Optimal Speedup of Las Vegas Algorithms

Michael Luby[†] Alistair Sinclair[‡]

David Zuckerman[§]

TR-93-010

March 1993

Abstract

Let A be a Las Vegas algorithm, i.e., A is a randomized algorithm that always produces the correct answer when it stops but whose running time is a random variable. We consider the problem of minimizing the expected time required to obtain an answer from A using strategies which simulate A as follows: run A for a fixed amount of time t_1, then run A independently for a fixed amount of time t_2, etc. The simulation stops if A completes its execution during any of the runs. Let $\mathcal{S} = (t_1, t_2, \ldots)$ be a strategy, and let $\ell_A = \inf_{\mathcal{S}} T(A, \mathcal{S})$, where $T(A, \mathcal{S})$ is the expected value of the running time of the simulation of A under strategy $\mathcal{S}$.

We describe a simple universal strategy $\mathcal{S}^{\mathrm{univ}}$, with the property that, for *any* algorithm A, $T(A, \mathcal{S}^{\mathrm{univ}}) = \mathrm{O}(\ell_A \log(\ell_A))$. Furthermore, we show that this is the best performance that can be achieved, up to a constant factor, by any universal strategy.

[†] Current address: International Computer Science Institute, 1947 Center Street, Berkeley CA 94704, luby@icsi.berkeley.edu. Research supported in part by NSF Grant CCR-9016468 and grant No. 89-00312 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. ICSI and UC Berkeley

[‡] Current address: International Computer Science Institute, 1947 Center Street, Berkeley CA 94704, sinclair@icsi.berkeley.edu. Research supported in part by grant GR/F 90363 of the UK Science and Engineering Research Council. University of Edinburgh and ICSI, Berkeley

[§] Current address: MIT Laboratory for Computer Science, 545 Technology Square, Cambridge MA 02139, diz@theory.lcs.mit.edu. Research supported by an NSF Postdoctoral Fellowship. Laboratory for Computer Science, MIT

1 Introduction

Let $A(x)$ be a randomized algorithm of the *Las Vegas* type, by which we mean that, on any input x, the output of A is always correct but its running time, $T_A(x)$, is a random variable. Our task is to minimize, for each input x, the expected time required to get an answer from A. In doing so we are viewing $A(x)$ as a black box, so the only experiments we are allowed to perform are of the following kind: run $A(x)$ for some number t_1 of steps; if $A(x)$ halts during this time then we are done, otherwise restart $A(x)$ from the beginning (using an independent sequence of random bits) and run it for t_2 steps, and so on. Any such experiment can be described by a *strategy* $\mathcal{S} = (t_1, t_2, t_3, \ldots)$, which is just an infinite sequence of values from the set $\mathbb{Z}^+ \cup \{\infty\}$. We will also consider rather more general strategies, in which the running times t_i are themselves random variables and runs may be suspended and then restarted at a later time.

If full knowledge is available about the distribution of $T_A(x)$, then it is possible to design a strategy that is *optimal*, in the sense that it achieves the minimum expected running time amongst all strategies for $A(x)$. We describe such a strategy in Section 2, and prove its optimality; it has the form $(t^*, t^*, t^*, \ldots)$ for a carefully chosen value t^* that depends on the entire distribution of $T_A(x)$. The expected running time of this optimal strategy, which we denote $\ell_A(x)$, is a natural and easily characterized quantity associated with the distribution of $T_A(x)$.

While the existence of an optimal strategy is an interesting theoretical observation, it is of little value in practice because it requires for its implementation detailed information about the distribution of $T_A(x)$. In practical applications, very little, if any, a priori information is available about this distribution, and its shape may vary wildly with x. Furthermore, since we only want the answer once for any x, there is no point in running experiments to gather information about the distribution: the only information that could be gathered from such a run is that $A(x)$ stops, in which case we also obtain the answer. Thus, the problem we address is that of designing an efficient *universal* strategy, i.e., one that is to be used for all distributions on running times.

In Section 3 we describe a simple universal strategy. As is natural, we measure the performance of this strategy in terms of $\ell_A(x)$, the expected running time of the optimal strategy described above. We show that the expected running time of the universal strategy is $O(\ell_A(x) \log(\ell_A(x)))$, which is only a logarithmic factor slower than the optimal strategy that assumes full information about the distribution. For a wide variety of distributions, this represents a dramatic speedup over the naïve strategy of running the algorithm till termination. We go on to show that this bound is optimal, i.e., for any universal strategy there is a distribution for which the expected running time of the strategy is slower by a logarithmic factor than that of the optimal strategy.

Our interest in this problem was sparked by work presented by Wolfgang Ertel at ICSI on the speedup of Las Vegas-type theorem proving algorithms using parallel processors [2]. Ertel considers strategies of the form: run $A(x)$ in parallel using k processors and stop when any one of the k runs terminates. In his case, the algorithm $A(x)$ consists of random search of a (highly unbalanced) tree. He demonstrates that, in many real-life examples of this kind, the distribution on the running time of $A(x)$ is wildly erratic and that the parallel approach can result in substantially faster theorem provers. By using the strategy we develop in this paper for sequential computation, close to optimal speedup can be achieved in the sequential setting without any a priori knowledge of the shape of the tree. Of course, our results apply equally to *any* Las Vegas algorithm.

This paper is similar in spirit to recent work of Alt *et al* [1], who consider simulation strategies for Las Vegas algorithms with the goal of minimizing the *tail probability* of the simulation, i.e., the probability that the simulation runs for more than t steps. Although tail probabilities are not the main focus of this paper, it turns out that our universal strategy achieves tail probabilities that are in many cases stronger than those obtained in [1], and in no case qualitatively weaker. In both papers, the tail probability is essentially of the form $\exp(-t/\alpha)$, but in [1] α is given in terms of the expectation $\mathrm{E}[T_A(x)]$, whereas in our paper it is given in terms of the quantity $\ell_A(x)$, which is never larger than $\mathrm{E}[T_A(x)]$. (For the precise bounds, see Theorem 6 in Section 3.) In many cases our bound is considerably sharper, since for many distributions $\ell_A(x) \ll \mathrm{E}[T_A(x)]$, and it is even possible for $\mathrm{E}[T_A(x)]$ to be infinite while $\ell_A(x)$ is very small. The worst situation from our point of view is when $\ell_A(x) \approx \mathrm{E}[T_A(x)]$, in which case the bounds in the two papers are similar. Thus, our universal strategy achieves both optimal expected running time *and* small tail probabilities.

2 An optimal strategy when the distribution is known

In the remainder of this paper, we identify a Las Vegas algorithm A, together with an input x, with the probability distribution p on its running time $T_A(x)$. Thus p is a probability distribution over $\mathbb{Z}^+ \cup \{\infty\}$, and $p(t)$ denotes the probability that $A(x)$ stops after exactly t steps. We will always assume that p is non-trivial in the sense that $p(\infty) < 1$, so that there exists a finite earliest time, $t = t_{\min}$ say, for which $p(t) > 0$. Our main focus of attention is the expected running time of $\mathcal{S}$ when applied to an algorithm $A(x)$ described by distribution p, which we denote $T(\mathcal{S}, p)$. In this section we will be considering a fixed distribution p, so we abbreviate $T(\mathcal{S}, p)$ to $T(\mathcal{S})$.

The first question we ask is the following. Suppose that we have full knowledge of the distribution p; is there some strategy $\mathcal{S}$ that is *optimal* for p, in the sense that $T(\mathcal{S}) = \inf_{\mathcal{S}} T(\mathcal{S})$? The answer turns out to be "yes," and moreover the optimal strategy is rather easy to describe: it is the repeating sequence $(t^*, t^*, t^*, \ldots)$ for a carefully chosen value t^*.

In fact, strategies of the special form $(t, t, t, \ldots)$ will play a central role in our analysis, so we begin by analyzing them. Our first observation is that the expected running time of the strategy $\mathcal{S}_t = (t, t, t, \ldots)$ is given by the quantity

$$\ell(t) = \frac{1}{q(t)}\Bigl(t - \sum_{t'<t} q(t')\Bigr), \tag{1}$$

where $q(t) = \sum_{t' \le t} p(t')$ is the cumulative distribution function of p.

Lemma 1 *For all finite $t \geq t_{\min}$, $T(\mathcal{S}_t) = \ell(t)$.*

Proof: Let $\mathcal{S} = (t_1, t_2, t_3, \ldots)$ be any strategy, and write $\mathcal{S}'$ for the same strategy with t_1 omitted, i.e., $\mathcal{S}' = (t_2, t_3, \ldots)$. If t_1 is finite, we may write $T(\mathcal{S})$ as follows:

$$\begin{aligned} T(\mathcal{S}) &= \sum_{t \le t_1} tp(t) + (1 - q(t_1))(t_1 + T(\mathcal{S}')) \\ &= q(t_1)\ell(t_1) + (1 - q(t_1))T(\mathcal{S}')\,, \end{aligned} \tag{2}$$

where we have used the relationship

$$\sum_{t' \leq t} t' p(t') = t q(t) - \sum_{t' < t} q(t') = q(t)\ell(t) - t(1 - q(t)). \tag{3}$$

Now in the case $\mathcal{S} = \mathcal{S}_t$, we have $t_1 = t$ and $\mathcal{S}' = \mathcal{S}$, so we can solve (2) for $T(\mathcal{S})$ to give $T(\mathcal{S}_t) = \ell(t)$, as claimed. □

In order to specify our optimal strategy, we define

$$\ell \equiv \ell_p = \inf_{t < \infty} \ell(t). \tag{4}$$

It is easy to see that ℓ is finite for any non-trivial distribution p. Let t^* be any finite value of t for which $\ell(t) = \ell$, if such a value exists, and $t^* = \infty$ otherwise. We shall see shortly that the strategy $\mathcal{S}_{t^*} = (t^*, t^*, t^*, \ldots)$ is optimal for p. First, we analyze its expected running time.

Lemma 2 *The strategy $\mathcal{S}_{t^*} = (t^*, t^*, t^*, \ldots)$ satisfies $T(\mathcal{S}_{t^*}) = \ell$.*

Proof: If t^* is finite then the statement is immediate by the definition of t^*, ℓ and Lemma 1. We turn now to the case where $t^* = \infty$. Here the strategy $\mathcal{S}_{t^*} = \mathcal{S}_\infty$ is simply the naïve one of running $A(x)$ until it halts, so $T(\mathcal{S}_{t^*}) = \mathrm{E}[p]$, the expectation of p. Using (3), we may write $\mathrm{E}[p]$ as

$$\mathrm{E}[p] = q(t)\ell(t) + \epsilon(t) \tag{5}$$

for any $t \geq t_{\min}$, where $\epsilon(t) = \sum_{t > t'}(t' - t)p(t')$. Note first that in this case $\mathrm{E}[p]$ must be finite. For if not, take t_0 sufficiently large that $q(t_0)\ell(t_0) \geq \ell(t_{\min})$. Then we have $\ell(t) \geq q(t)\ell(t) \geq q(t_0)\ell(t_0) \geq \ell(t_{\min})$ for all $t \geq t_0$, so $\ell(t)$ must attain its minimum value for some $t \leq t_0$, contradicting the fact that $t^* = \infty$. Now, since $\mathrm{E}[p]$ is finite, we have $\lim_{t\to\infty} q(t) = 1$ and $\lim_{t\to\infty} \epsilon(t) = 0$, and thus from (5) it follows that $\mathrm{E}[p] = \lim_{t\to\infty} \ell(t)$. But since $t^* = \infty$, we must have $\mathrm{E}[p] = \lim_{t\to\infty} \ell(t) = \inf_{t<\infty} \ell(t) = \ell$. □

We are now in a position to prove that the strategy $\mathcal{S}_{t^*}$ is optimal.

Theorem 3 *For any distribution p, the strategy $\mathcal{S}_{t^*} = (t^*, t^*, t^* \ldots)$ is an optimal strategy for p, and $T(\mathcal{S}_{t^*}) = \ell_p$.*

Proof: We have already seen in Lemma 2 that $T(\mathcal{S}_{t^*}) = \ell_p$. It suffices to show that no other strategy can beat this bound.

Let $\mathcal{S} = (t_1, t_2, t_3, \ldots)$ be any strategy. We may assume without loss of generality that $t_i \geq t_{\min}$ for all i, since smaller values are plainly redundant and can be removed from $\mathcal{S}$ without increasing $T(\mathcal{S})$. We shall show that $T(\mathcal{S}) \geq \ell = \ell_p$ by expressing $T(\mathcal{S})$ as a convex combination of the values $\ell(t_i)$. In doing so, it will be convenient to extend the definition (1) of $\ell(t)$ by setting $\ell(\infty) = \mathrm{E}[p]$, the expectation of p. Note that $\mathrm{E}[p]$ is precisely the expected running time of the strategy $\mathcal{S}_\infty = (\infty)$, so we have effectively extended Lemma 1 to all values of t. Moreover, (3) also holds for $t = \infty$ provided we interpret $0 \times \infty$ as 0.

Now let X be the random variable whose value is the running time of $\mathcal{S}$ when applied to p, and for each i write $T_i = \sum_{j=1}^{i} t_j$. Thus $T(\mathcal{S})$ is just the expectation $\mathrm{E}[X]$, which we may write in the following way:

$$T(\mathcal{S}) = \mathrm{E}[X] = \sum_{i=1}^{\infty} \Big\{ t_i \Pr[X > T_i] + \sum\nolimits_{t=1}^{t_i} t \Pr[X = T_{i-1} + t] \Big\}. \tag{6}$$

Now using (3) the inner summation may be expanded as follows:

$$\begin{aligned} \sum\nolimits_{t=1}^{t_i} t \Pr[X = T_{i-1} + t] &= \Pr[X > T_{i-1}] \sum\nolimits_{t=1}^{t_i} t p(t) \\ &= \Pr[X > T_{i-1}] \Big\{ q(t_i)\ell(t_i) - t_i(1 - q(t_i)) \Big\} \\ &= \Pr[X > T_{i-1}]\, q(t_i)\ell(t_i) \;-\; \Pr[X > T_i]\, t_i. \end{aligned}$$

(Note that this holds even if $t_i = \infty$ provided we interpret $0 \times \infty$ as 0.) Thus (6) simplifies to

$$T(\mathcal{S}) = \sum_i \Pr[X > T_{i-1}]\, q(t_i)\ell(t_i) = \sum_i g_i \ell(t_i), \tag{7}$$

where $g_i = \Pr[T_{i-1} < X \le T_i]$. Since the g_i form a probability distribution, (7) is a convex combination of the $\ell(t_i)$. Hence we must have $T(\mathcal{S}) \ge \inf_i \ell(t_i) \ge \ell$ and the proof is complete. □

Remarks: (a) Note from (5) that $\ell_p \le \mathrm{E}[p]$ for all distributions p.

(b) As an example where t^* is infinite, consider the distribution p with $p(1) = 0$ and $p(t) = 2^{1-t}$ for $t > 1$. Routine calculations show that $\ell(1) = \infty$ and $\ell(t) = \frac{3 \cdot 2^{t-1} - 2}{2^{t-1} - 1}$ for $t > 1$, which is a strictly decreasing sequence with limit 3. Theorem 3 confirms the intuition that, because of the overhead involved in the first step, it is always better to continue the first run of $A(x)$ than to stop and return to the beginning.

(c) Theorem 3 indicates that, if we know the value t^* for a given distribution p, we can design an optimal strategy for p. It is perhaps worth noting that it is also enough to know the value $q(t^*)$ of the cumulative distribution at t^* in order to design a strategy whose expected running time is within a constant factor of the optimal value ℓ. The strategy is simply the following: for each $i \ge 1$ in sequence, execute $\lfloor 1/q(t^*) \rfloor$ runs of length 2^i. The justification is straightforward and left to the reader. □

It might be argued that the class of strategies we have considered is somewhat restricted, and that substantially better performance could perhaps be obtained if we widened the class. Two natural extensions are *probabilistic* strategies, in which the length t_i of each run is itself a random variable, and *mixed* strategies, in which runs may be suspended and then restarted at a later time. In this latter case, a strategy is an infinite sequence of pairs of the form $\{m_i, t_i\}$, where m_i is the label of a run and t_i a time as before; such a pair indicates that, at stage i, the run m_i is to be continued for t_i steps starting from the configuration in which it was previously suspended (if any), and then suspended again. To avoid confusion, we shall use the term *pure* strategy to refer to a strategy of the simple kind considered up to now, i.e., deterministic and with no restarts. Is the pure strategy of Theorem 3 still optimal in this more general framework? It turns out that it is, as we now prove.

Theorem 4 *The strategy $\mathcal{S}_{t^*}$ of Theorem 3 remains optimal even when probabilistic mixed strategies are permitted.*

Proof: The reader should have little trouble convincing herself that the proof of Theorem 3 goes through essentially unchanged for probabilistic strategies, because any probabilistic strategy can be viewed as a convex combination of deterministic strategies. The same applies to the more general proof we give below for mixed strategies.

Let $\mathcal{S} = (\{m_i, t_i\})$ be a mixed strategy. We start from expression (6) for the expectation $T(\mathcal{S}) = \mathrm{E}[X]$ in the proof of Theorem 3. The expansion of the inner summation is now a little more complicated because of the possibility that $\{m_i, t_i\}$ is a continuation of a previously suspended run. We introduce some more notation to handle this. For each i, let $\bar{t}_i$ be the total time allocated to run m_i up to and including the current stage $\{m_i, t_i\}$, and let Q_i be the probability that no *other* run (distinct from m_i) has stopped *before* the current stage. As in the proof of Theorem 3, we may assume that $\bar{t}_i \geq t_i \geq t_{\min}$ for all i. Suppose first that $\{m_i, t_i\}$ is the continuation of a run whose previous stage was $\{m_{i'}, t_{i'}\}$. Then the inner summation in (6) becomes

$$\begin{aligned}\sum\nolimits_{t=1}^{t_i} t \Pr[X = T_{i-1} + t] &= Q_i \sum\nolimits_{t=1}^{t_i} p(\bar{t}_{i'} + t)t \\ &= Q_i \left\{ t_i q(\bar{t}_i) - \sum\nolimits_{t=\bar{t}_{i'}}^{\bar{t}_i - 1} q(t) \right\} \\ &= Q_i \left\{ q(\bar{t}_i)\ell(\bar{t}_i) - (1 - q(\bar{t}_i))t_i - q(\bar{t}_{i'})\ell(\bar{t}_{i'}) \right\},\end{aligned}$$

where we have used (3) freely. Since $\Pr[X > T_i] = Q_i(1 - q(\bar{t}_i))$, the overall contribution of the ith summand to (6) is

$$Q_i q(\bar{t}_i)\ell(\bar{t}_i) - Q_i q(\bar{t}_{i'})\ell(\bar{t}_{i'}). \qquad (8)$$

Suppose on the other hand that $\{m_i, t_i\}$ is the first stage of run m_i. Then the contribution is exactly as in the proof of Theorem 3, namely (from (7))

$$\Pr[X > T_{i-1}]\, q(t_i)\ell(t_i) = Q_i q(\bar{t}_i)\ell(\bar{t}_i). \qquad (9)$$

Putting all this together, we may write (6) as

$$T(\mathcal{S}) = \sum_i g_i \ell(\bar{t}_i), \qquad (10)$$

where

$$g_i = \begin{cases} (Q_i - Q_{i''})q(\bar{t}_i) & \text{if } \{m_i, t_i\} \text{ is continued by } \{m_{i''}, t_{i''}\}; \\ Q_i q(\bar{t}_i) & \text{if } \{m_i, t_i\} \text{ is not continued.} \end{cases}$$

From this it is clear that all coefficients g_i in the sum (10) are positive. Moreover, for each i the sum of the two coefficients in (8), or the single coefficient in (9), is easily seen to be $\sum_i \Pr[T_{i-1} < X \leq T_i]$, from which it is clear that $\sum_i g_i = 1$. Hence we have a convex combination of values $\ell(\bar{t}_i)$, so just as in the proof of Theorem 3 we conclude that $T(\mathcal{S}) \geq \ell$. Thus the mixed strategy cannot beat our previous bound. □

3 Unknown distributions

The optimal strategy described in the previous section clearly requires detailed knowledge of the distribution p for its implementation. As we have already explained, however, in the applications we have in mind there will be no information available about p. We are therefore led to ask what is the best performance we can achieve in the absence of any a priori knowledge of p.

It will be helpful to introduce one further function associated with a distribution p. For finite values of $t \geq t_{\min}$, define

$$L(t) = \frac{t}{q(t)},$$

where q is the cummulative distribution function of p as before, and by analogy with (4) define

$$L_p = \inf_{t<\infty} L(t). \tag{11}$$

Note that $\ell_p \leq L_p \leq 4\ell_p$; the first inequality is obvious, and the second may readily be checked. Furthermore, as can easily be verified, there is always some finite value $t = t_0$ such that $L(t_0) = L_p$. (Recall that this is not the case for ℓ_p.)

Our next theorem says that, with no knowledge whatsoever about the distribution p, we can always come surprisingly close to the optimum value for the case of full knowledge given in Theorem 3. Moreover, this performance is achieved by a pure strategy of a very simple form that is easy to implement in practice.

Consider the strategy $\mathcal{S}^{\text{univ}}$ indicated by

$$\mathcal{S}^{\text{univ}} = (1,1,2,1,1,2,4,1,1,2,1,1,2,4,8,1,\ldots).$$

One way to describe this strategy is to say that all run lengths are powers of two, and that each time a pair of runs of a given length has been completed, a run of twice that length is immediately executed. For a more formal definition we may write $\mathcal{S}^{\text{univ}} = (t_1, t_2, t_3, \ldots)$, where

$$t_i = \begin{cases} 2^{k-1}, & \text{if } i = 2^k - 1; \\ t_{i-2^{k-1}+1}, & \text{if } 2^{k-1} \leq i < 2^k - 1. \end{cases}$$

Theorem 5 *For all distributions p,*

$$T(\mathcal{S}^{\text{univ}}, p) \leq 192\, \ell_p(\log(\ell_p) + 5).^2$$

Proof: The intuition for the bound is as follows. Let t_0 be a value that achieves the minimum value L_p in (11). Once the strategy has performed about $1/q(t_0)$ runs of length about t_0, it will have stopped with fairly high probability. At this point, the total time spent on runs of this length will be about $t_0/q(t_0) = L_p$. But the strategy is "balanced", in the sense that the total time spent on runs of each length is roughly equal. Since the number of different run lengths used up to this time is about $\log(L_p)$, the total running time is about $\mathrm{O}(L_p \log(L_p)) = \mathrm{O}(\ell_p \log(\ell_p))$.

[2] All logarithms in this paper are base 2.

To make this intuition precise, note that $\mathcal{S}^{\mathrm{univ}}$ has the property that, for any j, if the total time spent on runs of length 2^j up to the end of some run in the sequence is W, then at most $(\log(W)+1)$ different run lengths have so far been used, and the total time spent on each one cannot exceed $2W$. Thus the total time spent on runs of *all* lengths up to this point is at most $2W(\log(W)+1)$.

With t_0 defined as above, set $i_0 = \lceil \log(t_0) \rceil$ and $m_0 = \lceil \log(1/q(t_0)) \rceil$. Consider the instant when 2^{m_0} runs of length 2^{i_0} have been executed. The probability that A has failed to halt on all of these runs is at most

$$\left(1 - q(2^{i_0})\right)^{2^{m_0}} \leq \left(1 - q(t_0)\right)^{1/q(t_0)} \leq \mathrm{e}^{-1}.$$

At this point, the total time spent on runs of length 2^{i_0} is

$$W = 2^{m_0+i_0} \leq 4L_p \leq 16\ell_p, \tag{12}$$

and by the observation above the total time spent up to this point is at most $2W(\log(W)+1)$. More generally, after $k\,2^{m_0}$ runs of length 2^{i_0} have been completed, the probability that A has failed to halt is at most e^{-k}, and the total time spent up to this point is at most $2kW(\log(kW)+1)$. Therefore,

$$T(\mathcal{S}^{\mathrm{univ}}, p) \leq \sum_{k\geq 1} 2kW(\log(kW)+1)\mathrm{e}^{-k+1}.$$

Since $\log(kW)+1 \leq k(\log(W)+1)$ for all values of $k \geq 1$, we have

$$T(\mathcal{S}^{\mathrm{univ}}, p) \leq 2W(\log(W)+1)\sum_{k\geq 1} k^2\mathrm{e}^{-k+1}.$$

The theorem now follows from (12) and the fact that

$$\sum_{k\geq 1} k^2\mathrm{e}^{-k+1} = \frac{\mathrm{e}^2(\mathrm{e}+1)}{(\mathrm{e}-1)^3} \leq 6. \qquad \square$$

Remarks: (a) For clarity of presentation, little attempt has been made to minimize the constant 192 in the above theorem, and thus the true behavior of $\mathcal{S}^{\mathrm{univ}}$ is obviously much better than stated.

(b) Note that the above theorem makes no assumptions about p, other than that ℓ_p be finite. In particular, the expectation $\mathrm{E}[p]$ need not even be finite.

(c) In the strategy $\mathcal{S}^{\mathrm{univ}}$, we chose to increase the run lengths at each stage by a factor of 2. Of course, any other constant factor would do just as well, provided the number of runs is adjusted to preserve the property that the total running time for runs of each length is balanced. □

From the above proof, we also immediately obtain the following exponential tail bound on the running time of $\mathcal{S}^{\mathrm{univ}}$.

Theorem 6 *The probability that $\mathcal{S}^{\mathrm{univ}}$ runs for more than t steps is at most*

$$\exp\{-t/64\ell_p \log(t)\}. \qquad \square \tag{13}$$

Remark: In a recent paper [1], Alt *et al* study strategies for minimizing the tail probability in simulations of Las Vegas algorithms. In Theorem 4 of their paper, they present a strategy that achieves a tail bound of the form

$$\exp\{-ct/(\mathrm{E}[p]\log^2(\mathrm{E}[p])) + \log(t/c)\} \tag{14}$$

for some constant $c > 0$. As discussed in the introduction, our tail bound (13) is in many cases considerably sharper than (14) because it is expressed in terms of the quantity ℓ_p whereas (14) is given in terms of the expectation $\mathrm{E}[p]$, which may be much larger. Indeed, $\mathrm{E}[p]$ may even be infinite while ℓ_p is very small. The worst case from our point of view is when $\ell_p \approx \mathrm{E}[p]$; even in this case our bound is rather better than (14) as long as $t \leq \mathrm{E}[p]^{O(\log(\mathrm{E}[p]))}$, but it becomes rather worse for larger values of t. □

Finally, we show that Theorem 5 is actually the best we can hope to achieve, to within a constant factor, for unknown distributions.

Theorem 7 *For any strategy $\mathcal{S}$,*

$$\sup_{p:\ell_p=\ell} T(\mathcal{S},p) \geq \frac{1}{8}\ell\log(\ell).$$

Proof: For an arbitrary fixed value of ℓ, we construct a finite family of distributions p with $\ell_p = \ell$ such that any strategy $\mathcal{S}$ must satisfy $T(\mathcal{S},p) \geq \frac{1}{8}\ell\log(\ell)$ for some p in the family. Informally, these distributions will have the maximum possible weight (consistent with the constraint $\ell_p = \ell$) concentrated at the points $1, 2, 4, \ldots$.

Let $k = \lfloor\log(\ell)\rfloor$. There will be $k+1$ distributions in the family, the ith of which, $p^{(i)}$, is of the form

$$p_i(t) = \begin{cases} 2^i/\ell & \text{for } t = 2^i; \\ 1 - 2^i/\ell & \text{for } t = \infty; \\ 0 & \text{otherwise,} \end{cases}$$

for $0 \leq i \leq k$. Note that $\ell_{p^{(i)}} = \ell$ for all i.

Now we claim that at least $\frac{1}{2}(\ell/2^i)$ runs of length at least 2^i are required to ensure that the strategy stops on $p^{(i)}$ with probability at least $\frac{1}{2}$. To see this, note that the probability that the strategy fails to stop after $t < \frac{1}{2}(\ell/2^i)$ runs of this length is $(1 - 2^i/\ell)^t > \frac{1}{2}$. If a particular strategy fails to stop with probability at least $\frac{1}{2}$ after t steps, then then expected running time of that strategy is at least $t/2$. Therefore we must have, for any strategy $\mathcal{S}$,

$$\max_i\{T(\mathcal{S},p^{(i)})\} \geq \tau/2,$$

where τ is the minimum time required to perform $\frac{1}{2}(\ell/2^i)$ runs of length at least 2^i for $0 \leq i \leq k$. But it is easy to see that $\tau \geq \frac{1}{4}\ell\log(\ell)$, which completes the proof. □

It is not hard to modify the above proof to obtain a similar theorem (with a slightly smaller constant) when probabilistic mixed strategies are allowed.

Acknowledgements

We are grateful to Wolfgang Ertel for motivating our consideration of this problem and for his comments on an earlier version of the paper, to Dana Randall for valuable discussions that helped to clarify the presentation, and to Oded Goldreich and Dick Karp for their comments and suggestions.

References

[1] Helmut Alt, Leo Guibas, Kurt Mehlhorn, Richard Karp, and Avi Wigderson. A method for obtaining randomized algorithms with small tail probabilities. Technical Report TR-91-057, International Computer Science Institute, Berkeley, September 1991.

[2] Wolfgang Ertel. OR-Parallel Theorem Proving with Random Competition. *Proceedings of Logic Programming and Automated Reasoning,* St. Petersburg, July 1992, Springer Lecture Notes in AI Vol. 624, pp. 226–237.