

Secure Block Ciphers Are Not Sufficient for One-Way Hash Functions in the Preneel-Govaerts-Vandewalle Model

Shoichi Hirose

Graduate School of Informatics, Kyoto University, Kyoto, 606–8501 JAPAN
hirose@i.kyoto-u.ac.jp

Abstract. There are many proposals of unkeyed hash functions based on block ciphers. Preneel, Govaerts and Vandewalle, in their CRYPTO'93 paper, presented the general model of unkeyed hash functions based on block ciphers such that the size of the hashcode is equal to the block size and is almost equal to the key size. In this article, it is shown that, for every unkeyed hash function in their model, there exist block ciphers secure against the adaptive chosen plaintext attack such that the unkeyed hash function based on them is not one-way. The proof is constructive: the secure block ciphers are explicitly defined based on which one-way unkeyed hash functions cannot be constructed. Some of the block ciphers presented are secure even against the adaptive chosen plaintext/ciphertext attack.

1 Introduction

Hash functions are very important primitives in cryptography. Hash functions in cryptography are classified in two types: unkeyed hash functions and keyed hash functions. The former ones are also called manipulation detection codes (MDCs). They are used for message digest in signature schemes. The latter ones are also called message authentication codes (MACs). Excellent surveys are presented in [4,8]. Unkeyed hash functions are discussed in this article.

There are many proposals of unkeyed hash functions. One of the approaches is to construct them based on block ciphers. Some of the proposals following this approach are found in [2,3,6,7]. The main motivation of this approach is the minimization of design and implementation effort, which is supported by the expectation that secure unkeyed hash functions can be constructed from secure block ciphers.

Secure unkeyed hash functions are classified in two types: one-way hash functions and collision resistant hash functions. One-way hash functions are further classified in preimage resistant hash functions and second-preimage resistant hash functions. Informally, preimage resistance means that, given an output, it is infeasible to obtain an input which produces the output. Second-preimage resistance means that, given an input, it is infeasible to obtain another input which produces the same output as the given input. Collision resistance means that it is infeasible to obtain two different inputs which produce the same output.

Preneel, Govaerts and Vandewalle studied the unkeyed hash functions based on block ciphers such that the size of the hashcode is equal to the block size and is almost equal to the key size [9]. They presented the general model of such hash functions. There are 64 schemes in their model. Let us call them PGVHFs. They considered the security of all PGVHFs against the existing five important attacks and concluded that 12 schemes are secure assuming that the underlying block cipher is ideal. However, “ideal” means that the block cipher is a keyed random permutation, which is quite impractical in a strict sense.

On the other hand, for collision resistance, Simon’s result [10] implies that no provable construction of a collision resistant hash function exists based on a “black box” one-way permutation, which means that the one-way permutation is used as a subroutine, that is, the internal structure is not used. PGVHFs are constructed based on a “black box” block cipher, and a block cipher can be constructed based on a “black box” one-way permutation.

It is still open if there exist any provable one-way PGVHFs based on secure block ciphers. In this article, a negative result is given to this problem. It is shown that, for every PGVHF, there exist block ciphers secure against the adaptive chosen plaintext attack such that the PGVHF based on them is not one-way. Some of the block ciphers presented are secure against the adaptive chosen plaintext/ciphertext attack. The proof is constructive: the secure block ciphers are explicitly defined based on which one-way PGVHFs cannot be constructed. These block ciphers have a small amount of non-randomness. Though the non-randomness cannot be used by adversaries of a block cipher, they can be used to break one-wayness of PGVHFs. Informally, the reason is that block ciphers have a secret input (secret key) randomly chosen, while unkeyed hash functions do not have any secret input.

Actually, it is mentioned, for example in [4], that some kinds of non-randomness of underlying block ciphers such as weak keys or fixed points may facilitate adversarial manipulation of PGVHFs. The contribution of this article is that it shows explicitly that security of block ciphers against the adaptive chosen plaintext(/ciphertext) attack is not sufficient for one-wayness of PGVHFs.

This article is organized as follows. Section 2 gives the definitions and basic notations necessary for subsequent discussions. The main result of this article is presented in Section 3. Further considerations are given in Section 4. Section 5 concludes the article with a few open questions.

2 Preliminaries

Let $\{0, 1\}^{\leq k} = \cup_{i=1}^k \{0, 1\}^i$. For $b \in \{0, 1\}$, let b^n represent the sequence of n b ’s. For a $\{0, 1\}$ -sequence x , let $|x|$ represent the length of x . For two $\{0, 1\}$ -sequences x, y , let $x \cdot y$ represent the concatenation of x and y . For two $\{0, 1\}$ -sequences x, y such that $|x| = |y|$, $x \oplus y$ represents bit-wise addition modulo 2.

Let \mathbb{Z}^+ be the set of non-negative integers and \mathbb{R}^+ be the set of non-negative reals. A function $\varepsilon : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ is said to be negligible, if, for every $c > 0$, there

exists some $n_c > 0$ such that $\varepsilon(n) < n^{-c}$ for all $n > n_c$. A function $\sigma : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ is said to be non-negligible if it is not negligible.

2.1 Hash Functions

In this article, unkeyed hash functions are simply called hash functions. A hash function is a function which maps an input of arbitrary length to an output of fixed length.

Definition 1. A function $H = \{h_n \mid h_n : \{0, 1\}^{\leq \ell(n)} \rightarrow \{0, 1\}^n\}$ is called a hash function if

1. $\ell(n) = n^{O(1)}$ and $\ell(n) > n$,
2. the length of the description of h_n is $n^{O(1)}$,
3. there exists a polynomial-time algorithm H such that $H(1^n, h_n, x) = h_n(x)$ for every $x \in \{0, 1\}^{\leq \ell(n)}$. \square

The length of the input is polynomially bounded by the length of the output. This is because only polynomially bounded adversary will be considered in the following discussion.

In this article, the description of a function is denoted by its name as in the above definition; the description of the function h_n is also denoted by h_n .

Three kinds of notions are provided with the security of a hash function; preimage resistance, second-preimage resistance, and collision resistance. Informally, preimage resistance means that, given a hash function and its output, it is intractable to find a preimage which produces the output. Second-preimage resistance means that, given a hash function and its input, it is intractable to find another preimage which produces the same output as the given input. Collision resistance means that, given a hash function, it is intractable to find two inputs which produce the same output.

In this article, only preimage resistance and second-preimage resistance are considered. Formal definitions are given below.

Definition 2. A hash function $H = \{h_n \mid h_n : \{0, 1\}^{\leq \ell(n)} \rightarrow \{0, 1\}^n\}$ is preimage resistant if, for every probabilistic polynomial-time algorithm F ,

$$\Pr[F(1^n, h_n, y) = x' \in \{0, 1\}^{\leq \ell(n)} \wedge h_n(x') = y]$$

is negligible, where $y = h_n(x)$ such that x is selected uniformly from $\{0, 1\}^{\leq \ell(n)}$. The probability is taken over the random selection of x and the random choices of F . \square

Definition 3. A hash function $H = \{h_n \mid h_n : \{0, 1\}^{\leq \ell(n)} \rightarrow \{0, 1\}^n\}$ is second-preimage resistant if, for every probabilistic polynomial-time algorithm F ,

$$\Pr[F(1^n, h_n, x) = x' \in \{0, 1\}^{\leq \ell(n)} \wedge x' \neq x \wedge h_n(x') = h_n(x)]$$

is negligible, where x is selected uniformly from $\{0, 1\}^{\leq \ell(n)}$ and the probability is taken over the random selection of x and the random choices of F . \square

A hash function is usually defined with a round function with fixed length of inputs, which is applied to the input iteratively to produce the output. Thus, this kind of hash function is called an iterative hash function.

Definition 4. A hash function $H = \{h_n \mid h_n : \{0, 1\}^{\leq \ell(n)} \rightarrow \{0, 1\}^n\}$ is called an iterative hash function if h_n is specified with

- a round function $f_n : \{0, 1\}^n \times \{0, 1\}^{m(n)} \rightarrow \{0, 1\}^n$, where $\ell(n) > m(n) > \lceil \log_2(\ell(n) + 1) \rceil$,
- a padding rule **Pad**, and
- an initial-value generator **IV**.

A padding rule **Pad** is a deterministic polynomial-time algorithm which takes 1^n and $x \in \{0, 1\}^{\leq \ell(n)}$ and outputs z defined below. It first divides x into $l = \lceil |x|/m(n) \rceil$ blocks x_1, x_2, \dots, x_l such that $|x_i| = m(n)$ for $i = 1, \dots, l - 1$ and $1 \leq |x_l| \leq m(n)$. Then it outputs $z = (z_1, z_2, \dots, z_{l+1})$ such that

$$z_i = \begin{cases} x_i & \text{for } i = 1, 2, \dots, l - 1, \\ x_l \cdot 0^{m(n)-|x_l|} & \text{for } i = l, \\ \text{bin}(|x|) & \text{for } i = l + 1, \end{cases}$$

where $\text{bin}(|x|)$ is the binary representation of the length of x .

An initial-value generator **IV** is a deterministic polynomial-time algorithm which takes 1^n for input and outputs $IV \in \{0, 1\}^n$.

For $x \in \{0, 1\}^{\leq \ell(n)}$, let $z = (z_1, \dots, z_{l+1})$ be the corresponding padded input. $h_n(x)$ is defined as follows: $h_n(x) = v_{l+1}$, where $v_0 = IV$ and $v_i = f_n(v_{i-1}, z_i)$ for $i = 1, 2, \dots, l + 1$. □

As the padding rule in the above definition, adding a block which contains the length of the input is called MD-strengthening after Merkle [5] and Damgård [1].

For the sake of generality, the initial-value generator is assumed in the above definition. The results presented in the next section is independent of the initial value.

2.2 Block Ciphers

In this article, for block ciphers, it is assumed that the length of a plaintext or a ciphertext is equal to that of the secret key.

Definition 5. A pair of functions $B = (E, D)$ is called a block cipher if

1. $E = \{e_n \mid e_n : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$ and for every $k \in \{0, 1\}^n$, $e_n(k, \cdot) : \{0, 1\}^n \rightarrow \{0, 1\}^n$ is a one-to-one mapping,
2. $D = \{d_n \mid d_n : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$ and for every $k \in \{0, 1\}^n$ and $x \in \{0, 1\}^n$, $d_n(k, e_n(k, x)) = x$,
3. the length of the description of e_n and d_n is $n^{O(1)}$,
4. there exists a polynomial-time algorithm **B** such that $B(1^n, e_n, k, x) = e_n(k, x)$ and $B(1^n, d_n, k, x) = d_n(k, x)$.

E is called an encryption function and D is called a decryption function. □

In the following, security of a block cipher is defined. Adversaries are assumed to make the adaptive chosen plaintext attack or the adaptive chosen plaintext/ciphertext attack. The security goal is indistinguishability of encryption against the attacks.

Definition 6. An adversary \mathcal{A} making the adaptive chosen plaintext attack is defined as follows:

\mathcal{A} is a probabilistic polynomial-time algorithm with an oracle \mathcal{B} . Both \mathcal{A} and \mathcal{B} take 1^n and e_n, d_n for input. Before the execution of \mathcal{A} , \mathcal{B} uniformly selects $k \in \{0, 1\}^n$. First, \mathcal{A} selects and asks $x_i \in \{0, 1\}^n$ to \mathcal{B} , which replies $e_n(k, x_i)$ for $i = 1, 2, \dots, q$. These queries are adaptive; \mathcal{A} may select and ask x_{i+1} after it receives $e_n(k, x_i)$. Then, \mathcal{A} chooses $x^{(0)}, x^{(1)} \in \{0, 1\}^n$ such that $x^{(0)} \neq x^{(1)}$ and $x^{(0)}, x^{(1)} \notin \{x_1, \dots, x_q\}$ and sends them to \mathcal{B} . \mathcal{B} selects $b \in \{0, 1\}$ at random and replies $e_n(k, x^{(b)})$. \mathcal{A} outputs the guess of b .

The advantage of an adversary \mathcal{A} is $|\Pr[\mathcal{A}(1^n, e_n, d_n) = b] - 1/2|$, where the probability is taken over the random choices of \mathcal{A} and \mathcal{B} . \square

Based on the above definition, a block cipher secure against the adaptive chosen plaintext attack is defined in the following way.

Definition 7. A block cipher is secure against the adaptive chosen plaintext attack if, for every adversary making the adaptive chosen plaintext attack, its advantage is negligible. \square

Definition 8. An adversary \mathcal{A} making the adaptive chosen plaintext/ciphertext attack is defined as follows:

\mathcal{A} is a probabilistic polynomial-time algorithm with an oracle \mathcal{B} . Both \mathcal{A} and \mathcal{B} take 1^n and e_n, d_n for input. Before the execution of \mathcal{A} , \mathcal{B} uniformly selects $k \in \{0, 1\}^n$. First, \mathcal{A} selects and asks (x_i, s_i) to \mathcal{B} such that $x_i \in \{0, 1\}^n$ and $s_i \in \{0, 1\}$ for $i = 1, 2, \dots, q$. Then, \mathcal{B} replies $e_n(k, x_i)$ if $s_i = 0$ and $d_n(k, x_i)$ if $s_i = 1$. These queries are adaptive; \mathcal{A} may select and ask (x_{i+1}, s_{i+1}) after it receives the answer to (x_i, s_i) . Then, \mathcal{A} chooses $x^{(0)}, x^{(1)} \in \{0, 1\}^n$ and $s \in \{0, 1\}$ such that $x^{(0)} \neq x^{(1)}$ and

$$x^{(0)}, x^{(1)} \notin \begin{cases} \{x_i \mid s_i = 0\} \cup \{d_n(k, x_j) \mid s_j = 1\} & \text{if } s = 0 \\ \{x_i \mid s_i = 1\} \cup \{e_n(k, x_j) \mid s_j = 0\} & \text{if } s = 1 \end{cases}$$

and sends them to \mathcal{B} . \mathcal{B} selects $b \in \{0, 1\}$ at random and replies $e_n(k, x^{(b)})$ if $s = 0$ or $d_n(k, x^{(b)})$ if $s = 1$. \mathcal{A} outputs the guess of b .

The advantage of an adversary \mathcal{A} is $|\Pr[\mathcal{A}(1^n, e_n, d_n) = b] - 1/2|$, where the probability is taken over the random choices of \mathcal{A} and \mathcal{B} . \square

Based on the above definition, a block cipher secure against the adaptive chosen plaintext/ciphertext attack is defined in the following way.

Definition 9. A block cipher is secure against the adaptive chosen plaintext/ciphertext attack if, for every adversary making the adaptive chosen plaintext/ciphertext attack, its advantage is negligible. \square

2.3 Iterative Hash Functions Based on Block Ciphers

The general model of iterative hash functions considered in this article is defined. This model was first defined by Preneel, Govaerts and Vandewalle [9].

Definition 10. An iterative hash function based on a block cipher is in the general model by Preneel, Govaerts and Vandewalle if its round function

$$f_n(v_{i-1}, z_i) \stackrel{\text{def}}{=} e_n(k, x) \oplus y,$$

where

1. $E = \{e_n \mid e_n : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$ is the encryption function of a block cipher, and
2. $k, x, y \in \{v_{i-1}, z_i, v_{i-1} \oplus z_i\} \cup \{0, 1\}^n$. □

Let us call the iterative hash functions defined above PGVHFs. With a block cipher, 64 kinds of PGVHFs are defined. Let a PGVHF be denoted by $\text{HF}_E^{(k,x,y)}$ if its round function is defined by $e_n(k, x) \oplus y$, where $E = \{e_n \mid e_n : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$. For example, $\text{HF}_E^{(z_i, v_{i-1}, v_{i-1})}$ is a PGVHF with the round function $f_n(v_{i-1}, z_i) = e_n(z_i, v_{i-1}) \oplus v_{i-1}$. This is known as the Davies-Meyer scheme.

3 Secure Block Ciphers Are Not Sufficient for PGVHFs

In this section, the following statement is disproved:

There exists some PGVHF such that, for every secure block cipher B , it is one-way if constructed with B .

The proof is constructive: for each PGVHF, some block cipher is explicitly defined such that the PGVHF is not one-way if it is constructed with the block cipher.

3.1 Secure Block Ciphers Used to Construct Counterexamples

Let $B^* = (E^*, D^*)$ be a block cipher, where $E^* = \{e_n^* \mid e_n^* : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$ and $D^* = \{d_n^* \mid d_n^* : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n\}$. For even n and $x = (x_1, \dots, x_n) \in \{0, 1\}^n$, let $x^L = (x_1, \dots, x_{n/2})$ and $x^R = (x_{n/2+1}, \dots, x_n)$. Let $W(x)$ denote the Hamming weight of $x \in \{0, 1\}^n$. $(0, 0, \dots, 0) \in \{0, 1\}^n$ is simply denoted by 0.

Eight block ciphers $B^{A0}, B^{A1}, B^{B0}, B^{B1}, B^{C0}, B^{C1}, B^{D0}, B^{D1}$ are defined in the following way based on B^* . Only the definitions of the encryption functions are presented. The definitions of the decryption functions can be derived easily from those of the corresponding encryption functions.

$$e_n^{A0}(k, x) = \begin{cases} x & \text{if } W(k) \leq 1, \\ e_n^*(k, x) & \text{otherwise.} \end{cases}$$

$$e_n^{A1}(k, x) = \begin{cases} k \oplus x & \text{if } W(k) \leq 1, \\ e_n^*(k, x) & \text{otherwise.} \end{cases}$$

$$\begin{aligned} &\text{if } k = 0, \text{ then } e_n^{B0}(k, x) = x, \\ &\text{otherwise } e_n^{B0}(k, x) = \begin{cases} k & \text{if } x = k, \\ e_n^*(k, k) & \text{if } x = d_n^*(k, k), \\ e_n^*(k, x) & \text{otherwise.} \end{cases} \end{aligned}$$

$$\begin{aligned} &\text{if } k = 0, \text{ then } e_n^{B1}(k, x) = x, \\ &\text{otherwise } e_n^{B1}(k, x) = \begin{cases} 0 & \text{if } x = k, \\ e_n^*(k, k) & \text{if } x = d_n^*(k, 0), \\ e_n^*(k, x) & \text{otherwise.} \end{cases} \end{aligned}$$

$$\begin{aligned} &\text{if } k = 0, \text{ then } e_n^{C0}(k, x) = (x^L \oplus x^R) \cdot x^L, \\ &\text{otherwise } e_n^{C0}(k, x) = \begin{cases} k & \text{if } x = k, \\ e_n^*(k, k) & \text{if } x = d_n^*(k, k), \\ e_n^*(k, x) & \text{otherwise.} \end{cases} \end{aligned}$$

$$\begin{aligned} &\text{if } k = 0, \text{ then } e_n^{C1}(k, x) = (x^L \oplus x^R) \cdot x^L, \\ &\text{otherwise } e_n^{C1}(k, x) = \begin{cases} 0 & \text{if } x = k, \\ e_n^*(k, k) & \text{if } x = d_n^*(k, 0), \\ e_n^*(k, x) & \text{otherwise.} \end{cases} \end{aligned}$$

$$e_n^{D0}(k, x) = \begin{cases} x & \text{if } W(k \oplus x) \leq 1, \\ e_n^*(k, e_n^*(k, x)) & \text{if } W(k \oplus e_n^*(k, x)) \leq 1, \\ e_n^*(k, x) & \text{otherwise.} \end{cases}$$

$$e_n^{D1}(k, x) = \begin{cases} k \oplus x & \text{if } W(k \oplus x) \leq 1, \\ e_n^*(k, k \oplus e_n^*(k, x)) & \text{if } W(e_n^*(k, x)) \leq 1, \\ e_n^*(k, x) & \text{otherwise.} \end{cases}$$

B^{A0} and B^{A1} have $(n + 1)$ weak keys. The Hamming weight of the weak keys are at most 1. B^{B0}, B^{B1} have a weak key $k = 0$. B^{B0} and B^{C0} as well as B^{B1} and B^{C1} are different from each other only when $k = 0$. B^{C0} and B^{C1} are defined only if n is even.

To simplify the definitions of B^{D0} and B^{D1} , it is assumed that $\{\hat{x} \mid W(k \oplus \hat{x}) \leq 1\} \cap \{\tilde{x} \mid W(k \oplus e_n^*(k, \tilde{x})) \leq 1\} = \phi$ and $\{\hat{x} \mid W(k \oplus \hat{x}) \leq 1\} \cap \{\tilde{x} \mid W(e_n^*(k, \tilde{x})) \leq 1\} = \phi$ for every $k \in \{0, 1\}^n$. In general, it is not restrictive for secure encryption functions.

For the security of the block ciphers defined above, the following two lemmas can be obtained.

Lemma 1. If B^* is secure against the adaptive chosen plaintext attack, then $B^{A0}, B^{A1}, B^{B0}, B^{B1}, B^{C0}, B^{C1}, B^{D0}, B^{D1}$ are secure against the adaptive chosen plaintext attack.

(Proof) It is obvious that both B^{A0} and B^{A1} are secure against the adaptive chosen plaintext attack because the probability that the weak keys are selected is $(n + 1)/2^n$ and negligible.

In the following part of this proof, it is only proved that B^{B^0} and B^{B^1} are secure against the adaptive chosen plaintext attack. It can be proved that B^{C^0} , B^{C^1} , B^{D^0} and B^{D^1} are secure against the adaptive chosen plaintext attack almost in the same way.

Suppose that \mathcal{A} is an adversary against B^* and let $X = \{x_i \mid x_i \in \{0, 1\}^n \text{ for } i = 1, \dots, q\}$ be the set of plaintexts \mathcal{A} asks to the oracle, where q is bounded by some polynomial in n .

If the probability that $\{k, d_n^*(k, k)\} \cap X \neq \phi$ is non-negligible, then \mathcal{A} can guess k correctly with non-negligible probability, which contradicts the assumption that B^* is secure against the adaptive chosen plaintext attack. If the probability that $d_n^*(k, 0) \in X$ is non-negligible, then the following adversary \mathcal{A}' against B^* can be constructed.

1. \mathcal{A}' guess j such that $x_j = d_n^*(k, 0)$.
2. \mathcal{A}' simulates \mathcal{A} and asks x_1, \dots, x_{j-1} to the oracle.
3. When \mathcal{A} generates x_j , \mathcal{A}' terminates the simulation of \mathcal{A} . \mathcal{A}' does not ask x_j to the oracle.
4. \mathcal{A}' randomly selects $x^{(0)}$ and sends $x^{(0)}$ and $x^{(1)} = x_j$ to the oracle.
5. If the reply from the oracle is equal to 0, then \mathcal{A}' outputs 1. Otherwise, \mathcal{A}' randomly selects a bit and outputs it.

The advantage of \mathcal{A}' is non-negligible, which contradicts the assumption that B^* is secure against the adaptive chosen plaintext attack. Thus, for any adversary, the probability that $\{k, d_n^*(k, k), d_n^*(k, 0)\} \cap X \neq \phi$ is negligible. Thus, any adversary for B^{B^0} can ask k or $d_n^*(k, k)$ to the oracle only with negligible probability because $e_n^{B^0}(k, x) = e_n^*(k, x)$ for every $x \notin \{k, d_n^*(k, k)\}$. Any adversary for B^{B^1} can ask k or $d_n^*(k, 0)$ to the oracle only with negligible probability because $e_n^{B^1}(k, x) = e_n^*(k, x)$ for every $x \notin \{k, d_n^*(k, 0)\}$. Thus, any adversary for B^{B^0} or B^{B^1} is only as powerful as the most powerful adversary for B^* . Consequently, both B^{B^0} and B^{B^1} is secure against the adaptive chosen plaintext attack. \square

The next lemma is presented without proof because it can be proved in the same way as the above lemma.

Lemma 2. If B^* is secure against the adaptive chosen plaintext/ciphertext attack, then $B^{A^0}, B^{A^1}, B^{B^0}, B^{C^0}, B^{D^0}$ are secure against the adaptive chosen plaintext/ciphertext attack. \square

It is obvious that B^{B^1}, B^{C^1} and B^{D^1} is not secure against the chosen plaintext/ciphertext attack even if B^* is secure against the chosen plaintext/ciphertext attack. For these block ciphers, if an adversary asks the ciphertext 0 to the oracle, then it can obtain the secret key.

In the following parts of this section, it is proved that each PGVHF is not one-way if its round function is composed with at least one of $B^{A^0}, B^{A^1}, B^{B^0}, B^{B^1}, B^{C^0}, B^{C^1}, B^{D^0}, B^{D^1}$.

3.2 Counterexamples

In the following discussion, for $\text{HF}_E^{(k,x,y)}$, when k , x or y are constant, they are always regarded as 0. This is just for simplicity of the discussion. The following discussion can be easily modified so as to be applied to the case when they are non-zero constant.

Table 1 summarizes the results. It shows with which encryption function each PGVHF is not second-preimage resistant. “-” represents that the PGVHF is not second-preimage resistant with any encryption function.

Theorem 1. The following PGVHFs are not second-preimage resistant:

- i. $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, 0, v_{i-1})}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, 0, z_i)}$, $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, v_{i-1}, 0)}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, v_{i-1}, v_{i-1} \oplus z_i)}$,
 $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, z_i, 0)}$, $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, z_i, v_{i-1} \oplus z_i)}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, v_{i-1})}$,
 $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, z_i)}$,
- ii. $\text{HF}_{EB0}^{(z_i, v_{i-1} \oplus z_i, v_{i-1})}$, $\text{HF}_{EB1}^{(z_i, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}$, $\text{HF}_{EB0}^{(v_{i-1} \oplus z_i, z_i, v_{i-1})}$, $\text{HF}_{EB1}^{(v_{i-1} \oplus z_i, z_i, z_i)}$,
- iii. $\text{HF}_{EC0}^{(v_{i-1}, z_i, z_i)}$, $\text{HF}_{EC1}^{(v_{i-1}, z_i, v_{i-1} \oplus z_i)}$, $\text{HF}_{EC1}^{(v_{i-1}, v_{i-1} \oplus z_i, z_i)}$, $\text{HF}_{EC0}^{(v_{i-1}, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}$,
- iv. $\text{HF}_{ED1}^{(z_i, v_{i-1}, v_{i-1})}$, $\text{HF}_{ED0}^{(z_i, v_{i-1}, v_{i-1} \oplus z_i)}$, $\text{HF}_{ED1}^{(v_{i-1} \oplus z_i, v_{i-1}, v_{i-1})}$, $\text{HF}_{ED0}^{(v_{i-1} \oplus z_i, v_{i-1}, z_i)}$.

(Proof) For each PGVHF listed above and a given preimage, a second preimage will be presented which is of the same length as the given preimage. It is apparent from the following proof that second preimages can be easily found for almost all given preimages.

Let $a_1, a_2, \dots, a_m \in \{0, 1\}^n$ be a padded input corresponding to the given preimage. Let $b_j = f_n(b_{j-1}, a_j)$ for $j = 1, \dots, m$, where b_0 is the initial value. Let $\alpha^{(l)} = (\alpha_1^{(l)}, \dots, \alpha_n^{(l)}) \in \{0, 1\}^n$ such that

$$\alpha_j^{(l)} = \begin{cases} 1 & \text{if } j = l \\ 0 & \text{otherwise,} \end{cases}$$

for $1 \leq l \leq n$.

- (i) For $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, 0, v_{i-1})}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, 0, z_i)}$, $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, v_{i-1}, 0)}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, v_{i-1}, v_{i-1} \oplus z_i)}$,
 $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, z_i, 0)}$, $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, z_i, v_{i-1} \oplus z_i)}$, $\text{HF}_{EA0}^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, v_{i-1})}$ and
 $\text{HF}_{EA1}^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, z_i)}$,

$$v_i = f_n(v_{i-1}, z_i) = \begin{cases} v_{i-1} & \text{if } z_i = v_{i-1}, \\ v_{i-1} \oplus \alpha^{(l)} & \text{if } z_i = v_{i-1} \oplus \alpha^{(l)}. \end{cases}$$

Suppose $m \geq n + 2$. Let $\delta = (\delta_1, \dots, \delta_n) = b_0 \oplus b_n$. Let $a'_1, \dots, a'_n, b'_0, \dots, b'_{n-1} \in \{0, 1\}^n$ such that $b'_0 = b_0$, $b'_j = f_n(b'_{j-1}, a'_j)$ and

$$a'_j = \begin{cases} b'_{j-1} & \text{if } \delta_j = 0, \\ b'_{j-1} \oplus \alpha^{(j)} & \text{if } \delta_j = 1 \end{cases}$$

for $j = 1, \dots, n$. Then, $b'_n = b_n$. Thus, $a'_1, \dots, a'_n, a_{n+1}, \dots, a_m$ is a second preimage if $(a'_1, \dots, a'_n) \neq (a_1, \dots, a_n)$.

(ii) Suppose $m \geq 4$. For $\text{HF}_{E^{B0}}^{(z_i, v_{i-1} \oplus z_i, v_{i-1})}$ and $\text{HF}_{E^{B1}}^{(z_i, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}$, let $a'_j = 0$ and $a'_{j+1} = b_{j+1}$ for some $j \leq m - 3$. Then,

$$\begin{aligned} f(f(b_{j-1}, a'_j), a'_{j+1}) &= f(f(b_{j-1}, 0), b_{j+1}) \\ &= f(0, b_{j+1}) \\ &= b_{j+1}. \end{aligned}$$

For $\text{HF}_{E^{B0}}^{(v_{i-1} \oplus z_i, z_i, v_{i-1})}$ and $\text{HF}_{E^{B1}}^{(v_{i-1} \oplus z_i, z_i, z_i)}$, let $a'_j = b_{j-1}$ and $a'_{j+1} = b_{j+1}$ for some $j \leq m - 3$. Then,

$$\begin{aligned} f(f(b_{j-1}, a'_j), a'_{j+1}) &= f(f(b_{j-1}, b_{j-1}), b_{j+1}) \\ &= f(0, b_{j+1}) \\ &= b_{j+1}. \end{aligned}$$

Thus, for the above four PGVHFs, $a_1, \dots, a_{j-1}, a'_j, a'_{j+1}, a_{j+2}, \dots, a_m$ is a second preimage if $(a'_j, a'_{j+1}) \neq (a_j, a_{j+1})$.

(iii) Suppose $m \geq 4$. For $\text{HF}_{E^{C0}}^{(v_{i-1}, z_i, z_i)}$ and $\text{HF}_{E^{C1}}^{(v_{i-1}, z_i, v_{i-1} \oplus z_i)}$, let $a'_j = b_{j-1}$ and $a'_{j+1} = (b_{j+1}^L \oplus b_{j+1}^R) \cdot b_{j+1}^L$ for some $j \leq m - 3$. Then,

$$\begin{aligned} f(f(b_{j-1}, a'_j), a'_{j+1}) &= f(0, a'_{j+1}) \\ &= (b_{j+1}^R \cdot (b_{j+1}^L \oplus b_{j+1}^R)) \oplus ((b_{j+1}^L \oplus b_{j+1}^R) \cdot b_{j+1}^L) \\ &= b_{j+1}. \end{aligned}$$

For $\text{HF}_{E^{C1}}^{(v_{i-1}, v_{i-1} \oplus z_i, z_i)}$ and $\text{HF}_{E^{C0}}^{(v_{i-1}, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}$, let $a'_j = 0$ and $a'_{j+1} = (b_{j+1}^L \oplus b_{j+1}^R) \cdot b_{j+1}^L$ for some $j \leq m - 3$. Then,

$$f(f(b_{j-1}, a'_j), a'_{j+1}) = f(0, a'_{j+1}) = b_{j+1}.$$

Thus, for the above four PGVHFs, $a_1, \dots, a_{j-1}, a'_j, a'_{j+1}, a_{j+2}, \dots, a_m$ is a second preimage if $(a'_j, a'_{j+1}) \neq (a_j, a_{j+1})$.

(iv) Suppose $m \geq n + 2$. Let $\delta = (\delta_1, \dots, \delta_n) = b_0 \oplus b_n$.

For $\text{HF}_{E^{D1}}^{(z_i, v_{i-1}, v_{i-1})}$ and $\text{HF}_{E^{D0}}^{(z_i, v_{i-1}, v_{i-1} \oplus z_i)}$,

$$v_i = f_n(v_{i-1}, z_i) = \begin{cases} v_{i-1} & \text{if } z_i = v_{i-1}, \\ v_{i-1} \oplus \alpha^{(l)} & \text{if } z_i = v_{i-1} \oplus \alpha^{(l)}. \end{cases}$$

Let $a'_1, \dots, a'_n, b'_0, \dots, b'_{n-1} \in \{0, 1\}^n$ such that $b'_0 = b_0$, $b'_j = f_n(b'_{j-1}, a'_j)$ and

$$a'_j = \begin{cases} b_{j-1} & \text{if } \delta_j = 0, \\ b_{j-1} \oplus \alpha^{(j)} & \text{if } \delta_j = 1 \end{cases}$$

for $j = 1, \dots, n$. Then, $b'_n = b_n$.

For $\text{HF}_{E^{D1}}^{(v_{i-1} \oplus z_i, v_{i-1}, v_{i-1})}$ and $\text{HF}_{E^{D0}}^{(v_{i-1} \oplus z_i, v_{i-1}, z_i)}$,

$$v_i = f_n(v_{i-1}, z_i) = \begin{cases} v_{i-1} & \text{if } z_i = 0, \\ v_{i-1} \oplus \alpha^{(l)} & \text{if } z_i = \alpha^{(l)}. \end{cases}$$

Let $a'_1, \dots, a'_n, b'_0, \dots, b'_{n-1} \in \{0, 1\}^n$ such that $b'_0 = b_0, b'_j = f_n(b'_{j-1}, a'_j)$ and

$$a'_j = \begin{cases} 0 & \text{if } \delta_j = 0, \\ \alpha^{(j)} & \text{if } \delta_j = 1 \end{cases}$$

for $j = 1, \dots, n$. Then, $b'_n = b_n$.

Thus, for the above four PGVHFs, $a'_1, \dots, a'_n, a_{n+1}, \dots, a_m$ is a second preimage if $(a'_1, \dots, a'_n) \neq (a_1, \dots, a_n)$. □

Theorem 2. The following PGVHFs are not preimage resistant:

- i. $\text{HF}_E^{(0, v_{i-1} \oplus z_i, z_i)}$ if $E \in \{E^{A0}, E^{A1}\}$,
- ii. $\text{HF}_{E^{A1}}^{(z_i, v_{i-1}, 0)}, \text{HF}_{E^{A0}}^{(z_i, v_{i-1}, z_i)}, \text{HF}_{E^{A0}}^{(z_i, v_{i-1} \oplus z_i, 0)}, \text{HF}_{E^{A1}}^{(z_i, v_{i-1} \oplus z_i, z_i)}$,
- iii. $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, v_{i-1})}$ and $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, z_i)}$ if $E \in \{E^{B0}, E^{C0}\}$.

(Proof) (i) It is obvious that $\text{HF}_E^{(0, v_{i-1} \oplus z_i, z_i)}$ is not preimage resistant if $E \in \{E^{A0}, E^{A1}\}$, because the round function $f_n(v_{i-1}, z_i) = e_n(0, v_{i-1} \oplus z_i) \oplus z_i = v_{i-1}$ for any v_{i-1}, z_i if $e_n \in \{e_n^{A0}, e_n^{A1}\}$.

(ii) In this part, a procedure is presented to compute a preimage of a given output only for $\text{HF}_{E^{A0}}^{(z_i, v_{i-1}, z_i)}$. Preimages can be obtained in the same way for the other three PGVHFs.

Let b_{n+2} be the given output. The padded input a_1, \dots, a_{n+2} , which consists of $n + 2$ blocks, is obtained in the following way.

1. Fix a_{n+2} and a_{n+1} . a_{n+2} is the binary representation of the length of the preimage. Without loss of generality, suppose that the length of the preimage is $n^2 + 1$. Then, $a_{n+1} \in \{(0, 0, \dots, 0), (1, 0, \dots, 0)\}$.
2. Compute $b_{n+1} = d^{A0}(a_{n+2}, b_{n+2} \oplus a_{n+2})$ and $b_n = d^{A0}(a_{n+1}, b_{n+1} \oplus a_{n+1})$.
3. Let $\delta = (\delta_1, \dots, \delta_n) = b_0 \oplus b_n$, where b_0 is the initial value of the PGVHF. For $j = 1, \dots, n$, let

$$a_j = \begin{cases} 0 & \text{if } \delta_j = 0, \\ \alpha^{(j)} & \text{if } \delta_j = 1. \end{cases}$$

It can be verified easily that a_1, \dots, a_{n+2} is a padded input which produces the output b_{n+2} from the fact that

$$v_i = f_n(v_{i-1}, z_i) = \begin{cases} v_{i-1} & \text{if } z_i = 0, \\ v_{i-1} \oplus \alpha^{(i)} & \text{if } z_i = \alpha^{(i)} \end{cases}$$

for $\text{HF}_{E^{A0}}^{(z_i, v_{i-1}, z_i)}$.

(iii) It is obvious that $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, v_{i-1})}$ and $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, z_i)}$ are not preimage resistant if $E \in \{E^{B0}, E^{C0}\}$. If $e_n \in \{e_n^{B0}, e_n^{C0}\}$, then, for any v_{i-1}, z_i , the round function of $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, v_{i-1})}$ is $e_n(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i) \oplus v_{i-1} = z_i$ and that of $\text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i, z_i)}$ is $e_n(v_{i-1} \oplus z_i, v_{i-1} \oplus z_i) \oplus z_i = v_{i-1}$. □

Table 1. With which block cipher each PGVHF is not second-preimage resistant. Especially, \sharp represents that the corresponding PGVHF is not preimage resistant with the specified block cipher. “-” represents that the corresponding PGVHF is not second-preimage resistant with any block cipher.

k	x	y			
		0	v_{i-1}	z_i	$v_{i-1} \oplus z_i$
0	0	-	-	-	-
0	v_{i-1}	-	-	-	-
0	z_i	-	-	-	-
0	$v_{i-1} \oplus z_i$	-	-	A0 \sharp , A1 \sharp	-
v_{i-1}	0	-	-	-	-
v_{i-1}	v_{i-1}	-	-	-	-
v_{i-1}	z_i	-	-	C0	C1
v_{i-1}	$v_{i-1} \oplus z_i$	-	-	C1	C0
z_i	0	-	-	-	-
z_i	v_{i-1}	A1 \sharp	D1	A0 \sharp	D0
z_i	z_i	-	-	-	-
z_i	$v_{i-1} \oplus z_i$	A0 \sharp	B0	A1 \sharp	B1
$v_{i-1} \oplus z_i$	0	-	A1	A0	-
$v_{i-1} \oplus z_i$	v_{i-1}	A1	D1	D0	A0
$v_{i-1} \oplus z_i$	z_i	A0	B0	B1	A1
$v_{i-1} \oplus z_i$	$v_{i-1} \oplus z_i$	-	A0, B0 \sharp , C0 \sharp	A1, B0 \sharp , C0 \sharp	-

4 Discussion

In this section, a few considerations are given to the results obtained in the previous section.

Twelve Schemes Secure against the Existing Attacks. It is interesting that only for PGVHFs regarded as secure by Preneel, Govaerts and Vandewalle [9], some non-randomness other than weak keys is required for block ciphers to disprove their second-preimage resistance. These are the PGVHFs whose corresponding entries in Table 1 include B0, B1, C0, C1, D0, or D1 without \sharp . Other PGVHFs may not be one-way with any block cipher or block ciphers only with some weak keys.

Preimage Resistance. For some of the PGVHFs in Theorem 1, preimages may be found for given outputs. However, it seems infeasible to find a preimage with polynomial length for such PGVHFs.

Example 1. For $\text{HF}_{EC0}^{(v_{i-1}, z_i, z_i)}$, a preimage can be found for a given output by the following algorithm.

1. Let $b_m \in \{0, 1\}^n$ be the given output.
2. Let $b_{m-1} = 0$ and $a_m = (b_m^L \oplus b_m^R) \cdot b_m^L$. Then, $f_n(b_{m-1}, a_m) = b_m$.
3. Select arbitrarily a_{m-1} such that it ends with at least $n - (\text{len mod } n)$ consecutive 0's, where len is the length of the input represented by a_m . Let $b_{m-2} = a_{m-1}$. Then, $f_n(b_{m-2}, a_{m-1}) = b_{m-1} = 0$.
4. Select a_1, a_2, \dots, a_{m-4} arbitrarily, and compute b_{m-4} .
5. Compute a_{m-3}, a_{m-2} for b_{m-4}, b_{m-2} with the technique used in the proof of Theorem 1.

However, if E^{C0} is random enough except for some non-randomness provided, then b_m is also random. Thus, the probability is negligible that the length of the preimage obtained by the above algorithm is polynomial in n . □

Padding Rules. In the above discussions, only the padding rule with MD-strengthening is considered. Thus, the length of every second preimage obtained in the proof of Theorem 1 is equal to that of the corresponding first preimage. Fixed points of the round function f_n such as $f_n(v_{i-1}, z_i) = v_{i-1}$ are not used in that proof. If the padding rules without MD-strengthening are adopted, that is, the padded inputs do not contain the length of the original inputs, then it is much easier to find the examples which are not one-way for some of PGVHFs using fixed points.

Example 2. For $\text{HF}_E^{(v_{i-1}, v_{i-1} \oplus z_i, z_i)}$, if $E = E^{B0}$ or E^{C0} , then $f_n(v_{i-1}, 0) = v_{i-1}$ for every v_{i-1} . Thus, for every input x , second preimages such as $0 \cdot x$ can be found easily. □

5 Conclusion

In this article, it has been shown that, for every PGVHF, there exist block ciphers secure against the adaptive chosen plaintext attack such that the PGVHF based on them is not one-way. The secure block ciphers have been explicitly defined based on which one-way PGVHFs cannot be constructed. Some of them are secure against the adaptive chosen plaintext/ciphertext attack.

The followings are some open questions.

- Are the following PGVHFs second-preimage resistant if their round functions are composed with any block cipher secure against the adaptive chosen plaintext/ciphertext attack?

$$\text{HF}_E^{(v_{i-1}, z_i, v_{i-1} \oplus z_i)}, \text{HF}_E^{(v_{i-1}, v_{i-1} \oplus z_i, z_i)}, \text{HF}_E^{(z_i, v_{i-1}, v_{i-1})},$$

$$\text{HF}_E^{(z_i, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}, \text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1}, v_{i-1})}, \text{HF}_E^{(v_{i-1} \oplus z_i, z_i, z_i)}.$$

- Are the following PGVHFs preimage resistant if their round functions are composed with any block cipher secure against the adaptive chosen plaintext attack?

$$\text{HF}_E^{(v_{i-1}, z_i, z_i)}, \text{HF}_E^{(v_{i-1}, v_{i-1} \oplus z_i, v_{i-1} \oplus z_i)}, \text{HF}_E^{(z_i, v_{i-1}, v_{i-1} \oplus z_i)},$$

$$\text{HF}_E^{(z_i, v_{i-1} \oplus z_i, v_{i-1})}, \text{HF}_E^{(v_{i-1} \oplus z_i, v_{i-1}, z_i)}, \text{HF}_E^{(v_{i-1} \oplus z_i, z_i, v_{i-1})}.$$

Acknowledgements

The author would like to thank the anonymous reviewers for their valuable comments.

References

1. I. B. Damgård. A design principle for hash functions. In *CRYPTO'89*, pages 416–427, 1990. Lecture Notes in Computer Science 435.
2. D. Davies and W. L. Price. Digital signatures, an update. In *Proceedings of the 5th International Conference on Computer Communication*, pages 845–849, 1984.
3. S. M. Matyas, C. H. Meyer, and J. Oseas. Generating strong one-way functions with cryptographic algorithm. *IBM Technical Disclosure Bulletin*, 27:5658–5659, 1985.
4. A. Menezes, P. van Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.
5. R. C. Merkle. A fast software one-way hash function. *Journal of Cryptology*, 3:43–58, 1990.
6. C. H. Meyer and M. Schilling. Secure program load with manipulation detection code. In *Proceedings of the 6th Worldwide Congress on Computer and Communications Security and Protection (SECURICOM'88)*, pages 111–130, 1988.
7. B. Preneel. *Analysis and Design of Cryptographic Hash Functions*. PhD thesis, Katholieke Universiteit Leuven, 1993.
8. B. Preneel. The state of cryptographic hash functions. In *Lectures on Data Security*, pages 158–182, 1998. Lecture Notes in Computer Science 1561.
9. B. Preneel, R. Govaerts, and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. In *CRYPTO'93*, pages 368–378, 1994. Lecture Notes in Computer Science 773.
10. D. R. Simon. Finding collisions on a one-way street: Can secure hash functions be based on general assumptions? In *EUROCRYPT'98*, pages 334–345, 1998. Lecture Notes in Computer Science 1403.