Optimal Robot Scheduling for Web Search Engines

E. G. Coffman, Jr.
Bell Labs, Lucent Technologies
Murray Hill, NJ 07974
USA
egc@research.bell-labs.com

Zhen Liu
INRIA
06902 Sophia-Antipolis
France
Zhen.Liu@sophia.inria.fr

Richard R. Weber
Statistical Laboratory
Cambridge University
England
R.R.Weber@statslab.cam.ac.uk

June 27, 1999

Abstract

A robot is deployed by a Web search engine in order to maintain the currency of its data base of Web pages. This paper studies robot scheduling policies that minimize the fractions r_i of time pages spend out-of-date, assuming independent Poisson page-change processes, and a general distribution for the page access time X. We show that, if X is decreased in the increasing-convex ordering sense, then r_i is decreased for all i under any scheduling policy, and that, in order to minimize expected total obsolescence time of any page, the accesses to that page should be as evenly spaced in time as possible.

We then investigate the problem of scheduling to minimize the cost function $\sum c_i r_i$, where the c_i are given weights proportional to the page-change rates μ_i. We give a tight bound on the performance of such a policy and prove that the optimal frequency at which the robot should access page i is proportional to $\ln(h_i)^{-1}$, where $h_i := \mathrm{E}e^{-\mu_i X}$. Note that this reduces to being proportional to μ_i when X is a constant, but not, as one might expect, when X has a general distribution.

Next, we evaluate randomized accessing policies whereby the choices of page access are determined by independent random samples from the distribution $\{f_i\}$. We show that when the weights c_i in the cost function are proportional to μ_i, the minimum cost is achieved when f_i is proportional to $(h_i)^{-1} - 1$. Finally, we present and analyze a heuristic policy that is especially suited to the asymptotic regime of large data bases.

Keywords: Web Search Engines, Web Robots, Spiders, Stochastic Scheduling, Stochastic Ordering.

1 Introduction

The Web's role as a major information publishing and retrieving mechanism on the Internet continues to grow extremely fast. Indeed, the amount of information on the Web has long since become too large for manually browsing through any significant portion of its hypertext structure. Instead, the Web offers search engines for this purpose; Alta Vista, Lycos, Infoseek, Magellan, Excite, and Yahoo are but a few of those available. These systems consist of indexing engines for constructing a data base of Web pages, and in many cases **robots** for bringing information to the indexing engine. To maintain currency and completeness of the data base, robots periodically make recursive traversals of the Web's hypertext structure by accessing pages, then the pages referenced by these pages, and so on. There are currently well over 150 robots on the Web. Most of them, including the ones of interest here, are designed for resource discovery, but others are used for mirroring, maintenance (link checking), statistics, etc. For an extensive discussion of Web robots, see the work of Martijn Koster at

`http:/info.webcrawler.com/mak/projects/robots/robots.html`

and for a recent discussion of page change rates, see [8]. In the literature one finds more colorful terms for robot, such as wanderer, crawler or spider, and the notion of a robot being 'routed to' or 'visiting' a page. This paper keeps with the 'robot' and 'accessing' terminology throughout.

We study problems of scheduling a robot that maintains the currency of existing pages in search-engine data bases. We assume that the set of data-base pages is fixed, but as we shall see, our results can be promoted as heuristics for data bases that acquire new pages and drop old pages over time. A specific objective will be to find robot schedules that minimize the obsolescence of the data base in some useful sense. For example, assume there are N Web pages, labeled $1, 2, \ldots, N$, which are to be accessed repeatedly by a robot, the duration of each access being an independent sample from a given distribution. Assume also that the contents of page i are modified at times that follow a Poisson process with parameter μ_i. A page is considered up-to-date by the indexing engine from the time it is accessed by the robot until the next time it is modified, at which point it becomes out-of-date until the robot's next access. Let r_i be the fraction of time page i spends out-of-date. The problem is to find relative page-access frequencies and a sequencing policy that realizes these frequencies such that the objective function $C = \sum_{1\le i\le N} c_i r_i$, is minimized, where the c_i are given weights. Under simplifying but plausible assumptions on the weights, page access times, and the class of allowed policies, we obtain explicit solutions to this problem.

From a theoretical point of view, our problem is closely related to those multiple-queue single-server systems usually called *polling systems* in the queueing literature. Indeed, the robot can be considered as the server and the pages as the stations in the polling system. The durations of consecutive page accesses correspond to switch-over times and the page modifications correspond to customer arrivals. The service times in this polling system are zero. Our two-stage approach of optimizing robot schedules (determining access frequencies and then finding a schedule that realizes them) is similar to the approach in [5]-[7] of optimizing visit sequences in polling systems.

An extensive literature exists on the analysis and control of polling systems. The interested reader is referred to the book of Takagi [17, 18] for general references; the special issue of the journal *Queueing Systems*, Vol. 11 (1992) on polling models and the recent thesis of Borst [4] can be consulted for more recent developments. In particular, the polling systems with zero service times were motivated by communication networks such as *teletext* and *videotex* where pages of information are to be broadcast to terminals connected to a computer network [1, 9, 14]. However, the problem here has not been analyzed. Indeed, in the usual analysis of polling systems with unbounded buffers, interest centers on mean waiting times and mean queue lengths, whereas in our problem, the performance measure of interest, viz., the obsolescence time, corresponds to the maximum waiting time of a customer during a visit cycle of the server. An alternative view of our model identifies it with a polling loss system having unit buffers, in which our obsolescence time becomes the waiting time. With this point of view, our model has potential use in mainenance applications.

The next section is devoted to a precise formulation of our model, and a review of some useful concepts in stochastic ordering theory. Section 3 begins by proving two properties of robot scheduling policies: *(i)* expected obsolescence times increase as the page-access time increases in the increasing-convex-ordering sense, and *(ii)*, by Schur-convexity results, accesses to any given page should be as evenly spaced as possible. We then derive a tight lower bound on the cost function C assuming that the weights c_i are proportional to the μ_i. These results yield a formula for optimal access frequencies. Our techniques can be extended to general c_i, but explicit formulas are not attainable in general.

To motivate the assumption on weights, note that a useful choice for the c_i is the customer page-access frequency, for in this case the total cost can be regarded as a customer total error rate. The special case where the customer access frequency c_i is proportional to the page-change rate μ is reasonable under this interpretation - the greater the interest (access frequency), the greater the frequency of page modification.

Sections 4 and 5 deal with the problem of sequencing page accesses optimally, or near optimally, so as to realize a given set of access frequencies. This material is prefaced by a discussion at the end of Section 3 which relates our scheduling problem to those that come under the heading of generalized round-robin or template-driven scheduling.

In Section 4, we introduce randomized page accessing, where each access is determined by an i.i.d. sample from a distribution $\{f_i\}$. We show how to find that choice for this distribution which minimizes C. In Section 5, we develop a policy that performs well when N is large. It is based on work of Itai and Rosberg [12] (in an entirely different setting) and yields a cost within 5% of optimal. Some directions for further research are mentioned in Section 6, which concludes the paper.

2 Preliminaries

Let $\{X_k\}$ be the sequence of durations of consecutive page accesses by the robot, each X_k being distributed independently as a random variable X. For scheduling policy π, let $\pi_n \in \{1, 2, \ldots, N\}$ be the scheduling decision for the n-th access, i.e., the index of the n-th page to be accessed by the robot under π. Define the *interaccess distance* $d^i_j(\pi) = n^i_j(\pi) - n^i_{j-1}(\pi)$, where $n^i_j(\pi)$ is the index of the j-th access of page i, i.e., $n^i_j(\pi) = \inf\{n > n^i_{j-1}(\pi) \mid \pi_n = i\}$, and where $n^i_0(\pi) \equiv 0$. Let $X^i_j = X^i_j(\pi)$ be the j-th *interaccess time* of page i, i.e., the time between the $(j-1)$-st and j-th page-i access completion times. We have $X^i_j = \sum_{k=n^i_{j-1}+1}^{n^i_j} X_k$, so the random variables X^i_j are mutually independent. Note that, if page access times X_k are exponentially distributed, then X^i_j has an Erlang distribution of d^i_j stages.

Hereafter, except in definitions, the policy π will normally be omitted from our notation; in such cases, the policy will always be clear in context.

Let $Z^i_j = Z^i_j(\pi)$ be the time that page i is out-of-date during the j-th interaccess time of page i. Let $m^i_n = m^i_n(\pi)$ be the number of accesses of page i among the first n accesses: $m^i_n = \sum_{k=1}^n \mathbf{1}\{\pi_k = i\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. Hereafter, we consider only stationary scheduling policies in the sense that, for each such policy, the limit

$$f_i = f_i(\pi) = \lim_{n\to\infty} \frac{m^i_n}{n} \tag{1}$$

exists and is strictly positive for all i, $1 \leq i \leq N$. We call f_i the access frequency of page i. We also require that the limits $\lim_{n\to\infty} \sum_{j=1}^n Z^i_j/n$ and $\lim_{n\to\infty} \sum_{j=1}^n E[Z^i_j]/n$ exist and be equal. These last assumptions hold under fairly mild conditions, e.g., when the sequence $\{d^i_j(\pi)\}_j$ is stationary and ergodic (cf. Kingman [13]).

The *obsolescence rate* $r_i = r_i(\pi)$ of page i is the limiting fraction of time that page i is out of date; precisely, it is defined as

$$r_i = \lim_{n\to\infty} \frac{\sum_{j=1}^{m^i_n} Z^i_j}{\sum_{j=1}^{m^i_n} X^i_j} = \frac{\lim\limits_{n\to\infty} \dfrac{\sum_{j=1}^{m^i_n} Z^i_j}{n}}{\lim\limits_{n\to\infty} \dfrac{\sum_{j=1}^{m^i_n} X^i_j}{n}} = \frac{1}{E[X]} \cdot \lim_{n\to\infty} \frac{\sum_{j=1}^{m^i_n} E[Z^i_j]}{n} \tag{2}$$

In particular, when policy π is cyclic with cycle length K, i.e., when $\pi_{nK+k} = \pi_{(n-1)K+k}$ for all $1 \leq k \leq K$ and all $n = 1, 2, \ldots$, then

$$r_i = \frac{1}{KE[X]} \sum_{j=1}^{m^i_K} E[Z^i_j], \tag{3}$$

where m^i_K is the number of page-i accesses during a cycle. The cost function to be minimized is the weighted sum of the obsolescence rates:

$$C = C(\pi) = \sum_{i=1}^N c_i \; r_i, \tag{4}$$

where c_i are given positive real numbers and the minimization is to be over all stationary scheduling policies.

A few basics in stochastic ordering conclude this section. For two m-dimensional real vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}$ majorizes $\mathbf{y}$, written $\mathbf{x} \succ \mathbf{y}$, if $\sum_{i=1}^{k} x_{[i]} \geq \sum_{i=1}^{k} y_{[i]}$, for $k = 1, \ldots, m-1$ and $\sum_{i=1}^{m} x_{[i]} = \sum_{i=1}^{m} y_{[i]}$, where $x_{[i]}$ is the i^{th} largest component of $\mathbf{x}$. Intuitively, $\mathbf{y}$ is better balanced than $\mathbf{x}$. A function h is said to be *Schur-convex* if $h(\mathbf{x}) \geq h(\mathbf{y})$ whenever $\mathbf{x} \succ \mathbf{y}$. See [15] for more details about this and related properties.

A random variable Y_1 is said to be no greater than a random variable Y_2 in the convex ordering sense, denoted $Y_1 \leq_{cx} Y_2$, if $E[h(Y_1)] \leq E[h(Y_2)]$ for all convex functions h, provided the expectations exist. If in this definition 'convex' is replaced everywhere by 'increasing and convex,' then we write $Y_1 \leq_{icx} Y_2$. As is easily verified, $Y_1 \leq_{cx} Y_2$ implies that Y_1 has the same mean but smaller variance than Y_2. It is also easy to see that $Y_1 \leq_{cx} Y_2$ implies $Y_1 \leq_{icx} Y_2$. See [16] for equivalent definitions and further properties.

3 Schur Convexity and a Lower Bound

Recall that a page is considered out-of-date from the time it is modified until the next time it is accessed by the robot. Thus, if page i is not modified during its j-th interaccess interval, then the obsolescence time is $Z_j^i = 0$. Otherwise, Z_j^i is the time that elapses from the first moment page i is modified during its j-th interaccess interval until the end of that interval. Recall also that the modification (or mutation) epochs of page i follow a Poisson process with parameter μ_i. By the memoryless property of the Poisson process, the time that elapses from the beginning of page i's j-th interaccess interval to the first subsequent mutation has an exponential distribution with parameter μ_i. Let $R_1^i, R_2^i, \ldots$ be an i.i.d. sequence of such random variables, so that

$$Z_j^i \stackrel{d}{=} \left(X_j^i - R_j^i\right)^+, \tag{5}$$

where x^+ denotes $\max(x, 0)$ and $\stackrel{d}{=}$ denotes equality in distribution.

As an immediate consequence, we obtain

Theorem 3.1 *If the page access time is decreased in the increasing convex ordering sense, then the obsolescence rate is decreased for all pages under any scheduling policy.*

Proof. Let $\{X'_k\}$ be a sequence of access times distributed independently as X', and define $\{X'^i_j\}_j$ and $\{Z'^i_j\}_j$ as for X_k. Assume that $X' \leq_{icx} X$. Then,

$$Z'^i_j \stackrel{d}{=} \left(X'^i_j - R^i_j\right)^+ \leq_{icx} \left(X^i_j - R^i_j\right)^+ \stackrel{d}{=} Z^i_j,$$

and so $E[Z'^i_j] \leq E[Z^i_j]$. Thus,

$$r'_i = \frac{1}{E[X']} \cdot \lim_{n\to\infty} \frac{\sum_{j=1}^{m^i_n} E[Z'^i_j]}{n} \leq \frac{1}{E[X]} \cdot \lim_{n\to\infty} \frac{\sum_{j=1}^{m^i_n} E[Z^i_j]}{n} = r_i,$$

ė

as desired. ∎

$\int^\infty$ Rd$^{\mu}$−

Returning to our main problem, where the distribution of page access times is assumed given, we now show that the obsolescence rate is a Schur convex function of the vector of interaccess distances. For this, we need the following calculation which will also be useful for later results. Define $h_i = E[e^{-\mu_i X}]$, the Laplace transform of X evaluated at μ_i.

Lemma 3.1 *For any page i,*

$$E[Z^i_j] = d^i_j E[X] - \frac{1}{\mu_i}\left(1 - h_i^{d^i_j}\right).$$

Proof. Let G^i_j be the probability distribution of X^i_j. We have from (5) that

$$\begin{aligned} E[Z^i_j] &= \int_0^\infty P(Z^i_j > z)dz \\ &= \int_0^\infty P(X^i_j - R^i_j > z)dz \\ &= \int_0^\infty \int^\infty \end{aligned}$$

Rd μ_i −

Rd μ_i −

which yields the lemma. ∎

We can conclude from the above proof that the result of Theorem 3.1 still holds when the increasing convex ordering is replaced by the weaker Laplace-transform ordering (see [16]). It follows from a result of Schur (cf. [15, Proposition 3.C.1, page 64]) and Lemma 3.1 that

Theorem 3.2 *For any fixed number n of page-i accesses, the expected total obsolescence time of page i, $\sum_{i=1}^{n} E[Z_j^i]$ is a Schur convex function of the distances d_j^i, $j = 1, \ldots, n$.*

Thus, in order to minimize the expected obsolescence time, the accesses to any particular page should be as evenly spaced as possible.

An algorithm that computes a schedule of the robot that implements a given set of access frequencies in the sense of (1) is called an *accessing* policy. In these terms, the scheduling policies proposed in this paper consist of two stages; the first computes a set of access frequencies $\{f_i\}$ and the second is an accessing policy that implements $\{f_i\}$. The even-spacing objective of accessing policies yields a lower bound, as follows.

Theorem 3.3 *The obsolescence rate under any accessing policy implementing the access frequencies $\{f_i\}$ satisfies for each i,*

$$r_i \geq \frac{1}{E[X]} \left(E[X] - \frac{f_i}{\mu_i} + \frac{f_i}{\mu_i} h_i^{1/f_i} \right).$$

Proof.

$$\begin{aligned}
r_i &= \frac{1}{E[X]} \cdot \lim_{n\to\infty} \frac{\sum_{j=1}^{m_n^i} E[Z_j^i]}{n} \\
&= \frac{1}{E[X]} \cdot \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{m_n^i} \left(d_j^i E[X] - \frac{1}{\mu_i} \left(1 - h_i^{d_j^i} \right) \right) \\
&= \frac{1}{E[X]} \cdot \lim_{n\to\infty} \frac{1}{n} \left(nE[X] - \frac{m_n^i}{\mu_i} + \frac{1}{\mu_i} \sum_{j=1}^{m_n^i} h_i^{d_j^i} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{E[X]}\left(E[X] - \frac{f_i}{\mu_i} + \frac{1}{\mu_i}\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{m_n^i} h_i^{d_j^i}\right) \\
&\geq \frac{1}{E[X]}\left(E[X] - \frac{f_i}{\mu_i} + \frac{1}{\mu_i}\lim_{n\to\infty}\frac{m_n^i}{n} h_i^{n/m_n^i}\right) \\
&= \frac{1}{E[X]}\left(E[X] - \frac{f_i}{\mu_i} + \frac{f_i}{\mu_i} h_i^{1/f_i}\right),
\end{aligned}
$$

where the inequality comes from the Schur convexity of $\sum_{j=1}^{m_n^i} h_i^{d_j^i}$ in the d_j^i's (cf. Theorem 3.2).

■

The above lower bound can be achieved only in special cases. For instance, if the frequencies are all equal, the policy that accesses pages $1, 2, \ldots, N$ cyclically yields this optimal obsolescence rate. Another example where we can find a feasible accessing policy achieving the lower bound is when the frequencies are of the form $f_i = 1/2^{k_i}$, where k_i is an integer for every i. We return to the general case after considering the cost-minimization theorem. The proof of the following theorem gives a solution technique applicable to general weights c_i and shows that the technique leads to explicit results in an interesting special case.

Theorem 3.4 *Assume that the weights in the cost function are proportional to the mutation rates of the pages, i.e., $c_i = c_0 \mu_i$ for all $i = 1, 2, \ldots, N$. Then for any scheduling policy,*

$$
C = c_0 \cdot \sum_{i=1}^{N} \mu_i r_i \geq c_0 \left(\mu - \frac{1}{E[X]} + \frac{1}{E[X]}\prod_{i=1}^{N} h_i\right) > 0, \tag{6}
$$

where $\mu = \sum_{i=1}^{N} \mu_i$.

Proof. For the moment, let the c_i be general. Following Theorem 3.3, we have $C \geq C^*$, where C^* is the solution to the following optimization problem:

$$
C^* = \min \sum_{i=1}^{N} c_i \left(1 - \frac{1}{E[X]\mu_i} x_i + \frac{1}{E[X]\mu_i} x_i h_i^{1/x_i}\right) \tag{7}
$$

subject to $x_i \geq 0$ and

$$\sum_{i=1}^{N} x_i = 1.$$

To solve the above problem, we use Lagrange multipliers and define

$$\mathcal{L}(x_1, \ldots, x_N, \lambda) = \sum_{i=1}^{N} c_i \left(1 - \frac{1}{E[X]\mu_i} x_i + \frac{1}{E[X]\mu_i} x_i h_i^{1/x_i}\right) + \lambda \left(\sum_{i=1}^{N} x_i - 1\right).$$

By the convexity of the function

$$\sum_{i=1}^{N} c_i \left(1 - \frac{1}{\mu_i E[X]} x_i + \frac{1}{\mu_i E[X]} x_i h_i^{1/x_i}\right)$$

in the vector $(x_1, \ldots, x_N)$, the solution satisfies the necessary and sufficient condition

$$\frac{\partial \mathcal{L}}{\partial x_i} = -\frac{c_i}{\mu_i E[X]} \left(1 - h_i^{1/x_i} + \frac{\ln h_i}{x_i} h_i^{1/x_i}\right) + \lambda = 0, \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{N} x_i - 1 = 0. \tag{9}$$

Observe that $h_i < 1$, so that $h_i^{1/x_i} < 1$. One can easily check that the function $1 - y + y \ln y$ is strictly decreasing in y for $y < 1$. Thus, under the assumption that c_i is proportional to μ_i, we conclude from (8) that all h_i^{1/x_i} are identical and that the minimum is achieved, by (9), when

$$x_i = \frac{\ln h_i}{\sum_{i=1}^{N} \ln h_i} = \frac{\ln(h_i)^{-1}}{\sum_{i=1}^{N} \ln(h_i)^{-1}}. \tag{10}$$

This solution is positive so it is also the solution to the minimization problem in (7). Hence,

$$\begin{aligned} C^* &= \sum_{i=1}^{N} c_0 \mu_i - \frac{c_0}{E[X]} \sum_{i=1}^{N} x_i + \frac{c_0}{E[X]} \sum_{i=1}^{N} x_i h_i^{1/x_i} \\ &= c_0 \left(\mu - \frac{1}{E[X]} + \frac{1}{E[X]} \exp\left\{\sum_{i=1}^{N} \ln h_i\right\}\right) = c_0 \left(\mu - \frac{1}{E[X]} + \frac{1}{E[X]} \prod_{i=1}^{N} h_i\right). \end{aligned}$$

Note that

$$\prod_{i=1}^{N} h_i = E\left[\exp\left\{-\sum_{i=1}^{N} \mu_i X_i\right\}\right] > 1 - E\left[\sum_{i=1}^{N} \mu_i X_i\right] = 1 - \mu E[X],$$

so

$$\mu - \frac{1}{E[X]} + \frac{1}{E[X]} \prod_{i=1}^{N} h_i > 0,$$

and the proof is complete. ∎

When the weights in the cost function are not proportional to the mutation rates of pages, one can still use Lagrange multipliers to solve the optimization problem. As noted earlier, however, we do not have closed-form solutions in general.

It is also worthwhile noticing that the optimal access frequencies (cf. (10)) in the above lower bound are not necessarily proportional to the page mutation rates μ_i, a fact that has emerged in the context of other polling systems (see, e.g., [5]-[7]). Rather, they are proportional to $\ln(h_i)^{-1} = \ln\left(E[e^{-\mu_i X}]\right)^{-1}$. Proportionality to the μ_i occurs only when X is a constant. Note also that the magnitude of the difference between $\mu_i E[X]$ and $\ln\left(E[e^{-\mu_i X}]\right)^{-1}$ is large if $Var(X)$ is large (or X is large in the convex ordering sense).

To summarize, the results of this section show that, if the weights in the cost function are proportional to the mutation rates of the pages, then an accessing policy that comes close to the lower bound in Theorem 3.3 with the f_i nearly proportional to the $\ln(h_i)^{-1}$ will come close to minimizing C.

Finding good accessing policies that realize a given set of access frequencies is the subject of the next two sections. In Section 5, we develop an optimal randomized accessing policy, and in Section 6, we adapt the well-studied golden-ratio policy to our problem, primarily as a candidate for good asymptotic performance; we will see that this policy gives an obsolescence rate within 5% of the lower bound, in the limit of large N.

We remark that this problem is closely related to the design and analysis of polling/splitting sequences in the context of queueing (and in particular, communication) systems [2, 3, 5, 6, 7], where algorithms are described as template driven or generalized round robin. With future research in mind, we note that these studies suggest other approaches worth investigating, e.g., extensions of the mathematical programming techniques in [5] and the algorithms [3] derived from Hajek's [10] results on regular binary sequences. Although the latter lack the established performance bounds of the golden-ratio policy, simulations in the earlier queueing models show they are superior algorithms. Thus, they make promising candidates for our page-accessing model.

4 Randomized Accessing and Its Optimal Solution

Let $f_1, f_2, \ldots, f_N$ be given access frequencies. According to the randomized scheduling policy, at each decision point, the robot chooses to access page i with probability f_i; the decision is made independently of all previous decisions. One can easily see that $\{d_j^i\}_j$, $\{X_j^i\}_j$ and $\{Z_j^i\}_j$ are three sequences of i.i.d. random variables for all i. Moreover, d_j^i has a geometric distribution: $P(d_j^i = n) = f_i(1-f_i)^{n-1}$. Thus, we have

Lemma 4.1 *For given frequencies $f_1, f_2, \ldots, f_N$,*

$$r_i = \frac{1}{E[X]} \left(E[X] - \frac{f_i}{\mu_i} + \frac{f_i}{\mu_i} \cdot \frac{f_i h_i}{1 - h_i + f_i h_i} \right).$$

Proof. As d_j^i has a geometric distribution, we obtain

$$E[X_j^i] = \sum_{n=1}^{\infty} f_i(1-f_i)^{n-1} n E[X] = \frac{E[X]}{f_i},$$

and by Lemma 3.1 we have,

$$E[Z_j^i] = \sum_{n=1}^{\infty} f_i(1-f_i)^{n-1} \left(nE[X] - \frac{1}{\mu_i} + \frac{1}{\mu_i} h_i^n \right) = \frac{E[X]}{f_i} - \frac{1}{\mu_i} + \frac{1}{\mu_i} \frac{f_i h_i}{1 - h_i + f_i h_i},$$

so elementary renewal theory and (2) imply

$$r_i(\rho) = \frac{E[Z_j^i(\rho)]}{E[X_j^i(\rho)]} = \frac{1}{E[X]} \left(E[X] - \frac{f_i}{\mu_i} + \frac{f_i}{\mu_i} \cdot \frac{f_i h_i}{1 - h_i + f_i h_i} \right).$$

$\blacksquare$

It is interesting to compare the lower bound of Theorem 3.3 with the obsolescence rate of the randomized policy. One can see that when f_i is small (close to 0) or large (close to 1), the difference between r_i and the lower bound tends to 0. More precisely, this difference is

$$\frac{h_i}{\mu_i E[X]} f_i^2 + o((f_i)^2)$$

when f_i goes to 0, and is

$$\frac{h_i}{\mu_i E[X]}(1 + \ln h_i)(f_i - 1) + o(1 - f_i)$$

when f_i goes to 1.

We now consider the problem of finding the optimal access frequencies under the randomized policy. First, we have the following lower bound over all frequencies.

Theorem 4.1 *Assume that the weights in the cost function are proportional to the mutation rates of the pages, i.e., $c_i = c_0 \mu_i$ for all $i = 1, 2, \ldots, N$. Then*

$$C = c_0 \cdot \sum_{i=1}^{N} \mu_i r_i \geq c_0 \left(\mu - \frac{1}{E[X]} \cdot \frac{\sum_{i=1}^{N} (h_i^{-1} - 1)}{1 + \sum_{i=1}^{N} (h_i^{-1} - 1)} \right). \tag{11}$$

Moreover, this lower bound is achieved when the access frequencies are proportional to $h_i^{-1} - 1$.

Proof. Following Lemma 4.1, we have

$$r_i = \frac{1}{E[X]} \left(E[X] - \frac{f_i}{\mu_i} + \frac{f_i}{\mu_i} \cdot \frac{f_i h_i}{1 - h_i + f_i h_i} \right),$$

so that $C \geq C^*$ where C^* is the solution to the following optimization problem:

$$C^* = \min \ \sum_{i=1}^{N} c_i - \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i x_i}{\mu_i} + \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i}{\mu_i} \cdot \frac{x_i^2 h_i}{1 - h_i + x_i h_i}, \tag{12}$$

subject to $x_i \geq 0$ and

$$\sum_{i=1}^{N} x_i = 1.$$

Again, we use Lagrange multipliers to solve the above problem. Define

$$\mathcal{L}(x_1, \ldots, x_N, \lambda) = \sum_{i=1}^{N} c_i - \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i x_i}{\mu_i} + \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i}{\mu_i} \cdot \frac{x_i^2 h_i}{1 - h_i + x_i h_i} + \lambda \left(\sum_{i=1}^{N} x_i - 1 \right).$$

By the convexity of the function

$$\sum_{i=1}^{N} c_i - \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i x_i}{\mu_i} + \frac{1}{E[X]} \sum_{i=1}^{N} \frac{c_i}{\mu_i} \cdot \frac{x_i^2 h_i}{1 - h_i + x_i h_i}$$

in $(x_1, \ldots, x_N)$, the solution satisfies the necessary and sufficient condition

$$\frac{\partial \mathcal{L}}{\partial x_i} = -\frac{c_i}{\mu_i E[X]} + \frac{c_i}{\mu_i E[X]} \cdot \frac{2h_i(1-h_i)x_i + h_i^2 x_i^2}{(1 - h_i + h_i x_i)^2} + \lambda = 0, \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{N} x_i - 1 = 0. \tag{14}$$

Equation (13) can be rewritten as

$$\lambda \mu_i (1 - h_i + h_i x_i)^2 = \frac{c_i}{E[X]} (1 - h_i)^2.$$

Since $h_i < 1$, we have

$$\frac{1 - h_i + h_i x_i}{1 - h_i} = \sqrt{\frac{c_i}{\lambda \mu_i E[X]}},$$

so

$$x_i = \left(\sqrt{\frac{c_i}{\lambda \mu_i E[X]}} - 1 \right) \frac{1 - h_i}{h_i}. \tag{15}$$

Now use the assumption that $c_i = c_0 \mu_i$ and obtain from (14) and (15) that

$$x_i = \frac{h_i^{-1} - 1}{\sum_{i=1}^{N} (h_i^{-1} - 1)}. \tag{16}$$

This solution is positive, so it is also the solution to the minimization problem of (12). Hence,

$$C^* = c_0\mu - \frac{c_0}{E[X]}\sum_{i=1}^{N} x_i \cdot \frac{1-h_i}{1-h_i+x_ih_i} = c_0\mu - \frac{c_0}{E[X]} \cdot \frac{\sum_{i=1}^{N}(h_i^{-1}-1)}{1+\sum_{i=1}^{N}(h_i^{-1}-1)},$$

which is the desired lower bound. ∎

Note that if the weights in the cost function are not proportional to the mutation rates of the pages, one can still use (14) and (15) to solve the problem. However, the solution is valid only when the right-hand side of (15) is positive. By solving for λ using (14) and (15) we obtain

$$\sqrt{\lambda} = \frac{1}{\sqrt{E[X]}} \cdot \frac{\sum_{i=1}^{N}\sqrt{\frac{c_i}{\mu_i}}\left(h_i^{-1}-1\right)}{1+\sum_{i=1}^{N}\left(h_i^{-1}-1\right)}$$

so the solution (15) is valid if

$$\min_{1\le i\le N}\sqrt{\frac{c_i}{\mu_i}} \ge \frac{\sum_{i=1}^{N}\sqrt{\frac{c_i}{\mu_i}}\left(h_i^{-1}-1\right)}{1+\sum_{i=1}^{N}\left(h_i^{-1}-1\right)}.$$

5 Asymptotic Optimality and the Golden Ratio Policy

In this section we consider the asymptotic large-N behavior of scheduling policies. A similar study was carried out by Itai and Rosberg [12] in the context of the control of a multiple-access channel. Some of our results here are analogous to theirs.

We define asymptotically optimal policies with respect to the lower bound in Theorem 3.4. Hence, we assume throughout this section that the weights in the cost function are proportional to the mutation rates of the pages, i.e., $c_i = c_0\mu_i$ for all $i = 1, 2, \ldots, N$. We say that a policy π is asymptotically optimal if

$$\lim_{N\to\infty} C(\pi) - C^* = 0.$$

Note first that if the total mutation rate μ tends to zero, then all *cyclic* policies are asymptotically optimal. Indeed, consider an arbitrary cyclic policy with cycle length K. It follows from (3) and Lemma 3.1 that

$$
\begin{aligned}
C &= c_0 \cdot \sum_{i=1}^{N} \mu_i r_i \\
&= \frac{c_0}{KE[X]} \cdot \sum_{i=1}^{N} \mu_i \sum_{j=1}^{m_K^i} \left\{ d_j^i E[X] - \frac{1}{\mu_i} \left(1 - h_i^{d_j^i} \right) \right\} \\
&= c_0 \mu - \frac{c_0}{KE[X]} \cdot \sum_{i=1}^{N} \sum_{j=1}^{m_K^i} \left(1 - (1 - d_j^i \mu_i E[X] + O(\mu_i^2)) \right) \\
&= c_0 \mu - \frac{c_0}{KE[X]} \cdot \sum_{i=1}^{N} \sum_{j=1}^{m_K^i} \left(d_j^i \mu_i E[X] + O(\mu_i^2) \right) \\
&= O(\mu^2),
\end{aligned}
$$

so if $\mu \to 0$, then $C \to 0$.

Thus, we assume that when $N \to \infty$, the total mutation rate μ, as well as the expected access time $E[X]$, is fixed. However, for any i, $1 \le i \le N$, we have $\mu_i \to 0$ when $N \to \infty$. Under such assumptions, the lower bound C^* in [illegible]

$i=1$

accesses in each cycle of $\gamma(k,N)$; these numbers satisfy

$$\lfloor f_i F_k \rfloor \leq M^i_{k,N} \leq \lceil f_i F_k \rceil$$

and $\sum_{i=1}^N M^i_{k,N} = F_k$, where f_i are the optimal access frequencies given by (10):

$$f_i = \frac{\ln h_i}{\sum_{i=1}^N \ln h_i}.$$

Thus,

$$\lim_{k\to\infty} \frac{M^i_{k,N}}{F_k} = f_i.$$

Let $frac(y) = y - \lfloor y \rfloor$ be the fractional part of y, and $A_k = \{frac(j\phi^{-1}) \mid j = 0, 1, \ldots, F_k - 1\}$. The s-th access of the robot is identified with the s-th smallest point of A_k. In the golden ratio policy $\gamma(k,N)$, the points

$$\left\{ frac(j\phi^{-1}) \mid \sum_{m=1}^{i-1} M^m_{k,N} \leq j < \sum_{m=1}^{i} M^m_{k,N} \right\}$$

correspond to the accesses of page i. As an example, let us suppose $N = 4$ and $f_1 = 2/13$, $f_2 = 3/13$, $f_3 = 3/13$ and $f_4 = 5/13$. Let $k = 8$ so that $F_k = 13$. Then, the golden ratio policy $\gamma(k,N)$ defines the access sequence $\{4, 2, 4, 1, 3, 4, 2, 4, 1, 3, 4, 2, 3\}$.

Thus, again from (3) and Lemma 3.1,

$$C(\gamma(k,N)) = c_0\mu - \frac{c_0}{F_k E[X]} \sum_{i=1}^N \sum_{m=1}^{M^i_{k,N}} \left(1 - h_i^{d^i_m}\right),$$

where the interaccess distance $d^i_m \in \{F_{j_i}, F_{j_i+1}, F_{j_i+2}\}$, where $j_i = \lceil \ln_\phi f_i \rceil$ (cf. [12]). Moreover, it can be shown by mimicking the proofs in [12] that

$$\begin{aligned} C(\gamma(N)) &:= \lim_{k\to\infty} C(\gamma(k,N)) \\ &= c_0\mu - \frac{c_0}{E[X]} \Bigg\{ 1 - \sum_{i=1}^N \Big[\left(f_i - \phi^{-j_i}\right) h_i^{F_{j_i}} \\ &\qquad + \left(f_i - \phi^{-j_i - 1}\right) h_i^{F_{j_i+1}} - \left(f_i - \phi^{-j_i+1}\right) h_i^{F_{j_i+2}} \Big] \Bigg\}. \end{aligned}$$

Theorem 5.1 *Assume for all i that $\mu_i \to 0$ as $N \to \infty$ and that $\sum_{i=1}^N \mu_i = \mu > 0$. Then,*

$$\limsup_{N\to\infty} C(\gamma(N)) \le c_0 \left\{ \mu - \frac{1}{E[X]} + \frac{1-\phi^{-1}}{E[X]} e^{-\frac{\mu\phi}{\sqrt{5}}E[X]} + \frac{\phi^{-1}}{E[X]} e^{-\frac{\mu\phi^2}{\sqrt{5}}E[X]} \right\} . \tag{18}$$

Proof. By mimicking the proof of Theorem 5.3 in [12], we can show that

$$C(\gamma(N)) \le c_0\mu - \frac{c_0}{E[X]} \left\{ 1 - \sum_{i=1}^N f_i \left[\left(1-\phi^{-1}\right) t_{i,N}^{\phi} + \phi^{-1} t_{i,N}^{\phi^2} \right] \right\} ,$$

where $t_{i,N} = h_i^{1/(f_i\sqrt{5})}$. Note that when $\mu_i \to 0$, $h_i = e^{-\mu_i E[X]} + o(\mu_i)$ so that $f_i = \mu_i/\mu + o(\mu_i)$. These imply that $t_{i,N} \to e^{-\mu E[X]/\sqrt{5}}$ when $N \to \infty$. Hence, by noting that $\sum_{i=1}^N f_i = 1$, we obtain

$$\limsup_{N\to\infty} C(\gamma(N)) \le c_0\mu - \frac{c_0}{E[X]} \cdot \left\{ 1 - \left(1-\phi^{-1}\right) e^{-\frac{\mu\phi E[X]}{\sqrt{5}}} - \phi^{-1} e^{-\frac{\mu\phi^2 E[X]}{\sqrt{5}}} \right\} .$$

∎

Finally we compare the right-hand side of (18) with (17).

Corollary 5.1 *Assume for all i that $\mu_i \to 0$ as $N \to \infty$ and that $\sum_{i=1}^N \mu_i = \mu > 0$. Then,*

$$\limsup_{N\to\infty} \frac{C(\gamma(N))}{C^*} \le \frac{2\phi^2}{5} = \frac{\sqrt{5}+3}{5} < 1.05 . \tag{19}$$

Proof. Let $q = \mu E[X]$. Define

$$H(q) := \limsup_{N\to\infty} \frac{C(\gamma(N))}{C^*} = \frac{q - 1 + (1-\phi^{-1})\, e^{-\frac{\phi}{\sqrt{5}}q} + \phi^{-1} e^{-\frac{\phi^2}{\sqrt{5}}q}}{q - 1 + e^{-q}} .$$

We will show below that $H(q)$ is decreasing in $q > 0$. Thus, by applying l'Hospital's rule twice, we obtain

$$\sup_{q\ge 0} H(q) \;=\; \lim_{q\to 0} H(q)$$

$$
\begin{aligned}
&= \lim_{q\to 0} \frac{1 - \frac{\phi-1}{\sqrt{5}} e^{-\frac{\phi}{\sqrt{5}}q} - \frac{\phi}{\sqrt{5}} e^{-\frac{\phi^2}{\sqrt{5}}q}}{1-e^{-q}} \\
&= \lim_{q\to 0} \frac{\frac{\phi^2-\phi}{5} e^{-\frac{\phi}{\sqrt{5}}q} + \frac{\phi^3}{5} e^{-\frac{\phi^2}{\sqrt{5}}q}}{e^{-q}} \\
&= \frac{\phi^2-\phi+\phi^3}{5} = \frac{2\phi^2}{5}
\end{aligned}
$$

Hence, (19) will hold if $H'(q) \le 0$ for all $q > 0$.

It is clear that $H'(q) \le 0$ if and only if $R(q) \le 0$, where

$$
R(q) = \left(1 - \frac{\phi-1}{\sqrt{5}} e^{-\frac{\phi}{\sqrt{5}}q} - \frac{\phi}{\sqrt{5}} e^{-\frac{\phi^2}{\sqrt{5}}q}\right)(q-1+e^{-q}) - \left(q - 1 + \left(1-\phi^{-1}\right) e^{-\frac{\phi}{\sqrt{5}}q} + \phi^{-1} e^{-\frac{\phi^2}{\sqrt{5}}q}\right)(1-e^{-q}) \ .
$$

After some simple algebra, we obtain

$$
-\frac{R(q)}{q} + e^{-q} = \left(\left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}-1}{2\sqrt{5}}\right) \cdot e^{-\frac{\phi}{\sqrt{5}}q} + \left(-\left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}+1}{2\sqrt{5}}\right) \cdot e^{-\frac{\phi^2}{\sqrt{5}}q} \ .
$$

Let

$$
\begin{aligned}
\lambda_1 &= \left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}-1}{2\sqrt{5}} \\
\lambda_2 &= -\left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}+1}{2\sqrt{5}}
\end{aligned}
$$

It is easy to see that $\lambda_1 + \lambda_2 = 1$ and that $0 < \lambda_1 \le 1$ (using the inequality $1 - e^{-x} \le x$). Thus, owing to the convexity of the exponential function e^{-qx}, we have

$$
\left(\left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}-1}{2\sqrt{5}}\right) \cdot e^{-\frac{\phi}{\sqrt{5}}q} + \left(-\left(1 - \frac{2}{\sqrt{5}}\right)\frac{1-e^{-q}}{q} + \frac{\sqrt{5}+1}{2\sqrt{5}}\right) \cdot e^{-\frac{\phi^2}{\sqrt{5}}q} \ge e^{-yq} \ ,
$$

where

$$
y = \lambda_1 \cdot \frac{\phi}{\sqrt{5}} + \lambda_2 \cdot \frac{\phi^2}{\sqrt{5}}.
$$

It is simple to check that $y < 1$ so that

$$
-\frac{R(q)}{q} + e^{-q} = \lambda_1 e^{-\frac{\phi}{\sqrt{5}}q} + \lambda_2 e^{-\frac{\phi^2}{\sqrt{5}}q} \ge e^{-yq} > e^{-q} \ .
$$

Therefore, $-R(q)/q > 0$ which implies that $H'(q) < 0$. ∎

6 Concluding Remarks

Many of the numerous questions left unaddressed by our robot-scheduling results lead to interesting avenues for future research. For example, there are the more general questions posed by search engines served by multiple robots, which may be distributed. Also, in a comprehensive data base such as Alta Vista, the number of pages varies with time; the creation of new Web pages continues apace, and many pages will become defunct. One can adapt to this dynamic version by delaying the first access to a newly arriving page until the beginning of the next cycle following its arrival time, at which point a new template is computed. (A page dropped from the data base is deleted from the template as soon as that fact becomes known to the robot; again, a new template must be computed for the next cycle.) Heuristics based on this technique should perform well when the *relative* variation of the data base over a cycle is small, or at least moderately so.

We have already pointed out in Section 3 the open problems connected with sequencing through a set of pages to achieve a given set of access frequencies and approach as closely as possible the lower bound in Theorem 3.3. Progress on this type of template driven scheduling problem has been frustrating in one important respect: Although there are very simple, e.g., greedy type, policies that seem to perform remarkably well, there has been little, if any, success in establishing this fact analytically. In general, one must expect these design and analysis problems to become even more complicated when the avoidance of "rapid-fire" (accessing a Web site too frequently) is taken into account.

Acknowledgment. The authors are pleased to acknowledge the many helpful comments of S. Borst and R. Righter.

References

[1] M. H. Ammar and J. W. Wong, "On the optimality of cyclic transmission in teletext systems," *IEEE Trans. Comm.*, vol. COM-35, no. 1, pp. 68-73, Jan. 1987.

[2] M. Andrews, A. Fernandez, M. Harchol-Balter, T. Leighton, and L. Zhang, "Template Algorithm for One-Hop Packet Routing," Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, 1997 (preprint).

[3] Y. Arian, Y. Levy, "Algorithms for Generalized Round Robin Routing", *Oper. Res. Letters*, Vol. 12, pp. 313-319, 1992.

[4] S. Borst, *Polling Systems*, Ph.D. Thesis, CWI, The Netherlands, 1994.

[5] S. Borst, O. J. Boxma, J. H. A. Harink, G. B. Huitema, "Optimization of Fixed Time Polling Schemes", *Telecommunications Systems*, Vol. 3, pp. 31-59, 1994.

[6] O. J. Boxma, H. Levy, J. A. Weststrate, "Efficient Visit Frequencies for Polling Tables: Minimization of Waiting Cost", *Queueing Systems*, Vol. 9, pp. 133-162, 1991.

[7] O. J. Boxma, H. Levy, J. A. Weststrate, "Efficient Visit Orders for Polling Systems", *Perf. Eval.*, Vol. 18, pp. 103-123, 1993.

[8] F. Douglis, A. Feldmann, B. Krishnamurthy, and J. Mogul, "Rate of change and other metrics: A live study of the World Wide Web," *USENIX*, Dec. 8-11, 1997, Monterey, CA.

[9] H. D. Dykeman, M. H. Ammar, and J. W. Wong, "Scheduling algorithms for videotex systems under broadcast delivery," *Proc. of ICC'86*, pp. 1847-1851, 1986.

[10] B. Hajek, "Extremal Splittings of Point Porcesses", *Math. Oper. Res.*, Vol. 10, pp. 543-556, 1985.

[11] M. Hofri, Z. Rosberg, "Packet Delay under the Golden Ratio Weighted TDM Policy in a Multiple-Acces Channel", *IEEE Trans. on Inform. Theory*, Vol. 33, pp. 341-349, 1987.

[12] A. Itai, Z. Rosberg, "A golden ratio control policy for a multiple-access channel", *IEEE Transactions on Automatic Control* **AC-29**, pp. 712-718, 1984.

[13] J. F. C. Kingman, "The Ergodic Theory of Subadditive Stochastic Processes", *J. Roy. Statist. Soc. Ser. B*, **30**, pp. 499-510, 1968.

[14] Z. Liu, P. Nain, "Optimal Scheduling in Some Multi-Queue Single-Server Systems", *IEEE Transactions on Automatic Control*, Vol. 37, No. 2, pp. 247–252, 1992.

[15] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications* Academic Press, New York, 1979.

[16] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models.* English translation (D.J. Daley editor), J. Wiley & Sons, New York, 1983.

[17] H. Takagi, *Analysis of Polling Systems.* MIT Press, 1986.

[18] H. Takagi, "Queueing analysis of polling models," *Stochastic analysis of computer and communication Systems*, H. Takagi (Ed.), North Holland, 1990.