Hash Functions Using a Block Cipher

Shoichi Hirose

University of Fukui

12th Dec. 2007

Cryptographic Hash Function

$H : \{0,1\}^* \rightarrow \{0,1\}^\ell$

Properties

Preimage resistance (PR)

It is difficult to obtain x such that $H(x) = y$ for given y.

Second preimage resistance (2ndPR)

It is difficult to obtain x' such that $H(x') = H(x)$ for given x.

Collision resistance (CR)

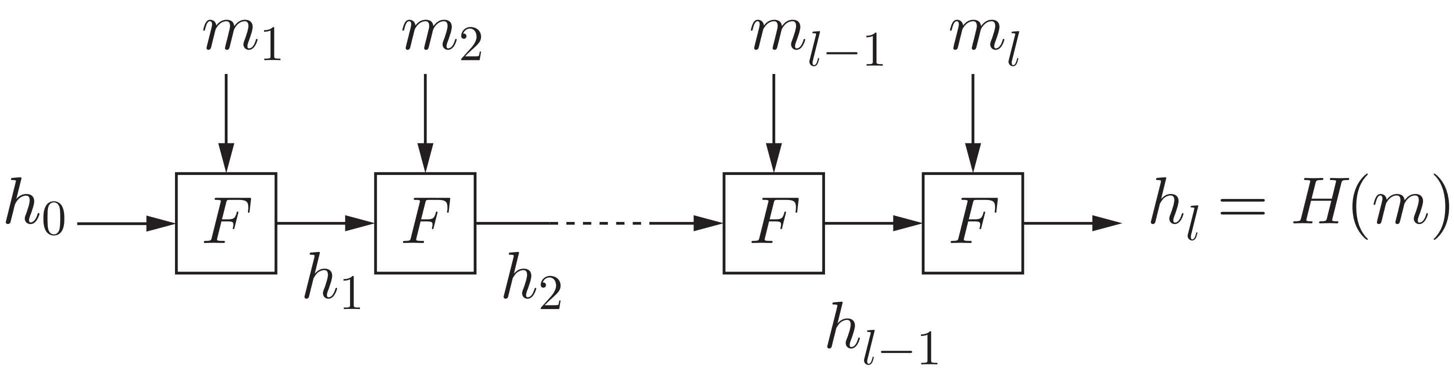
It is difficult to obtain x, x' such that $x \neq x'$ and $H(x) = H(x')$.

	PR	2ndPR	CR
Complexity	$O(2^\ell)$	$O(2^\ell)$	$O(2^{\ell/2})$

Iterated Hash Function (Merkle-Damgård)

- Compression function
  $F : \{0,1\}^\ell \times \{0,1\}^b \to \{0,1\}^\ell$
- Initial value $h_0 \in \{0,1\}^\ell$

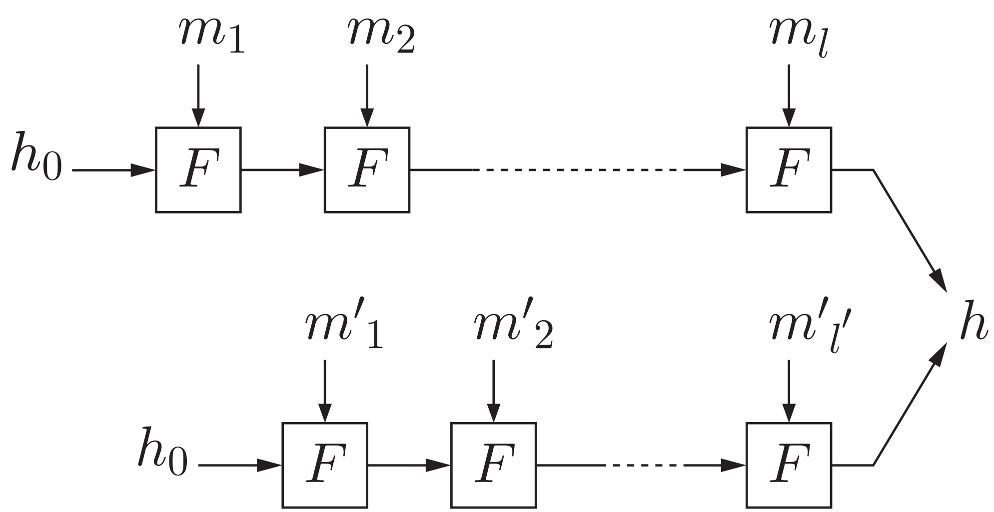
Input $m = (m_1, m_2, \ldots, m_l)$, $m_i \in \{0,1\}^b$ for $1 \le i \le l$



Iterated Hash Function

If MD-strengthening is used for padding, then

$$F \text{ is collision-resistant (CR)} \Rightarrow H \text{ is CR}$$



Advantage

- Only have to design a CR CF with fixed input length.
- Seems easier than to design a CR HF with variable IL from scratch.

Compression Function Construction

- Customized (1990–)
  - MDx family
    MD4, MD5; RIPEMD-160; SHA-1, SHA-224/256/384/512
  - Tiger
  - Whirlpool
- Using a block cipher
  - Single block length (SBL)
    output-length $=$ block-length
  - Double block length (DBL)
    output-length $= 2 \times$ block-length

Outline

- Brief overview of hash functions using a block cipher
  Single/Double-block-length constructions
- Our DBL constructions using
  - a smaller compression function
  - a block cipher
- Related DBL constructions

Motivation to Construct a Hash Function Using a Block Cipher

- MD5 and SHA-1 are vulnerable to Wang's collision attack.
- Hash functions using AES may be resistant to Wang's collision attack.
  - S-box and nonlinear key scheduling
- AES-based KDF for KEM (Jonsson & Robshaw 2005)
  - KEM-DEM using PKC and SKC without HF
- Useful for limited hardware

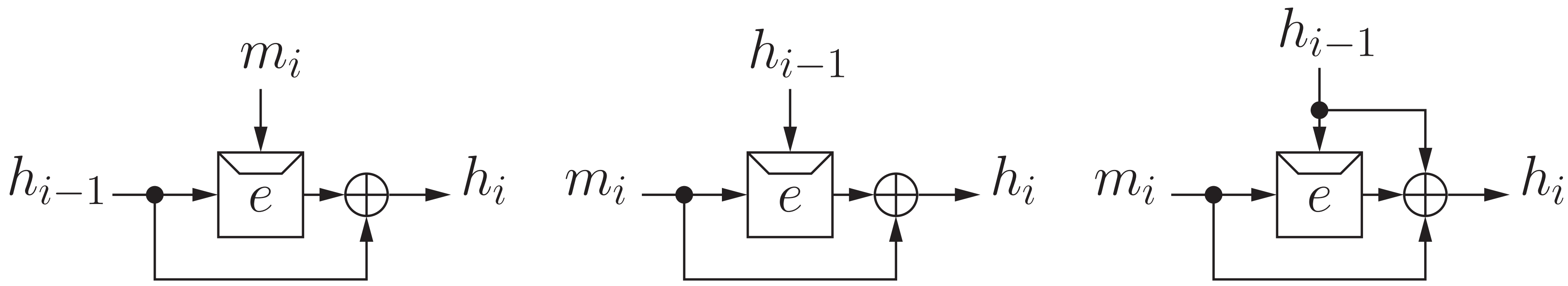
Rate

A measure of efficiency of a hash function using a block cipher e

$$\text{rate} = \frac{\text{length of the message block of the CF}}{(\text{number of invocations of } e) \times (\text{block-length of } e)}$$

Higher rate, higher efficiency.

Example: Constructions of SBL Compression Functions



Davies-Meyer Matyas-Meyer-Oseas Miyaguchi-Preneel

Note)

SHA-1: DM scheme using a dedicated 160-bit block cipher

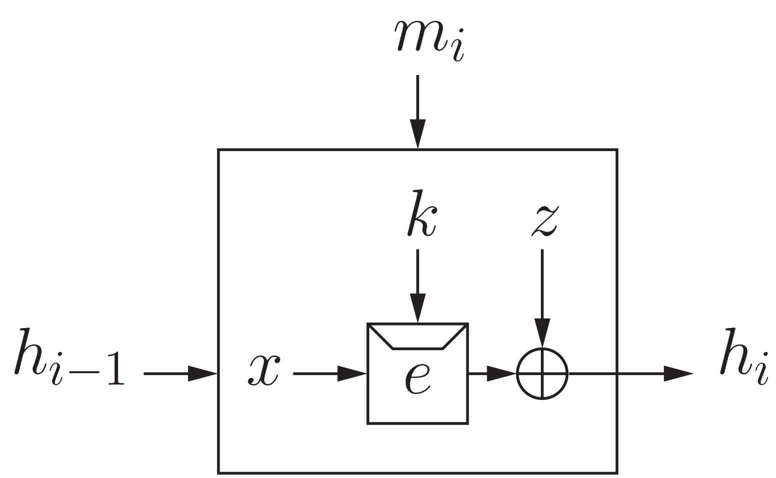
Whirlpool: MP scheme using a dedicated 512-bit block cipher W

Preneel-Govaerts-Vandewalle Model (PGV Model)

Preneel, Govaerts, Vandewalle 93

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

$x, k, z \in \{h_{i-1}, m_i, h_{i-1} \oplus m_i, 0\}$



- $4^3 = 64$ schemes
- rate $= 1$
- Some schemes are trivially insecure.

Security of the schemes in the PGV Model

- Preneel, Govaerts and Vandewalle 93
  - Security analysis against several generic attacks
  - 12 schemes are collision-resistant (CR).
- Black, Rogaway and Shrimpton 02
  - Provable security analysis in the ideal cipher model
  - The same 12 schemes are CR.

Note) DM, MMO and MP schemes are CR.

Ideal Cipher Model

Let e be an (n, κ) block cipher:

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

For each key k, $e(k, \cdot)$ is an invertible random permutation.

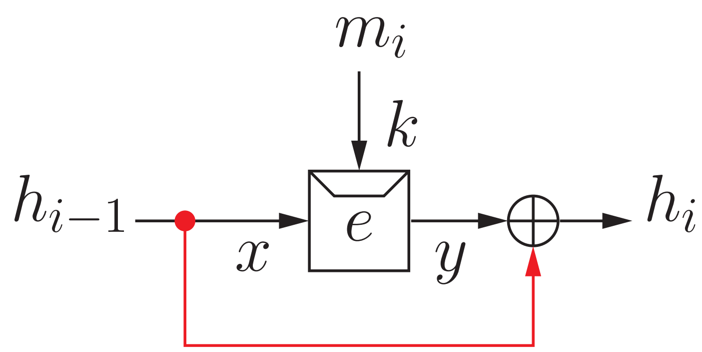
e is evaluated by two kinds of oracle queries:

oracle	query	answer
e	(key, plaintext)	ciphertext
e^{-1}	(key, ciphertext)	plaintext

Idea of the Proof

The DM compression function is CR in the ideal cipher model [Merkle 89]

$$h_i = e_k(x) \oplus x \quad \text{or} \quad y \oplus {e^{-1}}_k(y)$$



To compute h_i, we have to ask

- (k, x) to e or
- (k, y) to e^{-1}

h_i for a new input is random in the ideal cipher model.

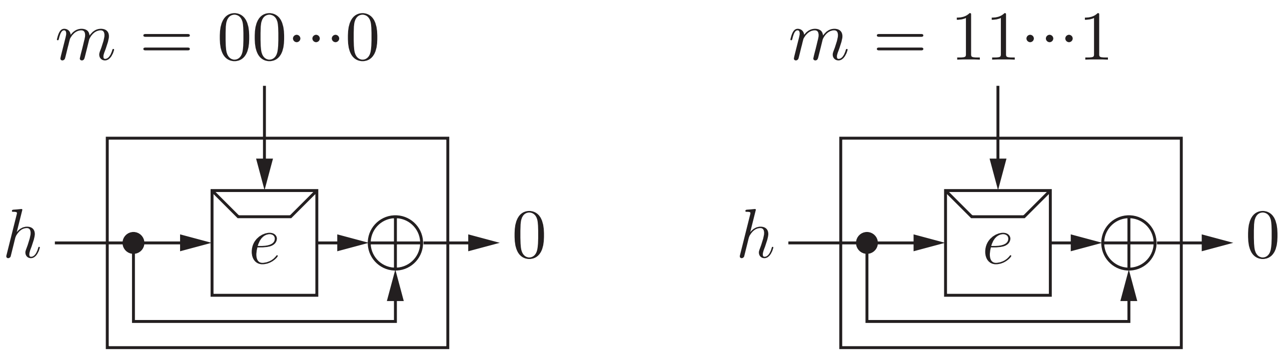
Any collision attack is at most as effective as the birthday attack.

Why Discuss CR in the Ideal Cipher Model?

An **almost** ideal cipher may not produce a CR compression function.

$$e_k(x) = \begin{cases} x & \text{if } k = 00\cdots0 \text{ or } 11\cdots1 \\ R_k(x) & \text{otherwise} \quad (R_k \text{ is a random permutation}) \end{cases}$$

There is a trivial collision of DM compression function using e:



Similar examples can be constructed for 12 CR schemes in PGV model.

Motivation to Investigate DBL Hash Function

Any SBL hash function using AES is **not secure**.

- Output length is 128 bit.
- Complexity of birthday attack $\approx 2^{64}$.

Goal

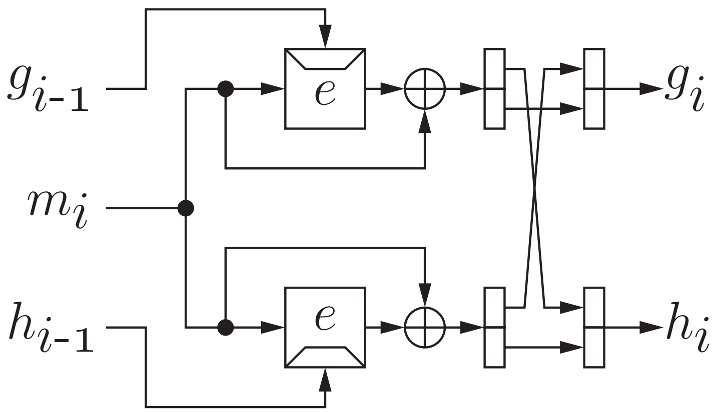
DBL hash function using $e : \{0,1\}^{\kappa} \times \{0,1\}^{n} \rightarrow \{0,1\}^{n}$

- Complexity of collision attack $\approx 2^{n}$

Example: Constructions of DBL Compression Functions (1/2)

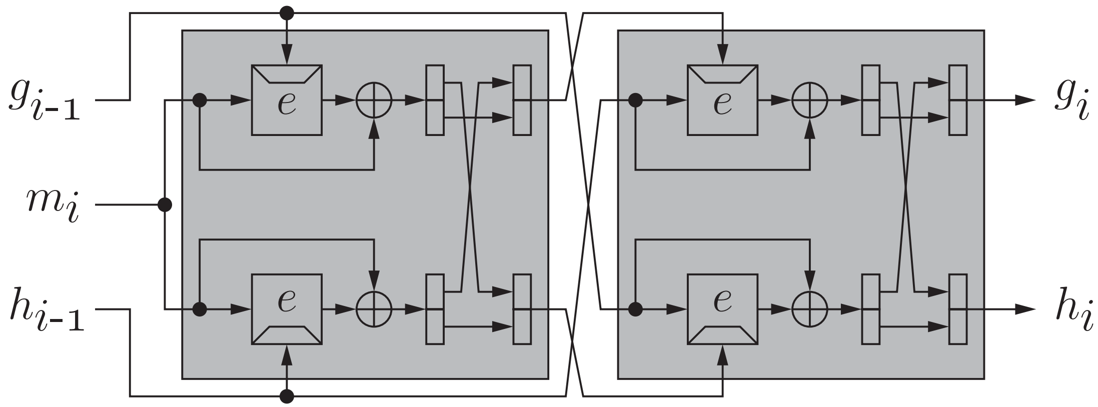
[Brachtl, Coppersmith, et.al. 88]

Using an (n, n) block cipher



MDC-2

rate $= 1/2$



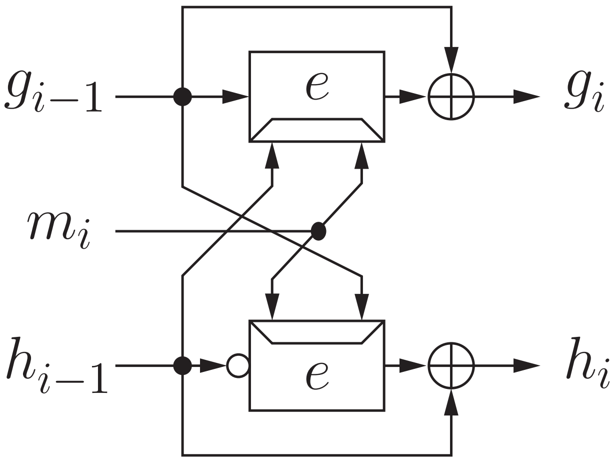
MDC-4

$1/4$

Example: Constructions of DBL Compression Functions (2/2)

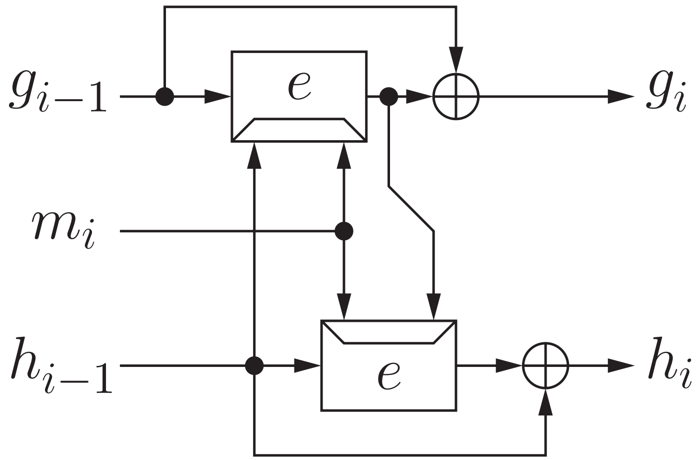
[Lai, Massey 92]

Using an $(n, 2n)$ block cipher (n-bit plaintext, $2n$-bit key)



abreast Davies-Meyer

rate $= 1/2$



tandem Davies-Meyer

$1/2$

New Constructions of DBL Compression Functions

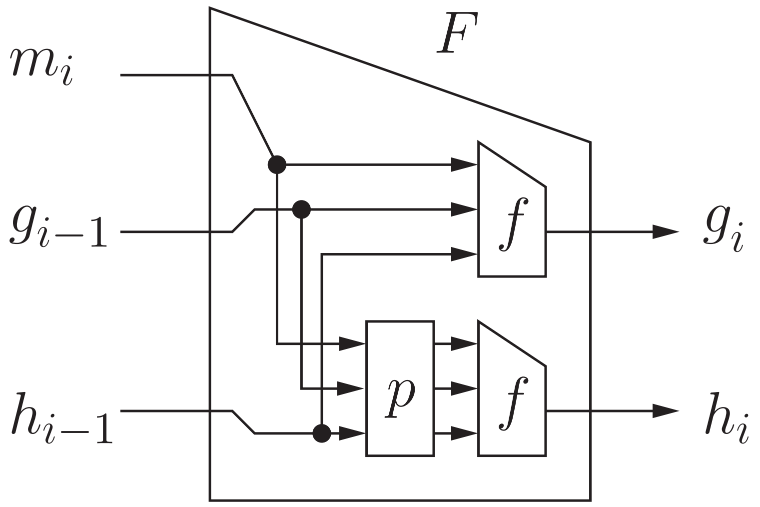
- Using a smaller compression function
  - $F(x) = (f(x), f(p(x)))$
    p is a permutation satisfying some properties
  - Collision-resistant (CR) in the **random oracle** model
- Using a block cipher with key-length $>$ block-length
  - AES with 192/256-bit key-length
  - CR in the **ideal cipher** model

Related Work

- Satoh, Haga and Kurosawa 99
  Attacks against rate-1 HFs using an $(n, 2n)$ block cipher
- Hattori, Hirose and Yoshida 03
  No CR rate-1 parallel-type CFs using an $(n, 2n)$ block cipher
- Lucks 05
  - $F(g, h, m) = (f(g, h, m), f(h, g, m))$
  - CR if f is a random oracle
- Nandi 05
  - $F(x) = (f(x), f(p(x)))$, where p is a permutation
  - CR schemes if f is a random oracle

DBL Hash Function Using a Smaller Compression Function

- f is a smaller CF
- p is a permutation
  - $p \circ p$ is an identity permutation



$$F(x) = (f(x), f(p(x)))$$
$$F(p(x)) = (f(p(x)), f(x))$$

$f(x)$ and $f(p(x))$ are used only for $F(x)$ and $F(p(x))$.

We can assume that an adversary asks x and $p(x)$ to f simultaneously in the random oracle model.

Collision Resistance

Th. 1 Let $F : \{0,1\}^{2n+b} \longrightarrow \{0,1\}^{2n}$ and $F(x) = (f(x), f(p(x)))$.

Let H be a hash function composed of F.

Suppose that

- $p \circ p$ is an identity permutation
- p has no fixed points: $p(x) \neq x$ for $\forall x$

$$\mathbf{Adv}_H^{\mathrm{coll}}(A) \stackrel{\mathrm{def}}{=} \text{success prob. of a collision finder } A \text{ for } H$$

which asks q pairs of queries to f.

Then, $\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq \frac{q}{2^n} + \left(\frac{q}{2^n}\right)^2$ for any A in the RO model.

Proof Sketch

F is CR $\Rightarrow$ H is CR

Two kinds of collisions for F:

$$\Pr[F(x) = F(x') \,|\, x' \neq p(x)]$$

$$= \Pr[f(x) = f(x') \wedge f(p(x)) = f(p(x'))] = \left(\frac{1}{2^n}\right)^2$$

$$\Pr[F(x) = F(x') \,|\, x' = p(x)] = \Pr[f(x) = f(p(x))] = \frac{1}{2^n}$$

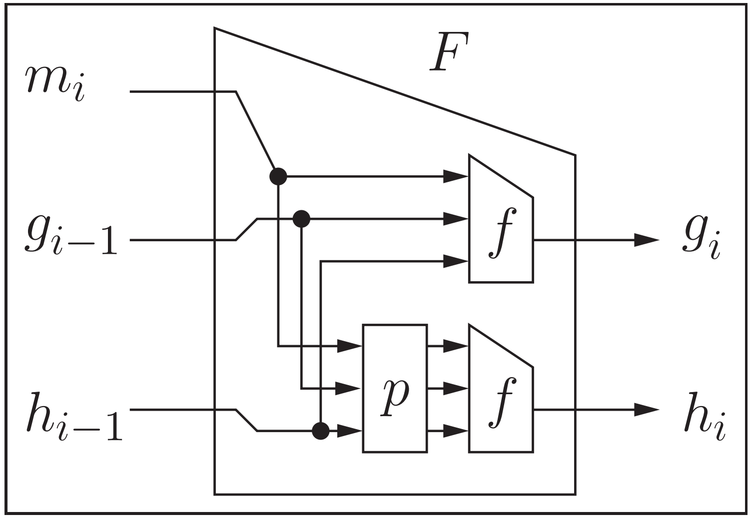
A asks q pairs of queries to f: x_j and $p(x_j)$ for $j = 1, 2, \ldots, q$.

$$\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq \frac{q}{2^n} + \left(\frac{q}{2^n}\right)^2$$

Collision Resistance: A Better Bound

Th. 2 Let H be a hash function composed of $F : \{0,1\}^{2n+b} \to \{0,1\}^{2n}$.

Suppose that



- $p \circ p$ is an identity permutation
- $p(g,h,m) = (p_{\mathrm{cv}}(g,h), p_{\mathrm{m}}(m))$
  - p_{cv} has no fixed points
  - $p_{\mathrm{cv}}(g,h) \neq (h,g)$ for $\forall (g,h)$

Then, $\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq 3\left(\frac{q}{2^n}\right)^2$ for any A in the RO model.

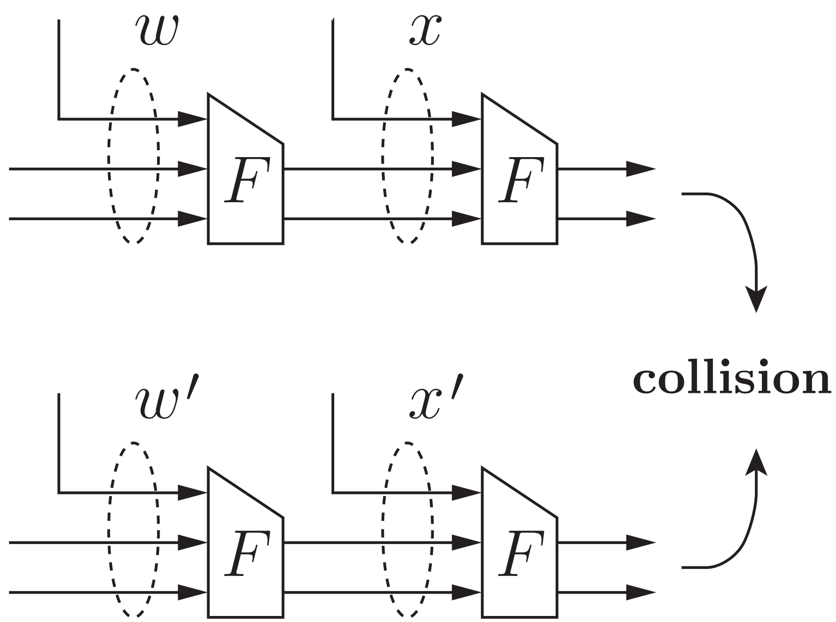
Proof Sketch (1/2)

Two kinds of collisions for F:

1. $\Pr[F(x) = F(x') \mid x' \neq p(x)] = \left(\frac{1}{2^n}\right)^2$
2. $\Pr[F(x) = F(x') \mid x' = p(x)] = \frac{1}{2^n}$

It is easier to find a type-2 collision.
However, a type-2 collision is accompanied by a **pseudo-collision** w, w' such that

- $\boldsymbol{F(w') = p_{\mathrm{cv}}(F(w))}$
- $\boldsymbol{w' \neq p(w)}$



Proof Sketch (2/2)

A collision for H implies

1. a collision for F such that

$$\Pr[F(x) = F(x') \,|\, x' \neq p(x)] = \left(\frac{1}{2^n}\right)^2$$

2. a **pseudo-collision** for F such that

$$\Pr[F(w') = p_{\mathrm{cv}}(F(w)) \,|\, w' \neq p(w)] = \left(\frac{1}{2^n}\right)^2$$

$$\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq 3\left(\frac{q}{2^n}\right)^2 = \left(\frac{q}{2^n}\right)^2 + 2\left(\frac{q}{2^n}\right)^2$$

Th. 1 vs. Th. 2

The difference between the upper bounds is significant.

E.g.) $n = 128$, $q = 2^{80}$

Th. 1 $\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq \frac{q}{2^n} + \left(\frac{q}{2^n}\right)^2 \approx 2^{\mathbf{-48}}$

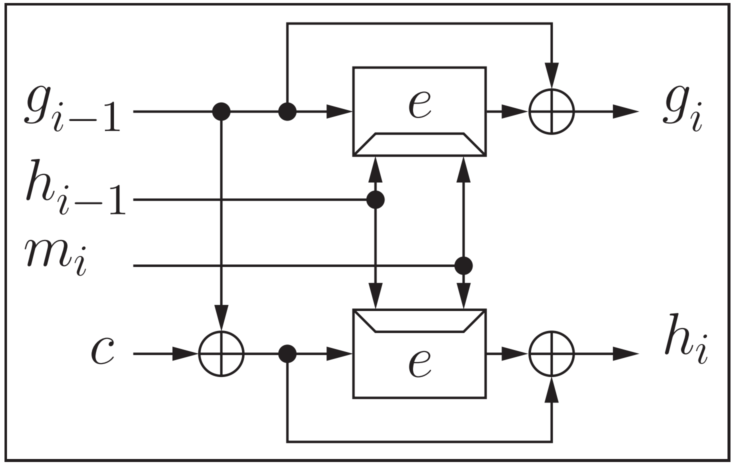
Th. 2 $\mathbf{Adv}_H^{\mathrm{coll}}(A) \leq 3\left(\frac{q}{2^n}\right)^2 \approx 2^{\mathbf{-94}}$

E.g.) A permutation p satisfying the properties in **Th. 2**

$$p(g, h, m) = (g \oplus c_1, h \oplus c_2, m), \text{ where } c_1 \neq c_2$$

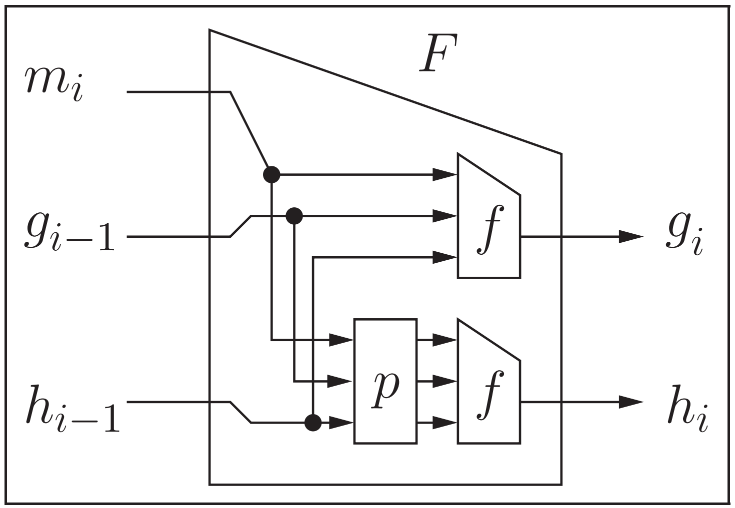
DBL Hash Function Composed of a Block Cipher

$F =$

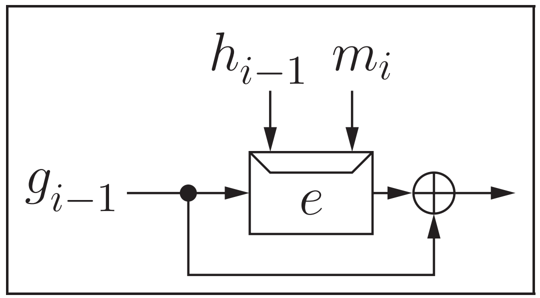


c is a non-zero constant.

Note)

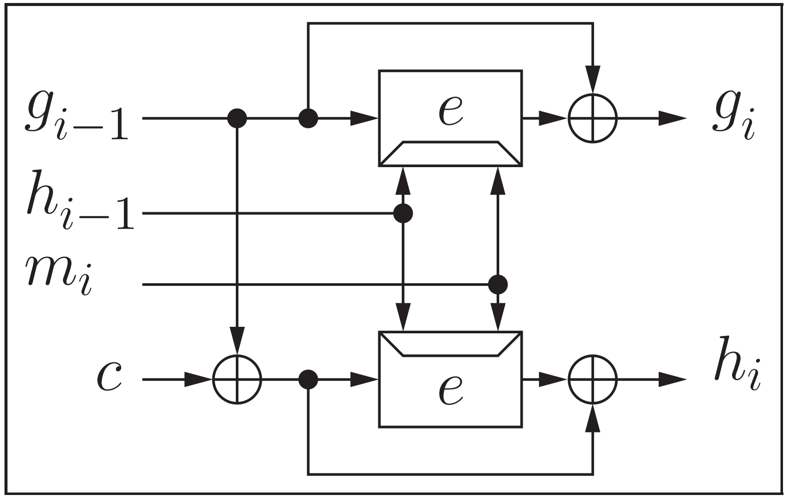


such that $f =$



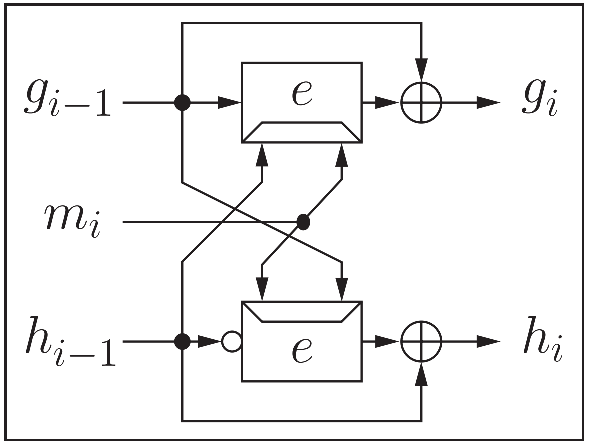
$$p(g, h, m) = (g \oplus c, h, m)$$

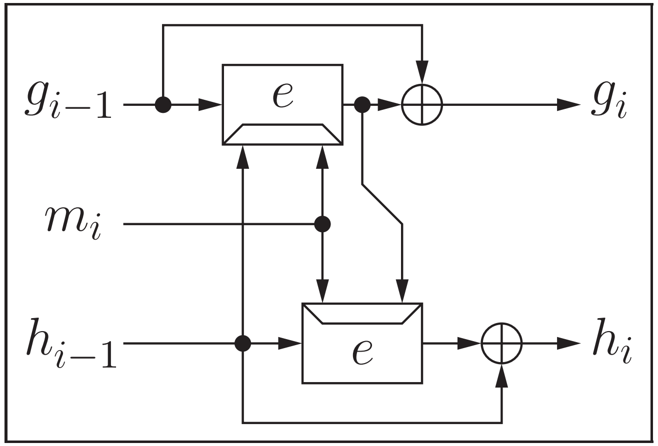
DBL Hash Function Composed of a Block Cipher



- can be constructed using AES with 192/256-bit key
- rate $= \begin{cases} 1/2 & \text{with AES-256} \\ 1/4 & \text{with AES-192} \end{cases}$
- requires only one key scheduling

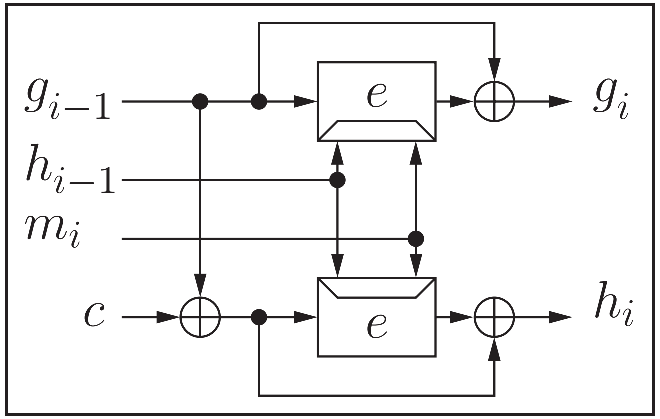
Simpler than abreast Davies-Meyer and tandem Davies-Meyer





Collision Resistance

Th. 3 Let H be a HF composed of $F : \{0,1\}^{2n+b} \to \{0,1\}^{2n}$ such that

$F =$ 

.

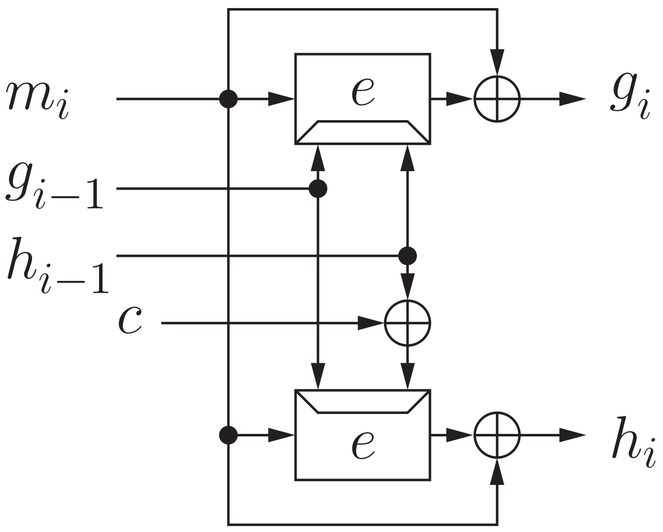
$$\mathbf{Adv}_H^{\mathrm{coll}}(A) \stackrel{\mathrm{def}}{=} \text{success prob. of a collision finder } A \text{ for } H$$

which asks q pairs of queries to (e, e^{-1}).

Then, in the ideal cipher model, for any A and $1 \le q \le 2^{n-2}$,

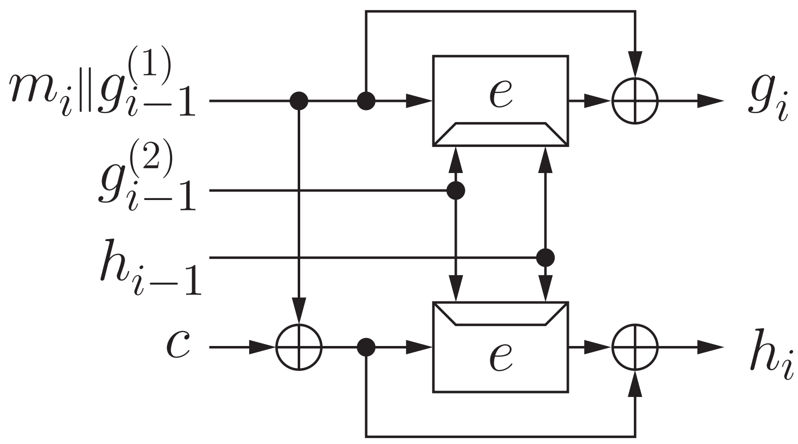
$$\mathbf{Adv}_H^{\mathrm{coll}}(A) \le 3 \left(\frac{q}{2^{n-1}} \right)^2$$

A Few More Examples of Compression Functions



For AES with 256-bit key

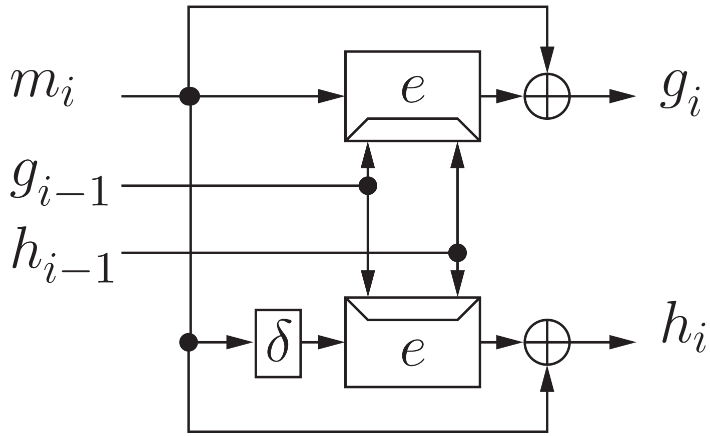
rate $= 1/2$



For AES with 192-bit key

$1/4$

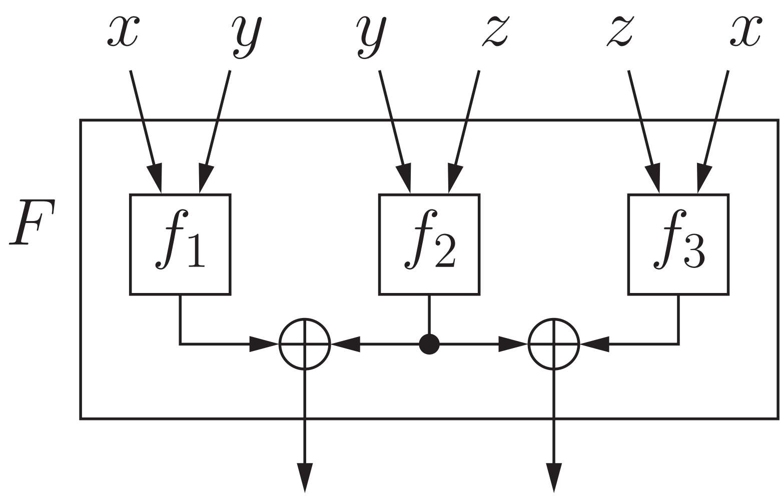
Jonsson & Robshaw (PKC05)



r	$00\|r'$	$01\|r'$	$10\|r'$	$11\|r'$
$\delta(r)$	$01\|r'$	$10\|r'$	$11\|r'$	$00\|r'$

$$\delta(r) = \delta((a)_2\|r') = (a + 1 \bmod 4)_2\|r'$$

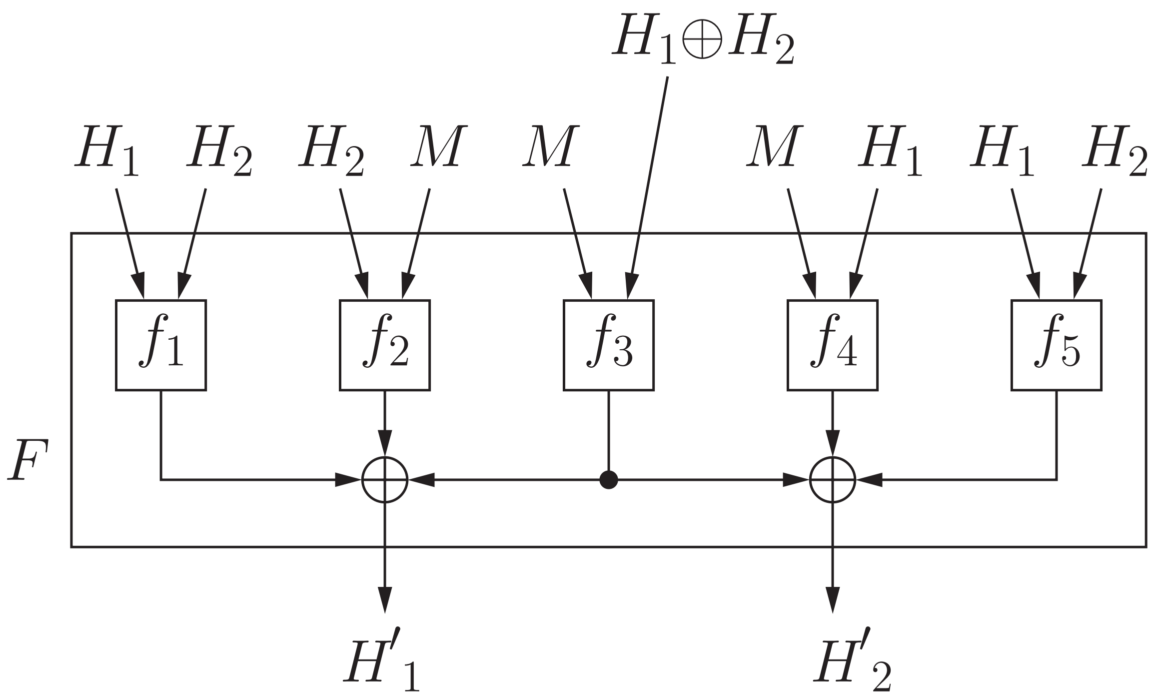
Nandi, Lee, Sakurai & Lee (FSE05)



$$f_i : \{0,1\}^{2n} \to \{0,1\}^n \qquad F : \{0,1\}^{3n} \to \{0,1\}^{2n}$$

- Rate $= 1/3$ using $\boldsymbol{(n, n)}$ block ciphers for f_i's
- Complexity of collision attack $= \Theta(2^{\frac{2n}{3}})$

Peyrin, Gilbert, Muller & Robshaw (ASIACRYPT06)



$$f_i : \{0,1\}^{2n} \to \{0,1\}^n \qquad F : \{0,1\}^{3n} \to \{0,1\}^{2n}$$

- Rate $= 1/5$ using $\boldsymbol{(n, n)}$ block ciphers for f_i's
- Complexity of collision attack $= \Theta(2^{\frac{2n}{3}})$ (Seurin, Peyrin FSE07)

ISO/IEC 10118

Consists of four parts:

1. General
2. Hash-functions using an n-bit block cipher
3. Dedicated hash-functions
4. Hash-functions using modular arithmetic

ISO/IEC 10118-2:2000 (Hash-functions using an n-bit block cipher)

- Cancels and replaces the first edition (ISO/IEC 10118-2:1994)
- Specifies four hash-functions

  Hash-function one: Matyas-Meyer-Oseas

  Hash-function two: MDC-2

  Hash-function three

  Hash-function four

Hash-functions 3/4 are complicated and inefficient

- Do not seem suitable for practical use

MDC-2 vs. Our Scheme

e: n-bit block, κ-bit key

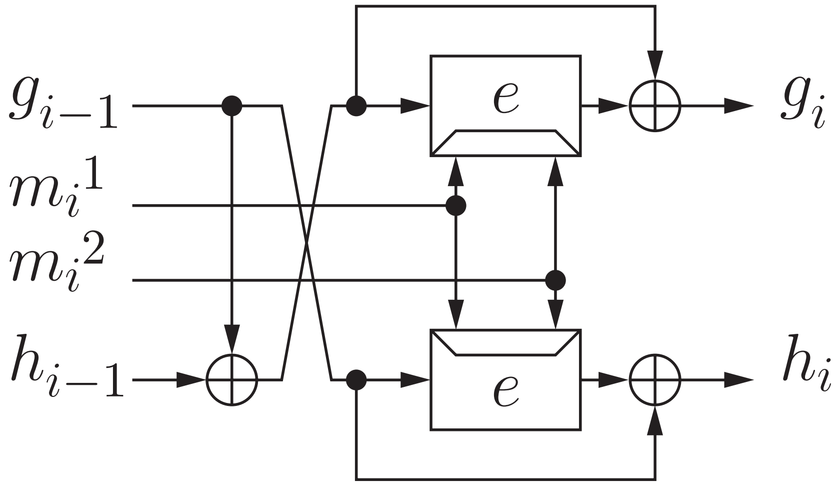
	MDC-2	Ours
Key Length κ	n	$> n$
Rate	$1/2$	$(\kappa - n)/(2n)$
Collision Attack	$\Omega(2^{0.6\,n})$	$\Theta(2^n)$

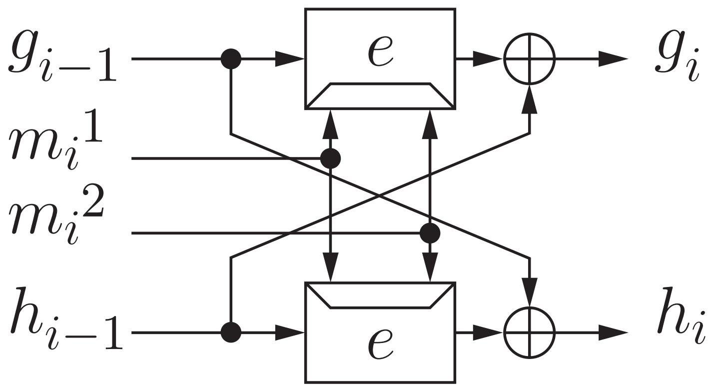
Complexity of collision attack on MDC-2 is from [Steinberger 06].

Conclusion

- Brief overview of hash functions using a block cipher
  Single/Double-block-length constructions
- Our DBL constructions using
  - a smaller compression function
  - a block cipher
- Related DBL constructions

Constructions As Efficient As MDC-2





- rate $= \dfrac{\kappa}{2n}$ with an (n, κ) block cipher
- As secure as MDC-2? [Satoh, Haga, Kurosawa 99]