Query Restart Strategies for Web Agents *

PRASAD CHALASANI
Carnegie Mellon University
Pittsburgh PA 15213
Office phone: (412) 268-3194
chal@cs.cmu.edu

SOMESH JHA
Carnegie Mellon University
sjha@cs.cmu.edu

ONN SHEHORY
Carnegie Mellon University
onn@cs.cmu.edu

KATIA SYCARA
Carnegie Mellon University
katia@cs.cmu.edu

Abstract

With the explosive growth of the internet, autonomous agents will increasingly need strategies for efficiently retrieving information. The time an agent (or server) takes to answer a query issued to it is often a random variable, whose distribution can be estimated by collecting statistics. When an agent A sends a query to another agent B and the query has not completed in some period of time, agent A faces the dilemma of whether to continue waiting or reissue the query (to agent B or to a different one). When some information is available about the probability distribution of the query completion time of the agents, it is possible to devise schemes for agents to reissue queries in such a way that the expected query completion time is reduced significantly. We design such schemes for various models of query-answering agents. Where meaningful, we take into account the cost of sending a query to an agent.

1 Introduction

It is a common observation that when a server, or agent, on the World Wide Web (WWW) is issued a query, the amount of time it takes to answer the query is not fixed. For the same query, the completion time can range from a few seconds to several minutes. This variability is attributable to, for example, network congestion, and the load on the server from concurrently processing several queries. It is therefore reasonable to assume that the completion time of the query is a random variable with a certain distribution. This distribution can be estimated by collecting statistics at various times during the day. Once we know (an approximation of) the completion time distribution, an interesting question arises. After an agent issues a query to another agent, if the query hasn't completed after a while, should it restart the query (on the same agent or a different one), or continue to wait? When is the optimal time to restart the query? Does it help to repeatedly restart the queries? In this paper, under various models of agents processing queries, we consider *restart* strategies with the aim of reducing the expected completion time.

Let us take an example. Suppose we are accessing a particular web site, and we have estimated that the page will arrive in 5 seconds with probability 0.6, and in 20 seconds with probability 0.4. If we do not restart the page request at any time, then the expected time for the page to arrive would be

$$(0.6)5 + (0.4)20 = 3 + 8 = 11 \text{ seconds.}$$

By contrast, consider using the following *restart strategy:*

> Abort and restart the page request if it has not completed in 5 seconds.

We make the reasonable assumption that the restarted request has the same distribution as the first one, and is independent of the first request. Suprisingly, with this strategy, the expected time for the page to arrive is significantly smaller:

$$(0.6)5 + (0.4)11 = 3 + 4.4 = 7.4 \text{ seconds}.$$

Now suppose we abort the *second* request if it hasn't completed in 5 seconds, and reissue the page request. The expected completion time now is

$$(0.6)5 + (0.4)7.4 = 3 + 2.96 = 5.96 \text{ seconds}.$$

which is even smaller than before!

The above type of strategy can be useful when an agent issues a query to another agent or server on the internet. In a typical scenario, an agent that needs some specific task done but is unable to perform the task by itself, will delegate that task to some other agent that has that capability. The requestee will process the task specified in the query and return the

* This research has been supported in part by ONR Grant #N-00014-96-1-1222, by NSF Grant IRI-9612131 and IRI-9712607.

results. There may be several agents with the required capability, each with its own performance characteristics. Now suppose the original agent sends its query to one of these agents, and this agent has not returned the results in, say, 20 seconds. Now should the querying agent continue to wait, or should it abort it and re-send the query, perhaps to a different agent?

There has been some previous work on strategies to query *web sites* (as opposed to web agents). Since some of the same issues arise in the case of agents, we briefly describe this work. For instance, Etzioni *et. al.* [5] consider a model where each web site charges a certain fee to answer a query, and has a deterministic completion time with a certain probability of successfully answering the query. In this model, there is a tradeoff between sequential querying (which is time-consuming but cheap) and concurrent querying (which is fast but expensive). The authors examine some strategies to keep the expected cost plus time small. One weakness of their model is that they assume each site has a fixed completion time. In this paper, we make the more realistic assumption that the completion time has some distribution that can be approximated by collecting statistics. In addition, we consider strategies for aborting a query while it is being handled by an agent, and re-issuing it to the same or another agent.

In [2], Huberman, Lukose and Hogg consider specifically the problem of restarting web-site queries. They briefly mention the possibility of collecting statistics of web-site access times, and mention that it is possible to devise restart strategies to reduce the expectation and variance of the access time. In [4], Lukose and Huberman carry out an analysis of restart strategies under a model very similar to ours. The differences are that (a) we design, using dynamic programming, an optimal restart strategy for arbitrary multi-valued distributions of completion times, (b) for two-valued, and continuous-valued distributions of completion time, we show a succinct characterization of when restarting reduces expected completion time, and (c) for two-valued distributions of completion time, we consider how expected completion time decreases as a function of the number of restarts.

It is common in the multi-agent literature to use *contracts* [7, 3, 6] as a model for agents delegating tasks to other agents. For example if an agent A delegates an information-retrieval task to agent B, the two agents enter into a contract: agent A agrees to pay agent B a certain fee for B's services. Most current work on multi-agent contracts considers contracts as binding. There has been a growing interest in *contingent contracts* [9], which can offer each party several agreed-upon flexibilities. For instance the querying agent may be allowed to abort the query (and perhaps send it to another agent) if the query-answering agent takes more than some agreed upon time to answer a query. Such flexible contracts have been shown to have several benefits [8]. Prior work in this area has not explicitly considered strategies for how best to take advantage of this type of decommitment flexibility. The present paper shows strategies for agents to use this flexibility, and how the results depend on the underlying distribution of query completion times.

The paper is organized as follows. In Section 2, we consider restart strategies on a single agent which has a simple discrete two-valued distribution of completion times. We characterize the conditions under which restarting reduces the expectation and variance of the completion time. Section 3 describes an optimal strategy for restarting queries on agents that have different multi-valued discrete distributions. In Section 4 we give conditions under which restarting helps, for arbitrary distributions. We show that for some familiar continuous distributions of the completion time, restarting can never reduce the expected completion time. We also establish necessary conditions for the existence of a locally optimum restart time. Section 5 concludes with a discussion of future research directions. The proofs of all our results appear in the technical report version of our paper [1].

2 Two-valued discrete distribution

Throughout this paper we will have in mind the following scenario. An agent A needs a certain query answered, and it has one or more specialized agents available to answer it. During its previous interactions with these agents with similar queries, agent A has collected statistics on how long these agents take to complete the queries. It can thus model the random completion time X of each agent by means of a certain distribution. Once this model is constructed, agent A needs to devise a *query restart strategy* that results in a smaller expected completion time (or even minimizes it) than the expected completion time of a single query.

In this section we will suppose that agent A has one or more *similar* agents available to answer the query, and that the completion times X of the agents are independent random variables with the *same* simple two-valued distribution: With probability p, $X = 1$, and with probability $1 - p$, $X = c$ where $c > 1$. Suppose agent A issues a query to one of these agents. If the query has not completed by time 1, should agent A abort the query and resend it (to the same or another agent). Can we reduce the expected completion time by means of an appropriate *restart strategy?* What happens to the variance of the completion time? If the second query also takes longer than time 1, should it again re-start the query? And in general, do multiple restarts help?

To study these questions, we introduce some notation. Let I denote the indicator random variable that equals 1 when $X = 1$, and 0 otherwise. Clearly $\mathbf{E}I = p$, and

$$\begin{aligned} \operatorname{var} I &= \mathbf{E}I^2 - (\mathbf{E}I)^2 \\ &= \mathbf{E}I - (\mathbf{E}I)^2 \\ &= p - p^2. \end{aligned}$$

The completion time X can be written as

$$X = I + (1 - I)c,$$

so the expected completion time for the query is:

$$\mu \stackrel{\triangle}{=} \mathbf{E}X = p + (1 - p)c,$$

and the variance is

$$\begin{aligned} \operatorname{var} X &= \operatorname{var}(1 + I(1 - c)) \\ &= (1 - c)^2 \operatorname{var} I \\ &= (1 - c)^2 p(1 - p). \end{aligned}$$

Now suppose agent A has made the query, and it hasn't completed by time 1. Let us consider the simplest restart strategy: wait for time 1 after the query has been issued, then restart it, and repeat, at most k times. We call this the *k-restart strategy.* We now show:

Theorem 1 *The k-restart strategy has expected completion time*

$$\mathbf{E}Y_k = 1/p + q^{k+1}(c - 1 - 1/p), \quad k \geq 0$$

which decreases with k for all $k \geq 0$ *if and only if $c > 1 + 1/p$. The variance of the completion time with k restarts is, for $k \geq 0$,*

$$\begin{aligned} \operatorname{var} Y_k = &\frac{q}{p^2} + q^{k+1}(c^2 + 1/p - 1 - 2c/p) \\ &+ 2kq^{k+1}(c - 1 - 1/p) - q^{2k+2}(c - 1 - 1/p)^2. \quad (1) \end{aligned}$$

Provided $c > 1 + 1/p$, this is smaller than $\operatorname{var} Y_0$ *for* all sufficiently large k *if and only if $c > 1 + 1/p^{3/2}$.*

How do we make use of this theorem? Clearly the parameters p and c completely characterize the agents that agent A can send the query to. If $c > 1 + 1/p$, then agent A knows that it should abort and restart the query if it hasn't completed by time 1, and the expected completion time will be reduced. However, if there is a cost to issuing queries, then agent A should not restart more often than its "querying" budget allows. The above theorem only says that provided $c > 1 + 1/p$, repeated restarting will reduce the expected completion time.

Examples. Suppose that each agent that is available to agent A has the following characteristics. With probability $p = 0.5$ the agent completes the query in 1 second, while with probability 0.5, the agent completes in c seconds. The two thresholds upon which the behavior of the expected completion time and its variance depend are (we express all times in units of seconds):

$$1 + 1/p = 1 + 1/.5 = 3s; \quad 1 + 1/p^{3/2} = 3.828s.$$

Let us first consider $c = 6$, so we have $c > 1 + 1/p^{3/2}$. The expected completion time for a single query is

$$\mathbf{E}X = p + (1 - p)c = 7/2 = 3.5s,$$

and its standard deviation is

$$\sqrt{\operatorname{var} X} = \sqrt{p(1 - p)(c - 1)^2} = 2.5s.$$

For $k = 1$ restart, the expected completion time is, from the above lemma,

$$\mathbf{E}Y_1 = 1/p + (1-p)^{k+1}(c-1-1/p) = 10 + (30-15)/4 = 2.75s,$$

already better than 3.5s. The standard deviation is 2.487s, which is also better than for $k = 0$. For $k = 5$, $\mathbf{E}Y_k = 2.047s$ and $\sqrt{\operatorname{var} Y_k} = 1.634s$, both significantly smaller than the corresponding quantities for the zero-restart case. It is easy to see that as $k \to \infty$, $\mathbf{E}Y_k \to 1/p = 2s$ and $\sqrt{\operatorname{var} Y_k} \to \frac{\sqrt{q}}{p} = 1.414s$.

In Fig. 1 we show a plot of the expectation $\mathbf{E}Y_k$ and standard deviation $\sqrt{\operatorname{var} Y_k}$ as a function of k. We show the plots for $c = 6s$, $c = 3.5s$ and $c = 2s$. When $c = 3.5s$, $c > 1 + 1/p$ but $c < 1 + 1/p^{3/2}$, so in accordance with Theorem 1, the expectation $\mathbf{E}Y_k$ decreases with k, but the variance $\operatorname{var} Y_k$ is larger than $\operatorname{var} Y_0$ for all $k > 0$. In case $c = 2s$, c is less than $1 + 1/p$ (and hence also less than $1 + 1/p^{3/2}$, so the expectation $\mathbf{E}Y_k$ increases with k.

■

3 Multi-valued discrete distributions

We now examine a more general situation. Suppose the querying agent A has access to r different *independent* agents $A_1, A_2, \ldots, A_r$. Let the completion time of agent A_i be the random variable X_i. By collecting statistics, agent A estimates the distribution of each X_i. In particular, for $i = 1, 2, \ldots, r$, X_i has m_i possible values $c_{i1}, c_{i2}, \ldots, c_{im_i}$, where

$$c_{i1} < c_{i2} < \ldots < c_{im_i},$$

and the probability $\mathbf{P}(X_i = c_{ij})$ is p_{ij}, for $j = 1, 2, \ldots, m_i$. We suppose that agent A is able to re-issue the query up to k times (Perhaps there is a certain fixed cost c dollars to issue a query, and agent A doesn't want to spend more than kc dollars on queries). When the agent reissues a query, we are assuming that the current query is aborted. Also, the query may be reissued to any of the r agents, including those that have been queried before. Our goal is to devise the optimal querying strategy for agent A, i.e., the strategy that minimizes the expected completion time of the query.

Let E^*_ℓ denote the minimum expected remaining time to complete the query, given that there are ℓ restarts remaining. When there is just $\ell = 1$ query remaining, agent A merely

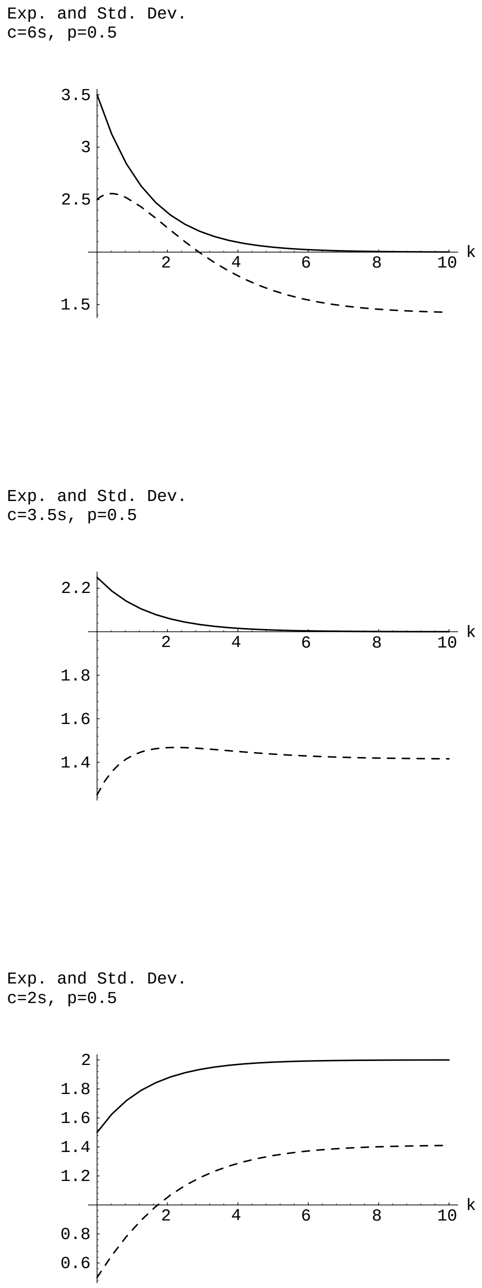

Figure 1: *Expectation and standard deviation of completion time with k restarts in the example, for $c = 6s, 3.5s$ and $2s$. The solid curve shows the expectation $\mathbf{E}Y_k$ and the dashed curve shows the standard deviation $\sqrt{\operatorname{var} Y_k}$.*

sends the query to the agent whose expected completion time $\mathbf{E}X_i$ is smallest, so

$$\mathbf{E}_1^* = \min_i \mathbf{E}X_i, \tag{2}$$

where

$$\mathbf{E}X_i = \sum_{j=1}^{m_i} p_{ij} c_{ij}.$$

Now when there are two query-restarts remaining, A first decides which of the r agents to send the query to, and also decides at which point to re-issue the query to another, *in case* the query hasn't completed by this time. If A decides to first send the query to agent A_i, there are m_i possible restart-points, namely at times $c_{i1}, c_{i2}, \ldots, c_{im_i}$ after issuing the query to agent A_i. Of course, if the "restart-point" c_{im_i} is reached, then A does not re-issue the query since agent A_i will have completed the query by this time. When the decided-upon restart-point is reached, A can re-issue the query according to the strategy that achieves the optimum $\mathbf{E}_1^*$ that A has already computed. For a given restart point c_{is}, the expected time taken by the query up to that point is

$$\sum_{j=1}^{s} c_{ij} p_{ij},$$

and with probability

$$1 - \sum_{j=1}^{s} p_{ij}$$

the query will not have completed by time c_{is}, and since the agents are independent, the expected remaining completion time is then $\mathbf{E}_1^*$. Thus A can compute

$$\mathbf{E}_2^* = \min_{1 \le i \le r} \left[\min_{1 \le s \le m_i} \left\{ \sum_{j=1}^{s} c_{ij} p_{ij} + \left(1 - \sum_{j=1}^{s} p_{ij} \right) \mathbf{E}_1^* \right\} \right] \tag{3}$$

Continuing in this way, A can compute $\mathbf{E}_k^*$

$$\mathbf{E}_k^* = \min_{1 \le i \le r} \left[\min_{1 \le s \le m_i} \left\{ \sum_{j=1}^{s} c_{ij} p_{ij} + \left(1 - \sum_{j=1}^{s} p_{ij} \right) \mathbf{E}_{k-1}^* \right\} \right] \tag{4}$$

Clearly, the computation of each $\mathbf{E}_i^*$ takes time proportional to $\sum_{i=1}^{r} m_i$, (the total number of possible completion times), so that A can compute the minimum expected completion time with up to k queries, $\mathbf{E}_k^*$, in time proportional to $k \sum_{i=1}^{r} m_i$.

We have shown how to compute the $\mathbf{E}_i^*$ values. What is the strategy for agent A that achieves the optimum $\mathbf{E}_k^*$? To extract the optimal strategy, A starts with the computation of $\mathbf{E}_k^*$. It sends the very first query to the agent A_i that achieves the minimum in expression (4) (let us call it i^*), and we pick the restart-point as the index $s = s^*$ that achieves the inner minimum in that expression for $i = i^*$. If the query hasn't completed by time $c_{i^*s^*}$, A re-issues the query. For re-issuing the query, it proceeds in a similar manner, i.e., pick the agent A_{i^*} and restart point $c_{i^*s^*}$ by examining the computation of $\mathbf{E}^*_{k-1}$. Similarly it determines the third agent and restart point, and so on until it reaches the k'th agent and restart point. Note that if at some stage the restart point s^* is m_{i^*} for a given agent A_{i^*}, then agent A does not restart the query, and lets the current query proceed to completion.

4 General Distributions

We now derive some general results that hold when the completion time has an arbitrary distribution, either discrete-valued or continuous-valued. We will use the notation developed here in the following sections. We first consider the case of a single restart, where agent A sends a query to agent A_1, and if the query has not completed by some time $b > 0$, it *aborts and resends* the query to agent A_2. We let X be the random completion time of agent A_1, and X' be the random completion time of agent A_2. We assume that X, X' are independent random variables. We let Y denote the (random) completion time under this strategy. Let I_b be the indicator random variable for the set $\{X \leq b\}$, i.e., $I_b = 1$ if agent A_1 completes the query by time b, and $I_b = 0$ otherwise. We denote the cumulative distribution function of X by $F(x)$, i.e., $F(x) = \mathbf{P}(X \leq x)$. Thus $\mathbf{E}I_b = \mathbf{P}(X \leq b) = F(b)$. Clearly Y is given by

$$Y = I_b X + (1 - I_b)(b + X').$$

We want to now compute the expectation of Y, which we denote by $g(b)$, to emphasize that this is the expected completion time when restarting at time b. Taking expectations in the above equation, and using the independence of I_b and X',

$$g(b) \equiv \mathbf{E}Y = \mathbf{E}XI_b + (1 - F(b))(b + \mathbf{E}X'). \qquad (5)$$

If X has a continuous-valued distribution with density $f(x)$, $\mathbf{E}XI_b$ is given by

$$\mathbf{E}XI_b = \int_0^b x f(x)\, dx,$$

and if X can take only discrete values $c_1 < c_2 < \ldots < c_m$, where $\mathbf{P}(X = c_i) = f(c_i)$, then $\mathbf{E}XI_b$ is given by

$$\mathbf{E}XI_b = \sum_{i: c_i \leq b} c_i f(c_i).$$

In the case of a discrete-valued distribution it is clear that A only needs to consider restart times b that coincide with the possible values of X. Note that as $b \to \infty$ we have $g(b) \to \mathbf{E}X$.

The question we would like to answer is: is there some $b > 0$ for which $g(b) < \mathbf{E}X$? To gain insight into this question we write the condition $g(b) < \mathbf{E}X$ as follows:

$$\mathbf{E}XI_b + (1 - F(b))(b + \mathbf{E}X') < \mathbf{E}X,$$

which is equivalent to

$$\frac{\mathbf{E}X(1 - I_b)}{(1 - F(b))} - b > \mathbf{E}X'$$

Notice that the first ratio term above is exactly the conditional expectation of X, *given* $X > b$, or $\mathbf{E}[X|X > b]$, so we can write the condition $g(b) < \mathbf{E}X$ as

$$\mathbf{E}[X - b|X > b] > \mathbf{E}X'. \qquad (6)$$

This condition is easy to understand intuitively. It says:

> The expected *remaining* time agent A_1 would take if it were allowed to continue, *given* that it has taken longer than b, is greater than the expected time agent A_2 would take.

If this condition is true for some b then aborting and resending the query to agent A_2 at time b will reduce the expected completion time. Of course this treatment also applies to aborting and resending the query to the *same* agent A_1, in which case we have $\mathbf{E}X = \mathbf{E}X'$.

In the following two sections, we apply the expressions derived above to some standard distributions.

4.1 Continuous distributions

We now consider some common continuous-valued distributions for the completion time X. Let $f(x)$ be the density of the distribution, and let $F(x) = \mathbf{P}(X \leq x)$ be the cumulative distribution function. We write $\mu = \mathbf{E}X$. Even for a single restart, it is not clear under what conditions a restart time exists that yields an expected completion time smaller than μ.

As we will see in the examples below, restarting does not reduce the expected completion time for some simple distributions: uniform over $[0, c]$ for some $c > 0$; the exponential distribution with density $f(x) = e^{-x}, x \geq 0$; and the one-sided normal distribution with density $f(x) = \sqrt{2/\pi}\exp(-x^2/2)$, $x \geq 0$. We first establish necessary conditions for the existence of a b such that $g(b) < g(0)$ and characterize a local minimum of the function $g(b)$.

Theorem 2 *If b is a restart time that results in a smaller expected completion time than the expected completion time μ of a single query, then b must satisfy*

$$\mu F(b) > (1 - F(b))b.$$

If b is a (local) minimum for the expected completion time $\mathbf{E}Y$, then b must satisfy

$$F(b) + \mu f(b) = 1$$

and

$$f'(b) < -(1 - F(b))/\mu^2.$$

Examples. We consider some simple distributions of X for which there does not exist a restart time $b > 0$ that reduces the expected completion time. First consider a **uniform distribution:** let X be uniformly distributed over the interval $[0, c]$ for some $c > 0$. Thus the density is $f(x) = 1/c$, $x \in [0, c]$, and the cumulative distribution function is

$$F(x) = \begin{cases} x/c, & x \in [0, c], \\ 1. & x > c \end{cases}$$

Then $\mu = \mathbf{E}X = c/2$, and

$$\mathbf{E}XI_b = \int_0^b xf(x)\, dx = (1/c)\int_0^b x\, dx = \frac{b^2}{2c}.$$

Clearly, the restart time b must lie in $[0, c]$. For such b, the expected completion time $g(b) = \mathbf{E}Y$ with restart at b is given by (5):

$$\begin{aligned} g(b) &= \frac{b^2}{2c} + (1 - b/c)(b + \mu) \\ &= \frac{b^2}{2c} + (1 - b/c)(b + c/2) \\ &= \frac{b + c}{2} - \frac{b^2}{2c}, \end{aligned}$$

and this is less than $\mu = c/2$ if and only if

$$\frac{c}{2} > \frac{b + c}{2} - \frac{b^2}{2c},$$

or equivalently,

$$b > c$$

which contradicts our assumption that $b \in [0, c]$. Thus there is *no* restart time b that reduces the expected completion time.

Consider now the **exponential distribution** with parameter $\lambda > 0$, whose density is $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, and whose cumulative distribution function is $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$. For this distribution, $\mu = \mathbf{E}X = 1/\lambda$,

$$\mathbf{E}XI_b = \int_0^b xf(x)\, dx = -be^{-\lambda b} - \frac{1}{\lambda}e^{-\lambda b} + \frac{1}{\lambda}.$$

The expected completion time with restart at any $b > 0$ is given by (5):

$$\begin{aligned} g(b) &= 1/\lambda - e^{-\lambda b}/\lambda - be^{-\lambda b} + e^{-\lambda b}(b + 1/\lambda) \\ &= 1/\lambda = \mu, \end{aligned}$$

so the expected completion time with restart at b is the same as the single-query expected completion time, regardless of the choice of b.

Consider next the **one-sided normal distribution** whose density is given by

$$f(x) = \sqrt{\frac{2}{\pi}}\exp(-x^2/2), \quad x \geq 0,$$

and whose cumulative distribution function is

$$F(x) = 2N(x) - 1, \quad x \geq 0,$$

where $N(x)$ is the familiar cumulative distribution function of the normal distribution:

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-u^2/2)\, du.$$

We then have

$$\begin{aligned} \mu = \mathbf{E}X &= \sqrt{\frac{2}{\pi}}\int_0^\infty ue^{-u^2/2}\, du \\ &= \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-u^2/2}\, du^2 \\ &= \sqrt{2/\pi}, \end{aligned}$$

and similarly

$$\mathbf{E}XI_b = \sqrt{2/\pi}\left(1 - e^{-b^2/2}\right).$$

So from (5), the expected completion time $\mathbf{E}Y$ with restart at b is

$$g(b) = \left(1 - e^{-b^2/2}\right)\sqrt{2/\pi} + 2(1 - N(b))(b + \sqrt{2/\pi}).$$

It is easily verified that $g(0)$ is indeed equal to $\mu = \sqrt{2/\pi}$. In Fig. 2 we show a Mathematica plot of $g(b) - \sqrt{2/\pi}$, and we see that this is never negative. Thus again there is no restart time b that reduces the expected completion time.

Despite what we have seen in the above examples, it should be kept in mind that there are continuous-valued distributions for which restarting reduces the expected completion time: simply take a 2-valued distribution for which restarting reduces the expected completion time, and "smooth" it to obtain a continuous-valued, bimodal distribution. Then restarting after the first "hump" in the distribution would result in an expected completion time smaller than the original expected completion time. This might lead us to conjecture that whenever the completion time X has a *unimodal* continuous-valued distribution, restarting cannot reduce the expected completion time. This, however, is not the case. For instance if X has the density (shown in Fig. 3)

$$f(x) = \begin{cases} 8/12, & 0 \leq x \leq 1, \\ 1/12, & 1 < x \leq 5, \end{cases}$$

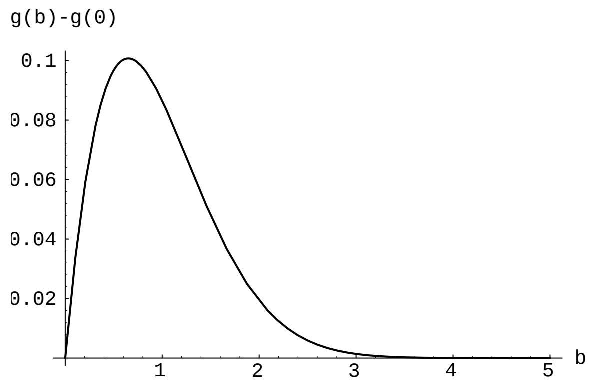

Figure 2: *Plot of* $g(b) - \sqrt{2/\pi}$ *vs* b.

then restarting at time 1 would reduce the expected completion time. Indeed, first we calculate $\mathbf{E}X$:

$$
\begin{aligned}
\mu = \mathbf{E}X &= \int_0^5 x f(x)\, dx \\
&= (8/12) \int_0^1 x\, dx + (1/12) \int_1^5 x\, dx \\
&= (8/12).(1/2) + (1/12)(24/2) = 4/3 = 1.333.
\end{aligned}
$$

Now if we restart the query if it hasn't completed by time 1, then from (5), the expected completion time now is

$$
\begin{aligned}
g(b) &= \mathbf{E}XI_1 + (1 - F(1))(1 + \mu) \\
&= (8/12) \int_0^1 x\, dx + \left(1 - (8/12) \int_0^1 dx\right)(1 + 4/3) \\
&= (8/12)(1/2) + (1 - 8/12)(1 + 4/3) \\
&= 10/9 = 1.111,
\end{aligned}
$$

which is smaller than $\mathbf{E}X$.

■

5 Conclusion

As agents proliferate on the internet, they will increasingly be able to choose among several agents to accomplish specific tasks such as information queries. Also, as network congestion increases, so will the variability of query completion times. Restart strategies attempt to reduce the variance of the access time, as well as the expectation. We presented several restart strategies and the conditions under which they reduce expected query completion time. The

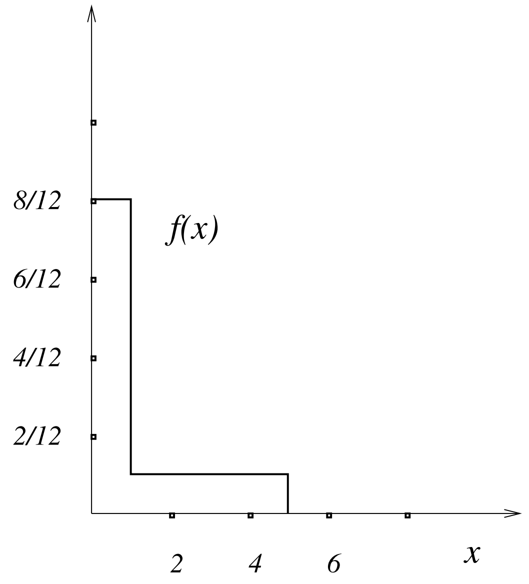

Figure 3: *A unimodal density* $f(x)$ *for the completion time* X*. When restarted at time* 1*, the resulting expected completion time is smaller than* $\mathbf{E}X$.

theoretical completion time distributions considered in this paper can be used as approximations to the actual distributions. Then we can apply the restart strategies described here to reduce the expected completion times. We are currently collecting statistics of access times in order to study restart strategies experimentally.

The mathematical model for restart strategies that we have presented can be used to make decommitment decisions in delegating tasks to agents through non-binding contracts. Currently, we are studying the effects of allowing agents to abort and restart queries on the overall multi-agent system performance. In particular we are studying under what conditions these restart strategies lead to increased multi-agent system efficiency, and whether these lead to thrashing and instabilities.

References

[1] P. Chalasani, S. Jha, O. Shehory, and K. Sycara. Restart strategies for web agents. Technical Report CMU-CS-97-180, Carnegie Mellon University, 1997.

[2] B. Huberman, R. Lukose, and T. Hogg. An economics approach to hard computational problems. *Science*, 275:51–54, 1997.

[3] S. Kraus. Agents contracting tasks in non-collaborative environments. In *Proc. 11th Nat. Conf. on Artificial Intelligence (AAAI)*, 1993.

[4] R. Lukose and B. Huberman. A methodology for managing risk in electronic transactions over the internet. In *3rd Int. conf. computational economics*, 1997.

[5] O.Etzioni, S. Hanks, T. Jiang, R. Kark, O. Madani, and O. Waarts. Efficient information gathering on the internet. In *Proc. Foundations of Comp. Sc.*, 1996.

[6] S. Park, E. H. Durfee, and W. P. Birmingham. Advantages of strategic thinking in multiagent contracts. In *Proceedings of the Second International Conference on Multiagent Systems (ICMAS-96)*, Kyoto, Japan., 1996.

[7] T. Sandholm and V. Lesser. Issues in automated negotation and electronic commerce: extending the contract net protocol. In *Proc. First Int. Conf. on Multiagent Systems (ICMAS-95)*, June 1995.

[8] T. Sandholm and V. Lesser. Advantages of a leveled commitment contracting protocol. In *Proc. AAAI*, pages 126–133, 1996.

[9] K. Sycara. Using option pricing to value commitment flexibility in multi-agent systems. Technical Report CMU-CS-97-169, Carnegie Mellon University, 1997.