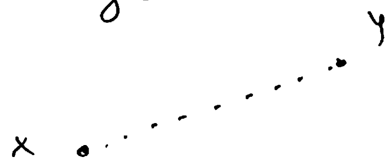
Analysis of Cuckoo Hashing [GT 6.4]

Given a set of keys S, we define the cuckoo graph to be the graph whose nodes are indices of T_1 and T_2 and whose edge set is $\{ (h_1(x), h_2(x)) \text{ for } x \in S \}$

<u>Fact</u>: The probability that there exists a path from x to y of lenght L in the cuckoo graph is at most $\frac{1}{2^L N}$ provided $N \geq 2n$

proof/ By induction on the lenght L

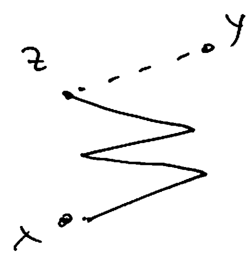
$(L = 1)$



same as the prob that $\exists\, k \in S$: $h_1(k) = x$ and $h_2(k) = y$

$$\text{Prob} \leq \sum_{k \in S} \Pr[h_1(k) = x \text{ and } h_2(k) = y] \leq \sum_{k \in S} \frac{1}{N^2} \leq \frac{n}{2N}$$

$(L > 1)$



$$\text{Prob} \leq \sum_{z} \Pr[\text{from } x \text{ to } z \text{ in } L-1 \text{ hops} \wedge \text{edge from } z \text{ to } y]$$

$$\leq \sum_{z} \frac{1}{2^{L-1} N} \cdot \frac{n}{2N} \leq \frac{n}{2^L N}$$

Fact: Expected length of eviction sequence ~~if~~ from entry x is $\frac{4}{N}$ if $N \geq 2n$

$$\mathbb{E}[\text{length of path from } x \text{ to } y] \leq \sum_{L \geq 1} \Pr[\text{path from } x \text{ to } y \text{ of length } L] \leq \sum_{L \geq 1} \frac{1}{2^L N} = \frac{1}{N}$$

Fact: Expected number of eviction for put provided put is successful.

$$\mathbb{E}[\text{length of path out of } x] \leq \sum_{k \in S} \mathbb{E}[\text{length from } x \text{ to } T_1[h_1(k)] \text{ or } T_2[h_2(k)]]$$

$$\leq \sum_{k \in S} \frac{2}{N} \leq \frac{2n}{N} \leq 1$$

Fact $\Pr[\exists \text{ cycle passing through } x] \leq \frac{1}{3N}$

$$\Pr[\ldots] \leq \sum_{L=1}^{\infty} \frac{1}{2^{2L} N} = \sum_{L \geq 1} \frac{1}{4^L N} = \frac{1}{N} \sum_{L \geq 1} \frac{1}{4^L} = \frac{1}{N} \frac{1/4}{1 - 1/4} = \frac{1}{3N}$$

Fact: $\Pr[\exists \text{ cycle on } n \text{ keys}] \leq \frac{2}{3}$ $\Longrightarrow$ $O(n)$ time to do n put operations

Fact: $E[\#\text{ of rehashes}] = 3$
(flip a biased coin)