Hash-Funktionen

Szenario:

Wir wollen Elemente eines Universums $U = \{0, 1, \dots, m-1\}$ (wobei m groß ist) auf ein Universum $V = \{0, 1, \dots, M-1\}$ abbilden (je nach Anwendungsszenario ist M manchmal klein und manchmal groß).

Am liebsten würden wir eine Abbildung

$$h : U \to V$$

komplett zufällig wählen, indem wir unabhängig und gleichverteilt für jedes $u \in U$ einen Wert $v_u \in V$ zufällig wählen und

$$h(u) := v_u \quad \text{für alle } u \in U$$

setzen.

$(*)_2$ Für jeden festen Wert $z \in V$ und alle festen Werte $u_1, u_2 \in U$ mit $u_1 \neq u_2$ gilt dann

$$\Pr_{h \text{ zufällig gewählt}} \left(h(u_1) = z \text{ und } h(u_2) = z \right) = \frac{1}{M^2}$$

und es gilt

$$\textcircled{\scriptsize$\Delta$}_2 \left\{ \Pr_{\substack{h \text{ zufällig} \\ \text{gewählt}}} \Big(h(u_1) = h(u_2) \Big) \leq \frac{1}{M} . \right.$$

Allgemein gilt für jede natürliche Zahl $k \geq 2$

$\circledast_k$ und alle paarweise verschiedenen $u_1, \dots, u_k \in U$
und alle $z \in V$, dass

$$\Pr_{\substack{h \text{ zufällig} \\ \text{gewählt}}} \Big(\bigwedge_{i=1}^{k} h(u_i) = z \Big) = \frac{1}{M^k}$$

und außerdem gilt

$$\textcircled{\scriptsize$\Delta$}_k \left\{ \Pr_{\substack{h \text{ zufällig} \\ \text{gewählt}}} \Big(h(u_1) = \dots = h(u_k) \Big) \leq \frac{1}{M^{k-1}} . \right.$$

<u>Problem:</u>

Eine komplett zufällige Funktion $h: U \to V$ zu erzeugen, kostet $\Theta(m \cdot \log_2 M)$ Zufallsbits; und die Funktion h zu speichern kostet $\Theta(m \cdot \log_2 M)$ Speicherbits.

Das ist uns viel zu teuer!

Glücklicherweise werden für viele Anwendungsszenarien gar keine "komplett zufälligen" Hash-Funktionen benötigt und es reicht aus, Funktionen zu betrachten, die für eine geeignete Zahl $k \geq 2$ die Eigenschaft $\textcircled{\Delta}_k$ oder $\textcircled{*}_k$ besitzen — und in den meisten Fällen genügt "$k=2$".

Statt h "komplett zufällig" zu wählen, wählen wir dann zufällig und gleichverteilt $h \in H$, wobei H eine geeignete Menge von Funktionen von V nach U ist. Was mit "geeignet" gemeint ist, wird in der folgenden Definition präzisiert.

<u>Definition</u>

Seien $m, M \in \mathbb{N}_{\geq 1}$, sei $U := \{0, \dots, m-1\}$ und $V := \{0, \dots, M-1\}$.

<u>(a)</u> Eine <u>Familie von Hash-Funktionen von U nach V</u> ist eine Menge H von Funktionen von U nach V.

<u>(b)</u> Sei H eine Familie von Hash-Funktionen von U nach V und sei $k \in \mathbb{N}$ mit $k \geq 2$.

<u>(i)</u> H heißt <u>k-universell</u>, wenn für alle paarweise verschiedenen $u_1, \dots, u_k \in U$ gilt:

$$(\triangle)_k \quad \Big\{ \quad \underbrace{\Pr_{h \in H}}_{} \big(h(u_1) = \dots = h(u_k) \big) \le \frac{1}{M^{k-1}}$$

Dabei wird h zufällig und gleichverteilt aus H gewählt.

<u>(ii)</u> H heißt <u>streng k-universell</u>, wenn für alle paarweise verschiedenen $u_1, \dots, u_k \in U$ und alle $z_1, \dots, z_k \in V$ gilt:

$$(*)_k \quad \Big\{ \quad \Pr_{h \in H} \Big(\bigwedge_{i=1}^{k} h(u_i) = z_i \Big) = \frac{1}{M^k}.$$

<u>(c)</u> Eine Familie H von Hash-Funktionen von U nach V wird <u>paarweise unabhängig</u> genannt, wenn sie <u>streng 2-universell</u> ist.

<u>Beachte</u>: Wenn H streng k-universell ist, so ist H auch k-universell, denn
$\Pr\big(h(u_1) = \dots = h(u_k)\big) = \Pr\Big(\bigvee_{z \in V} \big(\bigwedge_{i=1}^{k} h(u_i) = z \big) \Big) \le \sum_{z \in V} \Pr\Big(\bigwedge_{i=1}^{k} h(u_i) = z \Big) = |V| \cdot \frac{1}{M^k} = \frac{1}{M^{k-1}}$.

<u>Bemerkung</u>:
In der Literatur werden <u>streng k-universelle</u> Familien manchmal auch <u>k-fach unabhängige</u> Familien genannt.

Und manchmal wird das, was wir hier als "streng k-universell" bezeichnen, einfach nur "k-universell" genannt.

Beispiele für (streng) k-universelle Familien von Hash-Funktionen:

Sei $U := \{0, \dots, m-1\}$ und $V := \{0, \dots, M-1\}$

(a) Sei $m \in \mathbb{N}_{\geq 1}$ und sei M eine Primzahl mit $M \geq m$.

Für alle $a, b \in V$ sei $h_{a,b}$ die Funktion

$$h_{a,b} : U \to V \quad \text{mit}$$

$$h_{a,b}(x) := a \cdot x + b \pmod{M}$$

f.a. $x \in U$.

Dann ist die Familie

$$H_1 := \{ h_{a,b} : a, b \in V \}$$

streng 2-universell.

Beweis: Übung.

(b) Seien $m, M \in \mathbb{N}_{\geq 1}$ mit $m \geq M$ und sei p eine Primzahl mit $p \geq m$.

Für alle $a, b \in \mathbb{N}$ sei $h_{a,b} : U \to V$ die Funktion mit

$$h_{a,b}(x) := \big(ax + b \pmod{p} \big) \pmod{M} \quad \text{f.a. } x \in U.$$

Dann ist die Familie

$$H_2 := \{ h_{a,b} : a, b \in \{0, \dots, p-1\} \text{ und } a \neq 0 \}$$

2-universell. Beweis: Übung.

(c) Seien $m, M \in \mathbb{N}_{\geq 1}$, so dass M eine Primzahl ist.

Sei $\ell \in \mathbb{N}$ so gewählt, dass $m \leq M^\ell$ ist.

Somit ist $U \subseteq \{0, \ldots, M^\ell - 1\}$.

Für ein Element $u \in U$ ist die Darstellung $\overline{u}$ zur Basis M das Tupel $\overline{u} = (u_{\ell-1}, \ldots, u_0) \in \{0, \ldots, M-1\}^\ell$, für das gilt:

$$u = \sum_{i=0}^{\ell-1} u_i M^i .$$

Für jedes $\overline{a} = (a_{\ell-1}, \ldots, a_0) \in V^\ell$ und jedes $b \in V$ sei $h_{\overline{a},b} : U \to V$ die Funktion mit

$$h_{\overline{a},b}(x) := \left(\sum_{i=0}^{\ell-1} a_i \cdot x_i + b \right) \pmod{M}$$

für jedes $x \in U$ und dessen Darstellung $\overline{x} = (x_{\ell-1}, \ldots, x_0) \in V^\ell$ zur Basis M.

Dann ist die Familie

$$H_{s2} := \{ h_{\overline{a},b} : \overline{a} \in V^\ell, b \in V \}$$

<u>streng 2-universell</u>. <u>Beweis</u>: Übung.

(d) Sei p eine Primzahl, sei $m \in \mathbb{N}_{\geq 1}$ und sei $\ell \in \mathbb{N}_{\geq 1}$ so, dass $m \leq p^{\ell}$.

Setze $q := p^{\ell}$ und sei $\mathbb{F}_q$ der Körper mit (genau) q Elementen.

Der Einfachheit halber gehen wir davon aus, dass $U \subseteq \mathbb{F}_q$ ist.

Sei g eine (fest gewählte) Bijektion von $\mathbb{F}_q$ auf $\{0, 1, \ldots, q-1\}$.

Für $k \in \mathbb{N}$ mit $k \geq 2$ und $\bar{a} = (a_{k-1}, \ldots, a_0) \in \mathbb{F}_q^k$ sei $f_{\bar{a}} : U \to \mathbb{F}_q$ die Funktion mit

$$f_{\bar{a}}(x) := g\left(\sum_{i=0}^{k-1} a_i x^i \right) \quad \text{f.a. } x \in U.$$

Dann ist die Familie

$$H_q^k := \{ f_{\bar{a}} : \bar{a} \in \mathbb{F}_q^k \}$$

streng k-universell (hier ohne Beweis).

Sei $\ell' \in \mathbb{N}_{\geq 1}$ mit $\ell' \leq \ell$ und setze $M := p^{\ell'}$.
Für jedes $\bar{a} = (a_{k-1}, \ldots, a_0) \in \mathbb{F}_q^k$ sei $h_{\bar{a}} : U \to V$ die Funktion mit $h_{\bar{a}}(x) := f_{\bar{a}}(x) \pmod{M}$ f.a. $x \in U$.
Dann ist die Familie

$$H_{q,M}^k := \{ h_{\bar{a}} : \bar{a} \in \mathbb{F}_q^k \}$$

streng k-universell (hier ohne Beweis).