On the (In)Security of IDEA in Various Hashing Modes

Lei Wei[1], Thomas Peyrin[1], Przemysław Sokołowski[2], San Ling[1], Josef Pieprzyk[2], and Huaxiong Wang[1]

[1]Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

[2]Macquarie University, Australia

FSE 2012

Overview of attacks to IDEA hashing modes

Mode	hash output size	compression function			hash function
		free-start collision attack	semi-free-start collision attack	preimage attack complexity (s, p)	collision attack
Davies-Meyer	64	2^1		$2^{25.5}$ $(2^{17.5}, 2^{-17.5})$	$2^{16.13}$
Hirose	128	2^1		$2^{25.5}$ $(1, 2^{-64})$	
Abreast-DM	128	$2^{48.13}$		$2^{25.5}$ $(1, 2^{-64})$	
Tandem-DM	128	$2^{48.13}$		$2^{25.5}$ $(1, 2^{-64})$	
Peyrin *et al.*(II)*	128	2^1 / $2^{48.13}$	2^1 / $2^{48.13}$	$2^{25.5}$ $(1, 2^{-64})$	
MJH-Double	128	$2^{32.26}$	$2^{32.26}$	$2^{25.5}$ $(2^{17.5}, 2^{-17.5})$	

- The results are directly supported by experiments. Practical examples are computed for some of these attacks.
- The preimage complexity results find s preimages on average with a certain probability p, for a total average of $A = s \cdot p$ solutions.
- The attacks to Peyrin *et al.* (II) mode are valid only if the block cipher instances are used in certain ways.

Outline

- IDEA hashing modes
- Simple collision attacks
- Improved collision attacks
- Preimage attacks

Hash Functions from Merkle-Damgård Algorithm

An n-bit hash function with IV and m message blocks M_i

- uses n-bit compression function h as building block,
- processes M_i as $CV_{i+1} = h(CV_i, M_i)$, with $CV_0 := IV$,
- The final hash value is $H_m := CV_m$.

Collision security can be reduced to the compression function.

Attacks

- *free-start collision*: in less than $2^{n/2}$ computations, find $(CV, M) \neq (CV', M')$ s.t. $h(CV, M) = h(CV', M')$.
- *semi-free-start collision*: in less than $2^{n/2}$ computations, find CV and $M \neq M'$ s.t. $h(CV, M) = h(CV, M')$.
- *preimage*: in less than 2^n computations, find CV and M s.t. for a given output challange X: $h(CV, M) = X$.

n-bit block cipher $\longrightarrow$ n-bit compression function:

- Simple-length constructions: e.g. Davies-Meyer (DM), Miyaguchi-Preneel (MP), Matyas-Meyer-Oseas (MMO).

Block Cipher Based Hashing

IDEA the International Data Encryption Algorithm, designed by Xuejia Lai and James Massey in 1991.

- 64-bit block size, 128-bit key.
- Receives extensive cryptanalysis and is regarded as a very secure block cipher.

Double-block length (DBL) constructions: n-bit block ciphers of $2n$-bit key.

- Bigger hash sizes by making use of double-key block ciphers: e.g. IDEA, AES-256.
- DBL Constructions: Hirose DBL mode, Peyrin et al. (II), MJH-Double.
- Abreast-DM and Tandem-DM were initially proposed for hashing with IDEA.

The DBL Modes: Abreast-DM and Tandem-DM

Both are especially designed for IDEA, by Lai and Massey (Eurocrypt'92).

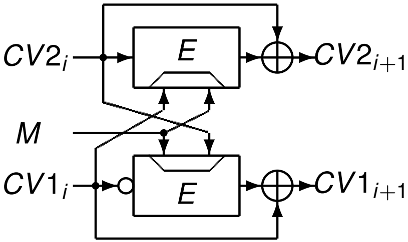


Figure: Abreast-DM

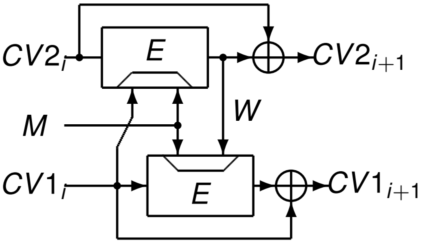
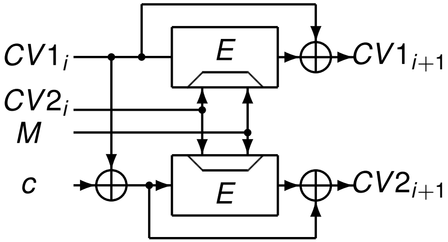


Figure: Tandem-DM

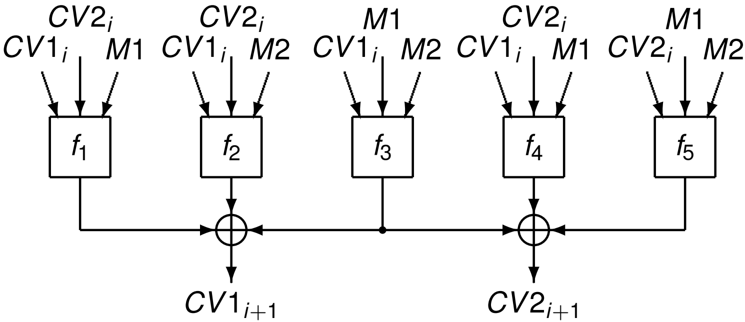
The DBL Modes: Hirose



- Proposed by Shoichi Hirose (ICISC'04, FSE'06).
- Using a constant *c* to simulate two independent ciphers.

The DBL Modes: Peyrin *et al.* (II)

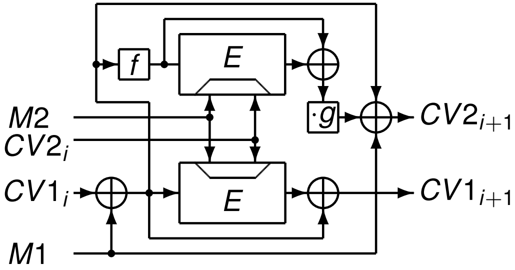
Proposed by Peyrin, Gilbert, Muller and Robshaw (Asiacrypt'06).



5 independent $3n$-to-n-bit compression functions are called, advising to be instantiated with double-key block ciphers such as AES-256 and IDEA.

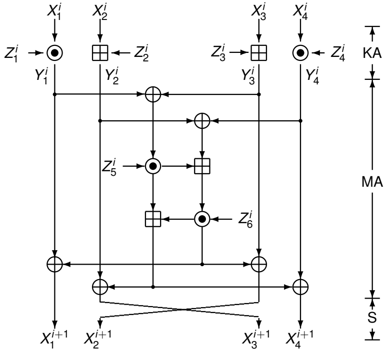
The DBL Modes: MJH-Double

Proposed by Lee and Stam (CT-RSA'11).



- f is an involution with no fixed point and $g \neq 0, 1$ is a constant.

IDEA Round Function



- 64-bit block, 128-bit key.
- Three operations: $\boxplus$, $\oplus$ and $\odot$.
- $a \boxplus b := (a + b) \bmod 2^{16}$.
- $a \odot b := (a \cdot b) \bmod (2^{16} + 1)$, 2^{16} as 0.
- With KA, MA, S, we have $C = KA \circ S \circ \{S \circ MA \circ KA\}^8(P)$.

Primitive Operations

When 0x0000 is mixed as subkey, $\boxplus$ can be removed. For mixing with $\odot$, since

$$\begin{aligned}
(a \odot 0) \bmod 2^{16} &= ((a \cdot 2^{16}) \bmod (2^{16}+1)) \bmod 2^{16} \\
&= (((a \cdot 2^{16} + a) + (2^{16}+1) - a) \bmod (2^{16}+1)) \bmod 2^{16} \\
&= (0 + 2^{16} + 1 - a) \bmod 2^{16} = 1 - a \bmod 2^{16} \\
&= 2 + (2^{16} - 1 - a) \bmod 2^{16} = (2 + \overline{a}) \bmod 2^{16}
\end{aligned}$$

and $\overline{a} = 0xffff \oplus a$, the diffusion is one way. There are many high probability differentials of the type $\delta \mapsto \delta$, for $\delta \in \mathbb{Z}_{2^{16}}$. E.g., $0x8000 \mapsto 0x8000$ with prob. 1.

The idea has been used by Daemen *et al.* (CRYPTO'93). When IDEA is keyed by the null-key, let $\Delta_{MSB} := (\delta_{MSB}, \delta_{MSB}, \delta_{MSB}, \delta_{MSB})$ where $\delta_{MSB} = 0x8000$, then we have a differential of probability 1:

$$\Delta_{MSB} \xrightarrow{\texttt{IDEA}_{K=0}} \Delta_{MSB}.$$

- The differential immediately allows free-start collisions on IDEA in Davies-Meyer mode, by setting $M = 0$.
- Free-start collisions as well for Hirose mode by setting $M = 0$ and $CV2 = 0$.
- Peyrin et al. (II) mode can be attacked if there is at least one $X \in \{CV1, CV2, M1, M2\}$ s.t. X is not used as key inputs in the 5 IDEA instances.
- Abreast-DM, Tandem-DM and MJH-Double cannot be attacked since null-key cannot be used on both instances.
- The differential probability remains close to 1 even if other higher bits in δ_{MSB} are active.
- Considering a collection of differentials in the form of $\Delta \mapsto \Delta$ where $\Delta = (\delta, \delta, \delta, \delta)$, we found the almost half-involution property.

Almost Half-involution

We show a special property of the null key (as a result, all subkeys are 0x0000).

$$\begin{aligned} C &= \mathrm{KA}_0 \circ \mathrm{S} \circ \{\mathrm{S} \circ \mathrm{MA}_0 \circ \mathrm{KA}_0\}^8(P) \\ &= \mathrm{KA}_0 \circ \mathrm{S} \circ \{\mathrm{S} \circ \mathrm{MA}_0 \circ \mathrm{KA}_0\}^3 \circ \mathrm{S} \circ \mathrm{MA}_0 \circ \mathrm{KA}_0 \circ \{\mathrm{S} \circ \mathrm{MA}_0 \circ \mathrm{KA}_0\}^4(P) \\ &= \underbrace{\mathrm{KA}_0 \circ \mathrm{MA}_0 \circ \{\mathrm{S} \circ \mathrm{KA}_0 \circ \mathrm{MA}_0\}^3}_{\sigma^{-1}} \circ \underbrace{\mathrm{KA}_0 \circ \mathrm{S}}_{\theta} \circ \underbrace{\{\mathrm{MA}_0 \circ \mathrm{KA}_0 \circ \mathrm{S}\}^3 \circ \mathrm{MA}_0 \circ \mathrm{KA}_0}_{\sigma}(P) \end{aligned}$$

If we write the encryption as $P \overset{\sigma}{\longleftarrow} U \overset{\theta}{\longrightarrow} V \overset{\sigma}{\longrightarrow} C$, then the *almost half-involution* property can be state as: for a pair of null-key encryptions that start from random plaintexts, $Pr[\Delta P = \Delta C]$ is around $2^{-16.26} \cdot 2^{-16}$.

The First Application

The almost half-involution property helps to find hash function collision of IDEA in Davies-Meyer mode by canceling ΔC with ΔP with the feed-forward.

We use two blocks M_0 and M_1, force $M_1 = 0$ to be the null-key block and randomize M_0. Hash collision can be found with around $2^{16.13}$ distinct message blocks of M_0.

This property also helps in finding improved results on the DBL hashing modes except Hirose mode.

Free-start Collisions for Abreast-DM and Tandem-DM

The idea is to force the null-key on one branch.

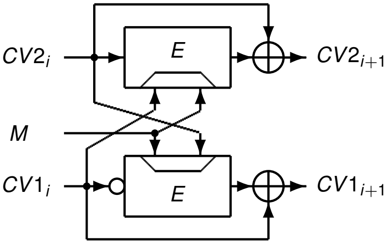


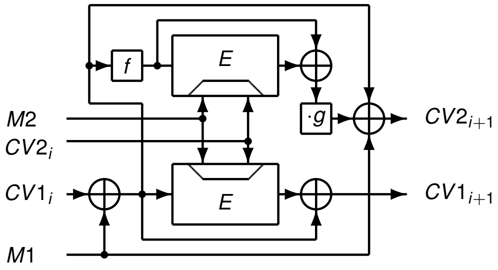
Figure: Abreast-DM

- Set $CV1 = 0$ and $M = 0$.
- Build $2^{48.13}$ distinct $CV2$.
- Check for collisions.

- The probability that a pair leads to a collision on the first (top) branch is $2^{-32.26}$.
- The probability that a pair leads to a collision on the second branch is 2^{-64}.

Semi-free-start Collision Attack on MJH-Double

The attacker may force the null-key for both branches.



- Set $CV2 = 0$ and $M2 = 0$.
- $CV1$ can be fixed as a challenge.
- Build $2^{32.26}$ distinct $M1$.
- Check for collisions.

Null-keyed IDEA as T-function

Used with a null-key, IDEA is a T-function (or triangular function), for which any output bit at position *i* depends only on the input bits of position *i* or lower.

- The primitive functions $\boxplus$ and $\oplus$ are both 16-bit T-functions.
- The modular multiplication $\odot$ is used only for subkey mixing. It is a T-function when the subkey is 0x0000.
- When IDEA uses the null-key, all the subkeys are 0x0000 and the encryption is a T-function.
- One can now search preimages by guessing the input words layer by layer.

Preimage Attack

We denote by

- p - the probability that given a random challenge, the attack algorithm outputs a preimage for this challenge.
- s - the average number of preimage solutions that the algorithm will output, given at least one is found.
- A - the average number of preimage solutions for each challenge. Then $A = p \cdot s$.

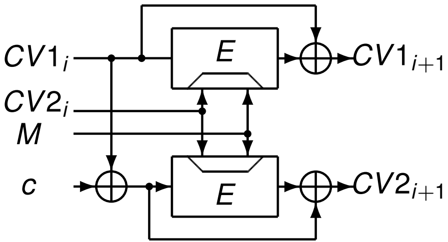
A generic attack restricted to C computations can generate $A = C \cdot 2^{-n}$ preimage solutions on average. We can thus consider that a preimage attack is found if we show an algorithm that outperforms this generic complexity.

Preimage Attack to IDEA in Davies-Meyer Mode

- Implemented as a recursive depth-first-search, from LSB to MSB of the four 16-bit state words.
- Wrong candidates are discarded as early as possible.
- We have $A = 1$ since the preimage space and image space are equal in size.
- We measure with 2^{32} random challenges that $p = 2^{-17.50}$.
- We can thus deduce that $s = A/p = 2^{17.5}$.
- For each of the 16 layers, 2^4 candidates are tried. Therefore, the total computations C to find s preimage solutions is bounded by $16 \cdot 2^4 \cdot s = 2^{25.5}$.
- A generic attack algorithm with $C = 2^{25.5}$ can only generate about $A = 2^{-38.5}$ solutions.

Preimage Attacks to DBL Modes

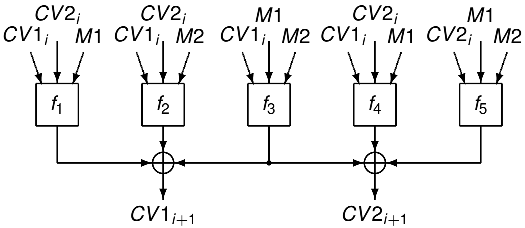
In the Hirose mode, we reuse the preimage attack to Davies-Meyer mode on one of the branches.



- Set $CV2 = 0$ and $M = 0$.
- Find preimage on the first (top) branch with a probability of $2^{-17.50}$.
- Use the $2^{17.5}$ solutions to match the second branch, with a probability of $2^{17.5-64} = 2^{-46.5}$.
- The attack has $A = 2^{-64}$ (since $p = 2^{-64}$ and $s = 1$) hence outperforms the generic attack with $A = 2^{-102.5}$.

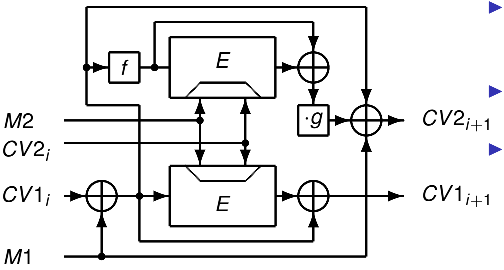
Abreast-DM and Tandem-DM can be attacked similarly.

Preimage Attacks to DBL Modes: Peyrin *et al.* (II)



If all of $CV1$, $CV2$, $M1$ and $M2$ appears in at least one `IDEA` key inputs in f_1, f_2, f_3 and at least one in f_3, f_4, f_5, then the attack cannot be applied. Otherwise, it can be attacked similarly to the Hirose case.

Preimage Attacks to DBL Modes: MJH-Double



- Set $CV2 = 0$ and $M2 = 0$. Find a preimage with $p = 2^{-17.5}$ for the bottom branch.
- The value of $M1 \oplus CV1$ is determined for this preimage.
- For each of the $s = 2^{17.5}$ preimages, $M1$ can be computed accordingly to make the top branch work as well.

- The attack has $A = 1$ and the generic attack has $A = 2^{-102.5}$ given that $C = 2^{25.5}$.

Conclusions

- Most of the constructions we considered are conjectured or proved to be secure in the ideal cipher model.
- Some ciphers, such as IDEA, have weak keys. Even a single weak key can be used to attack the block cipher based constructions.
- Our results indicate that one has to be cautious when hashing with a block cipher that presents any kind of non-ideal property (such as one or several weak keys) when the key is known or controlled by an attacker.
- Do not use IDEA for hashing purposes.

Q & A

Thank you !