$\mathcal{H}: X \longrightarrow Y$; silnie K-uniwersalna (*)

dla dowolnych $x_1, \dots, x_k \in X$ parami różnych

i dowolnych $y_1, \dots, y_k \in Y$ mamy

$$Pr_h [h(x_1) = y_1 \wedge \dots \wedge h(x_k) = y_k] = \frac{1}{|Y|^k}.$$

Terminologia: (*) = k-uniwersalna rodzina

KONSTRUKCJA 2-wiązek.

$X = Y = \mathbb{Z}_p$; p - pierwsza

$a, b \in \mathbb{Z}_p$: $\varphi_{a,b}(x) = ax + b$.

$\mathcal{H} = \{\varphi_{a,b} : a, b \in \mathbb{Z}_p\}$ $\quad |\mathcal{H}| = p^2$

FAKT 2-silnie uniwersalna.

$x_1, x_2 \in \mathbb{Z}_p$

$y_1, y_2 \in \mathbb{Z}_p$

$x_1 \neq x_2$

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \end{cases}$$

zmienne: a, b

$\exists!$ rozw.

$a(x_1 - x_2) = y_1 - y_2 \quad : \quad a = (y_1 - y_2) \cdot (x_1 - x_2)^{-1}$.

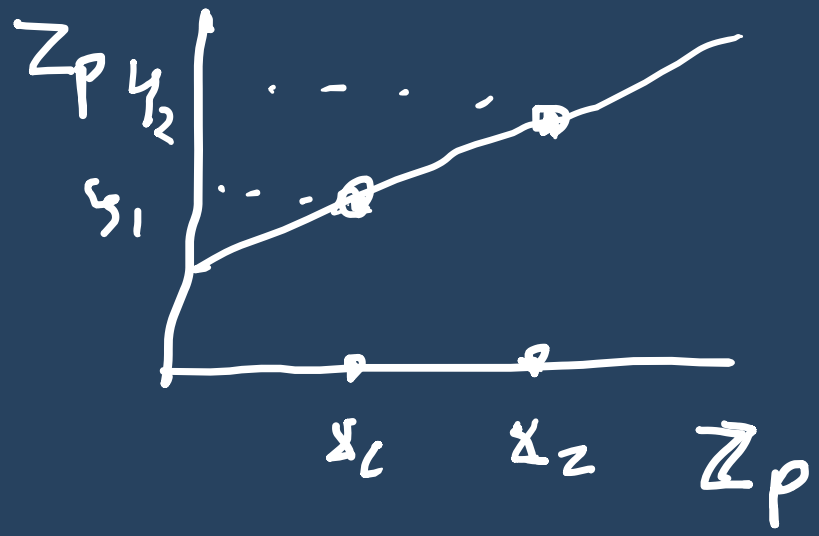
$|\{\varphi_{a,b} : \varphi(x_1) = y_1 \wedge \varphi(x_2) = y_2\}| = 1$ $\quad b = \dots$.

Konstr. k - silnie uniwers.

$$\varphi_{a_0 \dots a_{k-1}}(x) = a_0 + a_1 x + \dots + a_{k-1} x^{k-1} \in \mathbb{Z}_p[x]$$

$$\mathcal{H} = \{\varphi_{\bar{a}} : \bar{a} \in \mathbb{Z}_p^k\} \qquad |\mathcal{H}| = p^k$$

$$\begin{cases} \varphi_{\bar{a}}(x_1) = y_1 \\ \varphi_{\bar{a}}(x_2) = y_2 \\ \vdots \\ \varphi_{\bar{a}}(x_k) = y_k \end{cases}$$



wzór interp. Lagrange'a

$\mathcal{H} \subseteq \mathbb{Z}_p^{\mathbb{Z}_p}$ $\qquad \varphi : \mathbb{Z}_p \longrightarrow \mathbb{Z}_p \longrightarrow \mathbb{Z}_n$

$\varphi_a^*(x) = \varphi_a(x) \mod n.$

Problem :

Recepta na silne uniwersalności:

- weź ciało F o 2^m elementach, 2^m - wielkie
- weź podciało $K \subseteq F$ o 2^n elementach
  $K = GF(2^n)$ gdzie $n < m$

$$\varphi_a^*(x) = \left(\sum_{i=0}^{k-1} a_i x^{2^i} \; [GF(2^n)] \right) \bmod GF(2^n)$$

(P) Zróbmy ciało $GF(2^3)$:

- $\mathbb{Z}_2[x]$ – pierścień.
- szukamy wielomianu $w(x) \in \mathbb{Z}_2[x]$ stopnia 3 nierozkładalnego

$$w(x) = x^3 + x + 1.$$

$w(0) = 1 \neq 0$

$w(1) = 1+1+1 = 1 \neq 0$

wprowadzamy nowy symbol $\dot{2}$:

$$\dot{2}^3 + \dot{2} + 1 = 0$$

czyli

$$\dot{2}^3 = \dot{2} + 1$$

$$\mathbb{Q}F(8) = \{a + b\iota + c\iota^2 : a, b, c \in \mathbb{Z}_2\}$$

$$\iota^3 = \iota + 1$$

$$(1 + \iota^3)^2 = 1^2 + 2 \cdot 1 \cdot \iota^3 + \iota^6 =$$

$$= 1 + \iota \cdot \iota^3 = 1 + \iota(\iota + 1) = 1 + \iota + \iota^2.$$

RAKSYKALNA NIEZAL.

Założenie: $\mathcal{H} = Y^X$ ← k-niezależna dla dowolnego

HEURYSTYKA: jeśli $\mathcal{H} \subseteq Y^{X^k}$ jest $\log_2(|Y|)$ - niezależne to zachowuje się jak $|X|$-niezależna

(P) $|Y| \approx 2^{32} \rightsquigarrow \log_2|Y| = 32$

Uwaga: $h \in \mathcal{H}$ dobrany od razu.

- $h(x)$ szybko się liczy $\sim O(1)$.
- h nie musi być tzw. funkcją kryptograficzną (jednokierunkową).

czyli znając $y (= h(x))$ bardzo trudno jest wyzn. x.

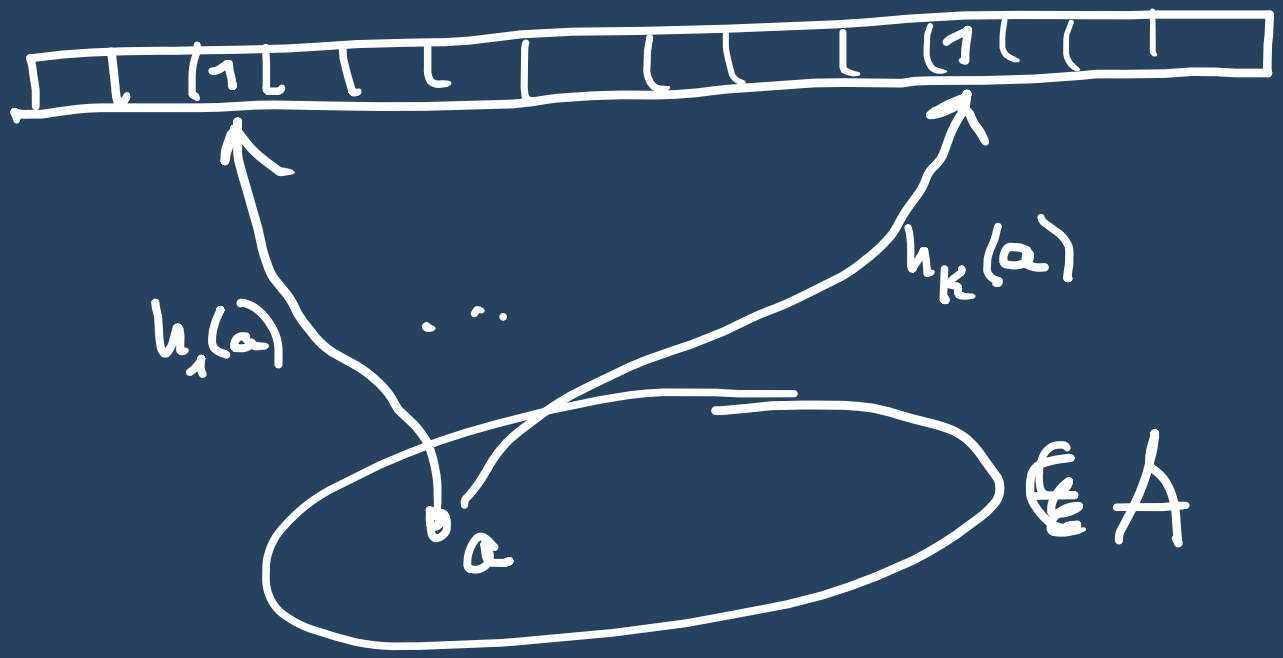
{MD5, SHA256

dobry wybór: MURMUR hash.

FILTRY BLOOMA

Q: mamy $\sigma \in \Sigma$.
czy $\sigma \in A$?

(P) mamy zbiór A = adresy e-mail pracowników firmy



$X[1 \dots n]$ tablica bitów

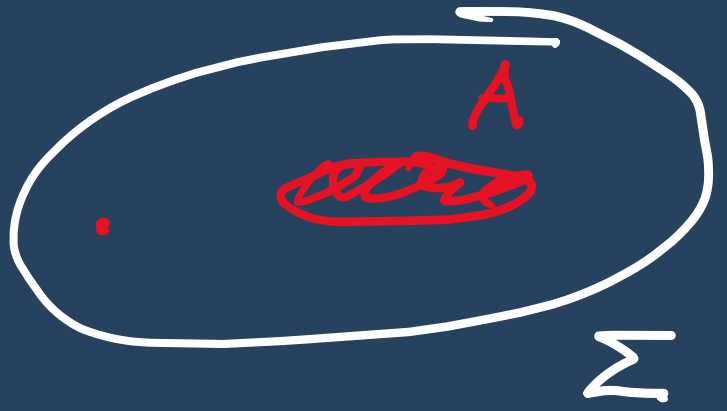
$X \leftarrow [0,0,0 \dots 0]$

$h_1, \dots, h_k : \sum \to \{1, \dots, n\}$

z dużo mniej. Uniw. rodziny fcji haszuj.

```
procedure PUT (a) {
  for (i=1; i≤k; i++) {
      X[h_i(a)] = 1
  }.
```

To rodziny dla wszystkich $a \in A$.



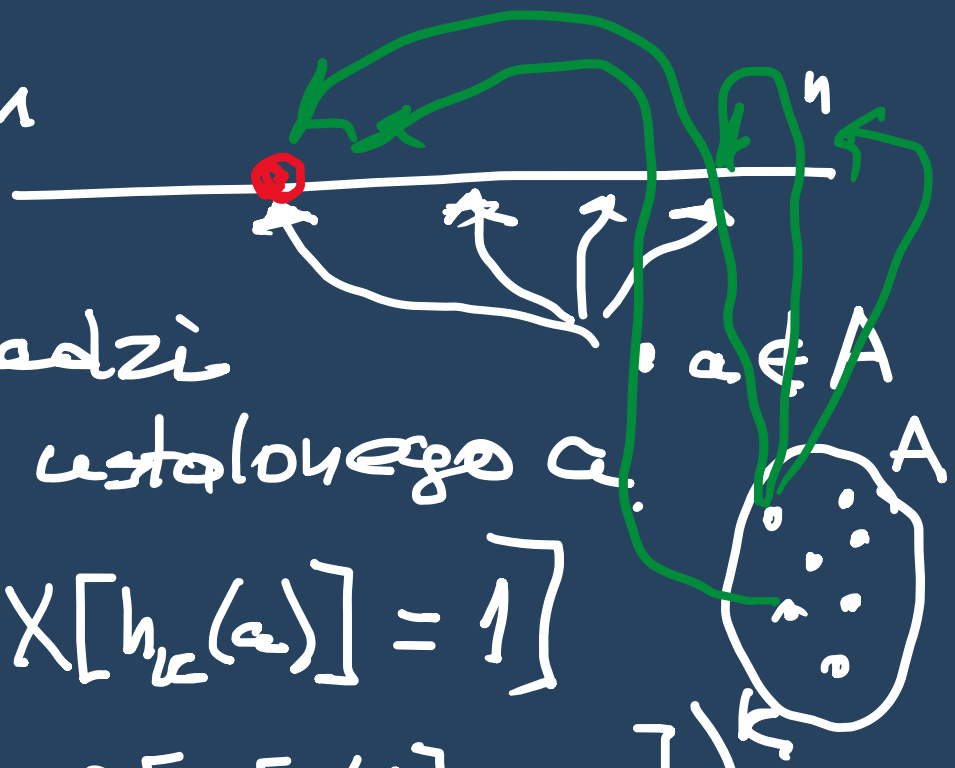

```
function Test(b) {
    ok = ~~false~~ TRUE;
    for (i=1; i ≤ k; i++){
        ok = ok && (X[h_i(b)] == 1)
    }
}
```

Jasne : $a \in A \longrightarrow Test(a) = TRUE$.

? $a \notin A$: $Test(a) = FALSE$: ok

$Test(a) = TRUE$: FALSE POSITIVE

- n - rozmiar filtru
- k - liczba funkcji hasz
- $m = |A|$.



FP = pozwalające odpowiedzi TRUE dla tego ustalonego a.

$$P[FP] = P[X[h_1(a)] = 1 \wedge \dots \wedge X[h_k(a)] = 1]$$

$$= (P[X[h_1(a)] = 1])^k = (1 - P[X[h_1(a)] = 0])^k$$

$$= \left(1 - \left(\frac{n-1}{n}\right)^{m \cdot k}\right)^k = \left(1 - \left(1 - \frac{1}{n}\right)^{m \cdot k}\right)^k$$

$$= \left(1 - \left(\left(1-\frac{1}{n}\right)^{n}\right)^{k\cdot\frac{m}{n}}\right)^{k}$$

$$k\cdot\frac{m}{n} \approx 1$$

n - dosyć duże

$$\approx \left(1 - \left(\frac{1}{e}\right)^{k\cdot\frac{m}{n}}\right)^{k} = (*)$$

$$\left(\frac{1}{e}\right) = u$$

$$0 < u < 1$$

$$(*) \cdot \left(\frac{*}{\frac{m}{n}}\right)$$

CEL: minimalizacja $(*)$.

$$CEL: \quad k\cdot\ln\left(1-\left(\frac{1}{e}\right)^{k}\right)$$

$$u = \left(\frac{1}{e}\right)^{k\cdot\frac{m}{n}}$$

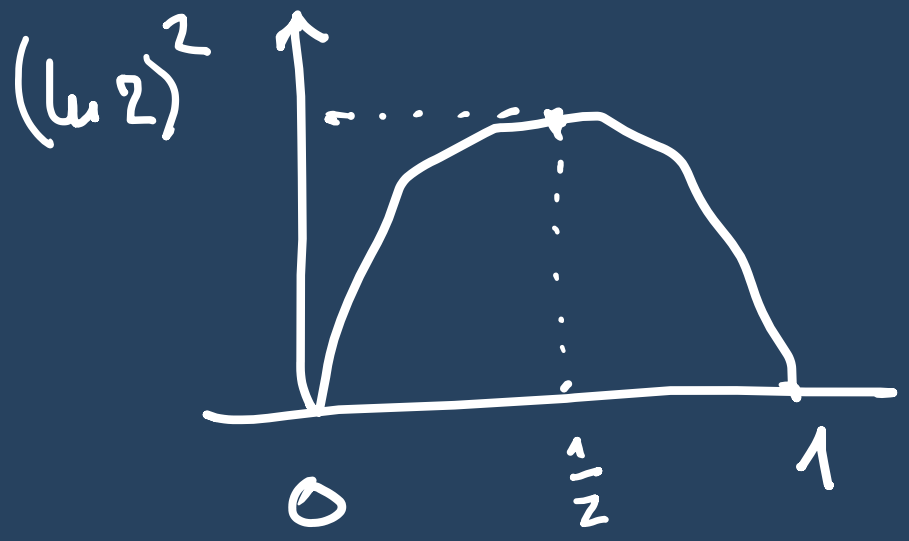
$$CEL'$$ czyli minimal.

$$\ln u = k\cdot\frac{m}{n}\ln\left(\frac{1}{e}\right)$$

$$k = -\frac{n}{m}\ln u.$$

$$-\frac{n}{m}\underbrace{\ln(u)\ln(1-u)}_{\varphi(u)}$$

$$\varphi(u) = \ln(u)\ln(1-u)$$



max. en $u = \frac{1}{2}$

$$\frac{1}{2} = \left(\frac{1}{e}\right)^{k\frac{m}{n}}$$

$$k\frac{m}{n} = \ln 2$$

$$k = \frac{n}{m}\ln 2$$

$$P[FP] = \left(\frac{1}{2}\right)^{\frac{n}{m}\ln 2} = \left(\left(\frac{1}{2}\right)^{\ln 2}\right)^{\frac{n}{m}} \simeq (0.62)^{\frac{n}{m}}$$

Dla takich parametrów

$n = 8 \cdot m \quad (= 8 \cdot |A|)$

$k = 6$

$Pr[FP] \approx 0.02 \approx \frac{1}{50}$.

(P) Speling check: język polski 40 000 słów

rozmiar filtra Bloom'a: ~~xxx bitów~~

$m = 4 \cdot 10^5$

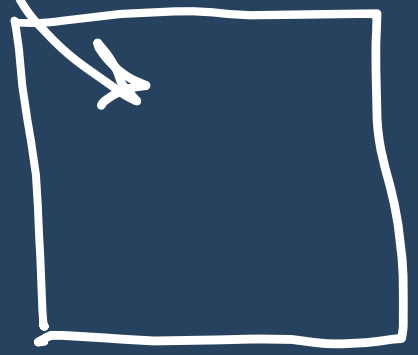
$n = 8 \cdot m$ bitów $=$ m bajtów

≈ 40 kB.

Typowe zastosowanie:

dokumenty : słowa kluczowe

Q? "number field".



$x_1, \ldots, x_n$

filtry Bloome tych samych rozm dla $A_1, \ldots, A_u$ z tymi samymi funkcjami.

$Y = X_1 \cup X_2 \cup \ldots \cup X_u$

filtr Bloome dla $A_1 \cup \ldots \cup A_u$.