Limited-Birthday Distinguishers for Hash Functions

Collisions Beyond the Birthday Bound can be Meaningful

Mitsugu Iwamoto[1], Thomas Peyrin[2] and Yu Sasaki[3]

1: The University of Electro-Communications, Japan

2:Nanyang Technological University, Singapore

3:NTT Secure Platform Laboratories, Japan

Asiacrypt 2013 (5/Dec/2013@Bengaluru)

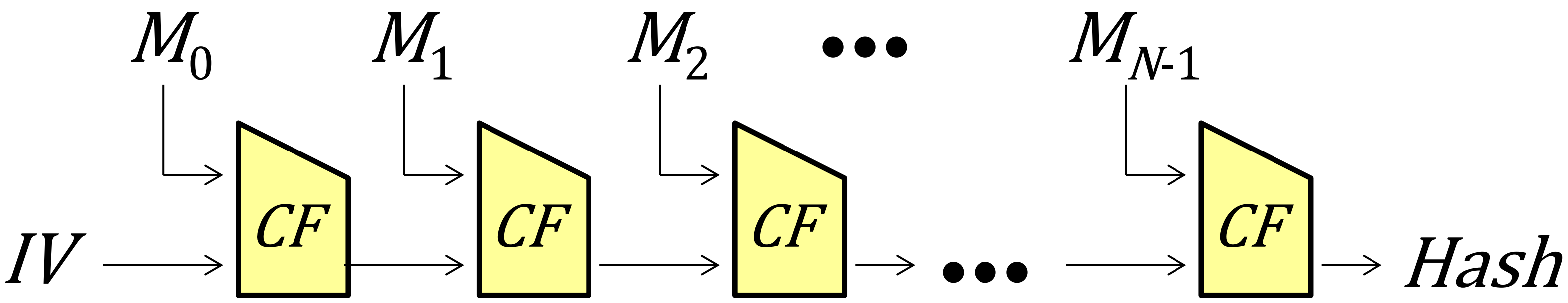
Research Summary

- Prove the generic attack cost of the LBD
  - the known generic attack [GP10] is optimal.
- LBD is useful
  - LBD for hash functions → breaking the dTCR notion.
- Constructing LBD on hash functions
  - Converting semi-free-start collisions (on the comp. func.) even with complexity beyond $2^{n/2}$.
- Find LBD for concrete designs
  - Some achieve the best attack for the hash setting: eg. RIPEMD128, Whirlpool

Hash Functions

- Hash Functions provide a fixed-size message fingerprint for arbitrary length message.

- Merkle-Damgård Construction



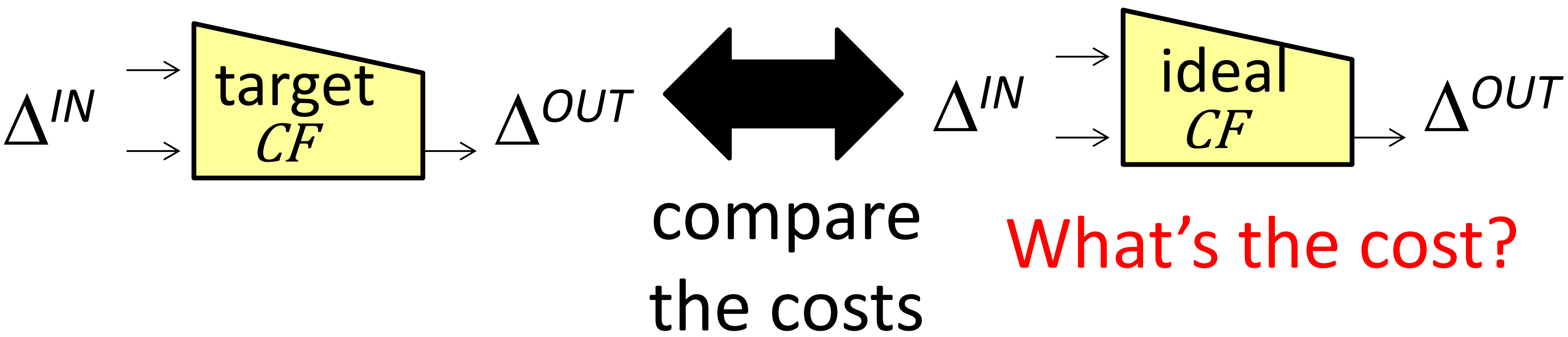
- Many schemes are proven to be secure by assuming the ideality of the underlying primitive.
  → Showing a non-ideality is important.

Limited Birthday Distinguishers (LBD)

- Recently, especially in the SHA-3 competition, many distinguishing attacks have been proposed.

  e.g. q-multi-coll., Rotational dist., subspace dist.

- Limited-Birthday Distinguisher [GP10] finds paired values satisfying the set of pre-specified input diffs Δ^{IN} and output diffs Δ^{OUT}.

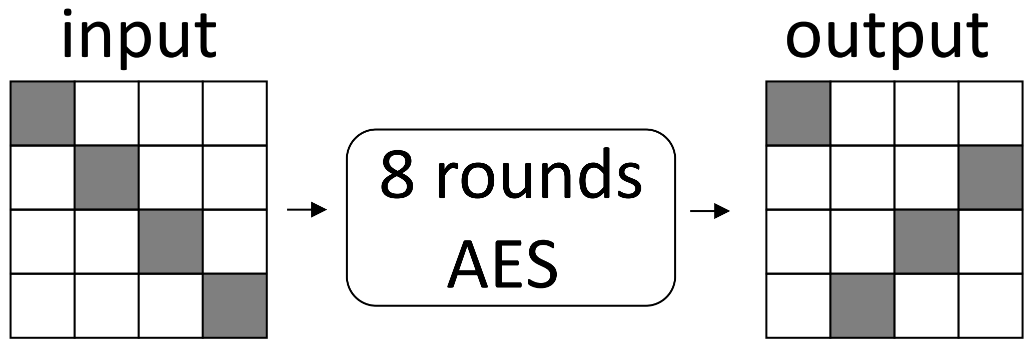


Known Generic Attack for LBD [GP10]

$n = 128$

$I = |\Delta^{\mathrm{IN}}| = 32$

$O = |\Delta^{\mathrm{OUT}}| = 32$



- Previous method conjectured to be the best
  - Fix 2^{n-I} inactive input bits
  - Choose all 2^{I} active input bits and make all (2^{2I-1}) pairs.
  - Repeat the above, by changing inactive input bits.

Theorem 1. *The limited-birthday attack complexity in [15]*

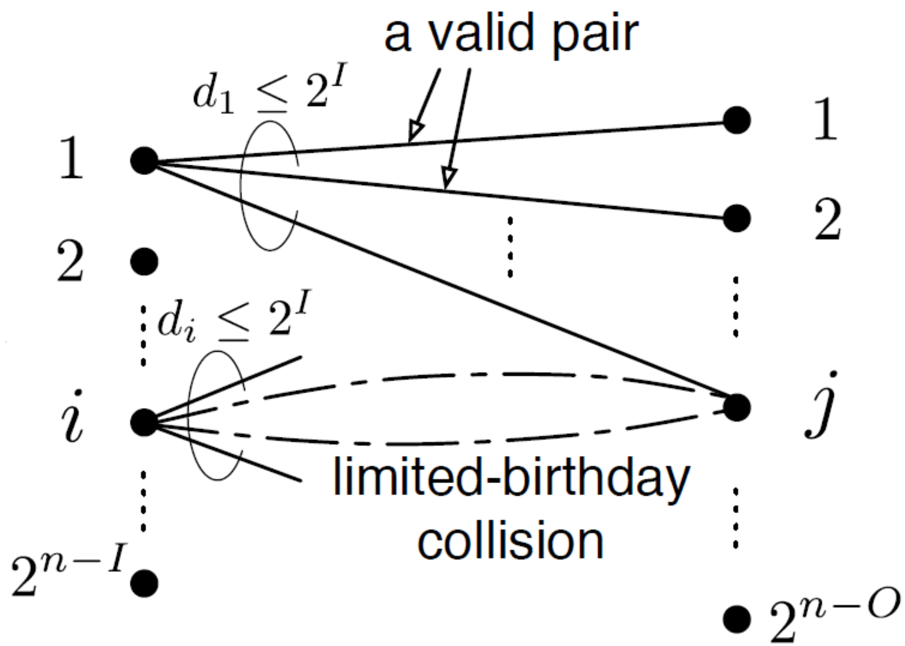
$$\max\left\{2^{\frac{n-O+1}{2}}, 2^{n-I-O+1}\right\}$$

Describing LBD with Bigraph

- Classify 2^n input values into $2^{n\text{-}I}$ groups indexed by non-active $n\text{-}I$ bits values. (Do the same for output.)
- Represent each input/output group by a nodes
- Represent the map from input to output by edges. Each input node can have 2^I edges in maximum.

Up to 2^I edges from each node

1 query to obtain 1 edge

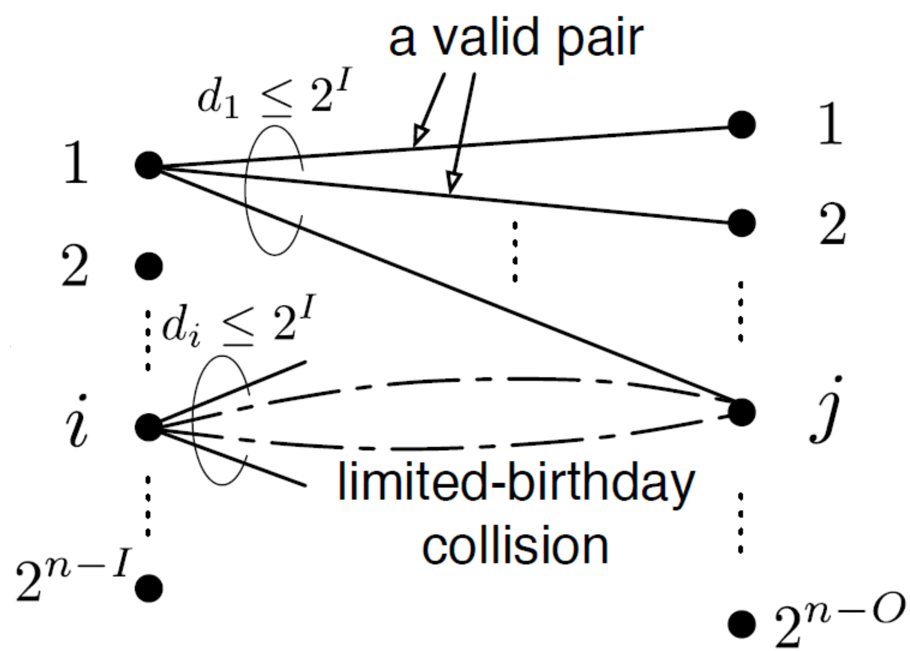


Describing LBD with Bigraph

- Achieving LBD is equivalent to find multiedges.
- *Valid pair*: a pair of edges sharing the same input node.
- If 2^{n-O} valid pairs are generated, multiedges will be found.

Up to 2^I edges from each node

1 query to obtain 1 edge



Describing LBD with Graph

- How many valid pairs can be generated with X queries?
- Suppose $d_i (1 \leq i \leq 2^{n-I})$ is the number of edges coming from the input node i.
- The number of valid pairs (#V) is:

$$\#\mathrm{V} = {d_1}^2/2 + {d_2}^2/2 + \ldots + d_{2^{n-I}}^2/2$$

- Constraint equations are:

$$\begin{cases} d_1 + d_2 + \ldots + d_{2^{n-I}} = X \\ 2^I \geq d_1 \geq d_2 \geq \ldots \geq d_{2^{n-I}} \geq 0. \text{ (Descendent order)} \end{cases}$$

Proof Approach

- Use the theory of majorization
- Proof is available in the paper.
- Interesting corollary: The proof can be extended to
  - limited-birthday multi-collisions
  - limited-birthday *k*-sums.

LBD for Hash Functions

- So far, LBD is mainly discussed only for a part of the hash function *i.e.*
  - underlying compression function
  - internal permutation

- We discuss LBD for the hash function *i.e.*
  - Fixed initial value
  - Δ^{IN} only exists in the input message before padding
  - Δ^{OUT} is defined on the hash digest

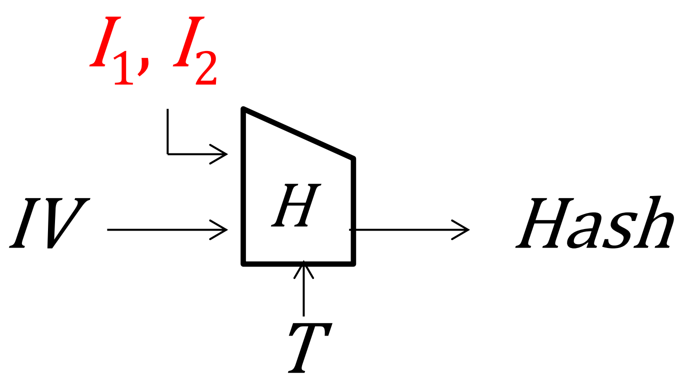
Applications of LBD for Hash Function

- Target collision resistance is a security notion for hash function with tweak value T.

Definition. *(Target Collision Resistance)*

The following attack must take 2^n cost.

- The adversary chooses an input value I_1.
- T is chosen without a control of the adversary.
- The adversary finds an input I_2 s.t. $H(I_1) = H(I_2)$.



A New Security Notion *dTCR*

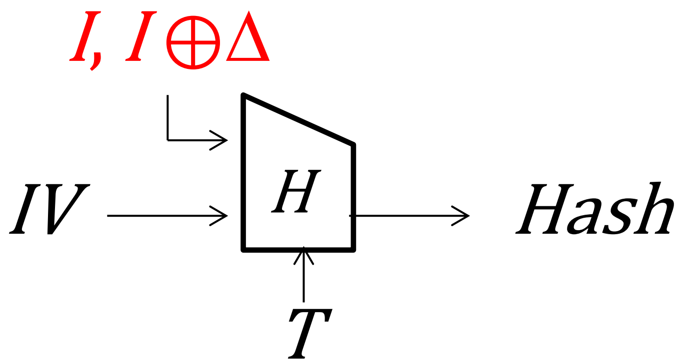
Definition. (*differential Target Collision Resistance*)

The following attack must take 2^n cost.

– The adversary chooses an input difference Δ.

– T is chosen without a control of the adversary.

– The adversary finds an input I s.t. $H(I) = H(I \oplus \Delta)$.

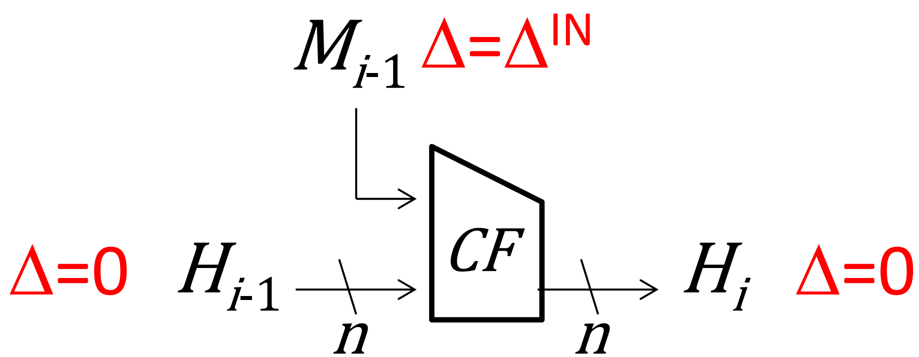


- A limited birthday distinguisher with $|\Delta^{IN}|=1$ and $\Delta^{OUT}=\{0\}$ immediately breaks the *dTCR* notion.

Converting Semi-Free-Start Collisions

- Semi-free-start collisions (on CF):

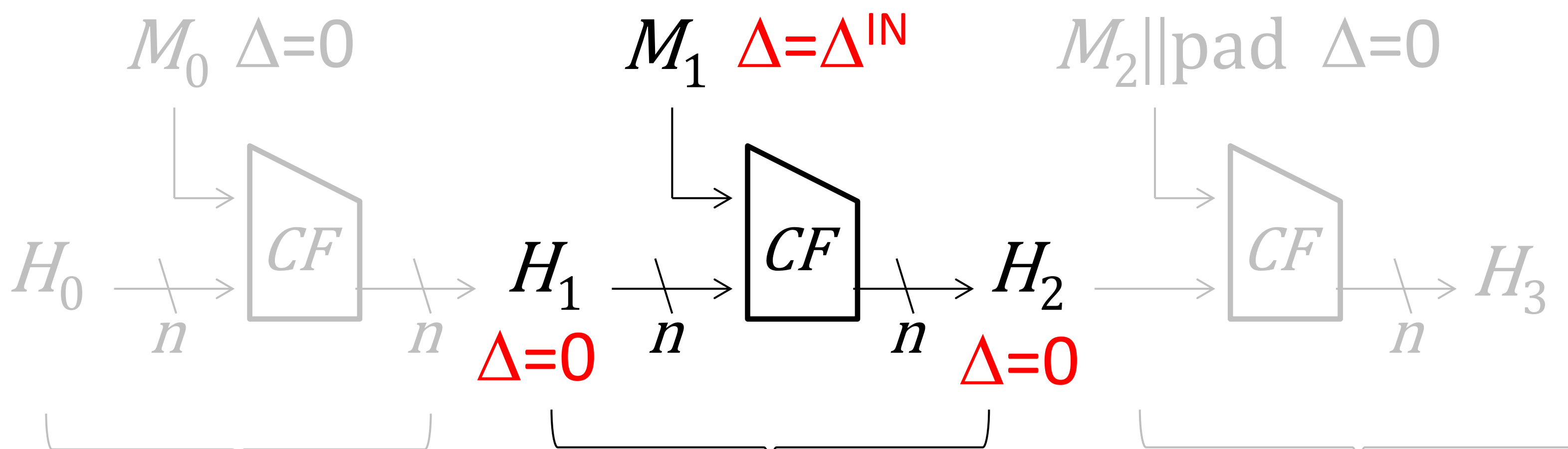
Find $(H_{i-1}, M_{i-1}, M'_{i-1})$ s.t. $CF(H_{i-1}, M_{i-1}) = CF(H_{i-1}, M'_{i-1})$



- In many cases, the input message difference Δ^{IN} is fixed in advance.
- This property is stronger than the collision attack with the birthday paradox.

Converting Semi-Free-Start Collisions

- 3-block LBD with Input difference $(0 \| \Delta^{\mathrm{IN}} \| 0)$
- Suppose the cost for semi-free-start coll is 2^x.



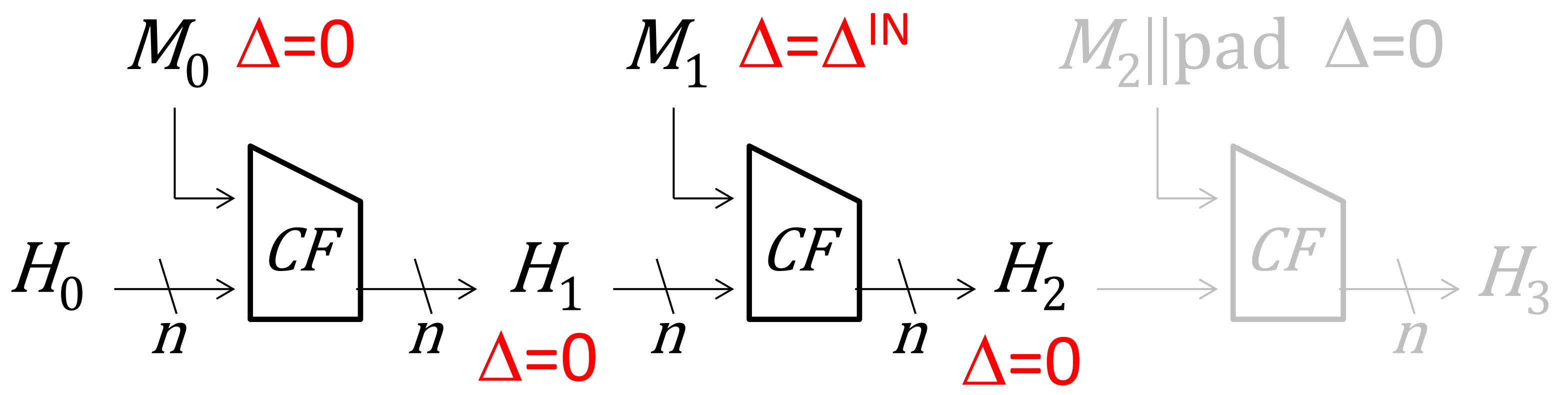
1. Generate $2^{(n-x)/2}$ semi-free-start collisions.

2. Generate $2^{(n+x)/2}$ random message blocks.

3. Collision is preserved for padding block.

Converting Semi-Free-Start Collisions

- 3-block LBD with Input difference $(0\|\Delta^{\mathrm{IN}}\|0)$
- Suppose the cost for semi-free-start coll is 2^x.



1. Generate $2^{(n-x)/2}$ semi-free-start collisions.

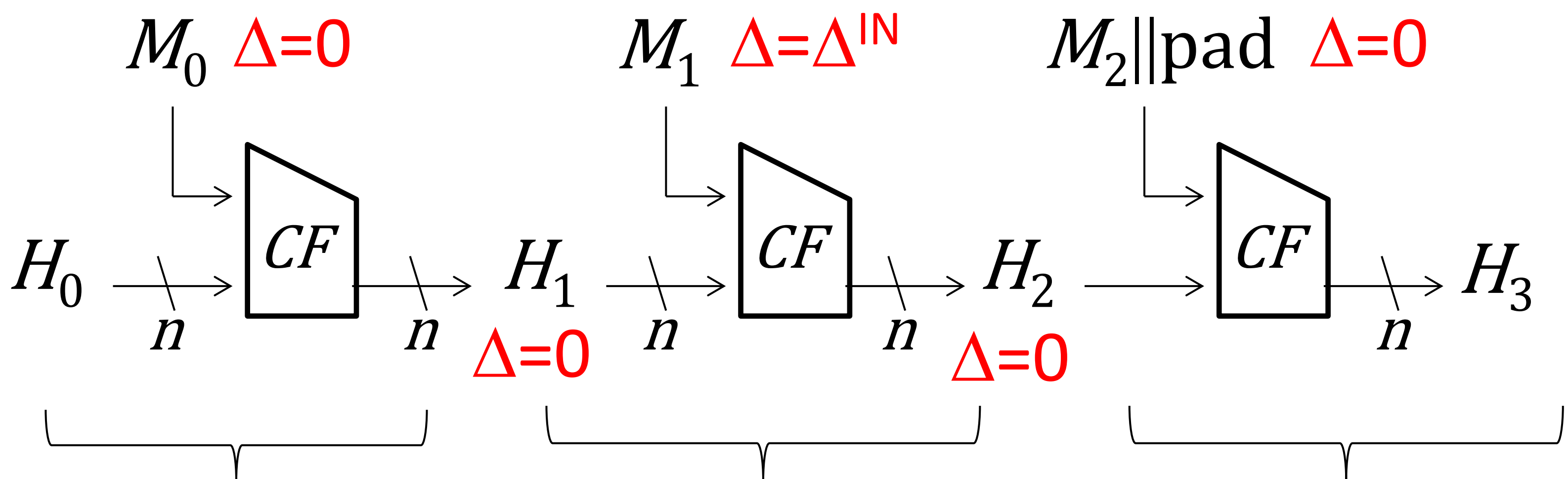
2. Generate $2^{(n+x)/2}$ random message blocks.

3. Collision is preserved for padding block.

NTT

Converting Semi-Free-Start Collisions

- 3-block LBD with Input difference $(0\|\Delta^{\mathrm{IN}}\|0)$
- Suppose the cost for semi-free-start coll is 2^{x}.



1. Generate $2^{(n-x)/2}$ semi-free-start collisions.

2. Generate $2^{(n+x)/2}$ random message blocks.

3. Collision is preserved for padding block.

Remarks for Conversion Method

- The attack complexity is $2^{(n+x)/2+1}$. Semi-free-start collisions with comp. beyond $2^{n/2}$ can be a valid LBD.
- Can be extended to (not too) wide-pipe, e.g. SHA224
- Be careful for the freedom degrees of the semi-free-start collision attack. Sometimes, generating $2^{(n-x)/2}$ of them is impossible.
- Can be extended to limited-birthday near-collisions (Δ^{OUT} can be other than {0}).
  - Differential path construction becomes easier.
  - Padding must be satisfied within the second block.

Applications to Concrete Designs

target	rounds	time	memory	type
AES-DM hash func.	7/10	2^{125}	2^{8}	preimage attack
AES-DM hash func.	6/10	2^{113}	2^{32}	limited-birthday dist.
AES-MP hash func.	7/10	2^{120}	2^{8}	2nd preimage attack
AES-MP hash func.	6/10	2^{89}	2^{32}	limited-birthday dist.
HAS-160 hash func.	68/80	$2^{156.3}$	2^{15}	preimage attack
HAS-160 hash func.	65/80	2^{81}	2^{80}	limited-birthday dist.
● LANE-256 hash func.	full	2^{169}	2^{88}	limited-birthday dist.
● LANE-512 hash func.	full	2^{369}	2^{144}	limited-birthday dist.
RIPEMD-128 hash func.	full	$2^{105.4}$	negl.	limited-birthday dist.
● RIPEMD-128 hash func.	full	$2^{95.8}$	$2^{33.2}$	limited-birthday dist.
SHA-256 hash func.	42/64	$2^{251.7}$	negl.	preimage attack
SHA-256 hash func.	38/64	2^{129}	2^{128}	limited-birthday dist.
Whirlpool hash func.	6/10	2^{481}	2^{256}	preimage attack
● Whirlpool hash func.	7/10	2^{440}	2^{128}	limited-birthday dist.

● : best attack in the hash function setting

Concluding Remarks

- Prove the optimality of the generic attack for LBD.
- LBD on hash functions can be used to attack the new security notion “differential-TCR”.
- LBD on hash functions can be constructed from semi-free-start collisions even with complexity beyond $2^{n/2}$.
- Apply the above conversion for several hash functions. Some achieved the best attack.

Thank for your attention !!

Concluding Remarks

- Prove the optimality of the generic attack for LBD.
- LBD on hash functions can be used to attack the new security notion “differential-TCR”.
- LBD on hash functions can be constructed from semi-free-start collisions even with complexity beyond $2^{n/2}$.
- Apply the above conversion for several hash functions. Some achieved the best attack.

Thank for your attention !!