CSE 312

Foundations of Computing II

Skip Lists

Administrivia

- TA's will run a review session in EEB 125 on Sunday from 4:30pm - 6pm.
- I will release a practice midterm later today or tomorrow.

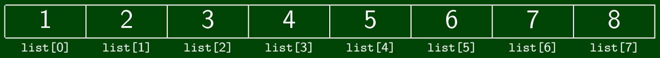
What Problem Are We Solving?

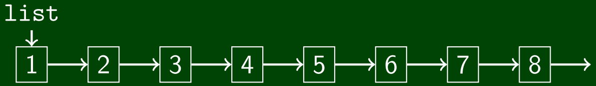
We would like to find a simple data structure to implement the `Set` ADT. Some possible choices include:

1	2	3	4	5	6	7	8
`list[0]`	`list[1]`	`list[2]`	`list[3]`	`list[4]`	`list[5]`	`list[6]`	`list[7]`

What Problem Are We Solving?

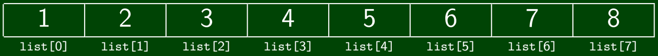
We would like to find a simple data structure to implement the `Set` ADT. Some possible choices include:

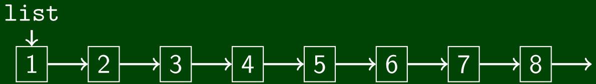


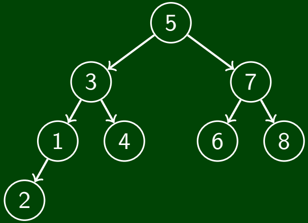


What Problem Are We Solving?

We would like to find a simple data structure to implement the `Set` ADT. Some possible choices include:

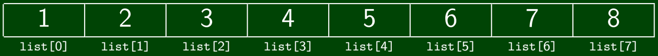


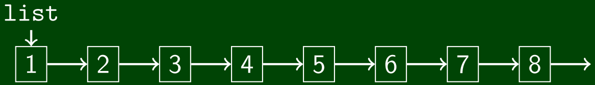


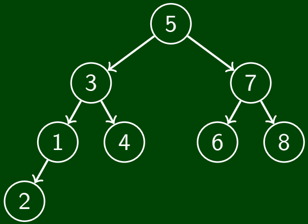


What Problem Are We Solving?

We would like to find a simple data structure to implement the `Set` ADT. Some possible choices include:

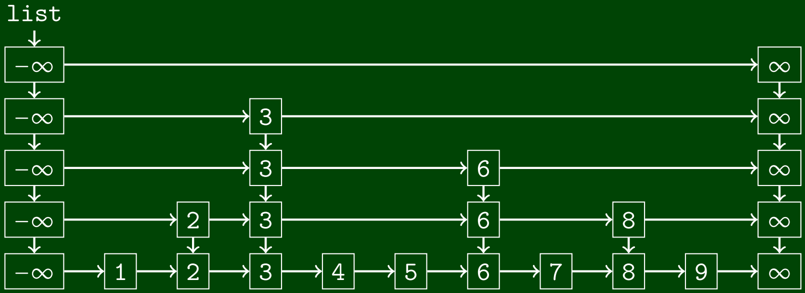




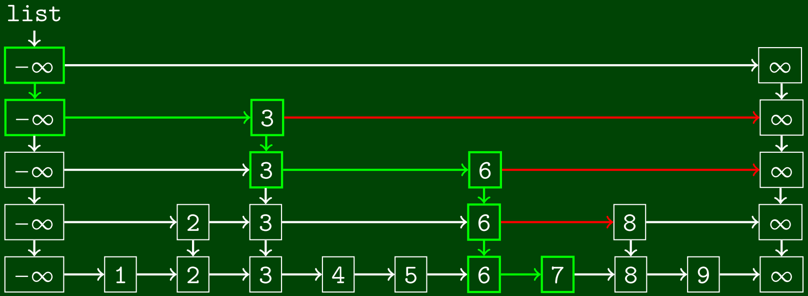


What if we combined these ideas? Could we get better behavior?

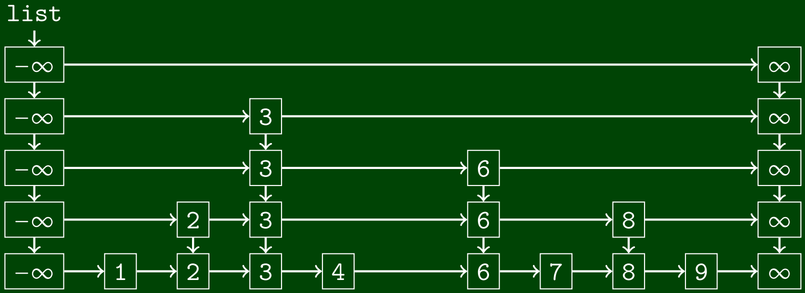
Skip Lists: The Idea



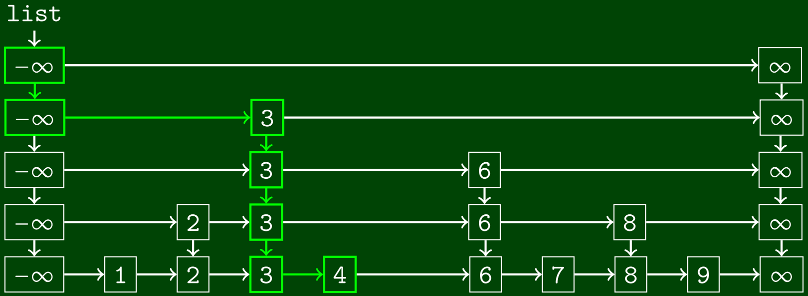
Skip Lists: Find



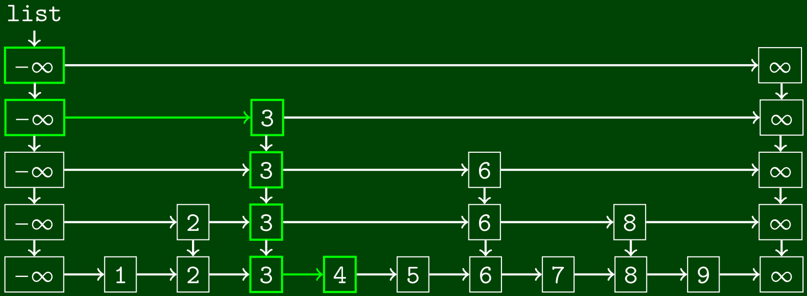
Skip Lists: Insert



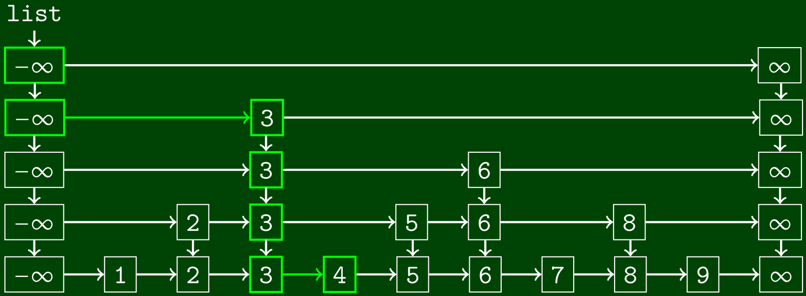
Skip Lists: Insert



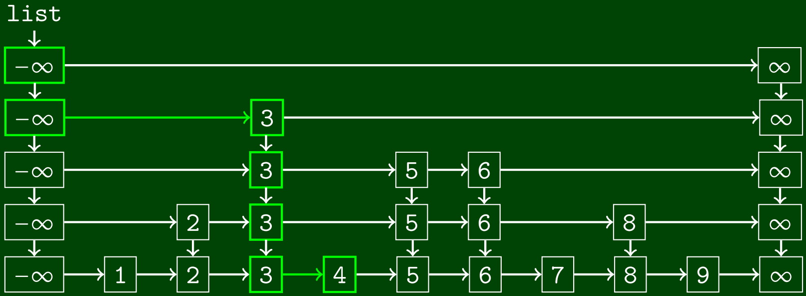
Skip Lists: Insert



Skip Lists: Insert



Skip Lists: Insert



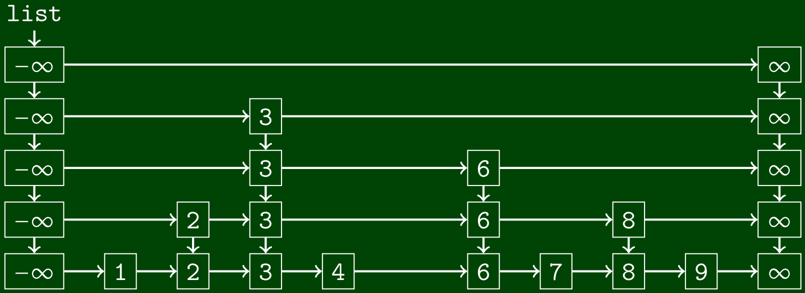
Definition (With High Probability)

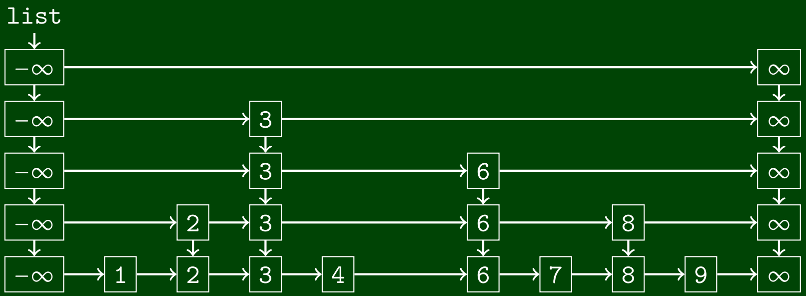
We say an event, parametrized on n, E_n, happens **with high probability** when $\lim_{n\to\infty} \Pr(E_n) = 1$.

Definition (Union Bound)

The probability that at least one of $\{A_1, \ldots, A_n\}$ happens is less than or equal to the sum of their probabilities. That is:

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n)$$

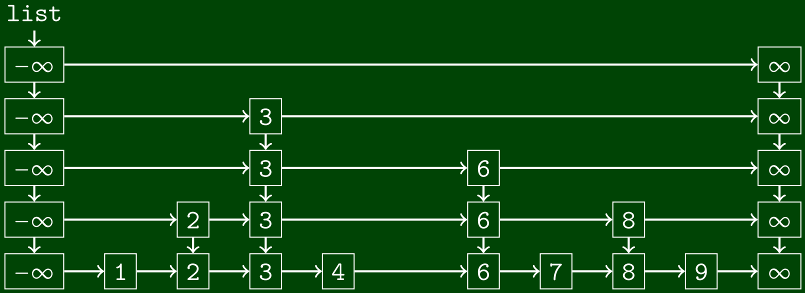

list
−∞
∞
−∞
3
∞
−∞
3
6
∞
−∞
2
3
6
8
∞
−∞
1
2
3
4
6
7
8
9
∞

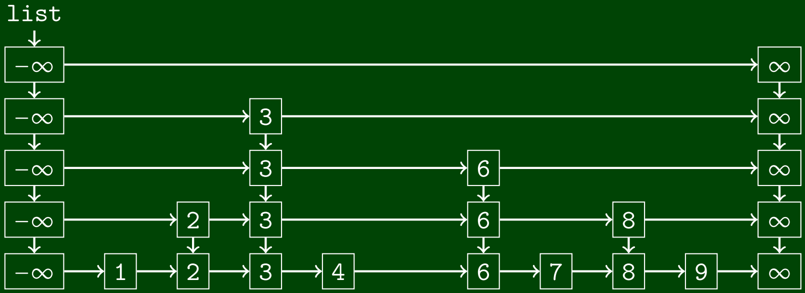


Let E_i be the event that some element is on the ith level. Note that this is the same as the Skip List having height i or larger.

Let x be a single entry of the Skip List, and $E_{x,i}$ be the event that x is on the ith level. Then, $\Pr(E_{x,i}) = \left(\frac{1}{2}\right)^i$.

Skip Lists: Analyzing Expected Height





By the union bound, the event that **any** entry reaches the ith level is less than the sum of their probabilities. That is: $\Pr(E_i) \leq \frac{n}{2^i}$.

Let $c \geq 2$ be a natural number constant.
Now, consider $\Pr\left(E_{c\lg n}\right) \leq \dfrac{n}{2^{c\lg n}} = \dfrac{n}{\left(2^{\lg n}\right)^c} = \dfrac{n}{n^c} = \dfrac{1}{n^{c-1}}$.

Then, $\lim\limits_{n\to\infty} \Pr\left(\overline{E_{c\lg n}}\right) \leq \lim\limits_{n\to\infty} 1 - \dfrac{1}{n^{c-1}} = 1$

Let E_i be the event that some element is on the ith level. Note that this is the same as the Skip List having height i or larger.

Let x be a single entry of the Skip List, and $E_{x,i}$ be the event that x is on the ith level. Then, $\Pr(E_{x,i}) = \left(\frac{1}{2}\right)^i$.

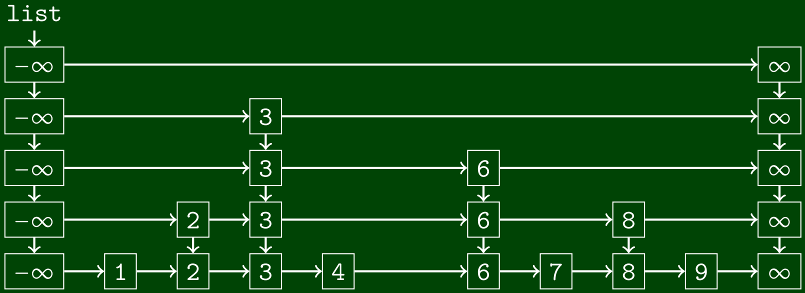
Let E_i be the event that some element is on the ith level. Note that this is the same as the Skip List having height i or larger.

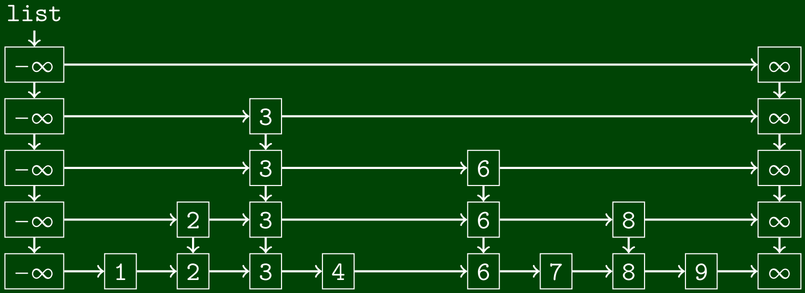
Let x be a single entry of the Skip List, and $E_{x,i}$ be the event that x is on the ith level. Then, $\Pr(E_{x,i}) = \left(\frac{1}{2}\right)^i$.

Let X be a r.v. for the number of nodes in the Skip List. Then, define X_i as an r.v. for the number of nodes on level i. Finally, let H be the height of the skip list.

Then, $X = \sum_{i=0}^{H} X_i$. So, $\mathbb{E}[X] = \sum_{i=0}^{H} \mathbb{E}[X_i] = \sum_{i=0}^{H} \sum_{x \in L} \Pr(E_{x,i}) < \sum_{i=0}^{\infty} \frac{n}{2^i} = 2n$.

Skip Lists: Analyzing Expected Runtime of Find

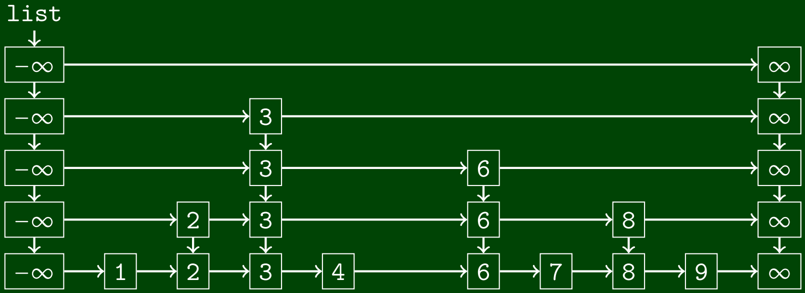


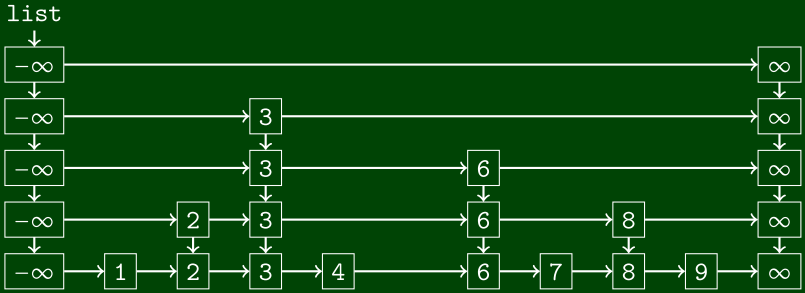


Let X be a r.v. for the number of node accesses in a single `find` operation. Then, $X = D + R$ where D is the number of "down" movements and R is the number of "right" movements. Furthermore, let R_i be the number of "right" movements on level i. Note that

$$\mathbb{E}[X] = \mathbb{E}[D] + \mathbb{E}[R] = \mathcal{O}(\lg n) + \sum \mathbb{E}[R_i]$$

Skip Lists: Analyzing Expected Runtime of Find





Consider a single row of the find operation. In particular, consider the keys that we visit on that row in reverse. We "go back left" until we hit a key that made it up to the $(i+1)$st level. Since we're considering level i, we already know this key made it to level i; so, the probability that it makes it to level $i+1$ is just $1/2$. So, the expected number of keys we visit on level i is just the number of times we flip a coin until we get a `HEADS` which means $\mathbb{E}[R_i] = 2$. So, $\mathbb{E}[X] = \mathcal{O}(\lg n)$.

Some Important Reminders for the Midterm

Important Distributions

- Bernoulli(p)
- Binomial(n, p)
- Geometric(p)
- Uniform(a, b)
- Poisson(λ)

Linearity

- Adding **Expectations DOES NOT** require independence.
- Adding **Variances DOES** require independence.