ANALYSIS OF SKIP LIST

Total number of nodes in a Skip List (Lemma 4.3):

Let's discuss about the probability of having a node for an element, x, at level r. The probability of having an element at level 0 is 1 – as all elements exist at level 0. The total number of expected nodes at level 1 (with high probability) can be defined as

$$L_r = L_0 * p$$

Where $p = \frac{1}{2}$ when we toss a fair coin to decide.

The total number of nodes, T, in a skip list can be calculated by summing up the expected number of nodes for each level:

$$T = L_0 + L_1 + L_2 + \cdots$$

$$T = n * p^0 + n * p^1 + n * p^2 + \cdots$$

$$T = \sum_{r=0}^{\infty} n * p^r$$

$$= \sum_{r=0}^{\infty} n * \frac{1}{2^r}$$

$$= n * \sum_{r=0}^{\infty} \frac{1}{2^r}$$

Using geometric sequence:

$$\sum_{r=0}^{\infty} \frac{1}{2^r} = \frac{1}{2^0} + \frac{1}{2^1} + \frac{1}{2^2} + \cdots = 2$$

Therefore, total number of expected nodes in a skip list is:

$$T = n * 2 = 2n$$

Height of a skip list (informal):

As discussed above, the number of nodes at level r can be defined as $\frac{n}{2^r}$. Accordingly, the number of nodes at level will be n ($\frac{n}{2^0} = n$). The number of nodes will be $\frac{n}{2^1}, \frac{n}{2^2}, \frac{n}{2^3}, \ldots$ for level 1, 2, 3, and so on. We know the last level should have at least one node. Therefore,

$$\frac{n}{2^r} = 1.$$

The number of nodes at each level can be written as:

$$\frac{n}{2^0}, \frac{n}{2^1}, \frac{n}{2^2}, \cdots, \frac{n}{2^r}$$

Whereas,

$$\frac{n}{2^r} = 1$$

Therefore,

$$r = \log n$$

Hence, the maximum level containing at least one node should not be more than $\log n$.

Formal proof: Expected Height of a Skip List (See Lemma 4.4 in Section 4.4 for detail discussion):

The height of a skip list can be defined as:

$$h = \sum_{r=1}^{\infty} I_r$$

Where I_r is an indicator variable. Its value is 1 when L_r is not empty, otherwise 0.

The expected height of h can be defined as:

$$E[h] = \sum_{r=1}^{\infty} E[I_r]$$

$$E[h] = \sum_{r=1}^{\lfloor \log n \rfloor} E[I_r] + \sum_{\lfloor \log n \rfloor + 1}^{\infty} E[I_r]$$

The value of I_r cannot be great than 1 for all levels less than or equal to $\log n$, and for the rest, $E[I_r]$ can never be more than total number of nodes at level r i.e. $\frac{n}{2^r}$:

$$E[h] \leq \sum_{r=1}^{\lfloor \log n \rfloor} 1 + \sum_{\lfloor \log n \rfloor + 1}^{\infty} \frac{n}{2^r} - (1)$$

For the right side of the equation,

$$\sum_{r=\lfloor \log(n) \rfloor + 1}^{\infty} \frac{n}{2^r}$$

$$\sum_{t=0}^{\infty} \frac{n}{2^{t+\lfloor \log n \rfloor + 1}}$$

$$\frac{n}{2^{\lfloor \log n \rfloor + 1}} \sum_{t=0}^{\infty} \frac{1}{2^t}$$

Using inequalities:

$$\lfloor \log n \rfloor + 1 > \log n$$

$$2^{\lfloor \log n \rfloor + 1} > n$$

$$n < 2^{\lfloor \log n \rfloor + 1}$$

$$\frac{n}{2^{\lfloor \log n \rfloor + 1}} < 1$$

Therefore:

$$\frac{n}{2^{\lfloor \log n \rfloor + 1}} \sum_{t=0}^{\infty} \frac{1}{2^t} < (1) \sum_{t=0}^{\infty} \frac{1}{2^t} = 2$$

Using in equation 1:

$$E[h] \leq \log n + 2 \ = O(\log n + 2)$$

Expected Length of a search path:

- Length of a search path has two components: steps to go downward and number of times need to move on the side. This can be proved similarly as the Lemma 4.4 as discussed above. See Lemma 4.6 in Section 4.4

Questions to attempt:

- Prove the expected height if the probability of increasing the height is 0.25 (instead of 0.5).
- If we are using an unfair coin with a probability of the head is 1/3, will the expected height increase or decrease? Would it have a similar impact on the search path or not?
- Prove that the expected number of nodes in a skiplist containing n elements, including all occurrences of the sentinel, is $2n + O(\log n)$.