Some applications of the Biham-Chen attack to SHA-like hash functions

CRYPTOGRAPHIC HASH WORKSHOP

NIST (Green Auditorium)

Gaithersburg, Maryland

Monday, October 31, 2005

Hirotaka Yoshida

Systems Development Laboratory, Hitachi, Ltd., Japan

Alex Biryukov, Bart Preneel

Katholieke Universiteit Leuven, Belgium

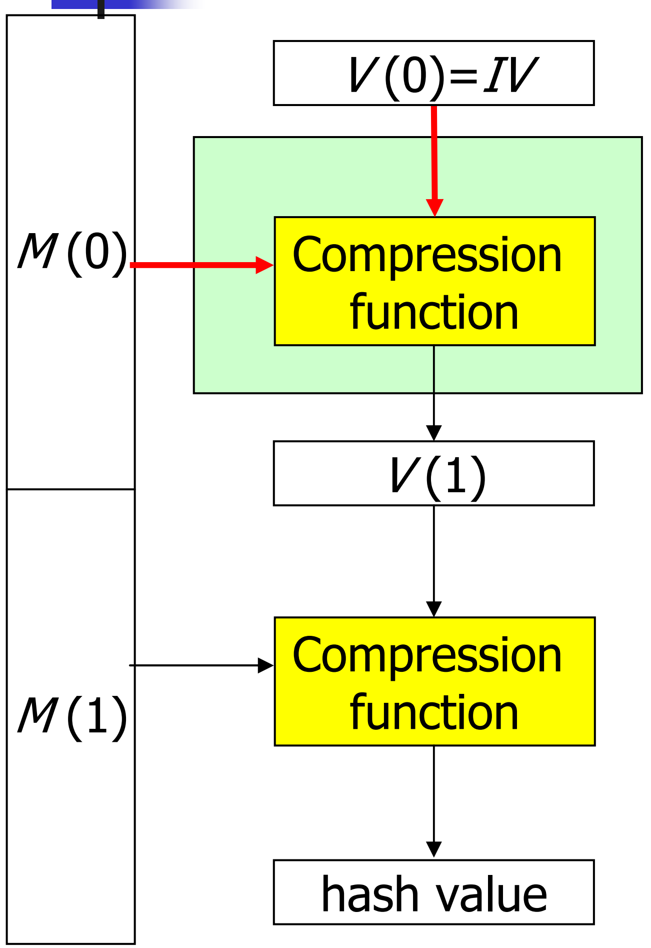
Overview of the talk

- Introduction to some resistance of hash functions
- Description of the Biham-Chen attack
- Cryptanalysis of hash functions in encryption mode
- Pseudo-collision attack on MD5
- Pseudo-collision attack on a SHA-256 variant
- Observation on SHA-256
- Conclusions

Some resistance of hash functions

- Near-collision resistance
  - Resistance against attacks finding a pair of hash values which differ in only small number of bit positions.
- Pseudo-collision resistance
  - Resistance against collision attacks where different initial vectors can be chosen.
- Pseudo-randomness
  - Indistinguishability from a random function.

Pseudo-collision resistance



MD-construction

- Resistance when 2 inputs controlled.
- Important in the theory of the MD-construction
- There could be some application which requires the underlying hash function to have this resistance
    - Knudsen *et al*, Preimage and pseudo-collision attack on MD2, FSE2005

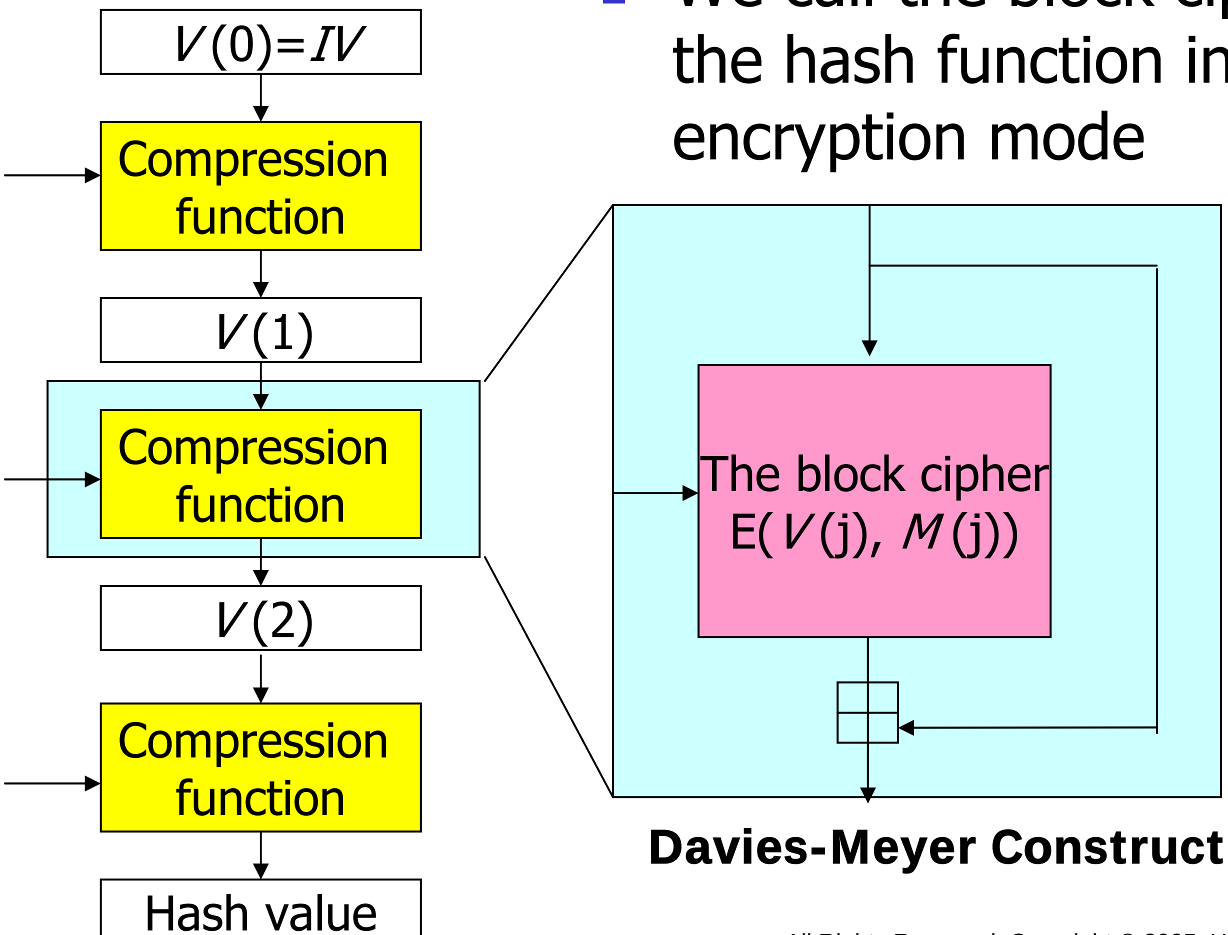
Biham and Chen attack

- Near-collision attack on SHA-0
  - Biham and Chen, near-collision of SHA-0, CRYPTO 2004

- Start collision search from some intermediate round *r*

- Introduce new technique called <u>neutral bits</u> to optimize attack complexity
  - Neutral bits do not affect the difference for *r* rounds
  - Use $2^{k(r)}$ messages generated from $k(r)$ neutral bits
  - Using this messages gives a better probability for *r* rounds than probability when using randomly chosen messages

Hash Function in encryption mode



- We call the block cipher *E* the hash function in encryption mode

SHA-256 in encryption mode was proposed in 2000 by Handschuh and Naccache and named SHACAL-2

Davies-Meyer Construction

Cryptanalysis of Hash functions in Encryption Mode

- Differential cryptanalysis of SHA-1
  - Handschuh *et al.*, SHACAL, Submission to the NESSIE project, 2000.
- Slide attack on SHA-1 and pseudo-collision attack on MD5
  - Saarinen, Cryptanalysis of Block Ciphers Based on SHA-1 and MD5, FSE2003.
- Attack which distinguishes HAVAL from a random function.
  - Yoshida *et al.*, Non-randomness of the Full 4 and 5-pass HAVAL, SCN2004.
- Attack on 32-round SHACAL-2 by Shin *at al* at ACISP 2004

Differential Cryptanalysis of a Hash Function in Encryption Mode

- Differential characteristic defines the expected differences $d(Y_i, Y_i)$ in each round.

- Definition

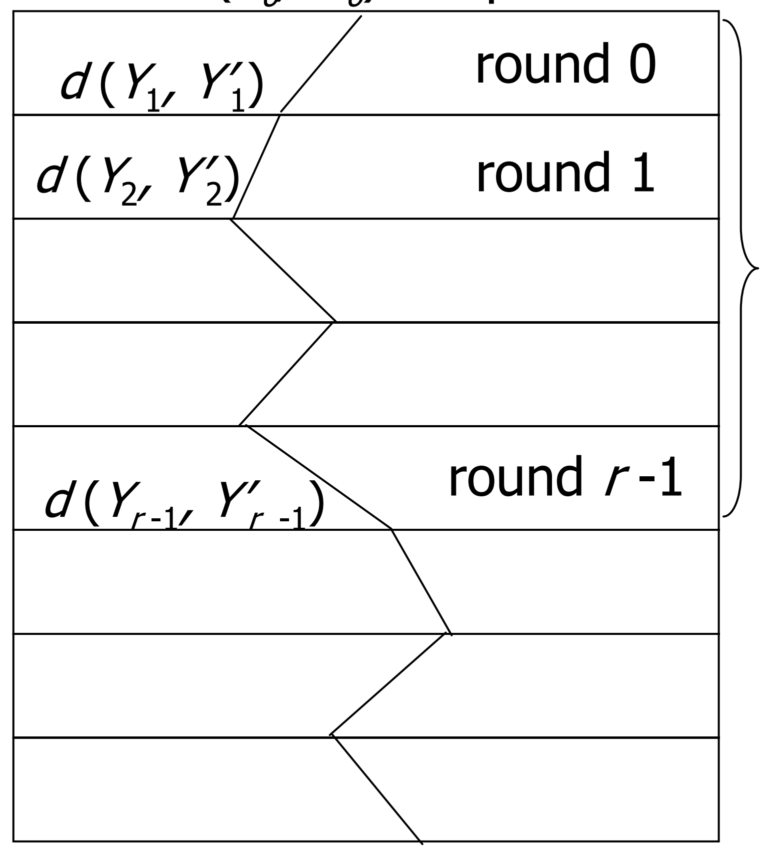
  A pair of plaintexts (P, P') <u>conforms</u> to the Differential characteristic if the differences at the output of the first r rounds are as expected.

- Assumption

  Differential characteristic has already been found

Differential characteristic

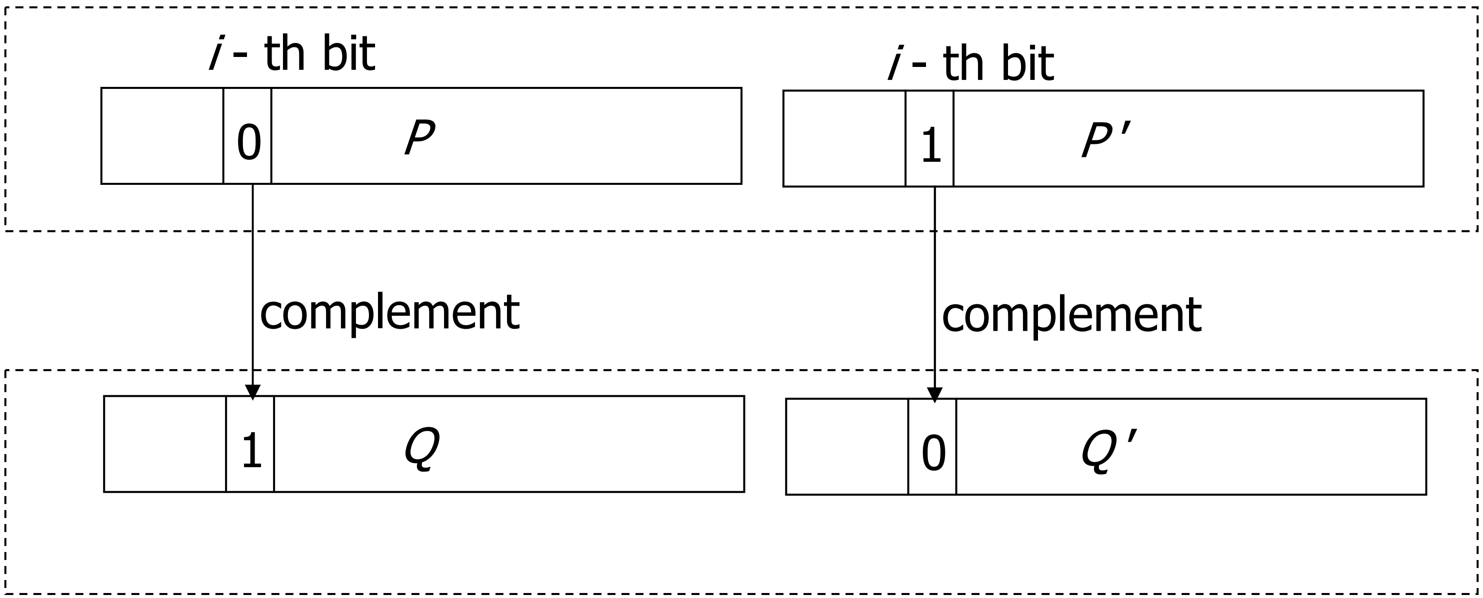
$d(Y_0, Y'_0)$= Input difference



$d(Y_R, Y'_R)$ = output difference

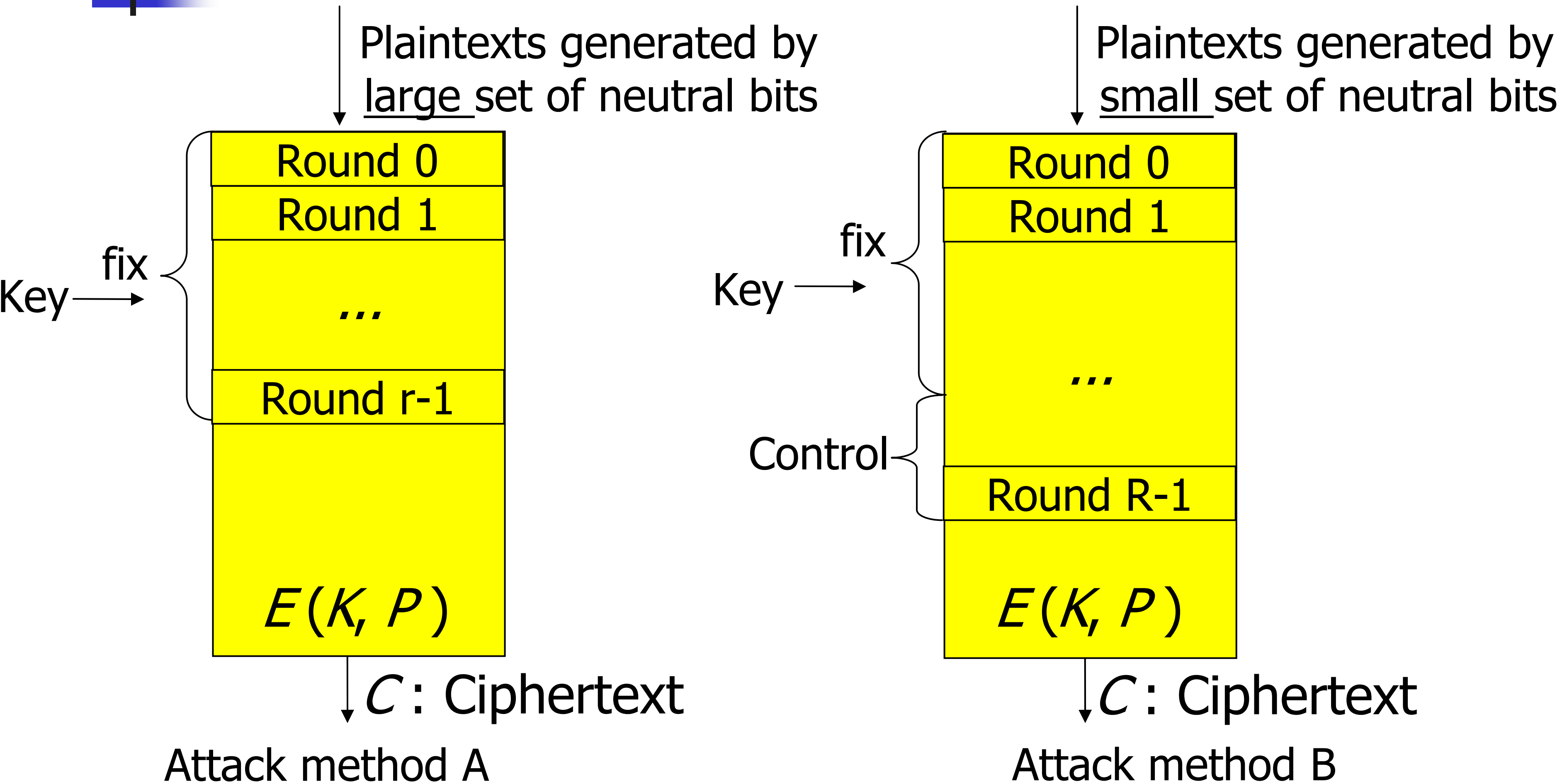
Neutral bit in case of Hash Functions in Encryption Mode

Assume that (P, P') conforms to some differential characteristic



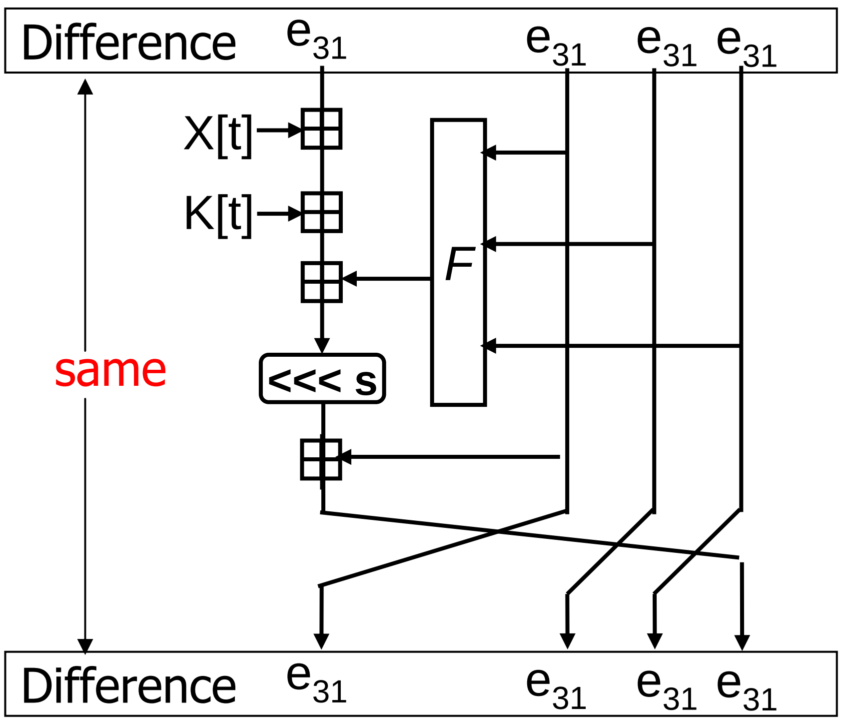
If (Q, Q') conforms to the differential characteristic, the *i* -th bit is called <u>neutral bit.</u>

Differential Cryptanalysis of a Hash Function in Encryption Mode

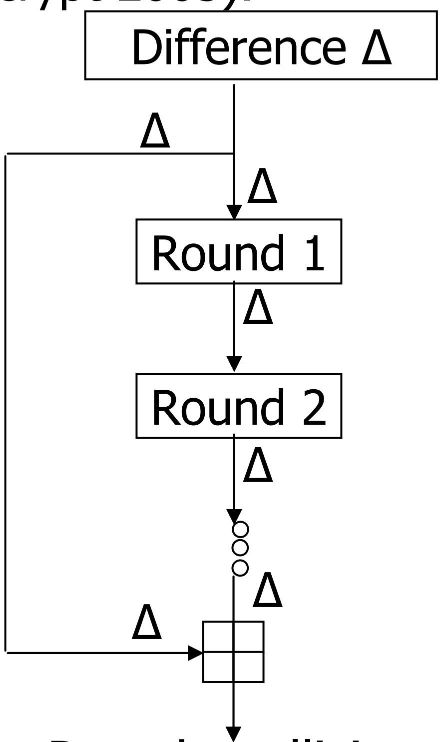


Application Biham-Chen attack to MD5 with Saarinen's characteristic

- Attacks on MD5
  - Pseudo-collision attacks (Dobbertin., Eurocrypt '96 rump session)
  - <u>Pseudo-collision attacks</u> (<u>Saarinen</u>, FSE 2003)
  - Attacks for finding collisions (Wang *et al.*, Eurocrypt 2005).



Saarinen's characteristic



Pseudo-collision

Experimental results on MD5

- Method A found a pseudo-near collision for MD5 with complexity 2^{42} which differs only in 1 bit position
- Method B found a pseudo-collision for MD5 with complexity 2^{39}
  - Probability of characteristic obtained from neutral set is about 2^{-39}, which is 2^{9} higher than original probability.

Original differential characteristic

Improved differential characteristic

Round	Original		Improved	
1^{st} round	2^{-16}	Improved!	2^{-7}	
2^{nd} round	2^{-16}		2^{-16}	
3^{rd} round	1		1	
4^{th} round	2^{-16}		2^{-16}	

Pseudo-collision Overall prob. 2^{-48} Improved! Overall prob. 2^{-39} Pseudo-collision

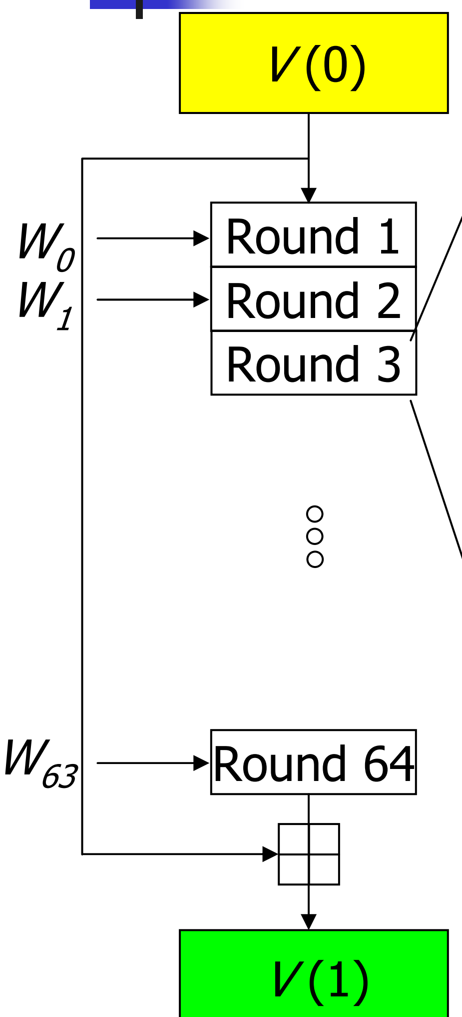
Introduction to SHA-256

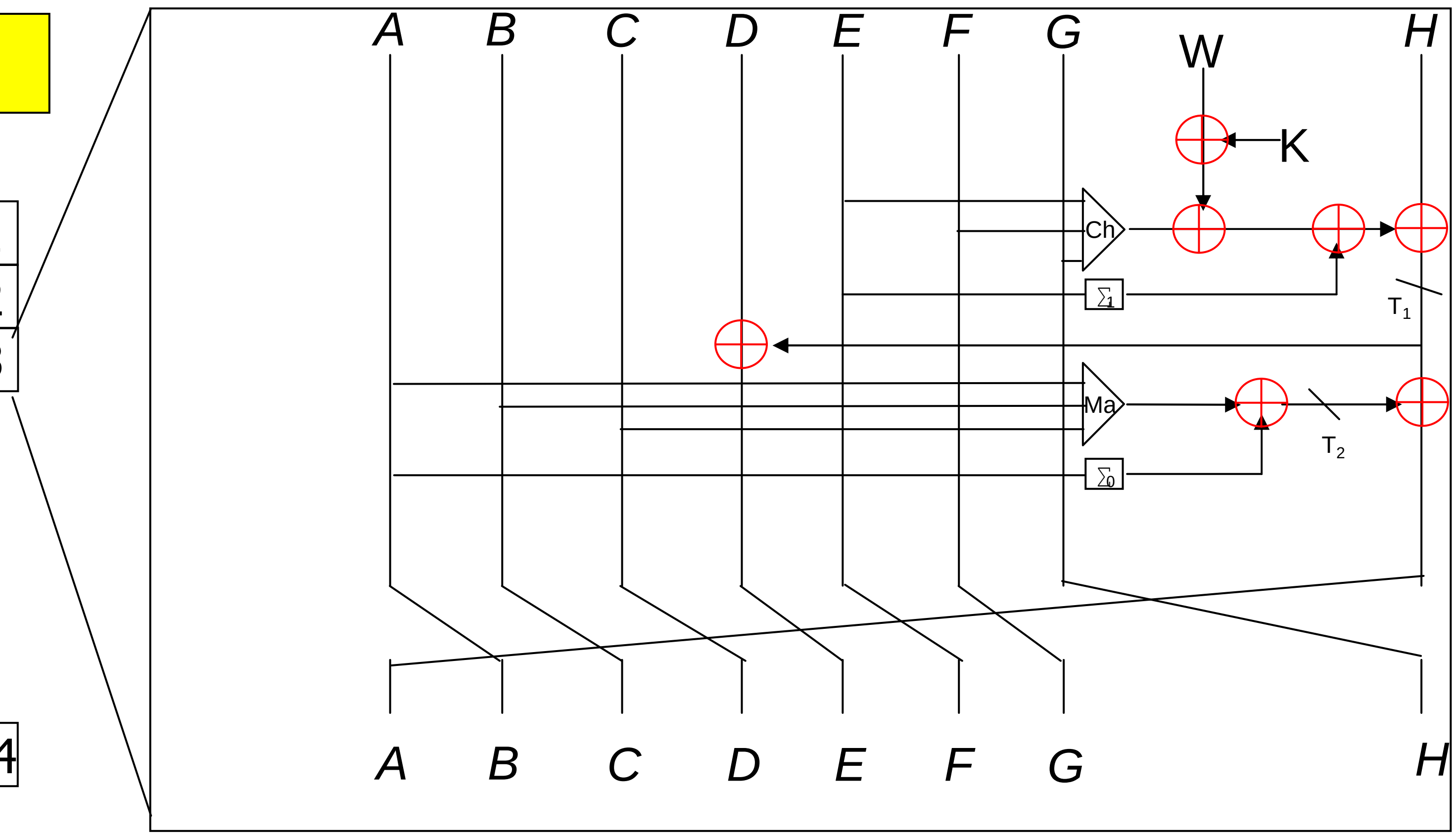
- Proposed in 2000 by NIST
  - Adopted as FIPS standard in 2002

- Resistance against known attacks studied
  - Security report at SAC 2003 by Gilbert and Handschuh
  - Property related to Chabaud-Joux attack by Hawkes *et al* in 2004

- Pseudo-collision attack on variant of SHA-256 with 34 rounds demonstrated at SAC 2005 by Yoshida and Biryukov

SHA-2-XOR, variant of SHA-256



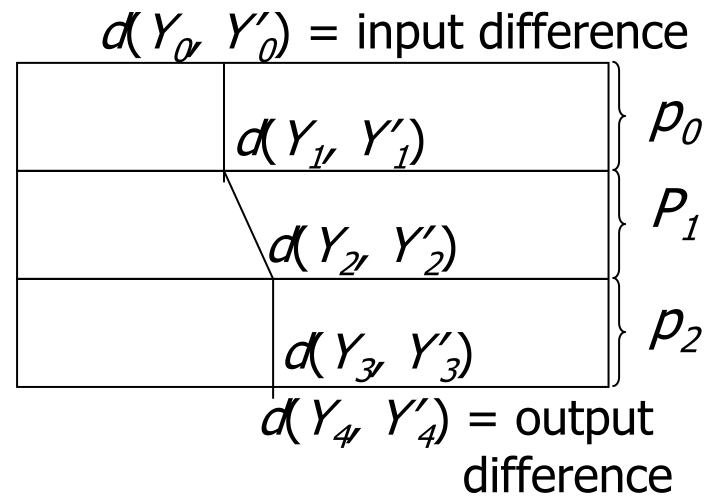


- Simplify SHA-256 by replacing ADD by XOR

Pseudo-collision attack on SHA-2-XOR at SAC 2005

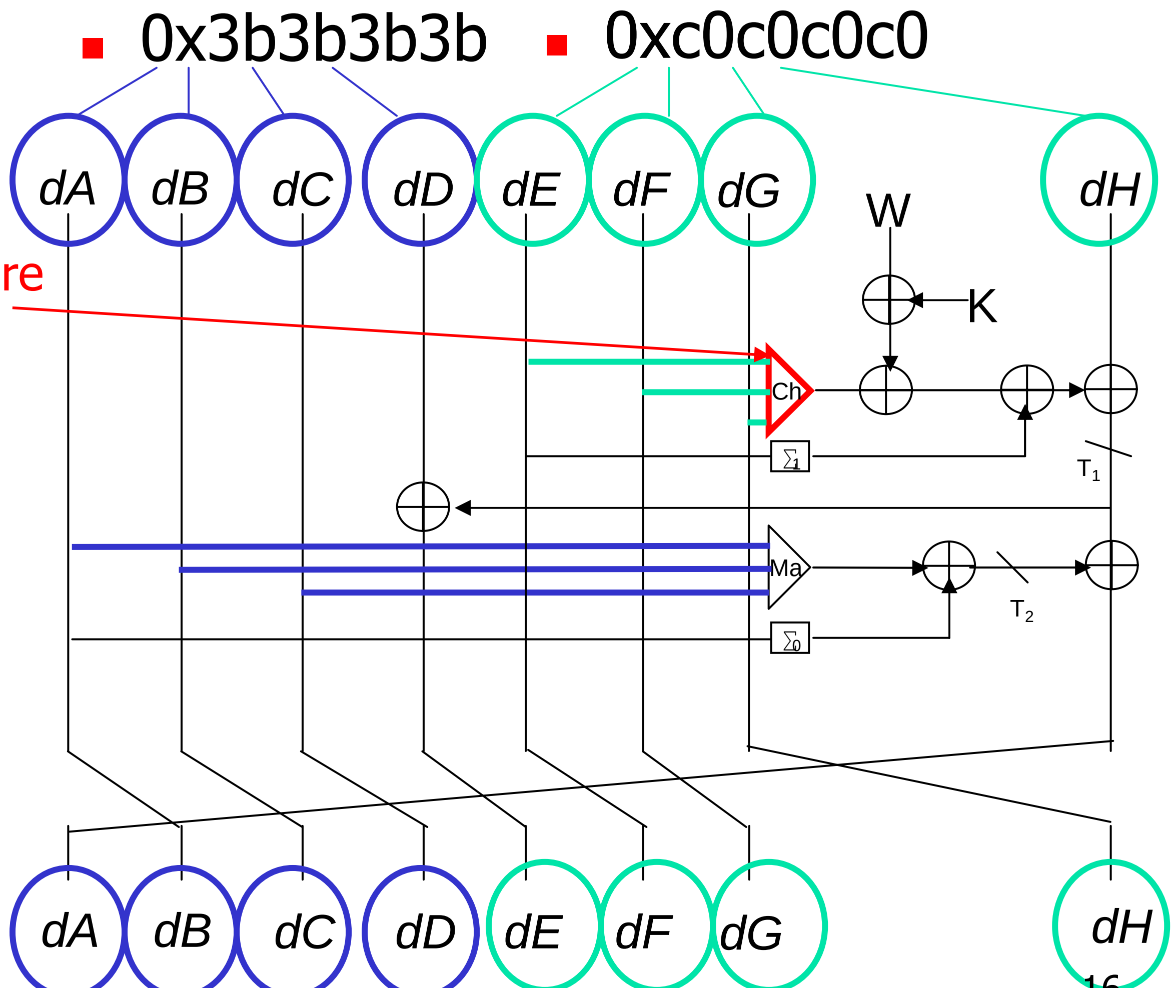
- Differential cryptanalysis
  - Biham, Shamir, Differential Cryptanalysis of the Data Encryption Standard, 1993.
  - The aim is to find <u>differential characteristics</u> for the whole cipher.

<u>A differential characteristic</u>



- Input modification
  - Select input values that follows the characteristic with probability 1 in the first several(or many) rounds.
    - Rijmen and Preneel, Improved characteristics for differential cryptanalysis of hash functions based on block ciphers, FSE 94
    - Wang *et al*, Cryptanalysis of the hash functions MD4 and RIPEMD, Eurocrypt 2005

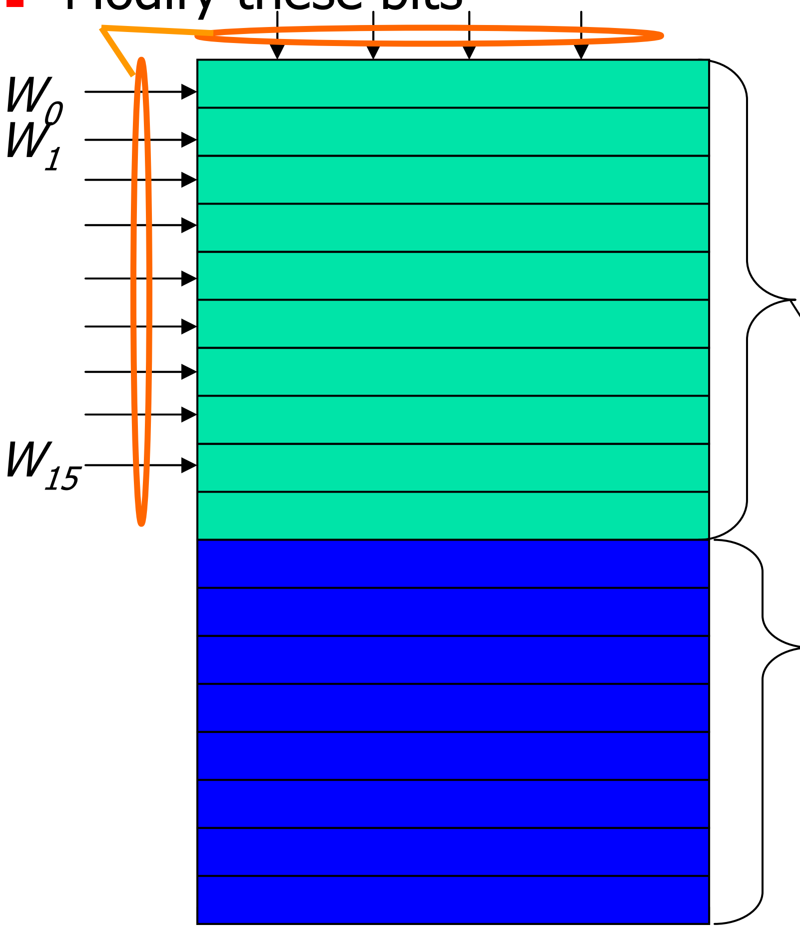
The best one-round iterative characteristics



- Best probability 2^{-8}
  - Properties used:
    - CH(0,0,0) = 0
    - CH(1,1,1) = 0/1 with probability 1/2
- MJ behaves <u>linearly</u>

Input modification

- Modify these bits



- Modify some bits of register values and $W_0, W_1,\ldots,W_{15}$ to ensure the following 152-bit conditions to hold:
  - For j =2,3,10,11,18,19,26,27
    - $(F_0 \oplus G_0)^{(j)}$ =0x08080808$^{(j)}$
    - $(E_0 \oplus F_0)^{(j)}$ =0x08080808$^{(j)}$
    - $(E_t \oplus E_{t+1})^{(j)}$ =0x08080808$^{(j)}$, t = 1,2...,16
- 19-round charactericstic with probability 1
- 15 round charactericstic with prob. 2^{-120}
- **34-round pseudo-collision with prob. 2^{-120} !!**

Experimental results on SHA-2-XOR

- Probability of the 10-round characteristic obtained from set of neutral bits is about $2^{-23.7}$, which is 2^{56} higher than the original probability.
  - The size of set is 27, $r = 7$.
- In practice, we found 10 pseudo-collisions for 10 rounds with complexity 2^{27}

Plaintext pair which produces a pseudo collision for 10 rounds:

Q =	0x4939a45a 0x79ec4172 0xf0ef52a9 0xa8161bbe 0xd92f76e4 0x21962dfe 0xd88e6416 0xfac1edb2
Q’=	0xfa8a17e9 0xca5ff2c1 0x435ce11a 0x1ba5a80d 0xd5237ae8 0x2d9a21f2 0xd482681a 0xf6cde1be

Theoretical results on SHA-2-XOR

- We can use 768 bits of input
- How many input bits we have used so far and will be able to control to add rounds?
- In order to obtain 10-round pseudo-collisions, what we did:
  - Fix each of the words W_0, W_1, …, W_6 to 0
  - Use the 2-neutral set of size 27
- We use 7 * 32bits to construct 10-round pseudo-collisions, therefore we can control the message words, W_7, W_8, …, W_{15} (=freedom of 9 * 32 bits.) to add 13 rounds.
- We find a pseudo-collision for 22-rounds of SHA-2-XOR with complexity 2^{120}

Comparison with previous attack

	Differential path	Optimization technique	# of rounds	Complexity
Pseudo-collision attack on SHA-2-XOR at SAC2005	One-round iterative	Input modification	34	2^{120}
Pseudo-collision attack on SHA-2-XOR in this talk	Same as above	Neutral bits	22	2^{120}

- Both attacks are based on one-round iterative differential characteristic whose Hammming weight iterative is high
- Unlikely to obtain a high probability for the same characteristic in SHA-256 as in SHA-2-XOR
- Not possible to apply both attack to actual SHA-256 in a straightforward way

Observation on real SHA-256

- The previous result:
- Attack on 32-round SHACAL-2 by Shin *et al* at ACISP 2004
- This is based on a 14-round truncated differential characteristic
  - Associated probability 2^{-32} which has been improved to $2^{-18.7}$ by:
  - Fixing some bits of plaintext pairs
  - Constructing multiple differential characteristics.

- Our results:
- We found a plaintext-pair with set of 20 neutral Bits for $r = 5$
- This set gave us a probability $\underline{2^{-8.01}}$ for the 14-round truncated characteristic, $\underline{2^{10}\text{ higher}}$ than previous probability $2^{-18.7}$

Conlusions

- We discussed some resistance and tried to apply the Biham-Chen attack to study well-known hash functions.
- Some improved results on MD5 and a SHA-256 variant were presented.
- The generic approach here may find interesting results on hash functions for which differential characteristics have been already found.

Some applications of the Biham-Chen attack to SHA-like hash functions *

Hirotaka Yoshida[1], Alex Biryukov[2], and Bart Preneel[2]

[1] Systems Development Laboratory, Hitachi, Ltd.,
1099 Ohzenji, Asao-ku, Kawasaki-shi, Kanagawa-ken, 215-0013 Japan
hyoshida@sdl.hitachi.co.jp
[2] Katholieke Universiteit Leuven, Dept. ESAT/SCD-COSIC
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium
{abiryuko,preneel}@esat.kuleuven.ac.be

Abstract. Biham and Chen proposed an attack on SHA-0 at Crypt 2004. In this paper, we apply the Biham-Chen attack to analyze SHA-like hash functions regarding pseudo-collision resistance and pseudo-randomness. We present a scenario about how to attack SHA-like hash functions applying the Biham and Chen attack. Using this scenario, we present a differential attack on the MD5 hash function and a differential attack on a variant of SHA-256 hash function. We also study certain several rounds' property of the real SHA-256 function.

Keywords: Differential attack, Pseudo-collision resistance, Pseudo-randomness.

1 Introduction

A cryptographic hash function is an algorithm that takes input strings of arbitrary (typically very large) length and maps these to short fixed length output strings.

Biham and Chen proposed an attack on the SHA-0 hash function at Crypt 2004 [1]. The attack seems to have a great influence on the future analysis and design of hash functions because it uses very generic technique which is the use of neutral bits. It is quite natural and interesting to apply this technique to other hash function whose structure and boolean functions are similar to SHA-0, such as the MD-family and the SHA-family for the next generation, which we call SHA-like hash functions.

In this paper, we apply the Biham-Chen attack to analyze such hash functions regarding pseudo-collision resistance and pseudo-randomness. The importance of pseudo-collision resistance is related to collision resistance. *Pseudo-collision* resistance is a resistance against finding a collision obtained from more relaxed condition that different initial vectors can be chosen. Pseudo-collision resistance has

* This work was supported in part by a consignment research from the National Institute on Information and Communications Technology (NiCT), Japan. This work was supported in part by the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government.

a particular importance for a hash function constructed by the MD-construction because in this case pseudo-collision resistance for the hash function can be translated into collision resistance for its compression function. The theory of the MD-construction, on which the securities of many popular hash functions rely, does not guarantee collision resistance for a hash function without pseudo-collision resistance for its compression function. It has also pointed out at FSE 2005 [12] that pseudo-collision resistance has an importance in some application.

The outline of this paper is as follows. In Section 5.1, we give a brief description of the Biham and Chen attack. In Section 3 we present an scenario how to attack SHA-like hash functions applying the Biham and Chen attack. Using this scenario, we present a differential attack on the MD5 hash function. in Section 4 and a differential attack on a variant of SHA-256 hash function in Section 5 where we also study the several rounds' property of the real SHA-256 function. Our conclusions are given in Section 6.

2 Description of the Biham and Chen attack

In this section, we give a brief description of the Biham and Chen attack. What the attack does is that for a given differential characteristic, the attack improves its probability by starting the collision search from some intermediate round r. The attack uses so called neutral bits which do not affect the difference for r rounds. If the attacker obtains $k(r)$ neutral bits, he can generate a set of $2^{k(r)}$ messages. For the characteristic from round 0 to $r-1$, using this set gives us a better probability than the probability when using a set of randomly chosen messages.

The Biham and Chen attack significantly reduces the complexity of the attack on SHA-0 by Chabaud and Joux and allowed to find near-collisions for the SHA-0 [1].

The Biham-Chen technique of neutral bits is a special case of the definition of "good pair oracle" (or space oracle) given in [6]. The difficulty in the case of block ciphers is that the attacker can not control the key and thus typically can not gain first rounds for free unless there is a property that holds with probability one for all the keys. In the case of hash functions the attacker may control the key (the message) and thus may prepare a large set of pairs with guaranteed propagation of the differences in the initial rounds. In the case of block ciphers similar phenomenon can be exploited only once the attacker has identified the first good pair for the full cipher.

3 Differential Cryptanalysis of hash functions in encryption mode

Any compression function of SHA-like hash function is constructed from a block cipher denoted by $E(K,P)$, using the Davies-Meyer mode. Therefore we obtain a block cipher $E(K,P)$ from such a compression function if the Davies-Meyer chaining is peeled off.

Several cryptanalytic techniques ranging from differential cryptanalysis [4] to slide attacks [5] have been applied to study the security of well-known hash functions in encryption mode. For example, differential cryptanalysis of SHA-1 has been shown in [11]. A slide attack on SHA-1 and an attack on MD5 which finds one high-probability differential characteristic were given in [16]. The strongest version of the HAVAL hash function in encryption mode was shown to be non-random[19].

In this paper, we apply the Biham-Chen attack to hash functions in encryption mode. First we assume that a differential characteristic Δ for the n-bit block cipher $E(K, P)$ has been already found and the key value K is fixed to one value $K = K_0$, we make the following definitions:

Definition 1. *The differential characteristic Δ defines the expected differences δ of the values of registers in each round. We say that a pair of plaintexts conforms to δ_r if $E_i(K_0, P) \oplus E_i(K_0, P') = \delta_i$ for every $i \in \{1, \ldots, r\}$, where $E_i(K_0, P)$ consists of the first i rounds of $E(K_0, P)$.*

Definition 2. *Let P and P' be a pair of plaintexts that conforms to δ_r for some r. We say that i-th bit of the plaintexts is a neutral bit with respect to P and P' if a pair of the plaintexts received by complementing the i-th bits of P and P' also conform to δ_r. We say that the pair of the i-th bit and j-th bit of the plaintexts is neutral with respect to P and P' if all the pairs of the plaintexts received by complementing the any subset of these bits $\{i\}, \{j\}, \{i, j\}$-th bits of P and P' also conform to δ_r. We say that a set of bits $S \in \{0, ..., n-1\}$ is neutral with respect to P and P' if all the pairs of the plaintexts received by complementing the any subset of the bits in S in both plaintexts P and P' also conform to δ_r. We say that a subset $S \in \{0, ..., n-1\}$ of bits of the plaintexts is 2-neutral with respect to P and P' if every bit $\in S$ is neutral, and every pair of bits in S is also neutral.*

In the Table 1, we show an algorithm for finding a 2-neutral set which we will use in the following in section. In this algorithm, we say that there is an edge between two bits, i-th bit and j-th bit if the pair of these bits is neutral.

Table 1. An algorithm for finding a 2-neutral set

```
Find a pair of plaintexts that conforms to δ_r for some r
Find the set S of singles of neutral bits
Find simultaneous neutral pairs in S
while  do
   Count the number of edges for each element of S
   If the resulting set is a neutral set, break
   Remove from S one of the elements which has the least number of edges.
   Let the resulting set be S
end while
```

4 Application to the MD5

4.1 Description of the MD5

In this section, we give a brief description of the MD5 hash function and the block cipher based on the hash function, which is sufficient to understand the concepts introduced in this paper. For a full description of MD5 we refer to [15].

MD5 is a cryptographic hash function which was proposed in 1992 and has been one of the most well-known hash function. MD5 is constructed from MD(Merkle-Damgård) –construction and Davis-Meier mode. MD5 has 64 rounds, three kinds of non-linear functions, cyclic rotations, and round-dependent constants. The hash value calculated by MD5 is 128 bits long.

The function obtained from the compression function of MD5 by removing the feed-forward operation of the Davis-Meier mode is invertible. This function can be used as a block cipher which is called MD5 in encryption mode. We denote it by $E(K, S)$. The block cipher was analyzed in FSE 2003 [16].

The function $E(K, S)$ is an iterated design that only uses simple operations on 32-bit words. The 128-bit input V_j is loaded into 4 registers (A, B, C, D) and the 512-bit message block is divided into 16 words of 32 bits $(W_0 \dots W_{15})$ and these words are expanded to a sequence of 64 words through the message schedule. MD5 encrypts the initial value using this sequence as a key.

The 4 registers are updated through a number of rounds. The MD5 compression function consists of 64 rounds and have the following four non-linear functions f_1, f_2, f_3, f_4. Every round function has arithmetic addition, a round-dependent constant K_i.

$$
\begin{aligned}
f_1(X, Y, Z) &= (X \wedge Y) \vee (\overline{X} \wedge Z); \\
f_2(X, Y, Z) &= (X \wedge Z) \vee (Y \wedge \overline{Z}); \\
f_3(X, Y, Z) &= X \oplus Y \oplus Z; \\
f_4(X, Y, Z) &= (X \vee \overline{Z}) \oplus Y;
\end{aligned}
$$

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

The t-th round of the compression function updates the 4 registers using input word X_t and the constant K_i as input.

4.2 Pseudo-collision Attack on the MD5 hash function

By definition, to find a pseudo-collision, an attacker can inject differences both into the message schedule and registers. The attacker would require a complexity 2^{64} to find a pseudo-collision if MD5 is a ideal hash function.

In the ideal case, if both of an input difference and an output difference are fixed, then the probability that a plaintext pair with the input difference results in the output difference is 2^{-128}.

In cryptanalysis of MD5 in encryption mode, Saarinen found a iterative differential characteristics with a high probability 2^{-48} at FSE 2003 [16].

$$\delta = \texttt{80000000} \quad \texttt{80000000} \quad \texttt{80000000} \quad \texttt{80000000}$$
$$\downarrow$$
$$E(K, P) \oplus E(K, P \oplus \delta) = \delta$$

This means that this function $E(P, K)$ which is the core function of MD5 does not behave as a random function. This characteristic leads to an attack finding a pseudo-collision with a complexity 2^{48} due to the feed-forward operation of the Davis-Meyer mode.

In this section, we will see how much the method presented here improves the probability of this characteristic by using the particular set of plaintexts , rather than using a set of randomly chosen plaintexts. We set the key value K to be 0 so we study the resulting function $E(P, 0)$.

Since MD5 uses four different non-linear functions, it is interesting to see how much the probability is improved for each 16 rounds by finding neutral bits. Here is the result on this which is shown in the Table 2.

Table 2. The best probability for each 16 rounds

Rounds	This paper	The previous result [16]
0-15	$2^{-6.49}$	2^{-16}
16-31	$2^{-9.33}$	2^{-16}
32-47	1	1
48-63	$2^{-7.22}$	2^{-16}

Next we used the algorithm shown in the Table 1 to find some good set of inputs to $E(P, K)$. In order to attack many rounds, we have to create a large value for r. The problem is that if r is larger, then the number of neutral bits is smaller. It turned out that the optimal value for r is 6 in this respect.

In practice we found a pseudo-near-collision which differ only in 1 bit position with with complexity 2^{42}:

$$P = A|||B\|C\|D = \texttt{4315524f} \quad \texttt{79ba3feb} \quad \texttt{51453fe2} \quad \texttt{e3af887c}$$
$$P \oplus \delta = \texttt{c315524f} \quad \texttt{f9ba3feb} \quad \texttt{d1453fe2} \quad \texttt{63af887c}$$
$$MD5(0, P) \oplus MD5(0, P \oplus \delta) = \texttt{00000000} \quad \texttt{00000000} \quad \texttt{00100000} \quad \texttt{00000000}$$

The running time was approximately half a week with 32 CPU's in parallel.

Next in order to find a pseudo collision, we take another approach where we pay attention to the key input to the block cipher. For r, we use a large value as possible. The number of neutral bits could be too small to obtain enough inputs used in an attack. Instead, we use the key input. In the following experiment, we used 10 for r and chose random values for the key words from round 0 to round 9. Therefore we had 2^{42} inputs which consist of 2^6 plaintext inputs and 2^{36} key inputs where the values for the first 10 words are fixed to 0 and the other ones are chosen randomly.

Our experiment confirmed that the probability of this characteristic obtained from a 2-neutral set of size 6 is 2^{-39}, which is 2^9 higher than the probability of the same rounds of the original characteristic. This means that with a probability 2^{-39} the following equation holds:

$\mathrm{MD5}(K_0,\ P\ ) = \mathrm{MD5}(K_0,\ P \oplus \delta)$

What this means to the security of MD5 ($E(K,P)$ with the Davies-Meyer chaining) that for such a plaintext P, a pair of chaining variable $(P, P \oplus \delta)$ and a pair of message block ($M = M' = K_0$) produce a pseudo-collision for MD5:

$$
\begin{aligned}
P = A\|B\|C\|D &= \texttt{0xe1b1c8f8} \quad \texttt{0x55143ae6} \quad \texttt{0x75babfe9} \quad \texttt{0x001558a1} \\
P \oplus \delta &= \texttt{0x61b1c8f8} \quad \texttt{0xd5143ae6} \quad \texttt{0xf5babfe9} \quad \texttt{0x801558a1} \\
K_0 &= \texttt{0x00000000} \quad \texttt{0x00000000} \quad \texttt{0x00000000} \quad \texttt{0x00000000} \\
& \texttt{0x00000000} \quad \texttt{0x00000000} \quad \texttt{0x00000000} \quad \texttt{0x00000000} \\
& \texttt{0x00000000} \quad \texttt{0x00000000} \quad \texttt{0x6009f204} \quad \texttt{0xd2bf6eee} \\
& \texttt{0xb52517de} \quad \texttt{0x2f1889c8} \quad \texttt{0x72417083} \quad \texttt{0xa1cf21a1}
\end{aligned}
$$

This means that MD5 has weakness in randomness and pseudo-collision resistance.

Since several techniques to find full collisions as well as pseudo-collisions for MD5 quite efficiently have been already known [7] [10] [18], here we attempt to explain about two reasons why our result is of interest. The first reason is that the attacker could expect more freedom. In the pseudo-collision attacks [7] [10], the values for several message words are determined during the process of the attacks. In our attacks, we can choose any value for message(key) freely before starting the attack. The second reason is its simplicity. we use a simple but good characteristic. To improve its probability, we do not perform a detailed analysis of the boolean function which is done in [18]. Our algorithm automatically finds us a good set of inputs of MD5.

5 Application to the SHA-256

5.1 Description of the SHA-256

In this section, we give a brief description of the SHA-256 hash function, which is sufficient to understand the concepts introduced in this paper. For a full description of SHA-256 we refer to [14].

The 256-bit chaining variable V_j is loaded into 8 registers (A, B, C, D, E, F, G, H) and the 512-bit message block is divided into 16 words of 32 bits $(W_0 \ldots W_{15})$

and these words are expanded to a sequence of 64 words through the message schedule:

$$\begin{aligned}\sigma_0(X) &= ROTR^7(X) \oplus ROTR^{18}(X) \oplus SHR^3(X);\\ \sigma_1(X) &= ROTR^{17}(X) \oplus ROTR^{19}(X) \oplus SHR^{10}(X);\\ W_t &= \sigma_1(W_{t-2}) + W_{t-7} + \sigma_0(W_{t-15}) + W_{t-16}\end{aligned}$$

where $ROTR^n$ is right rotation by n bits.

The 8 registers are updated through a number of rounds. The SHA-256 compression function consists of 64 rounds. Every round function has arithmetic addition, a round-dependent constant K_i, two linear functions Σ_0, Σ_1, and two non-linear functions CH, MJ.

$$\begin{aligned}CH(X,Y,Z) &= (X \wedge Y) \oplus (\overline{X} \wedge Z);\\ MJ(X,Y,Z) &= (X \wedge Y) \oplus (Y \wedge Z) \oplus (Z \wedge X);\\ \Sigma_0(X) &= ROTR^2(X) \oplus ROTR^{13}(X) \oplus ROTR^{22}(X);\\ \Sigma_1(X) &= ROTR^6(X) \oplus ROTR^{11}(X) \oplus ROTR^{25}(X),\end{aligned}$$

where $\overline{X}$ is bitwise complement of X. The t-th round of the compression function updates the 8 registers using the word W_t and the constant K_i as input. The compression function updates the 8 registers according to the following algorithm:

$$\begin{aligned}T1_t(E_t, F_t, G_t, H_t, K_t, W_t) &= H_t + \Sigma_1(E_t) + CH(E_t, F_t, G_t) + K_t + W_t\,;\\ T2_t(A_t, B_t, C_t) &= \Sigma_0(A_t) + MJ(A_t, B_t, C_t)\,;\\ H_{t+1} &= G_t; G_{t+1} = F_t; F_{t+1} = E_t; E_{t+1} = D_t + T1_t\,;\\ D_{t+1} &= C_t; C_{t+1} = B_t; B_{t+1} = A_t; A_{t+1} = T1_t + T2_t.\end{aligned}$$

5.2 Application to the SHA-2-XOR

We consider a SHA-256 variant in which every arithmetic addition is replaced by XOR operation. We call this variant SHA-2-XOR.

We discuss pseudo-collision resistance and pseudo-randomness of this function. In the ideal case, the attacker would require a complexity 2^{128} to find a pseudo-collision and if both of an input difference and an output difference are fixed, then the probability that a plaintext pair with the input difference results in the output difference is 2^{-256}.

At SAC 2005 [20], Yoshida and Biryukov presented a pseudo-collision attack on the reduced SHA-2-XOR with 34-rounds using the best one-round iterative characteristic with a high probability 2^{-8}. The attack is of complexity 2^{120} and uses the input modification technique which ensures some conditions to hold so that for the first 19-rounds no probability is paid.

Here we consider an attack which also uses this iterative characteristic but in order to improve the probability for the first several rounds, we apply the technique of neutral bits, instead of the input modification technique.

Here $P = A\|\|B\|C\|D\|E\|F\|G\|H$

$$\delta = \texttt{0xb3b3b3b3 \quad 0xb3b3b3b3 \quad 0xb3b3b3b3 \quad 0xb3b3b3b3}$$
$$\texttt{0x0c0c0c0c \quad 0x0c0c0c0c \quad 0x0c0c0c0c \quad 0x0c0c0c0c}$$
$$\downarrow$$
$$E(K, P) \oplus E(K, P \oplus \delta) = \delta$$

In this section, we will see how much the method presented here improves the probability of this characteristic by using the particular set of plaintexts , rather than using a set of randomly chosen plaintexts. We set the key value K to be 0 so we study the resulting function $E(0, P)$.

Table 3. The set of neutral bits of size 27 for r = 7, (the bits are numbered in the range 0, ..., 255)

.

P =	0x4939a45a 0x79ec4172 0xf0ef5249 0x29b5bb6f 0xd92f76e4 0x21962dfe 0xd88e64f6 0x7b624d63
P⊕δ =	0xfa8a17e9 0xca5ff2c1 0x435ce1fa 0x9a0608dc 0xd5237ae8 0x2d9a21f2 0xd48268fa 0x776e416f
Pairs:	(128 0), (129 1),(132 4),(133 5),(134 6),(135 7),(136 8), (137 9),(140 12),(141 13),(142 14),(143 15),(144 16),(145 17), (148 20),(149 21),(150 22),(151 23),(152 24),(153 25),(156 28), (157 29),(158 30),(159 31),(165 37),(166 38),(167 39)

For $r = 7$, an experiment using the algorithm1 gave us a 2-neutral set of size 53. In order to estimate probability with a practical complexity, we took a sub-set of size of 27 shown in the Table 3.

Our experiment confirmed that the probability of 10-round of this characteristic obtained from this 2-neutral set is $2^{-23.678072}$ which is slightly more than 2^{-24}. This is 2^{56} higher than the probability in the original characteristic.

This means that with a probability $2^{-23.678072}$ the following equation holds:

$$E_{10}(0, P) \oplus E_{10}(0, P \oplus \delta) = \delta$$

In practice we found 10 right pairs of plaintexts $(P, P \oplus \delta)$ with complexity 2^{27}. What this means to the security of SHA-2-XOR ($E(K, P)$ with the Davies-Meyer chaining) is that for such a plaintext P, a pair of chaining variable $(P, P \oplus \delta)$ and a pair of message $(M = 0, M' = M)$ produces a pseudo-collision for 10 rounds of SHA-2-XOR hash function. A pair of plaintexts which produce such a pseudo collision is as follows:

$$P = \texttt{0x4939a45a \quad 0x79ec4172 \quad 0xf0ef52a9 \quad 0xa8161bbe}$$

$$P \oplus \delta = \begin{matrix} \texttt{0xd92f76e4} & \texttt{0x21962dfe} & \texttt{0xd88e6416} & \texttt{0xfac1edb2} \\ \texttt{0xfa8a17e9} & \texttt{0xca5ff2c1} & \texttt{0x435ce11a} & \texttt{0x1ba5a80d} \\ \texttt{0xd5237ae8} & \texttt{0x2d9a21f2} & \texttt{0xd482681a} & \texttt{0xf6cde1be} \end{matrix}$$

Now the interesting question is how many rounds we could add to this 10-rounds from theoretical point of view. In principle, we can use 768 bits of input in the case of SHA-2-XOR. What we need to consider is that how many input bits we have used so far and will be able to control to add rounds. In order to obtain 10-round pseudo-collisions, we had to do two things:

1) Fix each of the words $W_0, W_1, \dots, W_6$ to 0

2) Use the 2-neutral set of size 27

This means that we use $7 \cdot$ 32bits to construct 10-round pseudo-collisions, therefore we can control the message words, $W_7, W_8, \dots, W_{15}$ (=freedom of $9 \cdot$ 32 bits.) to add 12 rounds.

This discussion above means that 38-rounds of SHA-2-XOR has weakness in randomness and 22-rounds of SHA-2-XOR has weakness in pseudo-collision resistance.

5.3 Application to the SHA-256

SHA-256 in encryption mode was proposed for use as a block cipher by Handschuh and Naccache and named SHACAL-2 [11]. The block cipher was selected as one of the NESSIE finalists.

In [17], an attack on the reduced 32-round SHACAL-2 using a 14-round truncated differential characteristic is presented. Since the round function of SHACAL-2 is exactly same as the round function of SHA-256, this 14-round characteristic shown in the Table 4 can be considered as some interesting property of SHA-256. In the Table 4 we denote by $e_{i_1,\dots,i_k,\sim}$ a 32-bit word that has 1's in the positions $i_1, \dots, i_k$, and unconcerned values in the positions of the bits $(i_k + 1) \sim 31$, and 0's in the rest of bit positions and we also denote by z_0 a 32-bit word that has 0 in the positions 0 and unconcerned values in the other bit positions.

This characteristic has a probability 2^{-32} which has been improved to approximately $2^{-18.7}$ in [17] using two kinds of technique:

1) Fixing some bits of plaintext pairs

2) Computing possible dE_{10} values and construct multiple differential characteristics.

Here we use the technique described before to improve the probability and compare the results with the ones in Table[17].

We found a pair of plaintexts with the set of 20 Neutral Bits for $r = 5$, which is shown in the Table 5.

Our experiment confirmed that using this set gave us a probability $2^{-8.01}$ for the 14-round truncated characteristic. This is about 2^{10} higher than the improved probability in [17].

Table 4. A 14-round truncated differential characteristic ($M_1 = \{9, 18, 29\}, M_2 = \{6, 9, 18, 25, 29\}, M_3 = \{6, 9, 18, 20, 25\}$)

Round	dA_t	dB_t	dC_t	dD_t	dE_t	dF_t	dG_t	dH_t	Prob.
Input($t = 0$)	0	0	e_{M_1}	0	0	e_{31}	e_{M_2}	0	2^{-10}
1	e_{31}	0	0	e_{M_1}	e_{31}	0	e_{31}	e_{M_2}	2^{-10}
2	0	e_{31}	0	0	0	e_{31}	0	e_{31}	2^{-2}
3	0	0	e_{31}	0	0	0	e_{31}	0	2^{-2}
4	0	0	0	e_{31}	0	0	0	e_{31}	1
5	e_{31}	0	0	0	0	0	0	0	2^{-4}
6	e_{M_1}	e_{31}	0	0	0	0	0	0	1
7	z_0	e_{M_1}	e_{31}	0	0	0	0	0	1
8	?	z_0	e_{M_1}	e_{31}	0	0	0	0	1
9	?	?	z_0	$e_{M_3,\sim}$	e_{31}	0	0	0	2^{-4}
10	?	?	?	z_0	$e_{M_3,\sim}$	e_{31}	0	0	1
11	?	?	?	?	z_0	$e_{M_3,\sim}$	e_{31}	0	1
12	?	?	?	?	?	z_0	$e_{M_3,\sim}$	e_{31}	1
13	?	?	?	?	?	?	z_0	e_{31}	1
14	?	?	?	?	?	?	?	z_0	

Table 5. The set of neutral bits of size 28 for r = 5, (the bits are numbered in the range 0, ..., 255)

.

P =	`0x2e76ad25`	`0x3c0d407b`	`0xd54f19d7`	`0xe8c7e1e3`
	`0xb25f725c`	`0x618fad55`	`0xb63b2fe8`	`0x9326a499`
$P \oplus \delta_0 =$	`0x2e76ad25`	`0x3c0d407b`	`0xf54b1bd7`	`0xe8c7e1e3`
	`0xb25f725c`	`0xe18fad55`	`0x942f2da8`	`0x9326a499`
Singles:	45,46,49,51,71,74,75,87,88,153,172, 176,186,192,198,199,200,209,214,220			

6 Conclusions

We applied the Biham-Chen attack to analyze SHA-like hash functions regarding pseudo-collision resistance and pseudo-randomness. Using our scenario, we presented a differential attack on the MD5 hash function and a differential attack on a variant of SHA-2-XOR hash function. We also studied the several rounds' property of the real SHA-256 function. We observed that in all the case, some previous results on the differential probability were improved. For the future work, we will use other kinds of neutral bits (triplets) to attack more rounds. We think that even better probabilities may be obtained with the resulting set.

Acknowledgements

The authors would wish to thank Dai Watanabe for helpful comments and useful discussions.

References

1. E. Biham, R. Chen "Near-Collision of SHA-0," in *Proceedings of CRYPT 2004*, LNCS 3152, M. Franklin, Ed., pp.290–305, 2004.
2. E. Biham, R. Chen, A. Joux, P. Carribault, C. Lemuet, and W. Jalby, "Collisions of SHA-0 and Reduced SHA-1," in *Proceedings of Eurocrypt 2005*, LNCS 3494, R. Cramer, Ed., Springer-Verlag, pp. 36–57, 2005.
3. E. Biham, A. Biryukov, A. Shamir, "Cryptanalysis of SkipJack Reduced to 31 Rounds Using Impossible Differentials," in *Proceedings of Eurocrypt'99*, LNCS 1592, Springer-Verlag, pp.12–23, 1999.
4. E. Biham, A. Shamir, *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
5. A. Biryukov, D. Wagner, "Advanced slide attacks," in *Proceedings of Eurocrypt 2000*, LNCS 1807, B. Preneel, Ed., Springer-Verlag, pp. 589–606, 2000.
6. A. Biryukov and E. Kushilevitz, "Improved Cryptanalysis of RC5", in *Proceedings of EUROCRYPT'98*, LNCS 1403, K. Nyberg, Ed., Springer-Verlag, pp. 85–99, 1998.
7. B. D. Boer, A. Bosselaers, "Collisions for the compression function of MD5," in *Proceedings of Eurocrypt 1993*, LNCS 765, T. Helleseth, Ed., Springer-Verlag, pp. 293–304, 1993.
8. F. Chabaud and A. Joux, "Differential Collisions in SHA-0," in *Proceedings of CRYPTO'98*, LNCS 1462, H. Krawczyk, Ed., pp.56-71, Springer-Verlag, 1998.
9. I. Damgård, "A design principle for hash functions," in *Proceedings of Crypto'89*, LNCS 435, G. Brassard, Ed., Springer-Verlag, pp. 416–427, 1990.
10. H. Dobbertin, "The status of MD5 after a recent attack,", Cryptobytes, Vol. 2, No. 2, pp. 1–6, Summer 1996.
11. H. Handschuh, D. Naccache, "SHACAL," Submission to the NESSIE project, 2000. Available from `http://www.gemplus.com/smart/r_d/publications/pdf/HN00shac.pdf`.
12. L. R. Knudsen and J. E. Mathiassen, "Preimage and collision attacks on MD2," in *Proceedings of FSE 2005*, LNCS 3557, H. Gilbert and H. Handschuh Ed., Springer-Verlag, pp. 255–267, 2005.

13. A. Menezes, P. van Oorschot and S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.
14. National Institute of Standards and Technology, FIPS-180-2: "Secure Hash Standard (SHS)," August 2002.
15. R. Rivest, "The MD5 message-digest algorithm," Request for Comments (RFC) 1321, Internet Activities Board, Internet Privacy Task Force, April 1992.
16. M. Saarinen, "Cryptanalysis of Block Ciphers Based on SHA-1 and MD5," in *Proceedings of FSE 2003*, LNCS 2887, T. Johansson, Ed., Springer-Verlag, pp. 36–44, 2003.
17. Y. Shin, J. Kim, G. Kim, S. Hong, and S. Lee, "Differential-Linear Type Attacks on Reduced Rounds of SHACAL-2," in *Proceedings of ACISP 2004*, LNCS 3108, H. Wang, J. Pieprzyk, and V. Varadharajan, Ed., Springer-Verlag, pp. 110–122, 2004.
18. X. Wang, X. Lai, D. Feng, H. Chen, and X. Yu, "Cryptanalysis of the Hash Functions MD4 and RIPEMD," in *Proceedings of Eurocrypt 2005*, LNCS 3494, R. Cramer, Ed., Springer-Verlag, pp. 1–18, 2005.
19. H. Yoshida, A. Biryukov, C. D. Cannière, J. Lano, and B. Preneel, "Non-randomness of the Full 4 and 5-pass HAVAL," in *Proceedings of SCN 2004*, LNCS 3352, C. Blundo and S. Climato, Ed., Springer-Verlag, pp. 324–336, 2005.
20. H. Yoshida and A. Biryukov, "Analysis of a SHA-256 Variant," in *Proceedings of SAC 2005.*