Universelles Hashing

Definition

Es sei $\mathcal{H}$ eine nicht-leere Menge von Funktionen $U \to \{1, \ldots, m\}$.

Wir sagen, daß $\mathcal{H}$ eine universelle Familie von Hashfunktionen ist, wenn für jedes $x, y \in U$, $x \neq y$ folgendes gilt:

$$\frac{|\{\, h \in \mathcal{H} \mid h(x) = h(y) \,\}|}{|\mathcal{H}|} \leq \frac{1}{m}$$

Theorem

Es sei $\mathcal{H}$ eine universelle Familie von Hashfunktionen $U \to \{1, \ldots, m\}$ für das Universum U und $S \subseteq U$ eine beliebige Untermenge.

Wenn $x \in U$, $x \notin S$ und $h \in \mathcal{H}$ eine zufällig gewählte Hashfunktion ist, dann gilt

$$E\Big(|\{\, y \in S \mid h(x) = h(y) \,\}|\Big) \leq \frac{|S|}{m}.$$

Beweis

$$E\Big(|\{\, y \in S \mid h(x) = h(y) \,\}|\Big) =$$

$$\sum_{y \in S} \Pr[h(x) = h(y)] = \sum_{y \in S} \frac{|\{\, h \in \mathcal{H} \mid h(x) = h(y) \,\}|}{|\mathcal{H}|} \leq \frac{|S|}{m}$$

Satz

Sei $x \in U$ beliebig, $\mathcal{H}$ eine universelle Familie von Hashfunktionen $U \to \{1, \ldots, m\}$ und k eine beliebige Zahl aus $\{1, \ldots, m\}$.

Dann gilt $\Pr[h(x) = k] = 1/m$, falls h zufällig aus $\mathcal{H}$.

Beweis

Nehmen wir an, es gibt ein y mit $h(y) = k$. Dann setze $S = \{y\}$ und wende das letzte Theorem an:

$$E\Big(|\{\, y \in S \mid h(x) = h(y)\,\}|\Big) \leq \frac{|S|}{m}.$$

Hier folgt daraus $\Pr[h(x) = k] \leq 1/m$.

Wenn es kein y mit $h(y) = k$ gäbe, dann wäre $\Pr[h(x) = k] = 0$.

Das geht aber nicht, denn $\sum_{k=1}^{m} \Pr[h(x) = k] = 1$.

Eine universelle Hashfamilie

Sei $U = \{0, \ldots, p-1\}$, wobei p eine Primzahl ist.

Es sei $h_{a,b}(x) = ((ax + b) \bmod p) \bmod m$.

Wir definieren

$$\mathcal{H} = \{\, h_{a,b} \mid 1 \leq a < p,\ 0 \leq b < p\}$$

Theorem

$\mathcal{H}$ ist eine universelle Familie von Hashfunktionen.

Es seien $x, y \in \{0, \ldots, p-1\}$, $x \neq y$.

Wir wollen zunächst zeigen, daß die Funktion

$$f : (a, b) \mapsto (ax + b \bmod p, ay + b \bmod p)$$

für $a, b \in \{0, \ldots, p-1\}$ injektiv und somit auch bijektiv ist.

$$\begin{aligned}
&(ax + b \bmod p, ay + b \bmod p) = (a'x + b' \bmod p, a'y + b' \bmod p) \\
\Leftrightarrow &(ax + b - b' \bmod p, ay + b - b' \bmod p) = (a'x \bmod p, a'y \bmod p) \\
\Leftrightarrow &(b - b' \bmod p, b - b' \bmod p) = ((a' - a)x \bmod p, (a' - a)y \bmod p) \\
\Leftrightarrow &(a' - a)x \bmod p = (a' - a)y \bmod p \Leftrightarrow a' = a \wedge b' = b
\end{aligned}$$

Nach wie vor gelte $x, y \in \{0, \ldots, p-1\}$, $x \neq y$.

Für wieviele Paare (a, b) haben $c_x := ax + b \bmod p$ und $c_y := ay + b \bmod p$ den gleichen Rest modulo m?

Wir haben auf der letzten Folie bewiesen, daß sich für jedes Paar (a, b) ein eindeutiges Paar (c_x, c_y) ergibt. Für ein festes c_x gibt es nur

$$\lceil p/m \rceil - 1 = \left\lfloor \frac{p+m-1}{m} \right\rfloor - 1 \leq \frac{p-1}{m}$$

viele mögliche Werte von c_y mit $c_x \equiv c_y \mod m$ und $c_x \neq c_y$.

Weil p verschiedene Werte für c_x existieren, gibt es insgesamt höchstens $p(p-1)/m$ Paare der gesuchten Art.

$$\frac{|\{\, h \in \mathcal{H} \mid h(x) = h(y) \,\}|}{|\mathcal{H}|} \leq \frac{p(p-1)/m}{p(p-1)} \leq \frac{1}{m}$$

Min-Cut

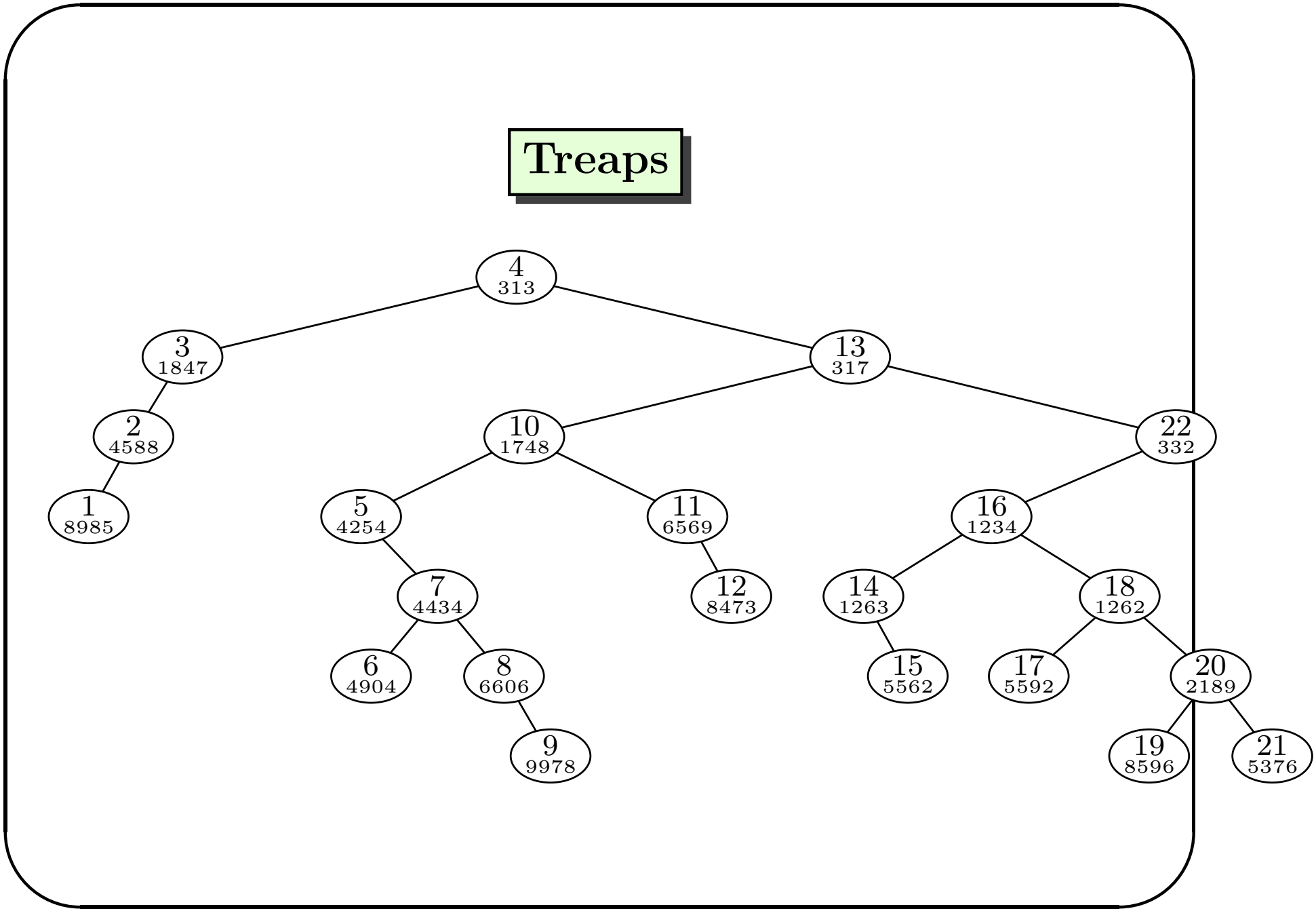
Einfacher Algorithmus:

Kontrahiere zufällige Kanten, bis nur zwei Knoten übrig.

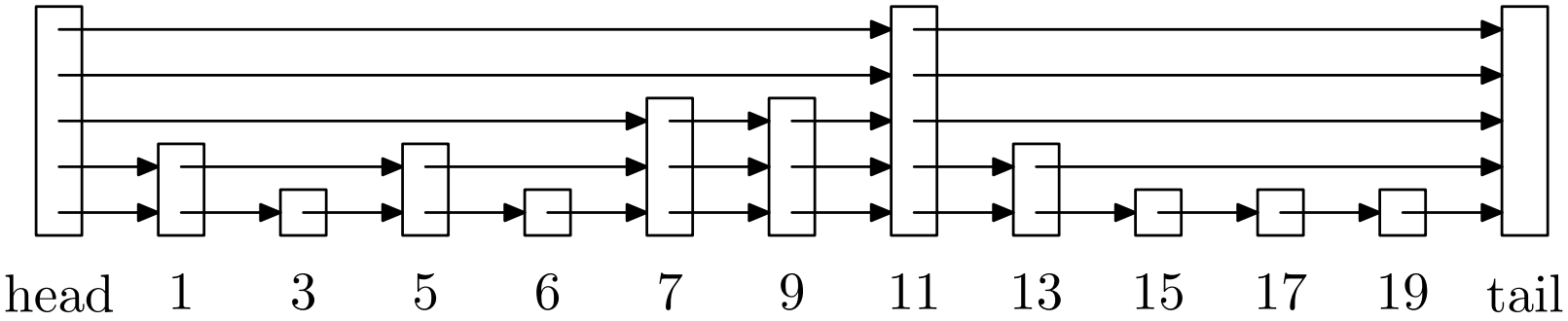
Mit Wahrscheinlichkeit $\Omega(n^{-2})$ ein Min-Cut.

Verwende Amplifizierung.

Verbesserung: Zwei Kontraktionssequenzen bis etwa $n/\sqrt{2}$ Knoten bleiben, dann rekursiv.


Treaps
4
313
3
1847
13
317
2
4588
10
1748
22
332
1
8985
5
4254
11
6569
16
1234
7
4434
12
8473
14
1263
18
1262
6
4904
8
6606
15
5562
17
5592
20
2189
9
9978
19
8596
21
5376

Skip-Lists



ENDE