Queues with service resetting

Ofek Lauber Bonomo[1,*] Uri Yechiali [2,†] and Shlomi Reuveni[1,‡]
[1]*School of Chemistry, Center for the Physics & Chemistry of Living Systems, Ratner Institute for Single Molecule Chemistry, and the Sackler Center for Computational Molecular & Materials Science, Tel Aviv University, 6997801, Tel Aviv, Israel*
[2]*Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel*
(Dated: October 31, 2023)

The interplay between the arrival rate of jobs to a queue and their mean service time is a key determinant of queue performance. Not less important are stochastic fluctuations in arrival and service times, which heavily affect queue performance causing long waiting times and queue backlogs. Various scheduling strategies and priority policies have been applied to mitigate the detrimental effects of these stochastic fluctuations. However, the implementation of such policies is mainly based on job size information which is not always available. Recently, it was shown by Bonomo et al. [1] that a simple service resetting policy can be used when scheduling policies are not applicable. Specifically, it was shown that when service times are solely determined by factors intrinsic to the server, service resetting can drastically improve queue performance. However, as observed by Gardner et al. [2], in many queueing systems, service times have two independent sources: the intrinsic server slowdown (S) and the jobs' inherent size (X). In these, so-called $S\&X$ queues, service resetting results in a newly drawn server slowdown while the inherent job size remains unchanged. To illuminate the advantages and limitations of service resetting in these realistic queueing systems, we present a comprehensive analysis of $S\&X$ queues with service resetting. To this end, we consider cases where the total service time is either a sum or a product of the service slowdown and the jobs' inherent size. We derive the total service time distribution under a generic service resetting policy for both the additive and multiplicative models. In addition, for the interesting special case of a constant-rate Poisson resetting process, we derive explicit conditions under which resetting reduces the mean service time and improves queue performance. All these results hold regardless of the arrival process. For the canonical M/G/1 queue, we further expand the analysis and derive the corresponding distribution of the number of jobs in the system and their sojourn time. Our analysis highlights situations where service resetting can be used as an effective tool to improve the performance of $S\&X$ queueing systems.

1. INTRODUCTION

Waiting in line is a frustrating yet common experience. Whether it is when standing in a long supermarket line or when waiting for a traffic light to turn green, queues appear ubiquitously in our life. Prominent examples include: call centers [3, 4], airplane boarding [5, 6], telecommunication and computer systems [7–9], production and manufacturing lines [10], enzymatic and metabolic pathways [11–14], gene expression [15–19], and transport phenomena [20–23].

The waiting time in a queue heavily depends on the arrival rate of customers and on the service rate of the server. Fluctuations in service time are another important factor that affect the waiting time. While the cashier at a supermarket works at a (roughly) constant rate, other servers, e.g., computer systems [9], and molecular machines like enzymes [24–26], often display pronounce fluctuations in service times. These large fluctuations are a major source of backlogs and delays [27].

To mitigate the detrimental effects large fluctuations in service times have on queue length and waiting times, various strategies have been developed. One major class is scheduling policies [9, 27, 28]. A scheduling policy is a rule which dictates the order in which jobs in a queue are served. For example, in a single server queue, the server can reduce the overall mean waiting time by serving shorter jobs first [29]. In cases where service can be stopped and resumed later on, performance can be further improved by applying the shortest remaining time discipline [30, 31].

Scheduling policies have proven useful in a wide variety of scenarios, yet they suffer from clear limitations. For example, the implementation of size-based scheduling policies is impossible when jobs' sizes are unknown a priori. A prominent example is given by situations where fluctuations in service times are *intrinsic* to the server itself. Such scenarios are fairly common, e.g., in stochastic optimization algorithms which can take markedly different times to run on two instances of the *exact same* problem [32]. Similarly, the loading time of an internet page is stochastic, due to the varying workload on the server, although the size of the page itself does not change [33]. In such cases, new strategies are needed in order to solve the problems caused by large service time fluctuations.

In a recent paper, Bonomo et al. [1] explored the prospect of mitigating long queues and waiting times with service resetting. It was shown that a simple service resetting mechanism can reverse the deleterious effects of large fluctuations in service times when these are solely *intrinsic* to the server. Nevertheless, it is clear that resetting may fail when service times are mainly determined by factors extrinsic to the server. A trivial example is a supermarket cashier. There, the teller serves the customers at a (roughly) constant rate, and the service duration is thus determined by the number of items purchased before arriving at the counter. Evidently, resetting service and starting the charging process anew would result in wasted time. Consequently, one should avoid resetting in cases where service times are completely determined by factors extrinsic to the server.

Having discussed the extreme cases of service times that are completely determined either by extrinsic or intrinsic factors, we hereby focus on the effect of service resetting in scenar-

ios where service times have contributions from both intrinsic and extrinsic factors, which is the case in numerous queues. In traditional queueing theory, a single random variable is assigned to represent service times, making it impossible to differentiate between extrinsic and intrinsic components. This issue was tackled in a paper by Gardner et al. [2]. There, the authors introduced the *S&X* model in which the service time is composed of contributions from two independent sources. The first source, the so-called server slowdown, is intrinsic to the server and denoted by S. The second source, extrinsic to the server, is the jobs' inherent size, denoted by X (hence the name of the model). The *S&X* model provides a framework to study the effects of service resetting on realistic queues in which service times are combinations of intrinsic and extrinsic factors, as illustrated in Fig. 1. Here, we generalize the results obtained in [1] to account for the aforementioned cases, by introducing resetting to the *S&X* model.

Consider a queueing process with a generic service time that is composed of the two sources, S and X, as discussed above. Following Gardner et al. [2], we let S and X be two independent, generic random variables, from which the server's slowdown and the jobs' sizes are taken, respectively. Resetting is introduced using a random time R, independent of both X and S. Namely, if service is not completed by R, it is halted, any previous progress is discarded, and the service is immediately restarted. Upon resetting, a new slowdown time, S', is drawn, where S and S' are independent and identically distributed. However, the inherent size of the job in service, $X = x$, remains the same. This combined process is repeated until the service is completed.

The paper is structured as follows. In Sec. 2, we briefly review recent developments in stochastic resetting and applications. In Sec. 3, we explore an additive *S&X* model, where the overall service time distribution is a sum of the server slowdown S and the jobs' inherent size X. For this model, we derive the service time distribution under a generally distributed resetting time. In Sec. 3a, we focus on the case of Poissonian resetting, i.e., resetting at a constant rate. For this case of interest, we derive an explicit condition for resetting to be advantageous, i.e., to improve queue performance by lowering the mean service time of a job. In Sec. 4, we explore a multiplicative model, where the overall service time distribution is given by a product of S and X. Here too, we first derive the service time distribution under a generally distributed resetting time, and in Sec. 4a concentrate on Poissonian resetting for which we obtain an explicit condition for resetting to be beneficial. We indicate that the results in Secs. 3 and 4 hold true regardless of the arrival process. In Sec. 5, we analyze the celebrated M/G/1 queue with service resetting, and obtain the full distribution of the number of jobs in the system for both the additive and multiplicative *S&X* models. We end in Sec. 6 with conclusions and outlook. The analytical results obtained in this paper are illustrated using several examples and are further corroborated by numerical simulations.

In what follows we use $f_Z(\cdot)$, $\mathbf{E}[Z]$, $\mathrm{Var}(Z)$, $\sigma(Z)$, and $\tilde{Z}(s) \equiv \mathbf{E}\left[e^{-sZ}\right]$ to denote, respectively, the probability density function (PDF), expectation, variance, standard deviation, and Laplace transform of a non-negative random variable Z.

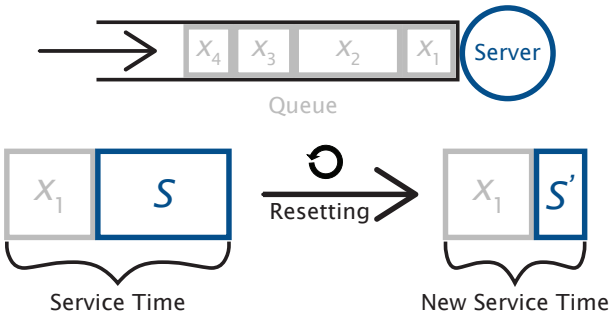


FIG. 1. *S&X* model with service resetting. Jobs of sizes $x_1, x_2, x_3, \ldots$ arrive at a service station. The initial required service time of the i-th job, $f(x_i, S)$, is a function of its inherent size x_i and the service slowdown S, e.g., $f(x_i, S) = x_i + S$. A service resetting policy is applied using a timer R that can be deterministic or random. If $f(x_i, S) < R$ the service is completed at the required service time. Otherwise, a new server slowdown S' is drawn, and service starts again with a new required service time $f(x_i, S')$ and a new resetting time R'. This process is repeated until service is completed.

2. STOCHASTIC RESETTING

Stochastic resetting has attracted a lot of attention in recent years and has been the subject of vigorous studies in physics [34–41], chemical and biological process [42–47], computer science [48, 49], operations research [50–55], and other cross-disciplinary fields. We refer the readers to [56] for extensive review and [57] for discussion on the connection with the inspection paradox. The subject has also seen advances through single particle experiments using optical tweezers [58, 59].

A hallmark of resetting is its ability to reduce the mean and fluctuations of the first passage time to a target. It was shown that by using different resetting mechanisms, namely, terminating a process and bringing it back to its initial state according to different time distributions, one can expedite the completion of stochastic processes [60–68]. This property of stochastic resetting was studied and utilized in the context of queueing theory as well.

The effects of constant time resetting (aka sharp resetting) were studied in [33]. There, the authors explored the effects of sharp resetting on the expected average waiting time required to execute electronic transactions in a network environment, e.g. internet. The authors show that for transactions with log-normally distributed service time, there exists some optimal resetting time τ^* for which the expected average waiting time is minimal and lower than the original expected average waiting time (without resetting). Using mean-field arguments and computer simulations the authors show that this resetting strategy remains advantageous, in terms of expected average waiting time, even when all the users in the network use it.

This constant time resetting strategy was also explored in [69] and [70]. In these works, customers have the option of resetting their service in constant time, i.e., resubmitting a service request (aka *timeout*). The optimal resetting time in terms of minimal total expected cost [69] or minimal expected waiting time [70] is then calculated. Before moving forward, we mention that the concept of resetting is also relevant in the context of reliability theory [71, 72] and income dynamics [73, 74]. Resetting also appears when considering queues with catastrophic events [75–78]. However, such catastrophic events lead to the clearing of jobs from the queue, whereas service resetting conserves the number of jobs in the queue.

3. ADDITIVE MODEL

Assume that the overall service time in the absence of resetting, B, is given as a sum of the two independent sources defined above, such that $B = S + X$. Let B_R denote the total service time until completion under a generic service resetting policy. Given an inherent job size of x, the conditional service time $B_R(x) \equiv \{B_R | X = x\}$ obeys the following renewal equation

$$B_R(x) = \begin{cases} x+S & \text{if} \quad x+S < R \\ R + B'_R(x) & \text{if} \quad x+S \geq R, \end{cases} \tag{1}$$

where R is a random resetting time drawn from a distribution with density $f_R(\cdot)$, and $B'_R(x)$ is an independent and identically distributed copy of $B_R(x)$. Under the additive model (indicated by the $+$ sign), it is useful to define the following conditional random variables

$$S_+(x) = \{S | x+S < R\}, \tag{2}$$

$$R_+(x) = \{R | x+S \geq R\}, \tag{3}$$

and their corresponding Laplace transforms $\tilde{S}_+(x;s)$ and $\tilde{R}_+(x;s)$

$$\tilde{S}_+(x;s) = \int_0^\infty e^{-st} f_{S_+}(t)\, dt, \tag{4}$$

$$\tilde{R}_+(x;s) = \int_0^\infty e^{-st} f_{R_+}(t)\, dt. \tag{5}$$

In the above, we used the probability density functions of the conditional random variables. These are given by

$$f_{S_+}(t) = \frac{f_S(t)}{\Pr(x+S<R)} \int_{t+x}^\infty f_R(\tau)\, d\tau, \tag{6}$$

$$f_{R_+}(t) = \frac{f_R(t)}{\Pr(x+S\geq R)} \int_{t-x}^\infty f_S(\tau)\, d\tau. \tag{7}$$

Note that in Eq. (7), for $t < x$ the lower integration limit is set to 0.

Taking the expectation of both sides of Eq. (1) we get

$$\begin{aligned} \mathbf{E}[B_R(x)] =& \mathbf{E}[(x+S)|x+S<R]\Pr(x+S<R) \\ &+ \mathbf{E}[(R+B'_R(x))|x+S\geq R]\Pr(x+S\geq R). \end{aligned} \tag{8}$$

Using Eqs. (2) and (3) and rearranging terms in Eq. (8), one obtains the mean conditional service time under resetting

$$\mathbf{E}[B_R(x)] = x + \mathbf{E}[S_+(x)] + \frac{1-\Pr(x+S<R)}{\Pr(x+S<R)} \mathbf{E}[R_+(x)], \tag{9}$$

where $\mathbf{E}[S_+(x)]$ and $\mathbf{E}[R_+(x)]$ are the means of the random variables $S_+(x)$ and $R_+(x)$ respectively.

Integrating Eq. (9) with respect to the probability density of the jobs' inherent size, $f_X(x)$, one obtains the mean service time under resetting

$$\begin{aligned} \mathbf{E}[B_R] =& \mathbf{E}[X] + \int_0^\infty \mathbf{E}[S_+(x)] f_X(x)\, dx \\ &+ \int_0^\infty \frac{\Pr(x+S\geq R)}{1-\Pr(x+S\geq R)} \mathbf{E}[R_+(x)] f_X(x)\, dx. \end{aligned} \tag{10}$$

The equation above for the mean service time under resetting can be explained as follows. Each service duration under resetting is composed of the following contributions: The first two terms on the right-hand side of Eq. (10) account for the mean service time of the last service trial. The last term accounts for the time wasted in the geometric sum of failed service attempts. Namely, this is the total service time of all uncompleted service trials.

We now turn to derive the Laplace transform of B_R. Taking the Laplace transform of both sides of Eq. (1) and rearranging terms we obtain

$$\tilde{B}_R(x;s) = \frac{e^{-sx}\tilde{S}_+(x;s)\Pr(x+S<R)}{1-\tilde{R}_+(x;s)\Pr(x+S\geq R)}. \tag{11}$$

Integrating Eq. (11) with respect to the probability density of the jobs' inherent size, $f_X(x)$, one obtains the Laplace transform of the service time under generic resetting

$$\tilde{B}_R(s) = \int_0^\infty \frac{e^{-sx}\tilde{S}_+(x;s)\Pr(x+S<R)}{1-\tilde{R}_+(x;s)\Pr(x+S\geq R)} f_X(x)\, dx. \tag{12}$$

3a. Constant-rate Poisson resetting

Consider the above-detailed service process, and further assume that resetting times are taken from an exponential distribution with rate parameter r. Namely, the probability density of the resetting time distribution is given by

$$f_R(t) = re^{-rt}, \tag{13}$$

for $t \geq 0$. In this case, we have

$$\begin{aligned} \Pr(x+S<R) &= \int_0^\infty f_R(t)\Pr(x+S<t)\, dt \\ &= \int_x^\infty re^{-rt}\left(\int_0^{t-x} f_S(\tau)\, d\tau\right) dt = e^{-rx}\tilde{S}(r), \end{aligned} \tag{14}$$

where $\tilde{S}(r)$ is the Laplace transform of S, evaluated at the resetting rate r. Note that in moving from the first to the second

equality in Eq. (14), we used $\Pr(x+S<t)=0$ for $t<x$. The last step then follows by a change of variable and order of integration.

Substituting the above result into Eqs. (6) and (7), allows us to obtain neat expressions for the Laplace transforms of the conditional random variables $S_+(x)$ and $R_+(x)$. These are given by

$$\begin{aligned}\tilde{S}_+(x;s) &= \int_0^\infty e^{-st} f_{S_+}(t)\,dt \\ &= \frac{\int_0^\infty e^{-st} f_S(t)\left(\int_{t+x}^\infty r e^{-r\tau}\,d\tau\right)dt}{\Pr(x+S<R)} \\ &= \frac{\int_0^\infty e^{-t(r+s)-rx} f_S(t)\,dt}{e^{-rx}\tilde{S}(r)} \\ &= \frac{\int_0^\infty e^{-t(r+s)} f_S(t)\,dt}{\tilde{S}(r)} \\ &= \frac{\tilde{S}(s+r)}{\tilde{S}(r)},\end{aligned} \tag{15}$$

$$\begin{aligned}\tilde{R}_+(x;s) &= \int_0^\infty e^{-st} f_{R_+}(t)\,dt \\ &= \frac{\int_0^\infty e^{-st}\, r e^{-rt}\left(\int_{t-x}^\infty f_S(\tau)\,d\tau\right)dt}{\Pr(x+S\geq R)} \\ &= \frac{r\int_0^\infty e^{-(s+r)t}\left(1-\int_0^{t-x} f_S(\tau)\,d\tau\right)dt}{1-e^{-rx}\tilde{S}(r)} \\ &= \frac{r}{s+r}\,\frac{1-e^{-(s+r)x}\tilde{S}(s+r)}{1-e^{-rx}\tilde{S}(r)},\end{aligned} \tag{16}$$

where in the last step in Eq. (16) we used a known formula for the Laplace transform of a time-domain integration: $\int_0^\infty dt\ \left(\int_0^t d\tau\ g(\tau)\right)e^{-rt}=\frac{\tilde{g}(r)}{r}$, where $\tilde{g}(r)$ denotes the Laplace transform of $g(t)$.

Substituting Eqs. (14-16) into Eq. (12) yields

$$\tilde{B}_r(s)=\int_0^\infty \frac{(s+r)e^{-(s+r)x}\tilde{S}(s+r)}{s+re^{-(s+r)x}\tilde{S}(s+r)}\,f_X(x)\,dx, \tag{17}$$

where, with a slight change of notation, $\tilde{B}_r(s)$ is the Laplace transform of the service time under a constant resetting rate r. The moments of the service time under resetting can now be readily computed from Eq. (17) by noting that

$$\mathbf{E}[B_r^n]=(-1)^n\frac{d^n}{ds^n}\tilde{B}_r(s)|_{s\to 0}. \tag{18}$$

For the first moment, we obtain

$$\mathbf{E}[B_r]=\frac{\tilde{X}(-r)-\tilde{S}(r)}{r\tilde{S}(r)}. \tag{19}$$

Note that Eq. (19) is well defined only when

$$\tilde{X}(-r)=\int_0^\infty e^{rx} f_X(x)\,dx \tag{20}$$

is finite. This condition is automatically violated for jobs' inherent size distributions with tails that are not exponentially bounded, i.e. heavy-tailed distributions [79]. In what follows we consider jobs' inherent size distributions, and resetting rates, for which $\tilde{X}(-r)$ is finite.

To better understand the effect of resetting on the mean service time, consider the introduction of an infinitesimal resetting rate δr. Expanding Eq. (19) in a Taylor series to first order in δr, we obtain (Appendix A)

$$\begin{aligned}\mathbf{E}[B_{\delta r}] \simeq\ & \mathbf{E}[X]+\mathbf{E}[S] \\ & +\delta r\Big(\mathbf{E}[X]\mathbf{E}[S]+\mathbf{E}[S]^2+\frac{1}{2}\left(\mathbf{E}[X^2]-\mathbf{E}[S^2]\right)\Big) \\ & +o(\delta r^2).\end{aligned} \tag{21}$$

The first term on the right hand side of Eq. (21) is the mean of the *original* service time, i.e., without resetting; and the second term gives the first order correction to this result. In particular, when this term is negative the introduction of a small resetting rate will reduce the mean service time. Rearranging the parenthesis in the second term, we find that this happens when

$$\Big(\mathbf{E}[X]+\mathbf{E}[S]\Big)^2<\sigma^2(S)-\sigma^2(X). \tag{22}$$

To better understand the above condition consider a scenario where the slowdown time S is taken from a generic distribution, and the jobs' inherent size X is fixed, i.e., $X=x$. In this scenario, $\sigma^2(X)=0$ and the above condition reduces to

$$x+\mathbf{E}[S]<\sigma(S). \tag{23}$$

The condition given in Eq. (23) generalizes a similar condition for the case $x=0$, which was previously derived in [1].

Note that the condition in Eq. (23) is satisfied when the standard deviation in the intrinsic component of the service time in the absence of resetting, $\sigma(S)$, is larger than the mean service time without resetting $\mathbf{E}[B]=x+\mathbf{E}[S]$. In such cases, the introduction of a small resetting rate is guaranteed to lower the mean service time. This is illustrated in Fig. 2a where we fix $x=2/3$ and plot the mean service time as a function of the resetting rate for intrinsic service times that are taken from the Gamma distribution. Parameters are taken such that $\mathbf{E}[S]=1/2$ is fixed and $\sigma^2(S)$ is varied in magnitude to illustrate the transition predicted by Eq. (23). Note that for $\sigma(S)>7/6$ the mean service time attains a minimum at an intermediate (optimal) resetting rate. Conversely, for lower variances, the mean service time is monotonically increasing with the resetting rate.

Next, we illustrate the effect of fluctuations in the intrinsic job size X. To this end, we plot in Fig. 2b the mean service time as a function of the resetting rate for intrinsic job sizes of equal mean and different variance. Specifically, we once again take the distribution of the server slowdown as Gamma with $\mathbf{E}[S]=1/2$ and $\sigma^2(S)=25$ (i.e., identical to the $\alpha=0.01$ case from Fig. 2a). Fixing this distribution for the server

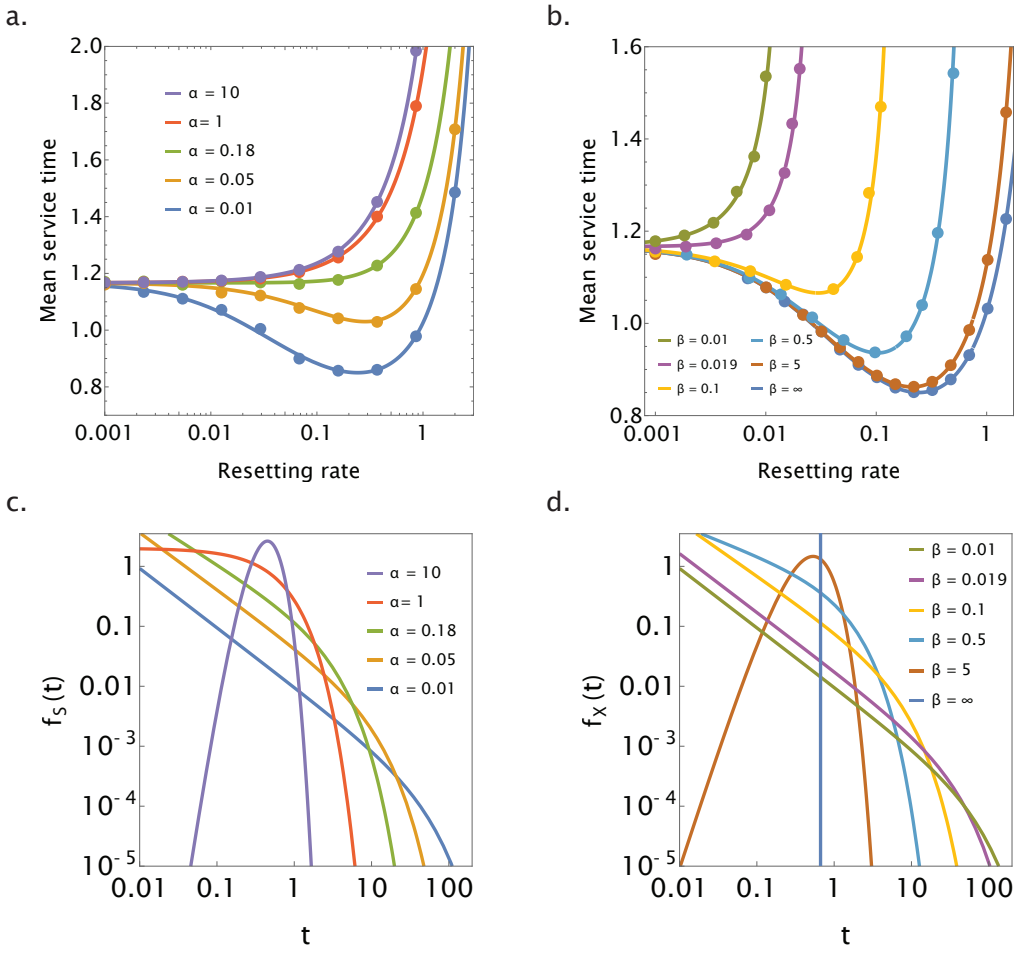


FIG. 2. The effect of constant-rate resetting in the additive model. Panel (a): The mean service time as a function of the resetting rate. Plots are made using Eq. (19) for a fixed inherent job size $X = x = 2/3$ and a server slowdown S that is taken from the Gamma distribution with density $f_S(t) = \frac{1}{\Gamma(\alpha)\theta^\alpha} t^{\alpha-1} e^{-t/\theta}$, $t > 0$. Here, we set $\mathbf{E}[S] = \theta\alpha = 1/2$ and vary α to control the variance $\sigma^2(S) = \theta^2\alpha = \alpha^{-1}/4$, as illustrated in panel (c). Observe that the introduction of a small resetting rate lowers the mean service time only when the condition in Eq. (23) is met, i.e., for $\alpha < 9/49 \simeq 0.184$. In such cases, the service time is minimized at some intermediate resetting rate. Analytical results (in solid lines) are corroborated with numerical simulations (full circles). Panel (b): The mean service time as a function of the resetting rate when both S and X come from the Gamma distribution. Here, as in panel (a), we take $S \sim \Gamma(\alpha, \theta = \frac{1}{2\alpha})$ and further set $\alpha = 0.01$. In addition, we take $X \sim \Gamma(\beta, \frac{2}{3\beta})$ such that $\mathbf{E}[X] = 2/3$ is fixed, and then tune β to control the variance $\sigma^2(X) = 4\beta^{-1}/9$ as illustrated in panel (d). Observe that the introduction of a small resetting rate lowers the mean service time only when the condition in Eq. (23) is met, i.e., for $\beta < 16/851 \simeq 0.019$.

slowdown, we consider an inherent job size X that comes from a Gamma distribution with mean $\mathbf{E}[X] = 2/3$ and a variance whose magnitude we gradually increase from an initial value of zero (the deterministic case considered in Fig. 2a). As predicted by Eq. (22), the introduction of resetting reduces the mean service time as long as the variance in the inherent job size is smaller that a certain threshold, which is here given by $\sigma^2(X) = 23\frac{23}{36}$. In such cases, the service time is minimized at some intermediate (optimal) resetting rate.

Summarizing, we see that service resetting is beneficial when the variance in the intrinsic component of the service time is large compared to the mean-squared service time. In addition, the variance in the inherent job size must be small by the same measure. As a result, the parameter phase space is divided into two regions: one where the introduction of resetting reduces the mean service time and one where the converse happens. This phase space is illustrated in Fig. 3 for the case of Gamma distributed $S\&X$ service times.

4. MULTIPLICATIVE MODEL

Following the above discussion on the additive $S\&X$ model with service resetting, we go on to discuss a multiplicative version. In what follows, we assume that the overall service time in the absence of resetting, V, is given by a product of the two independent sources S and X that were defined above. Namely, $V = S \times X$. Similar to the treatment of the additive

model, we let V_R denote the total service time under resetting. Given an inherent job size of x, the conditional service time $V_R(x) \equiv \{V_R|X = x\}$, obeys the following renewal equation

$$V_R(x) = \begin{cases} xS & if \quad xS < R \\ R + V'_R(x) & if \quad xS \geq R, \end{cases} \tag{24}$$

where, once again, R is a random resetting time drawn from a distribution with density $f_R(\cdot)$, and $V'_R(x)$ is an independent and identically distributed copy of $V_R(x)$. Under the multiplicative model (indicated below by the $\times$ sign), it is useful to define the following conditional random variables

$$S_\times(x) = \{S|xS < R\}, \tag{25}$$
$$R_\times(x) = \{R|xS \geq R\}, \tag{26}$$

and their corresponding Laplace transforms $\tilde{S}_\times(x;s)$ and $\tilde{R}_\times(x;s)$

$$\tilde{S}_\times(x;s) = \int_0^\infty e^{-st} f_{S_\times}(t)\, dt, \tag{27}$$
$$\tilde{R}_\times(x;s) = \int_0^\infty e^{-st} f_{R_\times}(t)\, dt. \tag{28}$$

The PDFs of these conditional random variables are given by

$$f_{S_\times}(t) = \frac{f_S(t)}{\Pr(xS < R)} \int_{xt}^\infty f_R(\tau)\, d\tau, \tag{29}$$
$$f_{R_\times}(t) = \frac{f_R(t)}{\Pr(xS \geq R)} \int_{\frac{t}{x}}^\infty f_S(\tau)\, d\tau. \tag{30}$$

Taking expectation of both sides of Eq. (24) and rearranging terms, one obtains the mean conditional service time under resetting

$$\mathbf{E}[V_R(x)] = x\mathbf{E}[S_\times(x)] + \frac{1 - \Pr(xS < R)}{\Pr(xS < R)} \mathbf{E}[R_\times(x)], \tag{31}$$

where $\mathbf{E}[S_\times(x)]$ and $\mathbf{E}[R_\times(x)]$ are the means of the random variables $S_\times(x)$ and $R_\times(x)$, respectively.

Integrating Eq. (31) with respect to the probability density of the jobs' inherent size, $f_X(x)$, one obtains the mean service time under resetting

$$\mathbf{E}[V_R] = \int_0^\infty x\mathbf{E}[S_\times(x)]\, f_X(x)\, dx + \int_0^\infty \frac{\Pr(xS \geq R)}{\Pr(xS < R)} \mathbf{E}[R_\times(x)]\, f_X(x)\, dx. \tag{32}$$

Note the similarities between Eq. (32) for the multiplicative case and Eq. (10) for the additive case. The first term on the right-hand side of Eq. (32) accounts for the mean service time of the last service trial, and the second term accounts for the time wasted in failed service attempts.

Finally, we turn to derive the Laplace transform of V_R. Taking the Laplace transform of both sides of Eq. (24) and rearranging terms we obtain

$$\tilde{V}_R(x;s) = \frac{\tilde{S}_\times(x;sx)\Pr(xS < R)}{1 - \tilde{R}_\times(x;s)\Pr(xS \geq R)}, \tag{33}$$

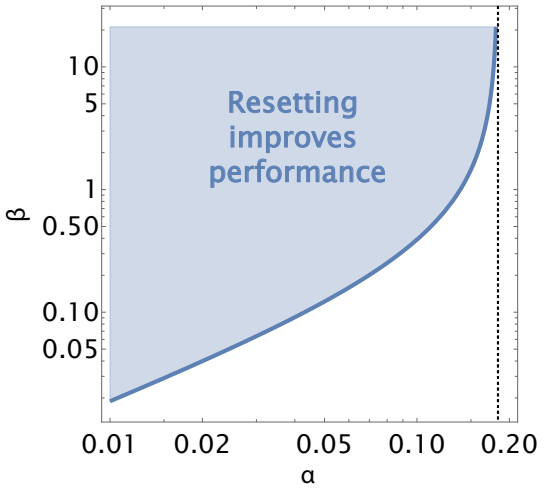


FIG. 3. Visualization of the condition in Eq. (22) for the case of Gamma distributed $S\&X$ service times in the additive model. Here, as in Fig. 2, we take $S \sim \Gamma(\alpha, \frac{1}{2\alpha})$ and $X \sim \Gamma(\beta, \frac{2}{3\beta})$ such that $\mathbf{E}[S] = 1/2$ and $\mathbf{E}[X] = 2/3$ are fixed. Tuning α and β we control the variance of these service time components as $\sigma^2(S) = \alpha^{-1}/4$ and $\sigma^2(X) = 4\beta^{-1}/9$. The condition in Eq. (22) is satisfied when $\sigma^2(S)$ is large enough and when $\sigma^2(X)$ is small enough. This part of the phase space, where the introduction of service resetting reduces the mean service time, is shaded in blue.

Integrating Eq. (33) with respect to the probability density of the jobs' inherent size, $f_X(x)$, one obtains the Laplace transform of the service time under generic resetting in the multiplicative model

$$\tilde{V}_R(s) = \int_0^\infty \frac{\tilde{S}_\times(x;sx)\Pr(xS < R)}{1 - \tilde{R}_\times(x;s)\Pr(xS \geq R)} f_X(x)\, dx. \tag{34}$$

4a. Constant-rate Poisson resetting

Consider the above-detailed service process and further assume that resetting times are taken from an exponential distribution with rate parameter r. Namely, for $t > 0$, the probability density of the resetting time distribution is given by Eq. (13). In this case, we have

$$\Pr(xS < R) = \int_0^\infty f_R(t)\Pr(xS < t)\, dt = \int_0^\infty re^{-rt} \left(\int_0^{t/x} f_S(\tau)\, d\tau \right) dt = \tilde{S}(rx), \tag{35}$$

where $\tilde{S}(rx)$ is the Laplace transform of S evaluated at rx. The last step in Eq. (35) follows by changing the order of integration.

Substituting the above result into Eqs. (29) and (30), allows us to obtain neat expressions for the Laplace transforms of the conditional random variables $S_\times(x)$ and $R_\times(x)$. These are

given by

$$\begin{aligned}\tilde{S}_{\times}(x;s) &= \int_0^{\infty} e^{-st} f_{S_{\times}}(t)\,dt \\ &= \frac{\int_0^{\infty} e^{-st} f_S(t)\left(\int_{xt}^{\infty} re^{-r\tau}\,d\tau\right)dt}{\Pr(xS<R)} \\ &= \frac{\int_0^{\infty} e^{-t(s+rx)} f_S(t)\,dt}{\tilde{S}(rx)} \\ &= \frac{\tilde{S}(s+rx)}{\tilde{S}(rx)},\end{aligned} \tag{36}$$

$$\begin{aligned}\tilde{R}_{\times}(x;s) &= \int_0^{\infty} e^{-st} f_{R_{\times}}(t)\,dt \\ &= \frac{\int_0^{\infty} e^{-st}\, re^{-rt}\left(\int_{\frac{t}{x}}^{\infty} f_S(\tau)\,d\tau\right)dt}{\Pr(xS\geq R)} \\ &= \frac{r\int_0^{\infty} e^{-t(s+r)}\left(\int_{\frac{t}{x}}^{\infty} f_S(\tau)\,d\tau\right)dt}{1-\tilde{S}(rx)} \\ &= \frac{r\int_0^{\infty} e^{-t(s+r)}\left(1-\int_0^{\frac{t}{x}} f_S(\tau)\,d\tau\right)dt}{1-\tilde{S}(rx)} \\ &= \frac{r}{r+s}\frac{1-\tilde{S}((s+r)x)}{1-\tilde{S}(rx)},\end{aligned} \tag{37}$$

where in the last step in Eq. (37) we once again used the known formula for the Laplace transform of a time-domain integration: $\int_0^{\infty} dt\,\left(\int_0^t d\tau\, g(\tau)\right)e^{-rt} = \frac{\tilde{g}(r)}{r}$, where $\tilde{g}(r)$ denotes the Laplace transform of $g(t)$.

Substituting Eqs. (35-37) into Eq. (34) yields

$$\tilde{V}_r(s) = \int_0^{\infty} \frac{(s+r)\tilde{S}((s+r)x)}{s+r\tilde{S}((s+r)x)} f_X(x)\,dx. \tag{38}$$

The moments of the service time under resetting can be computed from Eq. (38), by utilizing Eq. (18). For the first moment, we obtain

$$\mathbf{E}[V_r] = \int_0^{\infty} \frac{1-\tilde{S}(rx)}{r\tilde{S}(rx)} f_X(x)\,dx. \tag{39}$$

To better understand the effect of resetting on the mean service time, we once again consider the introduction of an infinitesimal resetting rate δr. Expanding Eq. (39) in a Taylor series to first order in δr, we obtain (Appendix B)

$$\begin{aligned}\mathbf{E}[V_{\delta r}] &\simeq \mathbf{E}[X]\mathbf{E}[S] \\ &\quad + \delta r\mathbf{E}[X^2]\left(\mathbf{E}[S]^2 - \frac{1}{2}\mathbf{E}[S^2]\right) + O(\delta r^2).\end{aligned} \tag{40}$$

The first term on the right hand side of Eq. (40) is the mean of the *original* service time, i.e., without resetting. The second term gives the first order correction. Rearranging the second term, we find that the introduction of a small ~~constant~~ resetting rate reduces the mean service time when the contribution of this term is negative. That is,

$$1 < CV_S^2, \tag{41}$$

where CV_S is the coefficient of variation of the random variable S

$$CV_S = \sigma(S)/\mathbf{E}[S]. \tag{42}$$

The condition in Eq. (41) is satisfied when the standard deviation in the intrinsic component of the service time in the absence of resetting, $\sigma(S)$, is larger than the mean of the intrinsic component, $\mathbf{E}[S]$. For example, if S comes from a distribution with a decreasing failure rate the condition in Eq. (41) will always hold [80]. In such cases, the introduction of a small resetting rate is guaranteed to lower the mean service time. Note, that the resetting condition in the multiplicative case depends solely on the service slowdown S and that it is independent of the jobs' inherent size X. This stands in contrast to to the condition in the additive case which depends on both S and X as can be seen in Eq. (22).

The condition in Eq. (41) is illustrated in Fig. 4a where we fix $x = 2/3$ and plot the mean service time as a function of the resetting rate for intrinsic service times that are taken from the Inverse Gaussian distribution. Parameters are taken such that $\mathbf{E}[S] = 3/2$ is fixed and $\sigma^2(S)$ is varied in magnitude to illustrate the transition predicted by Eq. (41). Note that for $\sigma(S) > 3/2$ the mean service time attains a minimum at an intermediate (optimal) resetting rate. Conversely, for lower variances, the mean service time is monotonically increasing with the resetting rate.

Next, we illustrate that fluctuations in the intrinsic job size X have no effect on whether resetting is beneficial or not in the multiplicative case. To this end, we plot in Fig. 4b the mean service time as a function of the resetting rate for intrinsic job sizes of equal mean and different variance. Specifically, we once again take the distribution of the server slowdown as Inverse Gaussian with $\mathbf{E}[S] = 3/2$ and $\sigma^2(S) = 9/2$ Is this correct? Ofek, please double check: (i.e., identical to the $\alpha_S = 0.75$ case from Fig. 4a). Fixing this distribution for the server slowdown, we consider an intrinsic job size X that comes from an Inverse Gaussian distribution with mean $\mathbf{E}[X] = 2/3$ and a variance whose magnitude we gradually increase from an initial value of zero, which corresponds to the deterministic X that was considered in Fig. 4a. As predicted by Eq. (41), since $CV_S^2 = 2$, the introduction of resetting reduces the mean service time regardless of the ~~magnitude of~~ the variance in the extrinsic job size. ~~Thus, the service time is always minimized at some intermediate (optimal) resetting rate when the condition for S is met.~~

Summarizing, we see that service resetting is beneficial when the standard deviation in the intrinsic component of the service time, $\sigma(S)$, is larger than the mean of the intrinsic component of the service time, $\mathbf{E}[S]$. The condition for resetting being beneficial is agnostic to~~independent of~~ the extrinsic component of the service time distribution. This sentence reads weird: As a result, the parameter phase space is now divided into two regions, each corresponding to a different parameter of the intrinsic component distribution. This phase space is illustrated in Fig. 5 for the case of Inverse Gaussian distributed S&X service times.

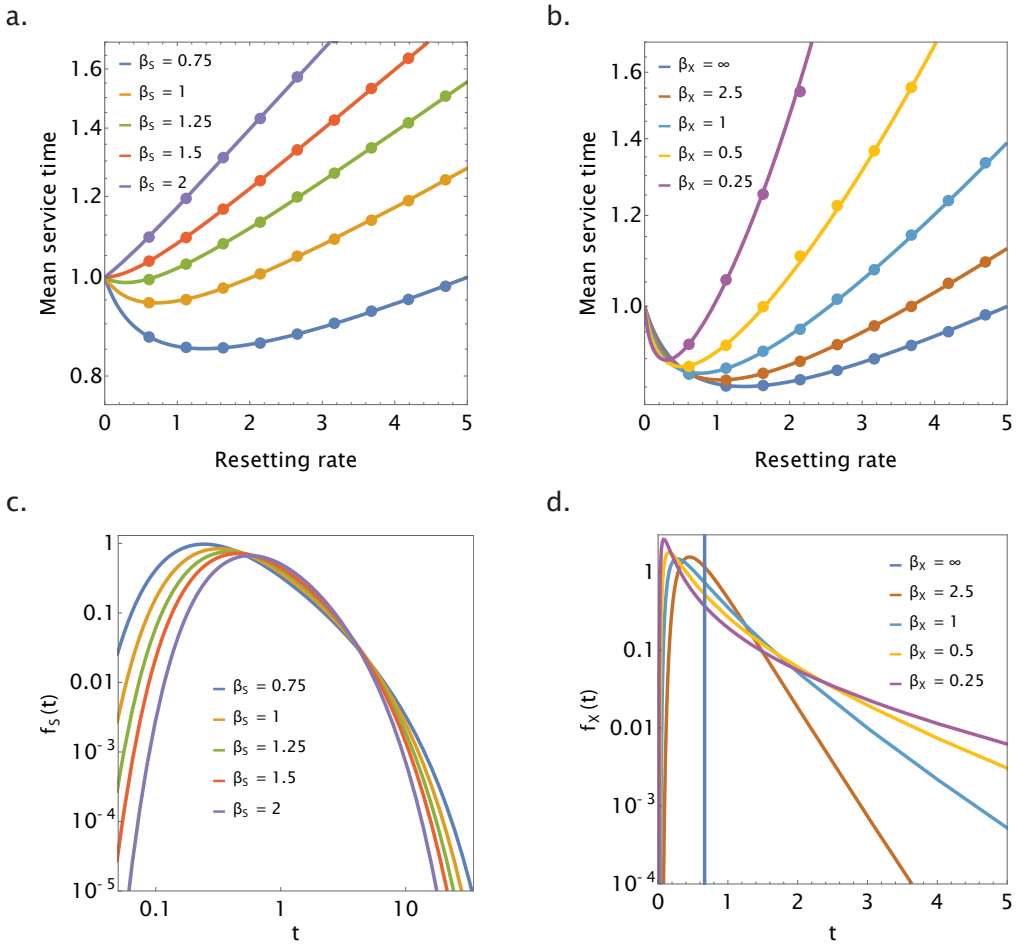


FIG. 4. The effect of constant-rate resetting in the multiplicative ~~additive~~ model. Panel (a): The mean service time as a function of the resetting rate. Plots are made using Eq. (39) for a fixed inherent job size $X = x = 2/3$ and a server slowdown S that is taken from the Inverse Gaussian distribution with density $f_S(t) = \sqrt{-\beta_S/2\pi t^3}\, e^{-\beta_S(t-\alpha_S)^2/2\alpha_S^2 t}$, $t > 0$. Here, we set $\mathbf{E}[S] = \alpha_S = 3/2$ and vary β_S to control the variance $\sigma^2(S) = \alpha_S^3/\beta_S = 27\beta_S^{-1}/8$, as illustrated in panel (c). Observe that the introduction of a small resetting rate lowers the mean service time only when the condition in Eq. (41) is met, i.e., for $\beta_S < 3/2$. In such cases, the service time is minimized at some intermediate resetting rate. Analytical results (in solid lines) are corroborated with numerical simulations (full circles). Panel (b): The mean service time as a function of the resetting rate when both S and X come from the Inverse Gaussian distribution. Here, as in panel (a), we take $S \sim IG(\alpha_S = 3/2, \beta_S)$ and further set $\beta_S = 3/4$ ~~$\beta_S = 0.75$~~. In addition, we take $X \sim IG(\alpha_X, \beta_X)$ such that $\mathbf{E}[X] = \alpha_X = 2/3$ is fixed, and then tune β_S to control the variance $\sigma^2(X) = 8\beta_X^{-1}/27$ as illustrated in panel (d). Observe that the introduction of a resetting lowers the mean service time regardless of the value β_X, as suggested by the condition in Eq. (41).

We emphasize that the results derived so far in Secs. 3 and 4 are agnostic to the details of the arrival process and make no assumptions about its stochastic behavior.

Consequently, in the following section we consider the celebrated M/G/1 queue for which we obtain explicit expressions for its probability generating function of the queue length and for the Laplace transform of the waiting times.

5. QUEUE LENGTH DISTRIBUTION UNDER RESETTING

So far, we were concerned only with the mean service time under resetting. We now turn our attention to deriving the distribution of the queue length. To this end, we consider an M/G/1 queue with Poisson arrivals with rate λ and overall service time U. In an M/G/1 queue, the probability mass function of the number of jobs in the system, L, is denoted as $P_L(n) = \Pr(L = n)$. The probability generating function (PGF)

$$G_L(z) = \sum_{n=0}^{\infty} P_L(n) z^n \,, \tag{43}$$

of the queue length, is given by the Pollaczek–Khinchine formula [81]

$$G_L(z) = \frac{(1-\rho)(1-z)\tilde{U}(\lambda(1-z))}{\tilde{U}(\lambda(1-z)) - z} \,, \tag{44}$$

where $\tilde{U}(\cdot)$ is the Laplace transform of the service time, and $\rho = \lambda \mathbf{E}[U]$ is the utilization of the queue. The probability to

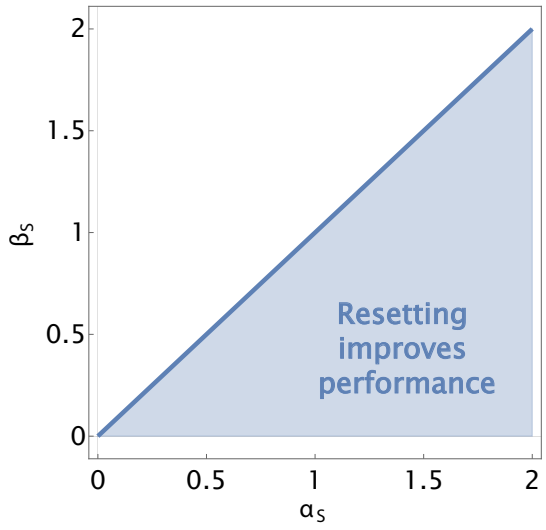


FIG. 5. Visualization of the condition in Eq. (41) for the case of Inverse Gaussian distributed $S\&X$ service times in the multiplicative model. Here, as in Fig. 4, we take $S \sim IG(\alpha_S, \beta_S)$, $X \sim IG(\alpha_X, \beta_X)$. Tuning α_S and β_S we control the mean $\mathbf{E}[S] = \alpha_S$, and the variance $\sigma^2(S) = \alpha_S^3/\beta_S$, of the intrinsic component of the service time. The condition in Eq. (41) is satisfied when $CV_S^2 = \alpha_S/\beta_S$ is larger than 1. This region ~~part~~ of the phase space, where the introduction of service resetting reduces the mean service time, is shaded in blue.

find n jobs in the system can then be computed from Eq. (44) via

$$P_L(n) = \Pr(L = n) = \frac{G_L^{(n)}(0)}{n!} , \tag{45}$$

where $G_L^{(n)}(0)$ stands for the n-th derivative of $G_L(z)$ evaluated at $z = 0$. Similarly, one can also compute the Laplace-Stieltjes transform, $\tilde{W}(s)$, of the total time a job spends in the system. This is given by the corresponding Pollaczek-Khinchin transform equation which reads

$$\tilde{W}(s) = \frac{\tilde{U}(s)(1-\rho)s}{s - \lambda(1-\tilde{U}(s))} . \tag{46}$$

The mean queue length is given by

$$\mathbf{E}[L] = \frac{\rho}{1-\rho} + \frac{\rho^2}{2(1-\rho)}\left(CV_U^2 - 1\right) , \tag{47}$$

where CV_U^2 is the squared coefficient of variation of the overall service time U, $CV_U^2 = \sigma^2(U)/\mathbf{E}^2[U]$.

To obtain corresponding formulas in the case of service resetting, we substitute the Laplace transforms of the service time with resetting into the Pollaczek–Khinchine formula. Namely, substituting $\tilde{B}_r(s)$ from Eq. (17) and $\tilde{V}_r(s)$ from Eq. (38) into the Pollaczek–Khinchine formula yields the PGF of the queue length under Poissonian resetting, in the additive and multiplicative models respectively.

In Fig. 6 we follow the aforementioned steps to derive the queue length distribution under resetting for both the additive and multiplicative models. In Fig. 6a, we we consider the additive movel and compare the queue length distribution under optimal Poissonian resetting to the queue distribution without resetting. Here, we take the distribution of the server slowdown to be Gamma with $\mathbf{E}[S] = 1/2$ and $\sigma^2(S) = 25$, and fix the inherent job size to $x = 2/3$ (i.e., identical to the $\alpha = 0.01$ case from Fig. 2a). The optimal resetting rate in this scenario is $r^* \simeq 0.2424$. The effect of resetting in this scenario is evident — under resetting, there is a higher probability of having shorter queues compared to the no resetting case. Furthermore, the tail of the distribution for a queue with resetting is significantly lighter. The effect is even more pronounced ~~visible~~ when comparing the mean queue lengths: the mean queue length under optimal resetting is $\mathbf{E}[L_{r^*}] \simeq 0.739$, which is more than ten fold ~~11 folds~~ smaller than its value $\mathbf{E}[L] \simeq 8.49$ without resetting.

In Fig. 6, we consider the multiplicative model and compare the queue length distribution under optimal Poissonian resetting to the queue distribution without resetting. Here, we take the distribution of the server slowdown to be Inverse Gaussian with $\mathbf{E}[S] = 3/2$ and $\sigma^2(S) = 9/2$, and fix the inherent job size as $x = 2/3$ (i.e., identical to the $\alpha_S = 0.75$ case from Fig. 4a). The optimal resetting rate in this scenario is $r^* \simeq 1.372$. The effect of resetting under the multiplicative model is qualitatively similar to the additive model — under resetting, there is a higher probability of having shorter queues compared to the no resetting case. Once again, the tail of the distribution for a queue with resetting is lighter, and we find that optimal resetting reduces the mean queue length from $\mathbf{E}[L] \simeq 4.12$ to $\mathbf{E}[L_{r^*}] \simeq 1.76$, i.e., more than two-fold.

Summarizing, in this section we see the effects of optimal Poissonian resetting on the queue length distribution of an M/G/1 queue in both the additive and multiplicative models. In both cases, we demonstrated that resetting has a significant impact on the queue length distribution: generating lighter tails, smaller means, and increasing the probability of seeing shorter queues.

6. CONCLUSIONS

In this paper, we further developed the theory of queueing systems with service resetting, extending the framework presented in [1] to $S\&X$ queues [2]. We considered scenarios where a job's service time is either a sum, or a product, of two components: one intrinsic to the server, the other extrinsic. We determined the total service time of a job in cases where times between service resetting events form a general renewal process. We then specialized to a constant rate resetting policy (aka Poissonian resetting) for which we derived explicit conditions under which resetting reduces the mean service time and improves queue performance. In the additive case, the condition is given by Eq. (22) which involves only the means and variances of the intrinsic and extrinsic components of the service time. In the multiplicative case, the condition is given by Eq. (41), which involves only the mean and variance of the intrinsic component of the service time. In both cases, the

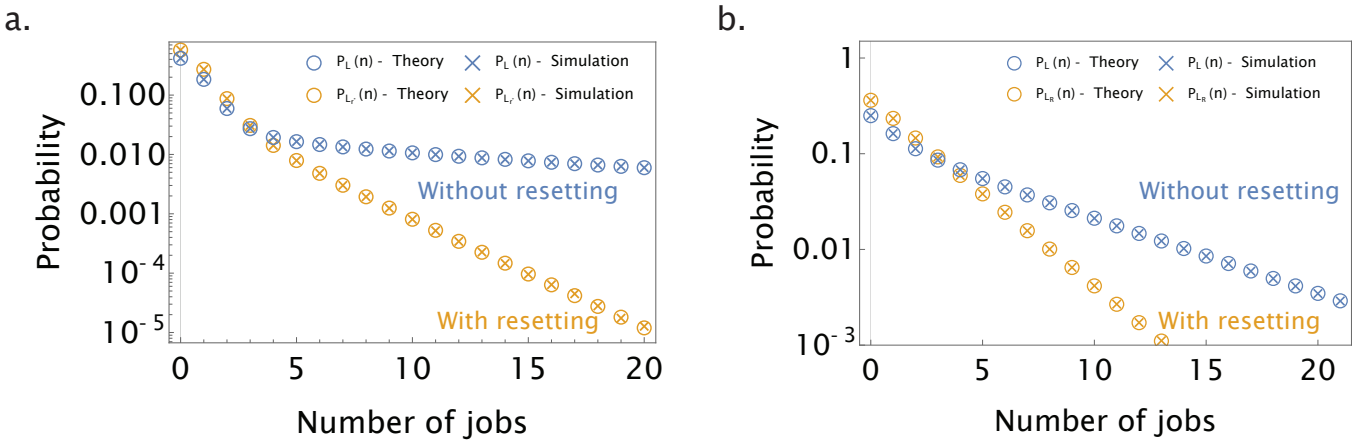


FIG. 6. The probability mass function of the number of jobs in an *S&X* queueing system (queue+server) without service resetting (blue) and with optimal Poissonian service resetting (orange). The probabilities $P_N(n)$ and $P_{L_{r^*}}(n)$ to find n jobs in a system without and with service resetting were computed by use of Eq. (45). Here, r^* was taken as the optimal resetting rate, i.e., that which minimizes the overall mean service time. The analytical results (circles) were further corroborated with numerical simulations (X marks). Panel (a): The additive model. Here, the arrival rate was set to $\lambda = 0.5$ and the service time Shoudn't this be *B*???: U was taken as the sum of: (i) S, a Gamma random variable with parameters $\alpha = 0.01$ and $\theta = 50$, and (ii) X, a deterministic random variable $\Pr(X = 2/3) = 1$. The optimal resetting rate is then found to be $r^* \simeq 0.2424$ (see Fig. 2a). Panel (b): The multiplicative model. Here, the jobs arrival rate was set to $\lambda = 0.75$, and the service time Shoudn't this be *V*???:U was taken to be a product of: (i) S, an Inverse Gaussian random variable with parameters $\alpha_S = 3/2$ and $\beta_S = 3/4$, and (ii) X, a deterministic random variable $\Pr(X = 2/3) = 1$. The optimal resetting rate is then found to be $r^* \simeq 1.372$ (see Fig. 4a). In both the additive and multiplicative models, it is evident that the introduction of optimal service resetting gives rise to lighter-tailed queue length distributions and consequently to shorter queues.

introduction of resetting is beneficial when the variance of the intrinsic component of the service time is large enough.

In our analysis, we have made the common assumption that the arrival process to the queue is independent of the service process. Thus, the results obtained in Chapters 3 & 4 hold regardless of the arrival process. For illustration we considered the canonical M/G/1 queue for which we have given explicit results for the mean and distribution of the queue length in *S&X* queuing systems with service resetting. Yet, we emphasize that results and conclusions coming from our analysis carry over to other systems such as the G/G/1 *S&X* queues. Indeed, under the conditions specified in Eq. (22) and Eq. (41) it is guaranteed that service resetting will reduce the mean service time irrespective of the arrival process.

In summary, this paper established the pivotal role of service resetting as a novel strategy for combating the harmful effects of service time fluctuations in *S&X* queues. Our comprehensive analysis, spanning both additive and multiplicative models, illuminates the consistent effectiveness of service resetting in significantly reducing mean service times and, consequently, shrinking queue lengths. The developed framework not only offers theoretical insights but also equips practitioners with a practical toolkit. By providing explicit conditions for the implementation of resetting, our paper enables precise decision-making, ensuring that service resetting becomes an invaluable asset in real-world queue management scenarios. Notably, our results transcend the confines of specific arrival processes, underscoring the universal applicability of service resetting. Through this work, we not only advance the theoretical foundations of queueing theory but also provide robust means for enhancing the efficiency and responsiveness of diverse queueing systems across various domains.

APPENDIX

A. Derivation of Eq. (21)

Equation (21) can be derived by straightforwardly expanding Eq. (19) in a Taylor series to first order in δr, where δr is infinitesimally small resetting rate. However, one can also derive this equation by expanding both numerator and denominator of Eq. (19) in a Taylor series, which is the method we present in the following appendix.

Utilizing the property of the Laplace transform given in Eq. (18), one can expand the numerator of the right hand side of Eq. (19) in a Taylor series to second order in δr

$$\begin{aligned}\tilde{X}(-r)-\tilde{S}(r) \simeq & \left(1+\delta r\mathbf{E}[X]+\frac{1}{2}\delta r^2\mathbf{E}[X^2]\right) \\ & -\left(1-\delta r\mathbf{E}[S]+\frac{1}{2}\delta r^2\mathbf{E}[S^2]\right) \\ & +O(\delta r^3)= \\ & \delta r\Big(\mathbf{E}[X]+\mathbf{E}[S]\Big)+ \\ & \frac{1}{2}\delta r^2\Big(\mathbf{E}[X^2]-\mathbf{E}[S^2]\Big)+O(\delta r^3). \qquad (48)\end{aligned}$$

Following the above steps, we obtain the Taylor series of the denominator of the RHS of Eq. (19) to second order in δr

$$r\tilde{S}(r) \simeq \ \delta r - \delta r^2 \mathbf{E}[S] + O(\delta r^3). \tag{49}$$

Substituting Eqs. (48) and (49) into Eq. (19), and dividing both numerator and denominator by δr yields the following Taylor expansion

$$\mathbf{E}[B_{\delta r}] \simeq \frac{\mathbf{E}[X] + \mathbf{E}[S] + \frac{1}{2}\delta r\Big(\mathbf{E}[X^2] - \mathbf{E}[S^2]\Big) + O(\delta r^2)}{1 - \delta r \mathbf{E}[S] + O(\delta r^2)}. \tag{50}$$

Multiplying by the conjugate, Eq. (50) reduces to

$$\mathbf{E}[B_{\delta r}] \simeq \Big(\mathbf{E}[X] + \mathbf{E}[S] + \frac{1}{2}\delta r\Big(\mathbf{E}[X^2] - \mathbf{E}[S^2]\Big) + O(\delta r^2)\Big) \times \Big(1 + \delta r \mathbf{E}[S] + O(\delta r^2)\Big). \tag{51}$$

Expanding the product given in the above equation and retaining terms up to first order in δr yields

$$\mathbf{E}[B_{\delta r}] \simeq \mathbf{E}[X] + \mathbf{E}[S] + \delta r\Big(\mathbf{E}[X]\mathbf{E}[S] + \mathbf{E}[S]^2 + \frac{1}{2}\left(\mathbf{E}[X^2] - \mathbf{E}[S^2]\right)\Big) + O(\delta r^2), \tag{52}$$

which is Eq. (21) in the main text.

B. Derivation of Eq. (40)

Equation (40) can be derived by straightforwardly expanding Eq. (19) in a Taylor series to first order in δr, where δr is infinitesimally small resetting rate. However, one can also derive this equation by expanding both numerator and denominator of Eq. (39) in a Taylor series, which is the method we present in the following appendix.

Utilizing the property of the Laplace transform given in Eq. (18), one can expand $\tilde{S}(rx)$ in a Taylor series to first order in δr

$$\tilde{S}(rx) \simeq \ 1 - \delta r x \mathbf{E}[S] + \frac{1}{2}(\delta r x)^2 \mathbf{E}[S^2] + O(\delta r^3). \tag{53}$$

Following the above steps, we obtain the Taylor series of $r\tilde{S}(rx)$ to second order in δr

$$r\tilde{S}(rx) \simeq \ \delta r - \delta r^2 x \mathbf{E}[S] + O(\delta r^3). \tag{54}$$

Substituting Eqs. (53) and (54) into Eq. (39), and dividing both numerator and denominator by δr yields the following Taylor expansion

$$\mathbf{E}[V_{\delta r}] \simeq \ \int_0^\infty \frac{x\mathbf{E}[S] - \frac{1}{2}\delta r x^2 \mathbf{E}[S^2] + O(\delta r^2)}{1 - \delta r x \mathbf{E}[S] + O(\delta r^2)} f_X(x)\, dx. \tag{55}$$

Multiplying by the conjugate, Eq. (55) reduces to

$$\mathbf{E}[V_{\delta r}] \simeq \int_0^\infty \Big(x\mathbf{E}[S] - \frac{1}{2}\delta r x^2 \mathbf{E}[S^2] + O(\delta r^2)\Big) \times \Big(1 + \delta r x \mathbf{E}[S] + O(\delta r^2)\Big) f_X(x)\, dx. \tag{56}$$

Expanding the product given in the above equation and retaining terms up to first order in δr yields

$$\mathbf{E}[V_{\delta r}] \simeq \int_0^\infty \Big(x\mathbf{E}[S] + \delta r x^2 \left(\mathbf{E}[S]^2 - \frac{1}{2}\mathbf{E}[S^2]\right) + O(\delta r^2)\Big) f_X(x)\, dx. \tag{57}$$

Calculating the integral with respect to x yields Eq. (40).

* ofekzvil@mail.tau.ac.il
† uriy@tauex.tau.ac.il
‡ shlomire@tauex.tau.ac.il

[1] Bonomo, O.L., Pal, A. and Reuveni, S., 2022. Mitigating long queues and waiting times with service resetting. PNAS Nexus, 1(3), p.pgac070.

[2] Gardner, K., Harchol-Balter, M., Scheller-Wolf, A. and Van Houdt, B., 2017. A better model for job redundancy: Decoupling server slowdown and job size. IEEE/ACM transactions on networking, 25(6), pp.3353-3367.

[3] Gans, N., Koole, G. and Mandelbaum, A., 2003. Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management, 5(2), pp.79-141.

[4] Koole, G. and Mandelbaum, A., 2002. Queueing models of call centers: An introduction. Annals of Operations Research, 113(1), pp.41-59.

[5] Bachmat, E., Berend, D., Sapir, L., Skiena, S. and Stolyarov, N., 2009. Analysis of airplane boarding times. Operations Research, 57(2), pp.499-513.

[6] Erland, S., Kaupužs, J., Frette, V., Pugatch, R. and Bachmat, E., 2019. Lorentzian-geometry-based analysis of airplane boarding policies highlights "slow passengers first" as better. Physical Review E, 100(6), p.062313.

[7] Cooper, R.B., 1981, January. Queueing theory. In Proceedings of the ACM'81 conference (pp. 119-122).

[8] Daigle, J.N., 2005. Queueing theory with applications to packet telecommunication. Springer Science & Business Media.

[9] Harchol-Balter, M., 2013. Performance modeling and design of computer systems: queueing theory in action. Cambridge University Press.

[10] Askin, R.G., and Standridge, C.R. 1993. Modeling and analysis of manufacturing systems. John Wiley & Sons Incorporated.

[11] Mather, W.H., Cookson, N.A., Hasty, J., Tsimring, L.S. and Williams, R.J., 2010. Correlation resonance generated by coupled enzymatic processing. Biophysical journal, 99(10), pp.3172-3181.

[12] Cookson, N.A., Mather, W.H., Danino, T., Mondragón-Palomino, O., Williams, R.J., Tsimring, L.S. and Hasty, J., 2011. Queueing up for enzymatic processing: correlated signaling through coupled degradation. Molecular systems biology, 7(1), p.561.

[13] Evstigneev, V.P., Holyavka, M.G., Khrapatiy, S.V. and Evstigneev, M.P., 2014. Theoretical description of metabolism using queueing theory. Bulletin of mathematical biology, 76(9), pp.2238-2248.

[14] Kloska, S., Pałczyński, K., Marciniak, T., Talaśka, T., Nitz, M., Wysocki, B.J., Davis, P. and Wysocki, T.A., 2021. Queueing theory model of Krebs cycle. Bioinformatics.

[15] Arazi, A., Ben-Jacob, E. and Yechiali, U., 2004. Bridging genetic networks and queueing theory. Physica A: Statistical Mechanics and its Applications, 332, pp.585-616.

[16] Gelenbe, E., 2007. Steady-state solution of probabilistic gene regulatory networks. Physical Review E, 76(3), p.031903.

[17] Jia, T. and Kulkarni, R.V., 2011. Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. Physical review letters, 106(5), p.058102.

[18] Kumar, N., Singh, A. and Kulkarni, R.V., 2015. Transcriptional bursting in gene expression: analytical results for general stochastic models. PLoS computational biology, 11(10), p.e1004292.

[19] Jun, S., Si, F., Pugatch, R. and Scott, M., 2018. Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. Reports on Progress in Physics, 81(5), p.056601.

[20] Helbing, D., 2001. Traffic and related self-driven many-particle systems. Reviews of modern physics, 73(4), p.1067.

[21] Evans, M.R. and Hanney, T., 2005. Nonequilibrium statistical mechanics of the zero-range process and related models. Journal of Physics A: Mathematical and General, 38(19), p.R195.

[22] Romano, M.C., Thiel, M., Stansfield, I. and Grebogi, C., 2009. Queueing phase transition: theory of translation. Physical review letters, 102(19), p.198104.

[23] Reuveni, S., Eliazar, I. and Yechiali, U., 2012. Asymmetric inclusion process as a showcase of complexity. Physical review letters, 109(2), p.020603.

[24] English, B.P., Min, W., Van Oijen, A.M., Lee, K.T., Luo, G., Sun, H., Cherayil, B.J., Kou, S.C. and Xie, X.S., 2006. Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited. Nature chemical biology, 2(2), pp.87-94.

[25] Moffitt, J.R. and Bustamante, C., 2014. Extracting signal from noise: kinetic mechanisms from a Michaelis–Menten-like expression for enzymatic fluctuations. The FEBS journal, 281(2), pp.498-517.

[26] Flomenbom, O., Velonia, K., Loos, D., Masuo, S., Cotlet, M., Engelborghs, Y., Hofkens, J., Rowan, A.E., Nolte, R.J., Van der Auweraer, M. and de Schryver, F.C., 2005. Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules. Proceedings of the National Academy of Sciences, 102(7), pp.2368-2372.

[27] Takagi, H., 1991. Vacation and priority systems (Vol. 1). North Holland.

[28] Conway, R.W., Maxwell, W.L. and Miller, L.W., 2003. Theory of scheduling. Courier Corporation.

[29] Cox, D.R., 2020. Queues. Chapman and Hall/CRC.

[30] Schrage, L.E. and Miller, L.W., 1966. The queue M/G/1 with the shortest remaining processing time discipline. Operations Research, 14(4), pp.670-684.

[31] Schrage, L., 1968. Letter to the editor—a proof of the optimality of the shortest remaining processing time discipline. Operations Research, 16(3), pp.687-690.

[32] Spall, J.C., 2005. Introduction to stochastic search and optimization: estimation, simulation, and control (Vol. 65). John Wiley & Sons.

[33] Maurer, S.M. and Huberman, B.A., 2001. Restart strategies and Internet congestion. Journal of Economic Dynamics and Control, 25(3-4), pp.641-654.

[34] Evans, M.R. and Majumdar, S.N., 2011. Diffusion with stochastic resetting. Physical review letters, 106(16), p.160601.

[35] Evans, M.R. and Majumdar, S.N., 2011. Diffusion with optimal resetting. Journal of Physics A: Mathematical and Theoretical, 44(43), p.435001.

[36] Majumdar, S.N., Sabhapandit, S. and Schehr, G., 2015. Dynamical transition in the temporal relaxation of stochastic processes under resetting. Physical Review E, 91(5), p.052131.

[37] Pal, A., 2015. Diffusion in a potential landscape with stochastic resetting. Physical Review E, 91(1), p.012113.

[38] Eule, S. and Metzger, J.J., 2016. Non-equilibrium steady states of stochastic processes with intermittent resetting. New Journal of Physics, 18(3), p.033006.

[39] Nagar, A. and Gupta, S., 2016. Diffusion with stochastic resetting at power-law times. Physical Review E, 93(6), p.060102.

[40] Pal, A., Kundu, A. and Evans, M.R., 2016. Diffusion under time-dependent resetting. Journal of Physics A: Mathematical and Theoretical, 49(22), p.225001.

[41] Gupta, S., Majumdar, S.N. and Schehr, G., 2014. Fluctuating interfaces subject to stochastic resetting. Physical review letters, 112(22), p.220601.

[42] Reuveni, S., Urbakh, M. and Klafter, J., 2014. Role of substrate unbinding in Michaelis–Menten enzymatic reactions. Proceedings of the National Academy of Sciences, 111(12), pp.4391-4396.

[43] Rotbart, T., Reuveni, S. and Urbakh, M., 2015. Michaelis-Menten reaction scheme as a unified approach towards the optimal restart problem. Physical Review E, 92(6), p.060101.

[44] Roldan, E., Lisica, A., Sanchez-Taltavull, D. and Grill, S.W., 2016. Stochastic resetting in backtrack recovery by RNA polymerases. Physical Review E, 93(6), p.062411.

[45] Bressloff, P.C., 2020. Modeling active cellular transport as a directed search process with stochastic resetting and delays. Journal of Physics A: Mathematical and Theoretical, 53(35), p.355001.

[46] Budnar, S., Husain, K.B., Gomez, G.A., Naghibosadat, M., Varma, A., Verma, S., Hamilton, N.A., Morris, R.G. and Yap, A.S., 2019. Anillin promotes cell contractility by cyclic resetting of RhoA residence kinetics. Developmental cell, 49(6), pp.894-906.

[47] Bressloff, P.C., 2020. Queueing theory of search processes with stochastic resetting. Physical Review E, 102(3), p.032109.

[48] Luby, M., Sinclair, A. and Zuckerman, D., 1993. Optimal speedup of Las Vegas algorithms. Information Processing Letters, 47(4), pp.173-180.

[49] Hamlin, P., Thrasher, W.J., Keyrouz, W. and Mascagni, M., 2019. Geometry entrapment in Walk-on-Subdomains. Monte Carlo Methods and Applications, 25(4), pp.329-340.

[50] Langville, A.N. and Meyer, C.D., 2004. Deeper inside pagerank. Internet Mathematics, 1(3), pp.335-380.

[51] Guan, Y. and Krone, S.M., 2007. Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing.

[52] Avrachenkov, K., Ribeiro, B. and Towsley, D., 2010. Improving random walk estimation accuracy with uniform restarts. In Algorithms and Models for the Web-Graph: 7th International Workshop, WAW 2010, Stanford, CA, USA, December 13-14, 2010. Proceedings 7 (pp. 98-109). Springer Berlin Heidelberg.

[53] Avrachenkov, K., Van Der Hofstad, R. and Sokol, M., 2014. Personalized pagerank with node-dependent restart. In Algorithms and Models for the Web Graph: 11th International Workshop, WAW 2014, Beijing, China, December 17-18, 2014, Proceedings 11 (pp. 23-33). Springer International Publishing.

[54] Avrachenkov, K., Piunovskiy, A. and Zhang, Y., 2018. Hitting times in Markov chains with restart and their application to network centrality. Methodology and Computing in Applied Probability, 20, pp.1173-1188.

[55] Borkar, V.S. and Chaudhuri, S., 2021. Accelerating MCMC by Rare Intermittent Resets. In Performance Evaluation Methodologies and Tools: 14th EAI International Conference, VALUETOOLS 2021, Virtual Event, October 30–31, 2021, Proceedings (pp. 107-125). Springer International Publishing.

[56] Evans, M.R., Majumdar, S.N. and Schehr, G., 2020. Stochastic

resetting and applications. Journal of Physics A: Mathematical and Theoretical, 53(19), p.193001.

[57] Pal, A., Kostinski, S. and Reuveni, S., 2022. The inspection paradox in stochastic resetting. Journal of Physics A: Mathematical and Theoretical, 55(2), p.021001.

[58] Tal-Friedman, O., Pal, A., Sekhon, A., Reuveni, S. and Roichman, Y., 2020. Experimental realization of diffusion with stochastic resetting. J. Phys. Chem. Lett. 2020, 11, 17, 7350–7355

[59] Besga, B., Bovon, A., Petrosyan, A., Majumdar, S.N. and Ciliberto, S., 2020. Optimal mean first-passage time for a Brownian searcher subjected to resetting: experimental and theoretical results. Physical Review Research, 2(3), p.032029.

[60] Pal, A. and Reuveni, S., 2017. First Passage under Restart. Physical review letters, 118(3), p.030603.

[61] Reuveni, S., 2016. Optimal stochastic restart renders fluctuations in first passage times universal. Physical review letters, 116(17), p.170601.

[62] Pal, A., Eliazar, I. and Reuveni, S., 2019. First passage under restart with branching. Physical review letters, 122(2), p.020602.

[63] Belan, S., 2018. Restart could optimize the probability of success in a Bernoulli trial. Physical review letters, 120(8), p.080601.

[64] Chechkin, A. and Sokolov, I.M., 2018. Random search with resetting: a unified renewal approach. Physical review letters, 121(5), p.050601.

[65] Pal, A. and Prasad, V.V., 2019. First passage under stochastic resetting in an interval. Physical Review E, 99(3), p.032123.

[66] Ray, S., Mondal, D. and Reuveni, S., 2019. Peclet number governs transition to acceleratory restart in drift-diffusion. Journal of Physics A: Mathematical and Theoretical, 52(25), p.255002.

[67] Ray, S., 2020. Space-dependent diffusion with stochastic resetting: A first-passage study. The Journal of Chemical Physics, 153(23), p.234904.

[68] Bonomo, O.L. and Pal, A., 2021. First passage under restart for discrete space and time: application to one-dimensional confined lattice random walks. Physical Review E, 103(5), p.052129.

[69] Libman, L. and Orda, A., 2002. Optimal retrial and timeout strategies for accessing network resources. IEEE/ACM Transactions on Networking, 10(4), pp.551-564.

[70] Zhang, L., Wu, F. and Huberman, B.A., 2008. Games and queues.

[71] Sheahan, R., Lipsky, L., Fiorini, P.M. and Asmussen, S., 2006. On the completion time distribution for tasks that must restart from the beginning if a failure occurs. ACM SIGMETRICS Performance Evaluation Review, 34(3), pp.24-26.

[72] Asmussen, S. and Rønn-Nielsen, A., 2010. Failure recovery via RESTART: Wallclock models. TN Thiele Centre, University of Aarhus.

[73] Stojkoski, V., Jolakoski, P., Pal, A., Sandev, T., Kocarev, L. and Metzler, R., 2022. Income inequality and mobility in geometric Brownian motion with stochastic resetting: theoretical results and empirical evidence of non-ergodicity. Philosophical Transactions of the Royal Society A, 380(2224), p.20210157.

[74] Jolakoski, P., Pal, A., Sandev, T., Kocarev, L., Metzler, R. and Stojkoski, V., 2022. The fate of the American dream: A first passage under resetting approach to income dynamics. arXiv preprint arXiv:2212.13176.

[75] Chao, X., 1995. A queueing network model with catastrophes and product form solution. Operations Research Letters, 18(2), pp.75-79.

[76] Kumar, B.K. and Arivudainambi, D., 2000. Transient solution of an M/M/1 queue with catastrophes. Computers & Mathematics with applications, 40(10-11), pp.1233-1240.

[77] Di Crescenzo, A., Giorno, V., Nobile, A.G. and Ricciardi, L.M., 2003. On the M/M/1 queue with catastrophes and its continuous approximation. Queueing Systems, 43(4), pp.329-347.

[78] Paz, N. and Yechiali, U., 2014. An M/M/1 queue in random environment with disasters. Asia-Pacific Journal of Operational Research, 31(03), p.1450016.

[79] Foss, S., Korshunov, D. and Zachary, S., 2011. An introduction to heavy-tailed and subexponential distributions (Vol. 6, pp. 0090-6778). New York: Springer.

[80] Barlow, R.E. and Proschan, F., 1996. Mathematical theory of reliability. Society for Industrial and Applied Mathematics.

[81] Kleinrock, L., 1975. Queueing Systems, Vol. 1: Theory. New York: John Wiley and Sons.

[82] Kingman, J.F.C., 1961, October. The single server queue in heavy traffic. In Mathematical Proceedings of the Cambridge Philosophical Society (Vol. 57, No. 4, pp. 902-904). Cambridge University Press.

[83] Pal, A., Kuśmierz, Ł. and Reuveni, S., 2020. Search with home returns provides advantage under high uncertainty. Physical Review Research, 2(4), p.043174.

[84] Evans, M.R. and Majumdar, S.N., 2018. Effects of refractory period on stochastic resetting. Journal of Physics A: Mathematical and Theoretical, 52(1), p.01LT01.