit representation for the Laplace transform of considered characteristics is found using a linear algebra-based approach. The results are illustrated using numerical examples, and the impact of the key parameters of the model is investigated.

Keywords: energy saving; queueing model; time to buffer overflow; transient analysis; working vacation policy

1. Introduction

Evidently, the problem of reducing energy consumption is global. This results in large-scale research on algorithms supporting power-saving control and the accompanying technical solutions. Energy-saving solutions are particularly desired in the area of computer and telecommunications networks, which is related to the rapidly growing share of wireless transmissions. Wireless network components, such as sensor network sensors, are powered by batteries. Such networks are designed to constantly monitor the air temperature, humidity, road traffic intensity, etc. According to their purpose, e.g., to warn about fire hazards, sensors (network nodes) are often located in hard-to-reach places.

Limiting energy consumption and, consequently, extending the possibility of powering a node from a single battery is, therefore, of key importance in ensuring reliable data transmission and the associated security. Queueing theory is widely used in traffic modelling in energy-efficient packet networks. Indeed, queueing models, especially those with a finite capacity of accumulating buffer in which a mechanism limiting the operation of the service station has been implemented (in the case of, e.g., low-intensity traffic), can be effectively used in the process of controlling the QoS (Quality of Service) and the energy consumption of individual nodes. The knowledge of stochastic characteristics changing in time, such as the distribution of the queue length, queueing delay, or the time to buffer overflow, allows for ongoing monitoring of the system and, thus, for control of the transmission quality.

The concept of a queue model was proposed for the first time in [1], in which a service station remains unavailable for job service for some time. Queueing systems with

server vacations quickly gained popularity, and many new models and a whole range of analytical results concerning such models appeared in the literature. An exhaustive study on queueing models with different types of vacation policies can be found, e.g., in [2] or in survey papers [3,4].

In [5], a model was proposed in which—instead of temporarily suspending the service—the server processes jobs with different intensities (speeds) depending, e.g., on the intensity of incoming traffic. On the one hand, such a policy, called a working vacation (WV), allows for energy savings (caused by a temporary reduction in job service intensity), and on the other hand, allows for better control of the queue length and reduces the risk of serial job loss. Moreover, it allows to redirect the unused resources for other tasks, e.g., for maintenance purpose or for redirecting traffic from other nodes.

Under a single working vacation policy, the service station takes only one working vacation when the queue is emptied. An alternative is multiple vacation policy, in which successive single working vacations are initialized as far as at least one job waiting in the queue is detected. In [6], a model of working vacation was studied in the context of energy saving and latency control in wireless sensor networks. The authors introduced a two-threshold working vacation policy, which is a combination of a vacation and working vacation policy.

In [7], the two–channel model $M/M/2$ with WV, negative customers, feedback, and N-strategy was proposed to reduce the energy consumption in wireless communication networks. The energy saving capabilities of WV models were also discussed from the cloud platform point of view, see, e.g., [8,9].

An $M/G/1$-type queueing model with single server working vacations was studied in [10], where the stationary queue-size distribution was obtained via the supplementary variable technique and matrix-analytic approach. A $G/M/1$ infinite-buffer system with a single working vacation policy was investigated in [11,12]. A discrete-time model with single working vacations was analysed in [13], where interarrival times and service times are geometrically distributed.

In [14], a model with a general independent input stream and single working vacation policy is studied in the case of memoryless service time distribution (both exponential and geometrical). A $M/G/1$-type queueing model with single working vacations and vacation interruption under a Bernoulli schedule was considered in [15].

In such a model, if there are jobs present in the system at the working vacation completion moment, the server can initialize the next working vacation period (with probability p) or it can return to the normal mode (with probability $1 - p$). The joint distribution of the steady-state queue size and service status is then derived by using the supplementary variable technique.

A similar model was investigated in [16], where the sojourn time distribution is obtained. Recently, in [17], the steady-state characteristics of a Markovian queue with working vacations and breakdowns were studied using the spectral expansion method. In [18], a discrete-time model with general batch input and geometric service time with multiple working vacations was studied with the supplementary variable technique.

A model with a single working vacation, customer impatience, and catastrophes was analysed in [19]. The steady-state distribution of the system size for a model with customer impatience and server breakdowns was obtained in [20]. Additionally, the authors solved a profit optimization problem using a particle swarm optimization algorithm.

As can be seen, the vast majority of the results obtained for models with working vacations concern the steady state of the system. In practice, however, there are often situations in which stationary analysis is insufficient. This is the case, for example, when observing a system immediately after its opening (when its steady state has not yet been established), after changing the control mechanism, or after removing a failure.

In the case of low traffic intensity (which is typical for, e.g., sensor networks), system stabilization may take a long time and, consequently, the steady state ceases to be an indicator of system operation. There are few results for the transient state, especially for

models with a general input stream. In [21], the study of a model with general input and phase type service was carried out using a simulation-based approach.

In [22], the transient results for the number served during a busy period in a $GI/M/1/N$ queue was obtained by approximating the interarrival distribution according to a two-phase Cox distribution. A more general model, $G/G/m/m+K$, was studied in [23] using the diffusion approximation technique.

With regards to working vacation policy, most of the papers concern only a $M/M/1$ queue, e.g., [24,25]. A $M/M/1$ queue with working vacation and impatient customers was studied in [25]. The transient system size probabilities were obtained using a continued fractions approach. In [24], the transient solution was found by solving differential equations using the Runge–Kutta algorithm.

In [26], the transient behaviour of a finite-capacity model with a general independent input flow of jobs and single working vacation policy was investigated. Using an analytic approach based on the idea of an embedded Markov chain and linear algebra, the compact-form representation for the Laplace transform of the queue-size distribution conditioned by the initial buffer state was derived.

An energy-saving mechanism based on a threshold-controlled multiple vacation policy was considered in [27] as a model for the operation a wireless sensor network node. The Laplace transform representations were obtained for queue-size distribution at an arbitrary fixed time and for idle and processing periods.

Moreover, the compact-form formulae for the distributions of the idle and processing period duration were found. A mathematical model for the node of a wireless sensor network with discrete-time parameters was proposed in [28]. An explicit formula for the tail cumulative distribution function of the first buffer overflow period duration was obtained. Hence, the corresponding result for the next buffer overflow periods was found.

In this paper, we study a finite-buffer $GI/M/1/N$ queueing model with general-type independent input flow of jobs, exponentially distributed service times, and a single working vacation policy. Applying an analytic approach based on the idea of embedded Markov chain and linear algebra, we find the closed-form representation for the Laplace transform of the time to the first buffer overflow distribution, conditioned by the initial system state and working mode of the service unit. The theoretical results are illustrated using numerical examples.

The remainder of the article is organized as follows. In Section 2, we provide a precise mathematical description of the considered queueing model and present an auxiliary algebraic result. In Section 3, systems of integral equations for conditional distributions of the time to the first buffer overflow, based on the idea of embedded Markov chain and the continuous version of the formula of total probability, are established for both the system start and operation in normal and working vacation mode. The closed-form solutions for corresponding systems written for Laplace transforms are found in Section 4. Section 5 is devoted to detailed numerical examples illustrating theoretical results. Finally, Section 6 contains a short conclusion.

2. Model Description and Auxiliary Result

Let us consider a $GI/M/1/N/WV$ model, where the times between successive arrivals are independent random variables with a common cumulative distribution function (CDF) $F(t)$, and the service times are exponential random variables with parameters μ and μ_v in normal mode and during a working vacation, respectively. The system is characterized by a finite buffer. At any given time, only N jobs can be present, namely one in the service unit and $N-1$ in the queue.

Every time the server finds the queue empty after the service completion epoch, it enters a single working vacation period, changing the service intensity to a lower value μ_v, and stays in this mode for a period of time that is exponentially distributed with parameter α. When the WV period ends, the server returns to normal mode and normal service speed $\mu > \mu_v$.

Let $X(t)$ be the number of jobs present in the system at a time epoch t. The random variable

$$\gamma = \min\{t \geq 0 : X(t) = N\}$$

denotes the time to the first buffer overflow. Our goal is to determine the conditional CDFs of γ given the initial state and working mode of the server, i.e.,

$$T_n(t) = P\{\gamma \leq t | X(0) = n, Y(0) = 0\},$$

and

$$T_n^v(t) = P\{\gamma \leq t | X(0) = n, Y(0) = 1\},$$

where $Y(t) = 0$ if at the time instant t, the server is in normal mode, and $Y(t) = 1$, otherwise.

In the following subsections, a system of integral equations for $T_n(t)$ and $T_n^v(t)$, $n = 0, 1, \ldots, N$ is stated using the formula of total probability and the method of embedded Markov chain. In the next section, the corresponding system for Laplace transforms of $T_n(t)$ and $T_n^v(t)$ is solved applying the method of potential.

For the rest of this paper, we use the following notations:

$$p_i(a) = \frac{a^i}{i!}e^{-a}, i = 0, 1, \ldots,$$

$$E_{i,a}(t) = 1 - \sum_{j=0}^{i-1} \frac{(at)^i}{i!}e^{-at}, i = 1, 2, \ldots,$$

i.e., $p_i(a)$ is the probability function of a Poisson distribution with parameter a, and $E_{i,a}(t)$ is the CDF of the Erlang distribution with a shape parameter i and a scale parameter a. We also assume that $N > 2$, $\mu \neq \alpha$, and $\alpha + \mu_v \neq \mu$. If any of these assumptions are not satisfied, the model is simplified, and those cases are not taken into consideration.

The concept of potential random walk is introduced in [29] as a tool for the analysis of compound Poisson processes. In particular, it is proven in [29] (see also [30]) that each solution of the infinite-size system of linear equations of the form

$$\sum_{k=-1}^{n-2} a_{k+1}x_{n-k} - x_n = \theta_n, \tag{1}$$

where $n \geq 2$, can be written as

$$x_n = MR_{n-1} + \sum_{k=2}^{n} R_{n-k}\phi_k, \quad n \geq 2, \tag{2}$$

where $M \in \mathbb{R}$, and the sequence (R_k) is defined as indicated below.

Consider the following generating functions:

$$r(\theta) \overset{def}{=} \sum_{k=0}^{\infty} \theta^k R_k \tag{3}$$

and

$$a(\theta) \overset{def}{=} \sum_{k=0}^{\infty} \theta^k a_k, \tag{4}$$

where $|\theta| < 1$.

It can be shown ([29,30]) that the following representation is true:

$$r(\theta) = \frac{\theta}{a(\theta) - \theta}.$$ (5)

As a consequence, applying Maclaurin's expansion, we obtain

$$R_k = \frac{1}{k!} \frac{d^k}{d\theta^k} \left(\frac{\theta}{a(\theta) - \theta} \right) \Big|_{\theta=0}, \quad k \geq 1.$$ (6)

Equivalently (see [29,30]), the sequence (R_k) can be defined in a recursive way as follows:

$$R_0 = 0, \quad R_1 = (a_0)^{-1},$$
$$R_k = R_1 \left(R_{k-1} - \sum_{j=0}^{k-1} a_{j+1} R_{k-1-i} \right),$$ (7)

where $k \geq 2$.

3. Transient Equations for the Time to the Buffer Overflow Distribution

3.1. Server Starting in Normal Mode

When the buffer is empty and in normal mode upon opening, we have

$$T_0(t) = \int_0^t T_1(t - x) \mathrm{d}F(x).$$ (8)

If the first arrival occurs before t (at some epoch x), then the probability that the buffer overflows before t can be expressed using the probability of the buffer overflow before $t - x$, given that there is one job present at the beginning, since the system renews at x and behaves like it has just started with one job present. If no new jobs enter the system before t, then there will clearly be no buffer overflow up to t.

For $1 \leq n < N$, the following is true:

$$T_n(t) = S_1(t, n) + S_2(t, n).$$ (9)

The summand $S_1(t, n)$ stands for the case of a new arrival before epoch t, and $0 \leq i < n$ jobs are finished before this event. Therefore, the system renews with $n - i + 1$ jobs present. This summand can be expressed by

$$S_1(t, n) = \int_0^t \sum_{i=0}^{n-1} p_i(\mu x) T_{n-i+1}(t - x) \mathrm{d}F(x).$$ (10)

The second summand, $S_2(t, n)$, results from the case where all jobs are finished before the new one arrives. This means that the server must change its operation mode to WV. Then, with the probability $1 - e^{-\alpha(x-y)}$, the system will switch back to the normal mode before the first arrival. Otherwise, the system will renew in WV mode:

$$S_2(t, n) = \int_0^t \mathrm{d}F(x) \int_0^x \left[\left(1 - e^{-\alpha(x-y)} \right) T_1(t - x) + e^{-\alpha(x-y)} T_1^v(t - x) \right] \mathrm{d}E_{n,\mu}(y).$$ (11)

The last considered case is $n = N$. It is clear that if N customers are present at the beginning, then the buffer overflow before any time instant t is a certain event; therefore,

$$T_N(t) = 1.$$ (12)

3.2. Server Starting in WV Mode

If the server starts in WV mode with no jobs, we have

$$T_0^v(t) = \int_0^t \left[\left(1 - e^{-\alpha x}\right) T_1(t - x) + e^{-\alpha x} T_1^v(t - x) \right] dF(x). \tag{13}$$

For $1 \le n < N$, the CDF $T_n^v(t)$ satisfies

$$T_n^v(t) = \sum_{i=1}^5 S_i^v(t, n). \tag{14}$$

In the first two summands, the case of the system still being in WV mode when the new job enters is considered. Allowing for the fact that not all of the customers were served before the first arrival, we have

$$S_1^v(t, n) = \int_0^t e^{-\alpha x} \sum_{i=0}^{n-1} p_i(\mu_v x) T_{n-i+1}^v(t - x) dF(x), \tag{15}$$

and, given all of the jobs were finished before this arrival, we conclude that

$$S_2^v(t, n) = \int_0^t e^{-\alpha x} E_{n,\mu_v}(x) T_1^v(t - x) dF(x). \tag{16}$$

For the remaining summands, we assume that the WV period ends before the first arrival epoch. Hence, we need to take into account the number of customers served in WV (i) and in normal mode (j).

If $i + j < n$, the following expression is obtained:

$$S_3^v(t, n) = \int_0^t dF(x) \int_0^x \alpha e^{-\alpha y} \sum_{i=0}^{n-1} p_i(\mu_v y) \sum_{j=0}^{n-i-1} p_j(\mu(x - y)) T_{n-i-j+1}(t - x) dy. \tag{17}$$

When $i = n$ at the arrival epoch, the system is operating in normal mode, since the system empties not in normal but in the WV period, and thus no new WV period is initialized; therefore,

$$S_4^v(t, n) = \int_0^t dF(x) \int_0^x \alpha e^{-\alpha y} E_{n,\mu_v}(y) T_1(t - x) dy. \tag{18}$$

If the last job is finished after switching to normal mode, the system starts a new WV period, we need to consider both cases of normal and WV mode at the arrival instant, and we can write

$$
\begin{aligned}
S_5^v(t, n) = &\int_0^t dF(x) \int_0^x \alpha e^{-\alpha y} \sum_{i=0}^{n-1} p_i(\mu_v y) \\
&\cdot \int_0^{x-y} \left[\left(1 - e^{-\alpha(x-(y+u))}\right) T_1(t - x) + e^{-\alpha(x-(y+u))} T_1^v(t - x) \right] dE_{n-i,\mu}(u) dy.
\end{aligned} \tag{19}
$$

For $n = N$, we have

$$T_N^v(t) = 1. \tag{20}$$

4. Solution of the System of Equations for $T_n(T)$ and $T_n^V(T)$

In this section, the system (8)–(20) is solved. It is divided into two subsections depending on the state of the server at the beginning. First, the solutions of (8) and (9) are found. Next, the solutions of (13) and (14) are explicitly obtained and introduced to

the former, which results in an explicit solution for the Laplace transforms of $T_i(t)$ and $T_i^v(t)$, $i = 0, \ldots, N - 1$.

4.1. Solution for the Normal Mode

Let us denote

$$\tilde{T}_n(s) = \int_0^\infty e^{-st} T_n(t) dt, \quad \tilde{F}(s) = \int_0^\infty e^{-st} dF(t).$$

Additionally, we introduce the following notation:

$$a_i(s) = \int_0^\infty e^{-st} p_i(\mu t) dF(t), \quad b_i(s) = \int_0^\infty e^{-st} E_{i,\mu}(t) dF(t),$$

$$c_i(s) = \left(\frac{\mu}{\mu - \alpha}\right)^i \int_0^\infty e^{-t(s+\alpha)} E_{i,\mu-\alpha}(t) dF(t).$$

Now, the Laplace transform of the system (8)–(12) can be written in the following form:

$$\tilde{T}_0(s) = \tilde{T}_1(s) \tilde{F}(s) \tag{21}$$

$$\tilde{T}_n(s) = \sum_{i=0}^{n-1} \tilde{T}_{n-i+1}(s) a_i(s) + \tilde{T}_1(s)(b_n(s) - c_n(s)) + \tilde{T}_1^v(s) c_n(s),$$

$$n = 1, \ldots N - 1 \tag{22}$$

$$\tilde{T}_N(s) = \frac{1}{s} \tag{23}$$

If we denote

$$\phi_n(s) = \tilde{T}_1(s)(c_n(s) - b_n(s)) - \tilde{T}_1^v(s) c_n(s), \tag{24}$$

we can rewrite system (22) for $n = 1, \ldots, N - 1$ in the form

$$\sum_{i=-1}^{n-2} a_{i+1}(s) \tilde{T}_{n-i}(s) - \tilde{T}_n(s) = \phi_n(s). \tag{25}$$

We can observe that the former system has the same form as system (1), and thus the solution can be obtained using (2), which leads to the following representation:

$$\tilde{T}_n(s) = M(s) R_{n-1}(s) + \sum_{i=2}^{n} R_{n-i}(s) \phi_i(s), \ n \geq 2, \tag{26}$$

where $R_k(s)$ is a sequence defined as follows (see (7)):

$$R_0(s) = 0, \ R_1(s) = a_0^{-1}(s), \ R_{k+1}(s) = R_1(s)\left(R_k(s) - \sum_{i=0}^{k} a_{i+1}(s) R_{k-i}(s)\right),$$

and $M(s)$ is some unknown function.

Taking $n = 1$ in (22) and $n = 2$ in (26), we can derive $M(s)$ as

$$M(s) = \tilde{T}_1(s)(1 - b_1(s) + c_1(s)) - \tilde{T}_1^v(s) c_1(s). \tag{27}$$

Now, the solution (26) can be rewritten in the form:

$$\tilde{T}_n(s) = \tilde{T}_1(s)\left(R_{n-1}(s) + \sum_{i=1}^{n} R_{n-i}(s)(c_i(s) - b_i(s))\right)$$
$$- \tilde{T}_1^v(s)\sum_{i=1}^{n} R_{n-i}(s)c_i(s), \ n \geq 2. \tag{28}$$

Combining (22) and (28) for $n = N - 1$ yields

$$\tilde{T}_1(s) = \frac{K(s)}{L(s)}\tilde{T}_1^v(s) + \frac{a_0(s)}{sL(s)}, \tag{29}$$

where

$$K(s) = \sum_{i=1}^{N-2}\left(R_{N-i-1}(s)c_i(s) - a_i(s)\sum_{j=1}^{N-i} R_{N-i-j}(s)c_j(s)\right) + c_{N-1}(s), \tag{30}$$

$$L(s) = \sum_{i=1}^{N-2}\left(R_{N-i-1}(s)(c_i(s) - a_i(s) - b_i(s)) - \sum_{j=1}^{N-i} R_{N-i-j}(s)\big(c_j(s) - b_j(s)\big)\right)$$
$$+ R_{N-2}(s) + c_{N-1}(s) - b_{N-1}(s). \tag{31}$$

Using the former expression in (28), we can state the solution (depending on $\tilde{T}_1^v(s)$) in the form

$$\tilde{T}_n(s) = \tilde{T}_1^v(s)\left(\frac{K(s)}{L(s)}R_{n-1}(s) + \sum_{i=1}^{n} R_{n-i}(s)\left(\frac{K(s)}{L(s)}(c_i(s) - b_i(s)) - c_i(s)\right)\right)$$
$$+ \frac{a_0(s)}{sL(s)}\left(R_{n-1}(s) + \sum_{i=1}^{n} R_{n-i}(c_i(s) - b_i(s))\right). \tag{32}$$

4.2. Solution for the WV Mode

In this subsection, the solution for the case of the server starting in WV mode is obtained.

The Laplace transform of the system (13)–(20) is given by

$$\tilde{T}_0^v(s) = \tilde{T}_1(s)\big(\tilde{F}(s) - \tilde{F}(s + \alpha)\big) + \tilde{T}_1^v(s)\tilde{F}(s + \alpha), \tag{33}$$

$$\tilde{T}_n^v(s) = \sum_{i=0}^{n-1} \tilde{T}_{n-i+1}^v A_i(s) + \tilde{T}_1^v(s)B_n(s)$$
$$+ \sum_{i=0}^{n-1}\sum_{j=0}^{n-i-1} \tilde{T}_{n-i-j+1}^v(s)C_{ij}(s) + \tilde{T}_1(s)D_n(s), \ n = 1, \ldots, N - 1, \tag{34}$$

$$\tilde{T}_N^v(s) = \frac{1}{s}, \tag{35}$$

where

$$A_i(s) = \int_0^\infty e^{-t(s+\alpha)} p_i(\mu_v t) dF(t), \tag{36}$$

$$B_i(s) = \int_0^\infty e^{-t(s+\alpha)} E_{i,\mu_v}(t) dF(t)$$
$$+ \alpha \sum_{j=0}^{i-1} \left(\frac{\mu}{\mu-\alpha}\right)^{i-j} \int_0^\infty e^{-t(s+\alpha)} \int_0^t p_j(\mu_v y) E_{i-j,\mu-\alpha}(t-y) dy\, dF(t), \tag{37}$$

$$C_{ij}(s) = \alpha \int_0^\infty e^{-st} \int_0^t e^{-\alpha y} p_i(\mu_v y) p_j(\mu(t-y)) dy dF(t), \tag{38}$$

$$D_i(s) = \alpha \int_0^\infty e^{-st} \int_0^t e^{-\alpha y} E_{i,\mu_v}(y) dy\, dF(t) + \alpha \int_0^\infty e^{-st} \int_0^t p_j(\mu_v y)$$
$$\cdot \sum_{j=0}^{i-1} \left(e^{-\alpha y} E_{i-j,\mu}(t-y) - \left(\frac{\mu}{\mu-\alpha}\right)^{i-j} E_{i-j,\mu-\alpha}(t-y)\right) dy\, dF(t). \tag{39}$$

The solutions (29) and (32) can be introduced into (34), which leads to the following form:

$$\tilde{T}_n^v(s) = \sum_{i=0}^{n-1} \tilde{T}_{n-i+1}^v(s) A_i(s) + \tilde{T}_1^v(s) G_n(s) + H_n(s), \tag{40}$$

$$\tilde{T}_N^v(s) = \frac{1}{s}, \tag{41}$$

where (we set $l = n - i - j$ for readability):

$$G_n(s) = B_n(s) + \frac{K(s)}{L(s)} D_n(s) + \sum_{i=0}^{n-1} \sum_{j=0}^{n-i-1} C_{ij}(s)$$
$$\cdot \left(\frac{K(s)}{L(s)} R_l(s) + \sum_{k=1}^l R_{l-k+1}(s) \left(\frac{K(s)}{L(s)}(c_k(s) - b_k(s)) - c_k(s)\right)\right) \tag{42}$$
$$- \mathcal{I}\{n = N-1\} C_{00}(s) \left(\frac{K(s)}{L(s)} R_{N-1}(s) + \sum_{k=1}^{N-1} R_{N-k}(s) \left(\frac{K(s)}{L(s)}(c_k(s) - b_k(s)) - c_k(s)\right)\right),$$

$$H_n(s) = \frac{a_0(s)}{sL(s)}$$
$$\cdot \left(D_n(s) + \sum_{i=0}^{n-1} \sum_{j=0}^{n-i-1} C_{ij}(s) \left(R_l(s) + \sum_{k=1}^{l+1} R_{l-k+1}(s)(c_k(s) - b_k(s))\right)\right) \tag{43}$$
$$- \mathcal{I}\{n = N-1\} C_{00}(s) \left(R_{N-1}(s) + \sum_{k=1}^{N-1} R_{N-k}(s)(c_k(s) - b_k(s))\right),$$

for $n < N - 1$, where $\mathcal{I}\{\cdot\}$ is an indicator function. The system (40) can be solved with the approach used in previous subsection. It can be rewritten in the following form (compare (1)):

$$\sum_{i=-1}^{n-2} A_{i+1}(s) \tilde{T}_{n-i}^v(s) - \tilde{T}_n^v(s) = \Phi_n(s), \ \ n = 1, 2, \ldots, N-1, \tag{44}$$

where $\Phi_n(s) = -\tilde{T}_1^v(s) G_n(s) - H_n(s)$.

Now, the solution is found in the form (see (2)):

$$\tilde{T}_n^v(s) = M^v(s) R_{n-1}^v(s) + \sum_{i=2}^n R_{n-i}^v(s) \Phi_i(s), \ \ n \geq 2, \tag{45}$$

where $M^v(s)$ is some function, and $R_n^v(s)$ is a sequence:

$$R_0^v(s) = 0, \quad R_1^v(s) = A_0^{-1}(s), \quad R_{k+1}^v(s) = R_1^v(s)\left(R_k^v(s) - \sum_{i=0}^{k} A_{i+1}(s)R_{k-i}^v(s)\right).$$

Comparing (40) for $n = 1$ with (45) for $n = 2$, we find

$$M^v(s) = \tilde{T}_1^v(s)(1 - G_1(s)) - H_1(s). \tag{46}$$

Introducing $M^v(s)$ to (45), we can write

$$\tilde{T}_n^v(s) = \tilde{T}_1^v(s)\left(R_{n-1}^v(s) - \sum_{i=1}^{n} R_{n-i}^v(s)G_i(s)\right) - \sum_{i=1}^{n} R_{n-i}^v(s)H_i(s). \tag{47}$$

Now, $\tilde{T}_1^v(s)$ can be obtained taking (47) for $n = N - 1$ and comparing it to $\tilde{T}_{N-1}^v(s)$ from (40) (after introducing the solutions $\tilde{T}_k^v(s)$ for $k < N - 1$). Finally, we obtain the explicit solution for $T_1^v(s)$:

$$T_1^v(s) = \frac{V(s)}{W(s)}, \tag{48}$$

where

$$V(s) = \frac{A_0(s)}{s} + H_{N-1}(s) - \sum_{i=1}^{N-2} A_i(s) \sum_{j=1}^{N-i} R_{N-i-j}^v(s)H_j(s) + \sum_{i=1}^{N-1} R_{N-i-1}^v(s)H_i(s), \tag{49}$$

$$W(s) = \sum_{i=1}^{N-2}\left(A_i(s)\sum_{j=1}^{N-i} R_{N-i-j}^v(s)G_j(s) - R_{N-i-1}^v(s)(G_i(s) + A_i(s))\right) + R_{N-2}^v(s) - G_{N-1}(s). \tag{50}$$

Utilizing (48) in (29), (32) and (47), the solution for the Laplace transforms $\tilde{T}_i(s)$ and $\tilde{T}_i^v(s)$ for $i = 0, \ldots, N - 1$ can be expressed as a function of the input parameters.

5. Numerical Examples

In this section, numerical examples are presented, and the impact of the model parameters is investigated. The CDFs $T_k(t)$ and $T_k^v(t)$ were computed using the Gaver–Wynn rho method of numerical Laplace transform inversion (see [31]).

Let us consider a $GI/M/1/N$ model with single working vacations with $N = 30$. We introduce the following notation for random variables X_i (interarrival times in the considered model):

$$X_1 \sim Exponential(15),$$
$$X_2 \sim Uniform(0, 2/15),$$
$$X_3 \sim Pareto(0.05, 4),$$
$$X_4 \sim Gamma(1/3, 1/5),$$
$$X_5 \sim Weibull(2, 0.0752),$$

and with $F_i(t)$, we denote the CDF of the corresponding variable X_i. For all of these distributions, we have $E(X_i) \approx 1/15$, and therefore, in all cases, the arrival intensity is $\lambda \approx 15$. These distributions are used through the examples.

The parameter λ can be interpreted as the number of packets arriving to the node per second. If a single packet has size 100 B, then $\lambda = 1500$ B/s. Similarly, the service intensity can be converted. This way, the mean time spent in WV mode $1/\alpha$ is expressed in seconds.

Example 1. *The interarrival times are independent, exponentially distributed random variables. The service speed is $\mu = 19$ in normal and $\mu_v = 9$ in WV mode.*
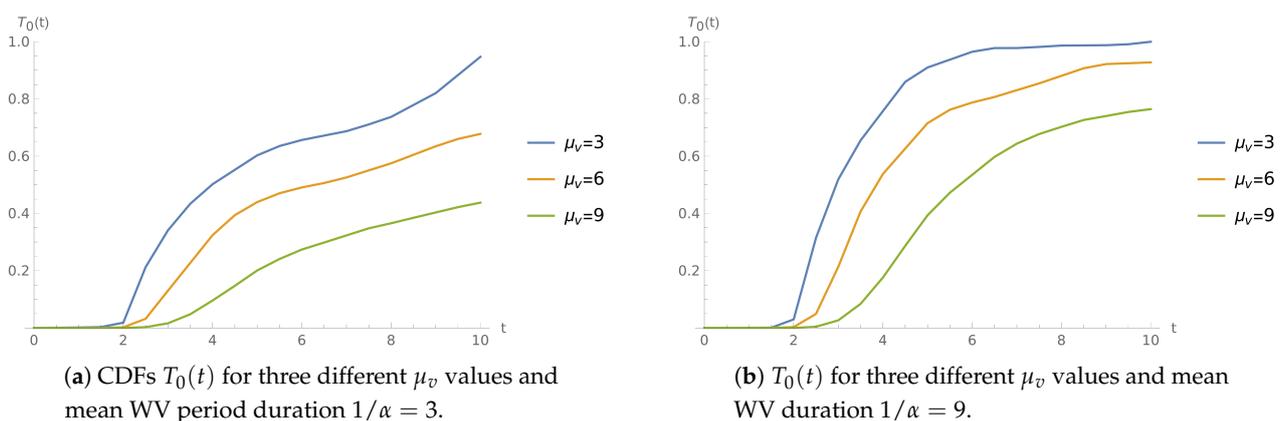
Figure 1 shows the pairs of CDFs $T_0(t)$ and $T_0^v(t)$ for the mean WV period duration $1/\alpha = 3$ (Figure 1a) and $1/\alpha = 9$ (Figure 1b). As we could expect, the $T_0^v(t)$ values tend to be higher than those of $T_0(t)$. In addition, we can observe that the growth is faster in the case of longer WV.

(a) $T_0(t)$ and $T_0^v(t)$ for the mean WV duration $1/\alpha = 3$.

(b) $T_0(t)$ and $T_0^v(t)$ for mean WV duration $1/\alpha = 9$.
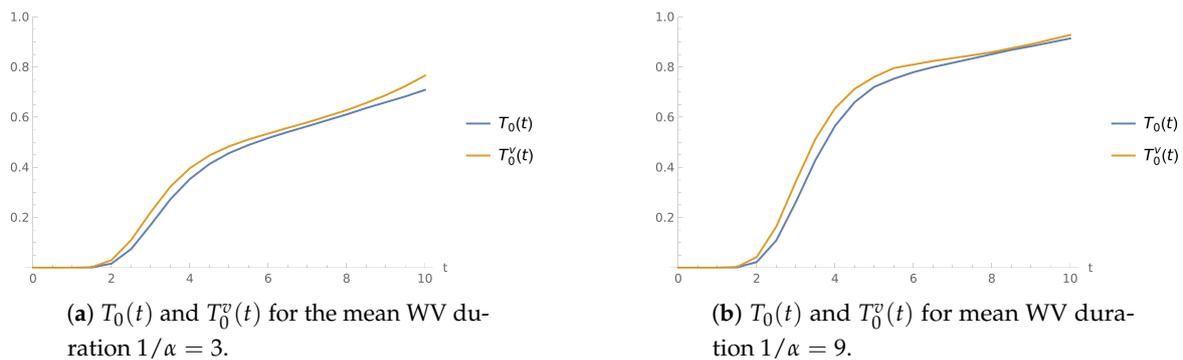
Figure 1. The time to buffer overflow CDFs for $\mu = 19$ and $\mu_v = 9$.

Example 2. *The interarrival times are uniformly distributed. The service speed in normal mode is $\mu = 19$.*

Figure 2 shows the CDFs $T_0(t)$ for three different values of WV service rate in case of shorter ($1/\alpha = 3$, Figure 2a) and longer ($1/\alpha = 9$, Figure 2b) WV periods. As expected, the values of $T_0(t)$ are greater for lower intensities μ_v. In addition, we can observe that they grow as the expected WV duration increases.

(a) CDFs $T_0(t)$ for three different μ_v values and mean WV period duration $1/\alpha = 3$.

(b) $T_0(t)$ for three different μ_v values and mean WV duration $1/\alpha = 9$.

Figure 2. The time to buffer overflow CDFs for $\mu = 19$.

Example 3. *The interarrival times follow a Weibull distribution. The service rates are $\mu = 21$ and $\mu_v = 9$ in normal and WV mode, respectively.*

The visualization of the impact of mean WV duration is shown in Figure 3. Clearly, the WV duration parameter strongly affects the length of the period with no packet losses.
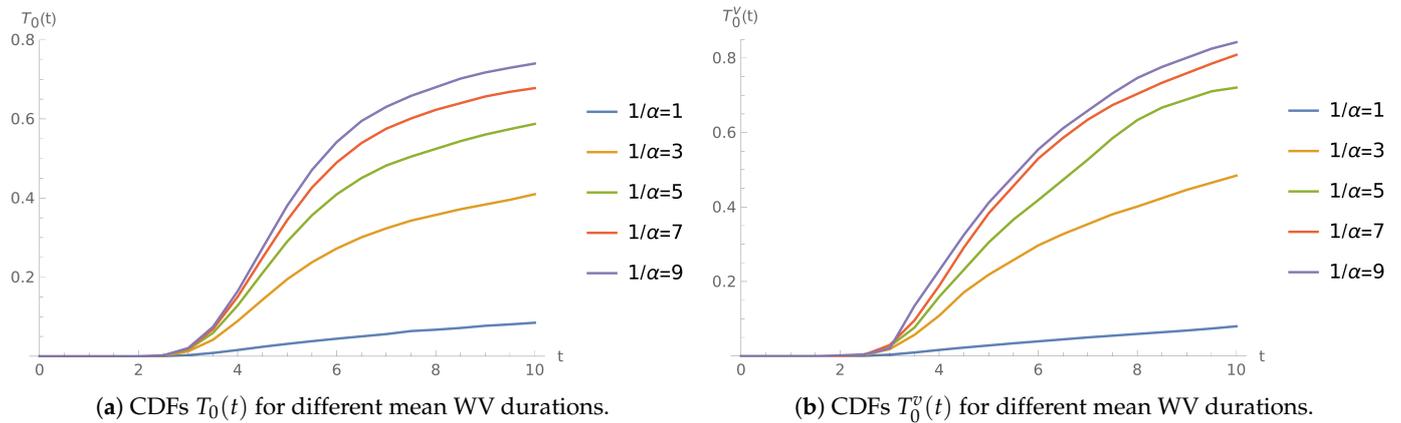
(**a**) CDFs $T_0(t)$ for different mean WV durations.

(**b**) CDFs $T_0^v(t)$ for different mean WV durations.

Figure 3. The time to buffer overflow CDFs for $\mu = 21$.

Example 4. *Interarrival times follow a Pareto distribution. The normal service speed is $\mu = 17$ and is reduced to $\mu_v = 9$ during WV.*

In Figure 4, we can see that, for the longer WV periods, if the server starts in normal mode, the CDF $T_0(t)$ grows faster than $T_{10}(t)$ and $T_{20}(t)$ (Figure 4c). Of course, if the buffer is initially empty, the server may enter the WV period sooner, changing the workload from $\rho \approx 0.88$ to $\rho_v \approx 1.67$. When the server is initially in WV mode (Figure 4b,d), the results seem more natural, i.e., the more packets that are initially present, the more probable that buffer overflow occurs before some time epoch t.
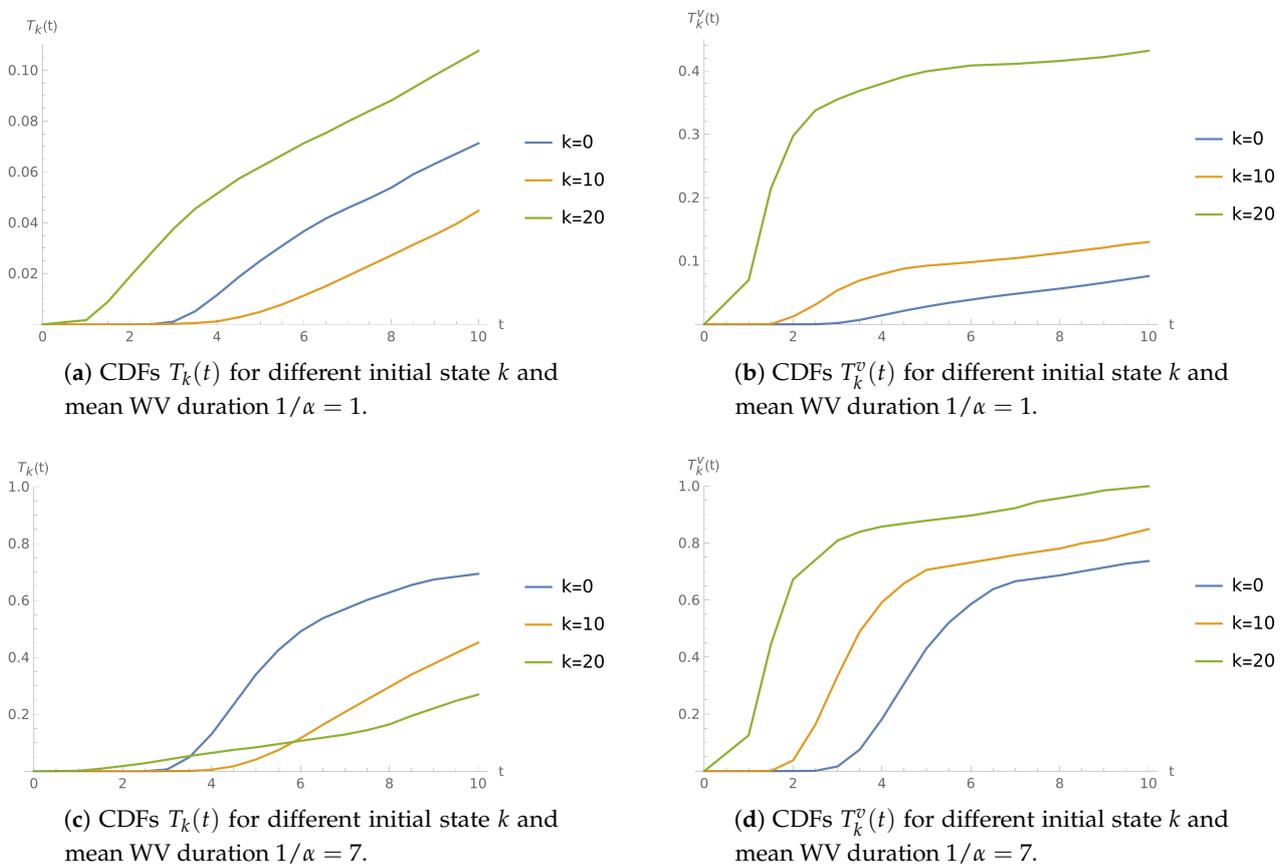

(**a**) CDFs $T_k(t)$ for different initial state k and mean WV duration $1/\alpha = 1$.

(**b**) CDFs $T_k^v(t)$ for different initial state k and mean WV duration $1/\alpha = 1$.

(**c**) CDFs $T_k(t)$ for different initial state k and mean WV duration $1/\alpha = 7$.

(**d**) CDFs $T_k^v(t)$ for different initial state k and mean WV duration $1/\alpha = 7$.

Figure 4. The time to buffer overflow CDFs for $\mu = 17$ and $\mu_v = 9$.

Example 5. *The arrival intensity is* $\lambda \approx 15$. *The incoming packets are processed with rate* $\mu = 21$ *in normal and* $\mu_v = 12$ *in WV mode.*

Figure 5 shows the CDFs $T_0(t)$ for different interarrival time distributions. For the gamma distributed interarrival times, the buffer tends to overflow sooner, compared to the other distributions taken into consideration. On the other side, the probability of overflow before t is the lowest, when the interarrival times follow a Pareto distribution. Note that X_4 has the highest variance and X_3 the lowest.
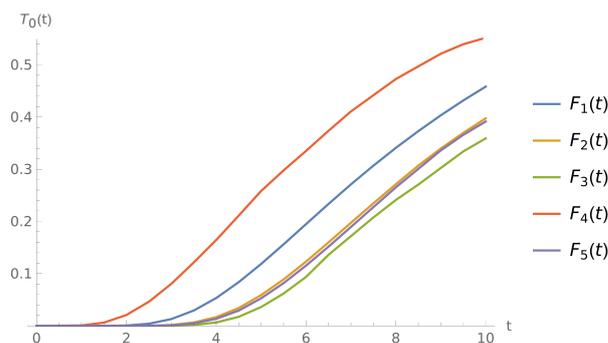
Figure 5. The CDFs $T_0(t)$ for $\mu = 21$, $\mu_v = 12$, $1/\alpha = 9$, and different types of input stream.

Example 6. *The arrival intensity is* $\lambda \approx 15$. *The server works with intensity* $\mu_v = 9$ *in WV mode, and the mean time spent in WV is* $1/\alpha = 9$.

In this example, we investigate the impact of μ on the expected value of the time to the first buffer overflow. As we can see in Figure 6, the expected time to the buffer overflow grows with μ for the lower range of μ, and then starts to decrease. This behaviour is linked to the fact that, as μ grows, the WV period occurrences tend to be more frequent. Another interesting observation is that, when the server starts in normal mode with 20 jobs, the mean time to the buffer overflow is greater than when it is initially empty (except for the case of $\mu < \lambda$). Note that if the server starts to empty, the normal working period will end sooner; therefore, this behaviour is not surprising.

For $k = 20$, we can observe that μ has a tenuous impact on the analysed characteristics for the server initializing its operation in WV mode. Presumably, the server would not complete the WV period before buffer overflow. Comparing the results for different interarrival distributions, namely, exponential (Figure 6a,b), gamma (Figure 6c,d), and uniform distribution (Figure 6e,f), we can draw an analogous conclusion as in the former example. For gamma distribution, the plotted expected values tend to be the lowest in the whole μ range considered.

Example 7. *The arrival intensity is* $\lambda \approx 15$. *The jobs are processed with rate* $\mu = 17$ *in normal and* $\mu_v = 9$ *in WV mode. The mean WV duration is* $1/\alpha = 7$.

The goal of this analysis is to validate the numerical results. Figure 7 contains the plots of CDFs $T_0(t)$ and $T_0^v(t)$ juxtaposed with corresponding simulated values in the case of interarrival times following uniform (Figure 7a,b), Pareto (Figure 7c,d), and Weibull distribution (Figure 7e,f). As one can note, the simulation results fit in with numerical results obtained by adopting the method described in this paper, which validate the correctness of the obtained formulae.
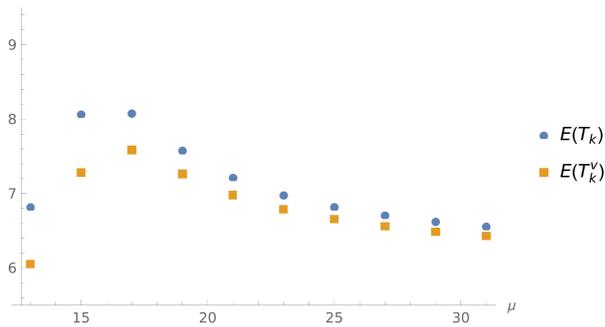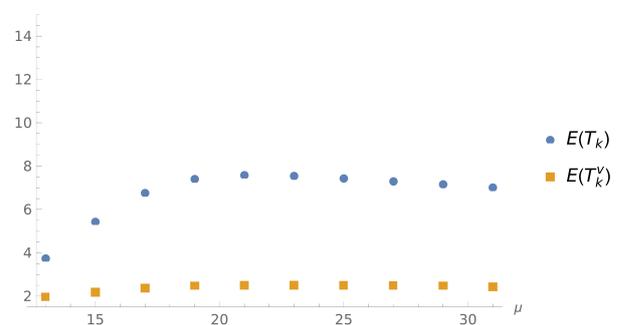
(**a**) Mean time to overflow for exponentially distributed interarrival times and $k = 0$.
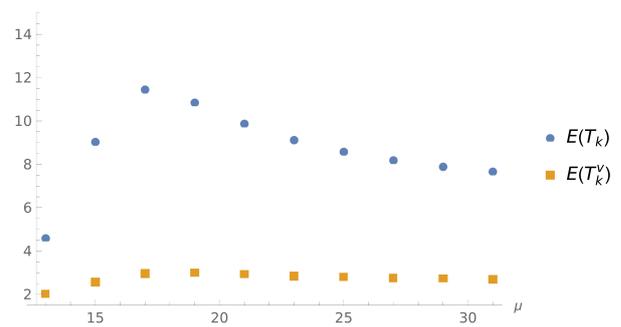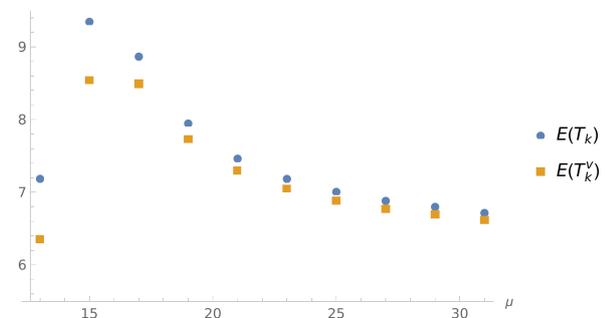
(**b**) Mean time to overflow for exponentially distributed interarrival times and $k = 20$.
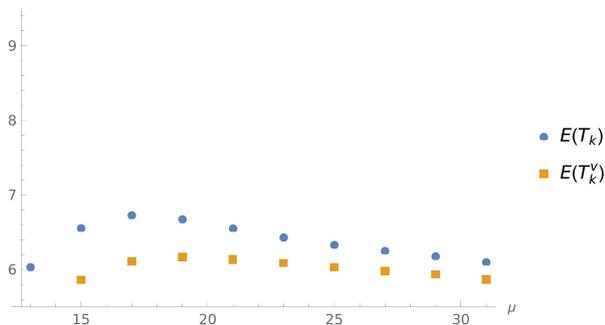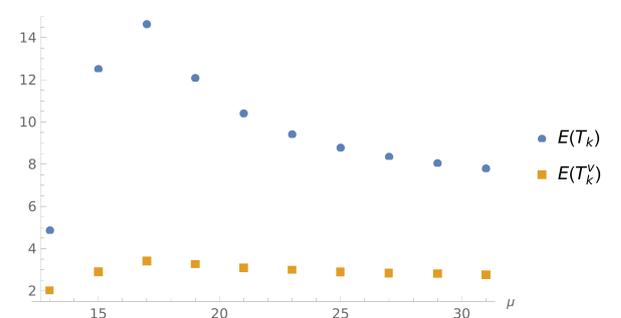
(**c**) Mean time to overflow for gamma distributed interarrival times and $k = 0$.

(**d**) Mean time to overflow for gamma distributed interarrival times and $k = 20$.
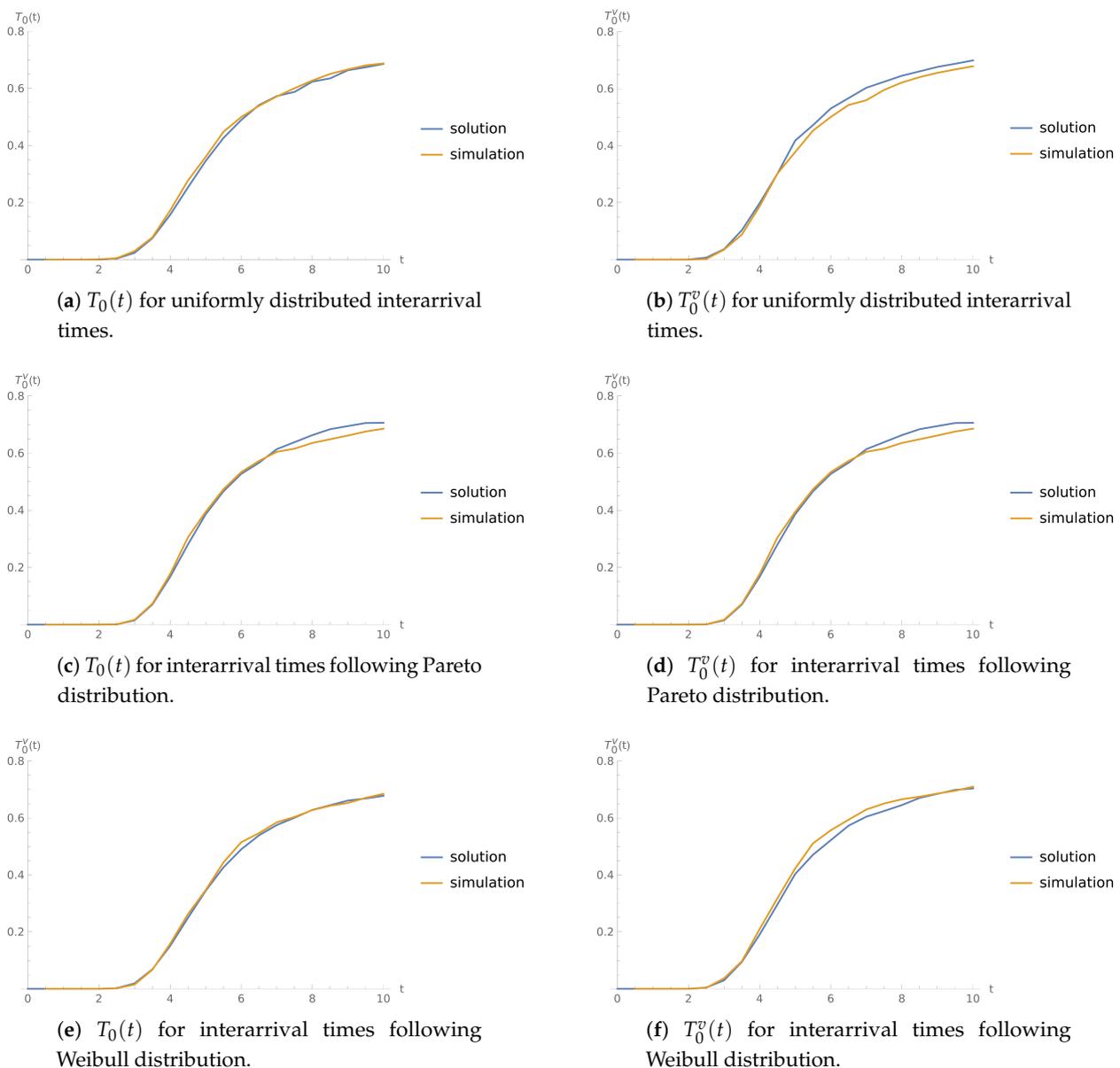
(**e**) Mean time to overflow for uniformly distributed interarrival times and $k = 0$.

(**f**) Mean time to overflow for uniformly distributed interarrival times and $k = 20$.

Figure 6. Mean value of the time to buffer overflow for different interarrival distributions with $\lambda \approx 15$, $\mu_v = 9$, and $1/\alpha = 9$.

(**a**) $T_0(t)$ for uniformly distributed interarrival times.

(**b**) $T_0^v(t)$ for uniformly distributed interarrival times.

(**c**) $T_0(t)$ for interarrival times following Pareto distribution.

(**d**) $T_0^v(t)$ for interarrival times following Pareto distribution.

(**e**) $T_0(t)$ for interarrival times following Weibull distribution.

(**f**) $T_0^v(t)$ for interarrival times following Weibull distribution.

Figure 7. CDFs of the time to buffer overflow for $\mu = 17$, $\mu_v = 9$, and $1/\alpha = 7$. Method results and simulated values.

6. Conclusions

We investigated a finite-capacity queueing model with an independent general input flow, exponential service times, and a single working vacation policy. Applying an analytic approach based on the idea of an embedded Markov chain and a continuous version of the total probability law and linear algebra, the closed form representations for Laplace transforms of the time to the first buffer overflow were found for the system starting operation in both normal and working vacation mode. A detailed numerical study was conducted in which the impact of the key system parameters was analysed, such as the type of probability distribution of the interarrival times, service speeds, and single working vacation duration on the time to buffer overflow distribution.

The considered queueing system has potential applications in the modelling of energy saving modes in wireless network nodes. Energy savings can be obtained by temporarily reducing the service speed. This approach can help to reduce the latency and packet loss ratio compared to the simple vacation policy and N-policy models. The influence of the model parameters on energy usage is a subject for future research.

References

1. Levy, Y.; Yechiali, U. Utilization of the idle period in an M/G/1 queueing systems. *Manag. Sci.* **1975**, *22*, 202–211. [CrossRef]
2. Tian, N.; Zhang, Z.G. *Vacation Queueing Models Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2006.
3. Doshi, B. Queueing systems with vacations—A survey. *Queueing Syst.* **1986**, *1*, 29–66. [CrossRef]
4. Teghem, J. Control of the service process in a queueing system. *Eur. J. Oper. Res.* **1986**, *23*, 141–158. [CrossRef]
5. Servi, L.D.; Finn, S.G. M/M/1 queues with working vacations (M/M/1/WV). *Perform. Eval.* **2002**, *50*, 41–52. [CrossRef]
6. Boutoumi, B.; Gharbi, N. Two Thresholds Working Vacation Policy for Improving Energy Consumption and Latency in WSNs. In Proceedings of the Queueing Theory and Network Applications (QTNA 2018), Tsukuba, Japan, 25–27 July 2018; pp. 168–181.
7. Yu, X.; Ma, Z.; Guo, S.; Chen, L. Performance analysis of wireless communication networks with threshold activation process and interference signals. *Int. J. Commun. Netw. Distrib. Syst. IJCNDS* **2020**, *25*, 78–94. [CrossRef]
8. Sahoo, C.N.; Goswami, V. Cost and energy optimisation of cloud data centres through dual VM modes—Activation and passivation. *Int. J. Commun. Netw. Distrib. Syst. IJCNDS* **2017**, *18*, 371–389.
9. Qin, B.; Jin, S.; Zhao, D. Energy-efficient virtual machine scheduling strategy with semi-sleep mode on the cloud platform. *Int. J. Innov. Comput. Inf. Control* **2019**, *15*, 337–349.
10. Zhang, M.; Hou, Z. M/G/1 queue with single working vacation. *J. Appl. Math. Comput.* **2012**, *39*, 221–234. [CrossRef]
11. Li, J.; Tian, N. Performance analysis of a GI/M/1 queue with single working vacation. *Appl. Math. Comput.* **2011**, *217*, 4960–4971. [CrossRef]
12. Ye, Q.; Liu, L. Performance analysis of the GI/M/1 queue with single working vacation and vacations. *Methods Comp. Appl. Prob.* **2017**, *19*, 685–714. [CrossRef]
13. Li, J.H.; Tian, N. Analysis of the discrete-time Geo/Geo/1 queue with single working vacation. *Qual. Technol. Quant. Manag.* **2008**, *5*, 77–89. [CrossRef]
14. Chae, K.C.; Lim, D.E.; Yang, W.S. The GI/M/1 queue and the GI/Geo/1 queue both with single working vacation. *Perform. Eval.* **2009**, *66*, 356–367. [CrossRef]
15. Gao, S.; Liu, Z. An M/G/1 queue with single working vacation and vacation interruption under Bernoulli schedule. *Appl. Math. Model.* **2013**, *37*, 1564–1579. [CrossRef]
16. Doo, H.L.; Bo, K.K. A note on the sojourn time distribution of an M/G/1 queue with a single working vacation and vacation interruption. *Oper. Res. Persp.* **2015**, *2*, 57–61.
17. Yang, D.-Y.; Chung, C.-H.; Wu, C.-H. Sojourn times in a Markovian queue with working breakdowns and delayed working vacations. *Comput. Ind. Eng.* **2021**, *156*, 107239. [CrossRef]
18. Barbhuiya, F.P.; Gupta, U.C. A Discrete-Time $GI^X/Geo/1$ Queue with Multiple Working Vacations Under Late and Early Arrival System. *Methodol. Comput. Appl. Probab.* **2020**, *22*, 599–624. [CrossRef]
19. Thilaka, B.; Poorani, B.; Udayabaskaran, S. Performance Analysis of an M/M/1 Queue with Single Working Vacation with Customer Impatience Subject to Catastrophe. In Proceedings of the International Conference on Mathematical Analysis and Computing (ICMAC 2019), Kalavakkam, India, 23–24 December 2019; pp. 271–292.
20. Vijaya Laxmi, P.; Rajesh, P.; Kassahun, T.W. Analysis of a variant working vacation queue with customer impatience and server breakdowns. *Int. J. Oper. Res.* **2021**, *40*, 437–459. [CrossRef]
21. Buchholz, P. A hybrid analysis approach for finite-capacity queues with general inputs and phase type service. *Queueing Syst.* **2000**, *35*, 167–183. [CrossRef]
22. Agarwal, M. Distribution of number served during a busy period of GI/M/1/N queues-lattice path approach. *J. Stat. Plan. Inference* **2002**, *101*, 7–21. [CrossRef]
23. Czachórski, T.; Grochla, K. Diffusion Approximation Models for Cloud Computations with Task Migrations. In Proceedings of the IEEE International Conference on Fog Computing (ICFC 2019), Prague, Czech Republic, 24–26 June 2019; pp. 27–30.
24. Kumar R.; Ghosh S.; Banik A.D. Numerical study on transient behaviour of finite bulk arrival or service queues with multiple working vacations. *Int. J. Math. Oper. Res.* **2021**, *18*, 384–403. [CrossRef]
25. Vijayashree, K.V.; Ambika, K. An M/M/1 Queueing Model Subject to Differentiated Working Vacation and Customer Impatience. In Proceedings of the International Conference on Computational Intelligence, Cyber Security, and Computational Models (ICC3 2019), Coimbatore, India, 19–21 December 2019; pp. 107–122.

26. Kempa, W.M.; Kobielnik, M. Transient solution for the queue-size distribution in a finite-buffer model with general independent input stream and single working vacation policy. *Appl. Math. Model.* **2018**, *59*, 614–628. [CrossRef]
27. Kempa, W.M. Analytical model of a wireless sensor network (WSN) node operation with a modified threshold-type energy saving mechanism. *Sensors* **2019**, *19*, 3114. [CrossRef] [PubMed]
28. Kempa, W.M. Probabilistic analysis of a buffer overflow duration in data transmission in wireless sensor networks. *Sensors* **2020**, *20*, 5772. [CrossRef]
29. Korolyuk, V.S. Boundary problems for a compound Poisson process. *Theory Prob. Appl.* **1974**, *19*, 1–13. [CrossRef]
30. Kempa, W.M. A comprehensive study on the queue-size distribution in a finite-buffer system with a general independent input flow. *Perform. Eval.* **2017**, *108*, 1–15. [CrossRef]
31. Valko, P.P.; Abate J. Comparision of Sequence Accelerators for the Gaver Method of Numerical Laplace Transform Inversion. *Comput. Math. Appl.* **2004**, *48*, 629–636. [CrossRef]