IEICE TRANS. FUNDAMENTALS, VOL.E95–A, NO.1 JANUARY 2012

LETTER *Special Section on Cryptography and Information Security*

Collision Resistance of Hash Functions in a Weak Ideal Cipher Model

Shoichi HIROSE[†a)] ***and*** **Hidenori KUWAKADO**[††], ***Members***

SUMMARY This article discusses the provable security of block-cipher-based hash functions. It introduces a new model called a weak ideal cipher model. In this model, an adversary is allowed to make key-disclosure queries to the oracle as well as encryption and decryption queries. A key-disclosure query is a pair of a plaintext and a ciphertext, and the reply is a corresponding key. Thus, in this model, a block cipher is random but completely insecure as a block cipher. It is shown that collision resistant hash functions can be constructed even in this weak model.

key words: *hash function, provable security, collision resistance*

1. Introduction

Background. Hash functions are widely used in cryptography. The most classical and well-known method to construct a hash function is based on a block cipher. A block-cipher-based hash function usually has iteration structure and is called an iterated hash function: it is composed of a compression function, which is successively applied to a given input of arbitrary length. The way of application is called a domain extension. Preneel, Govaerts and Vandewalle presented a model for block-cipher-based compression functions [7]. They also showed that 12 compression functions in the model are collision resistant against some generic attacks. Black, Rogaway and Shrimpton analyzed the compression functions in the Preneel-Govaerts-Vandewalle (PGV) model in terms of provable security [1]. They showed that the same 12 compression functions are collision resistant in the ideal cipher model up to the birthday bound. The 12 compression functions include the Davies-Meyer (DM) compression function, the Matyas-Meyer-Oseas (MMO), and the Miyaguchi-Preneel (MP) [5].

In response to the recent advances in the cryptanalysis of hash functions, NIST has opened a public competition, known as the SHA-3 competition, to select a new hash function standard [6]. Many SHA-3 candidates as well as MD4, MD5 and SHA-1/2 are based on the DM scheme. Whirlpool [8] uses the MP scheme. One of the five SHA-3 finalists, Skein, uses the MMO scheme. In contrast, other PGV schemes do not seem to be used in practice.

One of the fourteen second-round candidates, Blue Midnight Wish (BMW) [3], is one of the most efficient hash functions among them. It is also composed of a compression function using a block cipher. The designers claim that the structure of the BMW compression function is based on those of the 12 collision-resistant PGV compression functions. It is also remarkable that the underlying block cipher of the BMW compression function is weak: it is easy to compute a corresponding key for a given pair of a plaintext and a ciphertext. It implies that the block cipher of BMW is completely insecure as a block cipher.

Our Contribution. This article is inspired by the BMW compression function. It discusses the construction of compression functions using a weak block cipher, and the construction of iterated hash functions using such compression functions. A new model of an ideal primitive is introduced for the analyses, which we call a weak ideal cipher model. It is similar to the weak ideal compression function model introduced by Liskov [4].

In the weak ideal cipher model, a block cipher $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ is assumed to be a random element such that both $E(K,\cdot)$ and $E(\cdot,X)$ are permutations for every key K and every plaintext X, where $X, K \in \{0,1\}^n$. An adversary is allowed to make three kinds of queries to the oracle: encryption queries, decryption queries, and key-disclosure queries. Encryption and decryption queries are also allowed in the ideal cipher model. A key-disclosure query is a pair of a plaintext and a ciphertext, and the reply is a corresponding key.

We discuss the collision resistance (CR) in the weak ideal cipher model. We do not directly analyze the PGV compression functions. Instead of that, we analyze the compression functions in the framework by Stam [10], and then apply the results to the PGV compression functions, because it gives us a clearer view. We give a sufficient condition for CR compression functions. We also give a sufficient condition for compression functions with which a CR hash function is obtained using the strengthened Merkle-Damgård (SMD) domain extension.

For the PGV compression functions, 8 among the 12 CR compression functions in the ideal cipher model remain CR even in the weak ideal cipher model. The MP compression function is one of the 8 compression functions. On the other hand, neither the DM nor the MMO compression functions are found to be CR in the weak ideal cipher model.

Related Work. Liskov [4] introduced the weak ideal compression function model. He also proposed the zipper construction, and showed that it satisfies indifferentiability from

Manuscript received March 29, 2011.
Manuscript revised July 19, 2011.

[†]The author is with Graduate School of Engineering, University of Fukui, Fukui-shi, 910-8507 Japan.
[††]The author is with Faculty of Engineering, Kobe University, Kobe-shi, 657-8501 Japan.
a) E-mail: hrs_shch@u-fukui.ac.jp

DOI: 10.1587/transfun.E95.A.252

a random oracle (IRO) in this model, which implies CR. It is a non-streaming construction, and requires two independent weak compression functions. He also presented a double-pipe compression function based on the zipper construction.

The ideal cipher model dates back to Shannon, and it has been widely used in the analysis of the security of block-cipher-based hash functions [1], [2].

Stam [10] presented a general framework for block-cipher-based compression functions, which covers the PGV compression functions [7]. He studied the collision resistance and preimage resistance of block-cipher-based hash functions under this framework in the ideal cipher model.

Organization. Section 2 gives the definition of collision resistance and introduces the weak ideal cipher model. Section 3 discusses the collision resistance of compression functions and hash functions. Section 4 gives a concluding remark.

2. Definitions

2.1 Collision Resistance

Collision resistance of a block-cipher-based hash function is often discussed on the assumption that the underlying block cipher is an ideal primitive. It is mainly because the collision resistance of a hash function cannot be implied by the pseudorandomness of the underlying block cipher as a black-box [9].

Let C be a scheme to construct a hash function or a compression function using an ideal primitive $\mathcal{F}$. Let A be an adversary with access to the oracle $\mathcal{F}$. The col-advantage of A against $C^{\mathcal{F}}$, $\mathrm{Adv}^{\mathrm{col}}_{C^{\mathcal{F}}}(A)$, is given by

$$\Pr[(M, M') \leftarrow A^{\mathcal{F}} \wedge M \neq M' \wedge C^{\mathcal{F}}(M) = C^{\mathcal{F}}(M')],$$

where the probabilities are taken over the coin tosses by A and the distribution of $\mathcal{F}$. It can be assumed that A makes no repeated queries without loss of generality. $C^{\mathcal{F}}$ is said to be collision-resistant (CR) if $\mathrm{Adv}^{\mathrm{col}}_{C^{\mathcal{F}}}(A)$ is negligible for any efficient A.

2.2 Weak Ideal Cipher Model

A block cipher with block length n and key length κ is called an (n, κ) block cipher. Let $E : \{0,1\}^{\kappa} \times \{0,1\}^n \to \{0,1\}^n$ be an (n, κ) block cipher. We will assume that $\kappa = n$ and that E has the following properties:

1. $E(K, \cdot)$ is a permutation for every $K \in \{0,1\}^n$.
2. $E(\cdot, X)$ is a permutation for every $X \in \{0,1\}^n$.
3. It is easy to compute $E(K, X)$ for every $K, X \in \{0,1\}^n$.
4. For every $K, Y \in \{0,1\}^n$, it is easy to compute X such that $Y = E(K, X)$.
5. For every $X, Y \in \{0,1\}^n$, it is easy to compute K such that $Y = E(K, X)$.

E is obviously insecure as a block cipher from the last property.

Actually, the second property is not essential. We will assume it, however, to make the analyses in the next section easier. Notice that the weak block cipher used in Blue Midnight Wish has this property.

In the weak ideal cipher model, the oracle first selects an element from

$$\{E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n \mid \text{Both } E(K, \cdot) \text{ and } E(\cdot, X) \text{ are permutations}\}$$

under the uniform distribution. An adversary makes three kinds of queries to the oracle:

- Encryption query: it is of the form (e, K, X), and the oracle returns Y.
- Decryption query: it is of the form (d, K, Y), and the oracle returns X.
- Key-disclosure query: it is of the form (k, X, Y), and the oracle returns K.

3. Results

3.1 Collision Resistance

This section explores the collision resistance of compression functions represented by Stam's model [10] in the weak ideal cipher model. It also discusses the collision resistance of hash functions using the compression functions and the strengthened Merkle-Damgård (SMD) domain extension.

Stam's model is illustrated in Fig. 1. C^{PRE} is a preprocessing function: $C^{\mathrm{PRE}}(M, V) = (K, X)$, where M is a message block and V is a chaining variable. C^{POST} is a postprocessing function: $C^{\mathrm{POST}}(M, V, Y) = W$. An auxiliary function C^{AUX} is also defined by $C^{\mathrm{AUX}}(K, X, Y) \stackrel{\mathrm{def}}{=} C^{\mathrm{POST}}(C^{-\mathrm{PRE}}(K, X), Y)$, where $C^{-\mathrm{PRE}}$ is the inverse of C^{PRE}.

Theorem 1 gives a set of sufficient conditions for a block-cipher-based compression function to be CR in the weak ideal cipher model. It is well-known that a CR hash function can be constructed with a CR compression function and the SMD domain extension.

Theorem 1 Suppose that a block-cipher-based compression function $F : \{0,1\}^{2n} \to \{0,1\}^n$ in Stam's model satisfies the following conditions:

1. C^{PRE} is bijective.
2. For every M, V, $C^{\mathrm{POST}}(M, V, \cdot)$ is bijective.
3. For every K, Y, $C^{\mathrm{AUX}}(K, \cdot, Y)$ is bijective.
4. For every X, Y, $C^{\mathrm{AUX}}(\cdot, X, Y)$ is bijective.

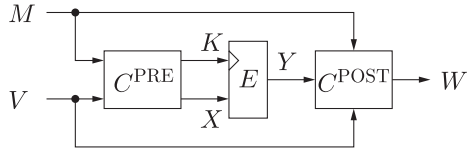

Fig. 1 Stam's model of compression functions [10].

Then, in the weak ideal cipher model, for any collision-finding adversary A asking at most $q \geq 1$ queries to the oracle (the underlying block cipher E),

$$\mathrm{Adv}_F^{\mathrm{col}}(A) \leq q^2/(2\,(2^n - q)).$$

Proof Let $\mathbb{N}_i$ be the set of positive integers smaller than i. Let (K_i, X_i, Y_i) be a pair of input and output of E obtained by the adversary A with the i-th query and the corresponding response from the oracle. Since C^{PRE} is bijective, a unique input to F corresponds to (K_i, X_i). Let (M_i, V_i) be the input. Since A makes no repeated queries, $(K_i, X_i) \neq (K_{i'}, X_{i'})$ if $i \neq i'$. So, $(M_i, V_i) \neq (M_{i'}, V_{i'})$.

Suppose that the i-th query is an encryption query (e, K_i, X_i). Then, the corresponding (M_i, V_i) is fixed by (K_i, X_i), and the oracle returns a random reply Y_i. Thus, since $C^{\mathrm{POST}}(M_i, V_i, \cdot)$ is bijective, for $W_i = C^{\mathrm{POST}}(M_i, V_i, Y_i)$,

$$\Pr\left[W_i = W_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq (i-1)/(2^n - (i-1)).$$

Suppose that the i-th query is a decryption query (d, K_i, Y_i). Then, the oracle returns a random reply X_i. Thus, since $C^{\mathrm{AUX}}(K_i, \cdot, Y_i)$ is bijective, for $W_i = C^{\mathrm{AUX}}(K_i, X_i, Y_i)$,

$$\Pr\left[W_i = W_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq (i-1)/(2^n - (i-1)).$$

Suppose that the i-th query is a key-disclosure query (k, X_i, Y_i). Then, the oracle returns a random reply K_i. Thus, since $C^{\mathrm{AUX}}(\cdot, X_i, Y_i)$ is bijective, for $W_i = C^{\mathrm{AUX}}(K_i, X_i, Y_i)$,

$$\Pr\left[W_i = W_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq (i-1)/(2^n - (i-1)).$$

Since A makes at most q queries,

$$\mathrm{Adv}_F^{\mathrm{col}}(A) \leq \sum_{i=1}^{q} (i-1)/(2^n - (i-1)) \leq q^2/(2\,(2^n - q)).$$

□

Theorem 2 gives a set of sufficient conditions for a block-cipher-based compression function with which a CR hash function is constructed using the SMD domain extension in the weak ideal cipher model.

Theorem 2 Let H be a hash function constructed with the SMD domain extension and a block-cipher-based compression function $F : \{0,1\}^{2n} \to \{0,1\}^n$ in Stam's model. Suppose that F satisfies the following conditions:

1. C^{PRE} is bijective.
2. For every M, V, $C^{\mathrm{POST}}(M, V, \cdot)$ is bijective.
3. For every K,
   a. for every Y, $C^{\mathrm{AUX}}(K, \cdot, Y)$ is bijective, or
   b. the mapping from X to V defined by $(M, V) \leftarrow C^{-\mathrm{PRE}}(K, X)$ is bijective.
4. For every X,
   a. for every Y, $C^{\mathrm{AUX}}(\cdot, X, Y)$ is bijective, or
   b. the mapping from K to V defined by $(M, V) \leftarrow C^{-\mathrm{PRE}}(K, X)$ is bijective.

Then, in the weak ideal cipher model, for any collision-finding adversary A asking at most $q \geq 1$ queries to the oracle (the underlying block cipher E),

$$\mathrm{Adv}_H^{\mathrm{col}}(A) \leq q^2/(2^n - q).$$

Proof The notations in the proof of Theorem 1 are also used here.

Suppose that A eventually finds a collision for the hash function with the i-th query. Let $W_i = F(V_i, M_i)$ and IV be the initial value of the hash function. Then, it is necessary for A to encounter one of the following events:

- W_i depends on the reply from the oracle, and it collides with W_j or V_j for some $j \in \mathbb{N}_i$.
- V_i depends on the reply from the oracle, and it collides with IV or W_j for some $j \in \mathbb{N}_i$.

Suppose that the i-th query is (e, K_i, X_i). Then, the corresponding (M_i, V_i) is fixed. Since $C^{\mathrm{POST}}(M_i, V_i, \cdot)$ is bijective, for $W_i = C^{\mathrm{POST}}(M_i, V_i, Y_i)$,

$$\Pr\left[W_i = W_j \vee W_i = V_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq \frac{2(i-1)}{2^n - (i-1)}.$$

Suppose that the i-th query is (d, K_i, Y_i). If $C^{\mathrm{AUX}}(K, \cdot, Y)$ is bijective for every K and Y, then for $W_i = C^{\mathrm{AUX}}(K_i, X_i, Y_i)$,

$$\Pr\left[W_i = W_j \vee W_i = V_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq \frac{2(i-1)}{2^n - (i-1)}.$$

On the other hand, if the mapping from X to V defined by $(M, V) \leftarrow C^{-\mathrm{PRE}}(K, X)$ is bijective for every K, then

$$\Pr\left[V_i = IV \vee V_i = W_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq \frac{i}{2^n - (i-1)}.$$

Suppose that the i-th query is (k, X_i, Y_i). If $C^{\mathrm{AUX}}(\cdot, X, Y)$ is bijective for every X and Y, then for $W_i = C^{\mathrm{AUX}}(K_i, X_i, Y_i)$,

$$\Pr\left[W_i = W_j \vee W_i = V_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq \frac{2(i-1)}{2^n - (i-1)}.$$

If the mapping from K to V defined by $(M, V) \leftarrow C^{-\mathrm{PRE}}(K, X)$ is bijective for every K, then

$$\Pr\left[V_i = IV \vee V_i = W_j \text{ for } \exists j \in \mathbb{N}_i\right] \leq \frac{i}{2^n - (i-1)}.$$

Since A makes at most $q \geq 1$ queries,

$$\mathrm{Adv}_H^{\mathrm{col}}(A) \leq \sum_{i=1}^{q} \frac{\max\{i, 2(i-1)\}}{2^n - (i-1)} \leq \frac{q^2}{2^n - q}.$$

□

Notice that the conditions 1, 2, 3(a) and 4(a) in Theorem 2 are identical with the conditions 1, 2, 3 and 4 in

Table 1 20 PGV compression functions for CR hash functions in the ideal cipher model [1]. M is a message block, and V is a chaining variable. $Z = M \oplus V$. c is a constant. "1" indicates that the corresponding compression function satisfies the conditions in Theorem 1. "2" indicates that the corresponding compression function satisfies the conditions in Theorem 2, but does not satisfy the conditions in Theorem 1. "×" indicates that the corresponding compression function does not satisfy the conditions in Theorem 2 (nor in Theorem 1).

h_1	$E_V(M) \oplus M$	2	h_{11}	$E_Z(M) \oplus V$	1
h_2	$E_V(Z) \oplus Z$	2	h_{12}	$E_Z(V) \oplus M$	1
h_3	$E_V(M) \oplus Z$	1	h_{13}	$E_Z(M) \oplus c$	2
h_4	$E_V(Z) \oplus M$	1	h_{14}	$E_Z(M) \oplus Z$	2
h_5	$E_M(V) \oplus V$	×	h_{15}	$E_M(V) \oplus c$	×
h_6	$E_M(Z) \oplus Z$	2	h_{16}	$E_Z(V) \oplus c$	×
h_7	$E_M(V) \oplus Z$	1	h_{17}	$E_M(V) \oplus M$	2
h_8	$E_M(Z) \oplus V$	1	h_{18}	$E_Z(V) \oplus Z$	2
h_9	$E_Z(M) \oplus M$	1	h_{19}	$E_M(Z) \oplus c$	2
h_{10}	$E_Z(V) \oplus V$	1	h_{20}	$E_M(Z) \oplus M$	2

Theorem 1, respectively.

3.2 PGV Compression Functions

We will see the implications of Theorems 1 and 2 for the PGV compression functions. They are summarized in Table 1. Among the twelve CR compression functions in the ideal cipher model (h_1 through h_{12}), eight functions labeled with "1" remain CR even in the weak ideal cipher model. The MP compression function (h_3) belongs to this class. Nine compression functions labeled with "2" are not CR by themselves. However, hash functions using them and the SMD domain extension are CR in the weak ideal cipher model. The MMO compression function (h_1) belongs to this class. In contrast, CR hash functions cannot be constructed with h_5, h_{15} nor h_{16}, where h_5 is the DM compression function.

4. Concluding Remark

This work is inspired by a SHA-3 candidate BMW. However, the results do not seem to have direct implications in its security. For example, the security of BMW also seems to depend on the preprocessing and postprocessing functions. They are as complex as the underlying weak block cipher. The results in this article still suggest that some PGV compression functions may compensate it if any vulnerability is found in the underlying block cipher as far as CR is concerned.

Acknowledgements

We would like to thank an anonymous reviewer for valuable comments. This work was supported in part by KAKENHI 20300003.

References

[1] J. Black, P. Rogaway, and T. Shrimpton, "Black-box analysis of the block-cipher-based hash-function constructions from PGV," CRYPTO, ed. M. Yung, Lect. Notes Comput. Sci., vol.2442, pp.320–335, Springer, 2002.

[2] J.-S. Coron, Y. Dodis, C. Malinaud, and P. Puniya, "Merkle-damgård revisited: How to construct a hash function," CRYPTO, ed. V. Shoup, Lect. Notes Comput. Sci., vol.3621, pp.430–448, Springer, 2005.

[3] D. Gligoroski, V. Klima, S.J. Knapskog, M. El-Hadedy, J. Amundsen, and S.F. Mjølsnes, "Cryptographic hash function BLUE MIDNIGHT WISH," Submission to NIST, 2008. http://people.item.ntnu.no/~danilog/Hash/BMW/Supporting_Documentation/BlueMidnightWishDocumentation.pdf

[4] M. Liskov, "Constructing an ideal hash function from weak ideal compression functions," in Selected Areas in Cryptography, ed. E. Biham and A.M. Youssef, Lect. Notes Comput. Sci., vol.4356, pp.358–375, Springer, 2006.

[5] A.J. Menezes, P.C. van Oorschot, and S.A. Vanstone, Handbook of Applied Cryptography, CRC Press, 1996.

[6] National Institute of Standards and Technology (NIST), Cryptographic hash algorithm competition. http://csrc.nist.gov/groups/ST/hash/sha-3/index.html

[7] B. Preneel, R. Govaerts, and J. Vandewalle, "Hash functions based on block ciphers: A synthetic approach," CRYPTO, ed. D.R. Stinson, Lect. Notes Comput. Sci., vol.773, pp.368–378, Springer, 1993.

[8] V. Rijmen and P.S.L.M. Barreto, "The Whirlpool hash function," http://www.larc.usp.br/~pbarreto/WhirlpoolPage.html, 2000.

[9] D.R. Simon, "Finding collisions on a one-way street: Can secure hash functions be based on general assumptions?," EUROCRYPT, ed. K. Nyberg, Lect. Notes Comput. Sci., vol.1403, pp.334–345, Springer, 1998.

[10] M. Stam, "Blockcipher-based hashing revisited," in FSE, O. Dunkelman, ed., Lect. Notes Comput. Sci., vol.5665, pp.67–83, Springer, 2009.