Power-law and exponential tails in a stochastic priority-based model queue

G. Grinstein and R. Linsker
IBM T. J. Watson Research Center, P. O. Box 218, Yorktown Heights, New York 10598, USA
(Received 18 September 2007; published 7 January 2008)

We derive exact asymptotic results for a stochastic queueing model in which tasks are executed according to a continuous-valued priority. The distribution $P(\tau)$ of the waiting times τ of executed tasks for this model is shown to behave asymptotically as a power law, $P(\tau)\sim\tau^{-3/2}$, when the average rates of task arrival λ and execution μ satisfy $\mu\le\lambda$ (as was earlier noted empirically). For $\mu>\lambda$, $P(\tau)\sim\tau^{-5/2}\exp[-(\sqrt{\mu}-\sqrt{\lambda})^2\tau]$.

DOI: 10.1103/PhysRevE.77.012101 PACS number(s): 02.50.Ey, 89.20.−a

I. INTRODUCTION

In an interesting recent series of papers [1–3], Barabási and co-workers analyzed data on activities such as web browsing, library use, and the exchange of letters and e-mail messages. Studying the distributions $P(\tau)$ of waiting times τ between successive events—for example, the time it takes for the recipient of a letter or e-mail message to reply to it—they reported that the distributions exhibited heavy tails consistent with power laws $P(\tau)\sim\tau^{-\alpha}$ over some range of τ. The reported values of α were close to 3/2 for written correspondence and close to 1 for the other activities. These results stand in contrast to much of the large literature on waiting-time distributions for real and model queues, where exponential decays of $P(\tau)$ are typical [4].

Several elementary queueing models were considered in Refs. [1–3] to try to explain the observed behavior. One of these models was the continuous-priority version of a classic model by Cobham [5]—a stochastic queue wherein tasks are selected for execution on the basis of a discrete task priority value. The continuous-priority version of this model was indeed found numerically to exhibit asymptotic power-law behavior of $P(\tau)$, with exponent 3/2, provided the average rate λ of task arrival equals [1–3] or exceeds [2,3] the average rate μ of task execution. We subsequently derived this result analytically [6], as well as the new result [6] $P(\tau)\sim\tau^{-5/2}\exp(-\tau/\tau_0)$ for $\mu>\lambda$, by approximating the number of tasks in the queue by a continuous variable, and thereby mapping the model onto the familiar problem of biased diffusion; here τ_0 is a characteristic time that diverges as $(\mu-\lambda)^{-2}$ as λ approaches μ from below.

In the present paper, we solve the model without approximation, deriving exact asymptotic results for the realistic case in which the number of tasks is an integer [7]. We show that indeed $P(\tau)\sim\tau^{-3/2}$ for $\mu\le\lambda$, and that $P(\tau)\sim\tau^{-5/2}\exp[-(\sqrt{\mu}-\sqrt{\lambda})^2\tau]$ for $\mu>\lambda$.

II. THE MODEL: STOCHASTIC, CONTINUOUS-PRIORITY QUEUE

In the continuous-priority version [1] of the Cobham model [5], new tasks, having priority x (with $0\le x\le 1$) chosen randomly from an arbitrary probability distribution $\rho(x)$ (which we make uniform by a simple transformation, without any loss of generality [6]), arrive in the queue with an average rate λ. The highest-priority task in the queue is executed with an average rate μ. As in Ref. [1], we consider the case in which the removal from the queue of a task that has been selected for execution occurs instantaneously (i.e., with zero service time).

III. ANALYSIS

In a manner similar to Ref. [6], the overall probability per unit time, $P(\tau)$, that a given task sits in the queue for a time τ before being executed is conveniently expressed in terms of two quantities: (a) the probability per unit time, $G(n,x,\tau)$, that a given task of priority x, which arrives in the queue at time $t=0$ with exactly n items of higher priority (i.e., larger x) already in the queue, gets executed at time τ; and (b) the probability $\tilde{Q}(n,x)$ of there being exactly n items in the queue with priority greater than x, once a steady state has been achieved [cf. Ref. [8], Eq. (2.10)]:

$$P(\tau)=\sum_{n=0}^{\infty}\int_0^1 dx\,\tilde{Q}(n,x)G(n,x,\tau). \tag{1}$$

As in [6], let $Q(m,x,t)$ be the probability that at time t there are exactly m tasks with priority greater than x in the queue, for $m=0,1,2,\ldots$. Then $Q(m,x,t)$ satisfies the following master equations, for $m>0$ and $m=0$, respectively:

$$\partial Q(m,x,t)/\partial t = aQ(m+1,x,t)+bQ(m-1,x,t) - (a+b)Q(m,x,t),$$

$$\partial Q(0,x,t)/\partial t = aQ(1,x,t)-bQ(0,x,t); \tag{2}$$

here $a=\mu$ and $b(x)=\lambda\int_x^1\rho(z)dz=\lambda(1-x)$ are the respective probabilities per unit time of the number of tasks with priorities greater than x in the queue decreasing by 1 (provided $m\ge 1$) and increasing by 1. (We will generally suppress the explicit x dependence of b in what follows.)

As in Ref. [6], it is simple to show that for $\mu>\lambda$ the steady-state solution of Eqs. (2) achieved as $t\to\infty$, where $\partial Q(m,x,t)/\partial t$ vanishes for all m, is

$$\tilde{Q}(m,x)=(1-b/a)(b/a)^m. \tag{3}$$

Since (again, as in [6]), $b(0)$ approaches a as λ approaches μ from below, the distribution $\tilde{Q}(m,0)$ becomes uniform in m, and the mean number of tasks in the queue in steady state, $\langle m(x=0)\rangle$, diverges as $1/(\mu-\lambda)$. Thus, though in this strict sense the steady-state distribution is ill defined for $\lambda=\mu$ [4,8], in fact the queue does have well-defined steady-state properties since, for $\lambda=\mu$, the mean number of tasks with priorities greater than x is $\langle m(x)\rangle=(1/x-1)$ and thus remains finite for any $x>0$. (The case $\mu<\lambda$, shown in [6] to be closely related to the case $\mu=\lambda$, will be summarized at the end of the paper.)

We continue with the analysis, restricting ourselves to $\mu \geq \lambda$ for now. To compute $G(n,x,\tau)$, consider $p(m,x,t)$, the probability that at time t there are exactly m tasks with priority greater than or equal to x in the queue, and that these tasks include the task with priority x that arrived at time $t=0$; here m is a positive integer. For $m>1$ and $m=1$, respectively, $p(m,x,t)$ satisfies master equations similar to those for $Q(m,x,t)$:

$$\partial p(m,x,t)/\partial t = ap(m+1,x,t) + bp(m-1,x,t) - (a+b)p(m,x,t),$$

$$\partial p(1,x,t)/\partial t = ap(2,x,t) - (a+b)p(1,x,t). \tag{4}$$

Since there are exactly n tasks with priority greater than x at $t=0$, the initial condition on these equations is $p(m,x,t=0)=\delta_{m,n+1}$.

The only difference between the equations for Q and for p is at the boundary: It is easy to see that the total probability $Q_{\text{tot}}(x,t)\equiv\Sigma_{m=0}^{\infty}Q(m,x,t)$ has a vanishing time derivative $\dot{Q}_{\text{tot}}(x,t)=0$, whereas the derivative $\dot{p}_{\text{tot}}(x,t)$ of the probability $p_{\text{tot}}(x,t)\equiv\Sigma_{m=1}^{\infty}p(m,x,t)$ satisfies $\dot{p}_{\text{tot}}(x,t)=-ap(1,x,t)$. This reflects the fact that $m=0$ is an absorbing boundary for $p(m,x,t)$: When $m=1$, only the task with priority x that arrived in the queue at $t=0$ is present. If that task is executed, then the process of emptying the queue of tasks whose priorities equal or exceed x is complete, which implies that $G(n,x,\tau)=-\dot{p}_{\text{tot}}(x,t)=ap(1,x,t)$.

Equations (4), which represent a random walk with a drift velocity (see [9] and [6]), are solved by standard Laplace transform methods (e.g., [9]), the chain of equations for the Laplace transform $\tilde{p}(m,x,s)$ of $p(m,x,t)$ taking the forms, for $m>1$ and $m=1$, respectively,

$$(s+a+b)\tilde{p}(m,x,s) = a\tilde{p}(m+1,x,s) + b\tilde{p}(m-1,x,s) + \delta_{m,n+1},$$

$$(s+a+b)\tilde{p}(1,x,s) = a\tilde{p}(2,x,s) + \delta_{1,n+1}. \tag{5}$$

It is easy to verify that the solution of these equations is given by $\tilde{p}(m,x,s)=c_m\tilde{p}(1,x,s)+d_m$, where $c_m\equiv(\beta_+^m-\beta_-^m)/(\beta_+-\beta_-)$ and $d_m\equiv\theta(m-n-1)(\beta_-^{m-n-1}-\beta_+^{m-n-1})/a(\beta_+-\beta_-)$. Here $\theta(n)\equiv0$ for $n\leq0$ and $\theta(n)\equiv1$ for $n>0$, and $\beta_\pm$ are the solutions of the quadratic equation $a\beta^2-(s+a+b)\beta+b=0$. We define β_+ to be the solution that approaches s/a for large $|s|$. It can then be shown [10] that $|\beta_+|\geq|\beta_-|$ for all s. It follows that, at all s for which $|\beta_+|>1$ (which is the case for all sufficiently large $|s|$), $\tilde{p}(1,x,s)$ must be given by $\tilde{p}(1,x,s)=(\beta_+)^{-n-1}/a$ in order for $\tilde{p}(m,x,s)$ to be bounded for large m. Hence $\tilde{G}(n,x,s)$, the Laplace transform of $G(n,x,t)$, is given by $\tilde{G}(n,x,s)=(\beta_+)^{-n-1}$. From Eq. (1), we then have the expression

$$P(\tau) = \sum_{n=0}^{\infty}\int_0^1 dx\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{ds}{2\pi i}e^{s\tau}(1-b/a)(b/a)^n(\beta_+)^{-n-1} \tag{6}$$

for the desired quantity $P(\tau)$. Here γ is an arbitrary positive real number that is chosen large enough so that $|\beta_+|>1$ for all s on the integration path. Performing the sum over n yields

$$P(\tau) = \int_0^1 dx\int_{\gamma-i\infty}^{\gamma+i\infty}\frac{ds}{2\pi i}e^{s\tau}(1-b/a)(\beta_+-b/a)^{-1}. \tag{7}$$

Aside from the factor $e^{s\tau}$, the only s dependence of the integrand comes from β_+. The only singularities of the integrand occur where β_+ itself has singularities, i.e., along a branch cut on the real axis between $s_-\equiv-(\sqrt{a}+\sqrt{b})^2$ and $s_+\equiv-(\sqrt{a}-\sqrt{b})^2$. It is easy to show that this branch cut lies entirely to the left of the integration path in Eq. (7). [The factor (β_+-b/a) vanishes only if $\mu=\lambda$ and $x=0$, and then only at $s=s_+=0$, which is the branch point at the right end of the branch cut.] The integral is performed by closing the contour at large $|s|$ in the left half plane. We now deform this contour into one that runs just above and below the branch cut and encloses it counterclockwise in standard fashion [11], yielding the following expression for $P(\tau)$:

$$P(\tau) = \int_0^1 dx\int_{s_-}^{s_+}ai\frac{ds}{2\pi i}e^{s\tau}(1-b/a)\frac{[(s-s_-)(s_+-s)]^{1/2}}{(-sb)}. \tag{8}$$

It remains only to extract the leading behavior of this double integral for asymptotically large τ. Note that, as τ gets large, the integral is dominated increasingly by values of s near the upper limit s_+ of the integration range. Writing $s=s_+-u$, we have

$$P(\tau) = \int_0^1 dx\int_0^{s_+-s_-}du\frac{e^{(s_+-u)\tau}(a-b)[u(s_+-u-s_-)]^{1/2}}{2\pi b(u-s_+)}. \tag{9}$$

We now distinguish two cases.

(a) $\mu=\lambda$. Here $s_+=-(\sqrt{a}-\sqrt{b})^2=-\lambda x^2/4+O(x^3)$, whereupon the largest values of the exponentials and hence the leading asymptotic behavior of the double integral come from the region where both x and u are near 0. Using $a-b=\lambda x$, we expand the integrand to leading order in x and u, obtaining, up to a multiplicative constant, $e^{-(\lambda x^2/4+u)\tau}xu^{1/2}/(Au+Bx^2)$, for positive constants A and B. Sending the upper limits on both the u and x integrals to infinity, we perform the rescalings $x=x'/\tau^{1/2}$ and $u=u'/\tau$ to yield $P(\tau)\sim\tau^{-3/2}$, the prefactor being a convergent double integral over x' and u'.

(b) $\mu>\lambda$. Here $a-b=\mu-\lambda+\lambda x$, and so approaches a nonzero constant as $x\to0$, as does s_+, which behaves like $s_+\sim-[(\sqrt{\mu}-\sqrt{\lambda})^2+vx+O(x^2)]$, where $v\equiv\sqrt{\lambda}(\sqrt{\mu}-\sqrt{\lambda})$. Thus the exponential factor in the integrand behaves as $e^{-(\sqrt{\mu}-\sqrt{\lambda})^2\tau}e^{-[vx+u+O(x^2)]\tau}$, from which it follows that, again, the asymptotic behavior comes from x and u both near 0. All the nonexponential factors of the integrand approach constants at $x=u=0$, except for the factor $u^{1/2}$. Sending the upper limits on the integrals to infinity and performing the rescalings $x=x'/\tau$ and $u=u'/\tau$, we find, for asymptotically large τ, $P(\tau)\sim\tau^{-5/2}e^{-\tau/\tau_0}$ [12], where $\tau_0\equiv(\sqrt{\mu}-\sqrt{\lambda})^{-2}$.

It is instructive to compare this result with that for the continuous-task approximation of Ref. [6] [see Model A, case (2), $\lambda<\mu$]. There we obtained $P(\tau)\sim\tau^{-5/2}\exp(-\tau/\tau_0)$ where $1/\tau_0=(\mu-\lambda)^2/[4\mu(1-\lambda)]$, μ and λ were probabilities per discrete time step, and τ_0 was the exponential time constant expressed as a number of time steps. If successive time steps are spaced Δt apart, the task execution and arrival rates are $\mu'=\mu/\Delta t$ and $\lambda'=\lambda/\Delta t$, and the exponential time constant is $\tau_0'=\tau_0\Delta t$. For the continuous-time limit, we take $\Delta t\to 0$, keeping $\mu'$and λ' constant, and obtain $1/\tau_0'=\mu'(1-\eta)^2/4$ where $\eta\equiv\lambda'/\mu'$. By way of comparison, the present paper yields (affixing primes to keep the notation consistent) $1/\tau_0'=(\sqrt{\mu'}-\sqrt{\lambda'})^2=\mu'(1-\sqrt{\eta})^2$. The critical point is at $\eta=1$, where this case (b) changes over to case (a). In the limit $\eta\to 1$ from below, we have $\mu'(1-\sqrt{\eta})^2\to\mu'[(1-\eta)^2/4+O((1-\eta)^3)]$, which agrees near the critical point, as it should, with the result from Ref. [6].

Finally, we consider the case $\mu<\lambda$, wherein tasks arrive more frequently than they are executed on average, producing a queue that grows in time t as $(\lambda-\mu)t$. As we pointed out in Ref. [6], however, tasks with priorities $x>x^\star\equiv(\lambda-\mu)/\lambda$ are executed more frequently than they arrive on average, since their rate of arrival, $\lambda(1-x)$, is less than μ. As in [6], therefore, the analysis of these tasks can be shown to be identical to that of case (a) above. In particular, the change of variable from x to w defined by $x=x^\star+(1-x^\star)w$ maps the range $x^\star<x<1$ onto the range $0<w<1$, and transforms the quantities $a\equiv\mu$ and $b(x)\equiv\lambda(1-x)$ in Eqs. (2) and (4) into $\tilde{a}=\tilde{\mu}$ and $\tilde{b}(w)=\lambda(1-x^\star)(1-w)=\tilde{\lambda}(1-w)$, respectively, where $\tilde{\mu}=\tilde{\lambda}=\mu$. Hence the problem of the tasks with $x>x^\star$ maps precisely onto the original problem with $\mu=\lambda$, i.e., case (a) above, for which $P(\tau)\sim\tau^{-3/2}$.

Since, moreover, the total number of tasks with $x<x^\star$ in the queue grows without bound as time progresses, the probability of executing such a task approaches 0 in the long-time limit. Asymptotically, therefore, all such tasks remain in the queue forever.

ACKNOWLEDGMENT

We thank Sid Redner for providing helpful information about first-passage probabilities.

[1] A.-L. Barabási, Nature (London) **435**, 207 (2005), and supplementary online information.

[2] A. Vázquez *et al.*, Phys. Rev. E **73**, 036127 (2006).

[3] J. G. Oliveira and A.-L. Barabási, Nature (London) **437**, 1251 (2005), and supplementary online information.

[4] For example, D. R. Cox and W. L. Smith, *Queues* (Methuen & Co., London, 1961); D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. (Wiley, New York, 1998).

[5] A. Cobham, J. Oper. Res. Soc. Am. **2**, 70 (1954).

[6] G. Grinstein and R. Linsker, Phys. Rev. Lett. **97**, 130201 (2006).

[7] These results were announced but not derived in Ref. [6].

[8] J. Abate and W. Whitt, Queueing Syst. **25**, 173 (1997), and references therein.

[9] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, New York, 2001).

[10] To see that $|\beta_+|\geq|\beta_-|$ for all s, define $z\equiv s+a+b\equiv re^{i\phi}$ and $z_\pm\equiv z\mp 2\sqrt{ab}\equiv r_\pm e^{i\phi_\pm}$. Then $\beta_\pm=(z\pm\sqrt{z_+z_-})/2a$, so that $|\beta_+|<|\beta_-|$ if and only if $|\phi_++\phi_--2\phi|>\pi$, which is impossible. Also, $|\beta_+|=|\beta_-|$ only for s on the branch cut of β_+ and β_-, i.e., on the interval $[s_-,s_+]$ of the real s axis, where $s_\pm\equiv-(\sqrt{a}\mp\sqrt{b})^2$. Note also that, since $\beta_-\beta_+=b/a\leq 1$, $|\beta_-|\leq 1$ for all s.

[11] See, e.g., G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable: Theory and Technique* (McGraw-Hill, New York, 1966).

[12] This result differs from the $P(\tau)\approx\tau^{-3/2}\exp(-\tau/\tau_0)$ form obtained numerically in Refs. [1–3]. Just as in Ref. [6], however, when λ is just slightly less than μ, making τ_0 large, there is an intermediate τ regime, $1\ll\mu\tau\ll\mu\tau_0$, where the x^2 term of s_+ dominates the term of order x for almost the entire range of x. Thus the scaling of the x integral in Eq. (9) proceeds as for $\mu=\lambda$ [case (a)], yielding the result $P(\tau)\approx\tau^{-3/2}\exp(-\tau/\tau_0)$, valid in the intermediate τ regime, but not for asymptotically large τ.