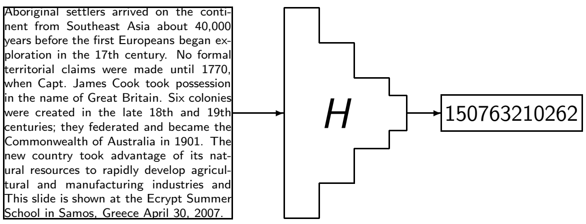
On the Design of Hash Functions

Lars R. Knudsen

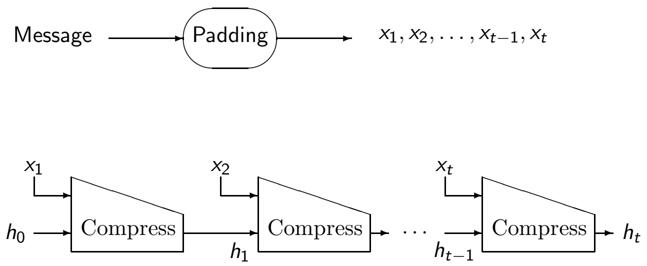
May 8, 2007

Definition - hash function



$H : \{0,1\}^* \to \{0,1\}^n$, for fixed value of n

Iterated hash functions



Damgård and Merkle (1989)

Build $H : \{0,1\}^* \to \{0,1\}^n$ from $h : \{0,1\}^m \to \{0,1\}^n$, $m > n$

1. apply padding such that $x = x_1 \mid \ldots \mid x_{t-1}$ and x_{t-1} full block
2. append to x integer $t-1$ as a string, $x = x_1 \mid \ldots \mid x_{t-1} \mid x_t$
3. define $h_0 = IV$ and $h_i = h(h_{i-1} \mid x_i)$ for $1 \le i \le t$
4. define $H(x) = h_t$

Theorem: collision for $H \Rightarrow$ collision for h

Generic attacks

For $H : \{0,1\}^* \to \{0,1\}^n$ and $h : \{0,1\}^m \to \{0,1\}^n$, $m > n$

attack	rough complexity
collisions	$\sqrt{2^n} = 2^{n/2}$
2nd preimages	2^n
preimage	2^n

Goal: generic attacks are best (known) attacks

Number-theoretic, difficult problems

- Factoring:

  given $N = pq$, find p and q,

  where p, q big, (odd) prime numbers, $p \neq q$
- Discrete logarithm:

  given $\beta = \alpha^a \bmod p$, find a,

  where p prime, a chosen random from Z_{p-1}, $\alpha \in Z_p^*$ primitive
- Note that not all instances of these problems are hard

Based on number-theoretic problems

- $N = pq$, $p \neq q$, large odd primes, α fixed, large order mod N.
- Public: N, α

$$H : \{0,1\}^* \to Z_N^*$$

$$H(x) = \alpha^x \bmod N$$

- Collision: $H(x) = H(x') \Rightarrow x - x' = k\phi(N)$.
- With $N = pq$ and $\phi(N) = (p-1)(q-1)$ easy to find p and q

Based on number-theoretic problems (2)

- Pfitzmann, Van Heijst
- Public primes: $p, q = \frac{p-1}{2}$, s.t. DLP(p) is hard
- Public primitive elements of Z_p: α, β (randomly chosen)

$$h : Z_q \times Z_q \to Z_p^*$$

$$h(x, y) = \alpha^x \beta^y \bmod p$$

- Find a collision for $h \Rightarrow$ compute $\log_\alpha(\beta)$

Based on number-theoretic problems (3)

- Goldwasser, Micali, Rivest
- $N = pq$, $p \neq q$, large primes, a_0, a_1 random squares modulo N
- Public: N, a_0, a_1

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

$$h(b, y) = y^2 \, a_0^b \, a_1^{1-b} \bmod N$$

- Collision gives x, x' such that $x^2 = x'^2 \bmod N \to$ factoring
- More efficient variants with more squares $a_0, \ldots, a_k$, Damgård

Based on number-theoretic problems (4)

- $N = pq$, $p \neq q$, large primes
- *MASH-1* (Modular Arithmetic Secure Hash)

$$h_i = ((m_i \oplus h_{i-1}) \vee a)^2 \pmod{N} \quad \oplus h_{i-1}$$

- m_i: 4 most significant bits in every byte are redundant: equal to 1111 (last byte 1010), $a = $ `0xf00...00`
- *MASH-2*: replace exponent 2 by $2^8 + 1$
- **Claims**: preimages $\sqrt{N} = N^{1/2}$, collisions $\sqrt{\sqrt{N}} = N^{1/4}$
- Both in ISO/IEC 10118-4:1998

Number-theoretic hash functions

- most schemes slow, e.g., no real speed-up for use in digital signature schemes
- some schemes have unfortunate algebraic properties (may interact badly with other public-key algorithms)
- open problem to devise efficient "provably" secure hash function

Newer constructions

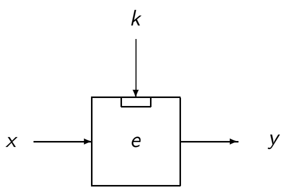
- VSH - Very Smooth Hash
  - Contini, Lenstra, Steinfeld, 2005
  - collision $\Rightarrow$ nontrivial modular square roots of very smooth numbers modulo N (composite)
  - efficient collision finder implies fast factoring algorithm
- LASH - A Lattice Based Hash Function
  - Bentahar, Page, Saarinen, Silverman, Smart 2006
  - based on the problem of finding small vectors in lattices

VSH - iterated hash function

- Let $N = pq$ be a public RSA modulus ($p \neq q$, both secret)
- Let $p_1, \ldots, p_k$ be public primes such that $\prod_{i=1}^{k} p_i < N$
  - Let $m = m_1, m_2, \ldots, m_{\ell k}$ be message, $m_i \in \{0, 1\}$
  - $x_0 = 1$
  - $x_1 = x_0^2 \, (p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}) \bmod N$
  - $x_{j+1} = x_j^2 \, \prod_{i=1}^{k} p_i^{m_{jk+i}} \bmod N$
  - $\text{Hash}(m) = x_\ell$

Block cipher - family of permutations

- $e : \{0,1\}^\kappa \times \{0,1\}^n \rightarrow \{0,1\}^n, \qquad m = \kappa + n > n$
- each κ-bit key specifies bijective mapping on n bits
- **must** hold for all x and k that $e_k^{-1}(e_k(x)) = x$.
- one-way function: given x and $e_k(x)$, hard to find k.



Product ciphers

- e most often some layers of substitutions and permutations
- example. SP-networks, ‘s’ for substitution, ‘p’ for permutation.

$$e_k(x) = s_k \circ p_k \circ s_k \circ p_k \circ \ldots \circ s_k \circ p_k \circ s_k(x)$$

- note that s_k and p_k must be invertible.

DES & AES

DES = Data Encryption Standard
AES = Advanced Encryption Standard

system	year	block size	key size
DES	1977	64	56
AES	2001	128	128, 192 or 256

Hash function using a block cipher

Why build on a block cipher?

- Advantages:
  - use existing technology
  - transfer security (trust?!) to hash construction
- Disadvantages:
  - if “keys” change often, schemes slow (due to key-schedules)
  - weaknesses of block cipher not relevant for encryption

Hash rate

Given hash function built from block cipher

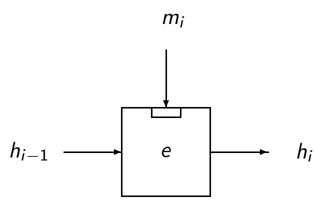
$$e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$$

Rate is defined as

$$\frac{\#\ n\text{-bit blocks hashed}}{\#\ \text{invocations of } e}$$

Single block hash (Rabin)

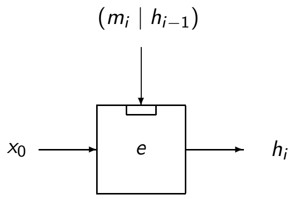
$e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$



- rate $= \kappa/n$
- one-way: no, given h_i easy to find (m_i, h_{i-1})
- attacker has full control over block cipher key

Single block hash, case: $\kappa > n$ (Merkle)

$e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$



- x_0 fixed block
- rate $= (\kappa - n)/n$
- one-wayness: given h_i, hard to find $(m_i \mid h_{i-1})$
- collision resistance ??

Single block hash

- $e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$
- 12 secure ones (Preneel 93, Black et al 2002), here three

$$\begin{aligned} h_i &= e_{m_i}(h_{i-1}) \oplus h_{i-1} && \text{Davies-Meyer} \\ h_i &= e_{h_{i-1}}(m_i) \oplus m_i && \text{Matyas-Meyer-Oseas} \\ h_i &= e_{h_{i-1}}(m_i) \oplus m_i \oplus h_{i-1} && \text{Preneel-Miyaguchi} \end{aligned}$$

- Hash rates. First one: κ/n, next two: 1
- Collisions (birthday attack) in $2^{n/2}$ operations
- Insufficient if e is DES or AES

Many hash functions have Davies-Meyer form

- Examples: MD4, MD5, SHAs
- Pros and cons of Davies-Meyer
  - Fixed points easy:

    $$h_i = e_{m_i}(h_{i-1}) \oplus h_{i-1}$$

    Choose arbitrary m_i, set $h_{i-1} := d_{m_i}(0)$. Then

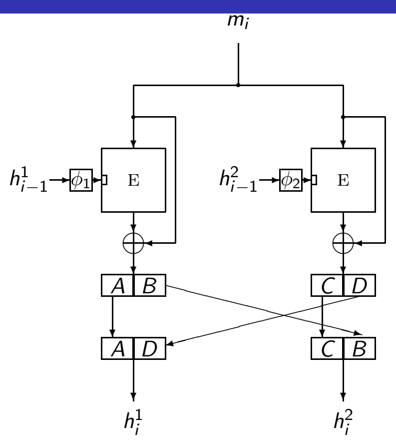
    $$h_i = h_{i-1}.$$

    Not possible in Matyas-Meyer-Oseas and Preneel-Miyaguchi
  - Hash rates for Davies-Meyer can be (arbitrarily) high

Double block hash

- Based on $e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$
- Length of hash, $2n$ bits
- Aim: 2^n security level for collisions
  - MDC-2, Brachtl, Coppersmith et al 1988/1990
  - PBGV, QG, LOKI-DBH, ...
  - Parallel-DM, 1993
  - Nandi, Hirose, 2005

MDC-2



MDC-2, MDC-4

- designed for DES
- initial values

$$h_0^1 = \{0x5252525252525252\},\ h_0^2 = \{0x2525252525252525\}.$$

- from text to key:

$$\phi_1(\cdot), \phi_2(\cdot) : \{0,1\}^{64} \to \{0,1\}^{56}$$

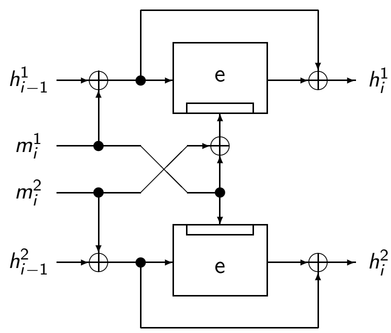
- $\phi_1(x)$, $\phi_2(y)$ never weak DES keys for any x, y
- hash rate 1/2
- MDC-4: variant using four encryptions per block

MCD-2 and MDC-4 used with DES

(Best known attacks)

	MDC-2	MDC-4
Preimage attack	2^{83}	2^{109}
2nd preimage attack	2^{83}	2^{109}
Collision attack	2^{55}	2^{56}
Hash rate	1/2	1/4

Parallel-DM, hash rate 1 - Lai et al (Crypto 93)



A large class of rate 1 hash functions

Consider the double block hash constructions

$$\begin{aligned} h_i^1 &= e_A(B) \oplus C \\ h_i^2 &= e_D(E) \oplus F \end{aligned}$$

where A, B, C linear combinations of m_i^1, m_i^2, h_{i-1}^1, and h_{i-1}^2, D, E, F are linear combinations of $h_i^1, m_i^1, m_i^2, h_{i-1}^1$, and h_{i-1}^2

- Knudsen-Lai (1993): preimages for all schemes in 2^n
- Knudsen-Lai-Preneel (1994-5): collisions $2^{n/2}$ or $2^{3n/4}$
- Ideal security not obtained by any schemes of above form

Abreast-DM & Tandem-DM - Lai, Massey 1990

$e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n,\ \kappa > n$ $\quad f(x, y) = e_x(y) \oplus y$

Abreast-DM scheme: $\left\{ \begin{aligned} h_i^1 &= f(h_{i-1}^2 \parallel m_i,\ h_{i-1}^1) \\ h_i^2 &= f(m_i \parallel h_{i-1}^1,\ \overline{h}_{i-1}^2) \end{aligned} \right.$

where $\overline{h}$ is bitwise complement of h.

Tandem-DM scheme: $\left\{ \begin{aligned} h_i^1 &= f(h_{i-1}^2 \parallel m_i,\ h_{i-1}^1) \\ h_i^2 &= f(m_i \parallel (h_i^1 \oplus h_{i-1}^1),\ h_{i-1}^2) \end{aligned} \right.$

Both hash rate 1/2, conjectured security level for collisions 2^n

Knudsen-Preneel 1996

- Compression function built from:
  - error-correcting codes
  - t small secure compression functions f_i
- Split input into small blocks, expand using code
- Different arguments to at least d of the t subfunctions
- Size of hash larger than security level
- Needs output transformation

Knudsen-Preneel, example $f_i(x, y) = e_x(y) \oplus y$

Compress: $(h^1_{i-1}, \ldots, h^5_{i-1}, m_i) \to (h^1_i, \ldots, h^5_i)$

$$
\begin{aligned}
h^1_i &= f_1(h^1_{i-1}\,,\ h^2_{i-1}) \\
h^2_i &= f_2(h^3_{i-1}\,,\ h^4_{i-1}) \\
h^3_i &= f_3(h^5_{i-1}\,,\ m_i) \\
h^4_i &= f_4(h^1_{i-1} \oplus h^3_{i-1} \oplus h^5_{i-1}\,,\ h^2_{i-1} \oplus h^4_{i-1} \oplus m_i) \\
h^5_i &= f_5(h^1_{i-1} \oplus h^3_{i-1} \oplus h^4_{i-1} \oplus m_i\,,\ h^2_{i-1} \oplus h^3_{i-1} \oplus h^5_{i-1} \oplus m_i)
\end{aligned}
$$

Constructed from $[5, 3, 3]$ Hamming code over GF(2^2): rate 1/5
Claimed security against collision attacks is 2^n
Higher rates by using codes over larger fields

Ideal cipher model

- Let $B_{n,k}$ be all block ciphers with a k-bit key and n-bit blocks,
$$\{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$$
- There are $2^n! \approx 2^{n2^n}$ bijections on n bits
- It holds that
$$|B_{n,k}| = \binom{2^n!}{2^k}$$
- An ideal cipher is randomly selected from $B_{n,k}$

Merkle's double block schemes with DES (1989)

- proof of security in ideal cipher model
- best rate about 1/4, inconvenient block sizes
- collisions $\approx 2^{55}$
- simplest scheme (rate $\simeq 1/18$):
$$h_i = \mathrm{chop}_{16}\left[f(0\|h^1_{i-1}\,,\ h^2_{i-1}\|m_i)\ \|\ f(1\|h^1_{i-1}\,,\ h^2_{i-1}\|m_i)\right]\ .$$
- $f(x, y) = e_x(y) \oplus y$ $\qquad h_{i-1} = (h^1_{i-1} \mid h^2_{i-1})$,
- $|h^1_{i-1}| = 55,\ |h^2_{i-1}| = 57,\ |m_i| = 7$

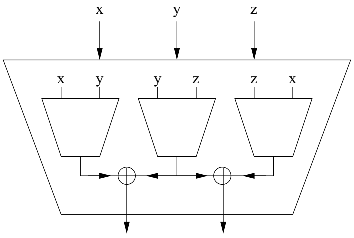
Ideal cipher model ? !

- proofs in model give protection against generic attacks
- no real-life cipher is an ideal cipher
- "nearly ideal" cipher can be strong for encryption but very weak when used for hashing
- attacker in control of key, can invest time in finding key(s) with certain properties

Ideal cipher model, cont.

- DES, weak keys, semi-weak keys
- SHACAL-1:
  - block cipher built from SHA-1
  - 160-bit blocks, 512-bit keys
  - best known attacks today:
    key-recovery attack on SHACAL-1 has complexity $\approx 2^{500}$
    collision attack on SHA-1 has complexity $\approx 2^{60}$

Nandi et al, 2005



Double length hash, rate 1/3
Collisions require $\geq 2^{2n/3}$ operations (proof, ideal cipher model)

Nandi et al, 2005

Variant based on block cipher with $\kappa = 2n$

$$e : \{0,1\}^{2n} \times \{0,1\}^n \to \{0,1\}^n$$

Yields compression function

$$h : \{0,1\}^{4n} \to \{0,1\}^{2n}$$

With $\kappa = 2n$, construction has rate 2/3

Knudsen-Muller, 2005

- collision in $2^{2n/3}$, preimages in time 2^n
- truncation to $2s$ bits: collisions in $2^{2s/3}$, preimages in 2^s

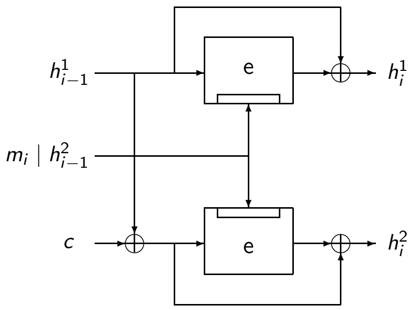
Hirose's double block mode 2006

$e : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$, $\kappa > n$, c nonzero constant

$$\begin{aligned} h_i^1 &= e_{h_{i-1}^2 \,|\, m_i}(h_{i-1}^1) \oplus h_{i-1}^1 \\ h_i^2 &= e_{h_{i-1}^2 \,|\, m_i}(h_{i-1}^1 \oplus c) \oplus h_{i-1}^1 \oplus c \end{aligned}$$

- Hash rate is $(\kappa - n)/2n$
- Collision requires 2^n operations assuming $e(\cdot,\cdot)$ is ideal cipher
- With AES-256 (128-bit block, 256-bit key), one gets hash rate 1/2 and security level 2^{128} for collisions

Hirose's double block mode, figure



Whirlpool - Barreto, Rijmen, 2003

- Based on 512-bit, 10-round block cipher W with a 512-bit key
- Preneel-Miyaguchi scheme:

$$h_i = W_{h_{i-1}}(m_i) \oplus m_i \oplus h_{i-1}$$

- W built in AES-style, 8 by 8 byte-matrix state, diffusion layer from MDS code
- ISO/IEC 10118-3:2004

Daemen-style hash constructions

- Iterated hash functions
- Compression function invertible or not hard to invert
- Invertible compression function $\rightsquigarrow$ meet-in-the-middle preimage attack with birthday attack complexity
- Cellhash, Subhash. Daemen 1991, 1992
- Radiogatun. Daemen, Peeters, Van Assche 2006
- Grindahl. Knudsen, Rechberger, Thomsen 2007

Concluding remarks

- 1980s: Hash functions based on block ciphers
- 1990s:
  - Dedicated, faster hash functions (Rivest-kickoff)
  - Many broken block cipher based hash function proposals
- 2000s:
  - Many dedicated schemes have been broken in later years
  - Many new constructions
- Future designs more conservative? (thereby slower?)
- Renaissance of block cipher based proposal?