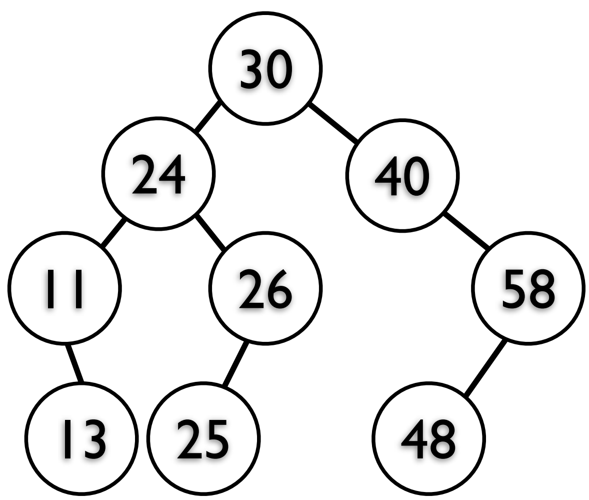
CISC235
Winter 2007
Homework for week 4
in preparation for quiz 2
Solutions

1. The tree resulting from the insertion sequence 30, 40, 24, 58, 48, 26, 11, 13, 25 is shown below.



2. The process of deleting the node with key 30 is illustrated in figure 1

3. The internal path length IPL is the sum of all pathlengths for all nodes in the tree. Note that the author of the text defines the pathlength of node v to be (level of) - 1. The quantities $path_{best}$ and $path_{worst}$ denote the average path length or the internal path length divided by the number of nodes in the tree. The worst case is realized when the binary tree degenerates into a linear linked list that is every node, except a single leaf node, has one child. Thus we have:

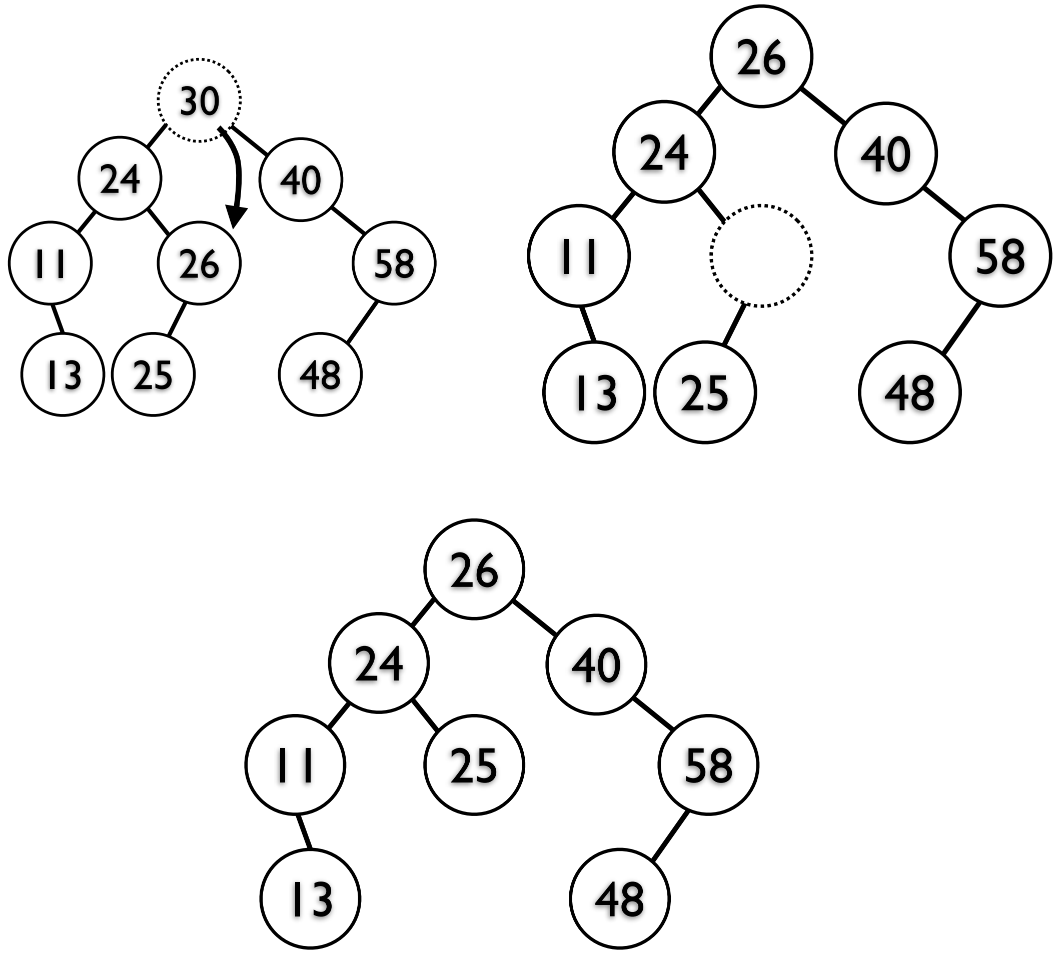


Figure 1: Find the predecessor node.Copy the contents.Delete the node.

$$path_{worst} = \frac{1}{n}\sum_{i=1}^{n}(i-1) = \frac{1}{n}i\sum_{i=1}^{n-1} i = \frac{n(n-1)}{2n} = \frac{n-1}{2}.$$

The best case is realized by a complete tree, that is where every non-leaf node has two children. The best internal path length BIPL, can be determined by observing that there are 2^i nodes with path length i. So for a binary tree of height h we obtain the sum

$$\sum_{i=0}^{h-1} i2^i.$$

The issue at hand is to evaluate the sum above. The author hints at a neat trick to evaluate the sum.

Let's look at a term by term expansion of the sum

$$\sum_{i=0}^{h-1} i2^i = 0 + 1 \times 2^1 + 2 \times 2^2 + 3 \times 2^3 + \ldots + (h-1) \times 2^{h-1} \tag{1}$$

and

$$2\sum_{i=0}^{h-1} i2^i = 0 + 1 \times 2^2 + 2 \times 2^3 + \ldots + (h-2) \times 2^{h-1} + (h-1) \times 2^h \tag{2}$$

Subtracting equation 1 from equation 2 (that is 2BIPL - BIPL = BIPL) we get

$$-1(2 + 2^2 + 2^3 + \ldots + 2^{h-1}) + (h-1)2^h = (h-1)2^h - \sum_{i=1}^{h-1} 2^i \tag{3}$$

We already know that

$$\sum_{i=0}^{h-1} 2^i = 2^h - 1$$

which also happens to be the value of n the number of nodes in a complete tree of height h. The sum $\sum_{i=1}^{h-1} 2^i = 2^h - 2$.

Therefore we conclude that:

$$\text{BIPL} = (h-2)2^h + 2$$

and that

$$path_{best} = \frac{(h-2)2^h + 2}{2^h - 1} \approx h - 2.$$

Since $n = 2^h - 1$ we conclude that $path_{best}$ is O($\log n$).

The sum $\sum_{i=1}^{h-1} i2^i$ will come up when we study heaps. I can't resist giving a much easier way to obtain upper and lower bounds. Observe that

$$\sum_{i=1}^{h-1} i2^i \leq (h-1)\sum_{i=1}^{h-1} 2^i = (h-1)(2^h - 1). \tag{4}$$

Also since approximately half of the terms in $\sum_{i=1}^{h-1} i2^i$ have i greater than $(h-1)/2$, we have

$$\sum_{i=1}^{h-1} i2^i \geq \frac{h-1}{2}\sum_{i=1}^{h-1} 2^i = \frac{h-1}{2}(2^h - 1). \tag{5}$$

For $n \approx 2^h$ both equations 4 and 5 are in O($n \log n$).