REFERENCES

1. G. E. Bergum & A. W. Kranzler, "Linear Recurrences, Identities and Divisibility Properties" (unpublished paper).
2. Marjorie Bicknell & V. E. Hoggatt, Jr., "A Primer for the Fibonacci Numbers," The Fibonacci Association, San Jose State University, San Jose, California.
3. Marjorie Bicknell-Johnson & V. E. Hoggatt, Jr., "Variations on Summing a Series of Reciprocals of Fibonacci Numbers," *The Fibonacci Quarterly* (to appear).
4. Marjorie Bicknell-Johnson & V. E. Hoggatt, Jr., "A Reciprocal Series of Fibonacci Numbers with Subscripts $2^n k$," *The Fibonacci Quarterly* (to appear).
5. I. J. Good, "A Reciprocal Series of Fibonacci Numbers," *The Fibonacci Quarterly* 12, No. 4 (1974):346.
6. I. J. Good & Paul S. Bruckman, "A Generalization of a Series of De Morgan with Applications of Fibonacci Type," *The Fibonacci Quarterly* (to appear).
7. D. A. Millin, Problem H-237, *The Fibonacci Quarterly* 12, No. 3 (1974): 309.

*****

A NOTE ON 3-2 TREES*

EDWARD M. REINGOLD

University of Illinois at Urbana-Champaign, Urbana, IL 61801

ABSTRACT

Under the assumption that all of the 3-2 trees of height h are equally probable, it is shown that in a 3-2 tree of height h the expected number of keys is $(.72162)3^h$ and the expected number of internal nodes is $(.48061)3^h$.

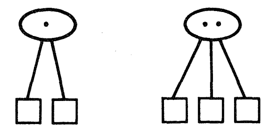
INTRODUCTION

One approach to the organization of large files is the use of "balanced" trees (see Section 6.2.3 of [3]). In particular, one such class of trees, suggested by J. E. Hopcroft (unpublished), is known as 3-2 trees. A 3-2 tree is a tree in which each internal node contains either 1 or 2 keys and is hence either a 2-way or 3-way branch, respectively. Furthermore, all external nodes (i.e., leaves) are at the same level. Figure 1 shows some examples of 3-2 trees.

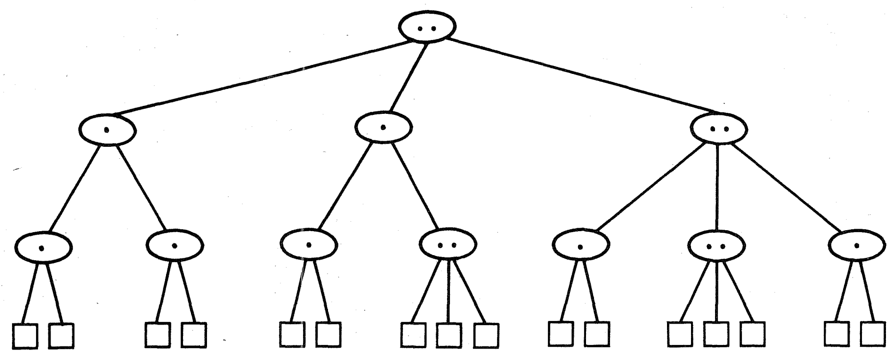
Insertion of a new key into a 3-2 tree is done as follows to preserve the 3-2 property: To add a new key into a node containing one key, simply insert it as the second key; if the node already contains two keys, split it into two one-key nodes and insert (recursively) the middle key into the parent node. This may cause the parent node to be split in a similar way, if it already contains two keys. For more details about 3-2 trees see [1] and [3].

This research was supported by the Division of Physical Research, U.S. Energy Research and Development Administration, and by the National Science Foundation (Grant GJ-41538).

(a) The unique 3-2 tree of height 0



(b) The only two 3-2 trees of height 1



(c) A 3-2 tree of height 3 with 15 keys, 11 internal nodes, and 16 external nodes (leaves)

FIGURE 1.—SOME EXAMPLES OF 3-2 TREES. THE SQUARES ARE EXTERNAL NODES (LEAVES), THE OVALS ARE INTERNAL NODES, AND THE DOTS ARE KEYS.

Yao [4] has studied the average number of internal nodes in a 3-2 tree with k keys, assuming that the tree was built by a sequence of k random insertions done by the insertion algorithm outlined above. He found the expected number of internal nodes to be between $.70k$ and $.79k$ for large k. Unfortunately, the distribution of 3-2 trees induced by the insertion algorithm is not well understood and Yao's techniques will probably not be extended to provide sharper bounds.

Using techniques like those in Khizder [2], some results can be obtained, however, for the (simpler) distribution in which all 3-2 trees of height are equally probable. In this paper we show that, under this simpler distribution, in a 3-2 tree of height h the expected number of keys and internal nodes are, respectively, $(.72162)3^h$ and $(.48061)3^h$.

ANALYSIS

Let $a_{n,k,h}$ be the number of 3-2 trees of height h with n nodes and k keys. Since there is a unique tree of height 0 (consisting of a single leaf—see Figure 1), and since a 3-2 tree of height $h > 0$ is formed from either two or three 3-2 trees of height $h - 1$, we have

$$a_{n,k,0} = \begin{cases} 1 & \text{if } n = k = 0 \\ 0 & \text{otherwise} \end{cases}$$

(1) $$a_{n,k,h} = \sum_{\substack{i+j=n-1 \\ u+v=k-1}} a_{i,u,h-1}\,a_{j,v,h-1} + \sum_{\substack{i+j+l=n-1 \\ u+v+w=k-2}} a_{i,u,h-1}\,a_{j,v,h-1}\,a_{l,w,h-1}$$

Let

$$A_h(x,y) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} a_{n,k,h}\, x^n y^k$$

be the generating function for $a_{n,k,h}$. From (1) we have

$$A_0(x,y) = 1$$

(2)

$$A_h(x,y) = xyA_{h-1}^2(x,y) + xy^2A_{h-1}^3(x,y)$$

and thus the number of 3-2 trees of height h is $A_h = A_h(1,1)$, the total number of keys in all 3-2 trees of height h is

$$K_h = \left.\frac{\partial A_h(x,y)}{\partial y}\right|_{x=y=1},$$

and the total number of internal nodes in all 3-2 trees of height h is

$$N_h = \left.\frac{\partial A_h(x,y)}{\partial x}\right|_{x=y=1}.$$

The table gives the first few values for A_h, K_h, and N_h as calculated from the recurrence relations arising from (2).

THE FIRST FEW VALUES FOR A_h, K_h, AND N_h

h	$A_h = A_h(1,1)$	$K_h = \left.\frac{\partial A_h(x,y)}{\partial y}\right\|_{x=y=1}$	$N_h = \left.\frac{\partial A_h(x,y)}{\partial x}\right\|_{x=y=1}$
0	1	0	0
1	2	3	2
2	12	68	44
3	1872	34608	21936
4	6563711232	377092654848	237180213504

Assuming that all of the 3-2 trees of height h are equally probable, the average number of keys in a 3-2 tree of height h is given by

$$\kappa_h = \frac{K_h}{A_h} = \left.\frac{\dfrac{\partial A_h(x,y)}{\partial y}}{A_h(x,y)}\right|_{x=y=1}$$

and the average number of internal nodes in a 3-2 tree of height h is given by

$$\nu_h = \frac{N_h}{A_h} = \left.\frac{\dfrac{\partial A_h(x,y)}{\partial x}}{A_h(x,y)}\right|_{x=y=1}$$

To determine κ_h, we use the recurrence relations for A_h and K_h arising from (2):

$$A_0 = 1$$

$$A_h = A_{h-1}^2 + A_{h-1}^3$$

and

$$K_0 = 0$$

$$K_h = 2A_{h-1}K_{h-1} + A_{h-1}^2 + 2A_{h-1}^3 + 3A_{h-1}^2K_{h-1}.$$

Rewriting the equation for K_h in terms of κ_h gives

$$K_h = \kappa_{h-1}(3A_h - A_{h-1}^2) + 2A_h - A_{h-1}^2$$

and so

$$\kappa_h = \frac{K_h}{A_h} = \kappa_{h-1}\left(3 - \frac{A_{h-1}^2}{A_h}\right) + 2 - \frac{A_{h-1}^2}{A_h}$$

$$= 3\kappa_{h-1} + 2 - \frac{A_{h-1}^2}{A_h}(\kappa_{h-1} + 1)$$

giving

$$(\kappa_h + 1) = 3(\kappa_{h-1} + 1) - \frac{K_{h-1} + A_{h-1}}{A_{h-1}^2 + A_{h-1}}.$$

Letting $\varepsilon_h = \dfrac{K_h + A_h}{A_h^2 + A_h}$, we get

$$(\kappa_h + 1) = 3^h(\kappa_0 + 1) - \sum_{i=1}^{h} 3^{i-1}\varepsilon_{h-i}.$$

But $\kappa_0 + 1 = \dfrac{K_0}{A_0} + 1 = \dfrac{0}{1} + 1 = 1$, and so

$$\frac{K_h}{A_h} + 1 = \kappa_h + 1 = 3^h\left(1 - \sum_{i=1}^{h} \frac{\varepsilon_{h-i}}{3^{h-i+1}}\right) = 3^h\left(1 - \sum_{i=0}^{h-1} \frac{\varepsilon_i}{3^{i+1}}\right) \tag{3}$$

i.e.,

$$\lim_{h\to\infty} \frac{1}{3^h}\left(\frac{K_h}{A_h} + 1\right) = 1 - \sum_{i=0}^{\infty} \frac{\varepsilon_i}{3^{i+1}}.$$

What is $\displaystyle\sum_{i=0}^{\infty} \frac{\varepsilon_i}{3^{i+1}}$? It is easy to show by induction that $A_i^2 > K_i$ and so

$$\varepsilon_i = \frac{K_i + A_i}{A_i^2 + A_i} < 1.$$

The comparison test thus insures that the summation converges:

$$\sum_{i=0}^{\infty} \frac{\varepsilon_i}{3^{i+1}} < \sum_{i=0}^{\infty} \frac{1}{3^{i+1}} = \frac{1}{2}.$$

Now, in order to use $\displaystyle\sum_{i=0}^{h} \frac{\varepsilon_i}{3^{i+1}}$ as an approximation to $\displaystyle\sum_{i=0}^{\infty} \frac{\varepsilon_i}{3^{i+1}}$ we need an upper bound on $\sum_{i=h+1}^{\infty} \frac{\varepsilon_i}{3^{i+1}}$. From the definition of ε_i, we have

(4) $$\sum_{i=h+1}^{\infty} \frac{\varepsilon_i}{3^{i+1}} = \frac{1}{3} \sum_{i=h+1}^{\infty} \frac{1}{3^i} \frac{K_i + A_i}{A_i^2 + A_i} = \frac{1}{3} \sum_{i=h+1}^{\infty} \frac{\frac{1}{3^i}\frac{K_i}{A_i} + \frac{1}{3^i}}{A_i + 1}.$$

From (3) and the fact that $0 < \varepsilon_i < 1$, we know that

$$\frac{1}{3^h}\frac{K_h}{A_h} + \frac{1}{3^h} = 1 - \sum_{i=0}^{h-1} \frac{\varepsilon_i}{3^{i+1}} < 1,$$

and so (4) becomes

$$\sum_{i=h+1}^{\infty} \frac{\varepsilon_i}{3^{i+1}} < \frac{1}{3} \sum_{i=h+1}^{\infty} \frac{1}{A_i + 1} < \frac{1}{3} \sum_{i=h+1}^{\infty} \frac{1}{A_i}.$$

But since $A_h = A_{h-1}^2 + A_{h-1}^3 > 2A_{h-1}^2$, we have by induction that $A_h > \frac{1}{2} 2^{2^h}$, and so

$$\sum_{i=h+1}^{\infty} \frac{\varepsilon_i}{3^{i+1}} < \frac{2}{3} \sum_{i=h+1}^{\infty} 2^{-2^i} < \frac{2}{3}\left(2^{-2^{h+1}} + 2^{-2^{h+1}-1}\right) = 2^{-2^{h+1}}.$$

Using the values in the table, we find that

$$\sum_{i=0}^{4} \frac{\varepsilon_i}{3^{i+1}} = .2783810593,$$

and thus

$$0 < \sum_{i=0}^{\infty} \frac{\varepsilon_i}{3^{i+1}} - .2783810593 < 2^{-2^5} < 3 \times 10^{-10}.$$

We conclude that

$$0 < \lim_{h\to\infty} \frac{1}{3^h}\left(\frac{K_h}{A_h} + 1\right) - .7216189407 < 3 \times 10^{-10}.$$

Thus, under the assumption that all the 3-2 trees of height h are equally probable, the expected number of keys in a 3-2 tree of height h is

$$\kappa_h = \frac{K_h}{A_h} \approx (.7216189407)3^h.$$

A similar analysis works for ν_h, the average number of internal nodes in a 3-2 tree of height h. We again use the recurrence relations arising from (2):

$$A_0 = 1$$

$$A_h = A_{h-1}^2 + A_{h-1}^3$$

as before, and

$$N_0 = 0$$

$$\begin{aligned} N_h &= 2A_{h-1}N_{h-1} + A_{h-1}^2 + A_{h-1}^3 + 3A_{h-1}^2 N_{h-1} \\ &= 2A_{h-1}N_{h-1} + 3A_{h-1}^2 N_{h-1} + A_h. \end{aligned}$$

Rewriting this last equation in terms of $\nu_h = N_h/A_h$ gives

$$N_h = \nu_{h-1}(3A_h - A_{h-1}^2) + A_h,$$

and so

$$\nu_h = \frac{N_h}{A_h} = \nu_{h-1}\left(3 - \frac{A_{h-1}}{A_h}\right) + 1 = 3\nu_{h-1} + 1 - \frac{N_{h-1}}{A_h},$$

giving

$$\left(\nu_h + \frac{1}{2}\right) = 3\left(\nu_{h-1} + \frac{1}{2}\right) - \frac{N_{h-1}}{A_h}.$$

Letting $\delta_h = \dfrac{N_h}{A_h}$, we get

$$\left(\nu_h + \frac{1}{2}\right) = 3^h\left(\nu_0 + \frac{1}{2}\right) - \sum_{i=1}^{h} 3^{i-1}\delta_{h-i}.$$

But $\nu_0 + \dfrac{1}{2} = \dfrac{N_0}{A_0} + \dfrac{1}{2} = \dfrac{0}{1} + \dfrac{1}{2} = \dfrac{1}{2}$, and so

$$\frac{N_h}{A_h} + \frac{1}{2} = \nu_h + \frac{1}{2} = 3^h\left(\frac{1}{2} - \sum_{i=1}^{h} \frac{\delta_{h-i}}{3^{h-i+1}}\right) = 3^h\left(\frac{1}{2} - \sum_{i=0}^{h-i} \frac{\delta_i}{3^{i+1}}\right), \tag{5}$$

i.e.,

$$\lim_{h\to\infty} \frac{1}{3^h}\left(\frac{N_h}{A_h} + \frac{1}{2}\right) = \frac{1}{2} - \sum_{i=0}^{\infty} \frac{\delta_i}{3^{i+1}}.$$

What is $\displaystyle\sum_{i=0}^{\infty} \frac{\delta_i}{3^{i+1}}$? It is easy to show by induction that $A_{i+1} > N_i$ and so $\delta_i = N_i/A_{i+1} < 1$; hence, the comparison test insures that the summation converges:

$$\sum_{i=0}^{\infty} \frac{\delta_i}{3^{i+1}} < \sum_{i=0}^{\infty} \frac{1}{3^{i+1}} = \frac{1}{2}.$$

In order to use $\displaystyle\sum_{i=0}^{h} \frac{\delta_i}{3^{i+1}}$ as an approximation to $\displaystyle\sum_{i=0}^{\infty} \frac{\delta_i}{3^{i+1}}$ we need an upper bound on $\displaystyle\sum_{i=h+1}^{\infty} \frac{\delta_i}{3^{i+1}}$. From the definition of δ_i, we have

$$\sum_{i=h+1}^{\infty} \frac{\delta_i}{3^{i+1}} = \frac{1}{3}\sum_{i=h+1}^{\infty} \frac{\frac{1}{3^i}\frac{N_i}{A_i}}{A_i + A_i^2}. \tag{6}$$

Since $0 < \delta_i < 1$, (5) tells us that

$$\frac{1}{3^h}\frac{N_h}{A_h} = \frac{1}{2}\left(1 - \frac{1}{3^h}\right) - \sum_{i=0}^{h-1} \frac{\delta_i}{3^{i+1}} < \frac{1}{2},$$

and so (6) becomes

$$\sum_{i=h+1}^{\infty} \frac{\delta_i}{3^{i+1}} < \frac{1}{6}\sum_{i=h+1}^{\infty} \frac{1}{A_i + A_i^2} < \frac{1}{6}\sum_{i=h+1}^{\infty} \frac{1}{A_i^2}.$$

Recalling that $A_i > \frac{1}{2}2^{2^i}$, this becomes

$$\sum_{i=h+1}^{\infty} \frac{\delta_i}{3^{i+1}} < \frac{1}{6}\sum_{i=h+1}^{\infty} 4 \cdot 2^{-2^{i+1}} = \frac{2}{3}\sum_{i=h+2}^{\infty} 2^{-2^i} < \frac{2}{3}\left(2^{-2^{k+2}} + 2^{-2^{k+2}-1}\right) = 2^{-2^{k+2}}.$$

Using the values in the table, we find that

$$\sum_{i=0}^{3} \frac{\delta_i}{3^{i+1}} = .0193890884,$$

and thus

$$0 < \sum_{i=0}^{\infty} \frac{\delta_i}{3^{i+1}} - .0193890884 < 2^{-2^5} < 3 \times 10^{-10}.$$

We conclude that

$$0 < \lim_{h\to\infty} \frac{1}{3^h}\left(\frac{N_h}{A_h} + \frac{1}{2}\right) - .4806109116 < 3 \times 10^{-10}.$$

Thus, under the assumption that all 3-2 trees of height h are equally probable, the expected number of internal nodes in a 3-2 tree of height h is

$$\nu_h = \frac{N_h}{A_h} \approx (.4806109116)3^h.$$

REFERENCES

1. A. V. Aho, J. E. Hopcroft, & J. D. Ullman, *The Design and Analysis of Computer Algorithms* (Reading, Mass.: Addison-Wesley, 1974).
2. L. A. Khizder, "Some Combinatorial Properties of Dyadic Trees," *Zh. Vychisl. Matem. i Matem. Fiziki* (Russian) 6, No. 2 (1966):389-394. English translation in *USSR Comput. Math. and Math. Phys.* 6, No. 2 (1966): 283-290.
3. D. E. Knuth, *The Art of Computer Programming*, Vol. III: *Sorting and Searching* (Reading, Mass.: Addison-Wesley, 1973).
4. A. C. Yao, "On Random 3-2 Trees," *Acta Informatica* 9 (1978):159-170.

*****