Skip Lists

William Pugh (1989)
- Easy to implement (as compared to balanced trees)
- Maintains a dynamic set of n elements in $O(\log n)$ time per operation in expectation and with high probability (w.h.p.)

One Linked List:
One (Sorted) linked list

14 $\rightarrow$ 23 $\rightarrow$ 34 $\rightarrow$ 42 $\rightarrow$ 50 $\rightarrow$ 59 $\rightarrow$ 66 $\rightarrow$ 72 $\rightarrow$ 79

Searches take $\Theta(n)$ time in worst case

Suppose we had two sorted linked lists
- each element can appear in one or both lists
Two Linked Lists

Express and local subway lines (à la New York City 7th Avenue Line)
- Express line connects a few of the stations
- Local line connects all stations
- Links between lines at common stations

Searching in Two Linked Lists

Search(x):
- Walk right in top linked list (L1) until going right would go too far
- Walk down to bottom linked list (L2)
- Walk right in L2 until element found (or not)

Search(59)
Analysis

Search cost \( \approx |L_1| \cdot \frac{|L_2|}{|L_1|} \)

Minimized when terms are equal

\( |L_1|^2 = |L_2| = n \)

\( |L_1| = \sqrt{n} \)  

Search is \( \Theta(\sqrt{n}) \)

More Linked Lists

2 sorted lists \( \Rightarrow 2 \cdot \sqrt{n} \)

3 sorted lists \( \Rightarrow 3 \cdot 3\sqrt{n} \)

k sorted lists \( \Rightarrow k \cdot k\sqrt{n} \)

\( \log n \) sorted lists \( \Rightarrow \log n \cdot \sqrt{n} \)

like a binary tree!
Searching in Ign Linked Lists

Try search(72)

Insert (x)

To insert an element x into a skip list

- Search(x) to see where x fits into bottom list
- Always insert into bottom list
- Insert into some of the lists above which ones?
- Flip fair coin
  - If HEADS: promote x to next level up
  - Else stop

this may be newly created
Warmup Lemma: \# levels in n-element skip list is $O(\lg n)$ w.h.p.
\[ c \cdot \lg n \quad \text{expected probability:} \quad \text{prob} \ 1 - \frac{1}{n^\alpha} \quad \text{related} \]

Proof: Failure probability (not $\leq c \lg n$ levels)
\[
= \Pr \{ \text{more than } c \lg n \text{ levels} \}^2 \\
= \Pr \{ \text{some element got promoted more than } c \lg n \text{ times} \}^2 \\
\leq n \cdot \Pr \{ \text{element } x \text{ got promoted more than } c \lg n \text{ times} \} \\
= n \cdot \left( \frac{1}{2} \right)^{c \lg n} \quad \text{by union bound} \\
= \frac{n^c}{n^c} = \frac{1}{n^\alpha} \quad \alpha = c - 1
\]
Theorem: Any search in an $n$-element skip list costs $O(\log n)$ w.h.p.

**Cool idea:** Analyze search backwards

- If search starts at node in bottom list.
- At each node visited:
  - If node wasn't promoted higher (tails here) then we go left
  - If node was promoted higher (heads here) then we go up
- Stop when we reach top level or $-\infty$

Look at arrows on page 4
Proof of Theorem

- Search makes "up" and "left" moves each with probability $\frac{1}{2}$

- Number of moves going "up" $\leq \# \text{ levels} \leq c \cdot \lg n \text{ w.h.p.}$ (by Warmup Lemma)

- Total number of moves $=$ number of coin flips until you get $c \cdot \lg n$ heads ("up" moves)

Claim: Number of coin flips until $c \cdot \lg n$ heads $= O(\lg n)$ w.h.p.
CHERNOFF BOUNDS

Theorem: Let $Y$ be a random variable representing the total number of heads in a series of $m$ independent coin flips, where each flip has a probability $p$ of coming up heads over tails. Then for all $r > 0$, we have

$$\Pr [Y > E[Y] + r] \leq e^{-\frac{2r^2}{m}}$$

Lemma: For any $c$, there is a constant $d$ such that with high probability (w.h.p.) the number of heads in flipping $d \lg n$ fair coins is at least $c \cdot \lg n$.  

Proof: Let $Y$ be the number of tails when flipping a fair coin $d \lg n$ times. $p = 1/2$. 

$m = d \lg n$, so $E[Y] = \frac{1}{2}m = \frac{d \lg n}{2}$.

We want to bound the probability of fewer than $c \cdot \lg n$ heads = the probability of getting at least $d \cdot \lg n - c \lg n$ tails.
Proof of Lemma (cont'd.)

\[ \Pr [Y > r (d-c) \lg n] = \Pr \left[ E[Y] + \frac{(d-c) \lg n}{r} \right] \]

Choose \( d = 8c \Rightarrow r = 3c \lg n \)

By Chernoff, prob of \( \leq c \cdot \lg n \) heads

\[ \leq e^{-\frac{2r^2}{m}} = e^{-\frac{2(3c \cdot \lg n)^2}{8c \cdot \lg n}} = e^{-\frac{9c \cdot \lg n}{4}} = e^{-c \cdot \lg n} = e^{-\frac{1}{c \cdot \lg n}} = \frac{1}{2^{c \cdot \lg n}} = \frac{1}{n^c} \]

😊 for Lemma
Proof of theorem (finally!)

Event A: number of levels \( \leq c \lg n \) w.h.p.
Event B: number of moves until \( c \lg n \) "up" moves \( \leq d \lg n \) w.h.p.

Event A and event B are not independent

Want to show \( \Pr(\text{event A} \& \text{event B}) \) high w.h.p.

\[
\Pr(\text{event A} \& \text{event B}) = \Pr(\overline{\text{event A}} + \overline{\text{event B}})
\]

\[
\leq \Pr(\overline{\text{event A}}) + \Pr(\overline{\text{event B}})
\]

\[
= \frac{1}{h^{c-1}} + \frac{1}{h^c}
\]

\[
= O\left(\frac{1}{h^{c-1}}\right)
\]

\[
\Pr(\text{event A} \& \text{event B}) \text{ w.h.p.}
\]

Search in \( O(\lg n) \) w.h.p.

☺️ --- for theorem.