(Vybrané) útoky proti hašovací funkci MD5

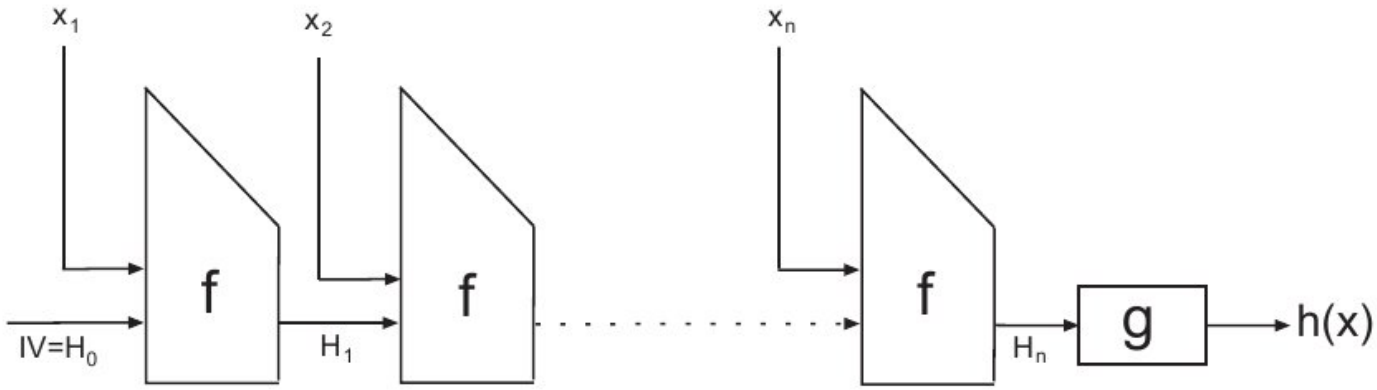
Karel Kohout (karel.kohout@centrum.cz)

18. května 2010

Hašovací funkce MD5

- MD5 = Message-Digest algorithm 5, vychází z MD4 (podobně jako SHA-1), autor prof. Ronald Rivest (RSA)
- Řetězec livobovolné délky na řetězec o 128 bitech.
- Algoritmus:
  1. Doplnění
  2. Merkle-Damgårdovo schéma po 512 bitových blocích
  3. Pro každý blok: kompresní funkce (4 kola, 16 kroků)

Merkle-Damgårdovo schéma (struktura)



Obrázek: Merkle-Damgårdovo schéma (struktura) (zdroj: [3, str. 200]); x_i jsou pro MD5 bloky o 512 bitech, H_i řetězce o délce 128 bitů a $h(x)$ výsledný hash.

Doplnění

1. Připojen jeden bit 1,
2. množství 0 tak, aby délka zprávy kongruentní s 448 modulo 512,
3. připojení délky zprávy (posledních 64 bitů[1].).

[1] V případě délky větší než 2^{64} je použito pouze spodních 64 bitů, neboli délka zprávy $mod 2^{64}$

Registry

Počáteční inicializační vektor – registry (pevně stanoveny ve standardu):

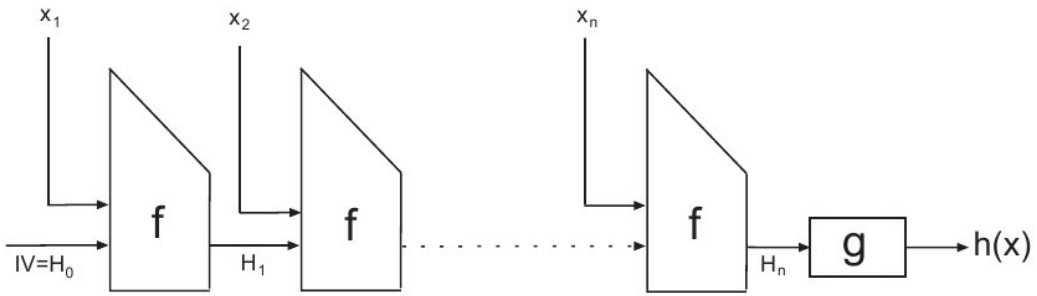
a_0 :	01	23	45	67
b_0 :	89	ab	cd	ef
c_0 :	fe	dc	ba	98
d_0 :	76	54	32	10

MD5

1. Rozdělení zprávy na N po sobě následujících bloků $M_1, M_2, ..., M_N$ délky 512 bitů,
2. pro zprávu o N blocích projde $N + 1$ stavy (mezivýsledky) IHV_i, pro $0 \leq i \leq N$.
   - Každý mezivýsledek IHV_i se skládá ze čtyř 32 bitových slov a_i, b_i, c_i, d_i (pro $i = 0$ viz výše),
   - mezivýsledky IHV_i jsou pro $i = 1, 2, ...N$ vypočítány kompresní funkcí MD5 ($MD5C$) takto:

$$IHV_i = MD5C(IHV_{i-1}, M_i).$$

Výsledný hash: IHV_N, (spojení a_N, b_N, c_N, d_N[2]).



[2] Převedený z little endian.

Kompresní funkce

Vstup do kompresní funkce $MD5C(IHV, B)$ je složený z mezivýsledku $IHV = (a, b, c, d)$ a 512 bitového bloku B. Každá runda t obsahuje modulární operace, rotaci (směrem vlevo), nelineární funkci f_t a konstanty AC_t, RC_t.

$$AC_t = \lfloor 2^{32} |sin(t+1)| \rfloor, 0 \leq t \leq 64$$

$$(RC_t, RC_{t+1}, RC_{t+2}, RC_{t+3}) = \begin{cases} (7, 12, 17, 22) & \text{pro } t = 0, 4, 8, 12, \\ (5, 9, 14, 20) & \text{pro } t = 16, 20, 24, 28, \\ (4, 11, 16, 23) & \text{pro } t = 32, 36, 40, 44, \\ (6, 10, 15, 21) & \text{pro } t = 48, 52, 56, 60. \end{cases}$$

Nelineární funkce

Nelineární funkce f_t závisí na kole:

$$f_t(X,Y,Z) = \begin{cases} F(X,Y,Z) = (X \wedge Y) \oplus (\overline{X} \wedge Z) & \text{pro } 0 \leq t < 16, \\ G(X,Y,Z) = (Z \wedge X) \oplus (\overline{Z} \wedge Y) & \text{pro } 16 \leq t < 32, \\ H(X,Y,Z) = X \oplus Y \oplus Z & \text{pro } 32 \leq t < 48, \\ I(X,Y,Z) = Y \oplus (X \vee \overline{Z}) & \text{pro } 48 \leq t < 64. \end{cases}$$

Kompresní funkce

Blok B je rozdělen do 16 po sobě následujících 32 bitových slov $m_0, \ldots, m_{15}$ a rozšířený na 64 slov W_t, pro $0 \leq t \leq 64$, každé o 32 bitech:

$$W_t = \begin{cases} m_t & \text{pro } 0 \leq t < 16, \\ m_{(1+5t) mod 16} & \text{pro } 16 \leq t < 32, \\ m_{(5+3t) mod 16} & \text{pro } 32 \leq t < 48, \\ m_{(7t) mod 16} & \text{pro } 48 \leq t < 64. \end{cases}$$

Kompresní funkce

Pro $t = 0, 1, ..., 63$ uchovává algoritmus kompresní funkce 4 stavová slova $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$. Ta jsou na začátku inicializována jako $Q_0, Q_{-1}, Q_{-2}, Q_{-3} = (b, d, d, a)$ a pro $t = 0, 1, ..., 63$ jsou postupně upravena následujícím způsobem:

$$\begin{aligned} F_t &= f_f(Q_t, Q_{t-1}, Q_{t-2}), \\ T_t &= F_t + Q_{t-3} + AC_t + W_t, \\ R_t &= RL(T_t, RC_t), \\ Q_{t+1} &= Q_t + R_t. \end{aligned}$$

Potom po provedení všech výpočtů jsou výsledná „stavová“ slova přidáva k mezivýsledku hašovací funkce a vrácena jako výsledek:

$$MD5C(IHV, B) = (a + Q_{61}, b + Q_{64}, c + Q_{63}, d + Q_{62})$$

Slovníkový útok

- Princip: znám hash a vím, že jde o krátký řetězec (typicky: dohledání hesel v databázi).
- Složitost:

Znaky (slova)	Maximální délka	Objem dat (GB), odhad
100 000 slov	$\sim$ 10 znaků	0.0026
a-z	6	7
a-z, A-Z	6	454
a-z, A-Z, 0-9	6	1281
a-z, A-Z, 0-9	8	5546713
a-z, A-Z, 0-9, 20 speciálních znaků	10	375 711 425 084

- Databáze: Google

Rainbow tables

Ukládá pouze některé výsledky ($\rightarrow$ nižší objemy dat).
Původně (Hellman):

$$C_0 = S_k(P_0)$$

P je otevřený text, C výsledek hašovací (šifrovací) funkce.
Po zavedení redukční funkce R a při procházení všech možných klíčů:

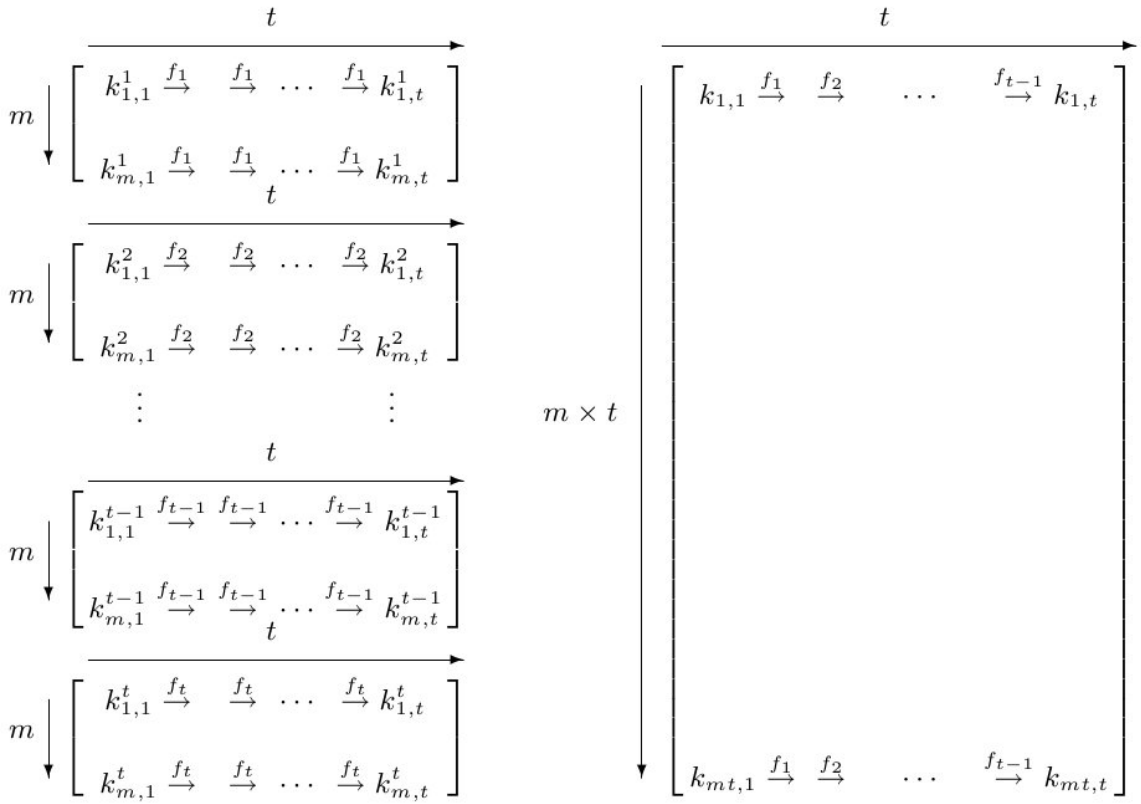
$$k_i \xrightarrow{S_{k_i}(P_0)} C_i \xrightarrow{R(C_i)} k_{i+1}$$

Označením $R(S_k(P_0))$ za $f(k)$:

$$k_i \xrightarrow{f} k_{i+1} \xrightarrow{f} k_{i+1}$$

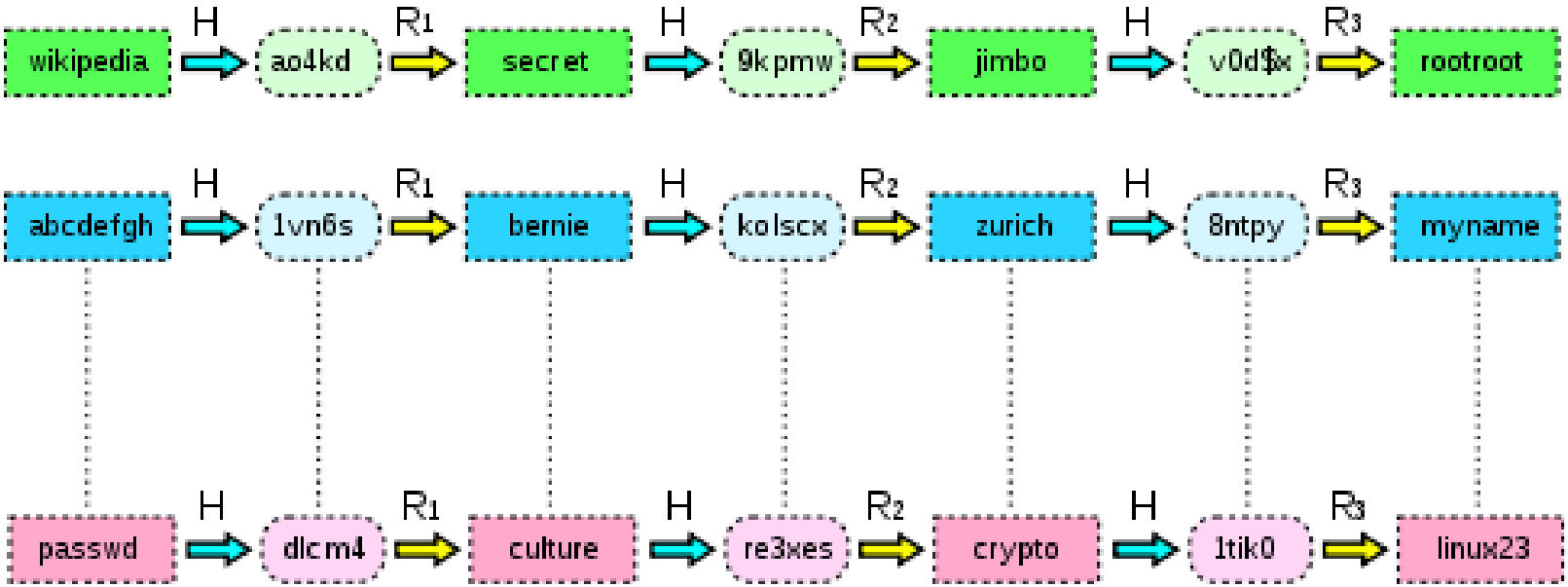
Problémy postupu: kolize řetězů (začátek jsou různé klíče, klíč nemusí být v tabulce).
Některá řešení: více tabulek, používání „distinguished points“.

Rainbow tables



Obrázek: Rainbow table (vpravo), vlevo klasické tabulky pro t-m (zdroj: [1]).

Rainbow tables

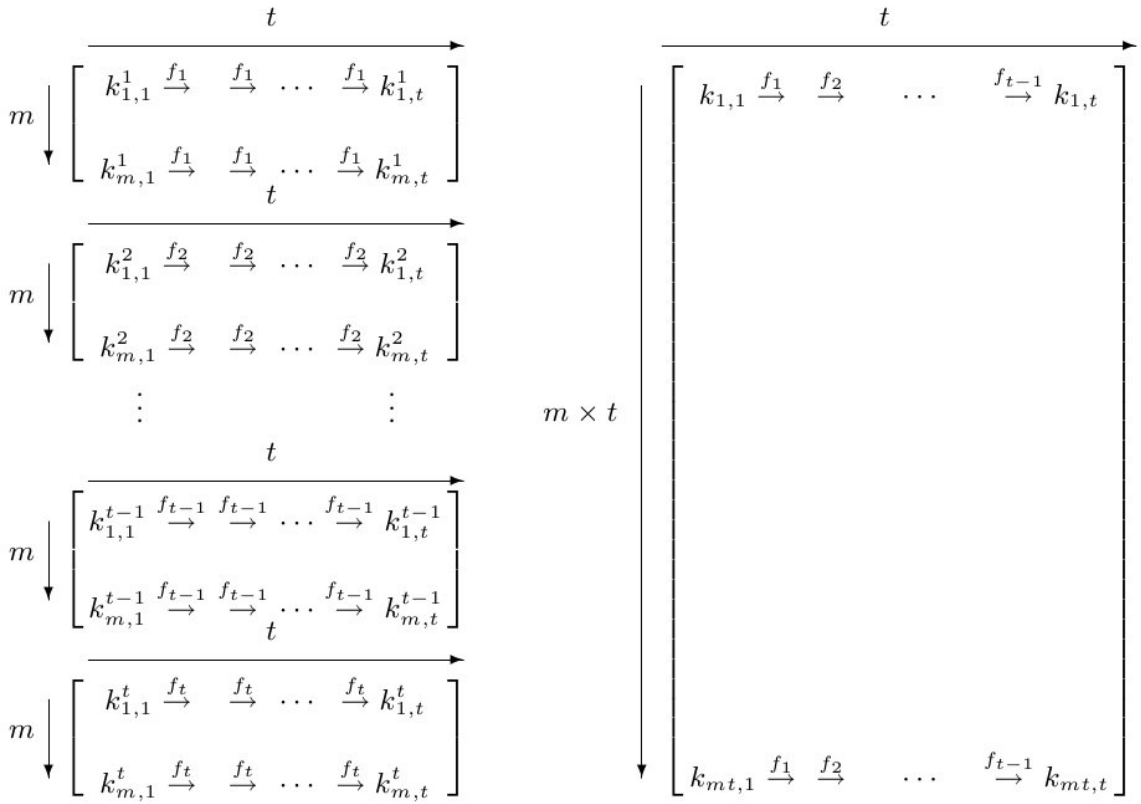


Obrázek: Rainbow table (zdroj: [4]).

Rainbow tables

	Slovníkový útok	Útok hrubou silou	Rainbow table
Prostor klíčů	23 109	$\sim$ 8 miliard	$\sim$ 8 miliard
Příprava	1.05 vteřiny	96 hodin (odhad)	20 hodin
Rychlost vyhledávání	$<$ 1 vteřina	Dle algoritmu	2.6 vteřiny (maximum)
Objem dat	$\sim$ 947 KB	300GB	$\sim$ 611 MB

Rainbow tables



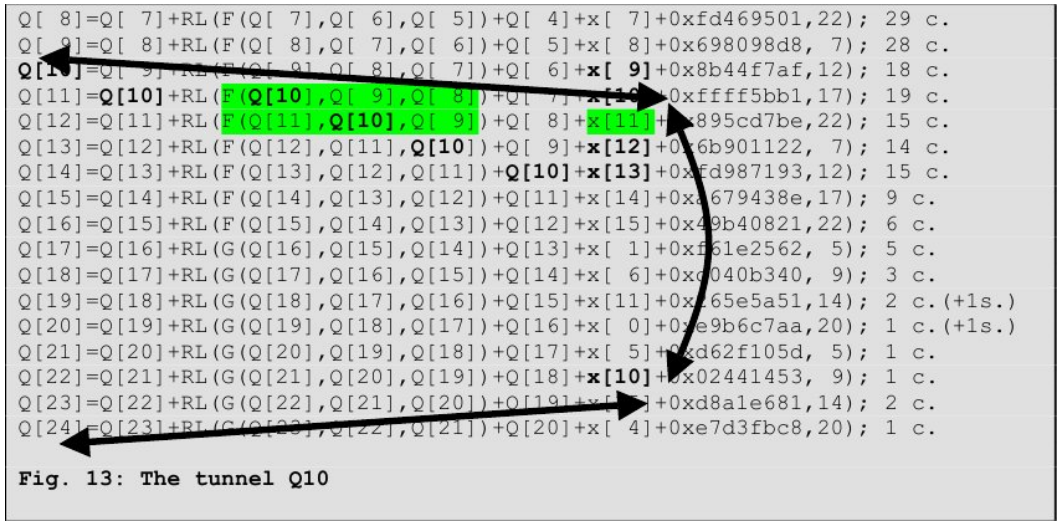
Obrázek: Rainbow table (vpravo), vlevo klasické tabulky pro t-m (zdroj: [1]).

Kolize MD5

Slabina: pouze jeden průchod přes data (vs. přidání společného konce dat). Tunely (matematické postačující podmínky pro vstup, část náhodná), závislosti v MD5.

- Náhodné kolize (Klíma, Wang), řádově desítky vteřin na běžném počítači.
- Kolize mezi dvěma určenými vstupy (Nostradamus, X.509 certikáty)
- Úpravy certifikátů (2008, mezičlánek pro RapidSSL);

Tunel (ukázka)



Obrázek: Tunel (zdroj: Klíma, Tunnels in Hash Functions: MD5 Collisions Within a Minute 1) `http://eprint.iacr.org/2006/105.pdf`).

Závěr

Nepoužívejte MD5[3].

[3] Náhrady: SHA-1, SHA-2, kolem 2012 SHA-3: http://csrc.nist.gov/groups/ST/hash/index.html.

Závěr

Fastcoll, Červená Karkulka?
HashClash: `http://www.win.tue.nl/hashclash/`
`http://www.win.tue.nl/hashclash/Nostradamus/`

For Further Reading I


Black, J., Cochran, M, A study of the md5 attacks: Insights and improvements
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.8007&rep=rep1&type=pdf


Menezes, Alfred J., Oorschot, Paul C. van, Vanstone, Scott A.: Handbook of applied cryptography, CRC Press, 1996, ISBN 9780849385230


Stinson, Douglas: Cryptography - Theory and practice, CRC Press, 1995, ISBN: 0849385210


Wikipedia, Rainbow table diagram
http://en.wikipedia.org/wiki/File:Rainbow_table1.svg


Stevens, M., Lenstra, A., Weger, B. de, Chosen-pre[U+FB01]x Collisions for MD5 and Applications
homepages.cwi.nl/~stevens/papers/stJOC-SLdW.pdf


Stevens, M., Lenstra, A., Weger, B. de, Chosen-prefix Collisions for MD5 andColliding X.509 Certi[U+FB01]cates for Di[U+FB00]erent Identities
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.3579&rep=rep1&type=pdf


Stevens, M. On Collisions for MD5 (Master's thesis)
http://www.win.tue.nl/hashclash/On%20Collisions%20for%20MD5%20-%20M.M.J.%20Stevens.pdf

For Further Reading II


Wang, X., Yu, H. How to Break MD5 and Other Hash Functions
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.6718&rep=rep1&type=pdf


The MD5 Message-Digest Algorithm, RFC 1321
http://www.ietf.org/rfc/rfc1321.txt