Packing Random Intervals

E. G. Coffman, Jr.,[1] *Bjorn Poonen*[2] and *Peter Winkler*[3]

Sept. 22, 1994

ABSTRACT. Let n random intervals $I_1, \dots, I_n$ be chosen by selecting endpoints independently from the uniform distribution on $[0,1]$. A *packing* is a pairwise disjoint subset of the intervals; its *wasted space* is the Lebesgue measure of the points of $[0,1]$ not covered by the packing.

In any set of intervals the packing with least wasted space is computationally easy to find; but its expected wasted space in the random case is not obvious. We show that with high probability for large n, this "best" packing has wasted space $O(\frac{\log^2 n}{n})$.

It turns out that if the endpoints 0 and 1 are identified, so that the problem is now one of packing random arcs in a unit-circumference circle, then optimal wasted space is reduced to $O(1/n)$. Interestingly, there is a striking difference between the *sizes* of the best packings: about $\log n$ intervals in the unit interval case, but usually only one or two arcs in the circle case.

1. INTRODUCTION

Let $\{I_i\}$ be a set of n intervals in $[0,1]$, $I_i \subseteq [0,1]$, $1 \leq i \leq n$. A nonempty, pairwise disjoint subset of $\{I_i\}$ is called a *packing*; its *length* is the sum of the lengths of its intervals, and its *wasted space* (in $[0,1]$) is one minus its length. The problem of finding packings with minimum wasted space, a problem with important applications to be mentioned later, can be solved in polynomial time. A simple way to see this is to convert the problem to a well-known "easy" problem on graphs. Let the vertices of a graph G be the intervals I_i together with $I_0 = [-\infty, 0]$ and $I_{n+1} = [1, \infty]$; if I_j lies to the left of I_k then an edge is directed from I_j to I_k of length equal to the distance from I_j to I_k. Then finding a packing in $\{I_i\}$ with minimum wasted space is equivalent to finding a shortest path in G from I_0 to I_{n+1}.

[1]AT&T Bell Laboratories, Murray Hill, NJ 07974; egc@research.att.com
[2]Mathematical Sciences Research Inst., Berkeley, CA 94720; poonen@msri.org
[3]AT&T Bell Laboratories, Murray Hill, NJ 07974; pw@research.att.com

This paper studies the typical or average-case behavior of the wasted space for an optimal packing taken from a set of n random intervals. A standard model of random intervals is adopted: the endpointe of the packing. Thereafter, interval I_i is selected, i.e., added to the current partial packing, if and only if it is disjoint from the intervals already selected from the sequence $I_1, \ldots, I_{i-1}$. Coffman, Mallows, and Poonen [2] proved that $\mathrm{E}[N_n] = cn^\alpha - 1 + o(1)$ as $n \to \infty$, where $\alpha = \frac{\sqrt{17}-3}{4}$ and where $c \approx 1.84$ is obtained from an explicit, though complicated, formula. They also give precise asymptotics for the distribution of the gap lengths between selected intervals.

Lipton and Tomkins [4] have examined combinatorial problems of online interval packing; their competitive analysis computes worst-case bounds on the performance of online algorithms relative to an optimal offline algorithm. In addition to problems on graphs and partial orders, there are a number of other cognate problems relative to interval packing; these include problems of covering, parking, partitioning, and splitting. The connections are briefly discussed by Coffman, Mallows, and Poonen (1994), who give many references. They also discuss the application of interval packing problems to one-dimensional communication networks, an application extended by Lipton and Tomkins (1994) to continuous media (see also Long and Thakur [5]). In general terms, the intervals I_i in such applications are time intervals during which a resource is requested by some "customer"; the packing objective is to maximize the number of requests satisfied, or to pack requests that maximize resource utilization, under the constraint that satisfied requests be disjoint in time.

2. Lower bound on the minimum wasted space

Our model for this and the next section is a set $\{I_1, \ldots, I_n\}$ of n random intervals formed as above, with endpoints chosen independently from the uniform distribution on [0,1]. It will be useful to regard a packing henceforth as a *sequence* of intervals from $\{I_1, \ldots, I_n\}$ which are disjoint and ordered left to right in $[0, 1]$.

Fix $s > 0$, and let $\mathbf{X}$ be the random variable that counts the number of such packings with wasted space at most s. By definition of W_n, we have

$$\Pr(W_n \leq s) = \Pr(\mathbf{X} \geq 1) \leq \mathrm{E}\mathbf{X}, \tag{1}$$

so a lower bound for W_n that holds with high probability can be obtained by choosing an s for which $\mathrm{E}\mathbf{X}$ can be shown to approach zero.

Theorem 1. For any $\epsilon > 0$, if $s = (\frac{1}{8} - \epsilon)\log^2 n/n$, then $\mathrm{E}\mathbf{X}$ approaches zero as n tends to infinity.

Combining the theorem with (1) yields

Corollary 1. As $n \to \infty$,

$$\Pr\left(W_n \geq (\frac{1}{8} - \epsilon)\log^2 n/n\right) \to 1 \ .$$

In this section and the next, all real variables will be non-negative, and we write

$$\int \cdots \int_{x_1 + \cdots + x_n = 1} f(x_1, \ldots, x_n)$$

for

$$\int \cdots \int_{x_1 + \cdots + x_{n-1} \leq 1, x_i \geq 0} f(x_1, \ldots, x_{n-1}, 1 - x_1 - \cdots - x_{n-1}) dx_1 \cdots dx_{n-1} \ .$$

We will repeatedly need the value of the integral in the lemma below, which is proved easily by induction.

Lemma 1. If $a_1, \ldots, a_n$ are non-negative integers, $n \geq 2$, then

$$\int \cdots \int_{x_1 + \cdots + x_n = 1} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} = \frac{a_1! a_2! \cdots a_n!}{(a_1 + a_2 + \cdots + a_n + n - 1)!} \ .$$

Proof of Theorem 1. Let $\mathbf{X}_k$ denote the random variable which counts the number of packings using k intervals, with wasted space at most s, so

(2) $$\mathbf{X} = \sum_{k \geq 1} \mathbf{X}_k \ .$$

Let p_k be the probability that a sequence of k random intervals, chosen without replacement from $\{I_1, \ldots, I_n\}$, forms a packing with wasted space at most s. Then

(3) $$\mathrm{E}\mathbf{X}_k = n(n-1) \cdots (n-k+1) p_k \ .$$

Next we compute p_k. The probability that the chosen intervals are disjoint and in left to right order is $2^k/(2k)!$, the factor of 2^k arising from the fact that the endpoints of each single interval are automatically ordered correctly. (See Figure 1.)

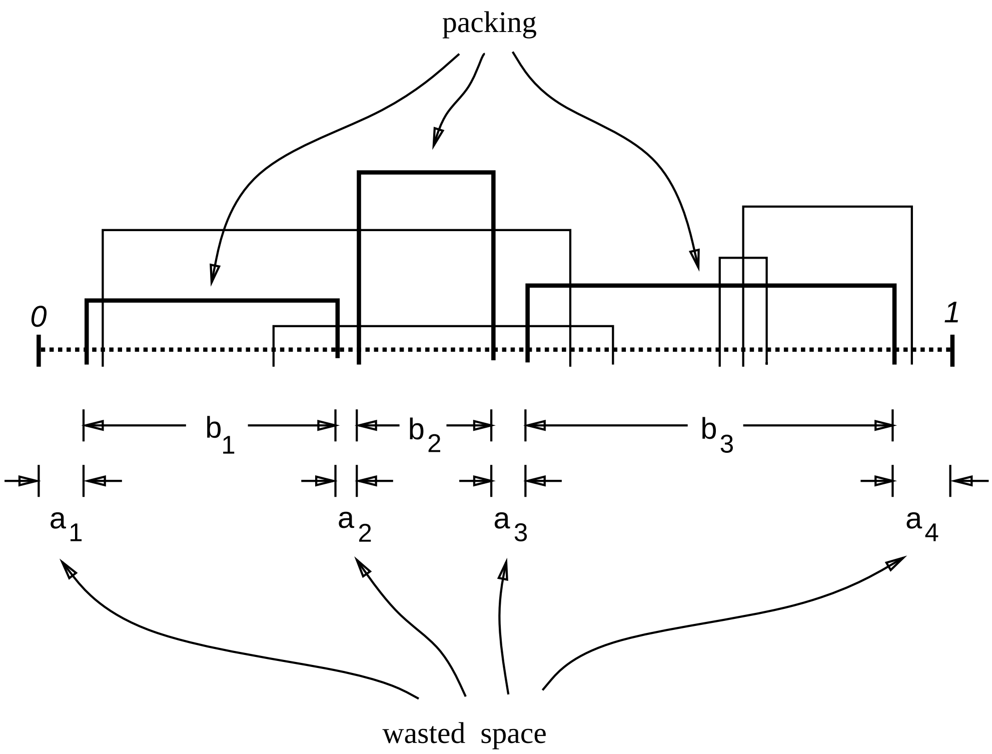

FIGURE 1. A 3-interval packing with lengths and gaps labelled

Suppose now that the chosen intervals are disjoint and in the correct order. Then their lengths $b_1, \ldots, b_k$ and gaps $a_1, \ldots, a_{k+1}$ are the coordinates of a random vector chosen uniformly from the simplex

$$\sum_{i=1}^{k+1} a_i + \sum_{j=1}^{k} b_j = 1, \qquad a_i, b_j \geq 0 \ , \tag{4}$$

since the a_i and b_j depend linearly on the endpoints of the intervals. This simplex

ic simple ∨ c $a \geq$ ∨ b J E ∨∨ c J J ∨∨ OF ∨∨ FO E 8 ∨ ∨ O E F F ∨∨ c O E 9 8 ∨∨ c N

$$\frac{\int_0^s x^k(1-x)^{k-1}dx}{\int_0^1 x^k(1-x)^{k-1}dx} .$$

Thus

$$p_k = \frac{2^k}{(2k)!} \cdot \frac{\int_0^s x^k(1-x)^{k-1}dx}{\int_0^1 x^k(1-x)^{k-1}dx} . \tag{5}$$

Combining (2), (3), and (5) yields

$$\begin{aligned}
\mathrm{E}\mathbf{X} &= \sum_{k=1}^{n} n(n-1)\cdots(n-k+1)\cdot\frac{2^k}{(2k)!}\cdot\frac{\int_0^s x^k(1-x)^{k-1}dx}{\int_0^1 x^k(1-x)^{k-1}dx} \\
&\le \sum_{k\ge 1} n^k\cdot\frac{2^k}{(2k)!}\cdot\frac{\int_0^s x^k dx}{k!(k-1)!/(2k)!} \qquad \text{(by Lemma 1)} \\
&= \sum_{k\ge 1}\frac{2^k n^k s^{k+1}}{k!(k-1)!(k+1)} \\
&\le s\sum_{k\ge 1}\frac{(2ns)^k}{k!^2} \\
&\le s\sum_{k\ge 1}\frac{(8ns)^k}{(2k)!} \qquad \left(\text{since } \binom{2k}{k}\le 2^{2k}\right) \\
&\le s\cdot\exp((8ns)^{1/2}) .
\end{aligned}$$

When $s = (\frac{1}{8}-\epsilon)\log^2 n/n$, this becomes

$$\begin{aligned}
\mathrm{E}\mathbf{X} &\le \frac{1}{8}\frac{\log^2 n}{n}\cdot\exp((1-8\epsilon)^{1/2}\log n) \\
&< \frac{1}{8}\left(n^{-\epsilon'}\log^2 n\right)
\end{aligned}$$

for $0<\epsilon'<1-(1-8\epsilon)^{1/2}$. Since this tends to zero as $n\to\infty$, the theorem is proved. □

3. Upper bound on the minimum wasted space

Fix $t>0$. We shall say that subintervals $J_1,\ldots,J_k$ of $[0,1]$ form a *good* packing of an interval $[\alpha,\beta]\subset[0,1]$ if they are disjoint subintervals of $[\alpha,\beta]$, they are in left to right order, the gaps between the intervals are of size between $t/2$ and t, and the distance from α to the first interval and the distance from the last interval to β are both less than or equal to t. A good packing has wasted space at most $(k+1)t$. The reason for the rather contrived definition is that it will be easier to prove the existence of good packings (in Theorem 2 below) than the existence of arbitrary packings with wasted space at most $(k+1)t$.

Now suppose $I_1, \ldots, I_n$ are random subintervals of $[0,1]$. Fix an integer $k \geq 1$ (which may depend on n). Let $\mathbf{Y}$ be the random variable that counts the number of k-tuples of intervals chosen from these n which define a good packing of $[0,1]$. We will show that $\mathrm{E}\mathbf{Y} \to \infty$ as $n \to \infty$ for $t = c \log n/n$ and appropriate k. However, this is not enough to conclude that $\Pr(\mathbf{Y} = 0) \to 0$.

One normally expects the second moment method (see e.g. [1]) to come to the rescue in this sort of situation. From Chebyshev's inequality, we have

$$\Pr(\mathbf{Y} = 0) \leq \frac{\sigma^2(\mathbf{Y})}{(\mathrm{E}\mathbf{Y})^2}$$

and therefore if we could show that the variance of $\mathbf{Y}$ vanishes relative to $(\mathrm{E}\mathbf{Y})^2$, we would have the desired result, giving an upper bound complementary to the lower bound of Corollary 1. It turns out, however, that $\sigma^2(\mathbf{Y})/(\mathrm{E}\mathbf{Y})^2$ is bounded away from 0 for fixed c. Hence we are forced to conduct a more thorough analysis of the second moment, estimating its dependence on c, in order to prove that $\mathrm{E}W_n = \mathrm{O}(\log^2 n/n)$.

Let $Q_m(x)$ denote the probability that a random m-tuple of subintervals of $[0,1]$ defines a good packing of a pre-specified interval of length x. The following lemma will be used in the estimation of $Q_m(x)$.

Lemma 2. Let $(x_1, x_2, \ldots, x_a, y_1, y_2, \ldots, y_b)$ be a point chosen randomly and uniformly from the $(a+b-1)$-dimensional simplex where $x_i, y_j \geq 0$ and

$$x_1 + x_2 + \cdots + x_a + y_1 + y_2 + \cdots + y_b = 1 \ .$$

Let $P_{a,b}(r)$ denote the probability that, for such a point,

(1) $x_1, x_a \leq r$,
(2) $r/2 \leq x_i \leq r$ for all i, $2 \leq i \leq a-1$, and
(3) $y_j \geq r$ for all j.

Then

$$P_{a,b}(r) \leq 2^{2-a} r^a (a+b-1)!/(b-1)! \ ,$$

with equality up to a factor of $1 + \mathrm{O}((a+b)^2 r)$.

Proof. For fixed b, the $(b-1)$-dimensional volume of the simplex $z_1+z_2+\cdots+z_b = c$, $z_j \geq 0$, is proportional to c^{b-1}. For fixed $(x_1,\ldots,x_a)$, the y_j's must satisfy $y_1+\cdots+y_b = 1-(x_1+\cdots+x_a)$, but the space measured by $P_{a,b}(r)$ consists of y_j's where $y_j = r+z_j$, $z_j \geq 0$, and $z_1+\cdots+z_b = 1-(x_1+\cdots+x_a+br)$, so

$$
\begin{aligned}
&P_{a,b}(r)\\
(6)\quad &\leq \frac{\int\cdots\int_{x_1+\cdots+x_a\leq 1,\ x_2,\ldots,x_{a-1}\geq r/2,\ x_i\leq r}[1-(x_1+\cdots+x_a+br)]^{b-1}dx_1\cdots dx_a}{\int\cdots\int_{x_1+\cdots+x_a\leq 1}[1-(x_1+\cdots+x_a)]^{b-1}dx_1\cdots dx_a}\\
(7)\quad &\leq \frac{\int\cdots\int_{x_2,\ldots,x_{a-1}\geq r/2,\ x_i\leq r} 1}{\int\cdots\int_{x_1+\cdots+x_a\leq 1}[1-(x_1+\cdots+x_a)]^{b-1}dx_1\cdots dx_a}\\
(8)\quad &= \frac{r^2\cdot(r/2)^{a-2}}{(b-1)!/(a+b-1)!} \qquad \text{(by Lemma 1)}\\
(9)\quad &= 2^{2-a}r^a(a+b-1)!/(b-1)!\ .
\end{aligned}
$$

Moreover, if $(a+b)r < 1$, we may drop the restriction $x_1+\cdots+x_a \leq 1$ from the range of integration in the numerator of (6), and substitute

$$[1-(x_1+\cdots+x_a+br)]^{b-1} = 1-\mathrm{O}((a+b)^2r)\ ,$$

since $x_1,\ldots,x_a \leq r$. Then we get equality up to a factor of $1+\mathrm{O}((a+b)^2r)$. $\square$

Lemma 3. For $m \geq 1$, and $0 < x \leq 1$,

$$Q_m(x) \leq 2t^{m+1}x^{m-1}/(m-1)!\ .$$

When $x = 1$, equality holds up to a factor of $1+\mathrm{O}(m^2t)$.

Proof. The probability that the m random subintervals of $[0,1]$ lie inside the prespecified interval I of length x is the probability that $2m$ independent random numbers in $[0,1]$ (the interval endpoints) lie in I, which is x^{2m}. The probability that the $2m$ numbers are in order, and hence define a collection of m disjoint intervals in order, is $2^m/(2m)!$ as in Section 2 above. The probability that m

random subintervals of $[0,1]$ define a good packing of I, given that they lie in I, are disjoint, and are in order, is exactly $P_{m+1,m}(t/x)$, since the distribution of the lengths of the intervals divided by x (which we call $y_1,\dots,y_m$) and the lengths of the gaps divided by x (which we call $x_1,\dots,x_{m+1}$) are distributed exactly as in the definition of P. To see this, note that a linear transformation takes the space of possible $(x_1,\dots,x_{m+1},y_1,\dots,y_m)$ to the space of possible endpoints of intervals, and P gives the probability that such a $(2m+1)$-tuple will correspond to a good packing. Thus

$$\begin{aligned} Q_m(x) &= 2^m/(2m)! \cdot x^{2m} \cdot P_{m+1,m}(t/x) \\ &\le 2^m/(2m)! \cdot x^{2m} \cdot 2^{-m+1}(t/x)^{m+1}(2m)!/(m-1)! \\ &= 2t^{m+1}x^{m-1}/(m-1)!, \end{aligned}$$

with equality up to a factor of $1 + \mathrm{O}((m+(m+1))^2(t/x)) = 1 + \mathrm{O}(m^2 t)$, when $x = 1$. $\square$

Corollary 2.

$$\mathrm{E}\mathbf{Y} = [1 + \mathrm{O}(k^2 t + k^2/n)] \cdot 2n^k t^{k+1}/(k-1)!\ .$$

Proof. The number of k-tuples of distinct intervals from the n randomly chosen ones is

$$n(n-1)\cdots(n-k+1) = [1-\mathrm{O}(k/n)]^k n^k = [1-\mathrm{O}(k^2/n)]n^k\ ,$$

and each k-tuple is a good packing with probability $Q_k(1)$, so by Lemma 3,

$$\begin{aligned} \mathrm{E}\mathbf{Y} &= [1-\mathrm{O}(k^2/n)]n^k \cdot [1+\mathrm{O}(k^2 t)] \cdot 2t^{k+1}/(k-1)! \\ &= [1+\mathrm{O}(k^2 t + k^2/n)] \cdot 2n^k t^{k+1}/(k-1)!\ . \end{aligned}$$

$\square$

Corollary 3. If $k = \lfloor \log n \rfloor$, and $t = c\log n/n$, for some constant $c > 1$, then $\mathrm{E}\mathbf{Y} \to \infty$ as $n \to \infty$.

Proof. From the previous corollary, we get

$$\begin{aligned}
\log \mathrm{E}\mathbf{Y} &= k\log(nt) + \log t - (k\log k - k) + \mathrm{O}(1)\\
&\geq k\log(ck) - \log n - k\log k + k + \mathrm{O}(1)\\
&= k\log c + \mathrm{O}(1)\\
&\to \infty.
\end{aligned}$$

□

Theorem 2. The probability that among $I_1, \ldots, I_n$ there exist $k = \lfloor \log n \rfloor$ intervals that form a good packing with maximum allowed gap size $t = c\log n/n$ for some constant $c > 0$, is at least $1 - 28/c$ for sufficiently large n.

Corollary 4. If c is any positive constant,

$$\Pr(W_n \leq c\log^2 n/n) \geq 1 - 28/c$$

for sufficiently large n.

Proof. Indeed, the theorem guarantees with probability at least $1 - 28/c$ the existence of a packing with $\lfloor \log n \rfloor + 1$ gaps each of size at most $c\log n$, and such a packing has total wasted space at most $c(\log^2 n + \log n)/n$. This is slightly weaker than what is needed, but it can be checked that the proof of the theorem goes through with 28 replaced by 27.999 and this strengthening is enough to imply the result. Alternatively, it can be checked that the proof of the theorem still goes through if the number of intervals to be used in the packing is one less. □

Proof of Theorem 2. In order to bound $\Pr(\mathbf{Y} = 0)$, we need to compute $\mathrm{E}(\mathbf{Y}^2)$ and compare it to $(\mathrm{E}\mathbf{Y})^2$. Now $\mathrm{E}(\mathbf{Y}^2)$ counts ordered pairs of k-interval good packings $(\mathcal{P}_1, \mathcal{P}_2)$ taken from the same random n subintervals of $[0,1]$. We will classify such pairs according to the pattern of intervals they share.

First there are the pairs in which the two packings share no intervals. The expected number of these is the number of $(2k)$-tuples of intervals chosen from the n intervals, times the probability that the first k intervals and the last k intervals of a $(2k)$-tuple each form a good packing, which is

$$n(n-1)\cdots(n-2k+1)\cdot Q_k(1)^2 \le [n(n-1)\cdots(n-k+1)\cdot Q_k(1)]^2 = (\mathrm{E}\mathbf{Y})^2, \tag{10}$$

by the proof of Corollary 2.

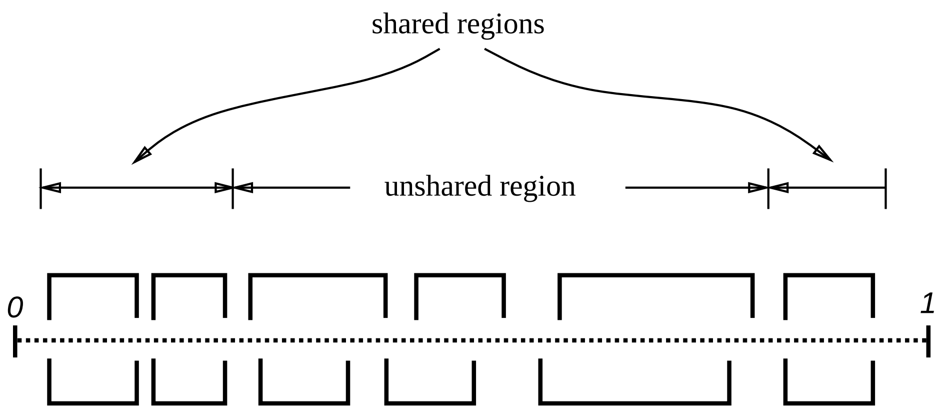

FIGURE 2. A pair of packings, with $u = 2$, $v = 1$

For the other pairs of packings, there will be alternating regions of shared and unshared intervals, as in Figure 2, for example. Let u be the number of regions of shared intervals, and v the number of unshared regions, for a particular configuration. (Then $|u-v| \le 1$). Let a_i denote the number of intervals used in the i-th shared region, and let b_j, c_j denote the number of intervals used in the j-th unshared region by $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. Let $\ell = a_1 + \ldots + a_u$ denote the total number of shared intervals. Then

$$b_1 + \ldots + b_v = c_1 + \ldots + c_v = k - \ell, \tag{11}$$

and the total number of intervals used in both packings is $2k - \ell$.

Now fix the order of shared and unshared regions, u, v, and all the a_i, b_j, and c_j (hence ℓ as well). We call this a *configuration.* Let p be the probability that a $(2k-\ell)$-tuple of random subintervals of $[0,1]$ forms a pair of good packings in this configuration, with the first k intervals in order forming $\mathcal{P}_1$, and the last $k-\ell$ in order being the remaining intervals (counted by the c_j's).

In order to bound p, consider the probability P that when $2k-\ell$ random subintervals of $[0,1]$ and $u+v-1$ random numbers $\alpha_1,\ldots,\alpha_{u+v-1}$ in $[0,1]$ are chosen, that the intervals in order form a pair of packings as above, and that the α_i in order demarcate the shared and unshared regions. Given that the intervals form a pair of packings as above, the conditional probability that $\alpha_1,\ldots,\alpha_{u+v-1}$ demarcate the $u+v$ regions is at least $(t/2)^{u+v-1}$, since the gap between the last interval of a region and the first interval of the next region (in either packing) is at least $t/2$ (by definition of a good packing), giving each α_i a range of length at least $t/2$ in which it may lie. These ranges are disjoint (and in order), since each region has length at least t (because the definition of good packing requires that any interval in the region has length at least t). Hence

$$P \geq p \cdot (t/2)^{u+v-1}. \tag{12}$$

On the other hand, for fixed $\alpha_1,\ldots,\alpha_{u+v-1}$, the probability that $2k-\ell$ random intervals form a pair of good packings of the specified configuration with regions demarcated by the α_i is at most the probability that the intervals of each packing with each region form a good packing (with the right number of intervals), which is the product of all $Q_{a_i}(x_i)$ and $Q_{b_j}(y_j)Q_{c_j}(y_j)$ where x_i is the difference of α's giving the length of the i-th shared region, and where y_j is similarly the length of the j-th unshared region. Hence P is bounded by a multiple integral over the α's of a product of Q functions in their differences, which after a change of variables to the x_i and y_j becomes

(13)

$$\begin{aligned}
P &\le \int\cdots\int_{x_1+\cdots+x_u+y_1+\cdots+y_v=1} \prod_{i=1}^{u} Q_{a_i}(x_i) \prod_{j=1}^{v} Q_{b_j}(y_j) Q_{c_j}(y_j) \\
&\le \int\cdots\int_{x_1+\cdots+x_u+y_1+\cdots+y_v=1} \prod_{i=1}^{u} \frac{2t^{a_i+1}x_i^{a_i-1}}{(a_i-1)!} \prod_{j=1}^{v} \frac{2t^{b_j+1}y_j^{b_j-1}\cdot 2t^{c_j+1}y_j^{c_j-1}}{(b_j-1)!\cdot(c_j-1)!} \\
&\qquad \text{(by Lemma 3)} \\
&= \frac{2^{u+2v} t^{\sum_{i=1}^{u}(a_i+1)+\sum_{j=1}^{v}(b_j+c_j+2)} \int\cdots\int_{\sum x_i+\sum y_j\ =1} \prod_{i=1}^{u} x_i^{a_i-1} \prod_{j=1}^{v} y_j^{b_j+c_j-2}}{\prod_{i=1}^{u}(a_i-1)! \prod_{j=1}^{v}(b_j-1)!(c_j-1)!} \\
&= \frac{2^{u+2v} t^{2k-\ell+u+2v} \prod_{i=1}^{u}(a_i-1)! \prod_{j=1}^{v}(b_j+c_j-2)!}{\prod_{i=1}^{u}(a_i-1)! \prod_{j=1}^{v}(b_j-1)!(c_j-1)! \cdot (\sum_{i=1}^{u} a_i + \sum_{j=1}^{v}(b_j+c_j-1)-1)!} \\
&\qquad \text{(by Lemma 1)} \\
&= 2^{u+2v} t^{2k-\ell+u+2v} \prod_{j=1}^{v} \binom{b_j+c_j-2}{b_j-1} / (2k-\ell-v-1)! \\
&\le 2^{u+2v} t^{2k-\ell+u+2v} \prod_{j=1}^{v} \binom{b_j+c_j-2}{b_j-1} \cdot (2k)^{\ell+v-1}/(2k-2)!\ ,
\end{aligned}$$

since

$$\begin{aligned}
(2k-2)! &= (2k-2)(2k-3)\cdots(2k-\ell-v)\cdot(2k-\ell-v-1)! \\
&\le (2k)^{\ell+v-1}\cdot(2k-\ell-v-1)!\ .
\end{aligned}$$

Combining (12) and (13) yields

$$p\cdot(t/2)^{u+v-1} \le 2^{u+2v} t^{2k-\ell+u+2v} \prod_{j=1}^{v} \binom{b_j+c_j-2}{b_j-1} \cdot (2k)^{\ell+v-1}/(2k-2)!$$

(14) $$p \le 2^{2u+3v-1} t^{2k-\ell+v+1} \prod_{j=1}^{v} \binom{b_j+c_j-2}{b_j-1} \cdot (2k)^{\ell+v-1}/(2k-2)!.$$

Now the number of $(2k-\ell)$-tuples is $n(n-1)\cdots(n-(2k-\ell)+1) \le n^{2k-\ell}$;

combining this with the bound for p in (14), the expected number of pairs of packings of the specified configuration coming from all $(2k-\ell)$-tuples of intervals chosen from $\{I_1, \ldots, I_n\}$ is at most

$$2^{2u+3v-1} n^{2k-\ell} t^{2k-\ell+v+1} \prod_{j=1}^{v} \binom{b_j + c_j - 2}{b_j - 1} \cdot (2k)^{\ell+v-1}/(2k-2)!, \tag{15}$$

and this divided by $(\mathrm{E}\mathbf{Y})^2$ is, by Corollary 2, at most

(16)

$$[1 + \mathrm{O}(k^2 t + k^2/n)] \cdot 2^{2u+3v-3} (nt)^{-\ell} t^{v-1} \prod_{j=1}^{v} \binom{b_j + c_j - 2}{b_j - 1} \cdot (2k)^{\ell+v-1} \Big/ \binom{2k-2}{k-1}.$$

Let N_v denote the expected number of pairs of packings of any configuration in which the number of unshared regions is v, so

$$\mathrm{E}(\mathbf{Y}^2) = \sum_{v=0}^{\infty} N_v. \tag{17}$$

First of all, N_0 counts the number of pairs of identical packings, so

$$N_0 = \mathrm{E}\mathbf{Y}. \tag{18}$$

The configurations in which $v = 1$ are those with a_1 shared intervals on the left and a_2 shared intervals on the right, and with each packing having $b_1 = c_1 = k - (a_1 + a_2)$ intervals in the middle unshared region, or those which are similar except that one or both of the shared regions is empty. The expected number of pairings with $v = 1, \ell = 0$ is at most $(\mathrm{E}\mathbf{Y})^2$ by (10). For $\ell \geq 1$, the number of such configurations with ℓ intervals shared is $\ell + 1$ (the number of ways to write ℓ as $a_1 + a_2$ with $a_1, a_2 \geq 0$), and from (16), the expected number of pairs of good packings in a particular configuration, divided by $(\mathrm{E}\mathbf{Y})^2$, is at most

$$\begin{aligned}
&[1 + \mathrm{O}(k^2 t + k^2/n)] \cdot 2^{2u} (nt)^{-\ell} \binom{b_1 + c_1 - 2}{b_1 - 1} \cdot (2k)^{\ell} \Big/ \binom{2k-2}{k-1} \\
&\qquad \leq [1 + \mathrm{O}(k^2 t + k^2/n)] \cdot (8k/nt)^{\ell} \ ,
\end{aligned}$$

since $u \leq \ell$ and $b_1, c_1 \leq k$ implies

$$\binom{b_1 + c_1 - 2}{b_1 - 1} \leq \binom{2k-2}{k-1} \ ,$$

so

$$(19)\qquad N_1/(\mathrm{E}\mathbf{Y})^2 \;\le\; 1+\sum_{\ell=1}^{\infty}(\ell+1)\cdot[1+\mathrm{O}(k^2t+k^2/n)]\cdot(8k/nt)^{\ell}$$

$$= \; 1+[1+\mathrm{O}(k^2t+k^2/n)]\cdot[(1-8k/nt)^{-2}-1],$$

provided we assume

$$(20)\qquad 8k/nt<1\ .$$

(Later we will specify more precise values for k and t.)

Now for fixed $v\ge 2$, there are 4 ways in which the regions at the ends of a configuration can be shared or unshared, and at most k possibilities for each of a_i, b_i and c_i; hence the number of configurations is at most $4k^{u+2v}$. Moreover $u\le v+1$, so the number of configurations for fixed v is at most $4k^{3v+1}$. From (16), the expected number of pairs of good packings in a particular one of these configurations, divided by $(\mathrm{E}\mathbf{Y})^2$, is at most

$$[1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{2u+3v-3}(nt)^{-\ell}t^{v-1}\prod_{j=1}^{v}\binom{b_j+c_j-2}{b_j-1}\cdot(2k)^{\ell+v-1}\Big/\binom{2k-2}{k-1}$$

$$\le\; \frac{[1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{2(v+1)+3v-3}k^{v-1}(2k/nt)^{\ell}t^{v-1}\prod_{j=1}^{v}2^{b_j+c_j-2}\cdot 2^{\ell+v-1}}{2^{2k-2}/k}$$

(since $\binom{2m}{m}\ge 2^{2m}/(m+1)$, as is easily proved by induction)

$$\le\; [1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{5v-1}k^{v-1}t^{v-1}\cdot 2^{2k-2\ell-2v}\cdot 2^{\ell+v-1}\cdot 2^{2-2k}k$$

(by (11) and (20))

$$=\; [1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{-\ell+4v}k^{v}t^{v-1}$$

$$\le\; [1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{4v}k^{v}t^{v-1}\ .$$

Thus

$$
\begin{aligned}
N_v/(\mathrm{E}\mathbf{Y})^2 &\le 4k^{3v+1}\cdot[1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{4v}k^vt^{v-1}\\
&\le [1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{4v+2}k^{4v+1}t^{v-1},
\end{aligned}
$$

$$
\begin{aligned}
\text{(21)}\qquad \sum_{v\ge 2} N_v/(\mathrm{E}\mathbf{Y})^2 &\le [1+\mathrm{O}(k^2t+k^2/n)]\cdot 2^{10}k^9t/(1-2^4k^4t)\\
&= [1+\mathrm{O}(k^2t+k^2/n)]\cdot \mathrm{O}(k^9t).
\end{aligned}
$$

Substituting (18), (19) and (21) into (17) yields

(22)

$$\mathrm{E}(\mathbf{Y}^2)/(\mathrm{E}\mathbf{Y})^2 = 1/\mathrm{E}\mathbf{Y} + 1 + [1+\mathrm{O}(k^2t+k^2/n)]\cdot[(1-8k/nt)^{-2}-1+\mathrm{O}(k^9t)]\ .$$

When we set $k=\lfloor \log n\rfloor$ and $t=c\log n/n$ for some constant $c\ge 28$ (the theorem is trivial for $c<28$), and use Corollary 3, this becomes

$$\mathrm{E}(\mathbf{Y}^2)/(\mathrm{E}\mathbf{Y})^2 = (1+o(1))\cdot(1-2/c)^{-2}$$

as $n\to\infty$. An easy calculation shows that

$$(1-8/c)^{-2} < 1+28/c$$

for $c\ge 28$, so

$$\mathrm{E}(\mathbf{Y}^2)/(\mathrm{E}\mathbf{Y})^2 < 1+28/c$$

provided n is sufficiently large. We now apply Chebyshev's inequality to get

$$\Pr(\mathbf{Y}=0)\le \frac{\sigma^2(\mathbf{Y})}{(\mathrm{E}\mathbf{Y})^2} = \frac{(\mathrm{E}\mathbf{Y}^2)-(\mathrm{E}\mathbf{Y})^2}{(\mathrm{E}\mathbf{Y})^2} < 28/c$$

and therefore

$$\Pr(\mathbf{Y}>0) > 1-28/c$$

for sufficiently large n. $\square$

The bound $1-28/c$ of the theorem was chosen to simplify the presentation. A more judicious choice of k leads to a tighter probability bound of the form $1-\mathrm{O}(\frac{1}{c\log c})$ in the theorem, and $1-\mathrm{O}(\frac{1}{c\log^2 c})$ in its corollary. This improvement is used in the proof of the following.

Theorem 3. As $n\to\infty$, $\mathrm{E}[W_n]=\Theta\left(\frac{\log^2 n}{n}\right)$.

Proof. Corollary 1 implies that $\mathrm{E}[W_n] = \Omega\left(\frac{\log^2 n}{n}\right)$. We will use

$$\mathrm{E}[W_n] = \int_0^1 \Pr(W_n \geq s)ds \tag{23}$$

to get an upper bound. We divide the range of integration into three regions, where $c_0 > 0$ is a constant to be chosen later:

$$0 \leq s \leq \frac{c_0 \log^2 n}{n}, \quad \frac{c_0 \log^2 n}{n} \leq s \leq n^{-3/5}, \quad \text{and } n^{-3/5} \leq s \leq 1 \ .$$

In the first region we simply use

$$\Pr(W_n \geq s) \leq 1, \tag{24}$$

which is enough to get

$$\int_0^{c_0 \log^2 n/n} \Pr(W_n \geq s)ds = \mathrm{O}(\log^2 n/n) \ . \tag{25}$$

For the second region, we borrow results from the proof of Theorem 2. The method there (in particular, (22) and Chebyshev's inequality) implies that

$$\begin{aligned} \Pr(W_n &\geq (k+1)t) \\ &\leq \Pr(\mathbf{Y} = 0) \\ &\leq \frac{1}{\mathrm{E}\mathbf{Y}} + \left[1 + \mathrm{O}\left(k^2 t + \frac{k^2}{n}\right)\right] \cdot \left[\left(1 - \frac{8k}{nt}\right)^{-2} - 1 + \mathrm{O}(k^9 t)\right] \\ &= \frac{1}{\mathrm{E}\mathbf{Y}} + \left[1 + \mathrm{O}\left(k^2 t + \frac{k^2}{n}\right)\right] \cdot \left[\mathrm{O}\left(\frac{k}{nt}\right) + \mathrm{O}(k^9 t)\right] \end{aligned} \tag{26}$$

provided k/nt is sufficiently small. Take

$$k = \lfloor \frac{3 \log n}{\log c} \rfloor, \ t = c\frac{\log^2 n/n}{k+1}$$

for some c, $c_0 \leq c \leq n^{2/5}/\log^2 n$. If c_0 is sufficiently large, then for c in this range

$$\frac{k}{nt} = \mathrm{O}\left(\frac{1}{c \log^2 c}\right)$$

is sufficiently small, $O(k^2 t)$ and $O(k^2/n)$ are $O(1)$, and $O(k^9 t)$ is at most $O(c \log^{10} n/n)$. Moreover, by Corollary 2, for c sufficiently large,

$$\begin{aligned} \log \mathrm{E}\mathbf{Y} &= k\log(nt) + \log t - (k\log k - k) + O(1) \\ &\geq k\log(nt/k) - \log n + k + O(1) \\ &\geq k\log c - \log n + O(1) \\ &\geq (2\log n/\log c)\log c - \log n + O(1) \\ &\geq \log n + O(1) . \end{aligned}$$

Thus $1/\mathrm{E}\mathbf{Y} = O(1/n)$ and (26) becomes

$$\Pr\left(W_n \geq c\log^2 n/n\right) = O\left(\frac{1}{n} + \frac{1}{c\log^2 c} + \frac{c\log^{10} n}{n}\right) . \tag{27}$$

Hence

$$\begin{aligned} &\int_{s=c_0\log^2 n/n}^{n^{-3/5}} \Pr(W_n \geq s)ds \\ &= \int_{c=c_0}^{n^{2/5}/\log^2 n} O\left(\frac{1}{n} + \frac{1}{c\log^2 c} + \frac{c\log^{10} n}{n}\right) d(c\log^2 n/n) \\ &= O\left(\frac{n^{-3/5}}{\log^2 n} + 1 + \left(\frac{n^{2/5}}{\log^2 n}\right)^2 \frac{\log^{10} n}{n}\right)\frac{\log^2 n}{n} \\ &= O\left(\frac{\log^2 n}{n}\right) . \end{aligned} \tag{28}$$

For the third region, we show that there exists a $\beta > 0$ such that

$$\Pr(W_n \geq n^{-3/5}) = O(e^{-\beta n^{1/5}}) . \tag{29}$$

In fact, we show that there exists a packing of *two* intervals with wasted space less than $n^{-3/5}$ with high probability (that is, with probability one minus a quantity exponentially small in $n^{1/5}$). In what follows we make tacit use of standard Chernoff bounds (or easy extensions of such bounds) on sums of Bernoulli random variables; see e.g. Appendix A of Alon and Spencer [1], in particular Corollary A.14, p. 239.

Set $t = \frac{1}{3}n^{-3/5}$. Each of the first $\lfloor n/2 \rfloor$ intervals has left endpoint in $[0,t]$ and right endpoint in $[1/3, 2/3]$ with probability $2t/3$, and these events are independent, so there will be at least $nt/6$ such intervals with high probability. Take the first $N \stackrel{\text{def}}{=} \lfloor nt/6 \rfloor$ of them, and associate to each the interval of length t to the right of its right endpoint. Since $Nt = \mathrm{O}(n^{-1/5}) < 1/6$, each of the associated intervals is disjoint from all the previous ones with probability at least $1/2$. Hence at least $N/4$ of them are disjoint, with high probability, and their union S is of total length at least $(N/4)t$. Each of the remaining $\lceil n/2 \rceil$ original intervals has its left endpoint in S and its right endpoint in $[1-t, 1]$ with probability at least $(N/4)t \cdot t = \mathrm{O}(n^{-4/5})$, so there will be $\Omega(n^{1/5})$ of them, with high probability. Any of these intervals I_2, together with the interval I_1 whose associated interval contains the left endpoint of I_2, forms a packing with wasted space at most $3t = n^{-3/5}$, proving (29).

Finally (29) implies that

$$\int_{n^{-3/5}}^{1} \Pr(W_n \geq s) ds = \mathrm{O}(e^{-\beta n^{1/5}}) \ .$$

Combining this with (23), (25), and (28) shows that $\mathrm{E}[W_n] = \mathrm{O}(\frac{\log^2 n}{n})$, as desired. $\square$

4. The circle case

For contrast—both in ease and in results—we consider the effect of replacing the unit interval by a unit-circumference circle, say by identifying 0 with 1 in [0,1]. Then we define arcs $I_1, \ldots, I_n$ instead of intervals, letting I_i be directed left to right from X_{2i-1} to X_{2i}. It turns out that the arc-packings behave better than the interval-packings with respect to independence, and thus a more precise quantitative analysis is possible; we content ourselves below with a rough statement of the behavior of optimal arc-packings.

Again we choose randomly a sequence of k arcs from among $I_1, \ldots, I_n$; here we obtain a packing (disjoint arcs in increasing order, mod 1) with probability $2k/(2k)!$ since the X_i's must be in circular order. In that case there are k lengths $b_1, \ldots, b_k$ but only k gaps $a_1, \ldots, a_k$, letting $a_1 = X_1 - X_{2n}$ (mod 1); the wasted space is $\sum a_i$. We have:

Theorem 4. Let n random arcs be chosen in a circle, their endpoints taken independently and uniformly. Then for any $\epsilon > 0$ there are reals α and β and an integer m such that for all n, with probability at least $1 - \epsilon$, the number of arcs in the optimal packing will be less than m and its wasted space W_n will satisfy $\alpha/n < W_n < \beta/n$.

Proof. Let $\mathbf{X}_k$ be the number of packings of k arcs with wasted space less than α/n, and let $\mathbf{X} = \sum_{k=1}^{n} \mathbf{X}_k$.

Proceeding as in Section 2, we have

$$\begin{aligned} \mathrm{E}\mathbf{X}_k &= n(n-1)\cdots(n-k+1) \cdot \frac{2k}{(2k)!} \cdot \frac{\int_0^{\alpha/n} x^{k-1}(1-x)^{k-1}dx}{\int_0^1 x^{k-1}(1-x)^{k-1}dx} \\ &\leq n^k \cdot \frac{2k}{(2k)!} \cdot \frac{\int_0^{\alpha/n} x^{k-1}dx}{(k-1)!^2/(2k)!} \qquad \text{(by Lemma 1)} \\ &= \frac{\alpha^k}{k(k-1)!^2} \end{aligned}$$

and thus in particular

$$\mathrm{E}\mathbf{X} = \sum_{k=1}^{n} \mathrm{E}\mathbf{X}_k \leq \sum_{k=1}^{n} \alpha^k \leq \alpha/(1-\alpha) \ .$$

Taking $\alpha = (\epsilon/3)/(1 + (\epsilon/3))$ ensures that $\Pr(W_n < \alpha/n) < \epsilon/3$.

On the other hand, the wasted space of a single random arc is uniform on [0,1]; hence taking $k = 1$ we already have

$$\Pr(W_n \leq \frac{\beta}{n}) \geq 1 - (1 - \frac{\beta}{n})^n > 1 - e^{-\beta}$$

so that, taking $\beta = -\log(\epsilon/3)$, we obtain $\Pr(W_n > \beta/n) < \epsilon/3$ as well.

Finally we choose m so that

$$\sum_{k=m}^{\infty} \frac{\beta^k}{k(k-1)!^2} < \epsilon/3$$

guaranteeing that with probability at least $\epsilon/3$ no packing of m or more arcs will have wasted space less than β/n. $\square$

ACKNOWLEDGMENT

This work was researched and written in part while the second author was visiting AT&T Bell Laboratories and the third was working for Bellcore.

References

1. Alon, N. and Spencer, J. H. (1991), *The Probabilistic Method*, Wiley & Sons, New York.
2. Coffman, E. G., Jr., Mallows, C. L., and Poonen, Bjorn (1994), "Parking Arcs on the Circle with Applications to One-Dimensional Communication Networks," *Ann. Appl. Prob.*, **4** (4), in press.
3. Justicz, J., Scheinerman, E. R., and Winkler, P. M. (1990), "Random Intervals," *Amer. Math. Monthly*, **97**, 881–889.
4. Lipton, R. J. and Tomkins, Andrew (1994), "Online Interval Scheduling," *Proc. 5th Ann. ACM-SIAM Symp. Disc. Alg.*, 302–311.
5. Long, D. and Thakur, M. (1993), "Scheduling Real-Time Disk Transfers for Continuous Media Applications," *12th Ann. IEEE Symp. Mass Stor. Sys.*, 227–232.