The Annals of Applied Probability
1999, Vol. 9, No. 2, 430–464

ASYMPTOTIC APPROXIMATION OF THE MOVE-TO-FRONT SEARCH COST DISTRIBUTION AND LEAST-RECENTLY USED CACHING FAULT PROBABILITIES

BY PREDRAG R. JELENKOVIĆ

Lucent Technologies

Consider a finite list of items $n = 1, 2, \ldots, N$, that are requested according to an i.i.d. process. Each time an item is requested it is moved to the front of the list. The associated search cost C^N for accessing an item is equal to its position before being moved. If the request distribution converges to a proper distribution as $N \to \infty$, then the stationary search cost C^N converges in distribution to a limiting search cost C.

We show that, when the (limiting) request distribution has a heavy tail (e.g., generalized Zipf's law), $\mathbb{P}[R = n] \sim c/n^\alpha$ as $n \to \infty$, $\alpha > 1$, then the limiting stationary search cost distribution $\mathbb{P}[C > n]$, or, equivalently, the least-recently used (LRU) caching fault probability, satisfies

$$\lim_{n\to\infty} \frac{\mathbb{P}[C > n]}{\mathbb{P}[R > n]} = \left(1 - \frac{1}{\alpha}\right)\left[\Gamma\left(1 - \frac{1}{\alpha}\right)\right]^\alpha \nearrow e^\gamma \quad \text{as } \alpha \to \infty,$$

where Γ is the Gamma function and γ ($= 0.5772\ldots$) is Euler's constant.

When the request distribution has a light tail $\mathbb{P}[R = n] \sim c \exp(-\lambda n^\beta)$ as $n \to \infty$ ($c, \lambda, \beta > 0$), then

$$\lim_{n\to\infty} \frac{\mathbb{P}[C_f > n]}{\mathbb{P}[R > n]} = e^\gamma,$$

independently of c, λ, β, where C_f is a fluid approximation of C.

We experimentally demonstrate that the derived asymptotic formulas yield accurate results for lists of finite sizes. This should be contrasted with the exponential computational complexity of Burville and Kingman's exact expression for finite lists. The results also imply that the fault probability of LRU caching is asymptotically at most a factor e^γ (≈ 1.78) greater than for the optimal static arrangement.

1. Introduction. One of the most commonly encountered problems in modern distributed network environment is efficient information retrieval (e.g., the Internet Web searching). As a solution to this problem, an entire spectrum of different heuristic dynamically organizing data structures have been proposed. Among the proposed algorithms, the most basic ones are the *move-to-front* (MTF) self-organizing searching algorithm and the corresponding *least-recently used* (LRU) caching scheme. The main objective of this paper is to obtain an analytic asymptotic characterization of the MTF search cost distribution function or, equivalently, the LRU caching fault probabilities.

Received May 1998; revised August 1998.

AMS 1991 *subject classifications*. Primary 60F99; secondary 68P05, 68P10.

Key words and phrases. Self-organizing lists, move-to-front searching, least-recently-used caching (paging), heavy-tailed distributions, light-tailed distributions.

As sketched in the abstract, the MTF algorithm can be informally described as follows. Assume that there is a finite linear list of items (say $1, 2, 3, \dots, N$, sequentially ordered from first to last) and a sequence of requests for the items of the list. Each time a requested item is found at the nth position in the list, it is brought to the first position and items in positions $1, \dots, n-1$ are moved one position down. One performance measure is the search cost function, which is defined to be the position of the requested item. The caching scheme that corresponds to MTF is the LRU algorithm. For this scheme, it is assumed that n items are kept in fast memory (cache) and that the remaining $N-n$ items are stored in slow memory. Each time a request for an item is made, fast memory is searched first. If the item is not found there it is brought from slow memory and replaced with the least recently used item in the cache. The performance quantity of interest for this algorithm is the LRU fault probability, that is, the probability that the requested item is not in the cache. It can be shown that computing the LRU fault probability is the same as computing the MTF search cost distribution (the details of this connection will be discussed in Section 2.1).

The performance analysis of self-organizing data structures (e.g., lists, trees) has a long history. Basic references can be found in [22] and [11]. In the analysis of self-organizing lists there have been two approaches: combinatorial and probabilistic analysis. For the combinatorial (amortized, competitive) analysis the reader is referred to [2] and [28]. Recent results and references for this approach can be found in [7] and [19]. In this paper we will concentrate on probabilistic analysis.

Early work on the probabilistic analysis of the MTF rule dates back to McCabe [23]. He computed the expected value and variance of the cost function for finite lists. In [18], a Markov chain on the state space of all permutations on the elements of the list is analyzed and the stationary distribution was derived. Rivest [27] showed that in stationarity the transposition rule (search algorithm in which the requested item is moved only one position closer to the front) is more efficient (in a certain sense) than the MTF rule. Bitner [5] investigated the transient behavior of the expected cost function. The n-step transition probabilities for the underlying Markov chain are derived in [24] (see also [26, 12]). Spectral analysis is conducted in [25] (see also [12]). A combinatorial expression for the distribution function of the search cost was first derived by Burville and Kingman [8]. An integral form of the Laplace transform of the search cost distribution function is computed in [15]. The authors derive this result using combinatorial techniques and formal languages. The same result was rederived in [13] using a Poisson embedding technique. A comprehensive list of references on the probabilistic analysis of the LRU and MTF algorithms can be found in [6].

The principal contribution of this paper is the complete asymptotic characterization, both for the light- and heavy-tailed case (as defined in the abstract), of the MTF search cost distribution and the LRU cache fault probabilities. One of the main mathematical techniques used for the heavy-tailed case is the Hardy–Littlewood–Karamata Tauberian theorem for an asymptotic inversion of the Laplace transform. In addition, we develop a novel fluid limit approach for analyzing self-organizing data structures. This approach, combined with a direct Laplace inversion, yields a characterization of the search cost distribution in the case of light tails. The practical implication of these results is that, for a large class of distributions considered in this paper, the LRU caching scheme is asymptotically only $e^{\gamma} \approx 1.78$ times worse than the optimal static arrangement.

The remainder of this paper is organized as follows. Section 2 formally defines the problem and gives a short technical note on existing results. A summary of the main results from the literature on the stationary distribution of the search cost distribution is given in Section 2.1. Optimality of MTF and LRU algorithms is discussed in Section 2.2. The main asymptotic results, Theorems 3 and 6, are presented in Section 3 and 4, respectively. Extensive simulation experiments that verify the accuracy of the asymptotic approximation formulas are presented in Section 5. The paper concludes in Section 6. To simplify the reading process, the majority of the technical proofs are given in Section 7.

2. Problem definition and historical notes. Consider a finite set of items $L = \{1, \dots, N\}$, and a sequence of i.i.d. requests $\{R_t^N,\ t = 0, 1, \dots\}$ distributed as $q_r = \mathbb{P}[R_t^N = r],\ 1 \le r \le N$. The dynamics of the MTF algorithm is described as follows. If at time t the requested item r ($R_t^N = r$) is at the nth position of the list, then it is brought to the first position and items in positions $1, \dots, n-1$ are moved one position down. The process of list updates can be modeled as a Markov chain $\{\sigma_t^N,\ t \ge 0\}$ on the state space of all list permutations. A search cost process $\{C_t^N,\ t \ge 0\}$ is defined such that C_t^N represents the position in the list of the item requested at time t. The notation R^N, C^N, σ^N will be used to denote the random variables that have the same distribution as the stationary distribution of R_t^N, C_t^N, σ_t^N, respectively. Our main objective is to derive a simple asymptotic characterization of the search cost distribution $\mathbb{P}[C^N > n]$.

McCabe [23] derived the following formula for the expected value of the search cost distribution:

$$\mathbb{E}C^N = 1 + 2\sum_{r<k} \frac{q_r q_k}{q_r + q_k}. \tag{2.1}$$

Bitner [5] obtained the transient expected cost

$$\mathbb{E}C_t^N = 1 + 2\sum_{r<k} \frac{q_r q_k}{q_r + q_k} + \sum_{r<k} \frac{(q_r - q_k)^2 (1 - q_r - q_k)^t}{2(q_r + q_k)}. \tag{2.2}$$

The stationary distribution for σ_t^N was first computed by Hendricks [18]:

$$\begin{aligned} \mathbb{P}[\sigma^N = x] = q_{x(1)} &\frac{q_{x(2)}}{1 - q_{x(1)}} \frac{q_{x(3)}}{1 - q_{x(1)} - q_{x(2)}} \\ &\cdots \frac{q_{x(N)}}{1 - q_{x(1)} - q_{x(2)} - \cdots - q_{x(N-1)}}, \end{aligned} \tag{2.3}$$

where x is a particular permutation (state) of the list, and $x(i)$ is the item at the ith position in the permutation. Spectral analysis of the Markov chain $\{\sigma_t^N,\ t \geq 0\}$ is performed in Phatarfod [25], where it was shown that the eigenvalues of the transition matrix of this Markov chain are the $2^N - N$ distinct numbers $\sum_{i \in S} q_i$, $S \subseteq \{1, \dots, N\}$, with $|S| \neq N-1$. In the same paper Phatarfod derived the multiplicities of eigenvalues (see also [12]).

2.1. *Stationary distribution of the search cost and the fault probability of the least recently used caching.* This subsection presents the results from the literature on the stationary distribution of the MTF search cost and the fault probability for the LRU caching. A formal connection between these two quantities is well known (e.g., see [15, 13]). Arguments in support of this connection are also supplied within this subsection.

Burville and Kingman [8] derived the following combinatorial formula:

$$\mathbb{P}[C^N = n] = \sum_{r=1}^{N} \sum_{a=0}^{n-1} (-1)^{n-1-a} \binom{N-1-a}{n-1-a} \sum_{A:\,|A|=a,\, r \notin A} \frac{q_r^2}{(1-Q_A)}, \tag{2.4}$$

where $Q_A = \sum_{r \in A} q_r$ as defined. The connection between LRU caching and MTF searching can be demonstrated as follows. Denote by $D(k, N)$ the fault probability in the LRU caching scheme with the cache size k. Then we claim

$$D(k, N) = \mathbb{P}[C^N > k].$$

Here is a simple argument that justifies this claim. We can imagine that k elements in the cache are arranged in increasing order of their last access times. Each time there is a request for an item that is not in the cache, the item is brought to the cache and the last element of the cache is moved to the slow memory. The claim is that the fault probability $D(k, N)$ stays the same if the remaining $N-k$ items in the slow memory are arranged in any specific order. In particular, they can be arranged in increasing order of their last access times. It is clear that the obtained algorithm is the same as the MTF algorithm and that $D(k, N) = \mathbb{P}[C^N > k]$. For those who are still not convinced by the preceding argument, one can obtain the expression for $D(k, N)$ directly from (2.4) as follows. Compute $\mathbb{P}[C^N \leq k] = \sum_{n=1}^{k} \mathbb{P}[C^N = n]$, by using (2.4), interchange the sums with respect to n and a, and use the identity (which can be proved easily by induction on k)

$$\sum_{n=a+1}^{k} (-1)^{n-1-a} \binom{N-1-a}{n-1-a} = (-1)^{k-1-a} \binom{N-2-a}{k-1-a}.$$

This derivation leads to Corollary 5.2 from [15], which, for convenience, is stated here:

$$\begin{aligned} 1 - D(k, N) &= \mathbb{P}[C^N \leq k] \\ &= \sum_{r=1}^{N} \sum_{a=0}^{k-1} (-1)^{k-1-a} \binom{N-2-a}{N-1-k} \sum_{A:\,|A|=a,\, r \notin A} \frac{q_r^2}{(1-Q_A)}. \end{aligned} \tag{2.5}$$

Unfortunately, except for relatively small N, n and a, the preceding formulas (2.4) and (2.5) are not suitable for numerical evaluation. This is due to a combinatorial explosion; as pointed out in [15] the evaluation of $\mathbb{P}[C^N > k]$ takes about $N^k/k!$ operations, which for example for $N = 1000$ and $k = 20$ computes to roughly 10^{40} operations. This is clearly infeasible. Furthermore, when $k = bN$, $0 < b \le 1$, application of Stirling's formula shows that $N^{bN}/(bN)!$ grows exponentially in N; that is, (2.5) [or equivalently (2.4)] has exponential complexity.

To alleviate this problem, in [15] a compact integral representation of the Laplace transform of the search cost distribution function is derived. In the same paper Cauchy contour integration was proposed for efficient inversion of the search cost distribution function. Fill [13] rederived the same result using the Poisson embedding technique. This result reads as

$$\begin{aligned}\mathbb{E}\exp(-sC^N) = e^{-s}\int_{t=0}^{\infty} e^{-t}\left[\sum_{r=1}^{N}\frac{q_r^2}{1+e^{-s}(\exp(q_r t)-1)}\right] \\ \times\left[\prod_{r=1}^{N}(1+e^{-s}(\exp(tq_r)-1))\right]dt,\end{aligned}$$

for any $s > 0$.

In [14] the limiting search cost C as $N \to \infty$ is investigated. In order to state this result, choose a probability distribution sequence $\mathbb{P}[R = r] = q_r$, $1 \le r < \infty$, $\sum_{r=1}^{\infty} q_r = 1$. Next, construct a sequence of MTF algorithms with finite number of elements N, whose request probabilities are given as $\mathbb{P}[R_t^N = r] = q_r/q_N^+$, $1 \le r \le N$, where $q_N^+ = \sum_{r=1}^{N} q_r$. Then Fill obtained the following result ([13], Proposition 4.4; in the same paper, he also considered the case when $q_N^+ \to \infty$ as $N \to \infty$ and showed that appropriately scaled C^N converges to a proper limit).

THEOREM 1. *The sequence of search costs C^N converges in distribution to C, as $N \to \infty$, and the Laplace transform of C is given as*

$$\begin{aligned}\mathbb{E}e^{-sC} = e^{-s}\int_{t=0}^{\infty}\sum_{i=1}^{\infty} q_i^2\exp(-q_i t) \\ \times\left[\prod_{r:\, r\neq i}(1-(1-e^{-s})(1-\exp(-q_r t)))\right]dt.\end{aligned} \tag{2.6}$$

This result is the basis of our further investigation.

2.2. *Optimality of MTF and LRU algorithms.* Before starting with the analysis, let us give a few known results about the optimality of the MTF and LRU algorithms. Note that without loss of generality we can assume that the request probabilities q_r form a monotonically nonincreasing sequence (if this is not the case we can always relabel the items in such a way that the new sequence q_r is nonincreasing). Now, it is clear that the optimal algorithm is the one that keeps the items static and ordered in decreasing order of their request probabilities. The search cost of this optimal algorithm is $\mathbb{E}R^N$. Next, by using $q_r/(q_r + q_k) < 1$ in (2.1) we easily derive an upper bound (see [22], page 399)

$$\mathbb{E}R^N \leq \mathbb{E}C^N \leq 2\mathbb{E}R^N.$$

Using Hilbert's inequalities in [9], the upper bound was improved from 2 to $\pi/2$; $\pi/2$ was also shown to be the best possible bound.

Similarly, for the caching system with a cache size n, the optimal algorithm keeps n most frequent items in the cache all of the time. Now, $\mathbb{P}[C^N > n]/\mathbb{P}[R^N > n]$ gives the ratio between the LRU fault probability and the fault probability for the optimal static arrangement. A bound on this quantity is obtained in [16], Theorem 3.

THEOREM 2.

$$\frac{\mathbb{P}[C^N > n]}{\mathbb{P}[R^N > n]} \leq 1 + n\frac{\mathbb{P}[R^N \leq n]}{1 + (n-1)\mathbb{P}[R^N > n]} = b(n, N)$$

as defined for arbitrary $q_1, \dots, q_N$.

At this point, and for the rest of the paper, we introduce the following customary notation. For any two real functions $a(t)$ and $b(t)$ and fixed $t_0 \in \mathbb{R} \cup \{\infty\}$, we will use $a(t) \sim b(t)$ as $t \to t_0$ to denote $\lim_{t\to t_0} a(t)/b(t) = 1$. Similarly, we say that $a(t) \gtrsim b(t)$ as $t \to t_0$ if $\liminf_{t\to t_0} a(t)/g(t) \geq 1$; $a(t) \lesssim b(t)$ has a complementary definition.

For the case when $\mathbb{P}[R^N \leq n]$ converges to a proper distribution $\mathbb{P}[R \leq n]$ as $N \to \infty$, the limit of the upper bound $b(n) = \lim_{N\to\infty} b(n, N)$, as defined, in Theorem 2 is equivalent to the following. If $\mathbb{E}R < \infty$, then

$$b(n) \sim n \quad \text{as } n \to \infty. \tag{2.7}$$

If $\mathbb{P}[R = n] \sim c/n^\alpha$, $1 < \alpha < 2$, then

$$b(n) \sim \frac{n^{2-\alpha}}{c} \quad \text{as } n \to \infty. \tag{2.8}$$

Under the additional tail conditions on $\mathbb{P}[R = n]$, our main results, Theorems 3 and 6, will show that the bounds in (2.7) and (2.8) can be replaced by a constant, namely, $e^\gamma \approx 1.78$, where γ is Euler's constant.

3. Heavy tails. This section presents a straightforward asymptotic characterization of the search cost distribution function $\mathbb{P}[C > n]$ for the case when the request distribution has a heavy (polynomial) tail. The main result is stated in Theorem 3. The primary technique that is used is Karamata's Tauberian–Abelian theory for the asymptotic inversion of the Laplace transforms.

In order to be able to obtain the asymptotic inversion of (2.6) we need to derive a simplified asymptotic representation of the infinite sum and infinite products that appear in (2.6). To this end we define the following density function

$$f(t) = \sum_{r=1}^{\infty} q_r^2 \exp(-q_r t) \quad \text{as defined,} \tag{3.1}$$

whose asymptotic behavior is described in the subsequent lemma.

LEMMA 1. *Assume that* $q_r \sim c/r^\alpha$ *as* $r \to \infty$, *with* $\alpha > 1$ *and* $c > 0$. *Then*

$$f(t) \sim \frac{c^{1/\alpha}}{\alpha} \Gamma\left(2 - \frac{1}{\alpha}\right) t^{-2+1/\alpha} \quad \text{as } t \to \infty,$$

where Γ *is the Gamma function.*

The proof is given in Section 7.1.

Our next object of investigation is

$$g(t) = \sum_{r=1}^{\infty} (1 - \exp(-q_r t)) \quad \text{as defined.} \tag{3.2}$$

Here $g(t)$ also has a straightforward asymptotic characterization.

LEMMA 2. *Assume that* $q_r \sim c/r^\alpha$ *as* $r \to \infty$, *with* $\alpha > 1$ *and* $c > 0$. *Then*

$$g(t) \sim \Gamma\left(1 - \frac{1}{\alpha}\right) c^{1/\alpha} t^{1/\alpha} \quad \text{as } t \to \infty.$$

The proof is given in Section 7.1.

Actually, a stronger version of Lemma 2 holds. Let A be any set of indexes $A \subset \mathbb{N}$ with cardinality $|A|$. Denote with $g_A(t) = \sum_{r:\, r \notin A} (1 - \exp(-q_r t))$.

COROLLARY 1. *Assume that* $q_r \sim c/r^\alpha$ *as* $r \to \infty$, *with* $\alpha > 1$ *and* $c > 0$. *Then for any fixed finite* ℓ,

$$g_A(t) \sim \Gamma\left(1 - \frac{1}{\alpha}\right) c^{1/\alpha} t^{1/\alpha} \quad \text{as } t \to \infty,$$

uniformly in all $A \subset \mathbb{N}$ *such that* $|A| \le \ell$.

PROOF. The proof follows immediately from Lemma 2 and

$$g(t) - \ell \le g_A(t) \le g(t). \qquad \square$$

The preceding technical results led up to the following main result of this section.

THEOREM 3. *Assume that* $q_r \sim c/r^\alpha$ *as* $r \to \infty$, *with* $\alpha > 1$ *and* $c > 0$. *Then*

$$\mathbb{P}[C > n] \sim \left(1 - \frac{1}{\alpha}\right)\left[\Gamma\left(1 - \frac{1}{\alpha}\right)\right]^\alpha \mathbb{P}[R > n] \quad \textit{as } n \to \infty. \tag{3.3}$$

Furthermore, if we denote the constant of proportionality in (3.3) *as* $K(\alpha)$, *then* $K(\alpha)$ *is monotonically increasing with*

$$\lim_{\alpha\to\infty} K(\alpha) = e^\gamma \approx 1.78107, \qquad \lim_{\alpha\downarrow 1} K(\alpha) = 1, \tag{3.4}$$

where γ *is the Euler constant, that is,*

$$\gamma = \lim_{n\to\infty}\left(\sum_{i=1}^{n} i^{-1} - \log n\right) = 0.5772156649\ldots.$$

REMARK. This theorem implies that

$$\mathbb{P}[C > n] \lesssim e^\gamma \mathbb{P}[R > n] \quad \text{as } n \to \infty. \tag{3.5}$$

A plot of $K(\alpha)$ is given in Figure 1.

A rigorous proof of Theorem 3 is given in Section 7.2.

HEURISTIC SKETCH OF THE PROOF. First, for all sufficiently small s, and all sufficiently large t,

$$\prod_{r: r\neq i} (1 - (1 - e^{-s})(1 - \exp(-q_r t))) \approx \exp\left(-s \sum_{r=1}^{\infty}(1 - \exp(-q_r t))\right). \tag{3.6}$$

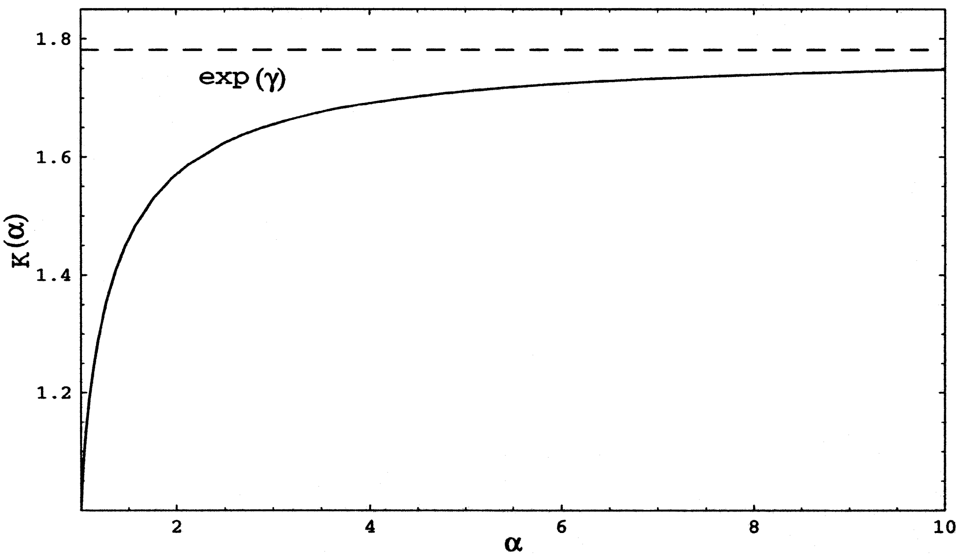

FIG. 1. *Function* $K(\alpha)$.

Now, by utilizing the above approximation and Lemmas 1 and 2, one can obtain the following informal approximation for sufficiently small s:

$$\begin{aligned} &\int_{t_0}^{\infty} \sum_{i=1}^{\infty} q_i^2 \exp(-q_i t)\Bigg[\prod_{r:\, r\neq i} (1-(1-e^{-s})(1-\exp(-q_r t)))\Bigg]\,dt \\ &\qquad \approx \int_{t_0}^{\infty} \frac{c^{1/\alpha}}{\alpha}\Gamma\bigg(2-\frac{1}{\alpha}\bigg) t^{-2+(1/\alpha)} \exp\bigg(-s\Gamma\bigg(1-\frac{1}{\alpha}\bigg)c^{1/\alpha}t^{1/\alpha}\bigg)\,dt. \end{aligned} \tag{3.7}$$

By changing the variable of integration to $u = \Gamma(1-1/\alpha)c^{1/\alpha}t^{1/\alpha}$ in the integral above, we conclude that the integral in (3.7) is approximately equal to

$$\bigg(1-\frac{1}{\alpha}\bigg)\bigg[\Gamma\bigg(1-\frac{1}{\alpha}\bigg)\bigg]^{\alpha} \int_{u_0}^{\infty} \frac{c}{u^{\alpha}} e^{-su}\,du,$$

where $u_0 = \Gamma(1-1/\alpha)c^{1/\alpha}t_0^{1/\alpha}$. From this, the observation that

$$\int_{u_0}^{\infty} \frac{c}{u^{\alpha}} e^{-su}\,du \approx \mathbb{E}(e^{-sR}1[R > u_0]),$$

and the uniqueness of the inverse of the Laplace transform, we can roughly conclude that the tail of C should be proportional to the tail of R with a constant of proportionality given in the equation above. Unfortunately, to make these arguments rigorous, much lengthier analysis is required. In particular, one has to investigate the asymptotic behavior of the derivatives of $\mathbb{E}e^{-sC}$. A complete proof that utilizes Karamata's Tauberian–Abelian theorem is provided in Section 7.2. □

4. Fluid limit and light tails. In Theorem 4 in this section we develop a fluid limit approximation of the search cost function. For the heavy-tailed case the validity of this approximation is demonstrated in Theorem 5. The main result is presented in Theorem 6. This theorem shows that for the light-tailed case the ratio between the tail of the search cost distribution and the request distribution is asymptotically *invariant* with respect to the shape of the request distribution function.

Consider a density function q on $[0,\infty)$, and the corresponding distribution function $Q(t) = \int_0^t q(u)\,du$. Assume that the request probabilities are given as $q_r = Q(r) - Q(r-1)$, $r \geq 1$. Now, construct a sequence of MTF algorithms with request probabilities

$$q_r^k = Q(r/k) - Q((r-1)/k), \qquad k, r \geq 1.$$

For each of the constructed MTF schemes, let $C(k)$ be the stationary search cost random variable. Informally, the kth MTF scheme is constructed from the original one by dividing each item r into k items with request probabilities $q_{(r-1)k+i}^k$, $1 \leq i \leq k$. In order to compare the derived schemes with the original one we will scale the search cost of the kth scheme as $C(k)/k$. Now we show that $C(k)/k$ converges in distribution to a proper (fluid) limit as $k \to \infty$. Assume that q is monotonically decreasing and continuous. [Note that this assumption is not restrictive; i.e., for any given monotonically decreasing sequence of request probabilities we can always choose a continuous function q such that $q_r = Q(r) - Q(r-1)$, $r \geq 1$.]

THEOREM 4. *The sequence $C(k)/k$ converges in distribution to the fluid limit C_f (finite with probability 1), as $k \to \infty$, whose Laplace transform is given as*

$$\mathbb{E}\exp(-sC_f) = \int_0^\infty \left(\int_0^\infty q^2(u)e^{-q(u)t}\,du\right) \times \exp\left(-s\int_0^\infty (1-e^{-q(u)t})\,du\right)dt, \qquad s > 0. \tag{4.1}$$

The proof is given in Section 7.3.

We term C_f the fluid limit approximation of C (recall that C_f is obtained by dividing each item into infinitely many smaller items; that is, the items become divisible like fluid). The accuracy of this fluid approximation is demonstrated in the following theorem. The theorem shows that in the heavy-tailed context the fluid limit search cost distribution behaves asymptotically the same as the original search cost distribution.

THEOREM 5. *If $q(u) \sim c/u^\alpha$, $\alpha > 1$, $c > 0$, then*

$$\mathbb{P}[C > n] \sim \mathbb{P}[C_f > n] \quad \text{as } n \to \infty.$$

In order to investigate the asymptotic behavior of the distribution function of C_f we define [in the same spirit as in (3.1) and (3.2)] the following functions:

$$\begin{aligned} f_f(t) &= \int_0^\infty q^2(u)e^{-q(u)t}\,du \quad \text{as defined,} \\ g_f(t) &= \int_0^\infty (1-e^{-q(u)t})\,du \quad \text{as defined,} \end{aligned} \tag{4.2}$$

where the subscript f refers to the fluid limit.

PROOF. Similarly to Lemmas 1 and 2, one can show that

$$f_f(t) \sim \frac{c^{1/\alpha}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right)t^{-2+(1/\alpha)} \quad \text{as } t \to \infty$$

and

$$g_f(t) \sim \Gamma\left(1-\frac{1}{\alpha}\right)c^{1/\alpha}\,t^{1/\alpha} \quad \text{as } t \to \infty.$$

By using exactly the same procedure as in the proof of Theorem 3, one can complete the proof of this theorem. In order to avoid duplications we omit this derivation. □

THEOREM 6 (e^γ law). *If* $q(u) \sim c\exp(-\lambda u^\beta)$ *as* $u \to \infty$, *for some positive constants* c, λ, β, *then*

$$\mathbb{P}[C_f > n] \sim e^\gamma \mathbb{P}[R > n] \quad \textit{as } n \to \infty,$$

for any (*fixed*) *choice of the parameters* c, λ, β.

REMARKS. (i) Note that $q(u) \sim c\exp(-\lambda u^\beta)$ is a large class of distributions, containing most of the well-known light-tailed distributions, for example, Weibull, exponential and Normal. (ii) Observe that polynomial ($\sim c/u^\alpha$) and Weibull distributions ($\sim c\exp(-\lambda u^\beta)$, $0 < \beta < 1$) belong to the class of "subexponential" distributions. In the queueing context these distributions demonstrate the same asymptotic behavior (see [20]). (iii) Based on the discussion in Section 2.2 we conclude that, under the conditions of this theorem and Theorem 3, the LRU fault probability is only a factor $e^\gamma \approx 1.78$ larger than for the optimal static setup.

LEMMA 3. *If* $q(u) \sim c\exp(-\lambda u^\beta)$ *as* $u \to \infty$, *where* $(c, \lambda, \beta) > 0$, *then the first derivative of* g_f *behaves asymptotically as*

$$g_f'(t) \sim \frac{(\log(ct))^{\beta^{-1}-1}}{t\beta\lambda^{1/\beta}} \quad \textit{as } t \to \infty.$$

The proof is given in Section 7.4.

LEMMA 4. *If* $q(u) \sim c\exp(-\lambda u^\beta)$ *as* $u \to \infty$, *for any* $(c, \lambda, \beta) > 0$, *then* f_f *of* (4.2) *behaves asymptotically as*

$$f_f(t) \sim \frac{(\log(ct))^{\beta^{-1}-1}}{t^2\beta\lambda^{1/\beta}} \quad \textit{as } t \to \infty.$$

The proof is given in Section 7.4.

COROLLARY 2. *If* $q(u) \sim c\exp(-\lambda u^\beta)$ *as* $u \to \infty$, *for any* $(c, \lambda, \beta) > 0$, *then*

$$\frac{f_f(t)}{g_f'(t)} \sim \frac{1}{t} \quad \textit{as } t \to \infty.$$

The proof follows directly from Lemmas 3 and 4.

LEMMA 5. *For any* $d > -1$, *and* $t > 0$,

$$\int_0^t \frac{1-e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx - \frac{(\log t)^{d+1}}{d+1} \sim \gamma(\log t)^d \quad \textit{as } t \to \infty.$$

The proof is given in Section 7.4.

LEMMA 6. *If* $q(u) \sim c \exp(-\lambda u^{\beta})$ *as* $u \to \infty$, $(c, \lambda, \beta) > 0$, *then the inverse of* $g_f(t)$ *behaves asymptotically as*

$$g_f^{-1}(v) \sim e^{-\gamma} c^{-1} \exp(\lambda v^{\beta}) \quad \text{as } v \to \infty.$$

PROOF. For any $\delta > 0$ we can choose u_0 such that for all $u > u_0$, $c(1-\delta)\exp(-\lambda u^{\beta}) \le q(u) \le c(1+\delta)\exp(-\lambda u^{\beta})$; let $c_{\delta} = (1+\delta)c$. Then

$$g_f(t) \le \int_0^{\infty} (1 - \exp(-c_{\delta} t \exp(-\lambda u^{\beta})))\, du + u_0.$$

Now, by changing the variable of integration to $x = c_{\delta} t \exp(-\lambda u^{\beta})$ in the integral above, we compute

$$g_f(t) \le \frac{1}{\beta \lambda^{1/\beta}} \int_0^{c_{\delta} t} \frac{1 - e^{-x}}{x} \left(\log\left(\frac{c_{\delta} t}{x} \right) \right)^{1/\beta - 1} dx + u_0. \tag{4.3}$$

Here, by choosing $d = 1/\beta - 1$ and applying Lemma 5 in (4.3) we obtain

$$\begin{aligned} g_f(t) \le{}& \lambda^{-1/\beta} (\log c_{\delta} t)^{1/\beta} \\ &+ \frac{\gamma}{\beta \lambda^{1/\beta}} (\log c_{\delta} t)^{1/\beta - 1} (1 + o(1)) \quad \text{as } t \to \infty. \end{aligned} \tag{4.4}$$

Consequently, by using the above expression we compute

$$\begin{aligned} g_f(e^{-\gamma} c_{\delta}^{-1} \exp(\lambda u^{\beta})) &\le \lambda^{-1/\beta} (\lambda u^{\beta} - \gamma)^{1/\beta} \\ &\quad + \frac{\gamma}{\beta \lambda^{1/\beta}} (\lambda u^{\beta} - \gamma)^{1/\beta - 1} (1 + o(1)) \quad \text{as } u \to \infty \\ &= u \left(1 - \frac{\gamma}{\beta \lambda u^{\beta}} (1 + o(1)) \right) \\ &\quad + \frac{\gamma}{\beta \lambda u^{\beta - 1}} (1 + o(1)) \quad \text{as } u \to \infty \\ &= u \left(1 + o\left(\frac{1}{u^{\beta}} \right) \right) \quad \text{as } u \to \infty. \end{aligned} \tag{4.5}$$

If we introduce a new variable $v \equiv v(u) = u(1 + o(1/u^{\beta}))$ as $u \to \infty$, then

$$\begin{aligned} v^{\beta} \left(\frac{v}{u} - 1 \right) &= v^{\beta} o\left(\frac{1}{u^{\beta}} \right) = o(1) \quad \text{as } u \to \infty \\ &\Rightarrow u = v \left(1 + o\left(\frac{1}{v^{\beta}} \right) \right) \quad \text{as } v \to \infty. \end{aligned}$$

Finally, by replacing the preceding expression in (4.5), it directly follows that

$$\begin{aligned} g_f^{-1}(v) &\ge e^{-\gamma} c_{\delta}^{-1} \exp(\lambda (v(1 + o(v^{-\beta})))^{\beta}) \quad \text{as } v \to \infty \\ &= e^{-\gamma} c_{\delta}^{-1} \exp(\lambda (v^{\beta} + o(1))) \quad \text{as } v \to \infty, \end{aligned}$$

which implies that

$$\liminf_{v\to\infty} g_f^{-1}(v)e^{\gamma}c\exp(-\lambda v^{\beta}) \geq (1+\delta)^{-1}.$$

Finally, by passing $\delta \to 0$ we obtain the lower bound of the proof of the lemma. To prove the upper bound we use

$$g_f(t) \geq \int_0^{\infty} (1-\exp(-c_{\delta}t\exp(-\lambda u^{\beta})))\,du - u_0,$$

where $c_{\delta} = (1-\delta)c$; by repeating exactly the same arguments as for the lower bound we obtain

$$\limsup_{v\to\infty} g_f^{-1}(v)e^{\gamma}c\exp(-\lambda v^{\beta}) \leq 1.$$

This completes the proof of the lemma. □

Finally, we are ready to supply the proof of the theorem.

PROOF OF THEOREM 6. By changing the variable of integration in Theorem 4 to $v = g_f(t)$ we obtain

$$\mathbb{E}\exp(-sC_f) = \int_0^{\infty} \frac{f_f(g_f^{-1}(v))}{g_f'(g_f^{-1}(v))}e^{-sv}\,dv.$$

Thus, by the uniqueness of the Laplace transform inverse, we conclude that the density of C_f is equal to

$$q_f(v) = \frac{f_f(g_f^{-1}(v))}{g_f'(g_f^{-1}(v))}.$$

Now by using Corollary 2 we derive

$$q_f(v) \sim \frac{1}{g_f^{-1}(v)} \quad \text{as } v\to\infty,$$

which by application of Lemma 6 yields

$$q_f(v) \sim e^{\gamma}c\exp(-\lambda v^{\beta}) \quad \text{as } v\to\infty.$$

This concludes the proof of the theorem. □

5. Simulation experiments. In this section we illustrate our main results (Theorems 3 and 6) with several simulation examples. Note that the asymptotic results were obtained first by passing the list size N to infinity in Theorem 1, then by investigating the tail of the limiting search cost distribution as n goes to infinity (or, equivalently, the LRU cache fault probability as the cache size n grows). Thus, it can be expected that the asymptotic expressions from Theorems 3 and 6 will give a reasonable approximation for $\mathbb{P}[C^N > n]$ when both n and N are large and N is significantly larger than n. However, it is surprising how accurately these approximations work for relatively *small* values of N and *almost all* values of $n \le N$.

The experiments were conducted on a modern multiprocessor Silicon Graphics computer. We have used C++ programming language with a standard 48-bit pseudorandom number generator. The initial position of the items in the list was chosen uniformly at random. In each experiment, before we have conducted the measurements, we allowed a certain amount of time τ_d for the system to reach its steady state. In general, we have adopted a heuristic for choosing τ_d such that $|\mathbb{E}C^N_{\tau_d} - \mathbb{E}C^N|/\mathbb{E}C^N < 1\%$ [recall that $\mathbb{E}C^N_{\tau_d}$ is given by (2.2)]. Typically, τ_d was smaller than 10^6 time units, where the only exception was the first experiment when the convergence to stationarity was very slow and we had to choose $\tau_d = 10^8$ to achieve $|\mathbb{E}C^N_{\tau_d} - \mathbb{E}C^N|/\mathbb{E}C^N < 5\%$. Then, after waiting τ_d units of time, in every experiment we have measured the search cost probabilities for a time interval τ which, depending on the experiment, was between 10^8 and 10^{10} time units. The measured data is presented in the remainder of this section.

EXAMPLE 1 (Heavy tails). In this example we will illustrate the heavy-tailed case from Theorem 3, that is, the case when the request distribution obeys a generalized Zipf's law, $\mathbb{P}[R^N = n] = c_N/n^\alpha$, $1 \le n \le N$. In this case we will use the approximation $\mathbb{P}[C^N = n] \approx (K(\alpha)c_N)/n^\alpha$.

In the first experiment we considered $\alpha = 1.4$, $N = 10^6$. The search cost was measured for $\tau = 10^8$ time units ($\tau_d = 10^8$). The expected value of the search cost is larger than $\mathbb{E}C^N \ge \mathbb{E}R^N \approx 2100$, from which the expected number of item lookups is greater than 10^{11}. Hence, it took more than *three days* on a modern high speed computer to complete this simulation. The simulation results are displayed with a solid line in Figure 2. The top part of the figure represents a zoomed-in view for small values of C^N. The bottom plot in the same figure represents a zoomed out view of the same experiment. On the other hand, it is needless to say that it takes negligible computer time to evaluate the normalization constant for the Zipf's law distribution $c_N = 0.322004$ and the asymptotic proportionality constant $K(\alpha) = 1.42362$, which together yield the approximation of the search cost density $K(\alpha)c_N/n^\alpha = 0.4584139/n^\alpha$. The plot of the approximation is represented on the same figure with a dashed line. From the figure we can see that the approximation converges very quickly to the actual distribution, that is, it already becomes accurate for $\mathbb{P}[C^N = 5]$. Hence, the approximation is *almost identical*, except for $n < 5$, to the simulation results.

Similar experiments are repeated for $\alpha = 3$, $N = 10^3$ and $\alpha = 4$, $N = 100$; the measurements were $\tau = 10^9$ and $\tau = 10^{10}$ time units long, respectively. The results are displayed on the top and bottom plots of Figure 3, respectively. The corresponding approximations are presented with dashed lines on the same figure. The accuracy of the approximation is apparent.

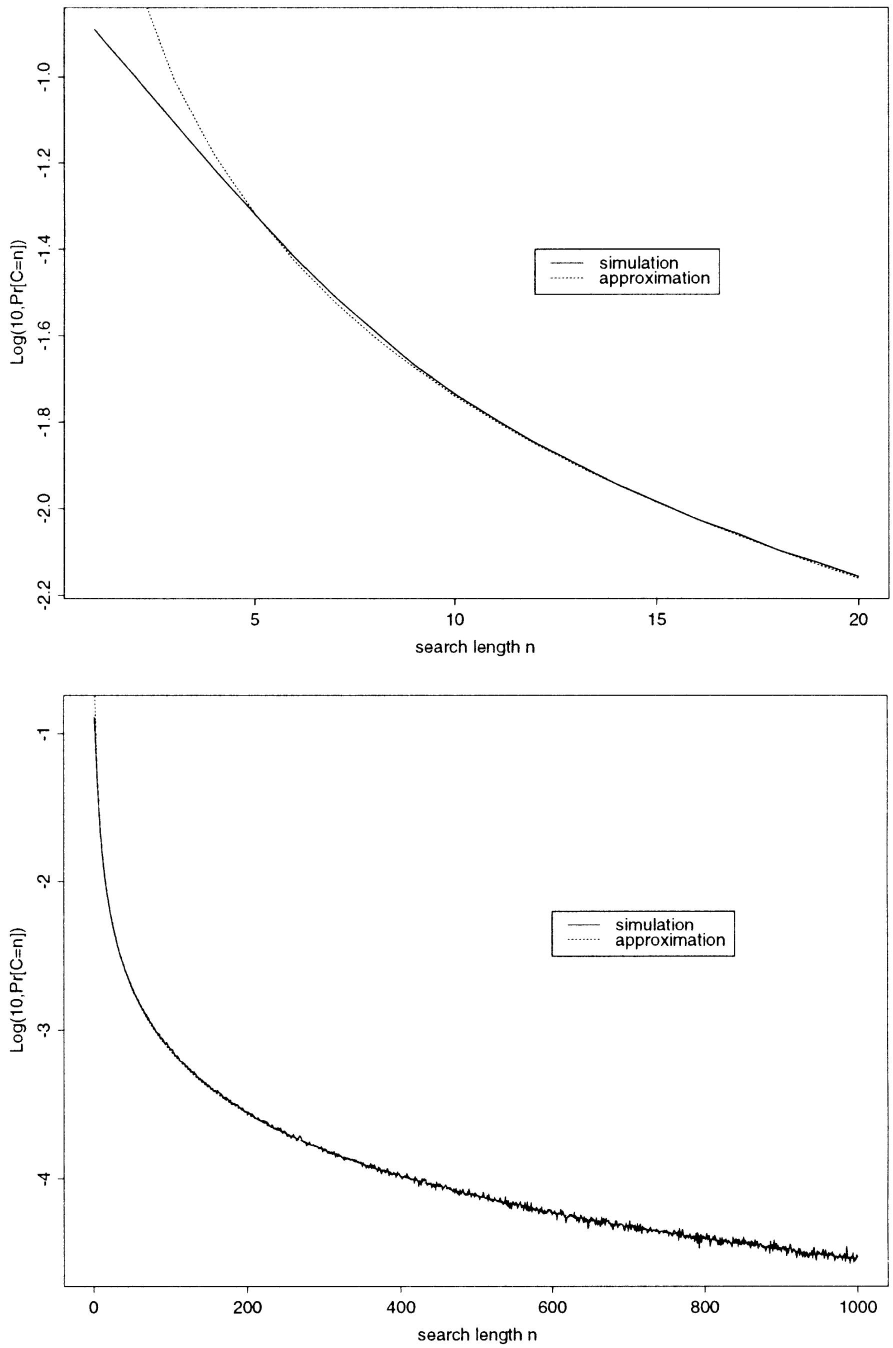

FIG. 2. *Illustration for Example* 1.

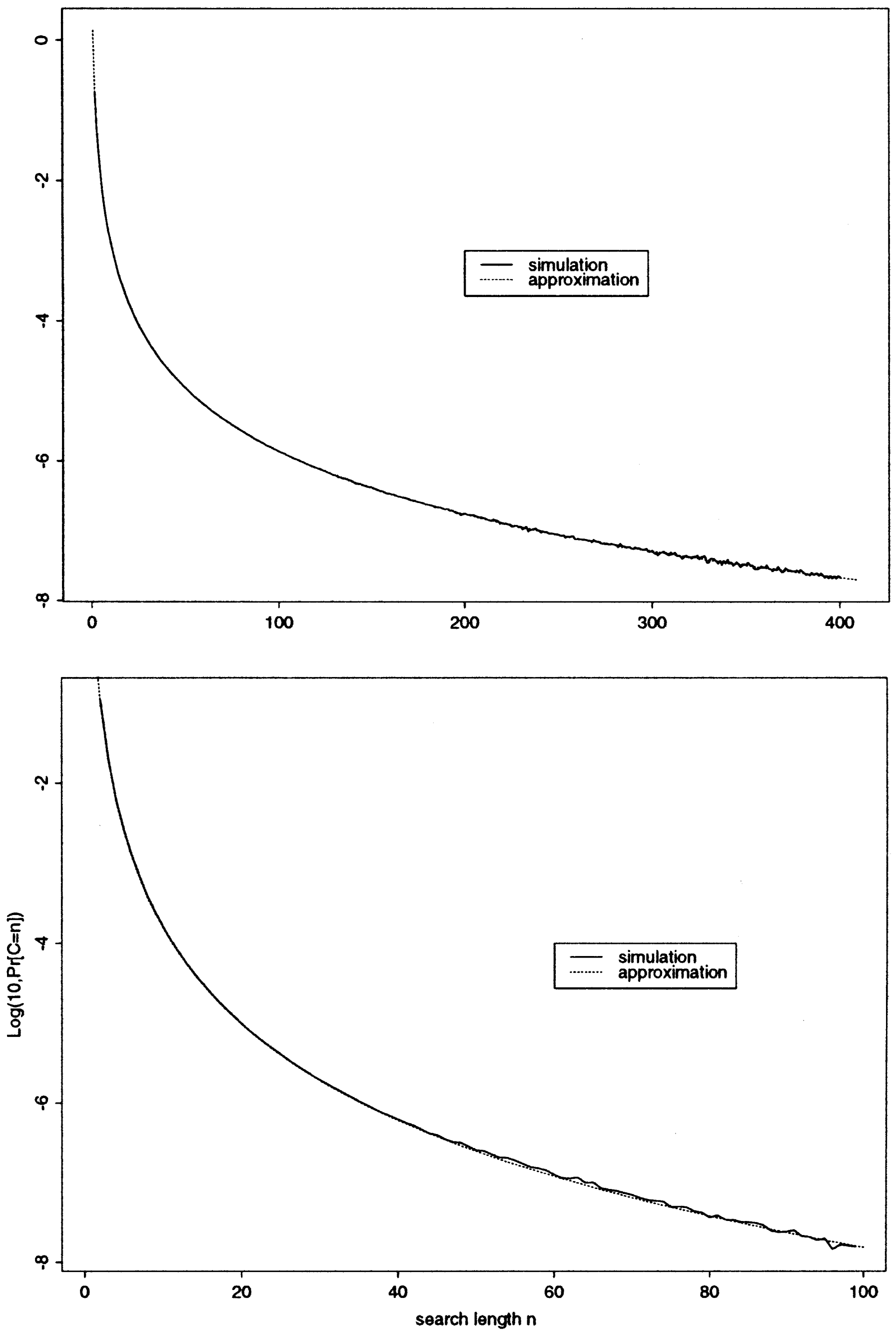

FIG. 3. *Illustration for Example* 1.

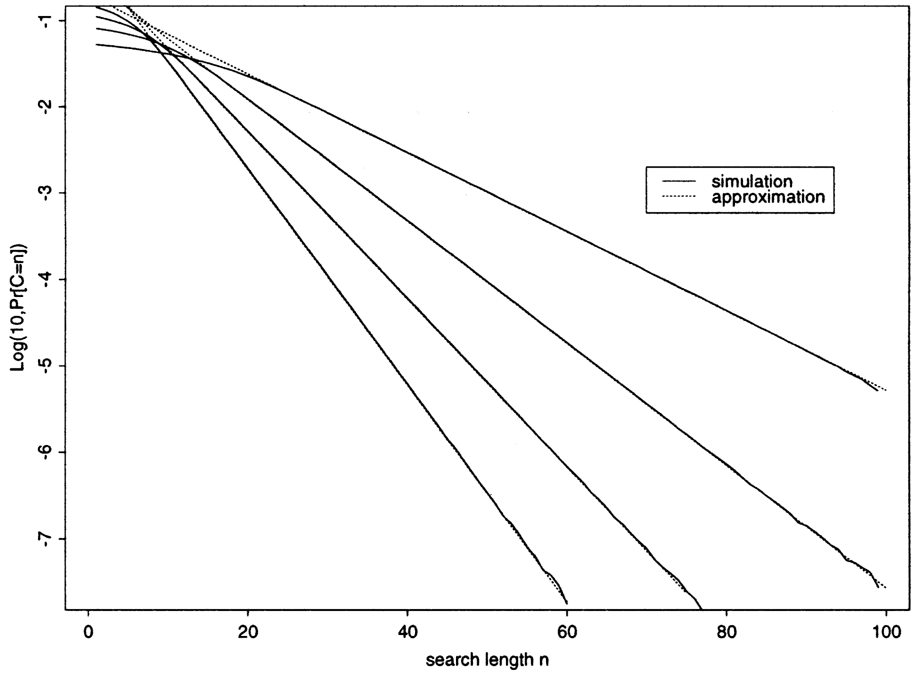

FIG. 4. *Illustration for Example* 2.

EXAMPLE 2 (Light tails). In this example we illustrate the light-tailed search distribution behavior that was asymptotically characterized by Theorem 6. As suggested by the asymptotic result for all the experiments we use the same asymptotic approximation $\mathbb{P}[C^N = n] \approx e^{\gamma}\mathbb{P}[R^N = n]$. Figure 4 contains the geometric (exponential) case for four different values of the geometric distribution parameter $p = 0.75, 0.8, 0.85, 0.9$; the observation interval was $\tau = 10^9$ time units for all the experiments. Again, the agreement between the approximation and the simulation results is evident.

Similar precision was observed for the distribution that has a Weibull tail $P[R^N = n] = c_N \exp(-\sqrt{n})$, $N = 1000$; see the top part of Figure 5. The experiment for a Normal-like tail $P[R^N = n] = c_N \exp(-0.005n^2)$, $N = 100$ is presented on the bottom part of the same figure. The measurements were conducted for $\tau = 10^9$ and $\tau = 10^{10}$ time units, respectively.

6. Conclusion. In this paper we obtained a complete asymptotic characterization of the MTF search cost distribution function or, equivalently, the LRU caching fault probability, for both heavy and light tails. In both cases the tail of the MTF search cost distribution is asymptotically directly proportional to the tail of the request distribution with an explicitly computable constant of proportionality.

In the heavy-tailed (polynomial) case, the constant is a function of the polynomial exponent. As the tail becomes lighter, the constant increases to

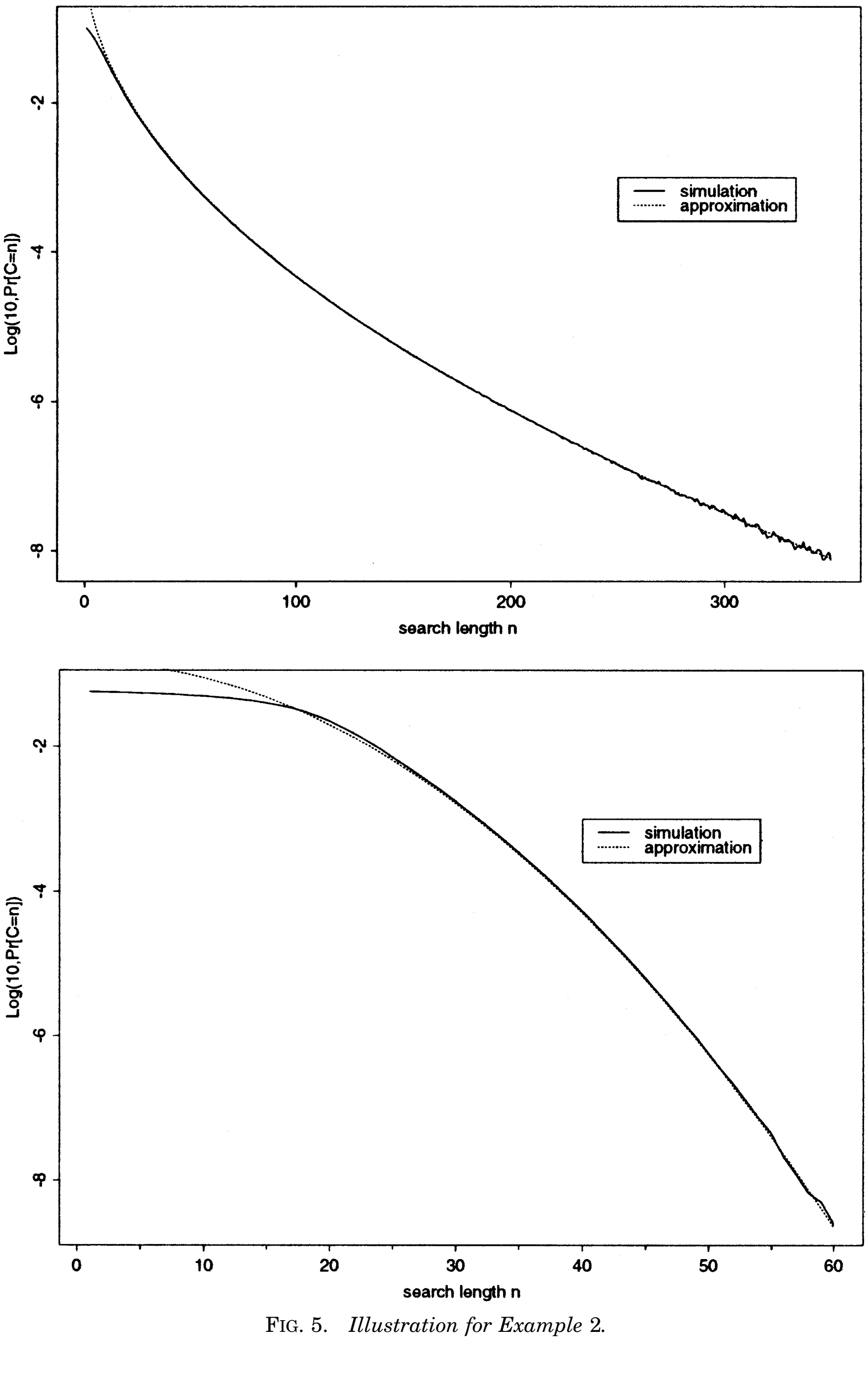

FIG. 5. *Illustration for Example 2.*

$e^\gamma \approx 1.78$, where γ is Euler's constant. In the light-tailed case the constant is *invariant* with respect to the request distribution shape and is always equal to e^γ.

We expect that the new asymptotic techniques developed in this paper will be useful for the analysis of more complex data structures.

7. Proofs.

7.1. *Proofs of Lemmas* 1 *and* 2.

PROOF OF LEMMA 1. Let us first prove the case $q_r = c/r^\alpha$. Observe that for $t > 0$, the function $(c^2/r^{2\alpha})\exp(-ct/r^\alpha)$ is increasing in r for $r < (ct/2)^{1/\alpha}$, it is decreasing for $r > (ct/2)^{1/\alpha}$ and it has its maximum $4e^{-2}/t^2$ for $r = (ct/2)^{1/\alpha}$. Using these observations we can obtain the following set of equations. Let $l(t) = \lfloor (ct/2)^{1/\alpha} \rfloor$ as defined, where $\lfloor x \rfloor$ represents the integer part of x. Then

$$\begin{aligned} f(t) &\le \sum_{r=1}^{l(t)-1} (c^2/r^{2\alpha})\exp(-ct/r^\alpha) \\ &\quad + \sum_{r=l(t)+1}^{\infty} (c^2/r^{2\alpha})\exp(-ct/r^\alpha) + 4e^{-2}/t^2 \\ &\le \sum_{r=1}^{l(t)-1} \int_r^{r+1} (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du \\ &\quad + \sum_{r=l(t)+1}^{\infty} \int_{r-1}^{r} (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du + 4e^{-2}/t^2 \\ &\le \int_1^\infty (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du + 4e^{-2}/t^2 \\ &\le \int_0^\infty (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du + 4e^{-2}/t^2. \end{aligned} \tag{7.1}$$

By changing variables to $v = ct/u^\alpha$ in the integral in (7.1), we compute

$$\int_0^\infty (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du = \frac{c^{1/\alpha}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right)t^{-2+1/\alpha}. \tag{7.2}$$

Finally, by substituting (7.2) in (7.1), and using $t^{-2}/t^{\alpha^{-1}-2} = o(1)$ as $t \to \infty$, we obtain

$$\limsup_{t\to\infty} f(t)t^{2-1/\alpha} \le \frac{c^{1/\alpha}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right). \tag{7.3}$$

Similarly, one can derive the lower bound

$$\begin{aligned} f(t) &\ge \sum_{r=1}^{l(t)} (c^2/r^{2\alpha})\exp(-ct/r^\alpha) + \sum_{r=l(t)+1}^{\infty} (c^2/r^{2\alpha})\exp(-ct/r^\alpha) \\ &\ge \int_0^\infty (c^2/u^{2\alpha})\exp(-ct/u^\alpha)\,du - 4e^{-2}/t^2. \end{aligned} \tag{7.4}$$

Thus, by replacing (7.2) in (7.4), and by taking the limit infimum with respect to t, we conclude

$$\liminf_{t\to\infty} f(t)t^{2-1/\alpha} \ge \frac{c^{1/\alpha}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right). \tag{7.5}$$

Now, the combination of (7.5) and (7.3) yields the proof for the case $q_r = c/r^\alpha$.

For the general case ($q_r \sim c/r^\alpha$), for any $c > \varepsilon > 0$ we can choose $r_0 > 0$ such that, for all $r \ge r_0$, $-\varepsilon < q_r r^\alpha - c < \varepsilon$. Using this we obtain

$$\begin{aligned} f(t) &\le \frac{r_0 4e^{-2}}{t^2} + \sum_{r=r_0+1}^{\infty} \frac{(c+\varepsilon)^2}{r^{2\alpha}} \exp\left(\frac{-(c-\varepsilon)t}{r^\alpha}\right) \\ &\le \frac{r_0 4e^{-2}}{t^2} + \frac{(c+\varepsilon)^2}{(c-\varepsilon)^2} \sum_{r=1}^{\infty} \frac{(c-\varepsilon)^2}{r^{2\alpha}} \exp\left(\frac{-(c-\varepsilon)t}{r^\alpha}\right) \end{aligned} \tag{7.6}$$

Consequently, by applying what we already have proved in (7.3)–(7.6) we arrive at

$$\limsup_{t\to\infty} f(t)t^{2-1/\alpha} \le \frac{(c+\varepsilon)^2}{(c-\varepsilon)^2}\frac{(c-\varepsilon)^{1/\alpha}}{\alpha}\Gamma\left(2-\frac{1}{\alpha}\right). \tag{7.7}$$

Finally, by passing $\varepsilon \to 0$, we prove the extension of (7.3). Similarly, starting with

$$f(t) \ge -\frac{r_0 4e^{-2}}{t^2} + \sum_{r=1}^{\infty} \frac{(c-\varepsilon)^2}{r^{2\alpha}} \exp\left(-\frac{(c+\varepsilon)t}{r^\alpha}\right)$$

we derive the analog of (7.5). This completes the proof of the lemma. □

PROOF OF LEMMA 2. Similarly to the proof of Lemma 1, let us first consider the case $q_r = c/r^\alpha$. Note that for $t > 0$, $1 - \exp(-ct/r^\alpha)$ is monotonically decreasing in r. Then,

$$\begin{aligned} g(t) &= \sum_{r=1}^{\infty} \int_r^{r+1} (1-\exp(-ct/r^\alpha))\,du \\ &\ge \sum_{r=1}^{\infty} \int_r^{r+1} (1-\exp(-ct/u^\alpha))\,du \\ &= \int_1^{\infty} (1-\exp(-ct/u^\alpha))\,du \\ &= \frac{1}{\alpha}c^{1/\alpha}t^{1/\alpha} \int_0^{ct} (1-e^{-v})v^{-(1/\alpha)-1}\,dv \\ &= 1 - e^{-ct} + c^{1/\alpha}t^{1/\alpha} \int_0^{ct} v^{-1/\alpha}e^{-v}\,dv, \end{aligned} \tag{7.8}$$

where in the last equality we have used integration by parts with $U = 1 - e^{-v}$, $dU = e^{-v}dv$, $V = -\alpha v^{-1/\alpha}$, $dV = v^{-(1/\alpha)-1}dv$. Since $\int_0^{ct} v^{-1/\alpha}e^{-v}\,dv \to$

$\Gamma(1-1/\alpha)$ as $t \to \infty$, from (7.8) it follows that

$$\liminf_{t\to\infty} \frac{g(t)}{t^{1/\alpha}} \geq \Gamma\left(1-\frac{1}{\alpha}\right) c^{1/\alpha}. \tag{7.9}$$

Similarly,

$$\begin{aligned} g(t)-(1-e^{-ct}) &= \sum_{r=2}^{\infty} \int_{r-1}^{r} (1-\exp(-ct/r^{\alpha}))\,du \\ &\leq \sum_{r=2}^{\infty} \int_{r-1}^{r} (1-\exp(-ct/u^{\alpha}))\,du \\ &= \int_{1}^{\infty} (1-\exp(-ct/u^{\alpha}))\,du. \end{aligned} \tag{7.10}$$

Finally, by replacing (7.8) in (7.10) we prove the upper bound, that is,

$$\limsup_{t\to\infty} \frac{g(t)}{t^{1/\alpha}} \leq \Gamma\left(1-\frac{1}{\alpha}\right) c^{1/\alpha}, \tag{7.11}$$

which together with (7.9) completes the proof for the case $q_r = c/r^{\alpha}$.

To prove the general case ($q_r \sim c/r^{\alpha}$), for any $0 < \varepsilon < c$, we choose $r_0 > 0$, such that for all $r \geq r_0$, $-\varepsilon < q_r r^{\alpha} - c < \varepsilon$. Since, $1-\exp(-ct/r^{\alpha})$ is monotonically increasing in c and bounded above by 1, it follows that

$$-r_0 + \sum_{r=1}^{\infty} (1-\exp(-(c-\varepsilon)t/r^{\alpha})) \leq g(t) \leq r_0 + \sum_{r=1}^{\infty} (1-\exp(-(c+\varepsilon)t/r^{\alpha})),$$

which, by applying (7.8), (7.9), (7.10), (7.11), and by passing $\varepsilon \to 0$, implies the conclusion of the lemma. □

7.2. *Proof of Theorem* 3. As we have already mentioned, the proof of this result is based on Karamata's Tauberian–Abelian theorem for distribution functions of regular variation. This theorem relates the tail behavior of a distribution function to the asymptotic behavior of its Laplace transform at the origin. For convenience we state the following result (Theorem 7) which is a weaker version of the theorem due to Bingham and Doney [3] ([4], page 333). This theorem has a wide application in probability; for example, some recent applications to queueing can be found in [21]. Let F be a distribution function on $[0,\infty)$, and let $\tilde{F}$ be its Laplace–Stieltjes transform.

THEOREM 7. *Let* $m \in \mathbb{N}_0$, *and* $\alpha = m + \beta$.

(i) *If* $0 < \beta < 1$, *then the following two asymptotic relations are equivalent*:

$$(-1)^{m+1} \tilde{F}^{(m+1)}(s) \sim \alpha\Gamma[1-\beta] \frac{c}{s^{1-\beta}} \quad \text{as } s \downarrow 0, \tag{7.12}$$

$$1 - F(x) \sim \frac{c}{x^{\alpha}} \quad \text{as } x \to \infty, \tag{7.13}$$

where Γ *stands for the gamma function and* $\tilde{F}^{(m+1)}$ *denotes the* $(m+1)$*st derivative of* $\tilde{F}$.

(ii) *If* $\beta = 1$, *then* (7.13) *is equivalent to*

$$(-1)^{m+1}\tilde{F}^{(m+1)}(s) \sim -\alpha c \log s \quad \textit{as } s \downarrow 0. \tag{7.14}$$

PROOF. (i) is just a special case of [4], Theorem 8.1.6, page 333.

(ii) For $\alpha = m+1$ the implication (7.13) $\Rightarrow$ (7.14) can be proved easily by direct evaluation of $(-1)^{m+1}\tilde{F}^{(m+1)}(s)$; we skip the details.

For the reverse implication (7.14) $\Rightarrow$ (7.13), we have that, by [4], Theorem 8.1.6, page 333, (7.14) implies

$$\int_0^x t^{m+1} dF(t) \sim (m+1)c \log x \quad \text{as } x \to \infty.$$

By changing the variables $u = t^{m+1}$, $y = x^{m+1}$, in the integral above, and by $F_1(y) = F(y^{1/(m+1)})$ as defined, we get that

$$\int_0^y u dF_1(u) \sim c \log y \quad \text{as } y \to \infty. \tag{7.15}$$

Now, by the remark after [4], Corollary 8.1.7, page 335, (7.15) is equivalent to

$$1 - F_1(y) \sim \frac{c}{y} \quad \text{as } y \to \infty,$$

which by $F_1(x^{m+1}) \equiv F(x)$ implies (7.13). This completes the proof. □

For any set of indices $A = \{i_1, \dots, i_k\}$ let

$$\Pi_A(s,t) = \prod_{r:\, r \notin A} (1 - (1 - e^{-s})(1 - \exp(-q_r t))) \quad \text{as defined;}$$

when A is the empty set we denote $\Pi_A(s,t)$ simply as $\Pi(s,t)$. Let $|A|$ denote the cardinality of A.

LEMMA 7. *For any* $\varepsilon > 0$ *and any set of indices* A, $|A| \le \ell < \infty$, *there exist* $s_0 > 0$, $t_0 < \infty$ *such that for all* $0 < s < s_0$, *and* $t > t_0$,

$$\exp(-s(1+\varepsilon)c_1 t^{1/\alpha}) \le \Pi_A(s,t) \le \exp(-s(1-\varepsilon)c_1 t^{1/\alpha}),$$

where $c_1 = \Gamma(1 - (1/\alpha))c^{1/\alpha}$.

PROOF. First let us observe that for any $|A| \le \ell$,

$$\Pi(s,t) \le \Pi_A(s,t) \le e^{s\ell}\Pi(s,t). \tag{7.16}$$

From this we see that for all sufficiently small s (and fixed ℓ) $\Pi(s,t)$ uniformly approximates $\Pi_A(s,t)$. Therefore, to complete the proof it is enough to prove that the lemma is satisfied for $\Pi(s,t)$. Now

$$\begin{aligned} \log \Pi(s,t) &= \sum_{r=1}^{\infty} \log(1 - (1 - e^{-s})(1 - \exp(q_r t))) \\ &\le -(1 - e^{-s}) \sum_{r=1}^{\infty} (1 - \exp(q_r t)), \end{aligned} \tag{7.17}$$

where in the inequality above we use the inequality $\log(1+x) \le x$. Here, by applying Lemma 2 in (7.17), and observing that $(1-e^{-s}) \sim s$, as $s \to 0$ we complete the proof of the upper bound, that is,

$$\Pi(s,t) \le \exp(-s(1-\varepsilon)c_1 t^{1/\alpha}) \tag{7.18}$$

Similarly, to prove the lower bound we can use the inequality $x - x^2 \le \log(1+x)$, $x \in [-0.683, 0]$. Therefore, by choosing s sufficiently small such that $1 - e^{-s} \le 0.683$, we obtain

$$\begin{aligned} \log \Pi(s,t) &\ge -(1-e^{-s}) \sum_{r=1}^{\infty} (1 - \exp(q_r t)) \\ &\quad - (1-e^{-s})^2 \sum_{r=1}^{\infty} (1-\exp(q_r t))^2 \\ &\ge -(1-e^{-s})(2-e^{-s}) \sum_{r=1}^{\infty} (1-\exp(q_r t)). \end{aligned} \tag{7.19}$$

Clearly, for any $\varepsilon > 0$ we can choose sufficiently small s such that $(1-e^{-s})(2-e^{-s}) \le (1+\varepsilon)s$. When this is replaced in (7.19) and by application of Lemma 2 we obtain the lower bound inequality of the lemma. This completes the proof. □

LEMMA 8. *For any* $\varepsilon > 0$, *for all* $i \ge 1$ *and for any fixed* $m \ge 0$, *there exist* $s_0 > 0$, $t_0 < \infty$ *such that for all* $0 < s < s_0$, *and* $t > t_0$,

$$\begin{aligned} (1-\varepsilon)(c_1 t^{1/\alpha})^m \exp(-s(1+\varepsilon)c_1 t^{1/\alpha}) &\le (-1)^m \frac{\partial^m}{\partial s^m} e^{-s} \Pi_i(s,t) \\ &\le (1+\varepsilon)(c_1 t^{1/\alpha})^m \exp(-s(1-\varepsilon)c_1 t^{1/\alpha}), \end{aligned}$$

where $c_1 = \Gamma(1-(1/\alpha))c^{1/\alpha}$ *and* $\Pi_i(s,t) \equiv \Pi_{\{i\}}(s,t)$.

PROOF. Let us first investigate the form of the mth derivative of $e^{-s}\Pi_i(s,t)$. For convenience of notation for any set $A = \{i_1, \dots, i_k\}$ we denote with $\Pi_{i_1,\dots,i_k}(s,t) \equiv \Pi_A(s,t)$; also, $\sum_{i_k/i_{k-1},\dots,i_1} \equiv \sum_{i_k : i_k \notin \{i_{k-1},\dots,i_1\}}$. For $m = 1$ simple algebra gives

$$\frac{\partial}{\partial s}\left[e^{-s}\Pi_i(s,t)\right] = -e^{-s}\Pi_i - e^{-2s} \sum_{k_1/i} (1-\exp(-q_{k_1} t))\Pi_{k_1 i}.$$

Similarly, for $m = 2$

$$\begin{aligned} \frac{\partial^2}{\partial s^2}\left[e^{-s}\Pi_i(s,t)\right] &= e^{-s}\Pi_i + 3e^{-2s} \sum_{k_1/i} (1-\exp(-q_{k_1} t))\Pi_{k_1 i} \\ &\quad + e^{-3s} \sum_{k_1/i} (1-\exp(-q_{k_1} t)) \sum_{k_2/k_1,\, i} (1-\exp(-q_{k_2} t))\Pi_{k_1 k_2 i}. \end{aligned}$$

Following this derivation, one can easily prove the following claim. We skip the details.

CLAIM 1. For any $m \geq 1$, there exist a set of nonnegative integers $d_1^m, \dots, d_{m-1}^m$, such that

$$\begin{aligned}
&\frac{\partial^m}{\partial s^m}\big[e^{-s}\Pi_i(s,t)\big] \\
&\quad = (-1)^m\Big[e^{-s}\Pi_i + d_1^m e^{-2s}\sum_{k_1/i}(1-\exp(-q_{k_1}t))\Pi_{k_1 i} \\
&\qquad + \cdots + d_{m-1}^m e^{-ms}\sum_{k_1/i}(1-\exp(-q_{k_1}t))\sum_{k_2/k_1,\,i}(1-\exp(-q_{k_2}t)) \\
&\qquad \cdots \sum_{k_{m-1}/k_1,\dots,k_{m-2},\,i}(1-\exp(-q_{k_{m-1}}t))\Pi_{k_1\cdots k_{m-1}i} \\
&\qquad + e^{-(m+1)s}\sum_{k_1/i}(1-\exp(-q_{k_1}t))\sum_{k_2/k_1,\,i}(1-\exp(-q_{k_2}t)) \\
&\qquad\qquad \cdots \sum_{k_m/k_1,\dots,k_{m-1},\,i}(1-\exp(-q_{k_m}t))\Pi_{k_1\cdots k_m i}\Big].
\end{aligned} \tag{7.20}$$

Finally, by applying Lemma 7 and Corollary 1 we complete the proof of this lemma. □

Here, we are ready to complete the proof of Theorem 3. Observe that if $\mathbb{P}[R=n] \sim c/n^{\alpha}$, $\alpha > 1$, then $\mathbb{P}[R > n] \sim (c/(\alpha-1))n^{-\alpha+1}$, as $n \to \infty$.

Let us first consider the case $\alpha = m+\beta$, $0<\beta<1$, $m \in \mathbb{N}$. Note that we can take the mth derivative of $\mathbb{E}e^{-sC}$ by interchanging the order of differentiation and integration–summation. Justification for this interchange follows by the dominated convergence theorem and the following bound.

LEMMA 9. *For any (fixed) $s > 0$, integer $\ell \geq 0$, there exist $h_0 > 0$, $\delta \equiv \delta(s) > 0$, $\theta \equiv \theta(s)$, such that for all $A = \{i_1, \dots, i_k\}$, $|A| \leq \ell$, $0 < |h| < h_0$,*

$$\left|\frac{\Pi_A(s,t) - \Pi_A(s+h,t)}{h}\right| \leq \theta t^{1/\alpha}\exp(-\delta t^{1/\alpha}). \tag{7.21}$$

Proof of this lemma is given at the end of this section.

Thus, by taking the mth derivative of $\mathbb{E}e^{-sC}$ and applying Lemma 1 and Lemma 8, we can choose, for any $\varepsilon > 0$, $t_0 < \infty$, $s_0 > 0$ such that for all $0 < s < s_0$ and $t > t_0$,

$$\begin{aligned}
(-1)^m\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\geq (1-\varepsilon)\int_{t_0}^{\infty}\frac{c^{1/\alpha}}{\alpha}\Gamma\Big(2-\frac{1}{\alpha}\Big)t^{-2+1/\alpha}(c_1 t^{1/\alpha})^m \\
&\qquad \times \exp(-s(1+\varepsilon)c_1 t^{1/\alpha})\,dt \\
&= (1-\varepsilon)c_1^n c_2\int_{t_0}^{\infty} t^{-2+(m+1)/\alpha}\exp(-s(1+\varepsilon)c_1 t^{1/\alpha})\,dt,
\end{aligned} \tag{7.22}$$

where $c_1 = \Gamma(1-(1/\alpha))c^{1/\alpha}$ and $c_2 = (c^{1/\alpha}/\alpha)\Gamma(2-1/\alpha)$; now, with the change of variables $u = s(1+\varepsilon)c_1 t^{1/\alpha}$ we arrive at

$$\begin{aligned}(-1)^m \frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\geq (1-\varepsilon)c_2 c_1^m \alpha(s(1+\varepsilon)c_1)^{\beta-1}\int_{s(1+\varepsilon)c_1 t_0^{1/\alpha}}^{\infty} u^{-\beta}e^{-u}\,du \\ &\gtrsim (1-\varepsilon)c_2 c_1^m \alpha(s(1+\varepsilon)c_1)^{\beta-1}\Gamma(1-\beta) \quad \text{as } s\downarrow 0.\end{aligned} \tag{7.23}$$

Finally by taking the limit with respect to ε we get

$$\liminf_{s\downarrow 0}\left\{(-1)^m\left[\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC}\right]s^{m+1-\alpha}\right\} \geq c_1^{\alpha-1}c_2\alpha\Gamma(m+1-\alpha). \tag{7.24}$$

For the upper bound we use

$$\begin{aligned}&(-1)^m \frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} \\ &\quad\leq \eta(t_0) + (1+\varepsilon)c_1^m c_2 \int_{t_0}^{\infty} t^{-2+(m+1)/\alpha}\exp(-s(1-\varepsilon)c_1 t^{1/\alpha})\,dt \\ &\quad\lesssim (1+\varepsilon)c_2 c_1^m \alpha\Gamma(m+1-\alpha)((1-\varepsilon)c_1 s)^{\alpha-m-1} \quad \text{as } s\downarrow 0,\end{aligned} \tag{7.25}$$

where $\eta(t_0)$ is a sufficiently large constant. Now, by letting $\varepsilon \to 0$ in (7.25) and combining it with (7.24) we obtain

$$\lim_{s\downarrow 0}\left\{(-1)^m\left[\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC}\right]s^{m+1-\alpha}\right\} = c_1^{\alpha-1}c_2\alpha\Gamma(1-\beta). \tag{7.26}$$

In conclusion, by applying Karamata's theorem 7(i) we derive

$$\mathbb{P}[C>n] \sim \frac{\alpha c_1^{\alpha-1}c_2}{c}\,\frac{c}{\alpha-1}n^{-\alpha+1} \sim \frac{\alpha c_1^{\alpha-1}c_2}{c}\mathbb{P}[R>n] \quad \text{as } n\to\infty,$$

which by replacing c_1 and c_2 yields the proof of the case $\alpha = m+\beta$, $0<\beta<1$.

For integer $\alpha = m+1$ by combining the same reasoning as in (7.23) and (7.25), one can easily obtain

$$\begin{aligned}\frac{\partial^m}{\partial s^m}\mathbb{E}e^{-sC} &\sim (-1)^m c_1^m c_2(m+1)\int_{sc_1 t_0^{1/\alpha}}^{\infty} u^{-1}e^{-u}\,du \\ &\sim (-1)^m c_1^m c_2(m+1)\log(1/s) \quad \text{as } s\downarrow 0.\end{aligned} \tag{7.27}$$

Finally, by applying Theorem 7(ii) we obtain the proof for integer α and conclude the proof of the expression (3.3) of the theorem.

At this point, we are going to prove that $K(\alpha)$ is monotonically increasing in α for $\alpha > 1$, with its limits at 1 and ∞ given by (3.4). Observe that

$$\begin{aligned}K(\alpha) &= \left(1-\frac{1}{\alpha}\right)\left[\Gamma\left(1-\frac{1}{\alpha}\right)\right]^{\alpha} \\ &= \left(1-\frac{1}{\alpha}\right)^{-\alpha+1}\left[\Gamma\left(2-\frac{1}{\alpha}\right)\right]^{\alpha},\end{aligned} \tag{7.28}$$

where in the last equality we have used the identity $\Gamma(x+1) = x\Gamma(x)$. Next, the monotonicity of $K(\alpha)$ will follow if we prove that

$$\log K(\alpha) = -(\alpha-1)\log\left(1-\frac{1}{\alpha}\right) + \alpha\log\Gamma\left(2-\frac{1}{\alpha}\right) \tag{7.29}$$

is monotonically increasing for $\alpha > 1$. By taking a derivative in (7.29) we arrive at

$$\frac{d}{d\alpha}\log K(\alpha) = \frac{-1}{\alpha} - \log\left(1-\frac{1}{\alpha}\right) + \log\Gamma\left(2-\frac{1}{\alpha}\right) + \frac{1}{\alpha}\psi^{(0)}\left(2-\frac{1}{\alpha}\right), \tag{7.30}$$

where $\psi^{(k)}$, $k = 0, 1, \ldots$, are Polygamma functions (see [1], equation 6.4.1, page 260). Furthermore, since $\psi^{(0)}(1) = -\gamma$ (Euler's constant) and $\Gamma(1) = 1$, by letting $\alpha \to \infty$ in (7.30) we conclude

$$\frac{d}{d\alpha}\log K(\alpha) \to 0 \quad \text{as } \alpha \to \infty. \tag{7.31}$$

Fortunately, the second derivative of $\log K(\alpha)$ takes the following relatively simple form:

$$\frac{d^2}{d\alpha^2}\log K(\alpha) = -\frac{1}{(\alpha-1)\alpha^2} + \frac{1}{\alpha^3}\psi^{(1)}\left(2-\frac{1}{\alpha}\right). \tag{7.32}$$

Now, we intend to prove that (7.32) is negative for all $\alpha > 1$; for this we use the following claim.

CLAIM 2. For any $\alpha \geq 1$,

$$\psi^{(1)}\left(2-\frac{1}{\alpha}\right) < \frac{4\alpha^2}{(2\alpha-1)^2}. \tag{7.33}$$

PROOF. Note that (7.33) is equivalent to

$$\psi^{(1)}(z) < \frac{4}{z^2}, \tag{7.34}$$

for all $1 \leq z < 2$. By using the integral representation given in [1], equation 6.4.1, page 260, of $\psi^{(1)}(z)$ we arrive at

$$\psi^{(1)}(z) = \int_0^\infty \frac{te^{-zt}}{1-e^{-t}}\,dt, \tag{7.35}$$

which by the change of variable $t = u/z$ is equivalent to

$$\psi^{(1)}(z) = \frac{1}{z^2}\int_0^\infty \frac{ue^{-u}}{1-e^{-u/z}}\,du. \tag{7.36}$$

For $1 \leq z \leq 2$ the integral in (7.36) is bounded by

$$\int_0^\infty \frac{ue^{-u}}{1-e^{-u/2}}du \leq \frac{20}{7}\int_0^1 e^{-u}\,du + \frac{20}{7}\int_1^\infty ue^{-u}\,du = \frac{20}{7}(1+e^{-1}) < 4,$$

since $1 - e^{-u/2}$ is monotonically increasing in u and $1 - e^{-u/2} > 7u/20$ for $0 \le u \le 1$. This completes the proof of Claim 2.

Next, by replacing (7.33) in (7.32) we arrive at

$$\begin{aligned}\frac{d^2}{d\alpha^2}\log K(\alpha) &\le \frac{1}{\alpha}\left(-\frac{1}{(\alpha-1)\alpha} + \frac{4}{(2\alpha-1)^2}\right)\\ &= \frac{-1}{\alpha^2(\alpha-1)(2\alpha-1)^2} < 0,\end{aligned}$$

for $\alpha > 1$. This implies that $d(\log K(\alpha))/d\alpha$ is strictly monotonically decreasing (for $\alpha > 1$), which in combination with (7.31) yields

$$\frac{d}{d\alpha}\log K(\alpha) > 0 \quad \text{for } \alpha > 1.$$

Thus, $\log K(\alpha)$ is strictly monotonically increasing for $\alpha > 1$, and therefore the same holds for $K(\alpha)$.

The limits in (3.4) follow by straightforward application of [1], equation 6.1.34, page 256. This concludes the proof of the theorem. □

PROOF OF LEMMA 9. Let $1 > h > 0$. Then, by applying elementary inequalities, we obtain

$$\begin{aligned}0 &< 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)}\\ &\le 1 - \frac{\Pi(s+h,t)}{\Pi(s,t)}\\ &= 1 - \exp\left(-\log\frac{\Pi(s,t)}{\Pi(s+h,t)}\right)\\ &\le \log\frac{\Pi(s,t)}{\Pi(s+h,t)} \qquad (7.37)\\ &= \sum_{r=1}^{\infty}\log\left(\frac{1-(1-e^{-s})(1-\exp(-q_r t))}{1-(1-e^{-s-h})(1-\exp(-q_r t))}\right)\\ &\le \sum_{r=1}^{\infty}\log\left(\frac{1-(1-e^{-s})(1-\exp(-q_r t))}{1-(1-e^{-s})(1-\exp(-q_r t)) - he^{-s}(1-\exp(-q_r t))}\right) \qquad (7.38)\\ &= -\sum_{r=1}^{\infty}\log\left(1 - \frac{he^{-s}(1-\exp(-q_r t))}{1-(1-e^{-s})(1-\exp(-q_r t))}\right)\\ &\le -\sum_{r=1}^{\infty}\log(1 - h(1-\exp(-q_r t))), \qquad (7.39)\end{aligned}$$

where in (7.37), (7.38) we have used $1-e^{-x} \le x$, $x \ge 0$, and in (7.39) we applied $1-(1-e^{-s})(1-\exp(-q_r t)) \ge e^{-s}$. By using the inequality $-\log(1-x) \le 2x$

for $0 \le x \le 0.79$, in (7.39) we derive

$$0 < 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)} \le 2h \sum_{r=1}^{\infty}(1-\exp(-q_r t)), \tag{7.40}$$

for all $0 < h < h_0 = 0.79$. Consequently, combining (7.40) with Lemma 2 yields

$$0 < 1 - \frac{\Pi_A(s+h,t)}{\Pi_A(s,t)} \le h\theta_1 t^{1/\alpha}, \tag{7.41}$$

for a sufficiently large constant θ_1 and all $0 < h < h_0$. In addition, equations (7.16), (7.17) and Lemma 2 produce for any fixed $s > 0$,

$$\Pi_A(s,t) \le \theta_2 \exp(-\delta t^{1/\alpha}), \tag{7.42}$$

for some finite $\theta_2 \equiv \theta_2(s)$, $\delta \equiv \delta(s) > 0$, and all $A = \{i_1, \dots, i_k\}$, $|A| \le \ell$. Finally, (7.41) and (7.42) give the proof of the lemma for $0 < h \le h_0 = 0.79$ ($\theta = \theta_1\theta_2$). The proof of the lemma when $h < 0$ is completely analogous, and therefore we leave it out. □

7.3. *Proof of Theorem* 4. To prove this result it is enough to show that $\mathbb{E}\exp(-sC(k)/k)$ converges to the expression in (4.1) and that $\mathbb{E}\exp(-sC_f) \to 1$ as $s \downarrow 0$ (see [10], Theorem 6.6.3, page 190). From (2.6) one obtains

$$\begin{aligned}\mathbb{E}\exp(-sC(k)/k) &= e^{-s/k}\int_{t=0}^{\infty}\sum_{i=1}^{\infty}(q_i^k)^2\exp(-q_i^k t)\\ &\qquad\times\Big[\prod_{r:\,r\ne i}(1-(1-e^{-s/k})(1-\exp(-q_r^k t)))\Big]\,dt\\ &= e^{-s/k}\int_{v=0}^{\infty}\sum_{i=1}^{\infty}k(q_i^k)^2\exp(-kq_i^k v)\\ &\qquad\times\Big[\prod_{r:\,r\ne i}(1-(1-e^{-s/k})(1-\exp(-kq_r^k v)))\Big]\,dv,\end{aligned} \tag{7.43}$$

where the last equality follows by the change of variable $v = t/k$. First, we show that for each (fixed) $v \ge 0$,

$$\lim_{k\to\infty}\sum_{i=1}^{\infty}k(q_i^k)^2\exp(-kq_i^k v) = \int_0^{\infty} q(u)^2 e^{-q(u)v}\,du. \tag{7.44}$$

In order to prove (7.44), observe that the monotonicity of q implies

$$\begin{aligned}\sum_{i=1}^{\infty}k(q_i^k)^2\exp(-kq_i^k v) &\le \sum_{i=1}^{\infty}\Big[\frac{q((i-1)/k)^2}{k}\Big]e^{-q(i/k)v}\\ &\le \frac{q(0)^2}{k} + \int_0^{\infty} q(u)^2 e^{-q(u+(2/k))v}\,du\\ &\to \int_0^{\infty} q(u)^2 e^{-q(u)v}\,du \quad \text{as } k\to\infty,\end{aligned}$$

where the last asymptotic relation follows by the dominated convergence theorem. Analogously, we prove the lower bound and finish the proof of (7.44). In a completely similar manner, one can prove that for each (fixed) $v \geq 0$, $s > 0$, and uniformly in i,

$$\begin{aligned}\prod_{r:r\neq i}(1-(1-e^{-s/k})(1-\exp(-kq_r^k v)))\,dv\\ \to \exp\Big(-s\int_0^\infty (1-e^{-q(u)v})\,du\Big) \quad \text{as } k\to\infty;\end{aligned} \tag{7.45}$$

we omit the details.

Now, by applying Fatou's lemma and (7.44) and (7.45) we derive

$$\begin{aligned}&\liminf_{k\to\infty}\mathbb{E}\exp\Big(-s\frac{C(k)}{k}\Big)\\ &\quad\geq \int_0^\infty \Big(\int_0^\infty q^2(u)e^{-q(u)t}\,du\Big)\exp\Big(-s\int_0^\infty (1-e^{-q(u)t})\,du\Big)\,dt.\end{aligned}$$

For the upper bound, note that for all i the infinite products in (7.43) are upper bounded by 1, and therefore, for any $v_0 > 0$,

$$\begin{aligned}\mathbb{E}\exp(-sC(k)/k) \leq e^{-s/k}\int_{v=0}^{v_0}\sum_{i=1}^\infty k(q_i^k)^2\exp(-kq_i^k v)\\ \times\Big[\prod_{r:r\neq i}(1-(1-e^{-s/k})(1-\exp(-kq_r^k v)))\Big]\,dv\\ +\int_{v=v_0}^\infty \sum_{i=1}^\infty k(q_i^k)^2\exp(-kq_i^k v);\end{aligned} \tag{7.46}$$

by applying Fubini's theorem in the second integral in (7.46) and using the monotonicity of q we derive

$$\begin{aligned}\int_{v_0}^\infty \sum_{i=1}^\infty k(q_i^k)^2\exp(-kq_i^k v)\,dv &= \sum_{i=1}^\infty q_i^k\exp(-kq_i^k v_0)\\ &\leq \sum_{i=1}^\infty \frac{q((i-1)/k)}{k}\exp\Big(-q\Big(\frac{i}{k}\Big)v_0\Big)\\ &\leq \frac{q(0)}{k}+\int_0^\infty q(u)\exp\Big(-q\Big(u+\Big(\frac{2}{k}\Big)\Big)v_0\Big)\,du\\ &\to \int_0^\infty q(u)\exp(-q(u)v_0)\,du \quad \text{as } k\to\infty;\end{aligned} \tag{7.47}$$

again the last asymptotic relation follows by dominated convergence. By replacing (7.47) in (7.46) and by using the dominated convergence theorem in the first integral in (7.46) we arrive at

$$\limsup_{k\to\infty} \mathbb{E}\exp\left(-s\frac{C(k)}{k}\right) \le \int_{v=0}^{v_0}\left(\int_0^\infty q^2(u)e^{-q(u)v}\,du\right) \times \exp\left(-s\int_0^\infty (1-e^{-q(u)v})\,du\right)dv + \int_0^\infty q(u)\exp(-q(u)v_0)\,du. \tag{7.48}$$

Finally, by letting $v_0 \to \infty$ in (7.48) and by using

$$\int_0^\infty q(u)\exp(-q(u)v_0)\,du \to 0 \quad \text{as } v_0 \to \infty,$$

we show that $\mathbb{E}\exp(-s(C(k)/k))$ converges to the desired limit in (4.1) as $k \to \infty$. To complete the proof we need to show that $\mathbb{E}\exp(-sC_f) \to 1$ as $s \downarrow 0$. But, this follows by the monotone convergence theorem and

$$\int_0^\infty \left(\int_0^\infty q^2(u)e^{-q(u)t}\,du\right)dt = 1,$$

where the last equality is implied by Fubini's theorem. □

7.4. *Proofs of Lemmas* 3, 4 *and* 5.

PROOF OF LEMMA 3. By dominated convergence, from the definition of $g_f(t)$ it follows that

$$g_f'(t) = \int_0^\infty q(u)e^{-q(u)t}\,du. \tag{7.49}$$

First, let us assume that $q(u) = c\exp(-\lambda u^\beta)$. By changing the variables to $x = t\exp(-\lambda u^\beta)$ in (7.49), we compute

$$g_f'\left(\frac{t}{c}\right) = \frac{c}{t\beta\lambda^{1/\beta}}\int_0^t e^{-x}\left(\log\left(\frac{t}{x}\right)\right)^d dx, \tag{7.50}$$

where $d = (1/\beta) - 1$ (> -1) as defined. Thus, to complete the proof, it is enough to show that for $d > -1$,

$$\int_0^t e^{-x}(\log(t/x))^d\,dx \sim (\log t)^d \quad \text{as } t \to \infty. \tag{7.51}$$

To finish this, let us decompose the integral above into three integrals,

$$\int_0^t e^{-x}(\log(t/x))^d\,dx = \int_0^{1/\log t} + \int_{1/\log t}^{\log t} + \int_{\log t}^t = I_1(t) + I_2(t) + I_3(t) \qquad \text{as defined.} \tag{7.52}$$

Let us first investigate the asymptotic behavior of $I_2(t)$. Assume first that $d \geq 0$. Then,

$$I_2(t) \leq (\log(t \log t))^d \int_{1/\log t}^{\log t} e^{-x}\, dx \sim (\log t)^d \tag{7.53}$$

as $t \to \infty$; similarly for the lower bound

$$I_2(t) \geq \left(\log\left(\frac{t}{\log t}\right)\right)^d \int_{1/\log t}^{\log t} e^{-x}\, dx \sim (\log t)^d \tag{7.54}$$

as $t \to \infty$. For $0 > d > -1$ the inequalities in (7.53) and (7.54) will hold with the inequalities being reversed. Thus, we have proved that

$$I_2(t) \sim (\log t)^d \quad \text{as } t \to \infty. \tag{7.55}$$

For $I_1(t)$ we have the following set of estimates:

$$\begin{aligned} I_1(t) &\leq \int_0^{1/\log t} \left(\log\left(\frac{t}{x}\right)\right)^d dx \\ &= t \int_{\log(t \log t)}^{\infty} u^d e^{-u}\, du \\ &\sim t(\log(t \log t))^d \frac{1}{t \log t} = o((\log t))^d \quad \text{as } t \to \infty, \end{aligned} \tag{7.56}$$

where the asymptotic equivalence follows from [1], equation (6.5.32), page 263. Similarly, one can easily prove that

$$I_3(t) = o((\log t)^d) \quad \text{as } t \to \infty. \tag{7.57}$$

Finally, by combining (7.55)–(7.57) we conclude the proof of the case $q(u) = c\exp(-\lambda u^\beta)$.

For the general case $q(u) \sim c\exp(-\lambda u^\beta)$ as $u \to \infty$, for any $\varepsilon > 0$ we can choose u_0, such that for all $u > u_0$, $(1-\varepsilon)c\exp(-\lambda u^\beta) \leq q(u) \leq (1+\varepsilon)c\exp(-\lambda u^\beta)$. Using this in conjunction with the inequality $xe^{-xt} \leq (1/t)e^{-1}$, $x \geq 0$, and the case $q(u) = c\exp(-\lambda u^\beta)$, we obtain

$$\begin{aligned} g'_f(t) &\leq \frac{u_0 e^{-1}}{t} + \frac{1+\varepsilon}{1-\varepsilon} \int_0^\infty (1-\varepsilon)c\exp(-\lambda u^\beta)\exp(-t(1-\varepsilon)c\exp(-\lambda u^\beta))\, du \\ &\sim \frac{1+\varepsilon}{1-\varepsilon} \frac{(\log((1-\varepsilon)ct))^{1/\beta-1}}{t\beta\lambda^{1/\beta}} \quad \text{as } t \to \infty. \end{aligned}$$

Similarly, one can obtain the lower bound. Finally by passing $\varepsilon \to 0$ we obtain the conclusion of the theorem. □

PROOF OF LEMMA 4. By assuming that $q(u) = c\exp(-\lambda u^\beta)$, and by changing the variable of integration to $x = t\exp(-\lambda u^\beta)$ in (4.2), as in the proof of Lemma 3, we compute

$$f_f\left(\frac{t}{c}\right) = \frac{c^2}{t^2 \beta \lambda^{1/\beta}} \int_0^t xe^{-x} \left(\log\left(\frac{t}{x}\right)\right)^d dx, \tag{7.58}$$

where $d = (1/\beta)-1\ (> -1)$ as defined. Thus, to complete the proof it is enough to show that for $d > -1$,

$$\int_0^t xe^{-x}(\log(t/x))^d\,dx \sim (\log t)^d \quad \text{as } t \to \infty. \tag{7.59}$$

To finish this, let us decompose the integral above into three integrals:

$$\begin{aligned}\int_0^t xe^{-x}(\log(t/x))^d\,dx &= \int_0^{1/\log t} + \int_{1/\log t}^{\log t} + \int_{\log t}^t \\ &= I_1(t) + I_2(t) + I_3(t) \quad \text{as defined.}\end{aligned} \tag{7.60}$$

The arguments from here are exactly the same as in the proof of Lemma 3. We skip the details. □

PROOF OF LEMMA 5. It is easy to compute that

$$\begin{aligned}&\int_0^t \frac{1-e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx - \frac{(\log t)^{d+1}}{d+1} \\ &\quad = \int_0^1 \frac{1-e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx \\ &\qquad - \int_1^t \frac{e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx \\ &\quad = I_1(t) - I_2(t) \quad \text{as defined.}\end{aligned} \tag{7.61}$$

By changing the variable of integration to $u = t/x$ in $I_1(t)$ we obtain

$$I_1(t) = \int_t^\infty \frac{1-e^{-t/u}}{u}(\log u)^d\,du,$$

which can be decomposed in

$$\begin{aligned}I_1(t) &= \int_t^{t\log^2 t} \frac{1-e^{-t/u}}{u}(\log u)^d\,du + \int_{t\log^2 t}^\infty \frac{1-e^{-t/u}}{u}(\log u)^d\,du \\ &= I_{11}(t) + I_{12}(t) \quad \text{as defined.}\end{aligned} \tag{7.62}$$

Consider first the case $d \geq 0$. It is easy to see that

$$\begin{aligned}I_{11}(t) &\leq (\log(t\log^2 t))^d \int_t^{t\log^2 t} \frac{1-e^{-t/u}}{u}\,du \\ &= (\log(t\log^2 t))^d \int_{1/(\log^2 t)}^1 \frac{1-e^{-x}}{x}\,dx \\ &\sim (\log t)^d \int_0^1 \frac{1-e^{-x}}{x}\,dx \quad \text{as } t \to \infty.\end{aligned} \tag{7.63}$$

Similarly, one gets the lower bound

$$I_{11}(t) \geq (\log t)^d \int_{1/(t\log^2 t)}^{1} \frac{1-e^{-x}}{x}\,dx \sim (\log t)^d \int_0^1 \frac{1-e^{-x}}{x}\,dx \quad \text{as } t \to \infty. \tag{7.64}$$

Also, by using the inequality $1 - e^{-x} \leq x$, for $x \geq 0$, we obtain

$$\begin{aligned} I_{12}(t) &\leq t \int_{t\log^2 t}^{\infty} \frac{1}{u^2}(\log u)^d\,du \\ &= t \int_{\log(t\log^2 t)}^{\infty} x^d e^{-x}\,dx \\ &\sim t(\log(t\log^2 t))^d \exp(-\log(t\log^2 t)) \quad \text{as } t \to \infty \\ &= o((\log t)^d) \quad \text{as } t \to \infty; \end{aligned} \tag{7.65}$$

the asymptotics in (7.65) follow from [1], page 263, equation 6.5.32. Next, we investigate

$$\begin{aligned} I_2(t) &= \int_1^{\log t} \frac{e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx + \int_{\log t}^{t} \frac{e^{-x}}{x}\left(\log\left(\frac{t}{x}\right)\right)^d dx \\ &= I_{21}(t) + I_{22}(t) \quad \text{as defined.} \end{aligned} \tag{7.66}$$

Here, the asymptotic behavior of $I_{21}(t)$ is determined by

$$I_{21}(t) \leq (\log t)^d \int_1^{\log t} \frac{e^{-x}}{x}\,dx \sim (\log t)^d \int_1^{\infty} \frac{e^{-x}}{x}\,dx, \tag{7.67}$$

and

$$I_{21}(t) \geq \left(\log\left(\frac{t}{\log t}\right)\right)^d \int_1^{\log t} \frac{e^{-x}}{x}\,dx \sim (\log t)^d \int_1^{\infty} \frac{e^{-x}}{x}\,dx. \tag{7.68}$$

The estimate for $I_{22}(t)$ is given by

$$I_{22}(t) \leq \left(\log\left(\frac{t}{\log t}\right)\right)^d \frac{1}{t\log t} = o((\log t)^d) \quad \text{as } t \to \infty. \tag{7.69}$$

Finally, by combining equations (7.61)–(7.69), it follows that

$$\begin{aligned} I_1(t) - I_2(t) &\sim (\log t)^d \left(-\int_1^{\infty} \frac{e^{-x}}{x} + \int_0^1 \frac{1-e^{-x}}{x}\,dx \right) \quad \text{as } t \to \infty \\ &= \gamma(\log t)^d, \end{aligned} \tag{7.70}$$

where the last equality follows from [17], page 946, equation 8.367 (12). This completes the proof for the case $d \geq 0$.

For the case $-1 < d < 0$, the inequalities in (7.63) and (7.64) are going to be reversed, but the asymptotic behavior of I_{11} is still going to be the same. Since (7.65) still holds, $I_1(t)$ will behave asymptotically the same as for the

$d \geq 0$ case. Similarly, the inequalities in (7.67) and (7.68) are going to be reversed, but the asymptotic behavior of $I_{21}(t)$ is unchanged. The asymptotic upper bound for $I_{22}(t)$ is given by

$$\begin{aligned} I_{22}(t) &\leq \frac{1}{t}\int_{\log(t)}^{t} \frac{1}{x}\left(\log\left(\frac{t}{x}\right)\right)^{d} dx \\ &= \frac{1}{t}\int_{0}^{\log(t/\log t)} u^{d}\, du \\ &= \frac{1}{t}\frac{(\log(t/\log t))^{d+1}}{d+1} = o((\log t)^{d}) \quad \text{as } t \to \infty. \end{aligned}$$

This implies that $I_2(t)$ will have the same asymptotics as in the $d \geq 0$ case, and therefore (7.70) holds. This completes the proof of the lemma. □

Acknowledgments. I thank Jim Fill and Alan Weiss for reading and providing valuable comments on the original draft of this paper. I am also grateful to the anonymous reviewer for detailed proofreading.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A. eds. (1974). *Handbook of Mathematical Functions*. Dover, New York.

[2] BENTLEY, J. L. and MCGEOCH, C. C. (1985). Amortized analyses of self-organizing sequential search heuristics. *Comm. ACM* **28** 404–411.

[3] BINGHAM, N. H. and DONEY, R. A. (1974). Asymptotic properties of super-critical branching processes I: the Galton–Watson process. *Adv. Appl. Probab.* **6** 711–731.

[4] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. Cambridge Univ. Press.

[5] BITNER, J. R. (1979). Heuristics that monotonically organize data structures. *SIAM J. Comput.* **8** 82–110.

[6] BODELL, J. (1997). Cost of searching: Probabilistic analysis of the self-organizing move-to-front and move-to-root sorting rules. PhD thesis, Dept. Mathematics, Royal Institute of Technology, Stockholm.

[7] BORODIN, W. A., IRANI, S., RAGHAVAN, P. and SCHIEBER, B. (1995). Competitive paging with locality of reference. *J. Comput. System Sci.* **50** 244–258.

[8] BURVILLE, P. J. and KINGMAN, J. F. C. (1973). On a model for storage and search. *J. Appl. Probab.* **10** 697–701.

[9] CHUNG, F. R. K., HAJELA, D. J. and SEYMOUR, P. D. (1996). Self-organizing sequential search and Hilbert's inequalities. *J. Comput. System Sci.* **36** 148–157.

[10] CHUNG, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.

[11] COFFMAN, E. G. JR. and DENNING, P. J. (1973). *Operating Systems Theory*. Prentice-Hall, Englewood Cliffs, NJ.

[12] FILL, J. A. (1996). An exact formula for the move-to-front rule for self-organizing lists. *J. Theoret. Probab.* **9** 113–159.

[13] FILL, J. A. (1996). Limits and rate of convergence for the distribution of search cost under the move-to-front rule. *Theoret. Comput. Sci.* **164** 185–206.

[14] FILL, J. A and HOLST, L. (1996). On the distribution of search cost for the move-to-front rule. *Random Structures Algorithms* **8** 179.

[15] FLAJOLET, P., GARDY, D. and THIMONIER, L. (1992). Birthday paradox, coupon collector, caching algorithms and self-organizing search. *Discrete Appl. Math.* **39** 207–229.

[16] FRANASZEK, P. A. and WAGNER, T. J. (1974). On some distribution-free aspects of paging algorithm performance. *J. ACM* **21** 31–39.

[17] GRADSHTEYN, I. S. and RYZHIK, I. M. (1967). *Table of Integrals Series and Products*. Academic Press, New York.

[18] HENDRICKS, W. J. (1972). The stationary distribution of an interesting Markov chain. *J. Appl. Probab.* **9** 231–233.

[19] IRANI, S., KARLIN, A. R. and PHILIPS, S. (1996). Strongly competitive algorithms for paging with locality of reference. *SIAM J. Comput.* **25** 477–497.

[20] JELENKOVIĆ, P. R. and LAZAR, A. A. (1998). Subexponential asymptotics of a Markov-modulated random walk with queueing applications. *J. Appl. Probab.* **35** 325–347.

[21] JELENKOVIĆ, P. R. and LAZAR, A. A. (1999). Asymptotic results for multiplexing subexponential on-off processes. *Adv. in Appl. Probab.* **31**. To appear.

[22] KNUTH, D. E. (1973). *The Art of Computer Programming* **3**. *Sorting and Searching*. Addison-Wesley, Reading, MA.

[23] MCCABE, J. (1965). On serial files with relocatable records. *Oper. Res.* **13** 609–618.

[24] NELSON, P. R. (1977). Single-shelf library-type Markov chains with infinitely many books. *J. Appl. Probab.* **14** 298–308.

[25] PHATARFOD, R. M. (1991). On the matrix occuring in a linear search problem. *J. Appl. Probab.* **28** 336–346.

[26] PHATARFOD, R. M. (1994). On the transition probabilities of the move-to-front scheme. *J. Appl. Probab.* **31** 570–574.

[27] RIVEST, R. (1976). On self-organizing sequential search heuristics. *Comm. ACM* **19** 63–67.

[28] SLEATOR, D. D. and TARJAN, R. E. (1985). Amortized efficiency of list update and paging rules. *Comm. ACM* **28** 202–208.

DEPARTMENT OF ELECTRICAL ENGINEERING
COLUMBIA UNIVERSITY
510 WEST 120TH STREET, MC 4712
NEW YORK, NEW YORK 10027
E-MAIL: predrag@ee.columbia.edu