Operational Research (2024) 24:18
https://doi.org/10.1007/s12351-024-00827-8

ORIGINAL PAPER

Orbit while in service

Gabi Hanukov[1] · **Uri Yechiali**[2]

Received: 18 March 2023 / Revised: 23 January 2024 / Accepted: 29 February 2024

Abstract

In various real-life queueing systems, part of the service can be rendered without involvement or presence of the customers themselves. In those queues, customers whose service order is still in process may leave the service station, go to 'orbit' for a random length of time, and then return to find out if their order has been completed. Common examples are car's annual maintenance works, food ordering, etc. In this paper, a thorough analysis of a single-server 'orbit while in service' queueing model with general service time is presented. Assuming an Exponentially distributed orbit time, we derive general formulae for the distributions of (i) a customer's total residence time in the system; (ii) a customer's net actual residence time in the system during service (not including orbit time); (iii) the time an orbiting customer is late to return, i.e., remains in orbit after his/her service has been completed; and (iv) the total number of customers in the system. Considering the family of Gamma-distributed service times (spanning the range of distributions between the Exponential and the Deterministic), as well as the Uniform distribution, we further derive explicit formulae for the distributions of the above variables. Under linear cost assumptions, the optimal mean orbit time is numerically calculated for each of the above service-time distributions. Figures depicting the behavior of the measures as functions of the parameters are presented.

Keywords Stochastic processes · Queueing · Impatient customers · Orbit

Mathematics Subject Classification 60K25 · 90B22

✉ Gabi Hanukov
gabih@ariel.ac.il

Uri Yechiali
uriy@tauex.tau.ac.il

1 Department of Industrial Engineering & Management, Ariel University, Ariel, Israel

2 Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, 6997801 Tel Aviv, Israel

Published online: 05 April 2024

1 Introduction

A common queueing system is the following: customers arrive at a service station and place an order. Part (or all) of the service may be conducted without the customer's presence or involvement. While service is rendered, a customer may become impatient and leave the system before the order is completed, only to return later, after a random (orbit) time. Upon return, a customer collects his/her order, if completed, or waits in the premises until the service is concluded. One example of this "orbit while in service" phenomenon is seen in food stores. Another example is an annual car maintenance, followed by a Department of Motor Vehicles inspection. In almost all countries, all private cars are required to pass an annual mechanical inspection. Before passing the inspection, the car owner usually drives his/her car to a service station where the car undergoes maintenance procedures. This latter process is usually lengthy. If the maintenance process extends beyond the customer's patience time, s/he leaves the station and goes elsewhere for a random "orbit" time. If the car is ready when the owner returns from orbit, the customer pays for the service and drives away. However, if the maintenance operations have not been completed, the customer waits on the premises until service is completed and only then leaves.

Such an "orbit while in service" policy allows customers to use their time efficiently so that rather than staying idle in the system during the service execution, they can perform other tasks while spending time "in orbit." For example, customers can shop while their car undergoes maintenance; alternatively, they can spend the time relaxing in a coffee shop. Many researchers state that a customer's satisfaction in service systems is significantly and implicitly affected by the time they physically spend in the system (see, e.g., Polas et al. 2018). Thus, by adopting the proposed policy and allowing, even encouraging, customers to use an orbit option while their service is in process, service system managers would gain a competitive advantage and increased demand. Despite the potential benefits of using the "orbit while in service" policy, this branch of queueing systems is barely investigated in the literature. To the best of our knowledge, only the current work and (Hanukov 2023) addressed such a policy. The latter work assumed a Markovian system and investigated the inventory of stored ready services. In the current work, we study the "orbit while in service" policy with a general service duration distribution and investigate the optimal time customers should orbit during their service.

Models of queueing systems with "impatient" customers who abandon the system before their service starts (or has been completed) have been widely investigated in the literature. For example, Bouchentouf et al. (2022) consider a finite population multi-server machine system with breakdowns, repairs, Bernoulli feedback, balking, reneging, and retention of reneged customers under multiple synchronous working vacations. The authors find the optimal system capacity, number of servers, and service rates during working vacations and regular busy periods by minimizing the system's total cost. Dong and Ibrahim (2021) investigate the shortest remaining processing time (SRPT) scheduling policy in multi-server queues with abandonment. The authors prove that the SRPT discipline asymptotically maximizes the system throughput among all scheduling disciplines. Bassamboo et al. (2023) study scheduling multi-class impatient customers in parallel server queueing systems. The authors use fluid approximations to analyze the multi-class scheduling problem and provide managers with ways to improve the quality of service to manage such systems. Further discussion on queueing systems with "impatient" customers is available in the literature (see e.g., Altman and Yechiali 2006; Yechiali 2007; Sherzer and Kerner 2018; Bouchentouf et al. 2021; Dong and Ibrahim 2021; Shajin and Krishnamoorthy 2021; Ayhan 2022; Firouz et al. 2022; Kumar et al. 2022; Yin et al. 2023; Cherfaoui et al. 2023; Manitz and Piehl 2023). In all these papers, customers abandon the system due to their impatience and relinquish receiving service, while in our work the customers go to orbit during the service execution and return to collect the completed service.

Also studied extensively are the so-called retrial models where blocked or delayed customers leave and "orbit" outside the system before returning to obtain service. For example, Dimitriou (2023) considers the single-server retrial queue with event-dependent arrival rates, investigates the impact of event dependency on performance measures and derives optimal joining probabilities. Zhang and Wang (2023) analyze retrial queueing systems with boundedly rational customers. They show that the revenue-optimal price is not generally socially efficient but depends on the retrial rate. Fiems (2023) investigates the M/D/1 retrial queueing system with constant retrial times. The author finds explicit expressions for various performance measures. Further discussion on retrial queues is available (see e.g., Avrachenkov and Yechiali 2010; Avrachenkov et al. 2014; Krishna Kumar et al. 2018; Do et al. 2020; Kumar et al. 2020; Gao et al. 2021; Lee et al. 2022; Nazarov et al. 2022; Templeton and Falin 2023; Zhang and Wang 2023; Melikov et al. 2023; Nithya et al. 2023). In all these studies, customers go to orbit *before entering the system*, while in our work customers go to orbit after their service has begun.

A special model where customers voluntarily go to orbit is the "ticket queue" process where impatient arriving customers who observe a long queue leave the service station for a random time and rejoin it after spending orbiting time outside the system (see e.g., Hanukov et al. (2020) and references there). Again, these studies, in contrast to ours, address the case where the customers go to orbit *before the beginning of their service*.

The model of customers orbiting while in service differs from the model of server vacations, which has been treated extensively in the literature (see e.g., Levy and Yechiali 1975, 1976; Kella and Yechiali 1988; Rosenberg and Yechiali 1993; Boxma et al. 2002; Tian and Zhang 2006; Liu and Wang 2017; Bouchentouf and Guendouzi 2019; Suranga Sampath and Liu 2020; Bouchentouf et al. 2020; Sakuma et al. 2021; Wang and Xu 2021; Jain et al. 2021; Kleiner et al. 2021; Xu et al. 2022; Economou et al. 2022; Afanasyev 2023; Sindhu et al. 2023). In the "orbit while in service" model, customers are "vacationing," while in the vacation models the server is the one who goes for (single or multiple) vacations.

In this paper, we thoroughly analyze the single-server "orbit while in service" queueing model with general service duration (which, to the best of our knowledge, as indicated above, has not been studied before) and investigate various variables characterizing the system. The main contributions of this paper are summarized below:

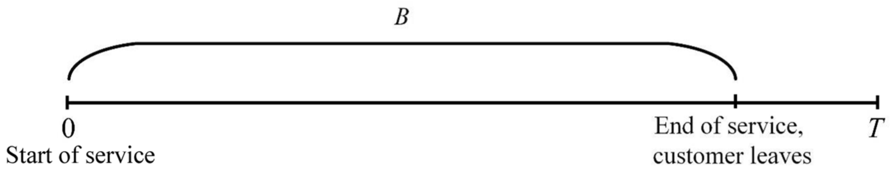
- General formulae are obtained for the distributions of (i) a customer's total residence time in the system; (ii) a customer's net actual residence time in the system during his/her service (not including orbit time); (iii) the time an orbiting customer is overdue, that is, remains in orbit after his/her service has been completed; and (iv) the number of customers in the system.
- Considering the family of Gamma-distributed service times (spanning the range of distributions between the Exponential and the Deterministic), as well as the Uniform distribution, we further derive explicit formulae for the distributions of the above variables.
- The optimal mean orbit time is calculated for each service-time distribution, considering (i) the utility customers gain from orbiting; (ii) the penalty incurred by a customer for every unit of time s/he is late returning from orbit; and (iii) the cost of queueing.
- Figures depicting the behavior of the measures as functions of the parameters are presented.

The remainder of this paper is organized as follows: Sect. 2 describes the model formulation. In Sect. 2.1, probability distribution functions for various sojourn times are obtained. Section 2.2 derives the probability distribution of the number of customers in the system, and Sect. 2.3 calculates closed-form expressions for various distributions. Section 3 provides a scheme to obtain the optimal mean orbit time. The findings are discussed in Sect. 4.

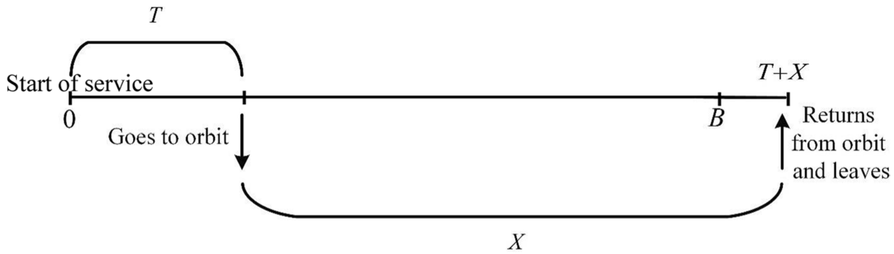
2 Model formulation

Customers arrive at a single-server queueing system according to a Poisson process with rate λ. Required net service times for individual customers, $B_1, B_2, B_3...$, are i.i.d, all distributed like B, with probability density function $f_B(t)$, cumulative distribution function $F_B(t)$, $\overline{F}_B(t) = 1 - F_B(t)$, and LST $\tilde{B}(s)$. After entering service, each customer is willing to wait an Exponentially distributed time T (with mean $1/\alpha$) for the service to be completed (T is called "patience time"). Several possible events (scenarios) can occur: (i) If the service time B is shorter than the customer's patience time T (i.e. $B < T$), the customer leaves the system upon service completion; (ii) If the service has not been completed by time T (i.e. $T < B$), the customer leaves for a random time X, called "orbit time", before returning to the system. The orbit time X is Exponentially distributed with parameter β. While the customer is in orbit, the server continues rendering the originally required service, B, for that customer (and possibly continues executing orders of other customers in case our customer's service has been completed). If the service has been completed before the customer's return ($B < T + X$), the customer picks his/her order and leaves; (iii) If a customer's service is not finished when s/he returns from orbit, the customer remains waiting in the service station until the service is completed, and only then leaves. The above possible scenarios are illustrated below:

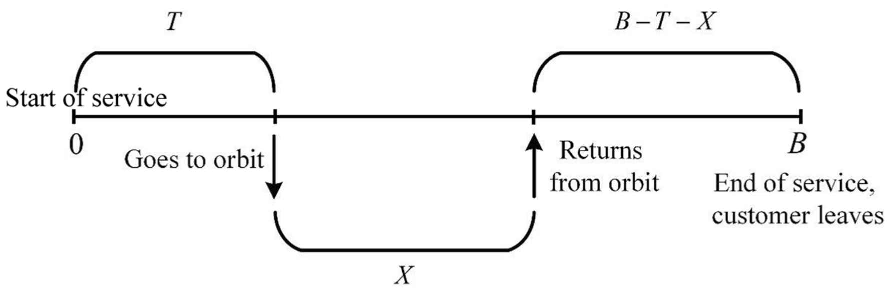
Scenario 1 The service time is shorter than the customer's patience time, that is, $B < T$. Consequently, the customer doesn't go to orbit, and leaves upon service completion.



Scenario 2 The customer's patience time is less than the service time, so the customer goes to orbit before service completion. Then, if the combined customer's patience time plus orbit time exceeds the service time, (that is, $T < B < T + X$), the customer leaves the station as soon as s/he is back from orbit.



Scenario 3 The customer's patience time plus orbit time is less than the service time, that is, $T + X < B$. After back from orbit, the customer waits until his/her service is completed.



Based on the above three scenarios, we investigate various key variables in the following section and derive their probability distribution functions.

It should be indicated that the ‘orbit while in service’ model differs from a regular M/G/1 queue in the following aspects: (i) service of orbiting customers continues even in their absence; (ii) sojourn time is not composed simply of queueing time and service time, but it is a combination of time before orbit; orbit time (if applicable); and waiting for service completion (if required) after returning from orbit. Thus, one cannot apply a standard argument to deduce the distribution of sojourn time from the corresponding probability generating function. Consequently, a different approach is required to derive the probability generating function of number of customers in the system (not including those in orbit).

Note: to make the following presentation and analysis easier to follow, we present a notation list in Appendix A.

2.1 Customers’ residence times

2.1.1 Total time in system

Let V denote a customer’s total residence time in the system measured from the instant s/he starts service until departure. Note that V does not include a customer’s waiting time for her/his service to start, but may include the customer’s time in orbit. We have (see the three scenarios above):

$$V=\begin{cases} B & \text{if } B<T \\ T+X & \text{if } T<B<T+X \\ B & \text{if } T+X<B \end{cases} \tag{1}$$

The Laplace-Stieltjes transform (LST) of V is derived in Theorem 1 below:

$$\tilde{V}(s)=\tilde{B}(s)+\frac{\alpha s}{(s+\beta)(\alpha-\beta)}\left(\tilde{B}(s+\alpha)-\tilde{B}(s+\beta)\right) \tag{2}$$

Theorem 1

Proof According to (1),

$$\tilde{V}(s)=E[e^{-sB}|B<T]P(B<T)+E[e^{-s(T+X)}|T<B<T+X]P(T<B<T+X)$$
$$+E[e^{-sB}|T+X<B]P(T+X<B).$$

Let $Y=T+X$ with probability distribution function $F_Y(y)$ and density $f_Y(y)$. Then, the above is translated to

$$\tilde{V}(s)=\int_{t=0}^{\infty}\int_{b=0}^{t}e^{-sb}f_B(b)f_T(t)dbdt+\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty}e^{-s(t+x)}f_B(b)f_T(t)f_X(x)dxdtdb$$
$$+\int_{y=0}^{\infty}\int_{b=y}^{\infty}e^{-sb}f_B(b)f_Y(y)dbdy. \tag{3}$$

Equation (3) comprises of three terms, calculated below.

Term 1

$\int_{t=0}^{\infty}\int_{b=0}^{t} e^{-sb}f_B(b)f_T(t)dbdt = \int_{b=0}^{\infty}\int_{t=b}^{\infty} e^{-sb}f_B(b)f_T(t)dtdb = \int_{0}^{\infty} e^{-sb}f_B(b)e^{-\alpha b}db = \tilde{B}(s+\alpha).$

Term 2

$$\begin{aligned}
&\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-s(t+x)}f_B(b)f_T(t)f_X(x)dxdtdb = \int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-st}f_B(b)f_T(t)e^{-sx}\beta e^{-\beta x}dxdtdb\\
&= \int_{b=0}^{\infty}\int_{t=0}^{b} e^{-st}f_B(b)f_T(t)\frac{\beta}{s+\beta}e^{-(s+\beta)(b-t)}dtdb\\
&= \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)e^{-st}\alpha e^{-\alpha t}\frac{\beta}{s+\beta}e^{-(s+\beta)(b-t)}dtdb\\
&= \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)\alpha e^{-(s+\alpha)t}\frac{\beta}{s+\beta}e^{-(s+\beta)(b-t)}dtdb = \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)e^{-(\alpha-\beta)t}\frac{\alpha\beta}{s+\beta}e^{-(s+\beta)b}dtdb\\
&= \int_{b=0}^{\infty} f_B(b)(1-e^{-(\alpha-\beta)b})\frac{\alpha\beta}{(s+\beta)(\alpha-\beta)}e^{-(s+\beta)b}db = \frac{\alpha\beta}{(s+\beta)(\alpha-\beta)}\left(\tilde{B}(s+\beta)-\tilde{B}(s+\alpha)\right).
\end{aligned}$$

Term 3

We first determine $F_Y(y)$ and $f_Y(y)$:

$F_Y(y) \equiv P(Y < y) = \int_{t=0}^{y} f_T(t)F_X(y-t)dt = \int_{t=0}^{y} \alpha e^{-\alpha t}(1-e^{-\beta(y-t)})dt = 1 + \frac{\beta e^{-\alpha y}-\alpha e^{-\beta y}}{\alpha-\beta}.$

Then,

$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{\alpha\beta(e^{-\beta y}-e^{-\alpha y})}{\alpha-\beta}.$

Consequently,

$$\begin{aligned}
&\int_{y=0}^{\infty}\int_{b=y}^{\infty} e^{-sb}f_B(b)f_Y(y)dbdy = \int_{b=0}^{\infty}\int_{y=0}^{b} e^{-sb}f_B(b)f_Y(y)dydb = \int_{0}^{\infty} e^{-sb}f_B(b)\left(1+\frac{\beta e^{-\alpha b}-\alpha e^{-\beta b}}{\alpha-\beta}\right)db\\
&= \tilde{B}(s) + \frac{\beta}{\alpha-\beta}\tilde{B}(s+\alpha) - \frac{\alpha}{\alpha-\beta}\tilde{B}(s+\beta).
\end{aligned}$$

Finally, the LST of V is given by

$$\begin{aligned}
\tilde{V}(s) =&\tilde{B}(s+\alpha) + \frac{\alpha\beta}{(s+\beta)(\alpha-\beta)}\left(\tilde{B}(s+\beta)-\tilde{B}(s+\alpha)\right)\\
&+\tilde{B}(s) + \frac{\beta}{\alpha-\beta}\tilde{B}(s+\alpha) - \frac{\alpha}{\alpha-\beta}\tilde{B}(s+\beta)\\
=&\tilde{B}(s) + \frac{\alpha s}{(s+\beta)(\alpha-\beta)}\left(\tilde{B}(s+\alpha)-\tilde{B}(s+\beta)\right).
\end{aligned}$$

Evidently, when the customer's patience time is unbounded ($\alpha \to 0$), $\tilde{V}(s) \to \tilde{B}(s)$. Moreover, when the orbit time tends to zero ($\beta \to \infty$), $\tilde{V}(s) \to \tilde{B}(s)$.

The customer's mean total residence time in the system is calculated by using $E[V] = -\frac{d\tilde{V}(s)}{ds}\Big|_{s=0}$. We obtain.

$E[V] = E[B] + \frac{\alpha(\tilde{B}(\beta)-\tilde{B}(\alpha))}{\beta(\alpha-\beta)}$.

It follows that for any $\alpha, \beta > 0$, $E[V] > E[B]$. Moreover, for small α and β, $E[V] \approx \left(1+\frac{\alpha}{\beta}\right)E[B]$.

2.1.2 Net actual residence time

An important variable is D, a customer's net actual residence time in the system during his/her service time (not including orbit time).

We have (see the corresponding illustrations for scenarios 1, 2, and 3):

$$D = \begin{cases} B & \textit{if} \;\; B < T \\ T & \textit{if} \;\; T < B < T + X \\ B - X & \textit{if} \;\; T + X < B \end{cases} \tag{4}$$

The LST of D is derived in Theorem 2:

$$\tilde{D}(s) = \frac{\beta}{\beta - s}\tilde{B}(s) + \frac{s}{s+\alpha-\beta}\tilde{B}(s+\alpha) - \frac{\alpha s}{(\beta-s)(s+\alpha-\beta)}\tilde{B}(\beta) \tag{5}$$

Theorem 2

Proof According to (4),

$$\begin{aligned}\tilde{D}(s) = {} & E[e^{-sB}|B < T]P(B < T) + E[e^{-sT}|T < B < T+X]P(T < B < T+X) \\ & + E[e^{-s(B-X)}|T+X < B]P(T+X < B).\end{aligned} \tag{6}$$

Equation (6) comprises three terms, calculated below.

Term 1

$$\begin{aligned}E[e^{-sB}|B < T]P(B < T) &= \int_{t=0}^{\infty}\int_{b=0}^{t} e^{-sb}f_B(b)f_T(t)dbdt = \int_{b=0}^{\infty}\int_{t=b}^{\infty} e^{-sb}f_B(b)f_T(t)dtdb \\ &= \int_{0}^{\infty} e^{-sb}f_B(b)e^{-\alpha b}db = \tilde{B}(s+\alpha).\end{aligned}$$

Term 2

$$E[e^{-sT}|T < B < T+X]P(T < B < T+X) = \int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-st}f_B(b)f_T(t)f_X(x)dxdtdb$$
$$= \int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-st}f_B(b)f_T(t)\beta e^{-\beta x}dxdtdb = \int_{b=0}^{\infty}\int_{t=0}^{b} e^{-st}f_B(b)f_T(t)e^{-\beta(b-t)}dtdb$$
$$= \int_{b=0}^{\infty}\int_{t=0}^{b} e^{-st}f_B(b)\alpha e^{-\alpha t}e^{-\beta(b-t)}dtdb = \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)\alpha e^{-(s+\alpha)t}e^{-\beta(b-t)}dtdb$$
$$= \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)e^{-\beta b}\alpha e^{-(s+\alpha-\beta)t}dtdb = \int_{b=0}^{\infty} f_B(b)e^{-\beta b}(1-e^{-(s+\alpha-\beta)b})\frac{\alpha}{(s+\alpha-\beta)}db$$
$$= \frac{\alpha}{(s+\alpha-\beta)}\left(\tilde{B}(\beta)-\tilde{B}(s+\alpha)\right).$$

Term 3

$$E[e^{-s(B-X)}|T+X < B]P(T+X < B) =$$
$$= \frac{\alpha\beta}{\beta-s}\left(\frac{1}{\alpha}\tilde{B}(s) - \frac{1}{(\alpha-\beta+s)}\tilde{B}(\beta) + \left(\frac{1}{(\alpha-\beta+s)} - \frac{1}{\alpha}\right)\tilde{B}(s+\alpha)\right).$$

The details of the derivation of Term 3 are given in Appendix B.

Finally, the LST of D is given by

$$\begin{aligned}\tilde{D}(s) =& \tilde{B}(s+\alpha) + \frac{\alpha}{(s+\alpha-\beta)}\left(\tilde{B}(\beta)-\tilde{B}(s+\alpha)\right)\\ &+ \frac{\alpha\beta}{\beta-s}\left(\frac{1}{\alpha}\tilde{B}(s) - \frac{1}{(\alpha-\beta+s)}\tilde{B}(\beta) + \left(\frac{1}{(\alpha-\beta+s)} - \frac{1}{\alpha}\right)\tilde{B}(s+\alpha)\right)\\ =& \frac{\beta}{\beta-s}\tilde{B}(s) + \frac{s}{s+\alpha-\beta}\tilde{B}(s+\alpha) - \frac{\alpha s}{(\beta-s)(s+\alpha-\beta)}\tilde{B}(\beta).\end{aligned}$$

Again, when the customer's patience time is unbounded ($\alpha \to 0$), $\tilde{D}(s) \to \tilde{B}(s)$. Moreover, when the orbit time tends to zero ($\beta \to \infty$), $\tilde{D}(s) \to \tilde{B}(s)$.

The mean of D, $E[D] = -\frac{d\tilde{D}(s)}{ds}\Big|_{s=0}$ is given by.

$$E[D] = E[B] - \frac{1}{\beta}\left(1 - \frac{\alpha\tilde{B}(\beta) - \beta\tilde{B}(\alpha)}{\alpha-\beta}\right) \tag{7}$$

A customer's mean queueing time (not including service or orbit) is the same as that of the classical M/G/1 queue, that is, $E[W_q] = \frac{\lambda E[B^2]}{2(1-\lambda E[B])}$. This follows since an arriving customer waits in line before entering service for the same length of time as s/he would have waited in a regular M/G/1 queue with no orbiting, since the orbiting of customers whose service has begun does not affect their service, which continues while they are orbiting. Thus, a customer's mean sojourn time is given by $E[W] = E[W_q] + E[D]$.

The mean total time of a customer in the system, from instant of arrival to that of departure, is $E[W_q] + E[V]$.

2.1.3 Time in orbit after service completion

Let H denote the time a customer remains in orbit after his/her service has been completed. We have.

$$H = \begin{cases} T + X - B \ \ if \ \ T < B < T + X \\ 0 \qquad\qquad Otherwise \end{cases} \tag{8}$$

The LST of H is derived in Theorem 3:

$$\tilde{H}(s) = \frac{\beta}{s+\beta}\frac{\alpha}{\alpha-\beta}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right) + 1 \cdot \left(1 - \frac{\alpha}{\alpha-\beta}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right)\right) \tag{9}$$

Theorem 3

Proof According to (8),

$$\tilde{H}(s) = E[e^{-s(T+X-B)}|T < B < T + X]P(T < B < T + X) \ + 1 \cdot (1 - P(T < B < T + X)), \tag{10}$$

the first term of which is calculated as follows:

$$\begin{aligned}
&\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-s(t+x-b)} f_B(b) f_T(t) f_X(x) dx dt db = \int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} e^{-s(t-b)} f_B(b) f_T(t) e^{-sx} \beta e^{-\beta x} dx dt db \\
&= \int_{b=0}^{\infty}\int_{t=0}^{b} e^{-s(t-b)} f_B(b) f_T(t) \frac{\beta}{s+\beta} e^{-(s+\beta)(b-t)} dt db = \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b) e^{-s(t-b)} \alpha e^{-\alpha t} \frac{\beta}{s+\beta} e^{-(s+\beta)(b-t)} dt db \\
&= \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b) e^{-(\alpha-\beta)t} \frac{\alpha\beta}{s+\beta} e^{-\beta b} dt db = \int_{b=0}^{\infty} f_B(b)(1 - e^{-(\alpha-\beta)b}) \frac{\alpha\beta}{(s+\beta)(\alpha-\beta)} e^{-\beta b} db \\
&= \frac{\alpha\beta}{(s+\beta)(\alpha-\beta)}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right).
\end{aligned}$$

The probability that a customer leaves to orbit and then is late to return is $P(T < B < T + X) = \frac{\alpha}{\alpha-\beta}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right)$ (see Appendix C). Thus, the second term in Eq. (10) is given as $1 \cdot (1 - P(T < B < T + X)) = 1 - \frac{\alpha}{\alpha-\beta}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right)$.

Note that the first term in (9) is equal to $\tilde{X}(s)P(T < B < T + X)$ which is explained by the memoryless property of the orbit time.

Using $E[H] = -\frac{d\tilde{H}(s)}{ds}\Big|_{s=0}$, we obtain.

$$E[H] = \frac{\alpha}{\beta(\alpha-\beta)}\left(\tilde{B}(\beta) - \tilde{B}(\alpha)\right) \tag{11}$$

2.2 Number of customers in the system

Let L denote the number of customers in the system (not including those in orbit), where $L \in \{0, 1, 2, ...\}$. A customer whose service has begun but has not been completed yet can be in one of three stages: stage 1—before entering orbit; stage 2—in orbit; or stage 3—after returning from orbit. Denote this set of stages by S, $S \in \{1, 2, 3\}$. Let U denote the elapsed time since the start of service of a tagged customer (including orbit if applicable). The system's state space is defined as $\{L, S, U\}$. For $L = n$, $S = m$ and $U = u$, let $p(n, m, u)$ denote the density of that state in steady state, $n = 0, 1, 2, ...$; $m = 1, 2, 3$; $u \geq 0$. Let $\mu(u)$ be the hazard rate function of the service time. We now apply a supplementary variable technique to obtain the probability generating function of the number of customers in the system. The analysis proceeds by deriving separately the PGF of the number of customers in the system in each of the three stages defined above. The probabilities $p(n, m, u)$ obey the following:

$p(n, 1, u + h) = \lambda hp(n - 1, 1, u) + [1 - (\lambda + \mu(u) + \alpha)h]p(n, 1, u) + o(h)$, $n = 1, 2, 3, ...$; $u > 0$, where $p(0, 1, u) = 0$.

$p(n, 2, u + h) = \lambda hp(n - 1, 2, u) + \alpha hp(n + 1, 1, u) + [1 - (\lambda + \mu(u) + \beta)h]p(n, 2, u) + o(h)$, $n = 0, 1, 2, ...$; $u > 0$, where $p(-1, 2, u) = 0$.

$p(n, 3, u + h) = \lambda hp(n - 1, 3, u) + \beta hp(n - 1, 2, u) + [1 - (\lambda + \mu(u))h]p(n, 3, u) + o(h)$, $n = 1, 2, 3, ...$; $u > 0$, where $p(0, 3, u) = 0$.

Dividing each equation by h and letting $h \to 0$, we obtain for $m = 1$, $\frac{d}{du}p(n, 1, u) = \lambda p(n - 1, 1, u) - (\lambda + \mu(u) + \alpha)p(n, 1, u)$, $n = 1, 2, 3, ...$, $u > 0$; for $m = 2$,

$\frac{d}{du}p(n, 2, u) = \lambda p(n - 1, 2, u) + \alpha p(n + 1, 1, u) - (\lambda + \mu(u) + \beta)p(n, 2, u)$, $n = 0, 1, 2, ...$; $u > 0$; for $m = 3$, $\frac{d}{du}p(n, 3, u) = \lambda p(n - 1, 3, u) + \beta p(n - 1, 2, u) - (\lambda + \mu(u))p(n, 3, u)$, $n = 1, 2, 3, ...$; $u > 0$.

Define the following three probability generating functions (PGFs) corresponding to the three possible values of $m = 1, 2, 3$ as follows:

$G_1(z, u) = \sum_{n=1}^{\infty} p(n, 1, u)z^n$, $G_2(z, u) = \sum_{n=0}^{\infty} p(n, 2, u)z^n$, $G_3(z, u) = \sum_{n=1}^{\infty} p(n, 3, u)z^n$

By multiplying each of the above differential equations (for $m = 1, 2, 3$) by the corresponding z^n and summing over all n, the following set of differential equations for the three PGFs, $G_m(z, u)$, $m = 1, 2, 3$, is derived:

$$\frac{d}{du}G_1(z, u) = [\lambda(z - 1) - \mu(u) - \alpha]G_1(z, u), \tag{12}$$

$$\frac{d}{du}G_2(z, u) = [\lambda(z - 1) - \mu(u) - \beta]G_2(z, u) + \alpha z^{-1}G_1(z, u), \tag{13}$$

$$\frac{d}{du}G_3(z, u) = [\lambda(z - 1) - \mu(u)]G_3(z, u) + \beta zG_2(z, u). \tag{14}$$

Equation (12) is solved using the identity (e.g., Ross 1996) relating the hazard function $\mu(u)$ to its corresponding probability distribution function, namely, $\overline{F}_B(u) = e^{-\int_0^u \mu(t)dt}$. We then obtain.

$G_1(z,u) = C_1 e^{(\lambda(z-1)-\alpha)u}\overline{F}_B(u)$

C_1 is calculated by using $\overline{F}_B(0) = 1$ and by utilizing the initial condition: $G_1(z,0) = C_1 e^{(\lambda(z-1)-\alpha)0}\overline{F}_B(0)$. Thus, $C_1 = G_1(z,0)$, leading to

$$G_1(z,u) = G_1(z,0)e^{(\lambda(z-1)-\alpha)u}\overline{F}_B(u) \tag{15}$$

We now solve Eqs. (13) and (14):

$$G_2(z,u) = G_1(z,0)(\beta-\alpha)^{-1}\alpha z^{-1}e^{(\lambda(z-1)-\alpha)u}\overline{F}_B(u) + C_2\overline{F}_B(u)e^{(\lambda(z-1)-\beta)u},$$

$$G_3(z,u) = G_1(z,0)(\beta-\alpha)^{-1}(-\beta e^{((\lambda(z-1)-\alpha)u} + \alpha e^{((\lambda(z-1)-\beta)u})\overline{F}_B(u) + C_3\overline{F}_B(u)e^{(\lambda(z-1)u}.$$

By using $G_2(z,0) = 0$, C_2 is calculated: $C_2 = -G_1(z,0)(\beta-\alpha)^{-1}\alpha z^{-1}$, so that.

$$\begin{aligned} G_2(z,u) = {} & G_1(z,0)(\beta-\alpha)^{-1}\alpha z^{-1}e^{(\lambda(z-1)-\alpha)u}\overline{F}_B(u) \\ & -G_1(z,0)(\beta-\alpha)^{-1}\alpha z^{-1}\overline{F}_B(u)e^{(\lambda(z-1)-\beta)u} \end{aligned} \tag{16}$$

Finally, using $G_3(z,0) = 0$, C_3 is calculated as $C_3 = G_1(z,0)$, leading to.

$$G_3(z,u) = G_1(z,0)\overline{F}_B(u)((\beta-\alpha)^{-1}(-\beta e^{(\lambda(z-1)-\alpha)u} + \alpha e^{(\lambda(z-1)-\beta)u}) + e^{(\lambda(z-1)u})) \tag{17}$$

Each of the above three PGFs, $G_m(z,u)$, is expressed as a function of $G_1(z,0)$. To obtain the latter, we proceed as follows:

$p(1,1,0) = \int_0^\infty [p(2,1,u) + p(1,2,u) + p(2,3,u)]\mu(x)du + \lambda p(0)$,

$p(n,1,0) = \int_0^\infty [p(n+1,1,u) + p(n,2,u) + p(n+1,3,u)]\mu(u)du$, $n = 2,3,4,....$

Here, $p(0)$ is the probability that the system is empty (and no customer is getting service at stage $S = 2$). By multiplying each equation by z^n and summing over all n, we get

$$\begin{aligned} & zG_1(z,0) = z^2\lambda p(0) \\ & + \int_0^\infty (G_1(z,u) - zp(1,1,u) + zG_2(z,u) - zp(0,2,u) + G_3(z,u) - zp(1,3,u))\mu(u)du. \end{aligned} \tag{18}$$

To evaluate the right-hand side of (18), we derive an equation for $p(0)$ as follows:

$$\lambda p(0) = \int_0^\infty [p(1,1,u) + p(0,2,u) + p(1,3,u)]\mu(u))dx \tag{19}$$

By substituting (19) in (18), we obtain.

$$zG_1(z,0) = \int_0^\infty \left(G_1(z,u) + zG_2(z,u) + G_3(z,u)\right)\mu(u)du - z\lambda p(0) + z^2\lambda p(0) \tag{20}$$

By substituting (15)–(17) in (20), we have.

$$G_1(z,0) = \frac{z\lambda(1-z)p(0)}{\widetilde{B}(\lambda(1-z)) - z} \tag{21}$$

Let $p(n,m) = P(L=n, S=m)$, and let.

$$G_1(z) = \sum_{n=1}^\infty p(n,1)z^n,\ G_2(z) = \sum_{n=0}^\infty p(n,2)z^n,\ G_3(z) = \sum_{n=1}^\infty p(n,3)z^n$$

Then, by substituting (21) in (15), (16), and (17), respectively, while using $G_m(z) = \int_0^\infty G_m(z,u)du,\ m = 1,2,3$, we obtain the three PGFs, each as a function of $p(0)$:

$$G_1(z) = \frac{z\lambda(1-z)\left(1 - \widetilde{B}(\lambda(1-z)+\alpha)\right)}{(\lambda(1-z)+\alpha)\left(\widetilde{B}(\lambda(1-z)) - z\right)}p(0), \tag{22}$$

$$\begin{aligned} G_2(z) = \Bigg[& \frac{\lambda(1-z)(\beta-\alpha)^{-1}\alpha\left(1 - \widetilde{B}(\lambda(1-z)+\alpha)\right)}{(\lambda(1-z)+\alpha)\left(\widetilde{B}(\lambda(1-z)) - z\right)} \\ & - \frac{\lambda(1-z)(\beta-\alpha)^{-1}\alpha\left(1 - \widetilde{B}(\lambda(1-z)+\beta)\right)}{(\lambda(1-z)+\beta)\left(\widetilde{B}(\lambda(1-z)) - z\right)}\Bigg]p(0), \end{aligned} \tag{23}$$

$$\begin{aligned} G_3(z) = \Bigg[& -\frac{z\lambda(1-z)\beta(\beta-\alpha)^{-1}\left(1 - \widetilde{B}(\lambda(1-z)+\alpha)\right)}{(\lambda(1-z)+\alpha)\left(\widetilde{B}(\lambda(1-z)) - z\right)} \\ & + \frac{z\lambda(1-z)\alpha(\beta-\alpha)^{-1}\left(1 - \widetilde{B}(\lambda(1-z)+\beta)\right)}{(\lambda(1-z)+\beta)\left(\widetilde{B}(\lambda(1-z)) - z\right)} \\ & + \frac{z\left(1 - \widetilde{B}(\lambda(1-z))\right)}{\widetilde{B}(\lambda(1-z)) - z}\Bigg]p(0). \end{aligned} \tag{24}$$

To calculate $p(0)$, we first calculate

$$
\begin{aligned}
&G_1(z)+G_2(z)+G_3(z)\\
&=\left[\frac{\lambda(1-z)\left(z+(\beta-\alpha)^{-1}(\alpha-z\beta)\right)\left(1-\widetilde{B}(\lambda(1-z)+\alpha)\right)}{(\lambda(1-z)+\alpha)\left(\widetilde{B}(\lambda(1-z))-z\right)}\right.\\
&\quad-\frac{\lambda(1-z)^2(\beta-\alpha)^{-1}\alpha\left(1-\widetilde{B}(\lambda(1-z)+\beta)\right)}{(\lambda(1-z)+\beta)\left(\widetilde{B}(\lambda(1-z))-z\right)}\\
&\quad\left.+\frac{z\left(1-\widetilde{B}(\lambda(1-z))\right)}{\widetilde{B}(\lambda(1-z))-z}\right]p(0).
\end{aligned}
$$

Since $1=p(0)+G_1(z=1)+G_2(z=1)+G_3(z=1)$, by using L'Hospital's rule, we get.

$p(0)=1-\lambda E[B]$.

By substituting $p(0)=1-\lambda E[B]$ in (22)–(24), we obtain complete expressions of the PGFs.

When the customer's patience time is unbounded ($\alpha\to 0$), then $p(0)+G_1(z)+G_2(z)+G_3(z)\to\frac{(1-z)\widetilde{B}(\lambda(1-z))}{\widetilde{B}(\lambda(1-z))-z}(1-\lambda E[B])$, which is the generating function of the number of customers in the classical M/G/1 system (e.g., Harchol-Balter 2013). The same occurs when the orbit time tends to zero ($\beta\to\infty$).

As $E[L]=\left.\frac{dG_1(z)}{dz}\right|_{z=1}+\left.\frac{dG_2(z)}{dz}\right|_{z=1}+\left.\frac{dG_3(z)}{dz}\right|_{z=1}$, using L'Hospital's rule, we obtain

$$
\begin{aligned}
E[L]&=\lambda E[B]+\frac{\lambda^2E[B^2]}{2(1-\lambda E[B])}-\frac{\lambda}{\beta}\left(1-\frac{\alpha\widetilde{B}(\beta)-\beta\widetilde{B}(\alpha)}{\beta(\alpha-\beta)}\right)\\
&=\lambda E[W]=E[L_{M/G/1}]-\frac{\lambda}{\beta}\left(1-\frac{\alpha\widetilde{B}(\beta)-\beta\widetilde{B}(\alpha)}{\beta(\alpha-\beta)}\right).
\end{aligned}
\tag{25}
$$

2.3 Examples

This section provides explicit formulae for the wide-ranging family of Gamma probability distribution functions (spanning the range between the Exponential and the Deterministic distributions), as well as for the Uniform distribution, each representing the distribution of service times.

2.3.1 Gamma distribution

Suppose the service time B is distributed according to the Gamma probability distribution, namely, $B \sim \Gamma(\gamma, \gamma\mu)$, with mean $E[B] = \frac{\gamma}{\gamma\mu} = \frac{1}{\mu}$, second moment $E[B^2] = \frac{1+\gamma}{\gamma\mu^2}$, variance $V[B] = \frac{1}{\gamma\mu^2}$, and LST $\tilde{B}(s) = \left(\frac{\gamma\mu}{\gamma\mu+s}\right)^{\gamma}$.

Case (i) $\gamma = 1 \quad (B \sim \mathit{Exponential}(\mu))$.

In this case,

$\tilde{B}(s) = \frac{\mu}{\mu+s}$,

$\tilde{V}(s) = \frac{\mu}{\mu+s} - \frac{\alpha\mu s}{(\beta+s)(\alpha+\mu+s)(\beta+\mu+s)}$,

$E[V] = \frac{1}{\mu} + \frac{\alpha\mu}{\beta(\alpha+\mu)(\beta+\mu)}$,

$\tilde{D}(s) = \frac{\mu(s(\alpha+\beta+\mu)+(\alpha+\mu)(\beta+\mu))}{(\mu+s)(\alpha+\mu+s)(\beta+\mu)}$,

$E[D] = \frac{1}{\mu} - \frac{\alpha}{(\alpha+\mu)(\beta+\mu)}$,

$\tilde{H}(s) = \frac{\alpha\mu}{(\beta+s)(\alpha+\mu)(\beta+\mu)}$,

$E[H] = \frac{\alpha\mu}{\beta(\alpha+\mu)(\beta+\mu)}$,

$G_1(z) = \frac{z\lambda(\lambda(1-z)+\mu)(\mu-\lambda)}{\mu(\mu-\lambda z)(\lambda(1-z)+\mu+\alpha)}$,

$G_2(z) = \frac{\lambda\alpha(\lambda(1-z)+\mu)(\mu-\lambda)}{\mu(\mu-\lambda z)(\lambda(1-z)+\mu+\alpha)(\lambda(1-z)+\mu+\beta)}$,

$G_3(z) = \frac{\lambda\alpha\beta z(\mu-\lambda)}{\mu(\mu-\lambda z)(\lambda(1-z)+\mu+\alpha)(\lambda(1-z)+\mu+\beta)}$,

$$E[L] = \frac{\lambda}{\mu-\lambda} - \frac{\lambda\alpha}{(\alpha+\mu)(\beta+\mu)} = E[L_{M/M/1}] - \frac{\lambda\alpha}{(\alpha+\mu)(\beta+\mu)} \tag{26}$$

Clearly, $E[L] < E[L_{M/M/1}]$ since during part of their service time, customers remain in orbit out of the main queue in the service station.

Case (ii) $1 < \gamma < \infty$.

$\tilde{V}(s) = \left(\frac{\gamma\mu}{\gamma\mu+s}\right)^{\gamma} + \frac{\alpha s}{(s+\beta)(\alpha-\beta)}\left(\left(\frac{\gamma\mu}{\gamma\mu+s+\alpha}\right)^{\gamma} - \left(\frac{\gamma\mu}{\gamma\mu+s+\beta}\right)^{\gamma}\right)$,

$E[V] = \frac{1}{\mu} + \frac{\alpha\left(\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma} - \left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}\right)}{\beta(\alpha-\beta)}$,

$\tilde{D}(s) = \frac{\beta}{\beta-s}\left(\frac{\gamma\mu}{\gamma\mu+s}\right)^{\gamma} + \frac{s}{s+\alpha-\beta}\left(\frac{\gamma\mu}{\gamma\mu+s+\alpha}\right)^{\gamma} - \frac{\alpha s}{(\beta-s)(s+\alpha-\beta)}\left(\frac{\gamma\mu}{\gamma\mu+s+\beta}\right)^{\gamma}$,

$E[D] = \frac{1}{\mu} - \frac{1}{\beta}\left(1 - \frac{\alpha\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma} - \beta\left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}}{\alpha-\beta}\right)$,

$\tilde{H}(s) = \frac{\alpha\beta}{(s+\beta)(\alpha-\beta)}\left(\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma} - \left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}\right)$,

$E[H] = \frac{\alpha}{\beta(\alpha-\beta)}\left(\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma} - \left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}\right)$,

$$E[L]=\lambda\frac{1}{\mu}+\frac{\lambda^2\frac{1+\gamma}{\gamma\mu^2}}{2\left(1-\lambda\frac{1}{\mu}\right)}-\frac{\lambda}{\beta}\left(1-\frac{\alpha\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma}-\beta\left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}}{\beta(\alpha-\beta)}\right)$$
$$=E[L_{M/\Gamma(\gamma,\gamma\mu)/1}]-\frac{\lambda}{\beta}\left(1-\frac{\alpha\left(\frac{\gamma\mu}{\gamma\mu+\beta}\right)^{\gamma}-\beta\left(\frac{\gamma\mu}{\gamma\mu+\alpha}\right)^{\gamma}}{\beta(\alpha-\beta)}\right).$$

Case (iii) $\gamma\to\infty$ ($B\sim$ Deterministic).

$\tilde{B}(s)=e^{-s/\mu}$,

$\tilde{V}(s)=e^{-\frac{s}{\mu}}-\frac{\alpha s\left(e^{-\frac{s+\beta}{\mu}}-e^{-\frac{s+\alpha}{\mu}}\right)}{(\beta+s)(\alpha-\beta)}$,

$E[V]=\frac{1}{\mu}+\frac{\alpha\left(e^{-\frac{\beta}{\mu}}-e^{-\frac{\alpha}{\mu}}\right)}{\beta(\alpha-\beta)}$,

$\tilde{D}(s)=\frac{s(s-\beta)e^{-\frac{s+\alpha}{\mu}}-\beta(s+\alpha-\beta)e^{-\frac{s}{\mu}}-\alpha se^{-\frac{\beta}{\mu}}}{(s-\beta)(s+\alpha-\beta)}$,

$E[D]=\frac{1}{\mu}-\frac{\beta e^{-\frac{\alpha}{\mu}}-\alpha e^{-\frac{\beta}{\mu}}+\alpha-\beta}{\beta(\alpha-\beta)}$,

$\tilde{H}(s)=\frac{\alpha\left(e^{-\frac{\beta}{\mu}}-e^{-\frac{\alpha}{\mu}}\right)}{(s+\beta)(\alpha-\beta)}$,

$E[H]=\frac{\alpha\left(e^{-\frac{\beta}{\mu}}-e^{-\frac{\alpha}{\mu}}\right)}{\beta(\alpha-\beta)}$,

$G_1(z)=\frac{z\lambda(1-z)(\mu-\lambda)\left(1-e^{-\frac{\lambda(1-z)+\alpha}{\mu}}\right)}{\mu(\lambda(1-z)+\alpha)\left(z-e^{-\frac{\lambda(1-z)}{\mu}}\right)}$,

$G_2(z)=\frac{\lambda\alpha(1-z)(\mu-\lambda)\left((\lambda(1-z)+\beta)e^{-\frac{\lambda(1-z)+\alpha}{\mu}}-(\lambda(1-z)+\alpha)e^{-\frac{\lambda(1-z)+\beta}{\mu}}+\alpha-\beta\right)}{\mu(\alpha-\beta)(\lambda(1-z)+\alpha)(\lambda(1-z)+\beta)\left(z-e^{-\frac{\lambda(1-z)}{\mu}}\right)}$,

$$G_3(z)=\frac{z(\mu-\lambda)\begin{pmatrix}\lambda\alpha(1-z)(\lambda(1-z)+\alpha)e^{-\frac{\lambda(1-z)+\beta}{\mu}}-\lambda\beta(1-z)(\lambda(1-z)+\beta)e^{-\frac{\lambda(1-z)+\alpha}{\mu}}\\-(\alpha-\beta)\left((\lambda(1-z)+\alpha)(\lambda(1-z)+\beta)e^{-\frac{\lambda(1-z)}{\mu}}-\alpha\beta\right)\end{pmatrix}}{(\alpha-\beta)(\lambda(1-z)+\alpha)(\lambda(1-z)+\beta)\left(z-e^{-\frac{\lambda(1-z)}{\mu}}\right)},$$

$E[L]=\frac{\lambda(2\mu-\lambda)}{2\mu(\mu-\lambda)}-\frac{\lambda\left(\beta e^{-\frac{\alpha}{\mu}}-\alpha e^{-\frac{\beta}{\mu}}+\alpha-\beta\right)}{\beta(\alpha-\beta)}=E[L_{M/D/1}]-\frac{\lambda\left(\beta e^{-\frac{\alpha}{\mu}}-\alpha e^{-\frac{\beta}{\mu}}+\alpha-\beta\right)}{\beta(\alpha-\beta)}.$

The following graphs exhibit the change in $E[D]$ (Fig. 1), $E[H]$ (Fig. 2), and $E[L]$ (Fig. 3), each as a function of γ, where $\lambda=8$, $\mu=10$, $\alpha=18$, and $\beta=20$. When γ increases, the variance of B decreases, so $E[D]$ and $E[L]$ monotonically decrease, while $E[H]$ monotonically increases.

An interesting phenomenon occurs regarding $E[V]$. When the service-time variance decreases (γ increases), $E[V]$ increases in some cases but decreases in others (see Figs. 4, 5 and 6). This result can be explained as follows: the decrease in the

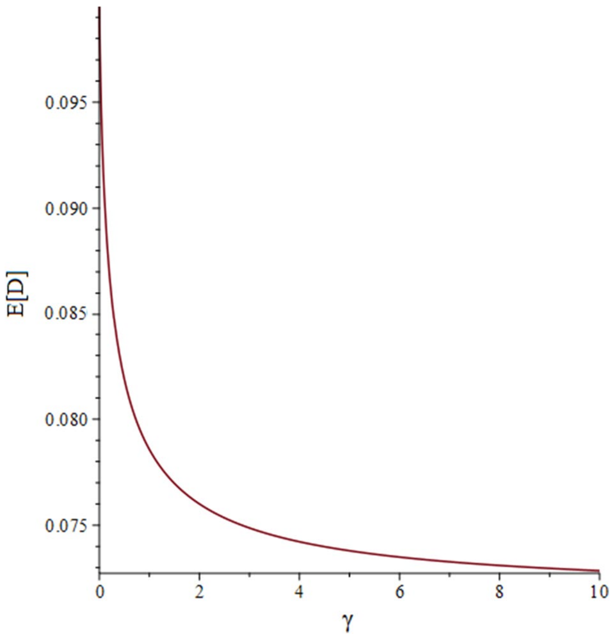


Fig. 1 $E[D]$ as a function of γ when $\lambda = 8$, $\mu = 10$, $\alpha = 18$, and $\beta = 20$

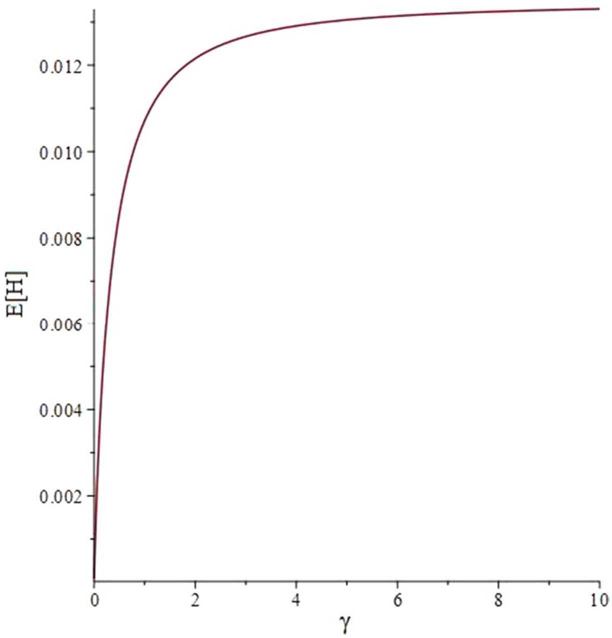


Fig. 2 $E[H]$ as a function of γ when $\lambda = 8$, $\mu = 10$, $\alpha = 18$, and $\beta = 20$

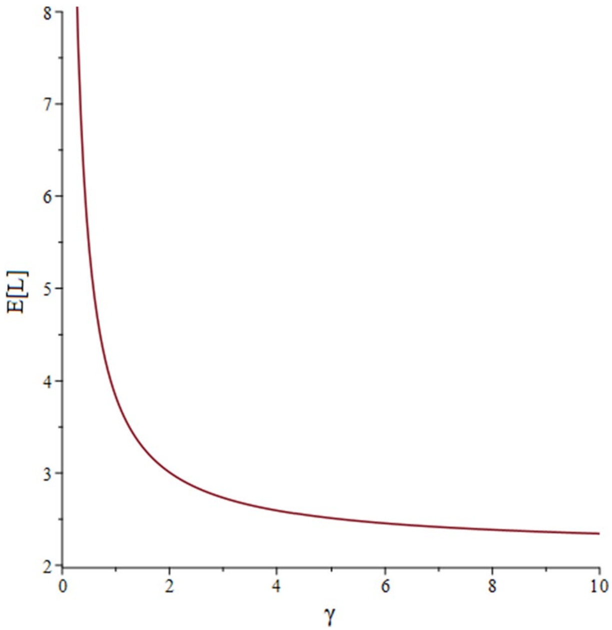


Fig. 3 $E[L]$ as a function γ when $\lambda = 8$, $\mu = 10$, $\alpha = 18$, and $\beta = 20$

service-time variance (when γ increases) has two contradicting effects on $E[V]$: (i) the probability of entering orbit decreases, causing $E[V]$ to decrease; however, (ii) the probability of returning from orbit after the service has been completed increases, causing $E[V]$ to increase. Consequently, $E[V]$ decreases when the first effect overtakes the second and increases vice versa. Moreover, Figs. 4, 5 and 6 show that the higher the value of α, the more significant is the first effect (i.e., the decrease in $E[V]$ starts sooner). Furthermore, for lower values of α, $E[V]$ monotonically increases (Fig. 6).

2.3.2 Uniform distribution

Consider now the case where B is Uniformly distributed, namely

$$B \sim U\left(0, \frac{2}{\mu}\right),\ E[B] = \frac{1}{\mu},\ E[B^2] = \frac{4}{3\mu^2},\ V[B] = \frac{1}{3\mu^2},\ \tilde{B}(s) = \frac{\mu}{2s}\left(1 - e^{-\frac{2s}{\mu}}\right)$$

Substituting those expressions in the general formulae derived above, we have.

$$\tilde{V}(s) = \frac{\mu\left(1-e^{-\frac{2s}{\mu}}\right)}{2s} - \frac{\alpha\mu s\left((s+\alpha)e^{-\frac{2(s+\beta)}{\mu}} - (s+\beta)e^{-\frac{2(s+\alpha)}{\mu}} - \alpha+\beta\right)}{(\beta+s)(\alpha-\beta)},$$

$$E[V] = \frac{1}{\mu} + \frac{\mu\left(\beta e^{-\frac{2\alpha}{\mu}} - \alpha e^{-\frac{2\beta}{\mu}} + \alpha-\beta\right)}{2\beta^2(\alpha-\beta)},$$

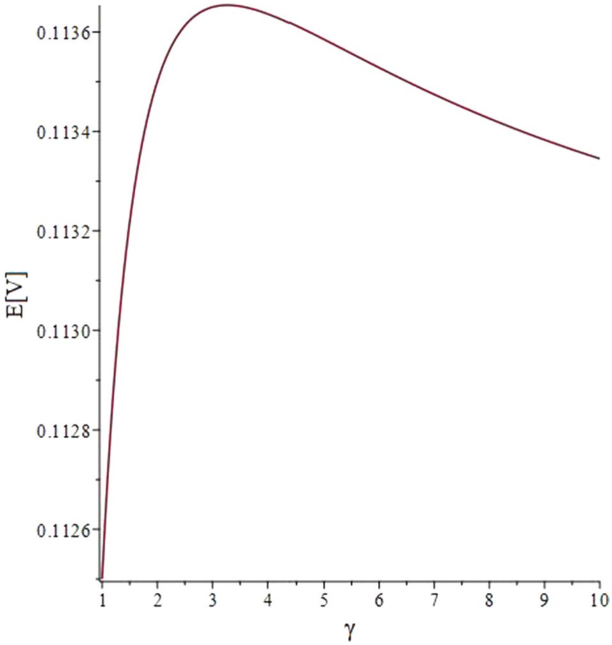


Fig. 4 $E[V]$ as a function of γ when $\lambda = 8$, $\mu = 10$, $\alpha = 30$, and $\beta = 20$

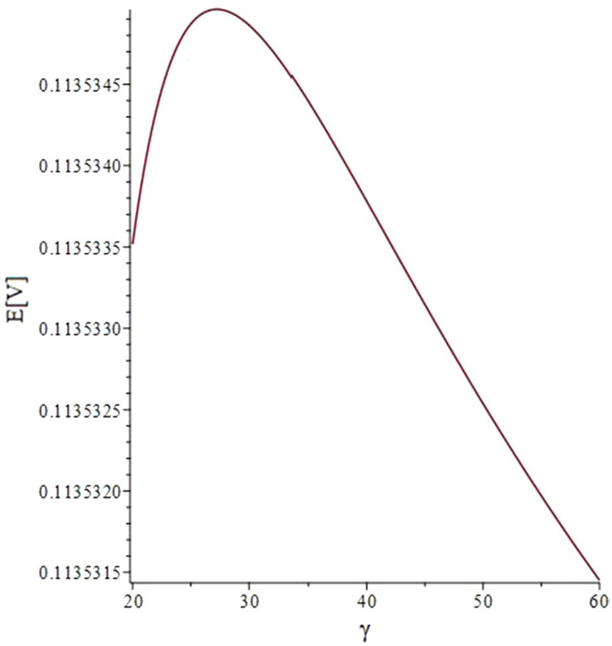


Fig. 5 $E[V]$ as a function of γ when $\lambda = 8$, $\mu = 10$, $\alpha = 21$, and $\beta = 20$

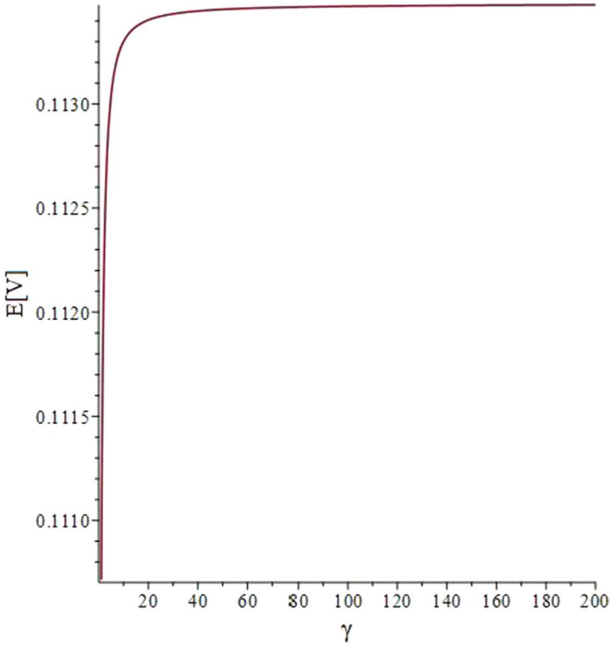


Fig. 6 $E[V]$ as a function of γ when $\lambda = 8$, $\mu = 10$, $\alpha = 18$, and $\beta = 20$

$$\tilde{D}(s) = \frac{s\mu\left(1-e^{-\frac{2(s+\alpha)}{\mu}}\right)}{2(s+\alpha)(s+\alpha-\beta)} + \frac{s\alpha\mu\left(1-e^{-\frac{2\beta}{\mu}}\right)}{2\beta(s-\beta)(s+\alpha-\beta)} - \frac{\beta\mu\left(1-e^{-\frac{2s}{\mu}}\right)}{2s(s-\beta)},$$

$$E[D] = \frac{1}{\mu} - \frac{\mu\left(\alpha^2 e^{-\frac{2\beta}{\mu}} - \beta^2 e^{-\frac{2\alpha}{\mu}}\right) - (\alpha-\beta)(\alpha(\mu-2\beta)+\mu\beta)}{2\alpha\beta^2(\alpha-\beta)},$$

$$\tilde{H}(s) = \frac{\mu\left(\beta e^{-\frac{2\alpha}{\mu}} - \alpha e^{-\frac{2\beta}{\mu}} + \alpha - \beta\right)}{2\beta(s+\beta)(\alpha-\beta)},$$

$$E[H] = \frac{\mu\left(\beta e^{-\frac{2\alpha}{\mu}} - \alpha e^{-\frac{2\beta}{\mu}} + \alpha - \beta\right)}{2\beta^2(\alpha-\beta)},$$

$$E[L] = \frac{\lambda(3\mu-\lambda)}{3\mu(\mu-\lambda)} - \frac{\lambda\left(\mu\left(\alpha^2 e^{-\frac{2\beta}{\mu}} - \beta^2 e^{-\frac{2\alpha}{\mu}}\right) - (\alpha-\beta)(\alpha(\mu-2\beta)+\mu\beta)\right)}{2\alpha\beta^2(\alpha-\beta)}.$$

3 Optimal mean orbiting time

This section provides a scheme to set the optimal mean orbiting time $1/\beta$ of an arbitrary customer. Let c be the cost rate of a customer's sojourn time in the service facility, let r be a customer's net rate of utility while in orbit, and let $g > r$ be the penalty each customer is levied for any unit of time s/he resides in orbit after his/her service has been completed. A customer's objective is to maximize the total expected net reward per unit time, Z, by controlling the orbiting rate, β. That is, a customer's optimization problem is

$$\max_{\beta>0}\{Z(\beta) = rE[X]P(T < B) - cE[D] - gE[H]\}, \tag{27}$$

where

$$P(T < B) = \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)f_T(t)dtdb = \int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)\alpha e^{-\alpha t}dtdb = \int_{b=0}^{\infty} f_B(b)(1-e^{-\alpha t})db = 1-\tilde{B}(\alpha). \tag{28}$$

Proposition 1. *For $\beta > 0$, the function $Z(\beta)$ has a global maximum point.*

Proof. The derivative of $Z(\beta)$ is given as follows.

$$\frac{dZ(\beta)}{d\beta} = \frac{-\alpha(c+g)\widetilde{B}'(\beta)\beta(\alpha-\beta) - (r+c)(1-\widetilde{B}(\alpha))\beta^2 + \alpha\Big((c+g)\widetilde{B}(\beta) - (g-r)\widetilde{B}(\alpha) - r - c\Big)(\alpha - 2\beta)}{\beta^2(\alpha-\beta)^2}.$$

Then, the claim is proved by two steps:

i. $\lim_{\beta\to 0^+}\frac{dZ(\beta)}{d\beta} = \lim_{\beta\to 0^+}\frac{(g-r)(1-\tilde{B}(\alpha))}{\beta^2} > 0$, implying that $Z(\beta)$ increases at the origin.
ii. $\lim_{\beta\to \infty}\frac{dZ(\beta)}{d\beta} = \lim_{\beta\to \infty}\frac{-(r+c)(1-\tilde{B}(\alpha))\beta^2}{\beta^4} < 0$, implying that from some value of $0 < \beta < \infty$, the function $Z(\beta)$ monotonically decreases.

From (i) and (ii) it follows that there exists at least one maximum point for $\beta > 0$, one of which is a global maximum point.

Following Proposition 1, the optimal mean orbiting time, $1/\beta^*$, is derived by solving $\frac{dZ(\beta)}{d\beta} = 0$.

For example, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = 1$:

$\frac{dZ(\beta)}{d\beta} = \frac{\alpha(\mu(g-r)(2\beta+\mu)-(c+r)\beta^2)}{(\alpha+\mu)\beta^2(\beta+\mu)^2}$,

and consequently, the unique optimal solution is.

$\beta^* = \frac{\mu\left((g-r)+\sqrt{(c+g)(g-r)}\right)}{c+r}$.

It follows that in this case β^* is a linear function of the service rate μ. That is, if the mean service time is longer so is the optimal mean orbit time. Also, β^* increases when the difference $g - r$ increases.

To illustrate the calculation of the optimal mean orbiting time, $1/\beta^*$, we use the following parameter values: $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, $c = 1$, and $\gamma = 1$. Since $\frac{dZ(\beta)}{d\beta} = 0$ when $\beta = 9.26$, and $\left.\frac{d^2Z(\beta)}{d^2\beta}\right|_{\beta = 9.26} < 0$, the optimal mean orbiting time is $1/\beta^* = 1/9.26$. That is, in the above example, if the mean service time is $E[B] = 1/4 = 0.25$ hours (15 min), and the mean patience time is $E[T] = 1/10 = 0.1$ hours (6 min), the optimal mean orbiting time is $1/\beta^* = 0.1079$ hours (6.5 min).

The behavior of the expected reward per unit time, $Z(\beta)$, as a function of the orbiting rate, β, is depicted in Fig. 7, which illustrates that $Z(\beta)$ is unimodal as a function of β in the current example.

To investigate the effect of each parameter on the optimal orbiting time and the maximal reward, we calculate β^* and $Z(\beta^*)$ for various values of the parameters. Let $\xi = g/r$. Figures 8 and 9 show that β^* increases monotonically with ξ for all $B \sim \Gamma(\gamma, \gamma\mu)$, as well as for $B \sim U(0, 2/\mu)$. That is, as the ratio between g, the penalty of delaying the return from orbit, and the gain r from orbiting, increases, the optimal orbiting time decreases (larger β^*). Moreover, Fig. 8 shows that β^* increases with γ for small values of ξ but decreases with γ for high values of ξ. Figures 9, 10, and 11 depict the values of β^* as a function of ξ when (i) $B \sim U(0, 2/\mu)$ and (ii) $B \sim \Gamma(\gamma, \gamma\mu)$ for given values of the parameters, where $\gamma = 1, 5, \infty$, respectively. It is shown that (i) for small values of γ (which is presented in Fig. 9), when $B \sim U(0, 2/\mu)$, β^* is higher than its value when $B \sim \Gamma(\gamma, \gamma\mu)$ for small values of ξ, but it is lower for high values of ξ; (ii) for intermediate values of γ (which is presented in Fig. 10), β^* for $B \sim U(0, 2/\mu)$ is smaller than β^* for $B \sim \Gamma(\gamma, \gamma\mu)$ for all values of ξ; and (iii) for high values of γ (which is presented in Fig. 11), when $B \sim U(0, 2/\mu)$, β^* is smaller than its corresponding value when $B \sim \Gamma(\gamma, \gamma\mu)$ for small values of ξ, but the opposite occurs for high values of ξ. Figure 12 shows that $Z(\beta^*)$ decreases monotonically with ξ but increases with γ. Figure 13 shows that, as function of ξ, the optimal value of $Z(\beta^*)$ when $B \sim U(0, 2/\mu)$ lies uniformly between its values under the two extreme cases of γ when $B \sim \Gamma(\gamma, \gamma\mu)$. That is, it is lower when $\gamma \to \infty$ (Deterministic service time) and it is higher when $\gamma = 1$ (Exponential service

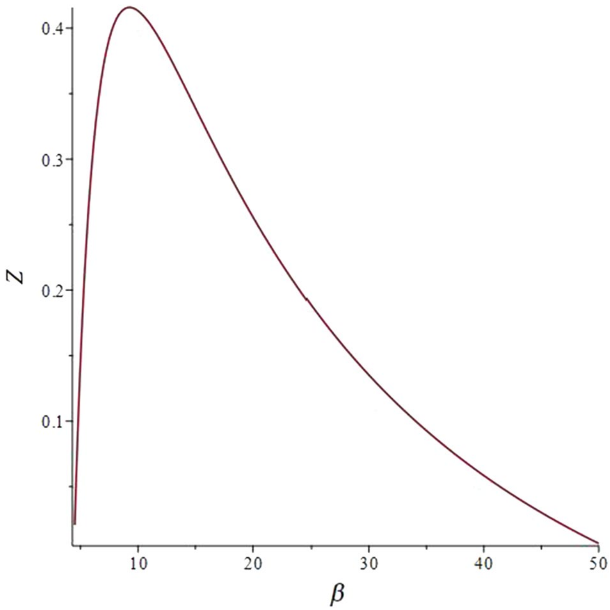


Fig. 7 $Z[\beta]$ as a function of β when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, $c = 1$, and $\gamma = 1$

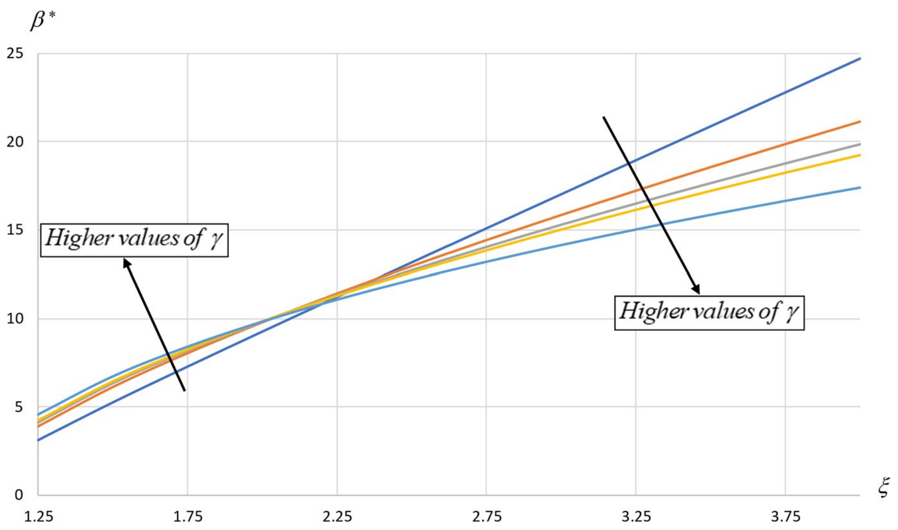


Fig. 8 β^* as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = \{1, 3, 5, 7, \infty\}$

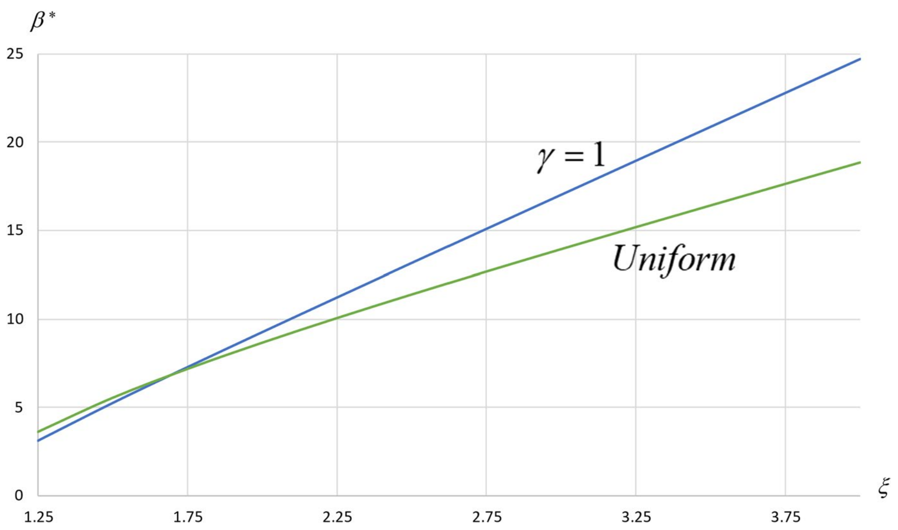


Fig. 9 β^* as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = 1$ and for $B \sim U(0, 2/\mu)$

time). Figure 14 shows that β^* remains constant with α for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = 1$. Conversely, β^* decreases monotonically with α for higher values of γ and for $B \sim U(0, 2/\mu)$. That is, the shorter the mean patience time, the higher is the mean orbit time. Figure 15 shows that $Z(\beta^*)$ increases monotonically with α.

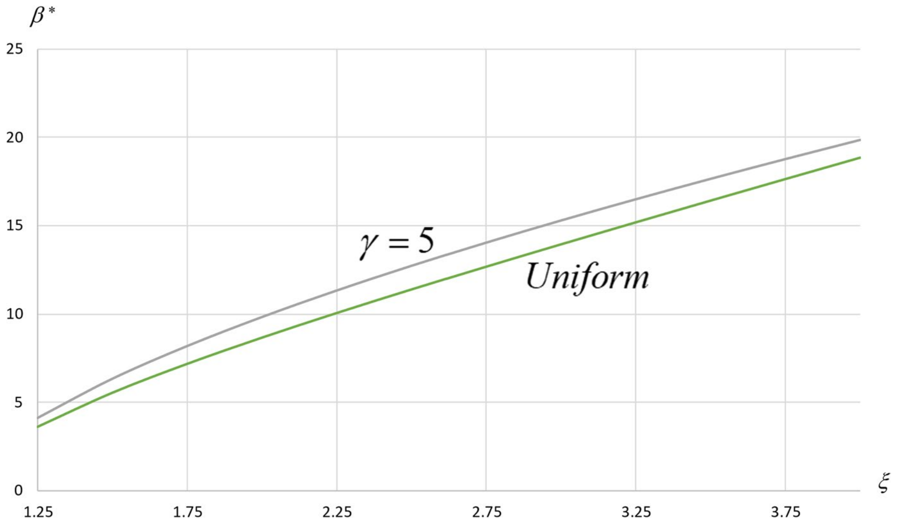


Fig. 10 β^* as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = 5$ and for $B \sim U(0, 2/\mu)$

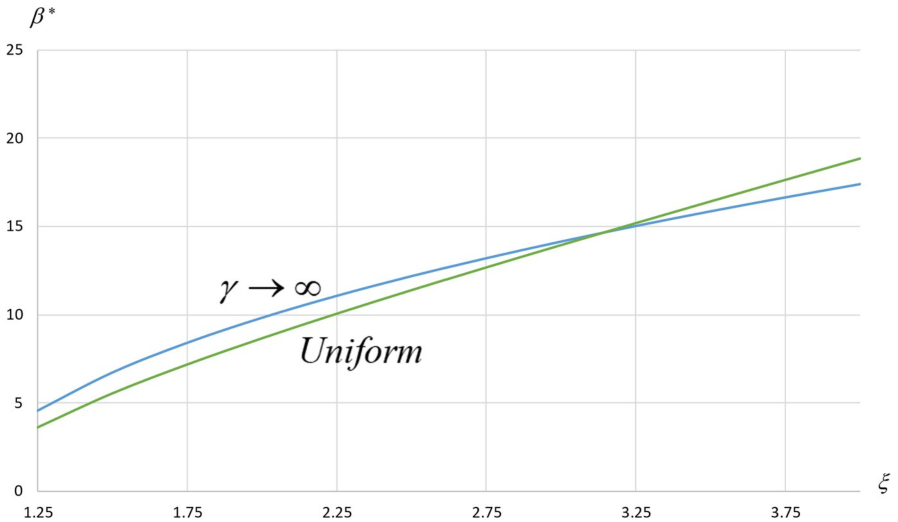


Fig. 11 β^* as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma \to \infty$, and for $B \sim U(0, 2/\mu)$

4 Discussion

This paper analyzes a previously unstudied real-life queueing system, where impatient customers may tentatively depart and "go to orbit" while their service is still being processed. While customers benefit from the time they spend in orbit, they may be penalized for any unit of time they are overdue. We provide a thorough

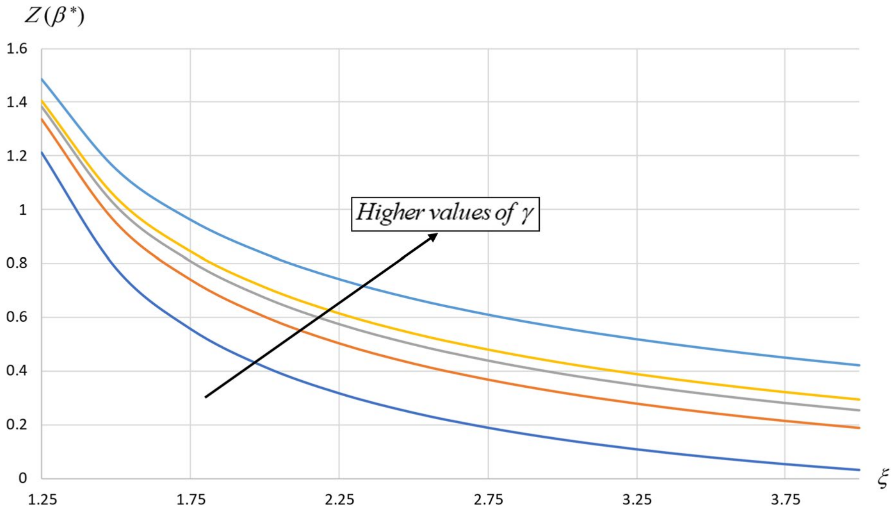


Fig. 12 $Z(\beta^*)$ as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = \{1, 3, 5, 7, \infty\}$

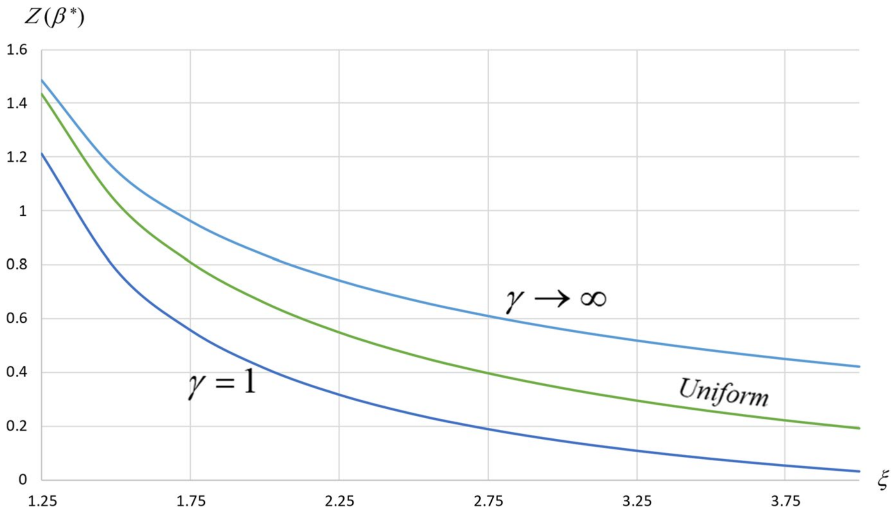


Fig. 13 $Z(\beta^*)$ as a function of ξ, when $\lambda = 3$, $\mu = 4$, $\alpha = 10$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = \{1, \infty\}$, and for $B \sim U(0, 2/\mu)$

probabilistic analysis for a wide range of service-time probability distributions. Explicit results are derived for various performance measures, such as a customer's total residence time in the system; a customer's net actual residence time in the system during service (not including orbit time); the time an orbiting customer is overdue, that is, remains in orbit after his/her service has been completed; and the total number of customers in the system. Numerical examples are presented, where

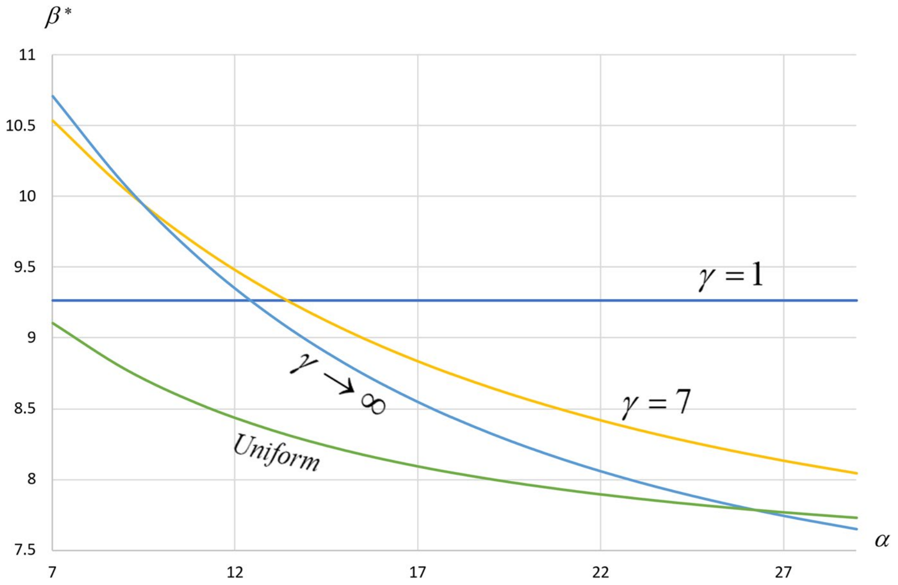


Fig. 14 β^* as a function of α, when $\lambda = 3$, $\mu = 4$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = \{1, 7, \infty\}$, and for $B \sim U(0, 2/\mu)$

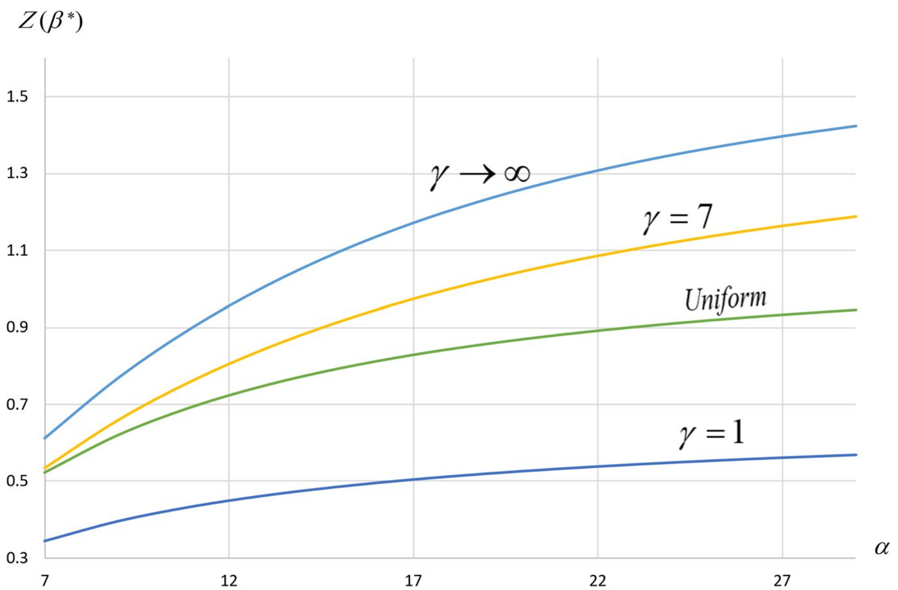


Fig. 15 $Z(\beta^*)$ as a function of α, when $\lambda = 3$, $\mu = 4$, $r = 20$, $g = 40$, and $c = 1$, for $B \sim \Gamma(\gamma, \gamma\mu)$, $\gamma = \{1, 7, \infty\}$, and for $B \sim U(0, 2/\mu)$

the impacts of the various parameters are visualized in several graphs. The optimal mean orbit time is calculated for each case of service-time distribution. Implementing our results is straightforward, requiring only the calculation of the mean optimal customer orbit time.

The investigation provided in this work shows that the proposed "orbit while in service" policy significantly increases the service system's efficiency by reducing customers waiting times, as well as increasing their utility by allowing them to use their time efficiently.

The study also shows that the mean time an orbiting customer is overdue ($E[D]$) increases with γ (the shape parameter of the Gamma service-time distribution). For example, $E[D]$ is higher in the Deterministic service-time case than in the Exponential case, which is explained by the higher variance of the service-time distribution. The latter result implies that customers prefer γ to be as low as possible. Conversely, a customer's net actual residence time in the system during service (not including orbit time) decreases with γ. Based on this observation, customers prefer that γ would be as high as possible. The combination of the two results raises the question regarding customers' preferences. Considering the customers' total utility, we show that the customers prefer γ to be as high as possible. Thus, service system managers are recommended to decrease the variance of the service time, which can be accomplished by establishing a suitable workplace.

Another interesting result is that for high values of the penalty rate of returning late from orbit ($\xi = g/r$), as γ increases (implying that the variance of the service time decreases), it is better for customers to spend more time in orbit. This counterintuitive result is explained as follows: when ξ is high, the optimal orbit time is relatively small, implying that the customer will return late from orbit only if the service duration is relatively low. Thus, when the variance of the service time decreases and consequently the probability of low service duration decreases, the customer can spend more time in orbit. Similar results are obtained when comparing the $B \sim U(0, 2/\mu)$ service time case with the $B \sim \Gamma(\gamma, \gamma\mu)$ case. Notably, the variance in the $B \sim U(0, 2/\mu)$ case is higher than that in the $B \sim \Gamma(\gamma, \gamma\mu)$ case when $\gamma > 3$.

Additionally, it is shown that the optimal orbiting time for the Exponential case is robust to changes in customer patience time, which is explained by the memoryless property. However, for other service-time distributions, the shorter mean patience time leads to a higher optimal mean orbit time. The latter result is intuitively understandable since the sooner the customer enters orbit, the more time remains for the service completion. Finally, it is shown that for the Exponential service time the optimal mean orbit time is a linear function of the mean service time.

For future research we suggest extending the model to a multi-server case and investigating the associated optimal staffing problem.

Appendix A: list of notations

λ—customers' arrival rate.
B—service time.
$f_B(t)$—probability density function of B.
$F_B(t)$—cumulative distribution function of B.
$\tilde{B}(s)$—LST of B.
T—customer's patience time, Exponentially distributed with mean $1/\alpha$.
X—orbit time, Exponentially distributed with parameter β

$$Y = T + X$$

V—customer's total residence time in the system measured from the instant s/he starts service until departure.
$\tilde{V}(s)$—LST of V.
D—customer's net actual residence time in the system during his/her service time (not including orbit time).
$\tilde{D}(s)$—LST of D.
W_q—customer's waiting time in queue.
W—customer's sojourn time in the system (not including orbit time).
H—length of time a customer remains in orbit after his/her service has been completed.
$\tilde{H}(s)$—LST of H.
L—number of customers in the system (not including those in orbit).
$S \in \{1, 2, 3\}$—customer's stage: stage 1—before entering orbit; stage 2—in orbit; stage 3—after returning from orbit.
U—elapsed time since start of service of a tagged customer (including orbit if applicable).
$\{L, S, U\}$—system's state.
$p(n, m, u)$—density of state $L = n$, $S = m$ and $U = u$.
$G_m(z, u)$—PGF of L as a function of u, given $S = m$.
$\mu(u)$—hazard rate function of the service time B.
$G_m(z)$—PGF of L, given $S = m$.
c—cost rate of a customer's sojourn time in the service facility.
r—customer's net rate of utility while in orbit.
g—a penalty each customer is levied for any unit of time s/he resides in orbit after his/her service has been completed.
$Z(\beta)$—customers' total expected net reward per unit time as a function of β.

Appendix B: derivation of Term 3 in Theorem 2

$$
\begin{aligned}
&E[e^{-s(B-X)}|T+X<B]P(T+X<B)= \\
&\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=0}^{b-t} e^{-s(b-x)}f_B(b)f_T(t)f_X(x)dxdtdb=\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=0}^{b-t} e^{-sb}f_B(b)f_T(t)e^{sx}\beta e^{-\beta x}dxdtdb \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=0}^{b-t} e^{-sb}f_B(b)f_T(t)\beta e^{-(\beta-s)x}dxdtdb=\int_{b=0}^{\infty}\int_{t=0}^{b} e^{-sb}f_B(b)f_T(t)\frac{\beta}{\beta-s}(1-e^{-(\beta-s)(b-t)})dtdb \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b} e^{-sb}f_B(b)\alpha e^{-\alpha t}\frac{\beta}{\beta-s}(1-e^{-(\beta-s)(b-t)})dtdb=\int_{b=0}^{\infty}\int_{t=0}^{b} e^{-sb}f_B(b)\alpha\frac{\beta}{\beta-s}(e^{-\alpha t}-e^{-(\beta-s)(b-t)-\alpha t})dtdb= \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b} e^{-sb}f_B(b)\alpha\frac{\beta}{\beta-s}(e^{-\alpha t}-e^{-(\beta-s)b}e^{-(\alpha-\beta+s)t})dtdb \\
&=\int_{b=0}^{\infty} e^{-sb}f_B(b)\frac{\alpha\beta}{\beta-s}\left((1-e^{-\alpha b})\frac{1}{\alpha}-(1-e^{-(\alpha-\beta+s)b})\frac{e^{-(\beta-s)b}}{(\alpha-\beta+s)}\right)db \\
&=\int_{b=0}^{\infty} e^{-sb}f_B(b)\frac{\alpha\beta}{\beta-s}\left((1-e^{-\alpha b})\frac{1}{\alpha}-(e^{-(\beta-s)b}-e^{-\alpha b})\frac{1}{(\alpha-\beta+s)}\right)db \\
&=\int_{b=0}^{\infty} f_B(b)\frac{\alpha\beta}{\beta-s}\left((e^{-sb}-e^{-(s+\alpha)b})\frac{1}{\alpha}-(e^{-\beta b}-e^{-(s+\alpha)b})\frac{1}{(\alpha-\beta+s)}\right)db \\
&=\int_{b=0}^{\infty} f_B(b)\frac{\alpha\beta}{\beta-s}\left(e^{-sb}\frac{1}{\alpha}-e^{-\beta b}\frac{1}{(\alpha-\beta+s)}+e^{-(s+\alpha)b}\left(\frac{1}{(\alpha-\beta+s)}-\frac{1}{\alpha}\right)\right)db \\
&=\frac{\alpha\beta}{\beta-s}\left(\frac{1}{\alpha}\tilde{B}(s)-\frac{1}{(\alpha-\beta+s)}\tilde{B}(\beta)+\left(\frac{1}{(\alpha-\beta+s)}-\frac{1}{\alpha}\right)\tilde{B}(s+\alpha)\right).
\end{aligned}
$$

Appendix C: calculation of $\widetilde{X}(s)P(T<B<T+X)$

$$
\begin{aligned}
\int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} f_B(b)f_T(t)f_X(x)dxdtdb &= \int_{b=0}^{\infty}\int_{t=0}^{b}\int_{x=b-t}^{\infty} f_B(b)f_T(t)\beta e^{-\beta x}dxdtdb \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)f_T(t)e^{-\beta(b-t)}dtdb \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)\alpha e^{-\alpha t}e^{-\beta(b-t)}dtdb \\
&=\int_{b=0}^{\infty}\int_{t=0}^{b} f_B(b)\alpha e^{-(\alpha-\beta)t}e^{-\beta b}dtdb \\
&=\int_{b=0}^{\infty} f_B(b)\frac{\alpha}{\alpha-\beta}(1-e^{-(\alpha-\beta)b})e^{-\beta b}db \\
&=\frac{\alpha}{\alpha-\beta}\left(\tilde{B}(\beta)-\tilde{B}(\alpha)\right).
\end{aligned}
$$

Funding Open access funding provided by Ariel University.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Afanasyev GA (2023) Cost optimization of queueing systems with vacations. Theory Probab Appl 68:308–315

Altman E, Yechiali U (2006) Analysis of customers' impatience in queues with server vacations. Queueing Syst 52:261–279

Avrachenkov K, Yechiali U (2010) On tandem blocking queues with a common retrial queue. Comput Oper Res 37:1174–1180

Avrachenkov K, Nain P, Yechiali U (2014) A retrial system with two input streams and two orbit queues. Queueing Syst 77:1–31

Ayhan H (2022) Server assignment policies in queues with customer abandonments. Queueing Syst. 1–3

Bassamboo A, Randhawa R, Wu C (2023) Optimally scheduling heterogeneous impatient customers. Manuf Serv Oper Manag 25:1066–1080

Bouchentouf AA, Guendouzi A (2019) Cost optimization analysis for an MX/M/c vacation queueing system with waiting servers and impatient customers. SeMA J 76:309–341

Bouchentouf AA, Cherfaoui M, Boualem M (2021) Analysis and performance evaluation of Markovian feedback multi-server queueing model with vacation and impatience. Am J Math Manag Sci 40:261–282

Bouchentouf AA, Boualem M, Yahiaoui L, Ahmad H (2022) A multi-station unreliable machine model with working vacation policy and customers' impatience. Qual Technol Quant Manag 19:766–796

Bouchentouf AA, Guendouzi A, Majid S (2020) On impatience in Markovian M/M/1/N/DWV queue with vacation interruption. Croat Oper Res Rev 21–37

Boxma OJ, Schlegel S, Yechiali U (2002) A note on an M/G/1 queue with a waiting server, timer and vacations. In: American Mathematical Society Translations. Citeseer

Cherfaoui M, Bouchentouf AA, Boualem M (2023) Modelling and simulation of Bernoulli feedback queue with general customers' impatience under variant vacation policy. Int J Oper Res 46:451–480

Dimitriou I (2023) A single server retrial queue with event-dependent arrival rates. Ann Oper Res 1–36

Do NH, Van Do T, Melikov A (2020) Equilibrium customer behavior in the M/M/1 retrial queue with working vacations and a constant retrial rate. Oper Res 20:627–646

Dong J, Ibrahim R (2021) SRPT scheduling discipline in many-server queues with impatient customers. Manage Sci 67:7708–7718

Economou A, Logothetis D, Manou A (2022) The value of reneging for strategic customers in queueing systems with server vacations/failures. Eur J Oper Res

Fiems D (2023) Retrial queues with constant retrial times. Queueing Syst 103:347–365

Firouz M, Li L, Keskin BB (2022) Managing equipment rentals: Unreliable fleet, impatient customers, and finite commitment capacity. Prod Oper Manag 31:3963–3981

Gao S, Dong H, Wang X (2021) Equilibrium and pricing analysis for an unreliable retrial queue with limited idle period and single vacation. Oper Res 21:621–643

Hanukov G, Anily S, Yechiali U (2020) Ticket queues with regular and strategic customers. Queueing Syst 95:145–171. https://doi.org/10.1007/S11134-020-09647-X/FIGURES/8

Hanukov G (2023) A queueing-inventory system with a repeated-orbit policy during the service. Ann Oper Res 1–33

Harchol-Balter M (2013) Performance modeling and design of computer systems: queueing theory in action. Cambridge University Press

Jain M, Kaur S, Singh P (2021) Supplementary variable technique (SVT) for non-Markovian single server queue with service interruption (QSI). Oper Res 1–44

Kella O, Yechiali U (1988) Priorities in M/G/1 queue with server vacations. Nav Res Logist 35:23–34

Kleiner I, Frostig E, Perry D (2021) A Decomposition Property for an MX/G/1 Queue with Vacations. arXiv Prepr arXiv211004557

Krishna Kumar B, Rukmani R, Thanikachalam A, Kanakasabapathi V (2018) Performance analysis of retrial queue with server subject to two types of breakdowns and repairs. Oper Res 18:521–559

Kumar MS, Dadlani A, Kim K (2020) Performance analysis of an unreliable M/G/1 retrial queue with two-way communication. Oper Res 20:2267–2280

Kumar A, Boualem M, Bouchentouf AA (2022) Optimal analysis of machine interference problem with standby, random switching failure, vacation interruption and synchronized reneging. In: Applications of advanced optimization techniques in industrial engineering. CRC Press, pp 155–168

Lee SW, Kim B, Kim J (2022) Analysis of the waiting time distribution in M/G/1 retrial queues with two way communication. Ann Oper Res 1–14

Levy Y, Yechiali U (1975) Utilization of idle time in an M/G/1 queueing system. Manage Sci 22:202–211

Levy Y, Yechiali U (1976) An M/M/s queue with servers' vacations. INFOR Inf Syst Oper Res 14:153–163

Liu J, Wang J (2017) Strategic joining rules in a single server Markovian queue with Bernoulli vacation. Oper Res 17:413–434

Manitz M, Piehl MP (2023) A fast staffing algorithm for multistage call centers with impatient customers and time-dependent overflow. Cent Eur J Oper Res 1–29

Melikov A, Chakravarthy SR, Aliyeva S (2023) A retrial queueing model with feedback. Queueing Model Serv Manag 6:63–95

Nazarov A, Phung-Duc T, Paul S, Lizyura O (2022) Diffusion limit for single-server retrial queues with renewal input and outgoing calls. Mathematics 10:948

Nithya N, Anbazhagan N, Amutha S et al (2023) Controlled arrivals on the retrial queueing-inventory system with an essential interruption and emergency vacationing server. Mathematics 11:3560

Polas MRH, Rahman MM, Miah MA, Hayash MMA (2018) The impact of waiting time towards customers satisfaction in fast food establishments: evidence from Bangladesh. IOSR J Bus Manag 20:11–21

Rosenberg E, Yechiali U (1993) The MX/G/1 queue with single and multiple vacations under the LIFO service regime. Oper Res Lett 14:171–179

Ross SM (1996) Stochastic processes. Wiley, New York

Sakuma Y, Boxma O, Phung-Duc T (2021) An M/PH/1 queue with workload-dependent processing speed and vacations. Queueing Syst 98:373–405

Shajin D, Krishnamoorthy A (2021) On a queueing-inventory system with impatient customers, advanced reservation, cancellation, overbooking and common life time. Oper Res 21:1229–1253

Sherzer E, Kerner Y (2018) Customers' abandonment strategy in an M/G/1 queue. Queueing Syst 90:65–87

Sindhu S, Krishnamoorthy A, Kozyrev D (2023) On Queues with working vacation and interdependence in arrival and service processes. Mathematics 11:2280

Suranga Sampath MIG, Liu J (2020) Impact of customers' impatience on an M/M/1 queueing system subject to differentiated vacations with a waiting server. Qual Technol Quant Manag 17:125–148

Templeton JGC, Falin GI (2023) Retrial queues. Taylor & Francis

Tian N, Zhang ZG (2006) Vacation queueing models: theory and applications. Springer Science & Business Media

Wang S, Xu X (2021) Equilibrium strategies of the fluid queue with working vacation. Oper Res 21:1211–1228

Xu J, Liu L, Wu K (2022) Analysis of a retrial queueing system with priority service and modified multiple vacations. Commun Stat Methods 1–25

Yechiali U (2007) Queues with system disasters and impatient customers when system is down. Queueing Syst 56:195–202

Yin M, Yan M, Guo Y, Liu M (2023) Analysis of a pre-emptive two-priority queuing system with impatient customers and heterogeneous servers. Mathematics 11:3878

Zhang Y, Wang J (2023) Managing retrial queueing systems with boundedly rational customers. J Oper Res Soc 74:748–761