Cuckoo Hashing

The probability that there exists an active cycle-structure is therefore at most

$$\sum_{s=3}^{\infty} s^3 \cdot n^{s-1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}} = \frac{\mu^2}{nm} \sum_{s=3}^{\infty} s^3 \left(\frac{m}{n}\right)^s \le \frac{\mu^2}{m^2} \sum_{s=3}^{\infty} s^3 \left(\frac{1}{1+\epsilon}\right)^s \le \mathcal{O}\left(\frac{1}{m^2}\right) \ .$$

Here we used the fact that $(1+\epsilon)m \le n$.

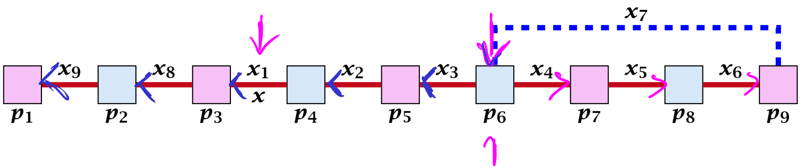
Hence,

$$\Pr[\text{cycle}] = \mathcal{O}\left(\frac{1}{m^2}\right) \ .$$

Cuckoo Hashing

Now, we analyze the probability that a phase is not successful without running into a closed cycle.

Cuckoo Hashing



Sequence of visited keys:

$x = x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_3, x_2, x_1 = x, x_8, x_9, \ldots$

Cuckoo Hashing

Consider the sequence of not necessarily distinct keys starting with x in the order that they are visited during the phase.

Lemma 30

If the sequence is of length p then there exists a sub-sequence of at least $\frac{p+2}{3}$ keys starting with x of distinct keys.

Cuckoo Hashing

Consider the sequence of not necessarily distinct keys starting with x in the order that they are visited during the phase.

Lemma 30

If the sequence is of length p then there exists a sub-sequence of at least $\frac{p+2}{3}$ keys starting with x of distinct keys.

Cuckoo Hashing

Proof.

Let i be the number of keys (including x) that we see before the first repeated key. Let j denote the total number of distinct keys.

The sequence is of the form:

$$x = x_1 \to x_2 \to \cdots \to x_i \to x_r \to x_{r-1} \to \cdots \to x_1 \to x_{i+1} \to \cdots \to x_j$$

As $r \le i - 1$ the length p of the sequence is

$$p = i + r + (j - i) \le i + j - 1 \ .$$

Either sub-sequence $x_1 \to x_2 \to \cdots \to x_i$ or sub-sequence $x_1 \to x_{i+1} \to \cdots \to x_j$ has at least $\frac{p+2}{3}$ elements. □

Cuckoo Hashing

Proof.

Let i be the number of keys (including x) that we see before the first repeated key. Let j denote the total number of distinct keys.

The sequence is of the form:

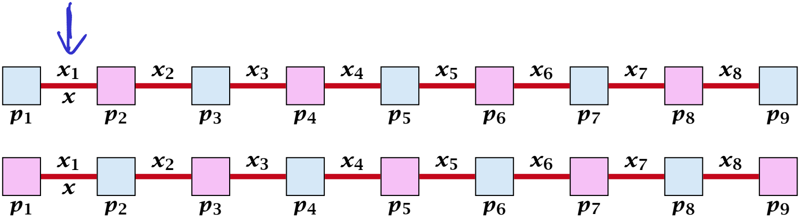
$$x = x_1 \to x_2 \to \cdots \to x_i \to x_r \to x_{r-1} \to \cdots \to x_1 \to x_{i+1} \to \cdots \to x_j$$

As $r \leq i-1$ the length p of the sequence is

$$p = i + r + (j - i) \leq i + j - 1 \ .$$

Either sub-sequence $x_1 \to x_2 \to \cdots \to x_i$ or sub-sequence $x_1 \to x_{i+1} \to \cdots \to x_j$ has at least $\frac{p+2}{3}$ elements. □

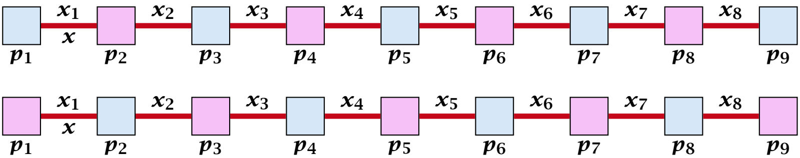
Cuckoo Hashing



A path-structure of size s is defined by

- $s+1$ different cells (alternating btw. cells from T_1 and T_2).
- s distinct keys $x = x_1, x_2, \ldots, x_s$, linking the cells.
- The leftmost cell is either from T_1 or T_2.

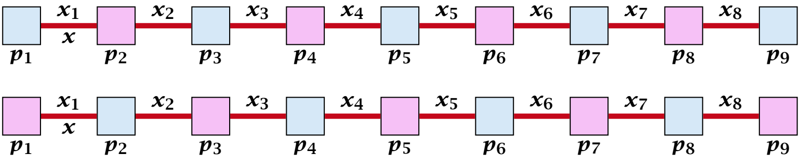
Cuckoo Hashing



A path-structure of size s is defined by

- $s+1$ different cells (alternating btw. cells from T_1 and T_2).
- s distinct keys $x = x_1, x_2, \ldots, x_s$, linking the cells.
- The leftmost cell is either from T_1 or T_2.

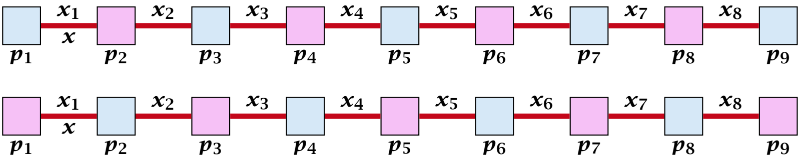
Cuckoo Hashing



A path-structure of size s is defined by

- $s+1$ different cells (alternating btw. cells from T_1 and T_2).
- s distinct keys $x = x_1, x_2, \ldots, x_s$, linking the cells.
- The leftmost cell is either from T_1 or T_2.

Cuckoo Hashing



A path-structure of size s is defined by

- n^{s+1} ▶ $s+1$ different cells (alternating btw. cells from T_1 and T_2).
- m^{s-1} ▶ s distinct keys $x = x_1, x_2, \ldots, x_s$, linking the cells.
- 2 ▶ The leftmost cell is either from T_1 or T_2.

Cuckoo Hashing

A path-structure is active if for every key x_ℓ (linking a cell p_i from T_1 and a cell p_j from T_2) we have

$$h_1(x_\ell) = p_i \quad \text{and} \quad h_2(x_\ell) = p_j$$

$\frac{\mu}{n^s}$ $\frac{\mu}{n^s}$ $\frac{\mu^2}{n^{2s}}$

Observation:

If a phase takes at least t steps without running into a cycle there must exist an active path-structure of size $(2t + 2)/3$.

p keys $\Rightarrow$ $\frac{p+2}{3}$

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}}$$

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}}$$

$$\leq 2\mu^2 \left(\frac{m}{n}\right)^{s-1}$$

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}}$$
$$\leq 2\mu^2 \left(\frac{m}{n}\right)^{s-1} \leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{s-1}$$

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}}$$

$$\leq 2\mu^2 \left(\frac{m}{n}\right)^{s-1} \leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{s-1}$$

Plugging in $s = (2t+2)/3$ gives

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}}$$

$$\leq 2\mu^2 \left(\frac{m}{n}\right)^{s-1} \leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{s-1}$$

Plugging in $s = (2t+2)/3$ gives

$$\leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{(2t+2)/3-1}$$

Cuckoo Hashing

The probability that a given path-structure of size s is active is at most $\frac{\mu^2}{n^{2s}}$.

The probability that there exists an active path-structure of size s is at most

$$2 \cdot n^{s+1} \cdot m^{s-1} \cdot \frac{\mu^2}{n^{2s}} \leq 2\mu^2 \left(\frac{m}{n}\right)^{s-1} \leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{s-1}$$

Plugging in $s = (2t+2)/3$ gives

$$\leq 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{(2t+2)/3-1} = 2\mu^2 \left(\frac{1}{1+\epsilon}\right)^{(2t-1)/3} .$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$.

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\Pr[\text{unsuccessful} \mid \text{no cycle}]$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\begin{aligned} &\Pr[\text{unsuccessful} \mid \text{no cycle}] \\ &\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \end{aligned}$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1]
\end{aligned}$$

Cuckoo Hashing

We choose $\mathrm{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$
\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad\leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\mathrm{maxsteps}+2}{3}] \\
&\quad\leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1] \\
&\quad\leq \Pr[\exists \text{ active path-structure of size exactly } \ell + 1]
\end{aligned}
$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$
\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size exactly } \ell + 1] \\
&\quad \leq 2\mu^2 \Big(\frac{1}{1+\epsilon}\Big)^{\ell}
\end{aligned}
$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size exactly } \ell + 1] \\
&\quad \leq 2\mu^2 \Big(\frac{1}{1+\epsilon}\Big)^{\ell} \leq \frac{1}{m^2}
\end{aligned}$$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size exactly } \ell + 1] \\
&\quad \leq 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{\ell} \leq \frac{1}{m^2}
\end{aligned}$$

by choosing $\ell \geq \log\left(\frac{1}{2\mu^2 m^2}\right)/\log\left(\frac{1}{1+\epsilon}\right) = \log\left(2\mu^2 m^2\right)/\log\left(1+\epsilon\right)$

Cuckoo Hashing

We choose $\text{maxsteps} \geq 3\ell/2 + 1/2$. Then the probability that a phase terminates unsuccessfully without running into a cycle is at most

$$\begin{aligned}
&\Pr[\text{unsuccessful} \mid \text{no cycle}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \tfrac{2\text{maxsteps}+2}{3}] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size at least } \ell + 1] \\
&\quad \leq \Pr[\exists \text{ active path-structure of size exactly } \ell + 1] \\
&\quad \leq 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{\ell} \leq \frac{1}{m^2}
\end{aligned}$$

by choosing $\ell \geq \log\big(\frac{1}{2\mu^2 m^2}\big)/\log\big(\frac{1}{1+\epsilon}\big) = \log\big(2\mu^2 m^2\big)/\log\big(1+\epsilon\big)$

This gives $\text{maxsteps} = \Theta(\log m)$.

Cuckoo Hashing

So far we estimated

$$\Pr[\text{cycle}] \le \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

and

$$\Pr[\text{unsuccessful} \mid \text{no cycle}] \le \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

Cuckoo Hashing

So far we estimated

$$\Pr[\text{cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

and

$$\Pr[\text{unsuccessful} \mid \text{no cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

Observe that

$$\Pr[\text{successful}] = \Pr[\text{no cycle}] - \Pr[\text{unsuccessful} \mid \text{no cycle}]$$

Cuckoo Hashing

So far we estimated

$$\Pr[\text{cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

and

$$\Pr[\text{unsuccessful} \mid \text{no cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

Observe that

$$\begin{aligned}\Pr[\text{successful}] &= \Pr[\text{no cycle}] - \Pr[\text{unsuccessful} \mid \text{no cycle}] \\ &\geq c \cdot \Pr[\text{no cycle}]\end{aligned}$$

Cuckoo Hashing

So far we estimated

$$\Pr[\text{cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

and

$$\Pr[\text{unsuccessful} \mid \text{no cycle}] \leq \mathcal{O}\Big(\frac{1}{m^2}\Big)$$

Observe that

$$\begin{aligned}\Pr[\text{successful}] &= \Pr[\text{no cycle}] - \Pr[\text{unsuccessful} \mid \text{no cycle}] \\ &\geq c \cdot \Pr[\text{no cycle}]\end{aligned}$$

for a suitable constant $c > 0$.

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$

Cuckoo Hashing

$E[X] = \sum_t Pr[X \geq t]$

The expected number of complete steps in the successful phase of an insert operation is:

$$E[\text{number of steps} \mid \text{phase successful}]$$

$$= \sum_{t \geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}]$$

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$
$$= \sum_{t\geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}]$$

We have

$$\Pr[\text{search at least } t \text{ steps} \mid \text{successful}]$$

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\begin{aligned} &\mathrm{E}[\text{number of steps} \mid \text{phase successful}] \\ &\quad = \sum_{t \geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}] \end{aligned}$$

We have

$$\begin{aligned} &\Pr[\text{search at least } t \text{ steps} \mid \text{successful}] \\ &\quad = \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{successful}] \end{aligned}$$

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\begin{aligned}
&\mathrm{E}[\text{number of steps} \mid \text{phase successful}] \\
&\qquad = \sum_{t\geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}]
\end{aligned}$$

We have

$$\begin{aligned}
&\Pr[\text{search at least } t \text{ steps} \mid \text{successful}] \\
&\qquad = \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{successful}] \\
&\qquad \leq \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{no cycle}]
\end{aligned}$$

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\begin{aligned}
&\mathrm{E}[\text{number of steps} \mid \text{phase successful}] \\
&\quad = \sum_{t\geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}]
\end{aligned}$$

We have

$$\begin{aligned}
&\Pr[\text{search at least } t \text{ steps} \mid \text{successful}] \\
&\quad = \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{successful}] \\
&\quad \leq \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{no cycle}] \\
&\quad \leq \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \wedge \text{no cycle}] / \Pr[\text{no cycle}]
\end{aligned}$$

Cuckoo Hashing

The expected number of complete steps in the successful phase of an insert operation is:

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$
$$= \sum_{t \geq 1} \Pr[\text{search takes at least } t \text{ steps} \mid \text{phase successful}]$$

We have

$$\Pr[\text{search at least } t \text{ steps} \mid \text{successful}]$$
$$= \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{successful}]$$
$$\leq \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \wedge \text{successful}] / \Pr[\text{no cycle}]$$
$$\leq \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \wedge \text{no cycle}] / \Pr[\text{no cycle}]$$
$$= \frac{1}{c} \Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}] \; .$$

Cuckoo Hashing

Hence,

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$

Cuckoo Hashing

Hence,

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$

$$\leq \frac{1}{c} \sum_{t \geq 1} \Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}]$$

Cuckoo Hashing

Hence,

$$\begin{aligned}
&\mathrm{E}[\text{number of steps} \mid \text{phase successful}] \\
&\quad \le \frac{1}{c}\sum_{t\ge 1}\Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}] \\
&\quad \le \frac{1}{c}\sum_{t\ge 1} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2t-1)/3}
\end{aligned}$$

Cuckoo Hashing

Hence,

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$

$$\leq \frac{1}{c}\sum_{t\geq 1}\Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}]$$

$$\leq \frac{1}{c}\sum_{t\geq 1} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2t-1)/3} = \frac{1}{c}\sum_{t\geq 0} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2(t+1)-1)/3}$$

Cuckoo Hashing

Hence,

$$
\begin{aligned}
&\mathrm{E}[\text{number of steps} \mid \text{phase successful}]\\
&\leq \frac{1}{c}\sum_{t\geq 1}\Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}]\\
&\leq \frac{1}{c}\sum_{t\geq 1} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2t-1)/3} = \frac{1}{c}\sum_{t\geq 0} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2(t+1)-1)/3}\\
&= \frac{2\mu^2}{c(1+\epsilon)^{1/3}}\sum_{t\geq 0}\Big(\frac{1}{(1+\epsilon)^{2/3}}\Big)^{t}
\end{aligned}
$$

Cuckoo Hashing

Hence,

$$\mathrm{E}[\text{number of steps} \mid \text{phase successful}]$$

$$\leq \frac{1}{c}\sum_{t\geq 1}\Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}]$$

$$\leq \frac{1}{c}\sum_{t\geq 1} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2t-1)/3} = \frac{1}{c}\sum_{t\geq 0} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2(t+1)-1)/3}$$

$$= \frac{2\mu^2}{c(1+\epsilon)^{1/3}}\sum_{t\geq 0}\Big(\frac{1}{(1+\epsilon)^{2/3}}\Big)^t = \mathcal{O}(1)\ .$$

Cuckoo Hashing

$$\sum_{i\ge 0} q^i = \frac{1}{1-q}$$

Hence,

$$\begin{aligned}
&\mathrm{E}[\text{number of steps} \mid \text{phase successful}] \\
&\le \frac{1}{c}\sum_{t\ge 1}\Pr[\text{search at least } t \text{ steps} \mid \text{no cycle}] \\
&\le \frac{1}{c}\sum_{t\ge 1} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2t-1)/3} = \frac{1}{c}\sum_{t\ge 0} 2\mu^2\Big(\frac{1}{1+\epsilon}\Big)^{(2(t+1)-1)/3} \\
&= \frac{2\mu^2}{c(1+\epsilon)^{1/3}}\sum_{t\ge 0}\Big(\frac{1}{(1+\epsilon)^{2/3}}\Big)^t = \mathcal{O}(1)\ .
\end{aligned}$$

This means the expected cost for a successful phase is constant (even after accounting for the cost of the incomplete step that finishes the phase).

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching $\mathrm{maxsteps}$ without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\geq 1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching $\mathrm{maxsteps}$ without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\geq 1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching $\mathrm{maxsteps}$ without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\geq 1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching maxsteps without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\geq 1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching $\mathrm{maxsteps}$ without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\ge1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Cuckoo Hashing

A phase that is not successful induces cost for doing a complete rehash (this dominates the cost for the steps in the phase).

The probability that a phase is not successful is $q = \mathcal{O}(1/m^2)$ (probability $\mathcal{O}(1/m^2)$ of running into a cycle and probability $\mathcal{O}(1/m^2)$ of reaching $\mathrm{maxsteps}$ without running into a cycle).

A rehash try requires m insertions and takes expected constant time per insertion. It fails with probability $p := \mathcal{O}(1/m)$.

The expected number of unsuccessful rehashes is $\sum_{i\geq 1} p^i = \frac{1}{1-p} - 1 = \frac{p}{1-p} = \mathcal{O}(p)$.

Therefore the expected cost for re-hashes is $\mathcal{O}(m) \cdot \mathcal{O}(p) = \mathcal{O}(1)$.

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m + 1$ insertions fails):

Formal Proof

Let Y_i denote the event that the i-th rehash ~~does not lead to a valid configuration (assuming i-th rehash occurs)~~ (i.e., one of the $m+1$ insertions fails):

occurs and

$$\Pr[Y_i \mid Z_i] \quad \Pr[Y_i] \le (m+1) \cdot \mathcal{O}(1/m^2) \le \mathcal{O}(1/m) =: p \ .$$

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m+1$ insertions fails):

$Y_i \mid Z_i$

$$\Pr[Y_i] \leq (m+1) \cdot \mathcal{O}(1/m^2) \leq \mathcal{O}(1/m) =: p \;.$$

Let Z_i denote the event that the i-th rehash occurs:

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m+1$ insertions fails):

$$\Pr[Y_i] \leq (m+1) \cdot \mathcal{O}(1/m^2) \leq \mathcal{O}(1/m) =: p \ .$$

Let Z_i denote the event that the i-th rehash occurs:

$$\Pr[Z_i] \leq \Pr[\wedge_{j=0}^{i-1} Y_j] \leq p^i$$

$$\leq \prod \Pr[Y_j \mid Z_j] \leq p^i$$

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m+1$ insertions fails):

$$\Pr[Y_i] \le (m+1)\cdot \mathcal{O}(1/m^2) \le \mathcal{O}(1/m) =: p \ .$$

Let Z_i denote the event that the i-th rehash occurs:

$$\Pr[Z_i] \le \Pr[\wedge_{j=0}^{i-1} Y_j] \le p^i$$

Let X_i^s, $s \in \{1,\dots,m+1\}$ denote the cost for inserting the s-th element during the i-th rehash (assuming i-th rehash occurs):

$$\mathrm{E}[X_i^s]$$

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m+1$ insertions fails):

$$\Pr[Y_i] \leq (m+1) \cdot \mathcal{O}(1/m^2) \leq \mathcal{O}(1/m) =: p \ .$$

Let Z_i denote the event that the i-th rehash occurs:

$$\Pr[Z_i] \leq \Pr[\wedge_{j=0}^{i-1} Y_j] \leq p^i$$

Let X_i^s, $s \in \{1, \ldots, m+1\}$ denote the cost for inserting the s-th element during the i-th rehash (assuming i-th rehash occurs):

$$\begin{aligned} \mathrm{E}[X_i^s] = \mathrm{E}[\text{steps} \mid \text{phase successful}] \cdot \Pr[\text{phase sucessful}] \\ + \text{maxsteps} \cdot \Pr[\text{not sucessful}] \end{aligned}$$

Formal Proof

Let Y_i denote the event that the i-th rehash does not lead to a valid configuration (assuming i-th rehash occurs) (i.e., one of the $m+1$ insertions fails):

$$\Pr[Y_i] \leq (m+1) \cdot \mathcal{O}(1/m^2) \leq \mathcal{O}(1/m) =: p \ .$$

Let Z_i denote the event that the i-th rehash occurs:

$$\Pr[Z_i] \leq \Pr[\wedge_{j=0}^{i-1} Y_j] \leq p^i$$

Let X_i^s, $s \in \{1, \ldots, m+1\}$ denote the cost for inserting the s-th element during the i-th rehash (assuming i-th rehash occurs):

$$\begin{aligned} \mathrm{E}[X_i^s] &= \mathrm{E}[\text{steps} \mid \text{phase successful}] \cdot \Pr[\text{phase sucessful}] \\ &\quad + \text{maxsteps} \cdot \Pr[\text{not sucessful}] = \mathcal{O}(1) \ . \end{aligned}$$

The expected cost for all rehashes is

$$\mathrm{E}\left[\sum_i \sum_s Z_i X_i^s\right]$$

Note that Z_i is independent of X_j^s, $j \geq i$ (however, it is not independent of X_j^s, $j < i$). Hence,

$\frac{1}{m^2}$ $\ll$ p

Z_i fails

probability of rehash failing is p

$\Pr[Z_i = 1] \leq p^i$

The expected cost for all rehashes is

$$\mathrm{E}\left[\sum_i \sum_s Z_i X_i^s\right]$$

Note that Z_i is independent of X_j^s, $j \geq i$ (however, it is not independent of X_j^s, $j < i$). Hence,

$$\begin{aligned}
\mathrm{E}\left[\sum_i \sum_s Z_i X_s^i\right] &= \sum_i \sum_s \mathrm{E}[Z_i] \cdot \mathrm{E}[X_s^i] \\
&\leq \mathcal{O}(m) \cdot \sum_i p^i \\
&\leq \mathcal{O}(m) \cdot \frac{p}{1-p} \\
&= \mathcal{O}(1) \ .
\end{aligned}$$

The expected cost for all rehashes is

$$\mathrm{E}\Big[\sum_i \sum_s Z_i X_i^s\Big]$$

Note that Z_i is independent of X_j^s, $j \geq i$ (however, it is not independent of X_j^s, $j < i$). Hence,

$$\begin{aligned}
\mathrm{E}\Big[\sum_i \sum_s Z_i X_s^i\Big] &= \sum_i \sum_s \mathrm{E}[Z_i] \cdot \mathrm{E}[X_s^i] \\
&\leq \mathcal{O}(m) \cdot \sum_i p^i \\
&\leq \mathcal{O}(m) \cdot \frac{p}{1-p} \\
&= \mathcal{O}(1) \enspace .
\end{aligned}$$

The expected cost for all rehashes is

$$\mathrm{E}\left[\sum_i \sum_s Z_i X_i^s\right]$$

Note that Z_i is independent of X_j^s, $j \geq i$ (however, it is not independent of X_j^s, $j < i$). Hence,

$$\begin{aligned}
\mathrm{E}\left[\sum_i \sum_s Z_i X_s^i\right] &= \sum_i \sum_s \mathrm{E}[Z_i] \cdot \mathrm{E}[X_s^i] \\
&\leq \mathcal{O}(m) \cdot \sum_i p^i \\
&\leq \mathcal{O}(m) \cdot \frac{p}{1-p} \\
&= \mathcal{O}(1)\ .
\end{aligned}$$

The expected cost for all rehashes is

$$\mathrm{E}\left[\sum_i \sum_s Z_i X_i^s\right]$$

Note that Z_i is independent of X_j^s, $j \geq i$ (however, it is not independent of X_j^s, $j < i$). Hence,

$$\begin{aligned}\mathrm{E}\left[\sum_i \sum_s Z_i X_s^i\right] &= \sum_i \sum_s \mathrm{E}[Z_i] \cdot \mathrm{E}[X_s^i] \\ &\leq \mathcal{O}(m) \cdot \sum_i p^i \\ &\leq \mathcal{O}(m) \cdot \frac{p}{1-p} \\ &= \mathcal{O}(1) \enspace .\end{aligned}$$

Cuckoo Hashing

What kind of hash-functions do we need?

Since $\mathrm{maxsteps}$ is $\Theta(\log m)$ the largest size of a path-structure or cycle-structure contains just $\Theta(\log m)$ different keys.

Therefore, it is sufficient to have $(\mu, \Theta(\log m))$-independent hash-functions.

Cuckoo Hashing

What kind of hash-functions do we need?

Since maxsteps is $\Theta(\log m)$ the largest size of a path-structure or cycle-structure contains just $\Theta(\log m)$ different keys.

Therefore, it is sufficient to have $(\mu, \Theta(\log m))$-independent hash-functions.

Cuckoo Hashing

What kind of hash-functions do we need?

Since $\mathrm{maxsteps}$ is $\Theta(\log m)$ the largest size of a path-structure or cycle-structure contains just $\Theta(\log m)$ different keys.

Therefore, it is sufficient to have $(\mu, \Theta(\log m))$-independent hash-functions.

Cuckoo Hashing

How do we make sure that $n \geq (1+\epsilon)m$?

- Let $\alpha := 1/(1+\epsilon)$.
- Keep track of the number of elements in the table. When $m \geq \alpha n$ we double n and do a complete re-hash (table-expand).
- Whenever m drops below $\alpha n/4$ we divide n by 2 and do a rehash (table-shrink).
- Note that right after a change in table-size we have $m = \alpha n/2$. In order for a table-expand to occur at least $\alpha n/2$ insertions are required. Similar, for a table-shrink at least $\alpha n/4$ deletions must occur.
- Therefore we can amortize the rehash cost after a change in table-size against the cost for insertions and deletions.

Cuckoo Hashing

How do we make sure that $n \geq (1+\epsilon)m$?

- Let $\alpha := 1/(1+\epsilon)$.
- Keep track of the number of elements in the table. When $m \geq \alpha n$ we double n and do a complete re-hash (table-expand).
- Whenever m drops below $\alpha n/4$ we divide n by 2 and do a rehash (table-shrink).
- Note that right after a change in table-size we have $m = \alpha n/2$. In order for a table-expand to occur at least $\alpha n/2$ insertions are required. Similar, for a table-shrink at least $\alpha n/4$ deletions must occur.
- Therefore we can amortize the rehash cost after a change in table-size against the cost for insertions and deletions.

Cuckoo Hashing

$m = \alpha n \rightarrow \boxed{m = \alpha \frac{n}{2}}$

How do we make sure that $n \geq (1+\epsilon)m$?

- Let $\alpha := 1/(1+\epsilon)$.
- Keep track of the number of elements in the table. When $m \geq \alpha n$ we double n and do a complete re-hash (table-expand).
- Whenever m drops below $\alpha n/4$ we divide n by 2 and do a rehash (table-shrink).
- Note that right after a change in table-size we have $m = \alpha n/2$. In order for a table-expand to occur at least $\alpha n/2$ insertions are required. Similar, for a table-shrink at least $\alpha n/4$ deletions must occur.
- Therefore we can amortize the rehash cost after a change in table-size against the cost for insertions and deletions.

Cuckoo Hashing

How do we make sure that $n \geq (1+\epsilon)m$?

- Let $\alpha := 1/(1+\epsilon)$.
- Keep track of the number of elements in the table. When $m \geq \alpha n$ we double n and do a complete re-hash (table-expand).
- Whenever m drops below $\alpha n/4$ we divide n by 2 and do a rehash (table-shrink).
- Note that right after a change in table-size we have $m = \alpha n/2$. In order for a table-expand to occur at least $\alpha n/2$ insertions are required. Similar, for a table-shrink at least $\alpha n/4$ deletions must occur.
- Therefore we can amortize the rehash cost after a change in table-size against the cost for insertions and deletions.

Cuckoo Hashing

How do we make sure that $n \geq (1+\epsilon)m$?

- Let $\alpha := 1/(1+\epsilon)$.
- Keep track of the number of elements in the table. When $m \geq \alpha n$ we double n and do a complete re-hash (table-expand).
- Whenever m drops below $\alpha n/4$ we divide n by 2 and do a rehash (table-shrink).
- Note that right after a change in table-size we have $m = \alpha n/2$. In order for a table-expand to occur at least $\alpha n/2$ insertions are required. Similar, for a table-shrink at least $\alpha n/4$ deletions must occur.
- Therefore we can amortize the rehash cost after a change in table-size against the cost for insertions and deletions.

Cuckoo Hashing

Lemma 31

Cuckoo Hashing has an expected constant insert-time and a worst-case constant search-time.

Note that the above lemma only holds if the fill-factor (number of keys/total number of hash-table slots) is at most $\frac{1}{2(1+\epsilon)}$.

Cuckoo Hashing

Lemma 31
Cuckoo Hashing has an expected constant insert-time and a worst-case constant search-time.

Note that the above lemma only holds if the fill-factor (number of keys/total number of hash-table slots) is at most $\frac{1}{2(1+\epsilon)}$.