PAPER Special Section on Information Theory and Its Applications

Efficient Pseudorandom-Function Modes of a Block-Cipher-Based Hash Function

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

SUMMARY This article discusses the provable security of pseudorandom-function (PRF) modes of an iterated hash function using a block cipher. The iterated hash function uses the Matyas-Meyer-Oseas (MMO) mode for the compression function and the Merkle-Damgård with a permutation (MDP) for the domain extension transform. It is shown that the keyed-via-IV mode and the key-prefix mode of the iterated hash function are pseudorandom functions if the underlying block cipher is a pseudorandom permutation under a related-key attack with respect to the permutation used in MDP. More precisely, the key-prefix mode also requires that $E_{IV}(K) \oplus K$ is pseudoramdom, where E is the underlying block cipher, IV is the fixed initial value of the hash function, and K is a secret key. It is also confirmed that the MMO compression function is the best choice with MDP among the block-cipher-based compression functions in the Preneel-Govaerts-Vandewalle model in terms of the provable security.

key words: *hash function, pseudorandom function, block cipher*

1. Introduction

(1) Background

In many textbooks on cryptography, a (cryptographic) hash function is defined to be a function mapping an input string of arbitrary length to an output string of fixed length, and satisfying preimage resistance, second-preimage resistance and collision resistance. However, hash functions are used in almost all cryptographic schemes, and required various security properties other than the three listed above. For example, a hash function is used to instantiate a random oracle. It is also used to construct a pseudorandom bit generator and a pseudorandom function. (Second-) preimage resistance and collision resistance do not validate such usage in general.

(2) Contribution

This article discusses the provable security of pseudorandom-function (PRF) modes of an iterated hash function using a block cipher. The iterated hash function uses the Matyas-Meyer-Oseas (MMO) mode [8] for the compression function and the Merkle-Damgård with a permutation (MDP) [7] for the domain extension transform. It is called MDP-MMO in this article.

The widely used PRF using a hash function is HMAC [2]. However, it is not very efficient. It invokes the hash function twice to process a given input. In this article, two more efficient PRF modes are considered. One is called the keyed-via-IV mode. It simply replaces the initial value of the underlying hash function with the secret key. The other is called the key-prefix mode. It first prepends the secret key to a given input, and then feeds it to the underlying hash function.

It is shown that the keyed-via-IV mode and the key-prefix mode of MDP-MMO are PRFs if the underlying block cipher is a pseudorandom permutation (PRP) under a related-key attack with respect to the permutation of MDP. The novelty of the result is that the PRF property of the modes is reduced to the PRP property of the underlying block cipher, not simply the PRF property of the compression function.

Actually, the key-prefix mode also requires that $E_{IV}(K) \oplus K$ is pseudoramdom, where E is the underlying block cipher, IV is the fixed initial value of the hash function, and K is a secret key. This property cannot be implied by the PRP property of E since the key of E is a fixed public constant. It does not seem difficult, however, to design a block cipher with the property.

The other contribution of the paper is that it confirms that the MMO compression function is the best choice with MDP among the block-cipher-based compression functions in the PGV model [11] in terms of the provable security.

(3) Related Work

Hirose and Kuwakado [6] discussed the following security properties of MDP-MMO: Collision resistance, indifferentiability from a variable-input-length (VIL) random oracle, and pseudorandomness of HMAC using MDP-MMO. Their results imply that the security of an iterated hash function is reduced to the security of the underlying block cipher to more extent with the MMO compression function than with the Davies-Meyer (DM) compression function.

Hirose, Park and Yun [7] proposed MDP, and showed that the keyed-via-IV mode and the key-prefix mode of an iterated hash function using MDP are PRFs on the assumption that the compression function is a PRF under a related key attack with respect to the permutation of MDP. MDP is one of the simplest and the most efficient domain extension transforms. Some of the first round candidates of NIST Cryptographic Hash Algorithm Competition [10] such as ARIRANG, Cheetah and CHI adopted the idea of MDP for

Manuscript received January 16, 2009.
Manuscript revised April 25, 2009.

[†]The author is with Graduate School of Engineering, University of Fukui, Fukui-shi, 910-8507 Japan. http://digcode2.fuee.fukui-u.ac.jp/~hirose/

[††]The author is with Graduate School of Engineering, Kobe University, Kobe-shi, 657-8501 Japan.

a) E-mail: hrs_shch@u-fukui.ac.jp

DOI: 10.1587/transfun.E92.A.2447

their domain extension transforms.

Bellare, Canetti and Krawczyk [2] showed that the cascade construction of a fixed-input-length PRF with prefix-free encoding is a variable-input-length PRF. Bellare [1] showed that HMAC is a PRF if the compression function of the underlying hash function is a PRF with two keying strategies. Yasuda proposed interesting PRF modes and provided the security proofs [13]–[15]. In his security proofs, it is also assumed that the compression function is a PRF. He did not consider its internal structure, either.

Preneel, Govaerts and Vandewalle defined a model of compression functions using a block cipher (the PGV model), which covers the Davies-Meyer, Matyas-Meyer-Oseas and Miyaguchi-Preneel modes [11]. They also provided a security analysis of the modes in their model against several generic attacks. Black, Rogaway and Shrimpton analyzed the modes in the PGV model in terms of provable security [4].

(4) Organization

Some notations and definitions are given in Sect. 2. MDP-MMO is described in Sect. 3. Two PRF modes and their security analysis are presented in Sect. 4. Section 5 shows that MMO is most compatible with MDP in terms of provable security.

2. Definitions

Let $\mathsf{Func}(D, R)$ be the set of all functions from D to R, and $\mathsf{Perm}(D)$ be the set of all permutations on D. Let $s \xleftarrow{\$} S$ represent that an element s is selected from the set S under the uniform distribution.

2.1 Pseudorandom Bit Generator

Let g be a function such that $g : \{0,1\}^n \to \{0,1\}^l$, where $n < l$. Let A be a probabilistic algorithm which outputs 0 or 1 for a given input in $\{0,1\}^l$. The prbg-advantage of A against g is defined as follows:

$$\mathrm{Adv}_g^{\mathrm{prbg}}(A) = \Big|\Pr[A(g(k)) = 1 \mid k \xleftarrow{\$} \{0,1\}^n] - \Pr[A(s) = 1 \mid s \xleftarrow{\$} \{0,1\}^l]\Big|,$$

where the probabilities are taken over the coin tosses by A and the uniform distributions on $\{0,1\}^n$ and $\{0,1\}^l$. g is called a pseudorandom bit generator (PRBG) if $\mathrm{Adv}_g^{\mathrm{prbg}}(A)$ is negligible for any efficient A.

We will abuse the definition for $n = l$. Actually, for example, the identity function is trivially a PRBG for $n = l$.

2.2 Pseudorandom Function

Let $f : K \times D \to R$ be a function family from D to R with a key space K. $f(k, \cdot)$ is often denoted by $f_k(\cdot)$. Let A be a probabilistic algorithm with oracle access to a function from D to R. A outputs 0 or 1. The prf-advantage of A against f is defined as follows:

$$\mathrm{Adv}_f^{\mathrm{prf}}(A) = \Big|\Pr[A^{f_k} = 1 \mid k \xleftarrow{\$} K] - \Pr[A^{\rho} = 1 \mid \rho \xleftarrow{\$} \mathsf{Func}(D, R)]\Big|,$$

where the probabilities are taken over the coin tosses by A and the uniform distributions on K and $\mathsf{Func}(D, R)$. f is called a pseudorandom function (PRF) if $\mathrm{Adv}_f^{\mathrm{prf}}(A)$ is negligible for any efficient A.

Let $p : K \times D \to D$ be a permutation family on D with a key space K. The prp-advantage of A against p is defined similarly:

$$\mathrm{Adv}_p^{\mathrm{prp}}(A) = \Big|\Pr[A^{p_k} = 1 \mid k \xleftarrow{\$} K] - \Pr[A^{\rho} = 1 \mid \rho \xleftarrow{\$} \mathsf{Perm}(D)]\Big|,$$

where the probabilities are taken over the coin tosses by A and the uniform distributions on K and $\mathsf{Perm}(D)$. p is called a pseudorandom permutation (PRP) if $\mathrm{Adv}_p^{\mathrm{prp}}(A)$ is negligible for any efficient A.

2.3 Pseudorandom Function under Related-Key Attack

Pseudorandom functions under related-key attacks are first formalized by Bellare and Kohno [3]. We only consider a related-key attack with respect to a permutation φ as in [7]. We refer to this type of attack as the φ-related-key attack. Let A be a probabilistic algorithm with oracle access to a pair of functions from D to R. Each query by A is an element in D. A sends it to one of the functions, which returns a corresponding element in R. The prf-rka-advantage of A against f under the φ-related-key attack is given by

$$\mathrm{Adv}_{\varphi,f}^{\text{prf-rka}}(A) = \Big|\Pr[A^{f_k, f_{\varphi(k)}} = 1 \mid k \xleftarrow{\$} K] - \Pr[A^{\rho,\rho'} = 1 \mid \rho, \rho' \xleftarrow{\$} \mathsf{Func}(D, R)]\Big|,$$

where the probabilities are taken over the coin tosses by A and the uniform distributions on K and $\mathsf{Func}(D, R)$. f is called a φ-rka-secure PRF if $\mathrm{Adv}_{\varphi,f}^{\text{prf-rka}}(A)$ is negligible for any efficient A.

For a permutation, the prp-rka-advantage and the φ-rka-secure PRP can also be defined similarly.

3. MDP with MMO

We denote concatenation of sequences by $\|$. For sequences $M_1, \ldots, M_N$, we often denote $M_1 \| M_2 \| \cdots \| M_N$ simply by $M_1 M_2 \cdots M_N$. Let $\mathcal{B} = \{0,1\}^n$ and $\mathcal{B}^+ = \cup_{i=1}^{\infty} \mathcal{B}^i$.

Let $E : \mathcal{B} \times \mathcal{B} \to \mathcal{B}$ be an (n, n) block cipher, where the first $\mathcal{B}$ is the key space. The Matyas-Meyer-Oseas (MMO) compression function [9] $h : \mathcal{B} \times \mathcal{B} \to \mathcal{B}$ with E is defined as follows: $h(s, x) = E_s(x) \oplus x$, where s is a chaining variable and x is a message block.

The MDP transform [7] of h with a permutation π is

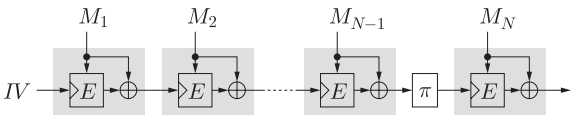

Fig. 1 MDP-MMO$[E, \pi, IV](M)$. E is an underlying (n, n) block cipher. $\mathsf{pad}(M) = M_1 M_2 \cdots M_N$.

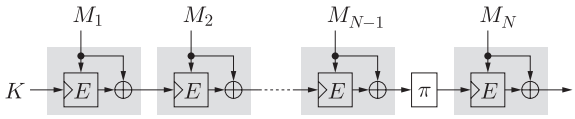

Fig. 2 KMDP-MMO. K is a secret key.

denoted by $h^\circ_\pi : \mathcal{B} \times \mathcal{B}^+ \to \mathcal{B}$ and defined as follows: For $s \in \mathcal{B}$ and $M_1 \| M_2 \| \cdots \| M_N$ ($M_i \in \mathcal{B}$),

1. $s_0 = s$,
2. if $N \geq 2$, then $s_i = h(s_{i-1}, M_i)$ for $1 \leq i \leq N-1$,
3. $s_N = h(\pi(s_{N-1}), M_N)$,
4. $h^\circ_\pi(s, M_1 M_2 \cdots M_N) \stackrel{\text{def}}{=} s_N$.

A padding function $\mathsf{pad} : \{0,1\}^* \to \mathcal{B}^+$ is also necessary for the preprocessing of a given message of arbitrary length.

Now, MDP-MMO is a scheme to construct a hash function using a block cipher $E : \mathcal{B} \times \mathcal{B} \to \mathcal{B}$, a permutation $\pi : \mathcal{B} \to \mathcal{B}$ and an initial value $IV \in \mathcal{B}$, which is defined as follows:

$$\mathsf{MDP\text{-}MMO}[E, \pi, IV](M) \stackrel{\text{def}}{=} h^\circ_\pi(IV, \mathsf{pad}(M)).$$

MDP-MMO is illustrated in Fig. 1.

4. PRF Modes of MDP-MMO

For the PRF modes given in this section, any unambiguous padding suffices. Thus, we will assume that the length of a message input M is always a multiple of n, and do without pad. Namely, $M = M_1 \| \cdots \| M_N$, where $M_i \in \mathcal{B}$ for $1 \leq i \leq N$. M_i is called a message block.

4.1 Keyed-via-IV Mode

A PRF is obtained from MDP-MMO by replacing the fixed initial value with a secret key. The function, KMDP-MMO, is illustrated in Fig. 2. It is simply $h^\circ_\pi(K, \cdot)$.

The security of KMDP-MMO is reduced to the security of the underlying block cipher. It resists any distinguishing attack that requires much fewer than $2^{n/2}$ queries if the underlying block cipher is a π-rka-secure PRP.

Theorem 1: Let A be a prf-adversary against h°_π. Suppose that A runs in time at most t, and makes at most q queries, and each query has at most ℓ message blocks. Then, there exists a prp-rka-adversary B against E such that

$$\mathrm{Adv}^{\text{prf}}_{h^\circ_\pi}(A) \leq \ell q \cdot \mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B) + \frac{\ell q(q-1)}{2^{n+1}}.$$

B makes at most q queries and runs in time at most $t + O(\ell q T_E)$, where T_E represents the time required to compute E.

π should be a permutation with at most a negligible number of fixed points. Otherwise, E cannot be a π-rka-secure PRP. Examples of possible candidates for π are bitwise addition of a nonzero constant or cyclic shift.

Let us consider the tightness of the bound given in Theorem 1. The attack in [12] can break KMDP-MMO with about $2^{n/2}/\sqrt{\ell}$ queries. Suppose that the best π-related-key attack against E as a PRP is exhaustive key search. Then, since the time complexity of B is $t + O(\ell q T_E)$, B can try $t/T_E + O(\ell q)$ keys, that is,

$$\mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B) \leq \frac{t/T_E + O(\ell q)}{2^n},$$

and

$$\mathrm{Adv}^{\text{prf}}_{h^\circ_\pi}(A) \leq \frac{\ell q t/T_E}{2^n} + \frac{O((\ell q)^2)}{2^n}.$$

The right side exceeds 1 if $q \approx 2^{n/2}/\ell$. Thus, the gap is the factor of $\sqrt{\ell}$ or more.

Theorem 1 directly follows from two lemmas given in the remaining part.

Let A be an adversary with access to m pairs of oracles $u_1, u'_1, u_2, u'_2, \ldots, u_m, u'_m$. Each query by A is directed to just one of the $2m$ oracles. Let us define the following notation:

$$\langle u_j, u'_j \rangle_{j=1}^m = u_1, u'_1, u_2, u'_2, \ldots, u_m, u'_m.$$

The m-prf-rka-advantage of A against h under the π-related-key attack is defined as follows:

$$\begin{aligned}&\mathrm{Adv}^{m\text{-prf-rka}}_{\pi,h}(A)\\ &= \Big|\Pr[A^{\langle h_{K_j}, h_{\pi(K_j)} \rangle_{j=1}^m} = 1 \mid K_1, \ldots, K_m \stackrel{\$}{\leftarrow} \mathcal{B}]\\ &\quad - \Pr[A^{\langle \rho_j, \rho'_j \rangle_{j=1}^m} = 1 \mid \langle \rho_j, \rho'_j \rangle_{j=1}^m \stackrel{\$}{\leftarrow} \mathsf{Func}(\mathcal{B}, \mathcal{B})]\Big|.\end{aligned}$$

Lemma 1 ([7]): Let A be a prf-adversary against h°_π. Suppose that A runs in time at most t, and makes at most q queries, and each query has at most ℓ message blocks. Then, there exists a prf-rka-adversary B against h with access to q pairs of oracles such that

$$\mathrm{Adv}^{\text{prf}}_{h^\circ_\pi}(A) \leq \ell \cdot \mathrm{Adv}^{q\text{-prf-rka}}_{\pi,h}(B).$$

B makes at most q queries and runs in time at most $t + O(\ell q T_h)$, where T_h represents the time required to compute h.

A proof of Lemma 1 is given in Appendix. It is based on the hybrid argument [5].

Lemma 2: Let $h_K(x) = E_K(x) \oplus x$. Let A be a prf-rka-adversary against h with m pairs of oracles. Suppose that A runs in time at most t, and makes at most q queries. Then, there exists a prp-rka-adversary B against E such that

$$\mathrm{Adv}^{m\text{-prf-rka}}_{\pi,h}(A) \leq m \cdot \mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B) + \frac{q(q-1)}{2^{n+1}}.$$

B makes at most q queries and runs in time at most $t + O(q\,T_E)$, where T_E represents the time required to compute E.

A proof of this lemma is given below. It is also based on the hybrid argument.

Proof : For a permutation $\varpi \in \mathsf{Perm}(\mathcal{B})$, let $\tilde{\varpi}(x) = \varpi(x) \oplus x$.

$$\begin{aligned}
&\mathrm{Adv}^{m\text{-prf-rka}}_{\pi,h}(A)\\
&= \Big|\Pr[A^{\langle h_{K_j}, h_{\pi(K_j)}\rangle^m_{j=1}} = 1 \mid K_1, \ldots, K_m \xleftarrow{\$} \mathcal{B}]\\
&\quad - \Pr[A^{\langle \rho_j, \rho'_j\rangle^m_{j=1}} = 1 \mid \langle \rho_j, \rho'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Func}(\mathcal{B},\mathcal{B})]\Big|\\
&\le \Big|\Pr[A^{\langle h_{K_j}, h_{\pi(K_j)}\rangle^m_{j=1}} = 1 \mid K_1, \ldots, K_m \xleftarrow{\$} \mathcal{B}]\\
&\quad - \Pr[A^{\langle \tilde{\varpi}_j, \tilde{\varpi}'_j\rangle^m_{j=1}} = 1 \mid \langle \varpi_j, \varpi'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\Big|\\
&\quad + \Big|\Pr[A^{\langle \tilde{\varpi}_j, \tilde{\varpi}'_j\rangle^m_{j=1}} = 1 \mid \langle \varpi_j, \varpi'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\\
&\quad - \Pr[A^{\langle \rho_j, \rho'_j\rangle^m_{j=1}} = 1 \mid \langle \rho_j, \rho'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Func}(\mathcal{B},\mathcal{B})]\Big|.
\end{aligned}$$

For $0 \le i \le m$, let O_i be m pairs of oracles such that $h_{K_1}, h_{\pi(K_1)}, \ldots, h_{K_i}, h_{\pi(K_i)}, \tilde{\varpi}_{i+1}, \tilde{\varpi}'_{i+1}, \ldots, \tilde{\varpi}_m, \tilde{\varpi}'_m$, where $K_1, \ldots, K_i \xleftarrow{\$} \mathcal{B}$ and $\varpi_{i+1}, \varpi'_{i+1}, \ldots, \varpi_m, \varpi'_m \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})$. Notice that $O_0 = \langle \tilde{\varpi}_j, \tilde{\varpi}'_j\rangle^m_{j=1}$ and $O_m = \langle h_{K_j}, h_{\pi(K_j)}\rangle^m_{j=1}$.

A prp-rka-adversary B is constructed using A as a subroutine. The algorithm of B with oracles u, u' is given below. u, u' are either $E_K, E_{\pi(K)}$ or ϖ, ϖ', where $K \xleftarrow{\$} \mathcal{B}$ and $\varpi, \varpi' \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})$.

1. $i \xleftarrow{\$} \{1, 2, \ldots, m\}$.
2. runs A with oracles $h_{K_1}, h_{\pi(K_1)}, \ldots, h_{K_{i-1}}, h_{\pi(K_{i-1})}, \tilde{u}, \tilde{u}', \tilde{\varpi}_{i+1}, \tilde{\varpi}'_{i+1}, \ldots, \tilde{\varpi}_m, \tilde{\varpi}'_m$, where $K_1, \ldots, K_{i-1} \xleftarrow{\$} \mathcal{B}$ and $\varpi_{i+1}, \varpi'_{i+1}, \ldots, \varpi_m, \varpi'_m \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})$.
3. outputs A's output.

Then,

$$\Pr[B^{E_K, E_{\pi(K)}} = 1 \mid K \xleftarrow{\$} \mathcal{B}] = \frac{1}{m}\sum_{i=1}^{m} \Pr[A^{O_i} = 1]$$

and

$$\begin{aligned}
&\Pr[B^{\varpi,\varpi'} = 1 \mid \varpi, \varpi' \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\\
&= \frac{1}{m}\sum_{i=0}^{m-1} \Pr[A^{O_i} = 1].
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B) &= \Big|\Pr[B^{E_K, E_{\pi(K)}} = 1 \mid K \xleftarrow{\$} \mathcal{B}]\\
&\qquad - \Pr[B^{\varpi,\varpi'} = 1 \mid \varpi, \varpi' \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\Big|\\
&= \frac{1}{m}\Big|\Pr[A^{O_m} = 1] - \Pr[A^{O_0} = 1]\Big|.
\end{aligned}$$

B makes at most q queries and runs in time at most $t + O(q\,T_E)$. There may exist an algorithm with the same resources and larger advantage. Let us also call it B. Then,

$$\begin{aligned}
&\Big|\Pr[A^{O_m} = 1] - \Pr[A^{O_0} = 1]\Big|\\
&= \Big|\Pr[A^{\langle h_{K_j}, h_{\pi(K_j)}\rangle^m_{j=1}} = 1 \mid K_1, \ldots, K_m \xleftarrow{\$} \mathcal{B}]\\
&\quad - \Pr[A^{\langle \tilde{\varpi}_j, \tilde{\varpi}'_j\rangle^m_{j=1}} = 1 \mid \langle \varpi_j, \varpi'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\Big|\\
&\le m \cdot \mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B).
\end{aligned}$$

It is possible to distinguish $\tilde{\varpi}_1, \tilde{\varpi}'_1, \ldots, \tilde{\varpi}_m, \tilde{\varpi}'_m$ and $\rho_1, \rho'_1, \ldots, \rho_m, \rho'_m$ only by the fact that there may be a collision for $\rho(x) \oplus x$ for $\rho \in \mathsf{Func}(\mathcal{B},\mathcal{B})$. Thus, since A makes at most q queries,

$$\begin{aligned}
&\Big|\Pr[A^{\langle \tilde{\varpi}_j, \tilde{\varpi}'_j\rangle^m_{j=1}} = 1 \mid \langle \varpi_j, \varpi'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Perm}(\mathcal{B})]\\
&\quad - \Pr[A^{\langle \rho_j, \rho'_j\rangle^m_{j=1}} = 1 \mid \langle \rho_j, \rho'_j\rangle^m_{j=1} \xleftarrow{\$} \mathsf{Func}(\mathcal{B},\mathcal{B})]\Big|\\
&\le \frac{q(q-1)}{2^{n+1}}.
\end{aligned}$$

□

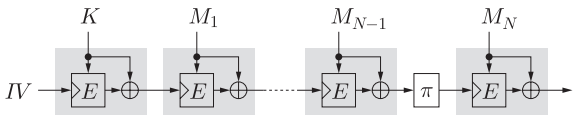

Fig. 3 KPMDP-MMO. K is a secret key.

4.2 Key-Prefix Mode

The key-prefix mode is a method to construct a PRF with a given hash function. It simply feeds $K\|M$ to the hash function as an input, where K is a secret key and M is a message input. The mode with MDP-MMO is illustrated in Fig. 3. K is assumed to be in $\mathcal{B}$. We call the function KPMDP-MMO. This mode uses MDP-MMO as a black box. In this sense, it is similar to HMAC. However, it is more efficient than HMAC.

Let $\nu_E : \mathcal{B} \to \mathcal{B}$ be a function such that $\nu_E(K) = E_{IV}(K) \oplus K$. KPMDP-MMO with a key $K \in \mathcal{B}$ and a message input $M \in \mathcal{B}^+$ is $h^\circ_\pi(\nu_E(K), M)$. Let us denote it by $(h^\circ_\pi \diamond \nu_E)(K, M)$.

KPMDP-MMO resists any distinguishing attack that requires much fewer than $2^{n/2}$ queries if the underlying block cipher E is a π-rka-secure PRP and ν_E is a PRBG.

Theorem 2: Let A be a prf-adversary against $h^\circ_\pi \diamond \nu_E$. Suppose that A runs in time at most t, and makes at most q queries, and each query has at most ℓ message blocks. Then, there exist a prp-rka-adversary B against E, and a prbg-adversary B' against ν_E such that

$$\begin{aligned}
&\mathrm{Adv}^{\text{prf}}_{h^\circ_\pi \diamond \nu_E}(A) \le\\
&\qquad \ell q \cdot \mathrm{Adv}^{\text{prp-rka}}_{\pi,E}(B) + \mathrm{Adv}^{\text{prbg}}_{\nu_E}(B') + \frac{\ell q(q-1)}{2^{n+1}}.
\end{aligned}$$

B makes at most q queries and runs in time at most $t +$

$O(\ell q T_E)$. B' runs in time at most $t + O(\ell q T_E)$, where T_E represents the time required to compute E.

Theorem 2 directly follows from Theorem 1 and the following lemma. It says that $h_\pi^\circ \diamond \nu_E$ is a PRF if h_π° is a PRF and ν_E is a PRBG. The proof is easy and omitted.

Lemma 3: Let A be a prf-adversary against $h_\pi^\circ \diamond \nu_E$. Suppose that A runs in time at most t and makes at most q queries, and each query has at most ℓ message blocks. Then, there exist a prf-adversary B against h_π° and a prbg-adversary B' against ν_E such that

$$\mathrm{Adv}^{\mathrm{prf}}_{h_\pi^\circ \diamond \nu_E}(A) \le \mathrm{Adv}^{\mathrm{prf}}_{h_\pi^\circ}(B) + \mathrm{Adv}^{\mathrm{prbg}}_{\nu_E}(B').$$

B runs in time at most $t + O(\ell q n)$, makes at most q queries, and each query has at most ℓ message blocks. B' runs in time at most $t + O(\ell q T_h)$, where T_h represents the time required to compute h.

5. Discussion

It is discussed in this section if other block-cipher-based compression functions are useful for PRF modes given in the previous section as well as MMO. Table 1 gives 20 PGV compression functions for collision-resistant hash functions in the ideal cipher model [4]. h_1 is MMO, h_5 is Davies-Meyer, and h_3 is Miyaguchi-Preneel.

If the key of E is M_i, then it is not secret and fully controlled by an adversary. Thus, it is impossible to reduce the security of the PRF modes in the previous section to the security of E for h_j, where $j \in \{5, 6, 7, 8, 15, 17, 19, 20\}$. In the remaining part, counterexamples are given which imply the impossibility to reduce the security of PRF modes to the security of E for h_j, where $j \in \{2, 3, 4, 9, 10, 11, 12, 13, 14, 16, 18\}$. The observations show that MMO is the best choice for the PRF modes in terms of provable security.

Example 1: Suppose that there exists some nonzero $d \in \{0,1\}^n$ such that $E_K(M) = E_{K\oplus d}(M \oplus d) \oplus d$ for every K and M. E can be a PRP under the chosen plaintext attack. It should be mentioned that DES has this kind of property for $d = 1^n$. If E is used for h_9, then

$$\begin{aligned} h_9(K, M_i \oplus d) &= E_{K\oplus M_i\oplus d}(M_i \oplus d) \oplus M_i \oplus d \\ &= E_{K\oplus M_i}(M_i) \oplus M_i \\ &= h_9(K, M_i). \end{aligned}$$

It implies that $h_9(K, \cdot)$ is not a PRF. h_{11}, h_{13} and h_{14} are not PRFs, either, if E is used for them.

Example 2: Suppose that there exists some nonzero $d \in \{0,1\}^n$ such that $E_K(M) = E_{K\oplus d}(M)$ for every K and M. E can be a PRP under the chosen plaintext attack. If E is used for h_{10}, then

$$h_{10}(K, M_i \oplus d) = h_{10}(K, M_i).$$

Thus, $h_{10}(K, \cdot)$ is not a PRF. h_{12}, h_{16} and h_{18} are not PRFs, either, if E is used for them.

Counterexamples in Examples 1 and 2 are block ciphers insecure under the related-key attack with respect to d. On the other hand, two PRF modes in the previous section require a block cipher secure under the related-key attack with respect to π. Major difference is as follows:

- The designer of E in MDP-MMO has only to worry about the related-key attack with respect to π, which is also set by the designer.
- The designer of E in h_9, for example, has to worry about all nonzero d.

Example 3: Suppose that $E_K(K) = K$ for every K. E can be a PRP under the chosen plaintext attack. If E is used for h_4, then

$$h_4(K, 0^n) = E_K(K) \oplus 0^n = K.$$

We can check if the oracle is $h_4(K, \cdot)$ or not with another query. Thus, $h_4(K, \cdot)$ is not a PRF. h_2 is not a PRF, either, if E is used for it.

Example 4: Suppose that $E_K(M) = K \oplus M$ for every K and M. E is a PRP under the chosen plaintext attack for any adversary making only one query (E is one-time pad). If E is used for h_3, then

$$h_3(K, 0^n) = E_K(0^n) \oplus K = 0^n.$$

It implies that $h_3(K, \cdot)$ is not a PRF against an adversary making only one query.

Example 4 may be insignificant. However, it still implies that MMO seems preferable to Miyaguchi-Preneel in terms of provable security.

Table 1 20 PGV compression functions for collision-resistant hash functions in the ideal cipher model. M_i is a message block, and v_{i-1} is a chaining variable. $w_i = M_i \oplus v_{i-1}$. c is a constant.

h_1	$E_{v_{i-1}}(M_i) \oplus M_i$	h_{11}	$E_{w_i}(M_i) \oplus v_{i-1}$
h_2	$E_{v_{i-1}}(w_i) \oplus w_i$	h_{12}	$E_{w_i}(v_{i-1}) \oplus M_i$
h_3	$E_{v_{i-1}}(M_i) \oplus w_i$	h_{13}	$E_{w_i}(M_i) \oplus c$
h_4	$E_{v_{i-1}}(w_i) \oplus M_i$	h_{14}	$E_{w_i}(M_i) \oplus w_i$
h_5	$E_{M_i}(v_{i-1}) \oplus v_{i-1}$	h_{15}	$E_{M_i}(v_{i-1}) \oplus c$
h_6	$E_{M_i}(w_i) \oplus w_i$	h_{16}	$E_{w_i}(v_{i-1}) \oplus c$
h_7	$E_{M_i}(v_{i-1}) \oplus w_i$	h_{17}	$E_{M_i}(v_{i-1}) \oplus M_i$
h_8	$E_{M_i}(w_i) \oplus v_{i-1}$	h_{18}	$E_{w_i}(v_{i-1}) \oplus w_i$
h_9	$E_{w_i}(M_i) \oplus M_i$	h_{19}	$E_{M_i}(w_i) \oplus c$
h_{10}	$E_{w_i}(v_{i-1}) \oplus v_{i-1}$	h_{20}	$E_{M_i}(w_i) \oplus M_i$

6. Conclusion

This article has discussed the provable security of two efficient PRF modes of an iterated hash function with the MMO compression function and the MDP domain extension transform. It has also been shown that the MMO compression

function is best suited to MDP among the compression functions in the Preneel-Govaerts-Vandewalle model in terms of provable security.

Combined with [6], this article implies that using MDP-MMO is a good strategy to construct a secure block-cipher-based hash function.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments. They would also like to thank Dr. Yoshida and Dr. Ideguchi at Hitachi, Ltd. and Prof. Ohta at The University of Electro-Communications for their fruitful discussions and comments on this research. This research was supported by the National Institute of Information and Communications Technology, Japan.

References

[1] M. Bellare, "New proofs for NMAC and HMAC: Security without collision-resistance," CRYPTO 2006 Proc., LNCS 4117, pp.602–619, 2006. The full version is Cryptology ePrint Archive: Report 2006/043 at http://eprint.iacr.org/

[2] M. Bellare, R. Canetti, and H. Krawczyk, "Keying hash functions for message authentication," CRYPTO'96 Proc., LNCS 1109, pp.1–15, 1996.

[3] M. Bellare and T. Kohno, "A theoretical treatment of related-key attacks: RKA-PRPs, RKA-PRFs, and applications," EUROCRYPT 2003 Proc., LNCS 2656, pp.491–506, 2003.

[4] J. Black, P. Rogaway, and T. Shrimpton, "Black-box analysis of the block-cipher-based hash-function constructions from PGV," CRYPTO 2002 Proc., LNCS 2442, pp.320–335, 2002.

[5] O. Goldreich, Foundations of Cryptography: Basic Tools, Cambridge University Press, 2001.

[6] S. Hirose and H. Kuwakado, "A scheme to base a hash function on a block cipher," Preproc. 15th Workshop on Selected Areas in Cryptography (SAC 2008), pp.243–256, 2008.

[7] S. Hirose, J. Park, and A. Yun, "A simple variant of the Merkle-Damgård scheme with a permutation," ASIACRYPT 2007 Proc., LNCS 4833, pp.113–129, 2007.

[8] S.M. Matyas, C.H. Meyer, and J. Oseas, "Generating strong one-way functions with cryptographic algorithm," IBM Technical Disclosure Bulletin, vol.27, pp.5658–5659, 1985.

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

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

[11] B. Preneel, R. Govaerts, and J. Vandewalle, "Hash functions based on block ciphers: A synthetic approach," CRYPTO'93 Proc., LNCS 773, pp.368–378, 1994.

[12] B. Preneel and P.C. van Oorschot, "On the security of iterated message authentication codes," IEEE Trans. Inf. Theory, vol.45, no.1, pp.188–199, 1999.

[13] K. Yasuda, "Boosting Merkle-Damgård hashing for message authentication," ASIACRYPT 2007 Proc., LNCS 4833, pp.216–231, 2007.

[14] K. Yasuda, "Multilane HMAC — security beyond the birthday limit," INDOCRYPT 2007 Proc., LNCS 4859, pp.18–32, 2007.

[15] K. Yasuda, ""Sandwich" is indeed secure: How to authenticate a message with just one hashing," ACISP 2007 Proc., LNCS 4586, pp.355–369, 2007.

Appendix: Proof of Lemma 1

Let $\mathcal{B}^{\le i} = \bigcup_{d=0}^{i} \mathcal{B}^d$. For $i \in \{0, 1, \ldots, \ell\}$ ($\ell \ge 1$) and two functions $\alpha : \mathcal{B}^{\le i} \to \mathcal{B}$ and $\beta : \mathcal{B}^i \to \mathcal{B}$, a function $I_i[\alpha, \beta] : \mathcal{B}^{\le \ell} \to \mathcal{B}$ is defined as follows:

$$I_i[\alpha,\beta](M_1 M_2 \cdots M_l) = \begin{cases} \alpha(M_1 \cdots M_l) & \text{if } l \le i, \\ h^\circ_\pi(\beta(M_1 \cdots M_i), M_{i+1} \cdots M_l) & \text{if } l > i. \end{cases}$$

Let P_i be the probability

$$\Pr[A^{I_i[\alpha,\beta]} = 1 \mid \alpha \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^{\le i}, \mathcal{B}) \wedge \beta \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^i, \mathcal{B})].$$

Then,

$$\mathrm{Adv}^{\mathrm{prf}}_{h^\circ_\pi}(A) = \left|P_0 - P_\ell\right|.$$

Notice that α and β are just random elements from $\mathcal{B}$ if $i = 0$.

A q-prf-rka-adversary B with q pairs of oracles $\langle u_j, u'_j \rangle_{j=1}^{q}$ is constructed using A as a subroutine. For $i \in \{1, \ldots, \ell\}$, a q-prf-rka-adversary $B_i^{\langle u_j, u'_j \rangle_{j=1}^{q}}$ is first defined. Then, B is constructed with B_i's.

B_i first picks $\gamma \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^{\le i-1}, \mathcal{B})$. Actually, B_i implements γ via lazy sampling. Then, B_i runs A. B_i has to answer q queries of A appropriately. In order to do that, B_i maintains a counter idx, which is initially set to 0. When B_i receives the k-th query $M^{(k)} = M_1^{(k)} M_2^{(k)} \cdots M_l^{(k)}$ of A, B_i returns

$$\begin{cases} \gamma(M_1^{(k)} \cdots M_l^{(k)}) & \text{if } l < i, \\ u'_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)}) & \text{if } l = i, \\ h^\circ_\pi(u_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)}), M_{i+1}^{(k)} \cdots M_l^{(k)}) & \text{if } l > i. \end{cases}$$

In the above, $\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})$ is a unique integer in $\{1, \ldots, q\}$ which depends on the query $M_1^{(k)} \cdots M_{i-1}^{(k)}$. If there is a previous query $M^{(p)}$ ($p < k$) such that $M_1^{(p)} \cdots M_{i-1}^{(p)} = M_1^{(k)} \cdots M_{i-1}^{(k)}$, then define $\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)}) = \mathsf{idx}(M_1^{(p)} \cdots M_{i-1}^{(p)})$, and otherwise increase idx by 1 and define $\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)}) = idx$.

Now, suppose that B_i is given oracles u_j, u'_j such that $u_j = h_{K_j}$ and $u'_j = h_{\pi(K_j)}$ with $K_j \xleftarrow{\$} \mathcal{B}$ for $1 \le j \le q$. Then, when A makes the k-th query $M^{(k)} = M_1^{(k)} M_2^{(k)} \cdots M_l^{(k)}$, B_i returns

$$\begin{cases} \gamma(M_1^{(k)} \cdots M_l^{(k)}) & \text{if } l < i, \\ h_{\pi(K_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})})}(M_i^{(k)}) & \text{if } l = i, \\ h^\circ_\pi(K_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}, M_i^{(k)} M_{i+1}^{(k)} \cdots M_l^{(k)}) & \text{if } l > i. \end{cases}$$

Since $K_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}$ is a random function of $M_1^{(k)} \cdots M_{i-1}^{(k)}$, we can say that A has oracle access to $I_{i-1}[\alpha, \beta]$ with $\alpha \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^{\le i-1}, \mathcal{B})$ and $\beta \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^{i-1}, \mathcal{B})$. Therefore,

$$\Pr[B_i^{\langle h_{K_j}, h_{\pi(K_j)} \rangle_{j=1}^{q}} = 1 \mid K_1, \ldots, K_q \xleftarrow{\$} \mathcal{B}] = P_{i-1}.$$

Next, suppose that B_i has oracle access to $\rho_1, \rho'_1, \ldots, \rho_q, \rho'_q \xleftarrow{\$} \mathsf{Func}(\mathcal{B}, \mathcal{B})$. Then, B_i returns

$$\begin{cases} \gamma(M_1^{(k)} \cdots M_l^{(k)}) & \text{if } l < i, \\ \rho'_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)}) & \text{if } l = i, \\ h_\pi^\circ(\rho_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)}), M_{i+1}^{(k)} \cdots M_l^{(k)}) & \text{if } l > i. \end{cases}$$

Since $\rho_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)})$ and $\rho'_{\mathsf{idx}(M_1^{(k)} \cdots M_{i-1}^{(k)})}(M_i^{(k)})$ are independent random functions of $M_1^{(k)} \cdots M_{i-1}^{(k)} M_i^{(k)}$, we can say that A has oracle access to $I_i[\alpha, \beta]$ with $\alpha \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^{\le i}, \mathcal{B})$ and $\beta \xleftarrow{\$} \mathsf{Func}(\mathcal{B}^i, \mathcal{B})$. Therefore,

$$\Pr[B_i^{\langle \rho_j, \rho'_j \rangle_{j=1}^q} = 1 \mid \langle \rho_j, \rho'_j \rangle_{j=1}^q \xleftarrow{\$} \mathsf{Func}(\mathcal{B}, \mathcal{B})] = P_i.$$

Finally, B is defined as follows: It first chooses $i \xleftarrow{\$} \{1, \ldots, \ell\}$, then behaves identically to B_i. Then,

$$\begin{aligned} &\mathrm{Adv}_{\pi,h}^{q\text{-prf-rka}}(B) \\ &= \Big| \Pr[B^{\langle h_{K_j}, h_{\pi(K_j)} \rangle_{j=1}^q} = 1 \mid K_1, \ldots, K_q \xleftarrow{\$} \mathcal{B}] - \\ &\qquad \Pr[B^{\langle \rho_j, \rho'_j \rangle_{j=1}^q} = 1 \mid \langle \rho_j, \rho'_j \rangle_{j=1}^q \xleftarrow{\$} \mathsf{Func}(\mathcal{B}, \mathcal{B})] \Big| \\ &= \frac{1}{\ell} \left| \sum_{i=1}^{\ell} P_{i-1} - \sum_{i=1}^{\ell} P_i \right| = \frac{1}{\ell} \big| P_0 - P_\ell \big| \\ &= \frac{1}{\ell} \mathrm{Adv}_{h_\pi^\circ}^{\mathrm{prf}}(A). \end{aligned}$$

B makes at most q queries and runs in time at most $t + O(\ell q T_h)$. There may exist an algorithm with the same resources and larger advantage. Let us also call it B. Then,

$$\mathrm{Adv}_{h_\pi^\circ}^{\mathrm{prf}}(A) \le \ell \cdot \mathrm{Adv}_{\pi,h}^{q\text{-prf-rka}}(B).$$

□

Shoichi Hirose received the B.E., M.E. and D.E. degrees in information science from Kyoto University, Kyoto, Japan, in 1988, 1990 and 1995, respectively. From 1990 to 1998, he was a research associate at Faculty of Engineering, Kyoto University. From 1998 to 2005, he was a lecturer at Graduate School of Informatics, Kyoto University. From 2005 to 2009, he was an associate professor at Faculty of Engineering, University of Fukui. From 2009, he is professor at Graduate School of Engineering, University of Fukui. His current interests include cryptography and information security. He received Young Engineer Award from IEICE in 1997. He is a member of ACM, IEEE, IACR and IPSJ.

Hidenori Kuwakado received the B.E., M.E. and D.E. degrees from Kobe University in 1990, 1992, and 1999 respectively. He worked for Nippon Telegraph and Telephone Corporation from 1992 to 1996. From 1996 to 2002 he was a Research Associate in the Faculty of Engineering, Kobe University. From 2002 to 2007, he was an Associate Professor in the Faculty of Engineering, Kobe University. Since 2007, he has been an Associate Professor in Graduate School of Engineering, Kobe University. His research interests are in cryptography and information security.