Priority queues with bursty arrivals of incoming tasks

N. Masuda,[1] J. S. Kim,[2] and B. Kahng[2]

[1]*Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan*
[2]*Department of Physics and Astronomy, Seoul National University, Seoul 151-747, Korea*
(Received 6 September 2008; published 16 March 2009)

Recently increased accessibility of large-scale digital records enables one to monitor human activities such as the interevent time distributions between two consecutive visits to a web portal by a single user, two consecutive emails sent out by a user, two consecutive library loans made by a single individual, etc. Interestingly, those distributions exhibit a universal behavior, $D(\tau) \sim \tau^{-\delta}$, where τ is the interevent time, and $\delta \simeq 1$ or $3/2$. The universal behaviors have been modeled via the waiting-time distribution of a task in the queue operating based on priority; the waiting time follows a power-law distribution $P_w(\tau) \sim \tau^{-\alpha}$ with either $\alpha = 1$ or $3/2$ depending on the detail of queuing dynamics. In these models, the number of incoming tasks in a unit time interval has been assumed to follow a Poisson-type distribution. For an email system, however, the number of emails delivered to a mail box in a unit time we measured follows a power-law distribution with general exponent γ. For this case, we obtain analytically the exponent α, which is not necessarily 1 or $3/2$ and takes nonuniversal values depending on γ. We develop the generating function formalism to obtain the exponent α, which is distinct from the continuous time approximation used in the previous studies.

I. INTRODUCTION

In the digital era, human activities can be easily monitored and quantified by analyzing digital records such as the dates of sending or replying to emails, and financial transactions. Interestingly, human activities generate emerging patterns: the interevent time distribution of human activities follows a power law, and its exponent is either 1 or $3/2$ in many cases [1–5]. Such a bursty nature of human dynamics has been understood to be a consequence of queuing processes driven by human decision making. Barabási introduced a queuing model operating in the priority-based protocol [1]. At each time step, a task arrives at such a queue and is assigned a priority x_i chosen randomly from a distribution $\rho(x)$. Then, with probability p, the task with the highest priority is selected for execution and removed from the list. With probability $1-p$, a task is randomly selected irrespective of its priority and is executed. This model was successful in analytically reproducing the empirical result [1–3]: the waiting time of a task in the queue before being executed, which is denoted by τ, follows a power-law distribution $P_w(\tau) \sim \tau^{-1}$. The result is independent of distribution $\rho(x)$. The power law $P_w(\tau) \sim \tau^{-3/2}$ is reproduced by allowing the queue length to vary in time [1,3].

To analyze both fixed-length and flexible-length queues, the Barabási's model was extended as follows. In each time step, a task arrives with probability λ, and the task with the highest priority in the queue list is executed with probability μ. Operation of this queue system is schematically shown in Fig. 1(a). Since the dynamics of the queue is stochastic if $0 < \lambda < 1$ or $0 < \mu < 1$, the queue length generally changes in time. This model is a type of the $M/G/1$ queuing system with a priority selection rule proposed in the seminal work of Cobham in 1954 [6]. This model was analytically studied recently. The waiting-time distribution of a task in the queue changes depending on λ and μ. (i) When $\lambda = \mu = 1$, the number of tasks in the queue is fixed, and the waiting time of

tasks obeys $P_w(\tau) \sim \tau^{-2}$ [7]. (ii) When $\lambda = \mu < 1$, $P_w(\tau) \sim \tau^{-3/2}$ [8]. (iii) When $\lambda < \mu < 1$, $P_w(\tau) \sim \tau^{-3/2} e^{-\tau/\tau_0}$ for $\tau \ll \tau_0$ and $P_w(\tau) \sim \tau^{-5/2} e^{-\tau/\tau_0}$ for $\tau \gg \tau_0$, where the characteristic time scales as $\tau_0 = 1/(\sqrt{\mu} - \sqrt{\lambda})^2$ [8]. (iv) When $\mu < \lambda < 1$, tasks with priority $x < (\lambda - \mu)/\lambda$ wait in the queue forever without being executed. Tasks with priority $x \geq (\lambda - \mu)/\lambda$ are executed with the waiting time τ following $P_w(\tau) \sim \tau^{-3/2}$ [8].

Previous studies focused on the case in which incoming tasks are independent of each other and delivered to the queue at a constant rate. Thus, the number of incoming tasks in a unit time follows the Poisson distribution. This is the case observed in, for example, the number of requests for wireless phone calls arriving at a cell station in a unit time (see inset of Fig. 2). However, we observe that the number of emails received by a single user in a unit time is heterogeneous and follows a power-law distribution (see Fig. 2). Time intervals between consecutive tasks arriving at a server computer [10–13] and between a user's hypertext markup language (HTML) requests [5], which are closely related to the number of incoming tasks per unit time, also show similar patterns. The origin of such nonuniform numbers of incoming tasks is not known yet, but may be consequences of multiple correspondences with multiple people or self-similar patterns in the number of data packets arriving at a given router [14]. Such bursty arrivals of tasks may significantly change the behavior of priority queue systems. For example, a more skewed distribution of the number of incoming tasks per unit time may result in a more skewed waiting-time distribution of a task $P_w(\tau)$, as briefly suggested in [12]. In this paper, we study the waiting-time distribution of a task in the queue for the case of heterogeneous numbers of incoming tasks. We find that the universal power-law exponent $\alpha = 3/2$ for $P_w(\tau) \sim \tau^{-\alpha}$ occurs as a limited case and obtain other values of α depending on the power-law exponent of the distribution of the number of incoming tasks.
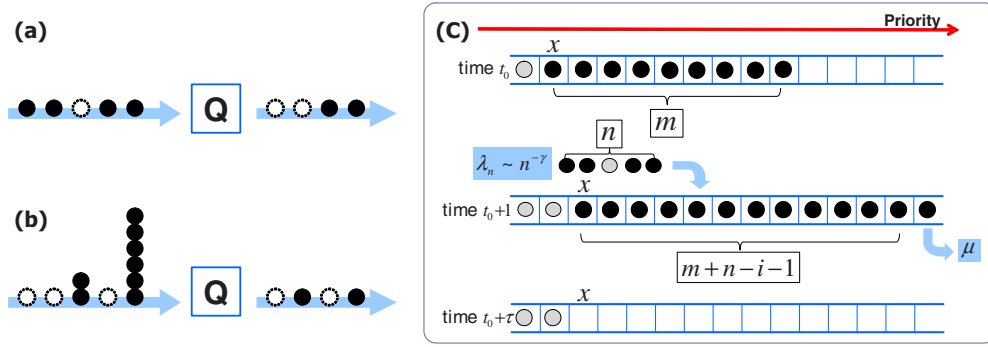
FIG. 1. (Color online) Schematic representation of queueing protocols. (a) A queue system proposed by Grinstein and Linsker, in which at most one task (filled circle) arrives in the system per time step. (b) The queue system we consider in this paper, in which input tasks (filled circles) can be bursty. (c) Operation of the queue system shown in (b): At time t_0, there are m tasks (black circles) with priority $\geq x$ in the queue. At time t_0+1, n tasks (black and gray circles) arrive in the queue with probability λ_n. Among them, $n-i$ tasks (black, not gray, circles) have priority $\geq x$. This event occurs with probability $\binom{n}{n-i}(1-x)^{n-i}x^i$, where $0 \leq i \leq n$. The task with the largest priority is executed with probability μ. No task is executed with probability $1-\mu$. At time $t_0+\tau$, the queue does not contain any tasks with priority $\geq x$ for the first time.

II. MODEL

We study the queue model defined as follows: in each discrete time step, n tasks are delivered to the queue, where n is distributed according to a power law $\lambda_n = \lambda n^{-\gamma}/\zeta(\gamma)$ ($n > 0$), $\lambda_0 = 1-\lambda$, where $0 \leq \lambda \leq 1$ and $\zeta(\gamma) \equiv \Sigma_{n'=1}^{\infty} n'^{-\gamma}$ is the Riemann ζ function. Each task is assigned a priority x uniformly distributed on $[0,1]$. At the same time, the task with the highest priority in a queue is executed with probability μ ($0 \leq \mu \leq 1$). Operation of this queue system is schematically depicted in Figs. 1(b) and 1(c). The queue length is unbounded so that the queue accommodates all incoming tasks. This model generalizes the model introduced by Grinstein and Linsker (GL) [8], which corresponds to $\lambda_0 = 1-\lambda$, $\lambda_1 = \lambda$, and $\lambda_n = 0$ for $n \geq 2$ in our model.

We will obtain the waiting-time distribution $P_w(\tau)$ for a task in the queue. To this end, we start with the probability that there are m tasks with priority larger than or equal to x in the queue at time t, which is denoted by $Q_x(m,t)$. We denote the queue-length distribution in the steady state by $\tilde{Q}_x(m) = \lim_{t\to\infty} Q_x(m,t)$. Note that the steady state exists only under a certain condition, as discussed later. We define $G_x(m,\tau)$ to be the probability that a given task with priority x arriving in the queue at time $t=t_0$ is executed at time $t=t_0+\tau$. When the task arrives in the steady state, there are already m tasks in the queue with priority larger than or equal to x, where m is distributed according to $\tilde{Q}_x(m)$. All of these m tasks are executed before the given task is executed. Then, the waiting-time distribution is obtained via the following formula [8]:

$$P_w(\tau) = \sum_{m=0}^{\infty} \int_0^1 dx \tilde{Q}_x(m) G_x(m,\tau), \tag{1}$$

where $G_x(m,\tau)$ is equivalent to the first passage probability that a random walker starting from position $m>0$ arrives at the origin at time τ for the first time. For a constant rate of incoming tasks, $\tilde{Q}_x(m)$, $G_x(m,\tau)$, and $P_w(\tau)$ can be obtained explicitly [8]. However, due to the complexity of our prob-

lem, we obtain them implicitly in terms of the generating functions. We define the generating function

$$\mathcal{P}_w(s) \equiv \sum_{\tau=1}^{\infty} P_w(\tau) s^{\tau}, \tag{2}$$

where $0 < s < 1$. Then,

$$\mathcal{P}_w(s) = \sum_{m=0}^{\infty} \int_0^1 dx \tilde{Q}_x(m) \mathcal{G}_x(m,s), \tag{3}$$

where $\mathcal{G}_x(m,s) \equiv \Sigma_\tau G_x(m,\tau) s^\tau$. Because the number of tasks in the queue decreases at most one per unit time, we obtain
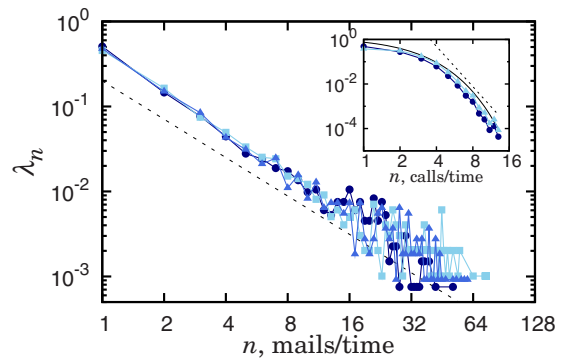

FIG. 2. (Color online) Distributions of the number of incoming tasks. The main panel shows the number of tasks delivered to an email box of an anonymous user per unit time [9], which follows a power-law distribution with slope -1.5. Different lines correspond to different bin sizes, namely, 500 ($\bigcirc$), 800 ($\triangle$), and 1000 ($\square$) seconds. Note that the slope -1.5 is not universal. It depends on users and can be as small as -3. We chose a user with the largest dataset. (Inset) The number of wireless phone calls arriving at a cell station in 10 seconds for the peak time (i.e., 12:00–20:00) ($\triangle$) and for the entire day ($\bigcirc$). Both data fit well to the Poisson distribution (black solid line), which decays even faster than the power law with exponent -6 (dotted line).

TABLE I. Power-law exponent α of the waiting-time distribution $P_w(\tau) \sim \tau^{-\alpha}$.

	$\langle n \rangle_\lambda < \mu$	$\langle n \rangle_\lambda \geq \mu$
$2 < \gamma \leq 3$	$\gamma - 1$	$\dfrac{2\gamma - 3}{\gamma - 1}$
$\gamma > 3$	$\gamma - 1$	$\dfrac{3}{2}$

$$G_x(m,t) = \sum_\tau G_x(m-1, t-\tau) f_x(\tau), \quad (4)$$

where $f_x(t) \equiv G_x(1,t)$. Equation (4) is expressed in terms of generating functions as

$$\mathcal{G}_x(m,s) = \mathcal{G}_x(m-1,s)\mathcal{F}_x(s), \quad (5)$$

where $\mathcal{F}_x(s) \equiv \Sigma_{t=1}^\infty f_x(t) s^t$. Applying Eq. (5) repeatedly, we obtain

$$\mathcal{G}_x(m,s) = \mathcal{F}_x^m(s). \quad (6)$$

Then, Eq. (3) is written as

$$\mathcal{P}_w(s) = \sum_{m=0}^\infty \int_0^1 dx \tilde{Q}_x(m) \mathcal{F}_x^m(s) = \int_0^1 dx \tilde{Q}_x(\mathcal{F}_x(s)), \quad (7)$$

where $\tilde{\mathcal{Q}}_x(z) \equiv \Sigma_{m=0} \tilde{Q}_x(m) z^m$.

Once we derive $\tilde{\mathcal{Q}}_x(z)$ and $\mathcal{F}_x(s)$ explicitly, we obtain the waiting-time distribution of a task in the queue, namely, $P_w(\tau)$. We will show that the waiting time exhibits a power-law behavior $P_w(\tau) \sim \tau^{-\alpha}$, where the values of α are shown in Table I. The analytic solutions are confirmed numerically in Fig. 3. Using our generating function formalism, we can also reproduce the results derived in Ref. [8], as shown in the Appendix.

III. QUEUE-LENGTH DISTRIBUTION

In this section, we calculate the queue-length distribution in the steady state by using the generating function $\tilde{\mathcal{Q}}_x(z)$. The master equation for $Q_x(m,t)$ is given by

$$Q_x(m, t+1) = \mu \sum_{j=0}^\infty \lambda_j x^j Q_x(m+1, t) \quad (8)$$

$$+ \sum_{i=0}^m (1-\mu) \sum_{j=i}^\infty \lambda_j \binom{j}{i}$$

$$\times (1-x)^i x^{j-i} Q_x(m-i, t) \quad (9)$$

$$+ \sum_{i=0}^m \mu \sum_{j=i+1}^\infty \lambda_j \binom{j}{i+1}$$

$$\times (1-x)^{i+1} x^{j-i-1} Q_x(m-i, t) \quad (10)$$

$$\equiv \sum_{i=-1}^m p_{m-i \to m} Q_x(m-i, t), \quad (m \geq 1). \quad (11)$$

In the above equation, the three terms in the right-hand side (RHS) correspond to different types of events that occur in a unit time. The first term (8) represents the case in which j ($j=0,1,\dots$) tasks arrive in the queue with probability λ_j, the priorities of all j tasks are smaller than x, and one task is executed with probability μ. The second term (9) represents the case in which j ($j=0,1,\dots$) tasks arrive in the queue with probability λ_j, i tasks out of the j tasks have priorities larger than or equal to x, and no task is executed with probability $1-\mu$. The third term (10) represents the case in which j ($j=0,1,\dots$) tasks arrive in the queue with probability λ_j, $i+1$ tasks out of the j tasks have priorities larger than or equal to x, and one task is executed with probability μ. For later discussion, we denote by $p_{m-i \to m}$ in Eq. (11) the transition probability of the random walk from position $m-i$ to position m in a unit time. The master equation at the boundary is given by

$$Q_x(0, t+1) = \mu \sum_{j=0}^\infty \lambda_j x^j Q_x(1,t) + \left[(1-\mu) \sum_{j=0}^\infty \lambda_j x^j \right.$$

$$\left. + \mu \sum_{j=1}^\infty \lambda_j j(1-x) x^{j-1} + \mu \sum_{j=0}^\infty \lambda_j x^j \right] Q_x(0, t)$$

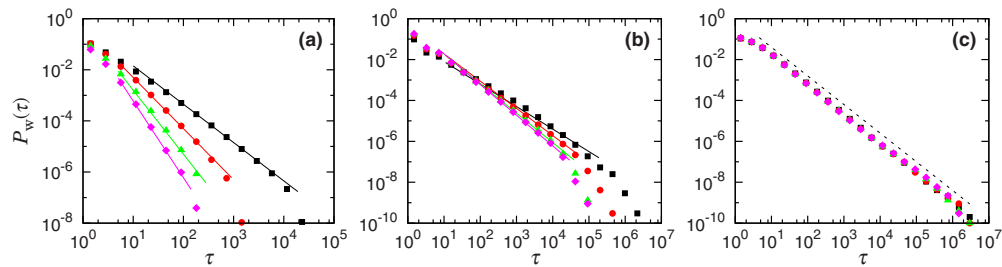$$\equiv p_{1 \to 0} Q_x(1,t) + p_{0 \to 0} Q_x(0,t). \quad (12)$$

FIG. 3. (Color online) The waiting-time distribution $P_w(t)$. (a) The case $\langle n \rangle_\lambda < \mu$. Given $\lambda = 0.3$ and $\mu = 1.0$, shown are numerically obtained $P_w(\tau)$ for $\gamma = 2.5$ ($\square$), 3.0 ($\bigcirc$), 3.5 ($\triangle$), and 4.0 ($\diamond$), yielding to $\langle n \rangle_\lambda \approx 0.58$, 0.41, 0.36, and 0.33, respectively. Solid lines indicate $P_w(\tau) \sim \tau^{-(\gamma-1)}$. (b) The case $\langle n \rangle_\lambda \geq \mu$ with $2 < \gamma \leq 3$. Given $\lambda = 0.5$ and $\mu = 0.5$, shown are numerically obtained $P_w(\tau)$ for $\gamma = 2.1$ ($\square$), 2.5 ($\bigcirc$), 2.8 ($\triangle$), and 3.0 ($\diamond$), yielding $\langle n \rangle_\lambda \approx 3.39$, 0.97, 0.75, and 0.68, respectively. Solid lines indicate $P_w(\tau) \sim \tau^{-(2\gamma-3)/(\gamma-1)}$. (c) The case $\langle n \rangle_\lambda > \mu$ with $\gamma > 3$. Given $\lambda = 0.5$ and $\mu = 0.3$, shown are numerically obtained $P_w(\tau)$ for $\gamma = 3.3$ ($\square$), 3.8 ($\bigcirc$), 4.0 ($\triangle$), and 4.5 ($\diamond$), yielding $\langle n \rangle_\lambda \approx 0.62$, 0.57, 0.56, and 0.53, respectively. The dotted line is a guideline with slope -1.4, close to the theoretical value -1.5.

Based on Eqs. (8)–(12), we calculate the generating function $\tilde{Q}_x(z)$ for the steady-state queue-length distribution $\tilde{Q}_x(m) \equiv \lim_{t \to \infty} Q_x(m, t)$. Specifically, the generating function of Eq. (8) is equal to $\mu \Lambda(x)[\tilde{Q}_x(z) - \tilde{Q}_x(0)]/z$ in the steady state, where $\Lambda(z) \equiv \Sigma_{j=0}^{\infty} \lambda_j z^j$. The generating function of Eq. (9) is equal to $(1-\mu)\tilde{Q}_x(z)\Lambda[(1-x)z+x]$. The generating function of Eq. (10) is equal to $\mu \tilde{Q}_x(z)\{\Lambda[(1-x)z+x] - \Lambda(x)\}$. The generating function of $\tilde{Q}_x(0)$ in Eq. (12) is equal to $\mu \Lambda(x)\tilde{Q}_x(0)$. Combining all these terms, we obtain

$$\tilde{Q}_x(z) = \frac{\mu \tilde{Q}_x(0)(z-1)\Lambda(x)}{z - (\mu + z - \mu z)\Lambda[(1-x)z+x]}. \qquad (13)$$

To eliminate $\tilde{Q}_x(0)$ from Eq. (13), we exploit the condition $\tilde{Q}_x(1)=1$. However, both the denominator and the numerator of Eq. (13) converge to zero as $z \to 1$. Thus, we apply the L'Hospital rule to Eq. (13) to derive

$$\tilde{Q}_x(0) = [\mu - (1-x)\langle n \rangle_\lambda]/(\mu \Lambda(x)), \qquad (14)$$

where $\langle n \rangle_\lambda \equiv \Sigma_{n=0}^{\infty} n \lambda_n$. Plugging Eq. (14) into Eq. (13) yields

$$\tilde{Q}_x(z) = \frac{[\mu - \langle n \rangle_\lambda(1-x)](z-1)}{z - (\mu + z - \mu z)\Lambda[(1-x)z+x]}. \qquad (15)$$

For the steady state to exist, the incoming rate of the task with larger than or equal to x [i.e., $\langle n \rangle_\lambda(1-x)$] must be smaller than the execution rate μ [1,8]; $A_1 \equiv \mu - \langle n \rangle_\lambda(1-x) > 0$ is required. In addition, $\langle n \rangle_\lambda$ must be finite, which is equivalent to the condition $\gamma > 2$.

The mean queue length denoted by $\langle m(x) \rangle_{\tilde{Q}}$ is derived as

$$\langle m(x) \rangle_{\tilde{Q}} = \left. \frac{\partial \tilde{Q}_x(z)}{\partial z} \right|_{z=1}$$

$$= \frac{2(1-\mu)\langle n \rangle_\lambda(1-x) + (\langle n^2 \rangle_\lambda - \langle n \rangle_\lambda)(1-x)^2}{2A_1}. \qquad (16)$$

Equation (16) implies that $\langle m(x) \rangle_{\tilde{Q}}$ diverges when $\langle n^2 \rangle_\lambda$ does, that is, when $\gamma \le 3$. When $\gamma > 3$, the queue length is finite for $x=0$ if and only if $\mu > \langle n \rangle_\lambda$ and diverges as $1/(\mu - \langle n \rangle_\lambda)$ as $\langle n \rangle_\lambda$ approaches μ from below, which extends the results in [8]. As $x \to 0$ and $\langle n \rangle_\lambda \to \mu$, $\langle m(x) \rangle_{\tilde{Q}}$ diverges as $1/x$, which is also consistent with the previous result [8].

To calculate the asymptotic behavior of the steady-state queue-length distribution $\tilde{Q}_x(m)$, we assume $\mu > \langle n \rangle_\lambda(1-x)$ and $\gamma > 2$, for which the steady state exists. When $2 < \gamma \le 3$, $\Lambda(z)$ is expanded near $z \to 1$ as follows [15]:

$$\Lambda(z) = 1 - \langle n \rangle_\lambda(1-z) + c_\gamma(1-z)^{\gamma-1} + o((1-z)^{\gamma-1}), \qquad (17)$$

where c_γ is a constant. Inserting Eq. (17) into Eq. (15) leads to

$$\tilde{Q}_x(z) = 1 - \frac{c_\gamma(1-x)^{\gamma-1}(1-z)^{\gamma-2}}{A_1} + o((1-z)^{\gamma-2}). \quad (18)$$

For $3 < \gamma \le 4$, we obtain

$$\Lambda(z) = 1 - \langle n \rangle_\lambda(1-z) + \frac{\langle n^2 \rangle_\lambda - \langle n \rangle_\lambda}{2}(1-z)^2 - c_\gamma(1-z)^{\gamma-1}$$

$$+ o((1-z)^{\gamma-1}), \qquad (19)$$

which leads to

$$\tilde{Q}_x(z) = 1 + \langle m(x) \rangle_{\tilde{Q}}(z-1) + \frac{c_\gamma(1-x)^{\gamma-1}}{A_1}(1-z)^{\gamma-2}$$

$$+ o((1-z)^{\gamma-2}). \qquad (20)$$

Similar expansions hold true for $\gamma > 4$. By applying the Tauberian theorem [15] to Eqs. (18) and (20), we obtain

$$\tilde{Q}_x(m) \sim \frac{1}{m^{\gamma-1}} (m \to \infty) \qquad (21)$$

for $\gamma > 2$. Equation (21) is consistent with the result under the first-in-first-out (FIFO) protocol [16]. This is because, when $\mu \gg \langle n \rangle_\lambda(1-x)$, tasks are executed upon its arrival in the steady state so that the priority-based protocol can be regarded as the FIFO-based one.

IV. FIRST-PASSAGE PROBABILITY

In this section, we derive $\mathcal{F}_x(s) = \Sigma_{t=1}^{\infty} f_x(t)s^t = \Sigma_{t=1}^{\infty} G_x(1,t)s^t$. Recall that $G_x(m,t)$ is the probability that a given task with priority x is executed at time t after its arrival, provided that there are m tasks in the queue with priority larger than or equal to x when this task arrives. This quantity can be interpreted as the first passage probability that a random walker on a half line starts from position m and arrives at the origin at time t for the first time. The probability that the random walker moves from i to j in a unit time is given by $p_{i \to j}$ [see Eq. (11)].

The generator of the one-step transition of the random walk before reaching the origin is represented by

$$\mathcal{P}(z) \equiv \sum_{i=-1}^{\infty} p_{m \to m+i} z^i = \left(1 - \mu + \frac{\mu}{z}\right)\Lambda[(1-x)z+x]. \qquad (22)$$

Note that the RHS of Eq. (22) is independent of m because the transition probability is homogeneous in space.

The amount of a single jump that the random walker makes to the right is unbounded, because it is equal to the number of incoming tasks with priority larger than or equal to x. However, the amount of a jump to the left is at most one, which yields a useful relation,

$$\mathcal{G}_x(i,s) = \mathcal{F}_x(s)^i. \qquad (23)$$

Using Eqs. (22) and (23), and the recursion relation [17,18],

$$f_x(t) = p_{1 \to 0} + p_{1 \to 1}G_x(1, t-1) + p_{1 \to 2}G_x(2, t-1) + \cdots, \qquad (24)$$

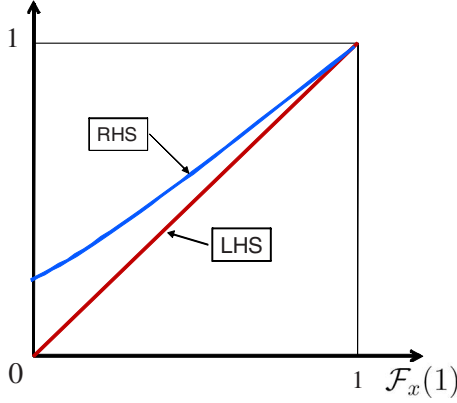we obtain the following self-consistent equation for the generating function:

FIG. 4. (Color online) Schematic representation of the LHS and the RHS of Eq. (26) as functions of $\mathcal{F}_x(1)$.

$$\mathcal{F}_x(s) = s\sum_{i=0}^{\infty} p_{1\to i}\mathcal{G}_x(i,s) = s\sum_{i=0}^{\infty} p_{1\to i}\mathcal{F}_x(s)^i = s\mathcal{F}_x(s)\mathcal{P}(\mathcal{F}_x(s))$$

$$= s[(1-\mu)\mathcal{F}_x(s) + \mu]\Lambda[(1-x)\mathcal{F}_x(s) + x]. \quad (25)$$

The first s in the RHS comes from the unit time spent by a single transition starting from $m=1$. After this transition, the generating function of the number of tasks with priority larger than or equal to x in the queue is $z\mathcal{P}(z)$. Since each such task incurs an execution time distributed according to $\{f_x(t)\}$, we replace z of $z\mathcal{P}(z)$ by $\mathcal{F}_x(s)$ to obtain Eq. (25).

We evaluate $\mathcal{F}_x(s)$ in the limit $s\to 1$ using Eq. (25). To guarantee that the task with priority x is eventually executed, $\mathcal{F}_x(s=1)=1$ has to be satisfied. To check if this condition is fulfilled, we put $s=1$ in Eq. (25) to obtain

$$\mathcal{F}_x(1) = s[(1-\mu)\mathcal{F}_x(1) + \mu]\Lambda[(1-x)\mathcal{F}_x(1) + x]. \quad (26)$$

The left-hand side (LHS) and the RHS of Eq. (26) are plotted in Fig. 4 as functions of $\mathcal{F}_x(1)$, where $\mathcal{F}_x(1)$ is regarded as a variable for the sake of this analysis. Note that the RHS of Eq. (26) is positive at $\mathcal{F}_x(1)=0$. Figure 4 implies that Eq. (26) has the unique solution $\mathcal{F}_x(1)=1$ if and only if the slope of the RHS of Eq. (26) at $\mathcal{F}_x(1)=1$ is less than or equal to unity, that is,

$$\frac{\partial}{\partial\mathcal{F}_x(1)}[(1-\mu)\mathcal{F}_x(1) + \mu]\Lambda[(1-x)\mathcal{F}_x(1) + x]\big|_{\mathcal{F}_x(1)=1} \le 1. \quad (27)$$

Equation (27) is equivalent to $A_1 \ge 0$, which is what we already assumed.

In the following, we obtain the solution of the self-consistent equation (25) by assuming that $f_x(t)$ follows a power law.

Case (i): $\mu > \langle n\rangle_\lambda$. In this case, $A_1 > 0$ holds for all x. When $2 < \gamma \le 3$, combining Eqs. (17) and (25) yields

$$\mathcal{F}_x(s) = 1 + \frac{1}{A_1}(s-1) + \frac{c_\gamma(1-x)^{\gamma-1}}{A_1^\gamma}(1-s)^{\gamma-1} + o((1-s)^{\gamma-1}). \quad (28)$$

When $3 < \gamma \le 4$, combining Eqs. (19) and (25) yields

$$\mathcal{F}_x(s) = 1 + \frac{1}{A_1}(s-1) + \frac{A_1 - A_1^2 + A_2}{A_1^3}(s-1)^2$$

$$- \frac{c_\gamma(1-x)^{\gamma-1}}{A_1^\gamma}(1-s)^{\gamma-1} + o((1-s)^{\gamma-1}), \quad (29)$$

where $A_2 \equiv (\langle n^2\rangle_\lambda - \langle n\rangle_\lambda)(1-x)^2/2 + (1-\mu)\langle n\rangle_\lambda(1-x) > 0$. Note that the coefficient of $(1-s)^2$ is positive. In a similar manner, we can show for $\gamma > 4$ that the leading singular term of $\mathcal{F}_x(s)$ is equal to $(-1)^{\lceil\gamma\rceil-1}c_\gamma(1-x)^{\gamma-1}(1-s)^{\gamma-1}/A_1^\gamma$, where $\lceil\gamma\rceil = \min\{i; i \ge \gamma, i \in \mathbf{Z}\}$. Thus, we obtain $f_x(t) \sim t^{-\beta}$ with $\beta = \gamma$ for $\gamma > 2$ using the Tauberian theorem [15].

Case (ii): $\mu = \langle n\rangle_\lambda$. Because $A_1 = 0$ for $x=0$, we cannot apply the results obtained for case (i). For example, $1/A_1 = 1/(\langle n\rangle_\lambda x)$ in the coefficient of $(s-1)$ in Eq. (28) diverges as $x \to 0$, implying that the exponent β is smaller than 2 near $x=0$. Actually the long-time behavior of $f_x(t)$ is dominated by the tasks whose priority is near $x=0$ [8]. Thus, we assume

$$\mathcal{F}_x(s) = 1 - c_\beta(1-s)^{\beta-1} + o((1-s)^{\beta-1}) \quad (30)$$

with $1 < \beta \le 2$.

When $2 < \gamma \le 3$, the RHS of Eq. (25) is written as

$$s[(1-\mu)\mathcal{F}_x(s) + \mu]\Lambda[(1-x)\mathcal{F}_x(s) + x]$$

$$= s\mathcal{F}_x(s) + \mu s[1 - \mathcal{F}_x(s)] + s\{\langle n\rangle_\lambda(1-x)[\mathcal{F}_x(s) - 1] + c_\gamma(1$$

$$- x)^{\gamma-1}[1 - \mathcal{F}_x(s)]^{(\gamma-1)} + \cdots\}. \quad (31)$$

Plugging Eq. (30) into the LHS and RHS of Eq. (25) leads to

$$(1-s)[1 - c_\beta(1-s)^{\beta-1} + \cdots]$$

$$= \langle n\rangle_\lambda x c_\beta(1-s)^{\beta-1} + c_\gamma(1-x)^{\gamma-1}c_\beta^{\gamma-1}(1-s)^{(\beta-1)(\gamma-1)}$$

$$+ \cdots . \quad (32)$$

If $\langle n\rangle_\lambda x \gg (1-s)^{(\gamma-2)/(\gamma-1)}$, the first term of the RHS of Eq. (32) is much larger than the second term as $s \to 1$ so that $\beta = 2$ and $c_\beta = 1/(\langle n\rangle_\lambda x)$. Conversely, if $\langle n\rangle_\lambda x \ll (1-s)^{(\gamma-2)/(\gamma-1)}$, the second term dominates the first term so that $\beta = 1 + 1/(\gamma-1)$ and $c_\beta = c_\gamma^{-1/(\gamma-1)}/(1-x) \simeq c_\gamma^{-1/(\gamma-1)}$.

When $3 < \gamma < 4$, as in the case of $2 < \gamma \le 3$, Eqs. (19), (25), and (30), with an appropriate assumption of $1 < \beta \le 2$, yield

$$(1-s) + o(1-s) = \langle n\rangle_\lambda x c_\beta(1-s)^{\beta-1} + A_2 c_\beta^2(1-s)^{2(\beta-1)}$$

$$- c_\gamma(1-x)^{\gamma-1}c_\beta^{\gamma-1}(1-s)^{(\beta-1)(\gamma-1)} + \cdots . \quad (33)$$

If $\langle n\rangle_\lambda x \gg \sqrt{A_2(1-s)}$, the first term in the RHS of Eq. (33) is much larger than the second term. Then $\beta = 2$ and $c_\beta = 1/(\langle n\rangle_\lambda x)$. Conversely, if $\langle n\rangle_\lambda x \ll \sqrt{A_2(1-s)}$, the second term is much larger than the first term so that $\beta = 3/2$ and $c_\beta = 1/\sqrt{A_2}$. The third term is always much smaller than the second term as $s \to 1$.

Case (iii): $\mu < \langle n\rangle_\lambda$. The task in the queue accumulates at rate $\langle n\rangle_\lambda - \mu$. In this case, only the tasks with priority $x > x_M \equiv (\langle n\rangle_\lambda - \mu)/\langle n\rangle_\lambda$ are executed, and the analysis can be ascribed to case (ii) [8]. Distributions of the priority of tasks in the queue in the steady state are shown in Fig. 5 for some values of $\langle n\rangle_\lambda$ and μ.
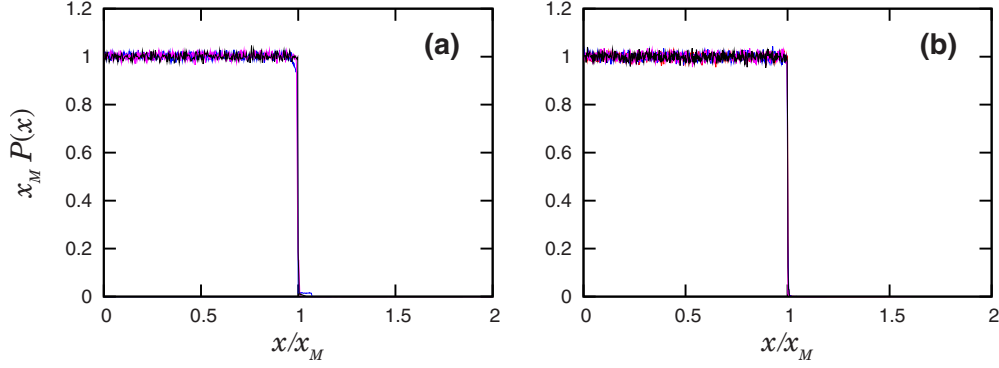
FIG. 5. (Color online) Distributions of the priority of tasks in the queue in the steady state. Distributions of x for different sets of $\langle n \rangle_\lambda$ and μ are shown. Only the tasks with priority $x > x_M \equiv (\langle n \rangle_\lambda - \mu)/\langle n \rangle_\lambda$ are executed. (a) $\lambda = 0.5$ and $\mu = 0.5$. We set $\gamma = 2.1$, 2.5, 2.8, and 3.0, which yield $x_M = 0.85$, 0.48, 0.33, and 0.26, respectively. (b) $\lambda = 0.5$ and $\mu = 0.3$. We set $\gamma = 3.3$, 3.8, 4.0, and 4.5, which yield 0.52, 0.47, 0.46, and 0.43, respectively. In each panel, the four plots almost collapse onto one.

V. WAITING-TIME DISTRIBUTION

Using Eqs. (7) and (15), and $\mathcal{F}_x(s)$ we obtained for the three cases, we calculate the waiting-time distribution as follows:

Case (i): $\mu > \langle n \rangle_\lambda$. The leading singular term of $\tilde{\mathcal{Q}}_x(\mathcal{F}_x(s))$ is equal to $(-1)^{\lceil \gamma \rceil} c_\gamma (1-x)^{\gamma-1}(1-s)^{\gamma-2}/A_1^{\gamma-1}$. Then, we obtain

$$\mathcal{P}_w(s) \sim (1-s)^{\gamma-2}, \tag{34}$$

which yields $P_w(\tau) \sim \tau^{-(\gamma-1)}$ for $\gamma > 2$.

Case (ii): $\mu = \langle n \rangle_\lambda$. In this case, we use Eq. (30) with values of β and c_β depending on γ and x.

For $2 < \gamma < 3$, we obtain

$$
\begin{aligned}
\mathcal{P}_w(s) &\simeq \int_0^{(1-s)^{(\gamma-2)/(\gamma-1)}} dx \langle n \rangle_\lambda x C_\gamma^{-1/(\gamma-1)}(1-s)^{1/(\gamma-1)-1} \\
&+ \int_{(1-s)^{(\gamma-2)/(\gamma-1)}}^1 dx + \cdots = 1 + \left(\frac{\langle n \rangle_\lambda c_\gamma^{-1/(\gamma-1)}}{2} - 1 \right) \\
&\times (1-s)^{(\gamma-2)/(\gamma-1)} + \cdots .
\end{aligned}
\tag{35}
$$

Therefore, $P_w(\tau) \sim \tau^{-(2\gamma-3)/(\gamma-1)}$.

For $3 < \gamma < 4$, we obtain

$$
\begin{aligned}
\mathcal{P}_w(s) &\simeq \int_0^{\sqrt{A_2(1-s)}/\langle n \rangle_\lambda} dx \frac{\langle n \rangle_\lambda x}{\sqrt{A_2(1-s)}} \\
&+ \int_{\sqrt{A_2(1-s)}/\langle n \rangle_\lambda}^1 dx + \cdots = 1 - \frac{\sqrt{A_2(1-s)}}{2\langle n \rangle_\lambda} + \cdots .
\end{aligned}
\tag{36}
$$

Therefore, $P_w(\tau) \sim \tau^{-3/2}$. Similar calculations yield $P_w(\tau) \sim \tau^{-3/2}$ for $\gamma > 4$.

Case (iii): $\mu < \langle n \rangle_\lambda$. Since the analysis can be ascribed to case (ii), we obtain $P_w(\tau) \sim \tau^{-(2\gamma-3)/(\gamma-1)}$ for $2 < \gamma \leq 3$ and $P_w(\tau) \sim \tau^{-3/2}$ for $\gamma > 3$.

VI. DISCUSSION AND SUMMARY

The analytic results are summarized in Table I and confirmed numerically in Fig. 3. The power-law behavior of the

waiting-time distribution $P_w(\tau) \sim \tau^{-\alpha}$ can be diverse in that α can take general values, rather than $\alpha = 1$ or $3/2$. Consistent with this, the intertransaction time of a stock broker obeys the power-law distribution with $\alpha \approx 1.3$ with an exponential cutoff [3].

Our results are compatible with those derived from the continuous time approximation [8] and the fractional derivative [19]. The generating function approach that we have developed can be useful for studying further problems. For example, we show in the Appendix that our approach considered in the limit $\lambda, \mu \to 0$ reproduces the results for the GL model [8]. Furthermore, GL as well as we are successful in deriving the exponential cutoff for $\lambda < \mu$ as $P_w(\tau) \sim \tau^{-3/2} e^{-\tau/\tau_0}$ with $\tau_0 = 1/(\sqrt{\mu} - \sqrt{\lambda})^2$. However, for the model with general distributions of the number of incoming tasks, the explicit form of the exponential correction factor is not obvious.

In our priority queue model, the jump distance of the equivalent random walk is unbounded to the right, whereas it is at most one to the left. In real queue systems, however, more than one tasks may be executed in a unit time. Therefore, a natural extension of our model is to allow the number of executed tasks in a unit time to exceed one. To be specific, in addition to the heterogeneity of the number of incoming tasks, i.e., n tasks are incoming with probability $\lambda_n \sim n^{-\gamma_{in}}$ per unit time, we can suppose that ℓ tasks are executed with probability $\mu_\ell \sim \ell^{-\gamma_{out}}$. Our numerical results for $P_w(\tau)$ seem to fit the formulas shown in Table I, with the exponent γ replaced by the minimum of γ_{in} and γ_{out}, as far as both γ_{in} and γ_{out} are larger than 2 (not shown). This suggests that the dominant tail determines the behavior of the waiting-time distribution in the priority queue system. In particular, when the distribution of the number of executed tasks is neither binary nor heavy-tailed (e.g., purely exponential), which may be true for many real queues, our results hold because $\gamma = \gamma_{in}$.

ACKNOWLEDGMENTS

APPENDIX: COMPARISON OF THE GRINSTEIN-LINSKER SOLUTION AND THE GENERATING-FUNCTION SOLUTION

GL analyzed a priority queue model in which, in a unit time, a new task arrives with probability λ and the task with the highest priority in the queue is executed with probability μ, which corresponds to $\lambda_0 = 1 - \lambda$, $\lambda_1 = \lambda$, and $\lambda_n = 0$ for $n \geq 2$ in our model [8]. They obtained the solution of the waiting-time distribution by analyzing the continuous-time dynamics. We compare the GL solution and the solution derived via the generating function in the GL limit.

1. Queue-length distribution

The generating function of the queue-length distribution in the steady state is given in Eq. (15) in the main text. By substituting $\Lambda(z) = 1 - \lambda + \lambda z$ and $\langle n \rangle_\lambda = \lambda$ into Eq. (15), we obtain

$$\widetilde{\mathcal{Q}}_x(z) = \frac{\mu - \lambda(1-x)}{\lambda(\mu-1)(1-x)z + (1-\lambda+\lambda x)\mu}, \qquad (A1)$$

which leads to

$$\widetilde{Q}_x(m) = \frac{\mu - \lambda(1-x)}{(1-\lambda+\lambda x)\mu}\left(\frac{\lambda(1-\mu)(1-x)}{(1-\lambda+\lambda x)\mu}\right)^m. \qquad (A2)$$

Using the continuous-time approach, GL derived

$$\widetilde{Q}_x(m) = \frac{\mu - \lambda(1-x)}{\mu}\left(\frac{\lambda(1-x)}{\mu}\right)^m. \qquad (A3)$$

Equations (A2) and (A3) are consistent in the limit $\lambda, \mu \rightarrow 0$.

2. Waiting-time distribution and the exponential cutoff

To obtain the waiting-time distribution of a task, we use the following theorem [20–22]:

Theorem. Suppose that, for real numbers s^* and $\mathcal{F}^*$, a power series $\mathcal{F}(s) = \sum_{t=1}^\infty a(t)s^t$ with nonnegative coefficients $a(1), a(2), \ldots$ satisfies the following equations (A4)–(A6):

$$F(s, \mathcal{F}) \text{ is analytic near } (s, \mathcal{F}) = (s^*, \mathcal{F}^*); \qquad (A4)$$

if $|s| \leq s^*, |\mathcal{F}| \leq \mathcal{F}^*$,

$$F(s, \mathcal{F}) = \frac{\partial F(s, \mathcal{F})}{\partial \mathcal{F}} = 0 \text{ if and only if } (s, \mathcal{F}) = (s^*, \mathcal{F}^*); \qquad (A5)$$

$$\frac{\partial F(s^*, \mathcal{F}^*)}{\partial s} \neq 0, \quad \frac{\partial^2 F(s^*, \mathcal{F}^*)}{\partial \mathcal{F}^2} \neq 0. \qquad (A6)$$

Then,

$$a(t) \approx \left(\frac{s^*\dfrac{\partial F(s^*, \mathcal{F}^*)}{\partial s}}{2\pi\dfrac{\partial^2 F(s^*, \mathcal{F}^*)}{\partial \mathcal{F}^2}}\right)^{1/2} t^{-3/2}s^{*-t}, \quad t \rightarrow \infty. \quad (A7)$$

To apply this theorem to the GL queue model, we define

$$F(s, \mathcal{F}) = s[(1-\mu)\mathcal{F} + \mu]\{1 - \lambda + \lambda[(1-x)\mathcal{F} + x]\} - \mathcal{F}, \qquad (A8)$$

so that $F(s, \mathcal{F}_x(s)) = 0$ holds, where $\mathcal{F}_x(s) = \sum_{t=1}^\infty f_x(t)s^t$ is the generating function of the first-passage time probability. Then the other main condition of the theorem [see Eq. (A5)] reads

$$\frac{\partial F(s, \mathcal{F})}{\partial \mathcal{F}} = s(1-\mu)\{1 - \lambda + \lambda[(1-x)\mathcal{F} + x]\}$$

$$+ s[(1-\mu)\mathcal{F} + \mu]\lambda(1-x) - 1 = 0. \quad (A9)$$

The solution to Eqs. (A8) and (A9) with the minimum absolute values is given by

$$s^* = \frac{1}{1 - \lambda - \mu + 2\lambda\mu + \lambda x - 2\lambda\mu x + 2\sqrt{\lambda(1-\lambda+\lambda x)(1-x)\mu(1-\mu)}}, \qquad (A10)$$

$$\mathcal{F}^* = \sqrt{\frac{\mu(1-\lambda+\lambda x)}{\lambda(1-\mu)(1-x)}}. \qquad (A11)$$

The rest of the conditions of the theorem are satisfied with s^* and $\mathcal{F}^*$ given by Eqs. (A10) and (A11). Equation (A7) implies that the tail of the first-passage time probability decays as $f(t) \propto t^{-3/2}s^{*-t}$. This asymptotic is also derived by

directly calculating $\mathcal{F}_x(s) \sim (s^* - s)^{1/2}$ as $s \uparrow s^*$ and using the Tauberian theorem [15,23,24].

The generating function of the waiting-time distribution of a task is equal to that of the queue-length distribution given by Eq. (15) with s replaced by $\mathcal{F}_x(s)$. To calculate the asymptotic of the waiting-time distribution, we erase Λ by combining Eq. (15) with s replaced by $\mathcal{F}_x(s)$ and Eq. (18),

which yields

$$\widetilde{\mathcal{Q}}_x[\mathcal{F}_x(s)] = \frac{A_1 s[1 - \mathcal{F}_x(s)]}{(1-s)\mathcal{F}_x(s)}. \tag{A12}$$

Inserting Eq. (A12) into Eq. (A8) results in

$$(1 - \lambda + \lambda x)\mu(1-s)\widetilde{\mathcal{Q}}_x^2(\mathcal{F}_x(s))$$
$$+ A_1[(1 - \lambda + \lambda x + \mu)s - 1]\widetilde{\mathcal{Q}}_x(\mathcal{F}_x(s)) - A_1^2 s = 0. \tag{A13}$$

Applying the theorem (A4)–(A7) with $\mathcal{F} \equiv \widetilde{\mathcal{Q}}_x(\mathcal{F}_x(s))$ leads to the same equation (A10). Therefore, the waiting-time distribution has the same asymptotic as the first-passage time probability, that is, $P_w(\tau) \sim \tau^{-3/2} s^{*-\tau}$. This asymptotic is also derived by solving Eq. (A13) as $\mathcal{P}_w(s) \sim (s^* - s)^{1/2}$ as $s \uparrow s^*$.

To evaluate s^*, we denote the denominator of the RHS of Eq. (A10) by $H(x)$, with λ and μ fixed. The existence of the exponential cutoff in the first-passage time and the waiting-time distribution is equivalent to $H(x) < 1$ ($0 \leq \forall\, x \leq 1$).

A straightforward calculation yields $d^2H/dx^2 < 0$, $\lim_{x \uparrow 1} dH/dx = -\infty$, and that $dH/dx = 0$ has a unique solution $x = (\lambda - \mu)/\lambda$. As explained in the main text and in previous literature [8], the analysis of case $\lambda > \mu$ is ascribed to that of case $\lambda = \mu$. Therefore, we assume $\lambda \leq \mu$ and obtain $dH/dx < 0$ ($0 < x \leq 1$). Then the maximum of $H(x)$ is realized at $x = 0$, so that the smallest s^* is equal to

$$s^* = \frac{1}{H(0)} = \frac{1}{1 - \lambda - \mu + 2\lambda\mu + 2\sqrt{\lambda(1-\lambda)\mu(1-\mu)}}. \tag{A14}$$

When $\mu = \lambda$, we obtain $s^* = 1$. The asymptotic of the waiting-time distribution is $P_w(\tau) \sim \tau^{-3/2}$, which is consistent with the results in [8] and coincides with our results for $\gamma > 3$.

When $\mu > \lambda$, we obtain $s^* > 1$ and $P_w(\tau) \sim \tau^{-3/2} e^{-\tau/\tau_0}$, where $\tau_0 = 1/\ln s^*$. In the limit $\lambda, \mu \to 0$, our discrete-time model tends to GL's continuous-time queue dynamics. By inserting $\lambda = \lambda' \Delta \tau$, $\mu = \mu' \Delta \tau$, and $\tau = \tau'/\Delta\tau$ into Eq. (A14) and letting $\Delta\tau \to 0$, we obtain $P_w(\tau) \sim \tau'^{-3/2} e^{-\tau'/\tau_0}$, where $\tau_0 = 1/(\sqrt{\mu} - \sqrt{\lambda})^2$. The predicted τ_0 agrees with the one derived by GL. They concluded $P_w(\tau) \sim \tau'^{-3/2} e^{-\tau'/\tau_0}$ for $\tau \ll \tau_0$ and $P_w(\tau) \sim \tau'^{-5/2} e^{-\tau'/\tau_0}$ for $\tau \gg \tau_0$. Our results only reproduce the asymptotic on the intermediate timescale (i.e., $\tau \ll \tau_0$) because τ_0 diverges as $\lambda, \mu \to 0$.

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

[2] A. Vázquez, Phys. Rev. Lett. **95**, 248701 (2005).

[3] A. Vázquez, J. G. Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási, Phys. Rev. E **73**, 036127 (2006).

[4] K.-I. Goh and A.-L. Barabási, Europhys. Lett. **81**, 48002 (2008).

[5] Z. Dezsö, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A.-L. Barabási, Phys. Rev. E **73**, 066132 (2006).

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

[7] A. Gabrielli and G. Caldarelli, Phys. Rev. Lett. **98**, 208701 (2007).

[8] G. Grinstein and R. Linsker, Phys. Rev. Lett. **97**, 130201 (2006); Phys. Rev. E **77**, 012101 (2008).

[9] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14333 (2004).

[10] C. Dewes, A. Wichmann, and A. Feldmann, in *Proceedings of the 2003 ACM SIGCOMM Conference on Internet Measurement (IMC'03)* (ACM Press, New York, 2003).

[11] V. Paxson and S. Floyd, IEEE/ACM Trans. Netw. **3**, 226 (1995).

[12] S. D. Kleban and S. H. Clearwater, in *Proceedings of the 2003 ACM/IEEE Conference on Supercomputing (SC2003)* (IEEE, Washington, 2003).

[13] U. Harder and M. Paczuski, Physica A **361**, 329 (2006).

[14] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation* (Wiley, New York, 2000).

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

[16] H. K. Lee, K.-I. Goh, B. Kahng, and D. Kim, Int. J. Bifurcation Chaos Appl. Sci. Eng. **17**, 2485 (2007).

[17] J. de Boer, B. Derrida, H. Flyvbjerg, A. D. Jackson, and T. Wettig, Phys. Rev. Lett. **73**, 906 (1994); J. de Boer, A. D. Jackson, and T. Wettig, Phys. Rev. E **51**, 1059 (1995).

[18] N. Masuda, K.-I. Goh, and B. Kahng, Phys. Rev. E **72**, 066106 (2005).

[19] A. V. Chechkin, R. Metzler, V. Y. Gonchar, J. Klafter, and L. V. Tanatarov, J. Phys. A **36**, L537 (2003).

[20] E. A. Bender, SIAM Rev. **16**, 485 (1974).

[21] A. M. Odlyzko, In *Handbook of Combinatorics*, edited by R. L. Graham, M. Groetschel, and L. Lovasz (Elsevier, Amsterdam, 1995), Vol. 2, pp. 1063–1229.

[22] D. A. Klarner and P. Woodworth, Aequ. Math. **23**, 236 (1981).

[23] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. (Wiley, New York, 1966), Vol. 2.

[24] G. H. Weiss, *Aspects and Applications of the Random Walk* (North-Holland, Amsterdam, 1994).