Teorija kodiranja in kriptografija 2013/2014

Kriptografske zgoščevalne funkcije

Arjana Žitnik

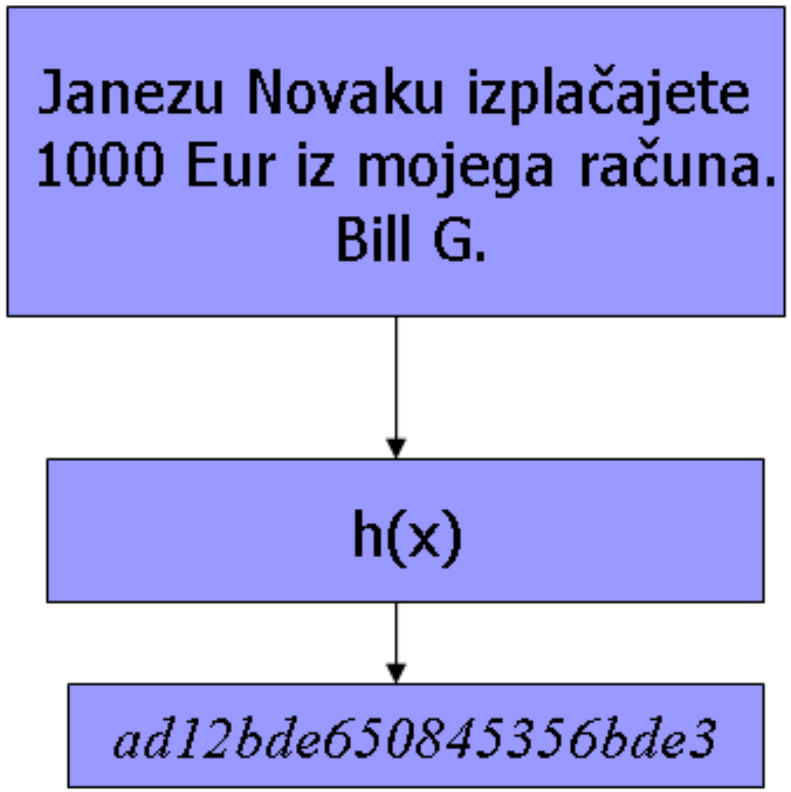
Univerza v Ljubljani, Fakulteta za matematiko in fiziko

Ljubljana, 1. 4. 2014

Kriptografske zgoščevalne funkcije

Kriptografska zgoščevalna funkcija h sporočilu poljubne dolžine priredi kratek izvleček fiksne dolžine:

$$h : \{0,1\}^* \to \{0,1\}^n$$



Uporaba zgoščevalnih funkcij

- Zagotavljanje celovitosti podatkov (dokumenti, programska oprema,...).
- Digitalni podpis: podpišemo le povzetek sporočila.
- Hranjenje gesel v računalniku.
- Uporaba v kriptografskih protokolih.
- ...

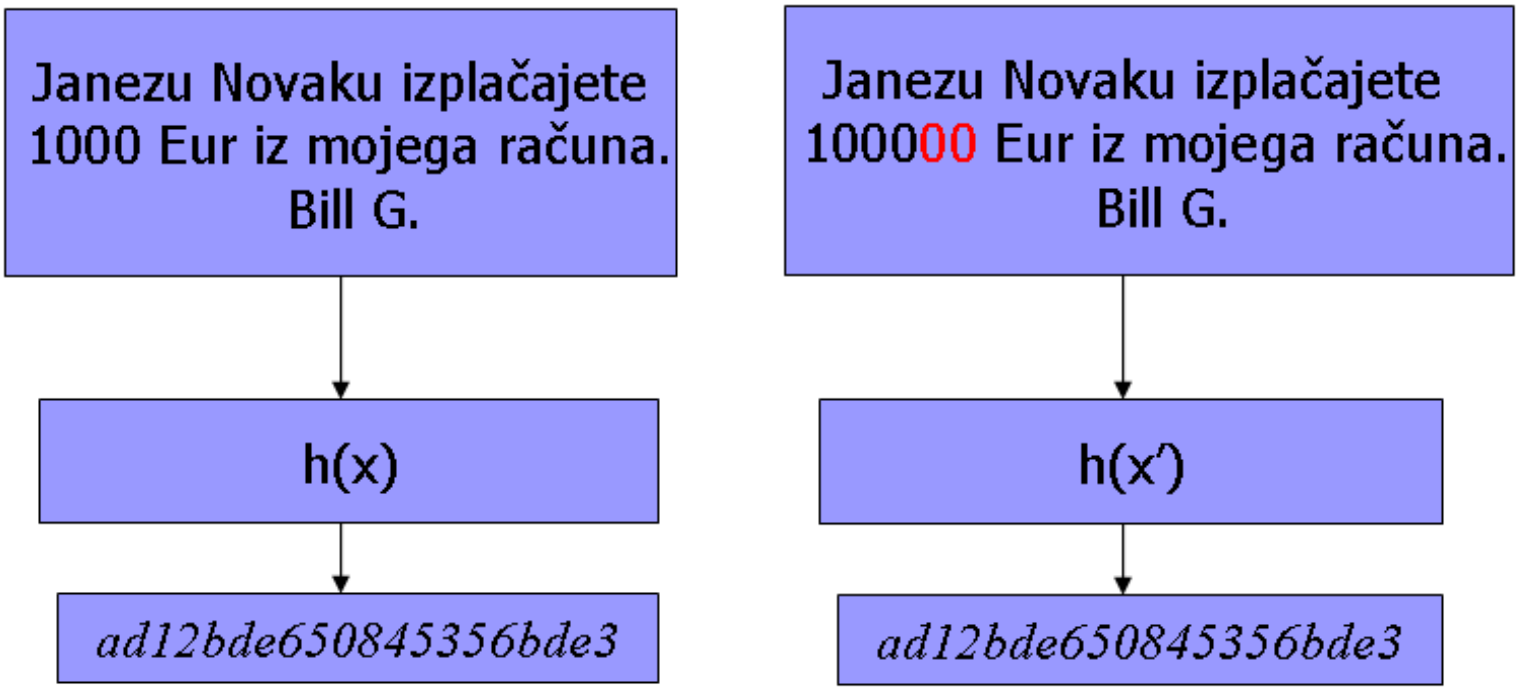
Za zagotavljanje celovitosti in avtentikacijo uporabimo zgoščevalno funkcijo s ključem (angl. message autenthication code - MAC).

Želene lastnosti zgoščevalnih funkcij

- Računanje izvlečka je enostavno in hitro.
- Naključnost: povzetek naj izgleda kot naključno število. Če se dve sporočili razlikujeta na enem mestu, naj povzetka izgledata kot neodvisno izbrani naključni števili.
- Odpornost praslik: za poljuben izvleček z je računsko nemogoče poiskati sporočilo x, da je $h(x) = z$; zgoščevalna funkcija je enosmerna funkcija .
- Odpornost drugih praslik: za dano sporočilo x je računsko nemogoče najti drugo sporočilo x', ki ima enak izvleček.
- Odpornost na trke: računsko nemogoče je poiskati dve različni sporočili x in x' z enakim povzetkom.

Trk je par različnih sporočil x in x' z enakim povzetkom.

Primer napada



Našli smo še eno sporočilo z istim povzetkom!

Paradoks rojstnega dne

Kolikšna je verjetnost, da imata v skupini 23 ljudi vsaj dva rojstni dan na isti dan?

Verjetnost je enaka

$$1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365}{343} \approx 0.507.$$

Čeprav je 23 majhno število, je med 23 osebami 253 različnih parov.

Paradoks rojstnega dne - nadaljevanje

Splošneje, k oseb, n dni. Ocenimo $1 - x \leq e^{-x}$ in dobimo

$$\prod_{i=0}^{k-1}\left(\frac{n-i}{n}\right) = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right) \leq \prod_{i=1}^{k-1} e^{\frac{-i}{n}} \quad = \quad e^{\frac{-k(k-1)}{2n}}.$$

Torej je verjetnost trčenja enaka vsaj

$$1 \; - \; e^{\frac{-k(k-1)}{2n}} \approx 1 \; - \; e^{\frac{-k^2}{2n}}.$$

Če je $k \geq \sqrt{2n\ln 2} \approx 1.17\sqrt{n}$, je ta verjetnost več kot 1/2.

Paradoks rojstnega dne in zgoščevalne funkcije

Torej: z verjetnostjo približno 1/2 najdemo trčenje, če izračunamo izvlečke za $\sqrt{2^n} = 2^{(n/2)}$ sporočil.

Nauk: v primeru, da lahko napadalec spreminja obe sporočili, mora biti dolžina izvlečka n dovolj velika, da je računsko nemogoč napad z grobo silo, ki preveri $2^{(n/2)}$ sporočil.

Zahtevnost iskanja trkov

Če je dolžina izvlečka n bitov, porabimo za

- iskanje trka za v naprej dano sporočilo $O(2^n)$ korakov
- iskanje dveh sporočil z istim izvlečkom $O(2^{(n/2)})$ korakov

V primeru, da najdemo napad, ki potrebuje bistveno manj korakov, smo zgoščevalno funkcijo razbili.

Število korakov, ki so potrebni za napad:

- $O(2^{64})$ - na meji zmogljivosti
- $O(2^{80})$ - se danes še smatra za varno
- $O(2^{128})$ - zaželena varnost

Tipična zgoščevalna funkcija

- Kompresijska funkcija:

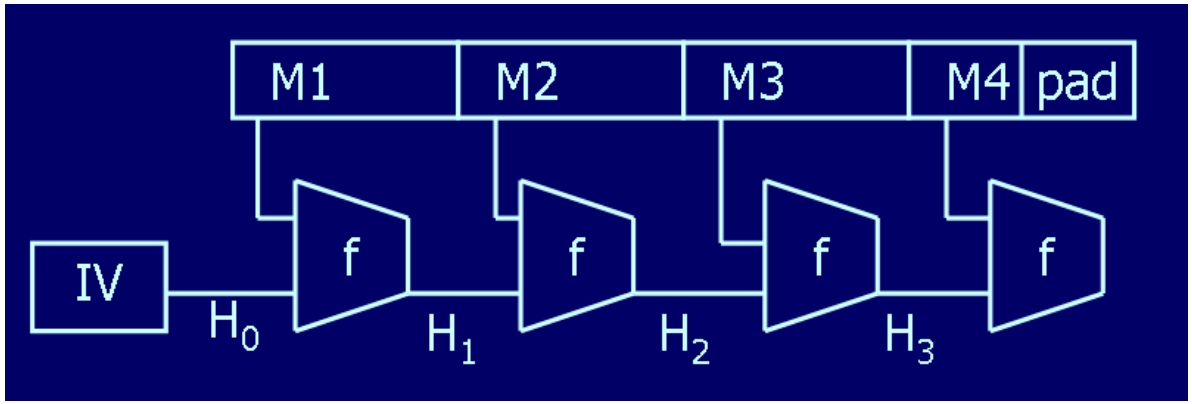
$$f : \{0,1\}^{r+n} \to \{0,1\}^n$$

- Zgoščevalna funkcija:

$$h : \{0,1\}^* \to \{0,1\}^n$$

Zgoščevalna funkcija iterativno kliče kompresijsko funkcijo.
Najbolj znana je iterativna konstrukcija Merkle-Damgård.
Pogledali bomo nekoliko poenostavljeno različico.

Iterativni postopek



Iterativni postopek:
$H_0 = IV$,
$i = 1, \ldots t$
$H_i = f(H_{i-1}||x_i)$
$h(x) = H_t$

Iterativni postopek, priprava besedila

Priprava besedila x:

- na konec dodamo najmanjše število ničel, da je dolžina besedila deljiva z r
- na konec besedila dodamo r ničel
- na konec dodamo še nekaj blokov, ki opisujejo dolžino besedila x: $|x|$ razdelimo v bloke dolžine $r-1$ in na začetek vsakega bloka dodamo 1

Besedilo razdelimo na bloke dolžine r:

$$x \mapsto x_1||x_2||...||x_t$$

Iterativni postopek: varnost

Trditev

Če je kompresijska funkcija f odporna na trke, je tudi zgoščevalna funkcija h odporna na trke.

Primeri zgoščevalnih funkcij - MD5

- Predlagana je bila leta 1991 (Ron Rivest - RSA).
- Je 128 bitna zgoščevalna funkcija.
- Algoritem je zasnovan tako, da je hiter in učinkovit na 32 bitnih računalnikih.
- Leta 2004 je bil odkrit algoritem za učinkovito iskanje trkov.

Secure Hash Algorithm (SHA)

- Razvit je bil v NIST-u kot FIPS PUB 180 leta 1993.
- Zasnovan je na strukturi MD4.
- Bloki dolžine 512 bitov, povzetek dolžine 160 bitov.
- **SHA-1** je izboljšana verzija, ki je nastala zaradi nekaterih odkritih slabosti SHA.
- Tudi SHA-1 danes ne velja več za (dolgoročno) varno. Leta 2005 je prof. Xiaoyun Wang s sodelavci objavila (teoretičen) napad na SHA-1 s pričakovano zahtevnostjo 2^{63} korakov.
- SHA-2: naslednik SHA-1 (SHA-224, SHA-256, SHA-384, SHA-512).

Zgoščevalna funkcija iz bločne šifre

Naslednji postopki veljajo za varne (12 varnih kombinacij):

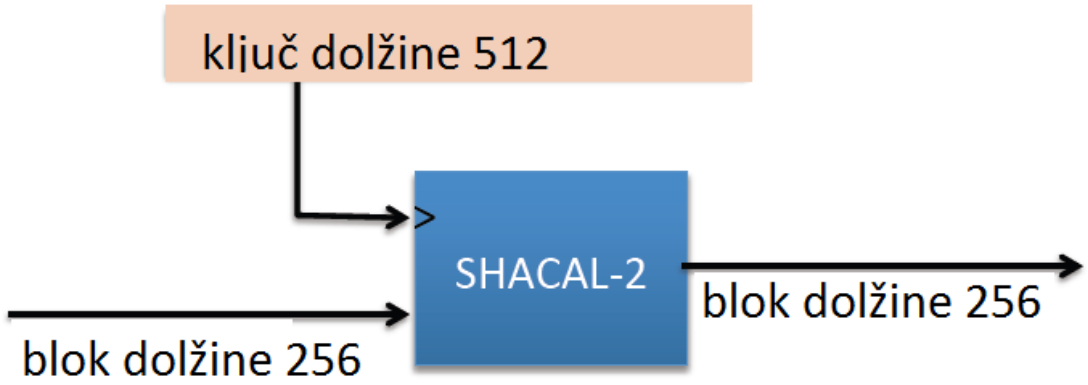
- $H_i = E_k(x_i) \oplus x_i$,
- $H_i = E_k(x_i) \oplus x_i \oplus k$,
- $H_i = E_{x_i}(H_{i-1}) \oplus H_{i-1}$ (Davies-Meyer),

kjer x_i blok besedila, $H_0 = IV$ in $k = H_{i-1}$.

Problem: prepočasi!

SHA-256

- Bloki dolžine 512 bitov, povzetek dolžine 160 bitov,
- Merkle-Damgårdova konstrukcija,
- Davies-Meyerjeva kompresijska funkcija,
- bločna šifra SHACAL-2.



Advanced Hash Standard

NIST - izbira standarda za zgoščevalno funkcijo
`http://csrc.nist.gov/groups/ST/hash/index.html`

Začetek natečaja oktober 2008: 64 predlogov
Finalisti december 2010: BLAKE, Grostl, JH, Keccak, Skein
Izbira standarda oktober 2012: Keccak

Zahteve

- besedilo dolžine vsaj 2^{64},
- dolžina izvlečka 224, 256, 384, 512 bitov,
- hitrost,
- varnost.