Analysis of a Recurrence Arising from a Construction for Non-Blocking Networks

Nicholas Pippenger*
e-mail: nicholas@cs.ubc.ca

Department of Computer Science
The University of British Columbia
Vancouver, British Columbia V6T 1Z4
CANADA

Abstract: Define f on the integers $n > 1$ by the recurrence

$$f(n) = \min\{n, \min_{m|n} 2f(m) + 3f(n/m)\}.$$

The function f has $f(n) = n$ as its upper envelope, attained for all prime n. Our goal in this paper is to determine the corresponding lower envelope. We shall show that this has the form $f(n) \sim C(\log n)^{1+1/\gamma}$ for certain constants γ and C, in the sense that for any $\varepsilon > 0$, the inequality $f(n) \leq (C+\varepsilon)(\log n)^{1+1/\gamma}$ holds for infinitely many n, while $f(n) \leq (C-\varepsilon)(\log n)^{1+1/\gamma}$ holds for only finitely many. In fact, $\gamma = 0.7878\ldots$ is the unique real solution of the equation $2^{-\gamma} + 3^{-\gamma} = 1$, and $C = 1.5595\ldots$ is given by the expression

$$C = \frac{\gamma\,(2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma})^{1/\gamma}}{(\gamma+1)\left(15^{-\gamma}\log^{\gamma+1}\frac{5}{2} + 3^{-\gamma}\sum_{5\leq k\leq 7}\log^{\gamma+1}\frac{k+1}{k} + \sum_{8\leq k\leq 15}\log^{\gamma+1}\frac{k+1}{k}\right)^{1/\gamma}}.$$

We also consider the function f_0 defined by replacing the integers $n > 1$ with the reals $x > 1$ in the above recurrence:

$$f_0(x) = \min\{x, \inf_{1<y<x} 2f_0(y) + 3f_0(x/y)\}.$$

We shall show that $f_0(x) \sim C_0(\log x)^{1+1/\gamma}$, where $C_0 = 1.5586\ldots$ is given by

$$C_0 = 6e\,(2^{-\gamma}\log 2^{-\gamma} + 3^{-\gamma}\log 3^{-\gamma})^{1/\gamma}\left(\frac{\gamma}{\gamma+1}\right)^{1+1/\gamma}$$

and is smaller than C by a factor of $0.9994\ldots$.

* This research was partially supported by an NSERC Operating Grant.

1. Introduction

Our goal in this paper is an analysis of the recurrence

$$f(n) = \min\{n, \min_{m|n} 2f(m) + 3f(n/m)\} \tag{1.1}$$

for the function $f : N \to N$, where N denote the set of integers exceeding 1. The value of $f(n)$ depends strongly on the factorization of n. Thus for example we have $f(n) = n$ whenever n is prime, since then the inner minimization is over an empty set of factorizations. This example characterizes the "upper envelope" of f, since the outer minimization ensures that $f(n) \le n$ always holds.

In the motivation for the study of this recurrence, which will be presented in Section 2, $f(n)$ is interpreted as a "cost" and n as a "benefit". We are thus led to seek the corresponding "lower envelope" of the function f, where the relationship between cost and benefit is most favorable. Our main result, Theorem 6.1, shows that this lower envelope takes the form

$$f(n) \sim C(\log n)^{1+1/\gamma} \tag{1.2}$$

(for certain constants γ and C), in the sense that for any $\varepsilon > 0$ the inequality

$$f(n) \le (C+\varepsilon)(\log n)^{1+1/\gamma} \tag{1.3}$$

is satisfied for infinitely many values of n, while

$$f(n) \le (C-\varepsilon)(\log n)^{1+1/\gamma} \tag{1.4}$$

is satisfied for only finitely many. The constant $\gamma = 0.78788\ldots$ is the unique real solution of the equation

$$2^{-\gamma} + 3^{-\gamma} = 1, \tag{1.5}$$

while the constant $C = 1.5595\ldots$ is given by

$$C = \frac{\gamma\,(2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma})^{1/\gamma}}{(\gamma+1)\left(15^{-\gamma}\log^{\gamma+1}\frac{5}{2} + 3^{-\gamma}\sum_{5\le k\le 7}\log^{\gamma+1}\frac{k+1}{k} + \sum_{8\le k\le 15}\log^{\gamma+1}\frac{k+1}{k}\right)^{1/\gamma}}. \tag{1.6}$$

It may seem surprising that a recurrence as simple as (1.1) can give rise to an expression as complicated as (1.6); nevertheless, we shall find a simple interpretation for each of the twelve terms that are summed in the denominator.

In preparation for the derivation of our main result, it will be convenient to analyze some related recurrences that provide upper and lower bounds for f, while being much easier to analyze. Firstly, for any integer $d > 1$, we may consider the function f_d that is defined by the same recurrence as f, but with the domain being restricted from the set N of all integers exceeding 1 to the set N_d of all integral powers of d exceeding 1:

$$f_d(n) = \min\{n, \min_{m|n} 2f_d(m) + 3f_d(n/m)\}. \tag{1.7}$$

The multiplicative semigroup formed by the integral powers of d constitutes a subsemigroup of the multiplicative semigroup of integers. Thus we have $f_d(n) \geq f(n)$ wherever the left-hand side is defined, since any factorization that participates in the minimization on the left-hand side also participates on the right-hand side. On the other hand, the factorizations that participate on the left-hand side are sufficiently uniform as to eliminate the discrepancy between the upper and lower envelopes, so that we shall obtain a simple asymptotic expression for f_d.

We shall show in Theorem 4.1 that for $d \geq 5$ we have

$$f_d(n) \sim C_d (\log n)^{1+1\gamma}, \tag{1.8}$$

where

$$C_d = \frac{4d\gamma}{\gamma+1} \left(\frac{2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma}}{\log^{\gamma+1} d} \right)^{1/\gamma}. \tag{1.9}$$

The expression (1.9) assumes its minimum for $d = 10$, with $C_{10} = 1.6296\ldots$.

For $2 \leq d \leq 4$ the situation is more complicated, since in these cases the first member of the outer minimization in (1.7) can minorize the second when n is a power of d, whereas this occurs only for $n = d$ when $d \geq 5$. Nevertheless, we shall show in Theorem 4.2 that (1.8) continues to hold, with $C_2 = 1.5909\ldots$ given by

$$C_2 = \frac{\gamma}{\gamma+1} \left(\frac{2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma}}{(4^{-\gamma} + 12^{-\gamma})\log^{\gamma+1} 2} \right)^{1/\gamma}, \tag{1.10}$$

$C_3 = C_9 = 1.6311\ldots$, and $C_4 = 1.6867\ldots$ given by

$$C_4 = \frac{\gamma}{\gamma+1} \left(\frac{2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma}}{(28^{-\gamma} + 36^{-\gamma})\log^{\gamma+1} 4} \right)^{1/\gamma}. \tag{1.11}$$

Thus the minimum of C_d over all d occurs for $d = 2$.

Finally, we may consider the function f_0 that is defined by the same recurrence as f, but with the domain being extended from the set N of all integers exceeding 1 to the set N_0 of all reals exceeding 1:

$$f_0(x) = \min\{x, \inf_{1<y<x} 2f_0(y) + 3f_0(x/y)\}. \tag{1.12}$$

(The infimum in (1.12) is in fact achieved as a minimum, as will become clear from the analysis, but we shall not need this fact.) Here we have a supersemigroup of the multiplicative semigroup of integers, so that we have $f_0(n) \leq f(n)$ for all integers $n > 1$. Again the discrepancy between upper and lower envelopes disappears, and we obtain a simple asymptotic formula for f_0.

We shall show in Theorem 5.1 that

$$f_0(x) \sim C_0 (\log x)^{1+1/\gamma}, \tag{1.13}$$

where $C_0 = 1.5586\ldots$ is given by

$$C_0 = 6e\,(2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma})^{1/\gamma}\left(\frac{\gamma}{\gamma+1}\right)^{1+1/\gamma}, \tag{1.14}$$

in which $e = 2.7182\ldots$ is the base of natural logarithms.

2. Non-Blocking Networks

The analysis of the recurrence (1.1) may be followed without reference to or knowledge of non-blocking networks. For the sake of motivation, however, we shall derive the recurrence against its historical background.

A "network" is an interconnection of "nodes" by means of "switches". In a network there are some distinguished nodes called "inputs", some other distinguished nodes called "outputs", and some distinguished sets of switches called "routes", each of which forms a path from an input to an output. A network is "non-blocking" if, given any disjoint set of routes (no two of which have a node or switch in common), and given any free input and free output (neither of which are involved in any of the given routes), there is a free route (disjoint from the given routes) from the given input to the given output. (The knowledgeable reader will recognize here the definition of a "strictly" non-blocking network. As this is the only type with which we shall have to deal in this paper, we shall omit the qualification "strictly".)

One of the basic questions concerning non-blocking networks is: given integers $n > 1$ and $m > 1$, what is the smallest possible number $G(n, m)$ of switches in a non-blocking network with n inputs and m outputs? Since inputs and outputs appear symmetrically in the definitions, we have

$$G(n, m) = G(m, n) \tag{2.1}$$

by taking "mirror images".

A non-blocking network can be constructed by letting the inputs and outputs be the only nodes, and by installing a separate switch between each input and each output. Such a network, which is called a "crossbar", shows that

$$G(n, m) \leq nm. \tag{2.2}$$

In 1953, Clos [C2] introduced what has become the most widely known method for the construction of non-blocking networks. His idea is to construct a large non-blocking network by interconnecting smaller subnetworks. In his construction the subnetworks are arranged in three "stages", as shown in Figure 1. The first stage, shown at the left, contains a subnetworks, each with b inputs and $2b$ outputs. The second stage contains $2b$ subnetworks, each having a inputs and a outputs. The inputs of the first-stage subnetworks are the inputs of the overall network; the outputs of the first-stage subnetworks are identified with (that is, connected by "wires" to) the inputs of the second-stage subnetworks, in such a way that each first-stage and each second-stage subnetwork have exactly one node in common. The third stage, shown on the right, contains a subnetworks, each having $2b$ inputs and b outputs. The outputs of the third-stage subnetworks are the outputs of the overall network; the outputs of the second-stage subnetworks are identified with the inputs of the third-stage subnetworks, in such a way that each second-stage and each third-stage subnetwork have exactly one node in common. Each route in the overall network consists of a route through a first-stage subnetwork, its extension through a second-stage subnetwork, and finally its extension through a third-stage subnetwork.

A simple argument based on the pigeon-hole principle shows that the overall network is non-blocking if each of the subnetworks is non-blocking. This construction thus shows that

$$\begin{aligned} G(ab, ab) &\leq aG(b, 2b) + 2bG(a, a) + aG(2b, b) \\ &\leq 2aG(b, 2b) + 2bG(a, a). \end{aligned} \tag{2.3}$$

(The attentive reader may have noticed that the argument remains valid even if $2b$ is replaced by $2b - 1$. We shall ignore this sharpening of the inequality, however, as it leads off the path we wish to follow.)

If crossbars are used in each of the three stages, and if the parameters a and b are each chosen to be about $n^{1/2}$, the resulting construction shows that $G(n,n) = O(n^{3/2})$. It is clear that further progress can be made by using the method recursively, but Clos did not succeed in finding the best way of doing this. In 1971, Cantor [C1] presented the two principles that underlie the best recursive use of Clos's method. Firstly, since the subnetworks in the outer stages have inputs and outputs in the proportion $1 : 2$ (or $2 : 1$, which is equivalent by taking mirror images), the recursion should be based entirely on such networks. This can be accomplished by giving the inner subnetworks inputs and outputs in the proportion $1 : 2$, whence the overall network will have inputs and outputs in the same proportion. When this has been done, the second-stage subnetworks have a inputs and $2a$ outputs, and there are $2a$ third-stage subnetworks. The resulting construction, shown in Figure 2, shows that

$$\begin{aligned} G(ab, 2ab) &\leq aG(b, 2b) + 2bG(a, 2a) + 2aG(2b, b) \\ &\leq 3aG(b, 2b) + 2bG(a, 2a). \end{aligned} \tag{2.4}$$

If the parameters a and b were given equal values, the outer subnetworks would be more numerous than the inner ones in the proportion $3 : 2$, and any diseconomy of scale would manifest itself more acutely in the outer stages. It follows that the sizes of the outer subnetworks should be reduced, and those of the inner subnetworks increased. To discover the optimal choices of a and b, let F denote the the largest function defined on the integers exceeding 1 and satisfying the inequalities

$$F(n) \leq 2n^2 \tag{2.5}$$

and

$$F(ab) \leq 3aF(b) + 2bF(a). \tag{2.6}$$

Comparing (2.2) and (2.4) with (2.5) and (2.6), we see that $F(n)$ is the smallest possible number of switches in a non-blocking network built according to the construction of Clos and Cantor. Furthermore, if we set $f(n) = F(n)/2n$, we see that f satisfies the recurrence (1.1). Thus the minimizations occurring in (1.1) correspond to the optimizations available in the construction of Clos and Cantor.

Cantor ([C1], Section 3) showed that for any $\varepsilon > 0$,

$$f(n) = O\big((\log n)^{1+1/\gamma+\varepsilon}\big) \tag{2.7}$$

for an infinite sequence of n, and Pippenger ([P1], Section 6) showed that

$$f(n) = O\big((\log n)^{1+1/\gamma}\big) \tag{2.8}$$

for an infinite sequence. Our Theorems 4.1, 4.2 and 6.1 can all be viewed as refinements of and complements to (2.7) and (2.8) for various sequences.

The construction for non-blocking networks that we study is not the best asymptotically. Indeed, Cantor ([C1], Section 4) gave a construction using $O\big(n(\log n)^2\big)$ switches, and Bassalygo and Pinsker [BP] gave a probabilistic argument showing the existence of non-blocking networks with $O(n \log n)$ switches. By an old result of Shannon [S], the rate of growth of this last result is the best possible. The result of Bassalygo and Pinkser has since been obtained through an explicit construction; see Pippenger [P1] for a presentation of all these results.

It is interesting to note that the results of Section 4 for fixed $d > 1$ correspond to the assumption that all crossbars in a non-blocking network have certain fixed sizes, d or a power of d; and it is curious that the choices $d = 10$ and $d = 2$ should have certain optimality properties, since precisely these values have been favored historically in the construction of telephone switching networks (following the widespread use of the decimal and binary number representations by humans and computers, respectively). The results of Section 5 similarly correspond to the assumption (contrary to fact) that crossbars could have any real (not necessarily integral) numbers of inputs and outputs; and it is curious how little could be gained in this way: C and C_0 differ by less than one part in one thousand!

3. Derivations

In this section we shall reinterpret our problem in terms of trees, which will become the main objects of our attention in later sections. To see the relevance of trees, consider the task of proving that

$$f(n) \le p \tag{3.1}$$

for some particular n and p. If $n \le p$, then (3.1) follows by the first member of the outer minimization of (1.1). Otherwise, we must have $f(n) = 2f(m) + 3f(n/m)$ for some $m \mid n$. In this case we can reduce the task of proving (3.1) to that of proving

$$f(m) \le q \tag{3.2}$$

and

$$f(l) \le r \tag{3.3}$$

for some m and l such that $ml = n$ and some q and r such that $p = 2q + 3r$. In either case, we may represent the proof of (3.1) in the form of a tree: in the first case the tree reduces to a single vertex, its root; in the second case, the root has two children, which are the roots of subtrees representing the proofs of (3.2) and (3.3). In the remainder of this section we present the combinatorial machinery that formalizes this representation.

For the purposes of this paper, the "infinite tree" is the set $V = \{2,3\}^*$ of finite words over the alphabet $\Sigma = \{2,3\}$. The words of V are called "vertices". The empty word λ is called the "root". If v is a word, the word $v2$ is called its "left child" and $v3$ is called its "right child", v is called the "parent" of $v2$ and $v3$, and $v2$ and $v3$ are called "siblings" of each other.

A "finite tree" (or simply a "tree") is a non-empty finite subset $T \subseteq V$ that is closed under taking parents and siblings. Every tree contains the root λ. If vertex v has a child in a tree T, then both its children are in T, and v is called an "internal vertex" of T. If v belongs to T but has no children in T, then v is called a "leaf" of T. The number of internal vertices in a tree is one less than the number of leaves.

The "weight" $W(v)$ of a vertex v is the product of the letters appearing in v, with each letter appearing as a factor with the same multiplicity that it has in v. (This definition accounts for our rather unorthodox use of 2 and 3 as the letters of a binary alphabet.)

A "derivation" $D = (T, l)$ is a tree $T = T_D$ together with an assignment $l = l_D$ of integers exceeding 1 to the leaves of T. If D is a derivation, the integer assigned to a leaf v will be called the "load" of v and will be denoted $l_D(v)$. The "capacity" $L(D)$ of a derivation D is the product of the loads of its leaves. The "cost" $C(D)$ of a derivation D is the sum, over all leaves, of the product of the weight of the leaf and the load of the leaf.

The main result of this paper is based on the following observation: the solution $f(n)$ of the recurrence (1.1) is equal to the minimum cost of a derivation with capacity n. This is easily proved by the inductive argument sketched in the opening paragraph of this section.

Furthermore, we can extend this reinterpretation to the recurrences (1.7) and (1.12) simply by restricting or extending the set of allowable loads. Specifically, if we define a "d-derivation" for $d > 1$ to be a derivation in which all the loads are integral powers of d, then $f_d(n)$ is the minimum possible cost of a d-derivation with capacity n. Similarly, if we define a "0-derivation" to be like a derivation, except that the loads may be any reals exceeding 1, then $f_0(x)$ is the minimum possible cost of a 0-derivation with capacity x.

4. Integral Powers of d

In this section we shall analyze the recurrence (1.7), starting with the case $d \geq 5$; later we shall also consider $2 \leq d \leq 4$. The case $d \geq 5$ could actually be solved by reduction to a recurrence dealt with by Fredman and Knuth [FK], but we shall use a slightly different analysis, in order to prepare for other cases treated later.

Theorem 4.1: For $d \geq 5$, the solution f_d to the recurrence (1.7) satisfies

$$f_d(n) \sim C_d (\log n)^{1+1\gamma}, \tag{4.1}$$

where

$$C_d = \frac{4d\gamma}{\gamma+1} \left(\frac{2^{-\gamma} \log 2^{\gamma} + 3^{-\gamma} \log 3^{\gamma}}{\log^{\gamma+1} d} \right)^{1/\gamma}. \tag{4.2}$$

As observed in Section 3, $f_d(n)$ is the minimum possible cost of a d-derivation of capacity n. When $d \geq 5$, our problem is simplified by the following observation: for every $n = d^k$, there exists a minimum-cost d-derivation of capacity n in which the load of every leaf is d. To see this suppose that every minimal-cost d-derivation with capacity n has a leaf with load at least d^l, where $l \geq 2$. Let D be a d-derivation with capacity n and the minimum possible number of loads equal to d^l, and let v be a leaf in T_D with load equal to $l_D(v) = d^l$. Consider the derivation D' obtained from D by making v an internal vertex with leaves as children. If we let $l_{D'}(v2) = d^{l-1}$ and $l_{D'}(v3) = d$, the capacity of D' is the same as that of D. Furthermore, the cost of D' is no greater than that of D, since the contribution $d^l W(v)$ to D has been replaced by the contribution $(2d^{l-1} + 3d)W(v)$ to D', and $2d^{l-1} + 3d \leq d^l$ when $d \geq 5$. This contradicts the assumption that D has the minimum possible number of loads equal to d^l, and completes the proof of the observation.

We shall refer to the number of leaves in a tree as the "scale" of the tree, and the sum of the weights of its leaves as its "total weight". When the load of every leaf is d, the capacity of a d-derivation is just d^k, where k is the scale of its tree, and the cost of a d-derivation is just d times the total weight of its tree. Thus a minimum-cost d-derivation is one based on a tree that, among those with a given scale, has the smallest possible total weight. This yields

$$f_d(d^k) = d\Psi(k), \tag{4.3}$$

where $\Psi(k)$ denotes the minimum possible total weight of leaves in a tree with k leaves.

If T is a tree, we shall call its set of internal vertices its "kernel", and denote it by $K(T)$. The kernel of a tree is closed under taking prefixes. Conversely, any set K closed under taking prefixes is the kernel of a tree $T(K)$, obtained from K by adjoining as leaves those vertices that are children of vertices in K but do not themselves appear in K. Thus there is a one-to-one correspondence between trees and their kernels.

For any tree T, the set $T \setminus K(T)$ is the set of leaves of T. It will be called the "frontier" of T and denoted $F(T)$.

If v is any vertex, we have $W(v2) + W(v3) = 5W(v)$. Summing this identity over all $v \in K(T)$, we obtain

$$\sum_{u \in F(T)} W(u) = 1 + 4 \sum_{u \in K(T)} W(u),$$

since a leaf $u \in F(T)$ appears once as $v2$ or $v3$, the root λ appears once as v, and each other internal vertex $u \in K(T)$ appears once as $v2$ or $v3$ and once as v. Thus, among trees of a given scale, those with the minimum total weight of their leaves are also those with the minimum total weight of their internal vertices. This yields

$$\Psi(k) = 1 + 4\Phi(k-1), \tag{4.4}$$

where $\Phi(k-1)$ denotes the minimum possible total weight of internal vertices in a tree with $k-1$ internal vertices.

We shall say that a tree is a "threshold tree" if the weight of every internal vertex is less than or equal to the weight of every leaf. If from the set of vertices we choose $k-1$ with the smallest weights, the resulting set of vertices is closed under taking prefixes, since the parent of a vertex v has a strictly smaller weight than v. Such a set thus constitutes the set of internal vertices of a threshold tree with k leaves. Thus there exist threshold trees of every scale. Furthermore, among trees of a given scale, threshold trees have minimum total weight of their internal vertices (since this is how their internal vertices were chosen), and thus have minimum total weight of their leaves. This yields

$$f_d(d^k) = d\big(1 + 4\Phi(k-1)\big), \tag{4.5}$$

where $\Phi(k-1)$ can now be interpreted as the sum of the weights of the $k-1$ smallest-weight vertices in the infinite tree. Our problem is now to determine the asymptotic behavior of Φ.

Let $h(x)$ denote the number of vertices of the infinite tree having weight at most x. This function satisfies the asymptotic formula

$$h(x) \sim \frac{x^\gamma}{H}, \tag{4.6}$$

where

$$H = 2^{-\gamma} \log 2^\gamma + 3^{-\gamma} \log 3^\gamma. \tag{4.7}$$

(Information theorists will recognize H as the entropy per independent flip of a biased coin that falls heads with probability $2^{-\gamma}$ and tails with probability $3^{-\gamma}$.) Formula (4.6) was proved by Fredman and Knuth [FK], who used an analytic argument; an elementary proof (in the technical sense) can be found in Pippenger [P2]. (This formula is the only point at which the present paper is not self-contained.)

Let W_j denote the weight of the j-th vertex of the infinite tree (when the vertices are arranged in non-decreasing order by weight). Inverting (4.6) by raising each side to the power $1/\gamma$, we see that

$$W_j \sim H^{1/\gamma} j^{1/\gamma}. \tag{4.8}$$

Summing over j we obtain

$$\begin{aligned} \Phi(k) &= \sum_{1 \le j \le k} W_j \\ &\sim \frac{\gamma}{\gamma+1} H^{1/\gamma} k^{1+1/\gamma}. \end{aligned} \tag{4.9}$$

Combining (4.9) with (4.5), and using $k = \log_d n$, yields Theorem 4.1.

For $2 \le d \le 4$, the analysis given above breaks down: there may be no minimum-cost d-derivations in which all loads equal d. This is best illustrated by the case $d = 2$, which we treat now.

Theorem 4.2: We have

$$f_2(n) \sim C_2 (\log n)^{1+1/\gamma}, \tag{4.10}$$

where

$$C_2 = \frac{\gamma}{\gamma+1} \left(\frac{2^{-\gamma} \log 2^\gamma + 3^{-\gamma} \log 3^\gamma}{(4^{-\gamma} + 12^{-\gamma}) \log^{\gamma+1} 2} \right)^{1/\gamma}. \tag{4.11}$$

We begin with a simple observation we shall call the "ordering principle". Suppose that we fix a tree T and a suite (that is, multiset) S of loads, and ask which loads should be assigned to which leaves in order to minimize the resulting cost. If $A_1 \leq \cdots \leq A_k$ are the weights in non-decreasing order and $B_1 \geq \cdots \geq B_k$ are the loads in non-increasing order, then the minimum possible cost is $\sum_{1\leq j\leq k} A_j B_j$ (this is simply Chebyshev's inequality). In particular, the smallest load should be assigned to the leaf with the largest weight, and *vice versa.*

Our next task is to determine what loads can appear on leaves of a minimum-cost 2-derivation; we claim that, excluding the trivial case $n = 2$, these are 4, 8 and 16.

Suppose there is an optimal 2-derivation D in which some leaf has load 2. By the odering principle, we may assume this leaf has the largest weight of any leaf in the tree; thus it is of the form $v3$ for some word v (this is the point at which we must exclude the case $n = 2$), and $v2$ is also a leaf (for if it were the root of a subtree, all the leaves of this subtree would have weight larger than that of $v3$). Let 2^l be the load of the leaf $v2$ in D. Let D' be the 2-derivation obtained from D by making v a leaf with load 2^{l+1}. Then D' has the same capacity as D. Furthermore, D' has smaller cost than D, since the contribution $2^{l+1}W(v)$ of v to $C(D')$ is less than the contribution $(2\cdot 2^l + 3\cdot 2)W(v)$ of $v2$ and $v3$ to D. This contradicts the assumption that an optimal 2-derivation can contain a leaf with load 2.

Suppose now that there is an optimal 2-derivation D in which some leaf v has load 2^l, where $l \geq 5$. Let D' be the 2-derivation obtained from D by making v an internal vertex, with leaves as children. If we let $l_{D'}(v2) = 2^{l-2}$ and $l_{D'}(v3) = 4$, then the capacity of D' is the same a that of D. Furthermore, the cost of D' is less than that of D, since the contribution $(2\cdot 2^{l-2} + 3\cdot 4)W(v)$ of $v2$ and $v3$ in D' is less than the contribution $2^l W(v)$ of v to D. This contradicts the assumption that an optimal 2-derivation can contain a leaf with load 2^l, where $l \geq 5$, and completes the proof of the claim that optimal 2-derivations contain only 4, 8 and 16 as loads.

Next we claim that in an optimal 2-derivation, no leaf of the form $v2$ (that is, no "left leaf") can have load 4. Suppose that D is an optimal 2-derivation in which $l_D(v2) = 4$. If the subtree rooted at $v3$ has capacity greater than 4, then we may obtain a 2-derivation with the same capacity as, but lower cost than, D by exchanging the subtrees rooted at $v2$ and $v3$. On the other hand, if $v3$ is a leaf with load 4, we may obtain a 2-derivation with the same capacity as, but lower cost than, D by making v a leaf with load 16. In either case we obtain a contradiction, proving that no left leaf can have load 4.

In what follows we shall confine our attention to 2-derivations in which all loads are 4, 8 and 16, and in which no left leaf has load 4; we shall call these "admissible" 2-derivations. Define D_0 to be the admissible 2-derivation in which the root λ is the only leaf, with load 8. Consider now three operations, which we shall call "promotions", that transform admissible 2-derivations into other admissible 2-derivations. The promotion $(v, 4)$ will be applicable to any admissible 2-derivation D in which v is a leaf with $l_D(v) = 4$; the result of applying $(v, 4)$ to D is the admissible 2-derivation obtained from D by increasing the load of v to 8. The promotion $(v, 8)$ will be applicable to any admissible 2-derivation D in which v is a leaf with $l_D(v) = 8$; the result of applying $(v, 8)$ to D is the admissible 2-derivation obtained from D by increasing the load of v to 16. The promotion $(v, 16)$ will be applicable to any admissible 2-derivation D in which v is a leaf with $l_D(v) = 16$; the result of applying $(v, 16)$ to D is the admissible 2-derivation obtained from D by making v an internal vertex with leaves as children, assigning 8 as the load of $v2$ and 4 as the laod of $v3$.

Any admissible 2-derivation D with capacity at least 8 can be obtained by starting with D_0 and applying a sequence of promotions; this is easily proved by induction on the capacity of D. (The basis is capacity 8. The inductive step breaks into three cases: if D has a leaf v with load 16, then D can be obtained by applying promotion $(v, 8)$ to an admissible 2-derivation with one-half the capacity (which can by the inductive hypothesis be obtained from D_0 by promotions); if D has a right leaf v with load 8, then D can be obtained by applying promotion $(v, 4)$ to an admissible 2-derivation with one-half the capacity; and otherwise D, if it is not the basis, must contain an internal vertex v with leaves as children, with 8 as load of $v2$ and 4 as load of $v3$, so that D can be obtained by applying promotion $(v, 16)$ to an admissible 2-derivation with one-half the capacity.)

Any promotion doubles the capacity of the admissible 2-derivation to which it is applied. We shall assign a "cost" $C(P)$ to each promotion P as follows: the cost of the promotion $(v, 4)$ is $4W(v)$, the cost of $(v, 8)$ is $8W(v)$, and the cost of $(v, 16)$ is $12W(v)$. Then if application of promotion P to admissible 2-derivation D yields D', then we have $C(D') = C(D) + C(P)$.

Among promotions, some are prerequisite to others: the promotion $(v, 8)$ is prerequisite to $(v, 16)$, the promotion $(v, 16)$ is prerequisite to both $(v2, 8)$ and $(v3, 4)$, and the promotion $(v3, 4)$ is prerequisite to $(v3, 8)$. In every case, however, if P is prerequisite to Q, then the cost of P is at most the cost of Q. In particular, we can order all possible promotions in a sequence $P_1, P_2, \ldots, P_j, \ldots$ in such a way that (1) the costs are non-decreasing and (2) each promotion is preceded by all of its prerequisites. It follows that the result of applying $P_1, \ldots, P_{k-3}$ in order to D_0 is a minimum-cost 2-derivation with capacity 2^k. This yields

$$f_2(2^k) = 8 + \sum_{1 \le j \le k-3} C(P_j). \tag{4.12}$$

Our problem is now to determine the asymptotic behavior of $C(P_j)$.

Let $g(x)$ denote the number of promotions with cost at most x. We can write

$$g(x) = g_4(x) + g_8(x) + g_{16}(x), \tag{4.13}$$

where $g_4(x)$, $g_8(x)$ and $g_{16}(x)$ denote the numbers of promotions of the form $(v, 4)$, $(v, 8)$ and $(v, 16)$, respectively, with cost at most x. Since the cost of $(v, 16)$ is $12W(v)$, we have

$$g_{16}(x) = h(x/12). \tag{4.14}$$

Since the cost of $(v, 8)$ is $8W(v)$, we have

$$g_8(x) = h(x/8). \tag{4.15}$$

There is a promotion $(v, 4)$ if and only if v is of the form $u3$; since the cost of $(v, 4)$ is $4W(v) = 12W(u)$, we have

$$g_4(x) = h(x/12). \tag{4.16}$$

This yields

$$\begin{aligned} g(x) &= 2h(x/12) + h(x/8) \\ &\sim (2 \cdot 12^{-\gamma} + 8^{-\gamma}) \frac{x^\gamma}{H} \\ &\sim (12^{-\gamma} + 4^{-\gamma}) \frac{x^\gamma}{H}, \end{aligned} \tag{4.17}$$

where we have used the identity (1.5) to obtain the last line from its predecessor.

Inverting (4.17) by raising each side to the power $1/\gamma$, we see that

$$C(P_j) \sim \left(\frac{H}{12^{-\gamma} + 4^{-\gamma}} \right)^{1/\gamma} j^{1/\gamma}. \tag{4.18}$$

Substituting this formula in (4.12) and summing yields (4.10) and (4.11), completing the proof of Theorem 4.2.

It is worth observing that the promotion $(v, 16)$ that creates a load of 4 has the same cost as the promotion $(v3, 4)$ that destroys the load of 4. It follows that the sequence of optimal promotions can be arranged so that the resulting optimal 2-derivations each have at most one leaf with load 4. Thus we can arrange that "almost all" the loads in an optimal 2-derivation are either 8 or 16.

The cases $d = 3$ and $d = 4$ are similar to $d = 2$, and we shall only describe the key points in the analyses. For $d = 3$, we easily show that no optimal 3-derivation can have a load as large as 81, and need not have any load as large as 27 (since this can be replaced without increasing the cost by children with loads 9 and 3). Futhermore, no left leaf can have a load of 3. Thus we need only consider "admissible" 3-derivations in which all loads are either 3 or 9, and no left leaf has a load of 3. We can analyze these by introducing promotions as before. We then observe that the promotion that creates a load of 3 has the same cost as the promotion that destroys the load of 3. Thus we can arrange that optimal 3-derivations have at most one leaf with load 3. Since almost all the loads are then 9, we obtain the same asymptotic result as in the case $d = 9$: $C_3 = C_9$.

For $d = 4$, a similar analysis shows that all loads in an optimal 4-derivation must be either 4 or 16. Furthermore, no left leaf can have a load of 4. We can then continue the analysis using promotions, and the result is

$$C_4 = \frac{\gamma}{\gamma+1} \left(\frac{2^{-\gamma} \log 2^\gamma + 3^{-\gamma} \log 3^\gamma}{(28^{-\gamma} + 36^{-\gamma}) \log^{\gamma+1} 4} \right)^{1/\gamma}. \tag{4.19}$$

It is worthwhile observing that for $2 \le d \le 4$ the trees underlying optimal d-derivations are threshold trees, just as they were for $d \ge 5$; this is easily seen by considering the promotions that increase the number of leaves in the tree.

5. Reals

In this section we shall analyze the recurrence (1.12) obtained by eliminating the integrality constraint from (1.1).

Theorem 5.1: We have

$$f_0(x) \sim C_0 (\log x)^{1+1/\gamma}, \tag{5.1}$$

where $C_0 = 1.5586\ldots$ is given by

$$C_0 = 6e \left(2^{-\gamma} \log 2^\gamma + 3^{-\gamma} \log 3^\gamma\right)^{1/\gamma} \left(\frac{\gamma}{\gamma+1}\right)^{1+1/\gamma}. \tag{5.2}$$

As indicated in Section 3, our quest is for optimal 0-derivations. To determine these, let us fix a tree T, and ask how the loads of its leaves should be assigned so as to minimize the cost, while achieving a prescribed capacity. (Later we shall determine how the tree T should be chosen.)

We first claim that, among 0-derivations based on a prescribed tree T and having a prescribed capacity x, a 0-derivation D with minimum cost must be such that there exists a constant c such that for all leaves $v \in F(T_D)$,

$$l_D(v)W(v) = c. \tag{5.3}$$

Thus the loads must vary as the reciprocal of the weights of their leaves. To see this, suppose to the contrary that $l_D(u)W(u) > l_D(v)W(v)$ for some leaves u and v. Set $c' = \sqrt{l_D(u)W(u)l_D(v)W(v)}$, and let D' be the 0-derivation obtained from D by changing the loads of u and v to $l_{D'}(u) = c'/W(u)$ and $l_{D'}(v) = c'/W(v)$. Then D' has the same capacity as D, but lower cost (as follows from the inequality between geometric and arithmetic means). This contradiction proves the claim.

We shall next ask which tree T, among those with k leaves, should be used to construct an optimal 0-derivation. (Later we shall determine how k should be chosen.)

If the capacity of D is to be x, we must have

$$\prod_{v \in F(T_D)} l_D(v) = x. \tag{5.4}$$

Multiplying (5.3) over all $v \in F(T_D)$ yields

$$\begin{aligned} c^k &= \prod_{v \in F(T_D)} l_D(v)W(v) \\ &= x \prod_{v \in F(T_D)} W(v), \end{aligned}$$

and thus

$$c = x^{1/k} \left(\prod_{v \in F(T_D)} W(v) \right)^{1/k}. \tag{5.5}$$

Since each leaf contributes c to the cost of D, we have

$$C(D) = kx^{1/k} \left(\prod_{v \in F(T_D)} W(v) \right)^{1/k}. \tag{5.6}$$

Thus the optimal tree T is one that minimizes the geometric mean of the weights of the leaves and therefore, given that the number of leaves is fixed, minimizes the product of the weights. And since the logarithm is an increasing function, it is equivalent to minimuize the sum of the logarithms of the weights of the leaves.

If v is any vertex, we have $\log W(v2) + \log W(v3) = \log 6 + 2 \log W(v)$. Summing this identity over all $v \in K(T)$, we obtain

$$\sum_{u \in F(T)} \log W(u) = \log 1 + (k-1) \log 6 + \sum_{u \in K(T)} \log W(u), \tag{5.7}$$

since each leaf $u \in F(T)$ appears once as $v2$ or $v3$, the root λ appears once as v, and each other internal vertex $u \in K(T)$ appears once as $v2$ or $v3$ and once as v. When k is fixed, the right-hand side of (5.7) is minimized by choosing the $k-1$ vertices v with the smallest $W(v)$ to be the internal vertices in $K(T)$. Thus threshold trees, which emerged as optimal for d-derivations ($d > 1$), are also optimal for 0-derivations.

It remains to determine the optimal value of k as a function of the capacity x. To do this we shall determine the asymptotic behavior of the geometric mean $U(k)$ of the weights of the leaves in a threshold tree with k leaves. We shall show that

$$U(k) \sim 6 e^{-1/\gamma} H^{1/\gamma} k^{1/\gamma}. \tag{5.8}$$

Using (5.7) we have

$$\begin{aligned} U(k) &= \exp \frac{1}{k} \sum_{v \in F(T)} \log W(v) \\ &= \exp \frac{1}{k} \left((k-1) \log 6 + \sum_{v \in K(T)} \log W(v) \right) \\ &= \exp \frac{1}{k} \left((k-1) \log 6 + \sum_{1 \le j \le k-1} \log W_j \right) \end{aligned} \tag{5.9}$$

From (4.8) we obtain

$$\log W_j = \frac{1}{\gamma} \log j + \frac{1}{\gamma} \log H + o(1). \tag{5.10}$$

Substituting (5.10) into (5.9) and estimating the sum by an integral yields (5.8).

From (5.8) we can complete the proof of Theorem 5.1 as follows. From (5.6) we have

$$\begin{aligned} C(D) &= k x^{1/k} U(k) \\ &\sim 6 e^{-1/\gamma} H^{1/\gamma} k^{1+1/\gamma} x^{1/k}. \end{aligned} \tag{5.11}$$

Choosing k to minimize $k^{1+1/\gamma}x^{1/k}$ yields

$$k \sim \frac{\gamma+1}{\gamma}\log x \tag{5.12}$$

and

$$k^{1+1/\gamma}x^{1/k} \sim \left(\frac{e\gamma}{\gamma+1}\right)^{1+1/\gamma}(\log x)^{1+1/\gamma}. \tag{5.13}$$

Substituting (5.13) into (5.11) yields

$$C(D) \sim 6eH^{1/\gamma}\left(\frac{\gamma}{\gamma+1}\right)^{1+1/\gamma}(\log x)^{1+1/\gamma}, \tag{5.14}$$

which completes the proof of Theorem 5.1.

6. Integers

We arrive in this section at our main result, the solution of the recurrence (1.1).

Theorem 6.1: For every $\varepsilon > 0$, we have

$$f(n) \le (C+\varepsilon)(\log n)^{1+1/\gamma} \tag{6.1}$$

for infinitely many values of n, but

$$f(n) \le (C-\varepsilon)(\log n)^{1+1/\gamma} \tag{6.2}$$

for only finitely many, where

$$C = \frac{\gamma\,(2^{-\gamma}\log 2^{\gamma} + 3^{-\gamma}\log 3^{\gamma})^{1/\gamma}}{(\gamma+1)\left(15^{-\gamma}\log^{\gamma+1}\frac{5}{2} + 3^{-\gamma}\sum_{5\le k\le 7}\log^{\gamma+1}\frac{k+1}{k} + \sum_{8\le k\le 15}\log^{\gamma+1}\frac{k+1}{k}\right)^{1/\gamma}}. \tag{6.3}$$

Because $f(n)$ is large when n is prime, we must focus attention on the lower envelope. We do this by defining

$$C' = \liminf_{n\to\infty} f(n)/(\log n)^{1+1/\gamma}, \tag{6.4}$$

so that our task is to prove that $C' = C$.

Let us say that an integer n is "good" if there is no larger integer $m > n$ such that $f(m) \le f(n)$. If n is not good, then for some larger m we have

$$f(m)/(\log m)^{1+1/\gamma} < f(n)/(\log n)^{1+1/\gamma}.$$

Thus the *limes inferior* in (6.4) remains unchanged if we confine attention to good n.

We begin as in Section 4 with an analysis of the possible load values; we shall not obtain the sharpest bounds here, but merely aim to reduce the range that must be considered later. Let us consider a minimum cost derivation D for a good integer n, and let us further suppose that, among derivations of this minimum cost, D has the largest possible number of leaves.

Firstly, we claim that D can have no load as small as 2. For if any leaf had load 2, this would certainly have to be the case for the leaf $v3$ of largest weight (by the ordering principle), and the sibling of $v3$ is another leaf $v2$ (else it would subtend a leaf of greater weight than $v3$). Suppose the load of $v2$ is l. Then the derivation obtained from D by making v a leaf with load $2l$ would have the same capacity as, but lower cost than, D.

Secondly, D cannot have a leaf v with load of 24: replacing v by leaves $v2$ with load 6 and $v3$ with load 4 would leave the capacity and cost unchanged, but increase the number of leaves.

Thirdly, D cannot have a leaf v with load of 25: replacing v by leaves $v2$ with load 5 and $v3$ with load 5 would leave the capacity and cost unchanged, but increase the number of leaves.

Finally, D cannot have a leaf v with load l as large as 26. To see this, it will suffice to show that we can find integers i and j such that $ij > l$ and $2i + 3j = l$, for then we could replace v by leaves $v2$ with load i and $v3$ with load j, and increase the capacity while leaving the cost unchanged; this contradicts the assumption that the capacity of D is good.

If we plot the line $2i + 3j = l$ and the hyperbola $ij = l$ in the real (i, j) plane, they intersect at two points with i-coordinates

$$i = \frac{l \pm \sqrt{l^2 - 24l}}{4}. \tag{6.5}$$

The difference between these i-coordinates is

$$\Delta i = \frac{\sqrt{l^2 - 24l}}{2}. \tag{6.6}$$

We will have $\Delta i > 3$ if $l^2 - 24l > 36$; this in turn holds when $l > 12 + 2\sqrt{45}$, and thus certainly when $l \geq 12 + 2\sqrt{49} = 26$.

The line $2i + 3j = l$ contains infinitely many lattice points (points with integral coordinates); the i-coordinates of successive such lattice points differ by 3, since adding

3 to i and subtracting 2 from j leaves the sum $2i+3j$ unchanged. Thus there must be a lattice point whose i-coordinate lies in the interval whose endpoints are given by (6.5). For this point (i,j) we have $2i+3j=l$ and $ij>l$, as desired.

Thus every load on an optimal derivation with a good capacity and a maximal number of leaves is at least 3 and at most 23. We also claim that, in such a derivation D, if v has leaves $v2$ with load i and $v3$ with load j for children, then $2i+3j \geq 24$. For the maximum of ij subject to the constraint $2i+3j=l$ is $l^2/24$. Thus if $2i+3j<24$, the derivation obtained from D by making v a leaf with load ij will have the same capacity as, but smaller cost than, D.

We next claim that the tree T underlying the derivation D is a threshold tree. Suppose to the contrary that T contains an internal vertex u and a leaf v with $W(u)>W(v)$. We may assume that u has leaves $u2$ and $u3$ as children (since if not we may transfer attention from u to one of its children). Let h, i and j be the loads of the leaves v, $u2$ and $u3$, respectively. As we have seen above, we must have $h \leq 23$ and $2i+3j \geq 24$. Thus we obtain

$$-h+2i+3j>0. \tag{6.7}$$

Consider now the tree T' obtained from T by making u a leaf and making v an internal vertex with children $v2$ and $v3$ as leaves. Let us now create a derivation D' from the tree T' by assigning the loads h, i and j in some order to the leaves u, $v2$ and $v3$, and letting the loads of all other leaves be the same as in D. Then D' has the same capacity as D. We shall show that there is some order of assignment that results in D' having a smaller cost than D. The analysis breaks into three cases, depending on how $W(u)$ ranks among $2W(v)<3W(v)$.

Firstly, suppose that $3W(v)<W(u)$. Then by the ordering principle we should assign h, i and j to $v2$, $v3$ and u, respectively. These three loads contribute $hW(v)+i2W(u)+j3W(u)$ to $C(D)$, and $h2W(v)+i3W(v)+jW(u)$ to $C(D')$. If this does not decrease the cost, that is if $C(D')-C(D) \geq 0$, then

$$hW(v)+i\big(3W(v)-2W(u)\big)-j2W(u) \geq 0. \tag{6.8}$$

Multiplying (6.7) by $W(v)$ and adding the result to (6.8) yields

$$i\big(5W(v)-2W(u)\big)+j\big(3W(v)-2W(u)\big)>0. \tag{6.9}$$

This is a contradiction, since $3W(v)<W(u)$ implies that the coefficients of i and j are each strictly negative.

Secondly, suppose that $2W(v) < W(u) \le 3W(v)$. Then we assign h, i and j to $v2$, u and $v3$, respectively. These three loads contribute $hW(v) + i2W(u) + j3W(u)$ to $C(D)$, and $h2W(v) + iW(u) + j3W(v)$ to $C(D')$. If this does not decrease the cost, then

$$hW(v) - iW(u) + j\big(3W(v) - 3W(u)\big) \ge 0. \tag{6.10}$$

Multiplying (6.7) by $W(v)$ and adding the result to (6.10) yields

$$i\big(2W(v) - W(u)\big) + j\big(6W(v) - 3W(u)\big) > 0. \tag{6.11}$$

This is a contradiction, since $2W(v) < W(u)$ implies that the coefficients of i and j are each strictly negative.

Finally, suppose that $W(v) < W(u) \le 2W(v)$. Then we assign h, i and j to u, $v2$ and $v3$, respectively. These three loads contribute $hW(v) + i2W(u) + j3W(u)$ to $C(D)$, and $hW(u) + i2W(v) + j3W(v)$ to $C(D')$. If this does not decrease the cost, then

$$-h\big(W(v) - W(u)\big) + 2i\big(W(v) - W(u)\big) + 3j\big(W(v) - W(u)\big) \ge 0. \tag{6.12}$$

But this contradicts (6.7), since $W(v) < W(u)$. Thus the assumption that an internal vertex v has greater weight than a leaf u leads to a contradiction, completing the proof that threshold trees are optimal.

Now that we know that optimal derivations are based on threshold trees, and that their loads are at least 3 and and most 23, it remains to determine the number of leaves that should be assigned each of these loads (since then the ordering principle will tell us which loads to assign to which leaves). Let T be a threshold tree and let y denote the largest weight of an internal vertex. We shall renormalize the weights of the leaves by setting $\eta_v = W(v)/y$ for each leaf $v \in F(T)$. Then we have $\eta_v \ge 1$, since T is a threshold tree. Furthermore, we have $\eta_v \le 3$, since the weight of a leaf is at most thrice the weight of its parent, which is an internal vertex. Finally, we have $\eta_v \le 2$ unless v is a "right" leaf (that is, a leaf of the form $u3$), since the weight of a left leaf is at most twice the weight of its parent, which is an internal vertex.

We shall show that, if we choose a leaf v at random from a threshold tree with k leaves, with all k leaves being equally likely, the value of η_v has a distribution that tends as $k \to \infty$ to a particular density function on the interval $1 \le \eta \le 3$, which is continuous except for a single jump at $\eta = 2$.

Firstly, let us fix η and ε such that $2 < \eta < \eta + \varepsilon < 3$, and consider the number $E(y, \eta, \varepsilon)$ of leaves v such that $\eta < \eta_v \le \eta + \varepsilon$. Such leaves are right leaves (since $\eta_v > 2$), and are in one-to-one correspondence with internal vertices u such that $\eta y/3 < W(u) \leq (\eta+\varepsilon)y/3$. Using (4.6), we obtain

$$\begin{aligned} E(y,\eta,\varepsilon) &= h\big((\eta+\varepsilon)y/3\big) - h\big(\eta y/3\big) \\ &\sim \frac{\big((\eta+\varepsilon)y/3\big)^{\gamma}}{H} - \frac{\big(\eta y/3\big)^{\gamma}}{H} \\ &\sim \frac{y^{\gamma}}{H}\big(3^{-\gamma}\gamma\eta^{\gamma-1}\varepsilon + O(\varepsilon^2)\big). \end{aligned} \tag{6.13}$$

Again using (4.6), we have $k \sim y^{\gamma}/H$. Thus the distribution of leaves in the interval $2 < \eta < 3$ aymptotically follows the density function

$$\phi(\eta) = 3^{-\gamma}\gamma\eta^{\gamma-1}. \tag{6.14}$$

We have normalized ϕ so that

$$\int_2^3 \phi(\eta)\, d\eta = 1 - (2/3)^{\gamma}; \tag{6.15}$$

the reason for this will become clear shortly.

Next, let us fix η and ε such that $1 < \eta < \eta+\varepsilon < 2$, and consider the number $E(y,\eta,\varepsilon)$ of leaves v such that $\eta < \eta_v \leq \eta+\varepsilon$. Such leaves may be either right leaves or left leaves. The right leaves are in one-to-one correspondence with internal vertices u such that $\eta y/3 < W(u) \leq (\eta+\varepsilon)y/3$, and the left leaves are in one-to-one correspondence with internal vertices u such that $\eta y/2 < W(u) \leq (\eta+\varepsilon)y/2$. Using (4.6) and (1.5), we obtain

$$\begin{aligned} E(y,\eta,\varepsilon) &= h\big((\eta+\varepsilon)y/3\big) - h\big(\eta y/3\big) \\ &\qquad + h\big((\eta+\varepsilon)y/2\big) - h\big(\eta y/2\big) \\ &\sim \frac{\big((\eta+\varepsilon)y/3\big)^{\gamma}}{H} - \frac{\big(\eta y/3\big)^{\gamma}}{H}. \\ &\qquad + \frac{\big((\eta+\varepsilon)y/2\big)^{\gamma}}{H} - \frac{\big(\eta y/2\big)^{\gamma}}{H} \\ &\sim \frac{y^{\gamma}}{H}\big(\gamma\eta^{\gamma-1}\varepsilon + O(\varepsilon^2)\big) \end{aligned} \tag{6.16}$$

Thus the distribution of leaves in the interval $1 < \eta < 2$ aymptotically follows the density function

$$\phi(\eta) = \gamma\eta^{\gamma-1}. \tag{6.17}$$

We have normalized ϕ so that

$$\int_1^2 \phi(\eta)\, d\eta = 2^\gamma - 1; \tag{6.18}$$

from (6.15), (6.18) and (1.5) we have $\int_1^3 \phi(\eta)\, d\eta = 1$, so that ϕ can be viewed as a probability density function on the interval $1 < \eta < 3$.

For each good integer n, let $D(n)$ denote an optimal derivation with capacity n. Let $k(n)$ denote the number of leaves in $T_{D(n)}$. For $3 \le m \le 23$, let $k_m(n)$ denote the number of leaves v of $T_{D(n)}$ such that $l_{D(n)}(v) = m$, and let $\mu_m(n) = k_m(n)/k(n)$ denote the fraction of such leaves. From the sequence of good integers, let us extract an infinite subsequence of "special" integers such that, as n runs through the special integers, (1) $f(n)/(\log n)^{1+1/\gamma}$ tends to C' (as defined in (6.4)), and (2) for each m in the range $3 \le m \le 23$, $\mu_m(n)$ tends to a limit μ_m. Condition (1) can be fulfilled by the definition of C', and condition (2) because for each of the finitely many values of m, $\mu_m(n)$ varies in the compact interval $0 \le \mu_m(n) \le 1$. Henceforth we confine our attention to these special n. Of course, we have

$$\sum_{3 \le m \le 23} \mu_m = 1. \tag{6.19}$$

For each m, define α_m and β_m such that

$$\int_1^{\alpha_m} \phi(\eta)\, d\eta = \sum_{m<l} \mu_l \tag{6.20}$$

and

$$\int_{\beta_m}^3 \phi(\eta)\, d\eta = \sum_{l<m} \mu_l. \tag{6.21}$$

Let M denote the set of m such that $\mu_m > 0$. Then we have $1 \le \alpha_m \le \beta_m \le 3$, and $\alpha_m < \beta_m$ if and only if $m \in M$. The half-open intervals $(\alpha_m, \beta_m]$ for $m \in M$ form a partition of the interval $(1, 3]$. Thus for $1 < \eta \le 3$ we may define a non-increasing left-continuous step function ψ on the interval $(1, 3]$ by letting $\psi(\eta)$ be the unique value of m such that $\alpha_m < \eta \le \beta_m$.

For special n we have

$$\log n = \sum_{v \in F(T_D(n))} \log l_{D(n)}(v) \tag{6.22}$$

and

$$f(n) = \sum_{v \in F(T_D(n))} l_{D(n)}(v) W(v). \tag{6.23}$$

The limiting distribution ϕ of the weights of leaves in threshold trees, the ordering principle, and the definition of ψ allow us to express the asymptotic behavior of the sums in (6.22) and (6.23) using integrals,

$$\log n \sim k \int_1^3 \phi(\eta) \log \psi(\eta)\, d\eta \tag{6.24}$$

and

$$f(n) \sim ky \int_1^3 \phi(\eta) \eta \psi(\eta)\, d\eta. \tag{6.25}$$

Since $k \sim y^\gamma / H$ and $f(n)/(\log n)^{1+1/\gamma} \sim C'$, we conclude that

$$C' = \frac{H^{1/\gamma} P}{Q^{1+1/\gamma}}. \tag{6.26}$$

where

$$P = \int_1^3 \phi(\eta) \eta \psi(\eta)\, d\eta \tag{6.27}$$

and

$$Q = \int_1^3 \phi(\eta) \log \psi(\eta)\, d\eta. \tag{6.28}$$

If now we let ψ be any any non-increasing left-continuous step function defined on $(1,3]$ and taking values in $\{3, \ldots, 23\}$, then for each m in the range M of ψ, there $\psi(\eta)$ takes on the value m for η in an interval of the form $(\alpha_m, \beta_m]$. From any threshold tree T_k with k leaves and threshold y, we can obtain a derivation D_k by assigning to each leaf $v \in F(T_k)$ the load $\psi\big(W(v)/y\big)$. Letting k (and with it y) tend to infinity, we obtain a sequence of derivations with capacities

$$\log n_k \sim k \int_1^3 \phi(\eta) \log \psi(\eta)\, d\eta \tag{6.29}$$

and costs

$$C(D_k) \sim ky \int_1^3 \phi(\eta) \eta \psi(\eta)\, d\eta. \tag{6.30}$$

Since we must have $f(n_k) \le C(D_k)$, we conclude that

$$C' = \min_{\psi} \Gamma(\psi), \tag{6.31}$$

where

$$\Gamma(\psi) = \frac{H^{1/\gamma} P(\psi)}{Q(\psi)^{1+1/\gamma}}, \tag{6.31'}$$

$$P(\psi) = \int_1^3 \phi(\eta)\eta\psi(\eta)\, d\eta \tag{6.32}$$

and

$$Q(\psi) = \int_1^3 \phi(\eta) \log \psi(\eta)\, d\eta, \tag{6.33}$$

and where the minimum is taken over all non-increasing left-continuous step functions ψ defined on $(1, 3]$ and taking values in $\{3, \ldots, 23\}$. Thus we have reduced the determination of C' to the solution of the variational problem (6.31).

Firstly, we claim that the range M of the function ψ minimizing (6.31) must be an interval of consecutive integers. Suppose to the contrary that for some $h > i > j$ we have $1 < \beta_h = \alpha_i = \beta_i = \alpha_j < 3$. Let us denote this common value by δ, and suppose for now that $\delta \neq 2$. Let us define a new function ψ' by choosing $\varepsilon > 0$, changing $\beta_h = \alpha_i = \delta$ to $\beta'_h = \alpha'_i = \delta - \varepsilon(i-j)$ and changing $\beta_i = \alpha_j = \delta$ to $\beta'_i = \alpha'_j = \delta + \varepsilon(h-i)$. The effect of this change on P is

$$\begin{aligned} P(\psi') - P(\psi) &= (i-h) \int_{\delta-\varepsilon(i-j)}^{\delta} \phi(\eta)\eta\, d\eta + (i-j) \int_{\delta}^{\delta+\varepsilon(h-i)} \phi(\eta)\eta\, d\eta \\ &= (i-h)\big(\varepsilon(i-j)\phi(\delta)\delta + O(\varepsilon^2)\big) + (i-j)\big(\varepsilon(h-i)\phi(\delta)\delta + O(\varepsilon^2)\big) \\ &= O(\varepsilon^2) \end{aligned} \tag{6.34}$$

for $\varepsilon > 0$ sufficiently small. The effect on Q is

$$\begin{aligned} Q(\psi') - Q(\psi) &= \left(\log \frac{i}{h}\right) \int_{\delta-\varepsilon(i-j)}^{\delta} \phi(\eta)\, d\eta + \left(\log \frac{i}{j}\right) \int_{\delta}^{\delta+\varepsilon(h-i)} \phi(\eta)\, d\eta \\ &= \left(\log \frac{i}{h}\right) \big(\varepsilon(i-j)\phi(\delta) + O(\varepsilon^2)\big) + \left(\log \frac{i}{j}\right) \big(\varepsilon(h-i)\phi(\delta) + O(\varepsilon^2)\big) \\ &= \big[(h-j)\log i - (i-j)\log h - (h-i)\log j\big]\phi(\delta)\varepsilon + O(\varepsilon^2) \end{aligned} \tag{6.35}$$

For sufficiently small $\varepsilon > 0$. The quantity in square brackets in (6.35) is strictly positive, by the concavity of the logarithm. Thus the change from ψ to ψ' increases Q to first-order in ε, but increases P only to second-order in ε. It follows that by choosing $\varepsilon > 0$ sufficiently small, we obtain a contradiction to the assumption that ψ minimizes (6.31). In the exceptional case that $\delta = 2$, the same argument works if we introduce a factor of 3^γ to compensate for the discontinuity of ϕ: if we set $\beta'_h = \alpha'_i = \delta - \varepsilon(i-j)$ as before, but now set $\beta'_i = \alpha'_j = \delta + \varepsilon(h-i)3^\gamma$, we again obtain cancellation to first-order in P but not in Q. Thus we conclude that M is an interval of consecutive integers.

Next we claim that if $m+1$ and m both belong to M, and if ψ minimizes $\Gamma(\psi)$ (and therefore also $\log\Gamma(\psi)$), then we must have

$$\beta_{m+1} = \alpha_m = \varrho \log \frac{m+1}{m}, \tag{6.36}$$

where

$$\varrho = \frac{\gamma+1}{\gamma}\frac{P(\psi)}{Q(\psi)}. \tag{6.37}$$

Suppose that $\beta_{m+1} = \alpha_m$ and denote this common value by δ. Suppose for now that $\delta \neq 2$. Let us define a new function ψ' by choosing a small number ϑ (of either sign) and setting $\beta'_{m+1} = \alpha'_m = \delta + \vartheta$. The effect of this change on P is

$$\begin{aligned} P(\psi') - P(\psi) &= \int_\delta^{\delta+\vartheta} \phi(\eta)\eta\, d\eta \\ &= \vartheta\phi(\delta)\delta + O(\vartheta^2), \end{aligned} \tag{6.38}$$

and the effect on Q is

$$\begin{aligned} Q(\psi') - Q(\psi) &= \left(\log\frac{m+1}{m}\right)\int_\delta^{\delta+\vartheta} \phi(\eta)\, d\eta \\ &= \vartheta\left(\log\frac{m+1}{m}\right)\phi(\delta) + O(\vartheta^2). \end{aligned} \tag{6.39}$$

The effect on $\log\Gamma$ is

$$\log\Gamma(\psi') - \log\Gamma(\psi) = \frac{\vartheta\phi(\delta)\delta}{P(\psi)} - \frac{\gamma+1}{\gamma}\frac{\vartheta\left(\log\frac{m+1}{m}\right)\phi(\delta)}{Q(\psi)} + O(\vartheta^2). \tag{6.40}$$

Thus if (6.36) did not hold, we could choose a small value of ϑ (with appropriate sign) and make the right-hand side of (6.40) strictly negative. This contradicts the assumption that ψ minimizes $\Gamma(\psi)$, and proves (6.36). In the exceptional case that $\delta = 2$, the same argument works if we interpret $\phi(\delta) = \phi(2)$ (which has not yet been defined) correctly. Specifically, if we wish to choose $\vartheta > 0$, we should set $\phi(2) = 3^{-\gamma}\gamma 2^{\gamma-1}$ (to make ϕ right-continuous at 2), and if we wish to choose $\vartheta < 0$, we should set $\phi(2) = \gamma 2^{\gamma-1}$ (to make ϕ left-continuous at 2).

Finally, we claim that if $\varrho\log\frac{m+1}{m}$ falls in the open interval $(1,3)$, then $m+1$ and m both belong to M. Suppose to the contrary that the largest element of M is $q \le m$. Let us define a new function ψ' by choosing $\varepsilon > 0$, changing $\alpha_q = 1$ to $\alpha_q = 1+\varepsilon$, and setting $\alpha'_{m+1} = 1$ and $\beta'_{m+1} = 1+\varepsilon$. The change to P is $(m+1-q)\varepsilon\phi(1) + O(\varepsilon^2)$, the change to Q is $\bigl(\log \frac{m+1}{q}\varepsilon\phi(1) + O(\varepsilon^2)$, and the change to $\log\Gamma$ is

$$\log\Gamma(\psi') - \log\Gamma(\psi) = \frac{\varepsilon\phi(1)}{P(\psi)} - \frac{\gamma+1}{\gamma}\frac{\varepsilon\left(\log\frac{m+1}{q}\right)\phi(1)}{Q(\psi)} + O(\varepsilon^2) \tag{6.40$'$}$$

which is strictly negative for $\varepsilon > 0$ sufficiently small because $\varrho\log\frac{m+1}{q} \geq \varrho\log\frac{m+1}{m} > 1$. This contradiction shows that $m+1$ belongs to M when $\varrho\log\frac{m+1}{m} > 1$; a similar argument (setting $\alpha'_m = 3-\varepsilon$ and $\beta'_m = 3$) shows that m belongs to M when $\varrho\log\frac{m+1}{m} < 3$.

At this point we have reduced the determination of the minimizing function ψ to the determination of the single parameter ϱ. Although ϱ is defined by (6.37), we do not yet know the values of $P(\psi)$ and $Q(\psi)$, and thus we seek a more explicit characterization of ϱ.

Assume for now that the set of values $\{\varrho\log\frac{m+1}{m} : 3 \leq m, m+1 \leq 23\}$ is disjoint from the set $\{1,2,3\}$. Then from (6.36) we can write

$$\psi(\eta) = \left\lceil \frac{1}{\exp\frac{\eta}{\varrho} - 1} \right\rceil. \tag{6.41}$$

Thus if we define $r = \psi(1)$, $s = \psi(2)$ and $t = \psi(3)$, we have

$$r = \left\lceil \frac{1}{\exp\frac{1}{\varrho} - 1} \right\rceil \tag{6.42}$$

$$s = \left\lceil \frac{1}{\exp\frac{2}{\varrho} - 1} \right\rceil \tag{6.43}$$

$$t = \left\lceil \frac{1}{\exp\frac{3}{\varrho} - 1} \right\rceil. \tag{6.44}$$

We can now evaluate the integrals in (6.32) and (6.33) by breaking the range of integration into intervals, over each of which ψ is constant; the results are

$$\begin{aligned} P(\psi) &= r\int_1^{\varrho\log\frac{r}{r-1}} \gamma\eta^\gamma\,d\eta + \sum_{r>q>s} q\int_{\varrho\log\frac{q+1}{q}}^{\varrho\log\frac{q}{q+1}} \gamma\eta^\gamma\,d\eta + s\int_{\varrho\log\frac{s+1}{s}}^{2} \gamma\eta^\gamma\,d\eta \\ &\quad + s\int_2^{\varrho\log\frac{s}{s-1}} 3^{-\gamma}\gamma\eta^\gamma\,d\eta + \sum_{s>q>t} q\int_{\varrho\log\frac{q+1}{q}}^{\varrho\log\frac{q}{q+1}} 3^{-\gamma}\gamma\eta^\gamma\,d\eta + t\int_{\varrho\log\frac{t+1}{t}}^{3} 3^{-\gamma}\gamma\eta^\gamma\,d\eta \\ &= P_0 + \varrho^{\gamma+1}P_{\gamma+1}, \end{aligned} \tag{6.45}$$

where

$$P_0 = \frac{\gamma+1}{\gamma}(3t+2s-r) \tag{6.46}$$

and

$$P_{\gamma+1} = \frac{\gamma+1}{\gamma}\left(\sum_{r>q\geq s}\log^{\gamma+1}\frac{q+1}{q} + 3^{-\gamma}\sum_{s>q\geq t}\log^{\gamma+1}\frac{q+1}{q}\right), \tag{6.47}$$

and

$$\begin{aligned} Q(\psi) &= (\log r)\int_1^{\varrho\log\frac{r}{r-1}}\gamma\eta^{\gamma-1}\,d\eta + \sum_{r>q>s}(\log q)\int_{\varrho\log\frac{q+1}{q}}^{\varrho\log\frac{q}{q+1}}\gamma\eta^{\gamma-1}\,d\eta \\ &\quad + (\log s)\int_{\varrho\log\frac{s+1}{s}}^{2}\gamma\eta^{\gamma-1}\,d\eta + (\log s)\int_2^{\varrho\log\frac{s}{s-1}}3^{-\gamma}\gamma\eta^{\gamma-1}\,d\eta \\ &\quad + \sum_{s>q>t}(\log q)\int_{\varrho\log\frac{q+1}{q}}^{\varrho\log\frac{q}{q+1}}3^{-\gamma}\gamma\eta^{\gamma-1}\,d\eta + (\log t)\int_{\varrho\log\frac{t+1}{t}}^{3}3^{-\gamma}\gamma\eta^{\gamma-1}\,d\eta \\ &= Q_0 + \varrho^\gamma Q_\gamma, \end{aligned} \tag{6.48}$$

where

$$Q_0 = \log\frac{st}{r} \tag{6.49}$$

and

$$Q_\gamma = \sum_{r>q\geq s}\log^{\gamma+1}\frac{q+1}{q} + 3^{-\gamma}\sum_{s>q\geq t}\log^{\gamma+1}\frac{q+1}{q}. \tag{6.50}$$

Note that $P_{\gamma+1} = \frac{\gamma+1}{\gamma}Q_\gamma$. Combining (6.37) with (6.45)—(6.50), we conclude that

$$\varrho = \frac{3t+2s-r}{\log\frac{st}{r}}. \tag{6.51}$$

We now observe that (6.42)—(6.44) and (6.51) have a unique solution, namely

$$r = 16, \qquad s = 8, \qquad t = 5 \tag{6.52}$$

and

$$\varrho = \frac{15}{\log\frac{5}{2}}. \tag{6.53}$$

To verify this, it is convenient to define $\varrho_{p,q} = p/\log\frac{q+1}{q}$, which is the value of ϱ for which ψ makes the step from $q+1$ to q at p. Since $r \leq 23$, we must have $\varrho \leq \varrho_{1,23} = 23.4964\ldots$; and if $\varrho = \varrho_{1,23}$, then (6.43) and (6.44) yield $s = 12$ and $t = 8$. Since $t \geq 3$, we must have $\varrho \geq \varrho_{3,2} = 7.3989\ldots$; and if $\varrho = \varrho_{3,2}$, then (6.42) and (6.43) yield $r = 7$ and $s = 4$. Thus we have $r \in \{7, \ldots, 23\}$, $s \in \{4, \ldots, 12\}$, $t \in \{3, \ldots, 8\}$. The transitions among these possibilities occur when ϱ takes on one of the values $\varrho_{1,7}, \ldots, \varrho_{1,22}, \varrho_{2,4}, \ldots, \varrho_{2,11}, \varrho_{3,3}, \ldots, \varrho_{3,7}$. These 29 points, when sorted into increasing order, divide the interval $[\varrho_{3,2}, \varrho_{1,23}]$ into 30 subintervals, and the values of r, s and t given by (6.42)—(6.44) are constant throughout each of these subintervals. Thus there is a unique value of ϱ given by (6.51) for each of these subintervals. In only one case does this value of ϱ fall into the subinterval: throughout the subinterval from $\varrho_{1,15} = 15.4949\ldots$ to $\varrho_{3,5} = 16.4544\ldots$, (6.42)—(6.44) give $r = 16$, $s = 8$ and $t = 5$, whence (6.51) gives $\varrho = 15/\log\frac{5}{2} = 16.3703\ldots$. (There are 30 cases to be considered here; the calculations could be done by hand with patience, but were in fact done by a computer.) Thus we have established (6.52) and (6.53), on the assumption that the set of values $\{\varrho \log\frac{m+1}{m} : 3 \le m, m+1 \le 23\}$ is disjoint from the set $\{1, 2, 3\}$, or equivalently that ϱ is not one of the values $\varrho_{1,7}, \ldots, \varrho_{1,23}, \varrho_{2,4}, \ldots, \varrho_{2,11}, \varrho_{3,2}, \ldots, \varrho_{3,7}$. To lift this assumption, we need only verify that (6.37) does not hold if ϱ takes on one of these values. (There are 31 cases to be considered here, and again the calculations were done by a computer.) Substituting (6.52) and (6.53) into (6.45)—(6.50) and (6.26) and simplifying yields $C' = C$ for C given by (6.3).

The outcome of the final search for r, s, t and ϱ may seem fortuitous or obscure, but it has a simple explanation. The complications of this section are due to the fact that loads, and thus the function ψ, can take on only integral values. If we drop this constraint, recovering the problem of Section 5, we may describe the solution by saying that $\psi(\eta) = 6e/\eta$ is then the minimizing choice of ψ. This corresponds to dropping the ceiling brackets, replacing $\exp(\eta/\varrho)$ by the first two terms $1 + (\eta/\varrho)$ of its power series expansion, and taking $\varrho = 6e = 16.3096\ldots$ in (6.41); and the values of r, s and t corresponding to this value of ϱ according to (6.42)—(6.44) are $r = 16$, $s = 8$ and $t = 5$.

We can also interpret the individual terms in the denominator of (1.6) in terms of "promotions". Specifically, we can associate the term $\log^{\gamma+1}\frac{q+1}{q}$ (for $8 \le q \le 15$) or $3^{-\gamma}\log^{\gamma+1}\frac{q+1}{q}$ (for $5 \le q \le 7$) with the promotion of a load from q to $q+1$; and we can associate the term $15^{-\gamma}\log\frac{5}{2}$ with the promotion of leaf with load 16 to a parent of leaves with loads 8 and 5. The expression for C is thus analogous to those for C_2 and C_4, though the justification is much more elaborate.

Finally, we observe that for any fixed $\varepsilon > 0$, (1.3) is fulfilled for infinitely many n in a geometric progression. To see this, observe that we may replace the optimal values of μ_m for $5 \le m \le 16$ by rational numbers p_m/q without increasing the value of $\Gamma(\psi)$ to more than $C + \varepsilon$. If we then consider a sequence T_i of trees having qi leaves, and form from these a sequence D_i of derivations in which there are $p_m i$ leaves with load m, the resulting derivations will have capacity $\left(\prod_{5\le m\le 16} m^{p_m}\right)^i$ and cost satisfying (1.3). It should be noted that this observation does not contradict the results of Section 4, which dealt with the behavior of (1.1) restricted to geometric progressions: in the observation the capacity n is confined to a geometric progression, but the divisor m of n in (1.1) ranges over all proper divisors; in Section 4, both n and m were confined to the geometric progression.

7. References

[BP] L. A. Bassalygo and M. S. Pinsker, "Complexity of an Optimum Nonblocking Network without Reconnections", *Prob. Info. Transm.*, 9 (1974) 64–66.

[C1] D. G. Cantor, "On Non-Blocking Switching Networks", *Networks*, 1 (1971) 367–377.

[C2] C. Clos, "A Study of Non-Blocking Switching Networks", *Bell Sys. Tech. J.*, 32 (1953) 406–424.

[FK] M. L. Fredman and D. E. Knuth, "Recurrence Relations Based on Minimization", *J. Math. Anal. Appl.*, 48 (1974) 534–559.

[P1] N. Pippenger, "Telephone Switching Networks", *AMS Proc. Symp. Appl. Math.*, 26 (1982) 101–133.

[P2] N. Pippenger, "An Elementary Approach to Some Analytic Asymptotics", *SIAM J. Math. Anal.*, 24 (1993) 1361–1377.

[S] C. E. Shannon, "Memory Requirements in a Telephone Exchange", *Bell Sys. Tech. J.*, 29 (1950) 343–349.

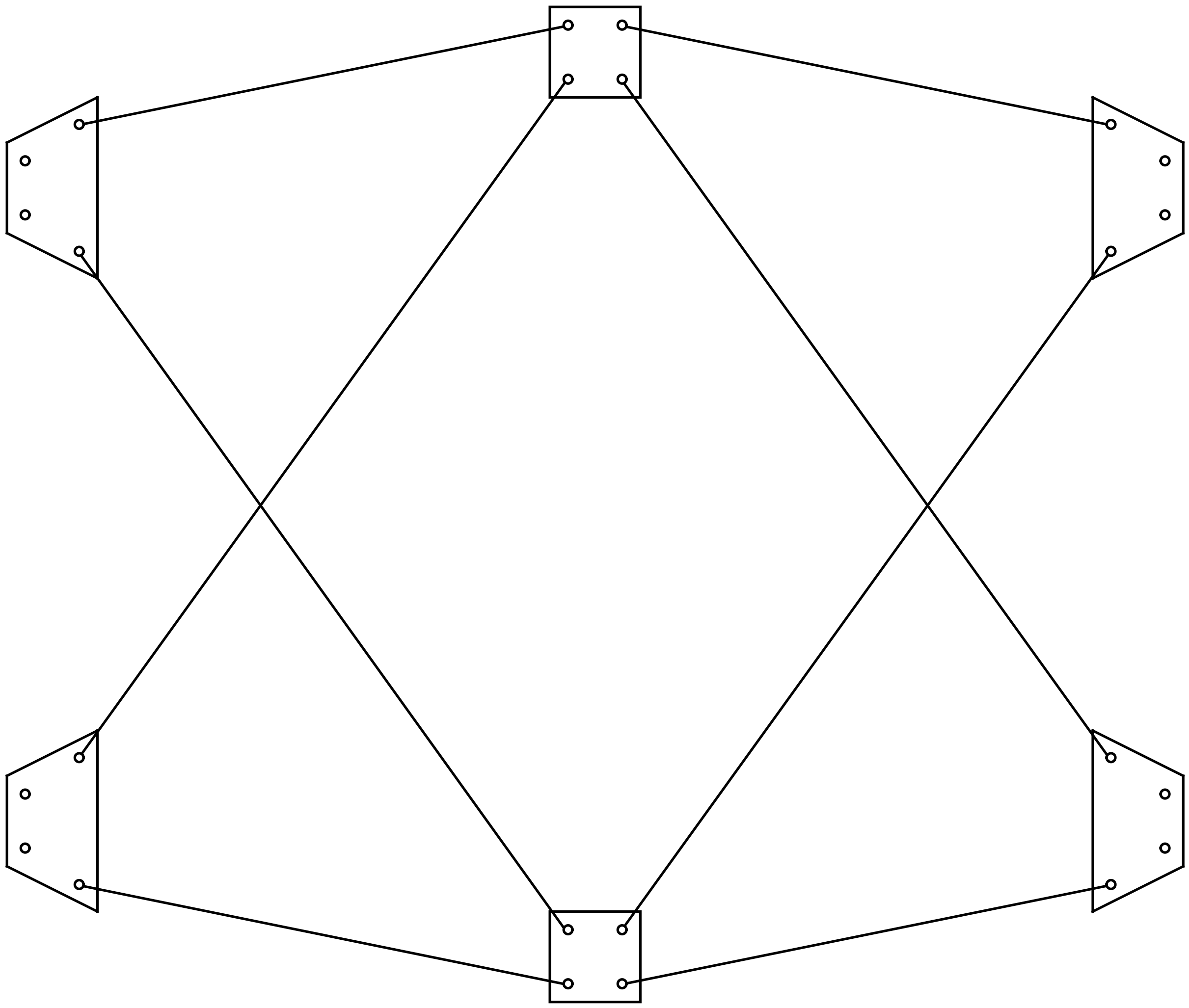

Figure 1

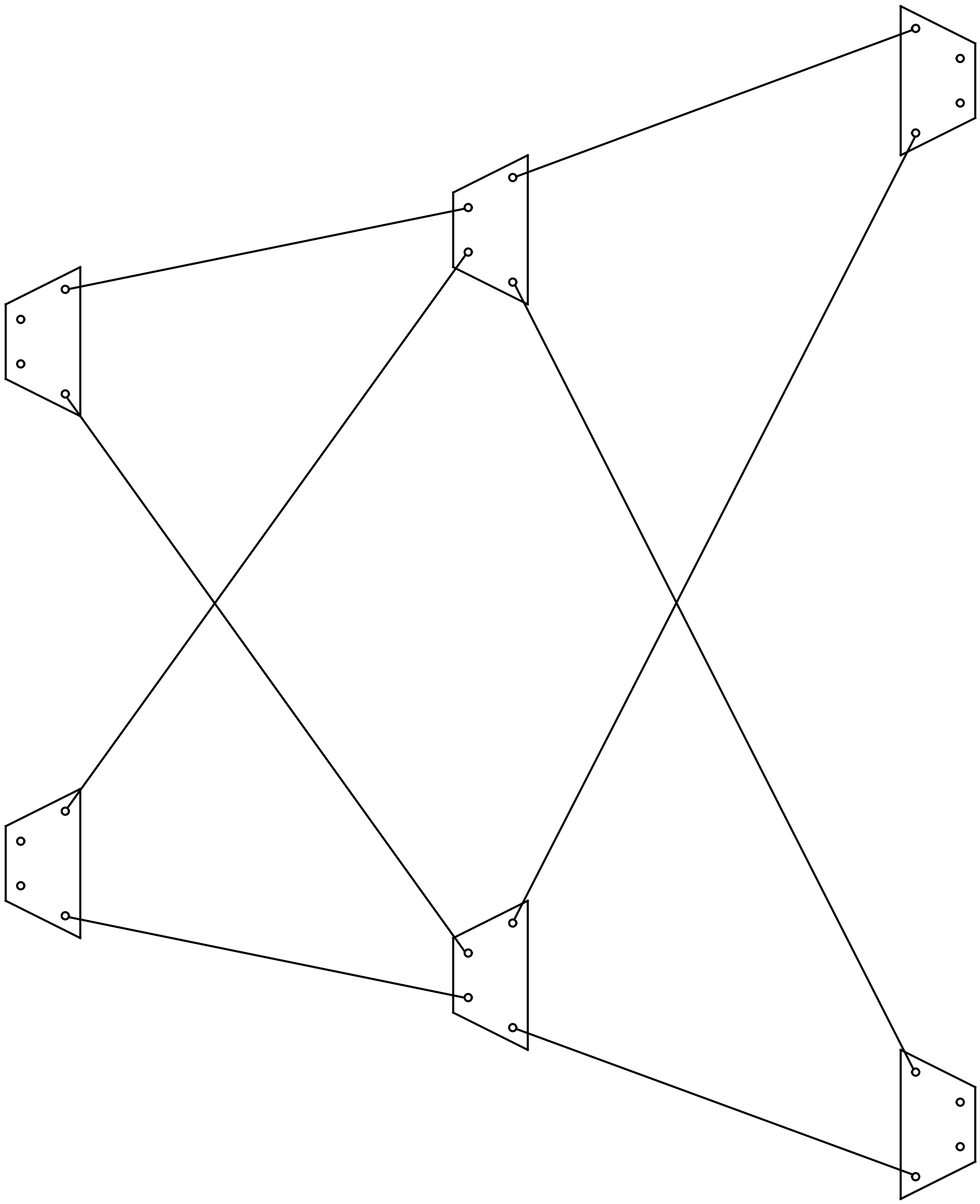

Figure 2