A New Bijection between Natural Numbers and Rooted Trees*

Peter R. Cappello
Department of Computer Science
University of California
Santa Barbara, CA 93106

Abstract

A bijection is defined recursively between the set of natural numbers and the set of finite rooted trees. It is based on the prime decomposition of a natural number and the rank of a prime. This bijection leads naturally to a canonical presentation of rooted trees. For trees in each of two classes, bounds are given for the number of nodes in terms of the number assigned to the tree.

1 Introduction

Investigations have been conducted into both the number of rooted trees (see, e.g., Cayley [2], and Pólya [6]), and their generation (see, e.g., Read [8], and Beyer and Hedetniemi [1]). Cayley [3, 4] proves that the number of labeled trees on n vertices is n^{n-2}. Prüfer [7] displays a bijection between these two sets. The generation algorithm presented by Beyer and Hedetniemi [1] can be used to construct a bijection between rooted trees and natural numbers. In this paper, another bijection is constructed between these two sets. It is intriguingly easy to describe, yet apparently hard to compute.

2 A bijection between natural numbers and rooted trees

Let $\mathbf{N}, \mathbf{P}, \mathbf{T}$ denote the sets of the natural numbers, the primes, and the finite, undirected rooted trees, respectively. Let function $p : \mathbf{N} \mapsto \mathbf{P}$ denote the nth prime (e.g., $p(4) = 7$).

Define the function $\tau : \mathbf{N} \mapsto \mathbf{T}$ recursively as follows:

1. $\tau(1)$ is the rooted tree comprised of exactly one node.
2. For $n > 1$, if the prime factorization of n is $f_1 f_2 \cdots f_j$, then $\tau(n)$ is the rooted tree in which the root is adjacent to the roots of the trees $\tau(p^{-1}(f_i))$.

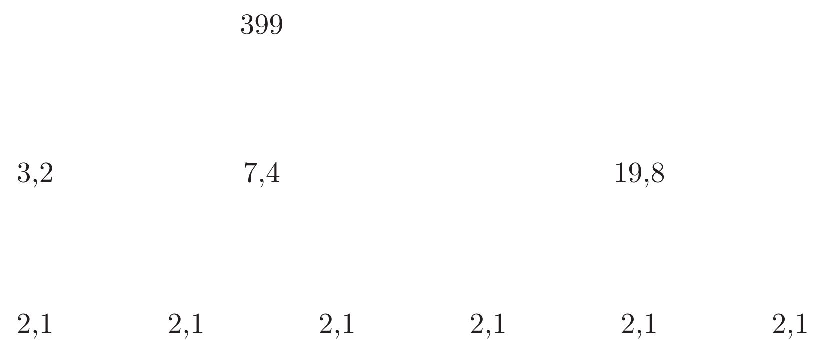


Figure 1: The root of each tree is bold, and in this figure contains the number associated with the tree; other nodes contain $f, p^{-1}(f)$, where f is a prime number.

Figure 1 depicts the tree $\tau(399)$. Figure 2 depicts the 15 trees associated with the natural numbers $[1, 15]$.

The map τ is a bijection. Its inverse is defined recursively as follows:

1. $\tau^{-1}(\bullet) = 1$.

2. If the root of tree t is adjacent to subtrees $t_1, t_2, \cdots, t_j$, then

$$\tau^{-1}(t) = \prod_{i=1}^{j} p(\tau^{-1}(t_i)).$$

A natural order (so to speak) of rooted finite trees suggests itself: If s and t are rooted finite trees, then is $s \preceq t$ when $\tau^{-1}(s) \leq \tau^{-1}(t)$.

A prime factorization of a number is in *canonical order* when the primes are presented in nondecreasing order. An analogue for rooted trees is offered below. Let $\tau(n) = t$, where the canonical order of the prime factorization of n is $f_1, f_2, \ldots, f_j$. The tree t is presented *canonically* when:

1. The rooted trees $t_1, t_2, \ldots, t_j$, corresponding to the factors $f_1, f_2, \cdots, f_j$, respectively, are presented from left to right.

2. Each rooted tree t_i is presented canonically.

The rooted trees in Figure 2 are presented canonically.

*Sections 2 and 3 were presented at the 4th SIAM Conf. on Discrete Mathematics, San Francisco, June 1988.

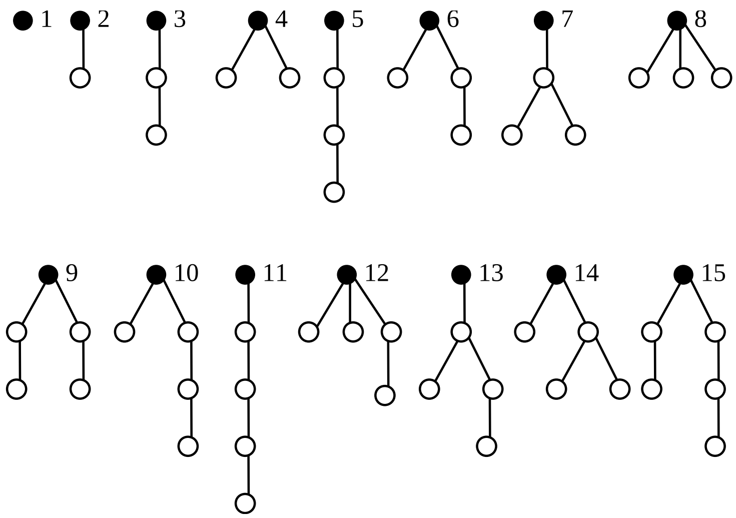


Figure 2: The first 15 rooted trees.

3 On the size of the trees

Two extreme classes of trees are considered:

Rooted stars: Let $B = \{2^k | k \in \mathbf{N}\}$. Then $\forall\, n \in B, \text{height}(\tau(n)) = 1$, and $|\tau(n)| = \log n + 1$.

Paths: Let the set S be defined recursively as follows:

- $1 \in S$,
- $n \in S$, if $n \in \mathbf{P}$ and $p^{-1}(n) \in S$.

The first 12 values of the set S are $1; 2; 3; 5; 11; 31; 127; 709; 5,381; 52,711; 648,391; 9,737,333$. The paths are the trees associated with the set S. Figure 2 includes the first 4 such rooted trees. For these trees, that is, for $n \in S$,

$$\text{height}(\tau(n)) = |\tau(n)| - 1 = \Omega(\log n / \log \log n).$$

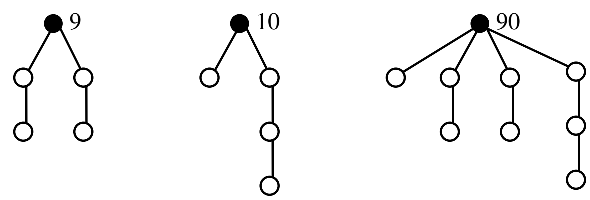
This can be seen as follows. The PRIME NUMBER THEOREM [5] states that the number of primes not exceeding n is asymptotic to $n/\log n$. Thus, for $n \in \mathbf{P}$, $p^{-1}(n) \sim n/\log n$, and if $p^{-1}(n) \in \mathbf{P}$ then

$$p^{-1}(p^{-1}(n)) \sim \frac{n/\log n}{\log(n/\log n)} = \Omega(n/\log^2 n).$$

Repeating this process, we see that the leaf of $\tau(n)$ corresponds to both the natural number 1, and $\Omega(n/\log^k n)$, where the $\text{height}(\tau(n)) = k$. The bound on $|\tau(n)|$ follows from this.

4 Some related algegraic structures

Let $\alpha = (\mathbf{N}, \cdot, 1)$ denote the commutative monoid of the natural numbers under product. Let the *product* of rooted trees t_1 and t_2, denoted $t_1 \cdot t_2$, be the tree that results from identifying their roots (see Figure 3). Let T be the set of all finite rooted trees, and $\beta = (T, \cdot, \bullet)$ be the commutative



(a) (b) (c)

Figure 3: (a) $\tau(9)$; (b) $\tau(10)$; (c) $\tau(9 \cdot 10) = \tau(9) \cdot \tau(10)$.

monoid of rooted trees under product. Since τ is a bijection, $\alpha \cong \beta$.

Finite trees associated with positive rational numbers

We now move to positive rational numbers. Let $\gamma = (\mathbf{N}^2, \cdot, (1,1))$ be the commutative monoid, where $(p,q) \cdot (r,s) = (pr, qs)$. Let $\tau_i(n)$ denote the rooted tree that results from directing all paths in $\tau(n)$ *towards* its root. We refer to $\tau_i(n)$ as an $in - tree$, and $\tau(n)$ as its underlying, rooted tree. This is a one-to-one correspondence. Let $\tau_o(n)$ denote the rooted tree that results from directing all paths in $\tau(n)$ *away from* its root. We refer to $\tau_o(n)$ as an $out - tree$, and $\tau(n)$ as its underlying, rooted tree, also a one-to-one correspondence. Let the *product* of in-tree $\tau_i(p)$ and out-tree $\tau_o(q)$, denoted $\tau(p,p)$ and referred to as an inout-tree, be the directed tree that results from identifying their roots (see Figure 4). The product of two inout-trees, $\tau(p,q)$ and $\tau(r,s)$, denoted $\tau(p,q)\cdot\tau(r,s)$, is the inout-tree that results from identifying the roots of $\tau(p,q)$ and $\tau(r,s)$ (see Figure 5):

$$\tau(p,q) \cdot \tau(r,s) = \tau(pr, qs).$$

Let IO be the set of all finite inout-trees. $\delta = (IO, \cdot, \bullet)$ forms a commutative monoid. Let $\upsilon : \mathbf{N}^2 \mapsto IO$, where $\upsilon(p,q) = \tau(p,q)$. Since υ is a bijection, $\gamma \cong \delta$.

Each underlying set has multiple representations of each particular positive rational number. Let $(p,q) \equiv_n (r,s)$ when $ps = qr$ and $\tau(p,q) \equiv_t \tau(r,s)$ when $ps = qr$. Let $[(p,q)]$ denote the equivalence class under $\equiv_n$ of ordered pairs (r,s) that are equivalent to (p,q). Let $[\tau(p,q)]$ denote the equivalence class under $\equiv_t$ of inout-trees $\tau(r,s)$ that are equivalent to $\tau(p,q)$ (see Figure 6.

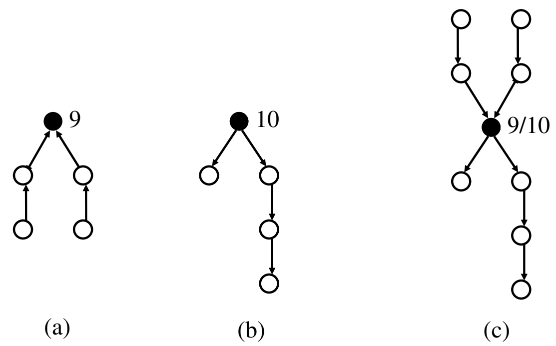


Figure 4: (a) $\tau_i(9)$; (b) $\tau_o(10)$; (c) $\tau(9,10)$.

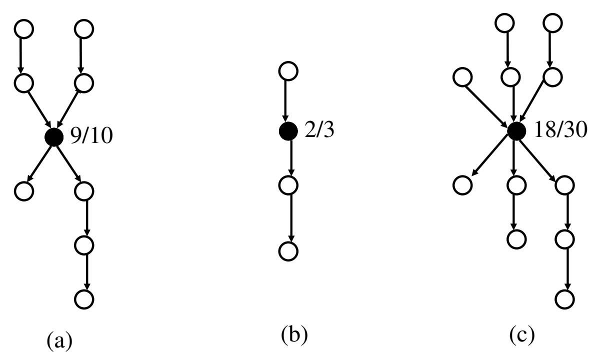


Figure 5: (a) $\tau(9,10)$; (b) $\tau(2,3)$; (c) $\tau(18,30)$.

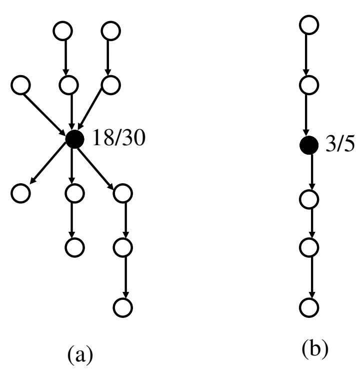


Figure 6: (a) $\tau(18,30) \equiv_t \tau(3,5)$ (b).

Then, referring to quotient sets, $\zeta = (\gamma/\equiv_n, \cdot, (1,1))$ is an abelian group, where $[(p,q)]^{-1} = [(q,p)]$; $\eta = (\delta/\equiv_t, \cdot, \bullet)$ is an abelian group, where $[\tau(p,q)]^{-1} = [\tau(q,p)]$. Moreover, since the resulting equivalence classes are in obvious one-to-one correspondence, $\zeta \cong \eta$. Each positive rational number thus corresponds to a unique equivalence class of ordered pairs of natural numbers and a unique equivalence class of inout-trees. For a particular positive rational number, there is a unique ordered pair, (p,q), where p and q have no common factors. Associate that positive rational number with $\tau(p,q)$. Each positive rational number thus has a unique representation as a finite inout-tree.

5 Recursive factor-exponent notation

In what follows, we are motivated to reduce replicated substructures in our number representation. Let $n = f_1^{e_1} f_2^{e_2} \cdots f_m^{e_m}$, where the f_i are the m unique prime factors of n and e_i is the number of times that f_i occurs in the prime decomposition of n. A number n thus can be represented by a multiset set of its prime factors and their exponents:

$$\{f_1 \cdot e_1, f_2 \cdot e_2, \ldots, f_m \cdot e_m\}.$$

In this representation, since each prime factor has an associated exponent, a notation that lists these pairs in increasing order of prime factor is an unambiguous representation of n:

$$(f_1\ e_1\ f_2\ e_2\ \ldots\ f_m\ e_m).$$

Since the first, third, fifth, etc. numbers in such a list are prime factors, it is unambiguous to replace them with their rank:

$$(p^{-1}(f_1)\ e_1\ p^{-1}(f_2)\ e_2\ \ldots\ p^{-1}(f_m)\ e_m).$$

This process can recursively be applied to each of these natural numbers in such a list, with 1 represented by (). We can formalize this representation as follows:

$$\begin{aligned} \lambda(n) &= (), \text{if } n = 1 && (1) \\ &= (\lambda(p^{-1}(f_1))\ \lambda(e_1)\ \lambda(p^{-1}(f_2))\ \lambda(e_2)\ \ldots\ \lambda(p^{-1}(f_m))\ \lambda(e_m), \text{if } n = f_1^{e_1} f_2^{e_2} \cdots f_m^{e_m} && (2) \end{aligned}$$

The application of these rules to the number 360 is given below:

$$\begin{aligned} \lambda(360) &= \lambda(2^3 3^2 5^1) \\ &= (\lambda(p^{-1}(2))\ \lambda(3)\ \lambda(p^{-1}(3))\ \lambda(2)\ \lambda(p^{-1}(5))\ \ \lambda(1)) \\ &= (()((()())())(()())(()())((()())())()) \end{aligned}$$

Figure 7 displays $\lambda(n)$, for $n = 1, 2, \ldots, 16$.

Each number is well-formed with respect to nesting of parentheses. In a sequence of such numbers, boundaries between successive numbers thus is easy to detect.

1	()	01
2	(() ())	001011
3	((() ()) ())	0001011011
4	(() (() ()))	0010010111
5	(((() ()) ()) ())	00001011011011
6	(() () (() ()) ())	00101001011011
7	((() (() ())) ())	00010010111011
8	(() ((() ()) ()))	00100010110111
9	((() ()) (() ()))	00010110010111
10	(() () ((() ()) ()) ())	001010001011011011
11	((((() ()) ()) ()) ())	000001011011011011
12	(() (() ()) (() ()) ())	001001011001011011
13	((() () (() ()) ()) ())	000101001011011011
14	(() () (() (() ())) ())	001010010010111011
15	((() ()) () ((() ()) ()) ())	0001011010001011011011
16	(() (() (() ())))	00100100101111

Figure 7: $\lambda(n)$, for $n = 1, 2, \ldots, 16$. Each right column entry is its left column entry transformed: "(" $\mapsto$ "0" and ")" $\mapsto$ "1".

Given two natural numbers in prime factor-exponent form, putting their product in the same form would seem to require adding the exponents of common factors. For this reason, λ does not appear to have a simple, structural, homomorphic product operator. Perhaps there are applications where a computationally complexity product is desirable (e.g., encryption/decryption).

Natural numbers and labeled trees

In this section, we represent the prime-factor-exponent description of a natural number as a rooted, undirected tree whose edges are labeled with natural numbers. In what follows, we refer to a rooted, undirected tree whose edges are labeled with natural numbers, perhaps with undue brevity, as a *labeled tree*. Let **L** be the set of finite labeled trees.

Define the function $\mu : \mathbf{N} \mapsto \mathbf{L}$ recursively as follows:

1. $\mu(1)$ is the labeled tree comprised of exactly one node.

2. For $n > 1$, let $n = f_1^{e_1} f_2^{e_2} \cdots f_m^{e_m}$, where the prime factors of n are $f_1, f_2 \ldots, f_m$ and e_i is the number of times that f_i is a factor in n. $\mu(n)$ is the labeled tree whose root is adjacent to labeled subtrees $\mu(p^{-1}(f_1)), \mu(p^{-1}(f_2)), \ldots, \mu(p^{-1}(f_m))$ and the edge joining the root to labeled subtree $\mu(p^{-1}(f_i))$ is labeled e_i.

Figure **??** displays the trees associated with $n = 1, 2, \ldots, 16$.

Figure 8: The first 16 labeled trees.

We will say that two labeled trees t and u are *isomorphic* when

1. the underlying trees are isomorphic

2. if edge $e \in t$ corresponds to edge $f \in u$ then these edges are labeled with the same natural number.

That is, two labeled trees are isomorphic when both their structure and their labeling match.

Let $\mathbf{M} \subset \mathbf{L}$ be the set of labeled trees such that, if u and v are distinct direct subtrees of $t \in \mathbf{M}$, then u is *not* isomorphic to v. We refer to $\mathbf{M}$ as the set of *distinct-sibling labeled trees.* μ is a bijection between the natural numbers and $\mathbf{M}$, a fact that can be proved by strong induction on n. Consequently, $\mu(\tau^{-1}(n))$ is a bijection between finite, rooted, undirected trees and distinct-sibling labeled trees.

Define the product of labeled trees t_1 and t_2, denoted, $t_1 \cdot t_2$, to be the labeled tree that results from:

1. identifying their roots

2. identifying s_1, labeled subtree directly beneath the root of t_1, with s_2, labeled subtree directly beneath the root of t_2, when s_1 is isomorphic with s_2. In this case, the label of the edge connecting the root of the product labeled tree and this labeled subtree is $e_1 + e_2$, where $e_1[e_2]$ is the label of the edge between the root of $t_1[t_2]$ and $s_1[s_2]$.

Figure 9 illustrates this definition.

Let $\kappa = (\mathbf{M}, \cdot, \bullet)$ denote the monoid of finite labeled trees under product. Since μ is a bijection, $\kappa \cong \alpha \cong \beta$. Thus, there is a bijection between finite rooted undirected trees and distinct-sibling labeled trees:

$$\mu \circ \tau^{-1}(t) \in \mathbf{M}, \text{ where t } \in .$$

Analogous to unlabeled, rooted trees, we can define labeled in-trees, labeled out-trees, and labeled inout-trees, and analogous algebraic structures and congruences, finally leading to an abelian group that is congruent to the positive rationals, and a representative finite labeled inout-tree that corresponds to a particular rational number.

Note: We could create a representation where the natural number labels are replaced with a labeled tree (except that they would be special trees whose edges were (or were associated with) labeled trees. We do not do this; it would lead to the same algebraic problem: The product operator requires adding exponents (i.e., edge labels), which is computationally complex - again, perhaps that is merely a feature.

Note: Related bijections between natural numbers and trees: Let $P \subset \mathbf{N}$ be a property possessed by some natural numbers such that every natural number [nonnegative integer] has a unique

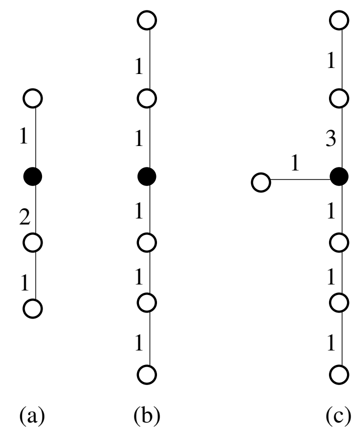


Figure 9: $\mu(18) \cdot \mu(15) = \mu(18 \cdot 15)$.

decomposition as a subset of P or a multisubset of P. Since natural numbers [nonnegative integers] are totally ordered, P is totally ordered. If the least element in P is greater than 1 [greater than 0], the index of an element $e \in P$ is less than p: $e_i \in P \Rightarrow i < p$. This immediately leads to a recursive map from the natural numbers to rooted undirected trees. Examples of such properties are

1. $\{b^n | n \geq 0\}$, for base $b > 1$ (decomposition is addition)
2. $\{n! | n \geq 1\}$ (decomposition is addition)
3. the set of prime numbers (decomposition is product)
4. others?

When is this a bijection between natural numbers [nonnegative integers] and rooted undirected trees? The set of such n bijections gives rise to a set n choose 2 bijections between the different indexed sets of trees. (There are many such bijections between countably infinite sets, but these are simply to describe.) When does the homomorphism $m(n_1 \cdot n_2) = m(n_1) \cdot m(n_2)$ have a natural structural correspondence (like primes: simply identify the roots)? (Hint: Show that it must be a decomposition into multisets.) For multiset decompositions, there is a labeled decomposition (e.g., prime-exponent representation) and the map is to distinct-sibling labeled trees.

6 Conclusion

This bijection between the natural numbers and the rooted trees is intriguing for at least two reasons:

1. its construction is easy to describe, yet apparently hard to compute;
2. the sequence of rooted trees corresponding to the increasing sequence of natural numbers is *not* monotonically nondecreasing in the number of vertices. For example, $|\tau(15)| = 6$, while $|\tau(16)| = 5$.

This bijection leads to a canonical presentation of rooted trees that is different from Beyer and Hedetniemi's [1]. By introducing the product operations, some algebraic structures on finite rooted trees arise naturally, including a unique representation of positive rational numbers as finite inout-trees. In addition to being an intriguing bijection, the apparent difficulty of computing τ and τ^{-1} may be of use practical use.

References

[1] T. Beyer and S. M. Hedetniemi. Constant time generation of rooted trees. *SIAM J. Comput.*, 9(4):706–712, 1980.

[2] A. Cayley. On the analytical forms called trees. *Mathematical Papers*, 11:365–367, 1896.

[3] A. Cayley. A theorem on trees. *Mathematical Papers*, 13:26–28, 1897.

[4] Shimon Even. *Graph Algorithms.* Computer Science Press, Inc, Rockville, MD, 1979.

[5] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Oxford University Press, Oxford, 5th edition, 1979.

[6] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.*, 68:145–254, 1937.

[7] H. Prüfer. Neuer Beweis eines Satzes über Permutationen. *Arch. Math. Phys.*, 27:742–744, 1918.

[8] R. C. Read. How to grow trees. In *Combinatorial Structures and Their Applications*, pages 343–347. Gordon and Breach, 1970.