# Collision-Resistant and Pseudorandom Function Based on Merkle-Damgård Hash Function

Shoichi Hirose[0000−0001−6723−722X]

University of Fukui, Fukui, Japan
hrs_shch@u-fukui.ac.jp

**Abstract.** This paper presents a keyed hash function satisfying collision resistance and the pseudorandom-function (PRF) property. It is based on the Merkle-Damgård hash function. It is shown to satisfy collision resistance under the ideal assumption that the underlying compression function is a random oracle. It is also shown to be a secure PRF if the underlying compression function is a secure PRF against related-key attacks in two keying strategies. The novel feature of the proposed keyed hash function is its efficiency. It achieves the minimum number of calls to the underlying compression function for any message input. Namely, constructed with the compression function accepting a $w$-bit message block, it processes any $l(\geq 0)$-bit massage with $\max\{1, \lceil l/w \rceil\}$ calls to the compression function. Thus, it is more efficient than the standardized keyed hash function HMAC, which also satisfies both collision resistance and the PRF property, especially for short messages. The proposed keyed hash function, as well as HMAC, can be instantiated with the SHA-256 compression function.

**Keywords:** Hash function · Collision resistance · Pseudorandom function · Related-key attack · Provable security

## 1 Introduction

*Background.* Cryptographic hash functions are an important primitive in cryptography and are used for various applications such as message digest, message authentication, and pseudorandom generation. Among them, some interesting cryptographic schemes were recently constructed with a keyed hash function satisfying both collision resistance and the pseudorandom-function property simultaneously: compactly committing authenticated encryption with associated data (ccAEAD) [10,16] and hash-based post-quantum EPID signatures [8]. We call such a keyed hash function a collision-resistant and pseudorandom hash function.

It is also well known that a collision-resistant and pseudorandom hash function $H$ directly instantiates computationally hiding and computationally binding string commitment. In the commit phase, for a message $M$, a sender chooses a key $K$ uniformly at random, computes $\sigma \leftarrow H_K(M)$, and sends $\sigma$ to a receiver. In the open phase, the sender reveals $M$ and $K$ to the receiver. This scheme is

computationally hiding since $H$ is a pseudorandom function (PRF). It is computationally binding since $H$ is collision-resistant: It is intractable to find a pair $(M, K)$ and $(M', K')$ such that $(M, K) \neq (M', K')$ and $H_K(M) = H_{K'}(M')$.

HMAC [4] is a collision-resistant and pseudorandom hash function, which is specified by NIST in FIPS PUB 198-1 [13] and by ISO in ISO/IEC 9797-2:2021 [20]. A drawback of HMAC is that it is not so efficient for short messages since it calls its underlying hash function twice.

*Our Contribution.* We present a new collision-resistant and pseudorandom hash function. It is a kind of Merkle-Damgård hash function $\mathsf{H}^F : \{0,1\}^{n/2} \times \{0,1\}^* \to \{0,1\}^n$ using a compression function $F : \{0,1\}^n \times \{0,1\}^w \to \{0,1\}^n$, which is depicted in Fig. 1. $\mathsf{H}^F$ is regarded as a keyed function with its key space $\{0,1\}^{n/2}$. It is based on MDP [18] and is very similar to Keyed-MDP (MDP keyed via IV). Thus, we call it KMDP$^+$.

$\mathsf{H}^F$ achieves the minimum number of calls to its compression function $F$ under the assumption that a message input is fed only into the message-block input ($\{0,1\}^w$) of its compression function. Namely, a message $M \in \{0,1\}^*$ is processed with $\lceil |M|/w \rceil$ calls to the compression function if $|M| > 0$, where $|M|$ is the length of $M$. If $|M| = 0$, then $M$ is processed with a single call to the compression function.

$\mathsf{H}^F$ is shown to be collision-resistant if $F$ is a random oracle, that is, $F$ is chosen uniformly at random: Any adversary needs $\Omega(2^{n/2})$ queries to $F$ to find a colliding pair of inputs of $\mathsf{H}^F$. $\mathsf{H}^F$ is shown to be a secure PRF if the compression function $F$ used as a keyed function with its key space $\{0,1\}^n$ is a secure PRF against related-key attacks [6] in two keying strategies. The proof uses the hybrid argument [15].

The construction is simple, and the techniques used for the security analyses are conservative. Nevertheless, as far as we know, KMDP$^+$ is the first keyed hash function satisfying collision resistance and the PRF property and achieving the minimum number of calls to its compression function for any message input. In addition, the SHA-256 compression function seems suitable for the instantiation of KMDP$^+$.

*Related Work.* An iterated hash function consisting of a compression function is called a Merkle-Damgård hash function [9,24]. Pseudorandom functions were first introduced by Goldreich, Goldwasser, and Micali [14].

Instantiated with a Merkle-Damgård hash function, HMAC is shown to be a secure PRF if its underlying compression function is a secure PRF in two keying strategies, keyed via the chaining value and keyed via the message [2]. AMAC [3] is also a hash-based PRF that calls its underlying hash function once. From its construction, it is easy to see that AMAC also satisfies collision resistance. However, due to its output transform such as truncation, SHA-256 does not seem suitable for the instantiation of AMAC with a sufficient level of collision resistance.

Bellare and Ristenpart [7] introduced the notion of multi-property preservation and presented the domain extension EMD shown to produce a hash

function satisfying collision resistance, PRF property, and pseudorandom-oracle property (indifferentiability) from a compression function satisfying the corresponding property. Their construction assumes padding with Merkle-Damgård-strengthening and is not so efficient as our construction.

The domain extension MDP was presented by Hirose, Park, and Yun [18]. It is implicit in [18] that Keyed-MDP is also a collision-resistant and pseudorandom hash function. In terms of efficiency, though Keyed-MDP is very competitive with KMDP$^+$, the latter is still better than the former that assumes padding with Merkle-Damgård-strengthening.

A PRF based on a Merkle-Damgård hash function and achieving the minimum number of calls to its compression function was presented by Hirose and Yabumoto [19]. A Merkle-Damgård hash function achieving the minimum number of calls to its compression function was proposed and shown to satisfy indifferentiability [17]. We unify these two constructions and obtain KMDP$^+$.

The Merkle-Damgård hash function keyed via the initial value with prefix-free padding for message input is shown to be a secure PRF if the underlying compression function is a secure PRF [5]. Our proof on PRF is based on this proof.

Quite recently, Andreeva et al. [1] and Dodis et al. [11] presented similar domain extension schemes which outperform the Merkle-Damgård domain extension in terms of the number of calls to the underlying compression function. However, their domain extension schemes are not effective in processing a short message consisting of a few message blocks.

CMAC [25] is a CBC-MAC function designed by Iwata and Kurosawa [21,22], which achieves the minimum number of calls to its underlying block cipher. It is not designed to satisfy collision resistance.

A distinguishing attack on the SHA-256/512 compression functions keyed via the chaining value was presented by Kuwakado and Hirose [23].

*Organization.* Section 2 gives notations and definitions. Section 3 describes the proposed keyed hash function. Section 4 shows the collision resistance of the proposed keyed hash function under the assumption that the underlying compression function is a random oracle. Section 5 shows that the proposed keyed hash function is a secure PRF if the underlying compression function is a secure PRF against related-key attacks in two keying strategies. Section 6 discusses the instantiation with the SHA-256 compression function and its efficiency. Section 7 gives a concluding remark.

## 2 Preliminaries

### 2.1 Notations

Let $\Sigma := \{0,1\}$. Let $(\Sigma^n)^* := \bigcup_{i=0}^{\infty} \Sigma^{ni}$ and $(\Sigma^n)^+ := \bigcup_{i=1}^{\infty} \Sigma^{ni}$. Let $\varepsilon \in \Sigma^0$ be an empty sequence.

For a binary sequence $u \in \Sigma^*$, let $|u|$ be the length of $u$. For binary sequences $z_i, z_{i+1}, \ldots, z_{i+j} \in \Sigma^*$, let $z_i \| z_{i+1} \| \cdots \| z_{i+j}$ be their concatenation, which is also denoted by $z_{[i,i+j]}$ for simplicity.

Let $s \twoheadleftarrow \mathcal{S}$ denote that $s$ is assigned an element chosen uniformly at random from set $\mathcal{S}$.

For integers $a$, $b$, and $d$, $a \equiv b \pmod{d}$ is denoted as $a \equiv_d b$.

A random function $\rho : \mathcal{D} \to \mathcal{R}$ is called a random oracle if, for any $x \in \mathcal{D}$, $\rho(x)$ is assigned an element chosen uniformly at random from $\mathcal{R}$.

### 2.2 Collision Resistance

Let $H^F : \mathcal{X} \to \mathcal{Y}$ be a hash function using a compression function $F$. The collision resistance of a hash function is often discussed under the assumption that its compression function is a random oracle.

Let $\mathbf{A}$ be an adversary trying to find a collision for $H^F$, that is, a pair of distinct inputs mapped to the same output. The col-advantage of $\mathbf{A}$ against $H^F$ is defined as

$$\mathrm{Adv}^{\mathrm{col}}_{H^F}(\mathbf{A}) := \Pr[(X, X') \leftarrow \mathbf{A}^F : H^F(X) = H^F(X') \wedge X \neq X'] \ .$$

It is assumed that $\mathbf{A}$ makes all the queries necessary to compute $H^F(X)$ and $H^F(X')$. Let $\mathrm{Adv}^{\mathrm{col}}_{H^F}(q)$ be the maximum col-advantage over all adversaries making at most $q$ queries.

### 2.3 Pseudorandom Function

Let $f : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ be a keyed function with its key space $\mathcal{K}$ and $f_K(\cdot) := f(K, \cdot)$. Let $\mathbf{A}$ be an adversary against $f$. The goal of $\mathbf{A}$ is to distinguish between $f_K$ and a random oracle $\rho : \mathcal{X} \to \mathcal{Y}$, where $K \twoheadleftarrow \mathcal{K}$. $\mathbf{A}$ is given either $f_K$ or $\rho$ as an oracle and makes adaptive queries in $\mathcal{X}$. $\mathbf{A}$ outputs 0 or 1. The prf-advantage of $\mathbf{A}$ against $f$ is defined as

$$\mathrm{Adv}^{\mathrm{prf}}_f(\mathbf{A}) := \left| \Pr\left[\mathbf{A}^{f_K} = 1\right] - \Pr\left[\mathbf{A}^{\rho} = 1\right] \right| \ ,$$

where $\mathbf{A}$ is regarded as a random variable that takes values in $\{0, 1\}$. $f$ is called a secure pseudorandom function (PRF) if no efficient adversary $\mathbf{A}$ has any significant prf-advantage against $f$.

The prf-advantage can be extended to adversaries with multiple oracles. The prf-advantage of adversary $\mathbf{A}$ with access to $p$ oracles is defined as

$$\mathrm{Adv}^{p\text{-}\mathrm{prf}}_f(\mathbf{A}) := \left| \Pr[\mathbf{A}^{f_{K_1}, f_{K_2}, \ldots, f_{K_p}} = 1] - \Pr[\mathbf{A}^{\rho_1, \rho_2, \ldots, \rho_p} = 1] \right| \ ,$$

where $(K_1, \ldots, K_p) \twoheadleftarrow \mathcal{K}^p$ and $\rho_1, \ldots, \rho_p$ are independent random oracles.

### 2.4 PRF under Related-Key Attack

Let $\mathbf{A}$ be an adversary against $f : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$. Let $\Phi$ be a set of functions from $\mathcal{K}$ to $\mathcal{K}$. $\mathbf{A}$ makes a $\Phi$-related-key attack ($\Phi$-RKA) [6]: $\mathbf{A}$ is given $g[K] : \Phi \times \mathcal{X} \to \mathcal{Y}$ such that $g[K](\varphi, X) := g(\varphi(K), X)$ as an oracle, where $g$ is either $f$ or a random oracle $\rho : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ and $K \twoheadleftarrow \mathcal{K}$. $\mathbf{A}$ makes adaptive queries to the oracle and outputs 0 or 1. The prf-rka-advantage of $\mathbf{A}$ making a $\Phi$-RKA on $f$ is given by

$$\mathrm{Adv}^{\text{prf-rka}}_{f,\Phi}(\mathbf{A}) := \left|\Pr[\mathbf{A}^{f[K]} = 1] - \Pr[\mathbf{A}^{\rho[K]} = 1]\right| \ .$$

$f$ is called a secure PRF under $\Phi$-RKAs if no efficient adversary $\mathbf{A}$ has any significant prf-rka-advantage.

The prf-rka-advantage of $\mathbf{A}$ with access to $p$ oracles is defined as

$$\mathrm{Adv}^{p\text{-prf-rka}}_{f,\Phi}(\mathbf{A}) := \left|\Pr[\mathbf{A}^{f[K_1],\ldots,f[K_p]} = 1] - \Pr[\mathbf{A}^{\rho_1[K_1],\ldots,\rho_p[K_p]} = 1]\right| \ ,$$

where $(K_1, \ldots, K_p) \twoheadleftarrow \mathcal{K}^p$ and $\rho_1, \ldots, \rho_p$ are independent random oracles.

**Lemma 1 ([19]).** *For any adversary $\mathbf{A}$ against $f$ taking at most $t$ time and making at most $q$ queries in total, there exists an adversary $\mathbf{A}'$ such that*

$$\mathrm{Adv}^{p\text{-prf-rka}}_{f,\Phi}(\mathbf{A}) \leq p \cdot \mathrm{Adv}^{\text{prf-rka}}_{f,\Phi}(\mathbf{A}') \ .$$

*$\mathbf{A}'$ takes at most about $t + q \cdot \tau_f$ time and makes at most $q$ queries, where $\tau_f$ is time required to compute $f$.*

## 3 Proposed Hash Function

Let $n > 1$ be an even integer and $w > 1$ be an integer. The proposed hash function KMDP$^+$ uses a compression function $F : \Sigma^n \times \Sigma^w \to \Sigma^n$. It also uses the following padding function:

$$\mathtt{pad}(M) := \begin{cases} M & \text{if } |M| > 0 \text{ and } |M| \equiv_w 0 \\ M \| 10^d & \text{otherwise} \ , \end{cases}$$

where $d$ is the smallest non-negative integer satisfying $|\mathtt{pad}(M)| \equiv_w 0$. Notice that $\mathtt{pad}(\varepsilon) = 10^{w-1}$.

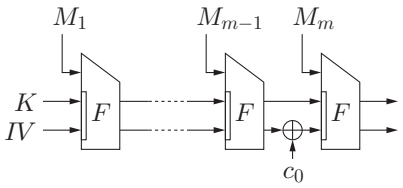
KMDP$^+$ is the hash function $\mathsf{H}^F : \Sigma^{n/2} \times \Sigma^* \to \Sigma^n$, which is described in Algorithm 1. It is also depicted in Figures 1 and 2. To specify $\mathsf{H}^F$, three fixed constants $IV, c_0, c_1 \in \Sigma^{n/2}$ are used. $c_0$ and $c_1$ are assumed to satisfy the following conditions: $c_0 \neq 0^{n/2}$; $c_1 \neq 0^{n/2}$; $c_0 \oplus c_1 \neq 0^{n/2}$.

For the PRF property, $\mathsf{H}^F$ is regarded as a keyed function with its key space $\Sigma^{n/2}$.

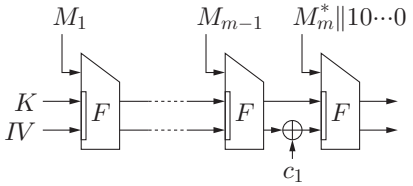
For $\mathsf{H}^F$, the number of calls to its compression function $F$ required to process an input message $M$ is 1 if $M$ is the empty sequence and $\lceil |M|/w \rceil$ otherwise.

**Algorithm 1:** The proposed hash function $\mathsf{H}^F : \Sigma^{n/2} \times \Sigma^* \to \Sigma^n$

**input :** $(K, M)$
**output:** $\mathsf{H}^F(K, M)$
$M_1 \| M_2 \| \cdots \| M_m \leftarrow \mathtt{pad}(M)$; `/*` $|M_i| = w$ `for` $1 \le i \le m$ `*/`
$V_0 \leftarrow K \| IV$;
**for** $i = 1$ **to** $m - 1$ **do** $V_i \leftarrow F(V_{i-1}, M_i)$;
**if** $|M| > 0 \wedge |M| \equiv_w 0$ **then** $V_m \leftarrow F(V_{m-1} \oplus (0^{n/2} \| c_0), M_m)$;
**else** $V_m \leftarrow F(V_{m-1} \oplus (0^{n/2} \| c_1), M_m)$;
**return** $V_m$;




(a) If $|M| > 0$ and $|M| \equiv_w 0$




(b) If $|M| = 0$ or $|M| \not\equiv_w 0$. The last block $M_m^*$ is incomplete: $1 \le |M_m^*| < w$.

Fig. 1: The proposed hash function. A message input $M$ is divided into blocks of length $w$.

## 4 Collision Resistance

The collision resistance of $\mathsf{H}^F$ is discussed under the assumption that $F$ is a random oracle. A pair of inputs $(K, M)$ and $(K', M')$ for $\mathsf{H}^F$ are colliding if $(K, M) \neq (K', M')$ and $\mathsf{H}^F(K, M) = \mathsf{H}^F(K', M')$. The following theorem implies that any adversary needs $\Omega(2^{n/2})$ queries to find a colliding pair of inputs for $\mathsf{H}^F$.

**Theorem 1.** *For collision resistance of* $\mathsf{H}^F$*,*

$$\mathrm{Adv}^{\mathrm{col}}_{\mathsf{H}^F}(q) \le q/2^{n/2-1} + q(q-1)/2^n \ .$$

*Proof.* Suppose that a colliding pair, $M$ and $M'$, are found for $\mathsf{H}^F$. Namely, $\mathsf{H}^F(M) = \mathsf{H}^F(M')$ and $M \neq M'$. It is assumed that $|M| \le |M'|$ without loss of generality. Let $\mathtt{pad}(M) = M_1 \| M_2 \| \cdots \| M_m$ and $\mathtt{pad}(M') = M'_1 \| M'_2 \| \cdots \| M'_{m'}$.

(i) Suppose that $m = m' = 1$.


(a) If $|M| = w$ (b) If $0 \leq |M| < w$

Fig. 2: The proposed hash function for at most a single-block message

If $|M| < w$ and $|M'| < w$, or $|M| = |M'| = w$, then a colliding pair are found for $F$ since $\mathtt{pad}(M) \neq \mathtt{pad}(M')$.

If $|M| < w$ and $|M'| = w$, then a colliding pair are also found for $F$ since $c_0 \neq c_1$.

(ii) Suppose that $m = 1$ and $m' \geq 2$.

If $|M| < w$ and $|M'| \equiv_w 0$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV \oplus c_0 \oplus c_1$.

If $|M| < w$ and $|M'| \not\equiv_w 0$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV$.

If $|M| = w$ and $|M'| \equiv_w 0$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV$.

If $|M| = w$ and $|M'| \not\equiv_w 0$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV \oplus c_0 \oplus c_1$.

(iii) Suppose that $m \geq 2$ and $m' \geq 2$.

If $|M| \equiv_w 0$ and $|M'| \not\equiv_w 0$, or $|M| \not\equiv_w 0$ and $|M'| \equiv_w 0$, then a colliding pair are found for $F$ or $F$ wrt $c_0 \oplus c_1$. A pair of inputs $(V_{i-1}, M_i)$ and $(V'_{i-1}, M'_i)$ are called colliding wrt $c_0 \oplus c_1$ if $F(V_{i-1}, M_i) = F(V'_{i-1}, M'_i) \oplus c_0 \oplus c_1$.

Suppose that $|M| \equiv_w 0$ and $|M'| \equiv_w 0$. If $m = m'$, then a colliding pair are found for $F$. If $m < m'$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV$.

Suppose that $|M| \not\equiv_w 0$ and $|M'| \not\equiv_w 0$. If $m = m'$, then a colliding pair are found for $F$. If $m < m'$, then a colliding pair are found for $F$ or an input for $F$ such that the least significant $n/2$ bits of the corresponding output equals $IV$.

Thus, a colliding pair for $\mathsf{H}^F$ implies

1. a colliding pair,
2. a colliding pair wrt $c_0 \oplus c_1$, or
3. an input mapped to an output whose least significant $n/2$ bits equals $IV$ or $IV \oplus c_0 \oplus c_1$

for $F$. The probability that the $j$-th query induces 1 or 2 above for $F$ is at most $2(j-1)/2^n$. The probability that the $j$-th query induces 3 above for $F$ is at most

$1/2^{n/2-1}$. Since an adversary makes at most $q$ queries,

$$\sum_{i=1}^{q}(2(i-1)/2^n + 1/2^{n/2-1}) \le q/2^{n/2-1} + q(q-1)/2^n \ .$$

□

## 5 Pseudorandom-Function Property

We treat the compression function $F : \Sigma^n \times \Sigma^w \to \Sigma^n$ as a keyed function with its key space $\Sigma^n$ in two keying strategies. In one strategy, a secret key is simply chosen uniformly at random from $\Sigma^n$. In the other strategy, it is chosen uniformly at random from $\Sigma^{n/2} \times \{IV\} (\subset \Sigma^n)$. To make the distinction clear, we denote the keyed compression function $F$ in the latter keying strategy by $\tilde{F}$.

For both $F$ and $\tilde{F}$, we consider $\{\mathsf{id}, \mathsf{x}_{c_0}, \mathsf{x}_{c_1}\}$-related-key attacks, where $\mathsf{id}$ is the identity permutation over $\Sigma^n$, and, for $b \in \{0,1\}$, $\mathsf{x}_{c_b}$ is a permutation over $\Sigma^n$ such that $x \mapsto x \oplus (0^{n/2} \| c_b)$.

The following theorem implies that $\mathsf{H}^F$ is a secure PRF if both $F$ and $\tilde{F}$ are secure PRFs under $\{\mathsf{id}, \mathsf{x}_{c_0}, \mathsf{x}_{c_1}\}$-related-key attacks.

**Theorem 2.** *For any adversary* $\mathbf{A}$ *taking at most* $t$ *time and making at most* $q$ *queries each of which has at most* $\ell$ *blocks after padding, there exist adversaries* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *such that*

$$\mathrm{Adv}^{\mathrm{prf}}_{\mathsf{H}^F}(\mathbf{A}) \le \mathrm{Adv}^{\text{prf-rka}}_{\tilde{F},\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{A}_1) + (\ell-1)q\,\mathrm{Adv}^{\text{prf-rka}}_{F,\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{A}_2) \ .$$

*Both* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *take at most about* $t + O(\ell q \tau_F)$ *time and make at most* $q$ *queries, where* $\tau_F$ *is time required to compute* $F$*.*

*Proof.* Let $\mathsf{I}^c : \Sigma^n \times (\Sigma^w)^+ \to \Sigma^n$ be a keyed function specified in Algorithm 2. For an integer $k \ge 0$ and functions $\mu : (\Sigma^w)^* \to \Sigma^n$ and $\bar{\mu} : \Sigma^* \to \Sigma^n$, let $\mathsf{Hy}[k]^{\mu,\bar{\mu}} : \Sigma^* \to \Sigma^n$ be a function specified as follows: For $M \in \Sigma^*$ such that $\mathsf{pad}(M) = M_1 \| \cdots \| M_m$,

$$\mathsf{Hy}[k]^{\mu,\bar{\mu}}(M) := \begin{cases} \bar{\mu}(M) & \text{if } m \le k, \\ \mathsf{I}^{c_0}(\mu(M_{[1,k]}), M_{[k+1,m]}) & \text{if } m > k \wedge (|M| > 0 \wedge |M| \equiv_w 0), \\ \mathsf{I}^{c_1}(\mu(M_{[1,k]}), M_{[k+1,m]}) & \text{if } m > k \wedge (|M| = 0 \vee |M| \not\equiv_w 0). \end{cases}$$

Notice that $M_{[1,0]} = \varepsilon$.

Suppose that $\bar{\mu}$ is a random oracle and $\mu$ is a random oracle with a restriction that $\mu(\varepsilon)$ is chosen uniformly at random from $\Sigma^{n/2} \times \{IV\}$. Then,

$$\mathsf{Hy}[0]^{\mu,\bar{\mu}}(M) := \begin{cases} \mathsf{I}^{c_0}(\mu(\varepsilon), M_{[1,m]}) & \text{if } |M| > 0 \wedge |M| \equiv_w 0, \\ \mathsf{I}^{c_1}(\mu(\varepsilon), M_{[1,m]}) & \text{if } |M| = 0 \vee |M| \not\equiv_w 0, \end{cases}$$

which is equivalent to $\mathsf{H}^F$. $\mathsf{Hy}[\ell]^{\mu,\bar{\mu}}$ works as a random oracle for any $M \in \Sigma^*$ such that $\mathtt{pad}(M)$ consists of at most $\ell$ blocks. Since every query made by $\mathbf{A}$ is assumed to consist of at most $\ell$ blocks after padding,

$$\begin{aligned}\mathrm{Adv}^{\mathrm{prf}}_{\mathsf{H}^F}(\mathbf{A}) &= \left|\Pr[\mathbf{A}^{\mathsf{Hy}[0]^{\mu,\bar{\mu}}} = 1] - \Pr[\mathbf{A}^{\mathsf{Hy}[\ell]^{\mu,\bar{\mu}}} = 1]\right| \\ &\le \left|\Pr[\mathbf{A}^{\mathsf{Hy}[0]^{\mu,\bar{\mu}}} = 1] - \Pr[\mathbf{A}^{\mathsf{Hy}[1]^{\mu,\bar{\mu}}} = 1]\right| \\ &\quad + \sum_{k=1}^{\ell-1}\left|\Pr[\mathbf{A}^{\mathsf{Hy}[k]^{\mu,\bar{\mu}}} = 1] - \Pr[\mathbf{A}^{\mathsf{Hy}[k+1]^{\mu,\bar{\mu}}} = 1]\right| \ . \qquad (1)\end{aligned}$$

For the first term of the upper bound in Inequality (1), let $\mathbf{D}_0$ be a prf-rka-adversary against $\tilde{F}$. $\mathbf{D}_0$ runs $\mathbf{A}$ and simulates the oracle of $\mathbf{A}$ using its oracle. $\mathbf{D}_0$ outputs the output of $\mathbf{A}$. Let $\tilde{G}[\tilde{K}]$ be the oracle of $\mathbf{D}_0$, which are either $\tilde{F}[\tilde{K}]$ or $\tilde{\rho}[\tilde{K}]$, where $\tilde{K} \twoheadleftarrow \Sigma^{n/2} \times \{IV\}$ and $\tilde{\rho} : \Sigma^n \times \Sigma^w \to \Sigma^n$ is a random oracle. For the $j$-th query $M$ made by $\mathbf{A}$ such that $\mathtt{pad}(M) = M_1 \| \cdots \| M_m$, $\mathbf{D}_0$ acts as follows:

- If $m = 1$, then $\mathbf{D}_0$ returns to $\mathbf{A}$

$$\begin{cases} \tilde{G}_{\tilde{K} \oplus (0^{n/2} \| c_0)}(M_1) & \text{if } |M| > 0 \wedge |M| \equiv_w 0, \\ \tilde{G}_{\tilde{K} \oplus (0^{n/2} \| c_1)}(M_1) & \text{if } |M| = 0 \vee |M| \not\equiv_w 0. \end{cases}$$

  $\mathbf{D}_0$ gets $\tilde{G}_{\tilde{K} \oplus (0^{n/2} \| c_b)}(M_1)$ by asking $(\mathsf{x}_{c_b}, M_1)$ to its oracle for $b \in \Sigma$.
- If $m \ge 2$, then $\mathbf{D}_0$ returns to $\mathbf{A}$

$$\begin{cases} \mathsf{I}^{c_0}(\tilde{G}_{\tilde{K}}(M_1), M_{[2,m]}) & \text{if } |M| > 0 \wedge |M| \equiv_w 0, \\ \mathsf{I}^{c_1}(\tilde{G}_{\tilde{K}}(M_1), M_{[2,m]}) & \text{if } |M| = 0 \vee |M| \not\equiv_w 0. \end{cases}$$

  $\mathbf{D}_0$ gets $\tilde{G}_{\tilde{K}}(M_1)$ by asking $(\mathsf{id}, M_1)$ to its oracle.

$\mathbf{D}_0$ implements $\mathsf{Hy}[0]^{\mu,\bar{\mu}}$ as the oracle of $\mathbf{A}$ if its oracle is $\tilde{F}[\tilde{K}]$. It implements $\mathsf{Hy}[1]^{\mu,\bar{\mu}}$ if its oracle is $\tilde{\rho}[\tilde{K}]$ since $\tilde{\rho}_{\tilde{K}}$, $\tilde{\rho}_{\tilde{K} \oplus (0^{n/2} \| c_0)}$, and $\tilde{\rho}_{\tilde{K} \oplus (0^{n/2} \| c_1)}$ are independent. Thus,

$$\begin{aligned}\left|\Pr[\mathbf{A}^{\mathsf{Hy}[0]^{\mu,\bar{\mu}}} = 1] - \Pr[\mathbf{A}^{\mathsf{Hy}[1]^{\mu,\bar{\mu}}} = 1]\right| &= \left|\Pr[\mathbf{D}_0^{\tilde{F}[\tilde{K}]} = 1] - \Pr[\mathbf{D}_0^{\tilde{\rho}[\tilde{K}]} = 1]\right| \\ &= \mathrm{Adv}^{\text{prf-rka}}_{\tilde{F}, \{\mathsf{id}, \mathsf{x}_{c_0}, \mathsf{x}_{c_1}\}}(\mathbf{D}_0) \ . \qquad (2)\end{aligned}$$

$\mathbf{D}_0$ takes at most about $t + O(\ell q \tau_F)$ time and makes at most $q$ queries.

For the second term of the upper bound in Inequality (1), let $\mathbf{D}_k$ be a prf-rka-adversary against $F$ for $k \in [1, \ell - 1]$. $\mathbf{D}_k$ runs $\mathbf{A}$ and simulates the oracle of $\mathbf{A}$ using its oracle. $\mathbf{D}_k$ outputs the output of $\mathbf{A}$. Let $G_1[K_1], \ldots, G_q[K_q]$ be the oracle of $\mathbf{D}_k$, which are either $F[K_1], \ldots, F[K_q]$ or $\rho_1[K_1], \ldots, \rho_q[K_q]$, where $K_i \twoheadleftarrow \Sigma^n$ and $\rho_i : \Sigma^n \times \Sigma^w \to \Sigma^n$ is a random oracle for $1 \le i \le q$. Notice that $\mathbf{A}$ makes at most $q$ queries. For the $j$-th query $M$ made by $\mathbf{A}$, let $\mathtt{pad}(M) = M_1 \| \cdots \| M_m$. Suppose that $m \le k$. Then, $\mathbf{D}_k$ simulates $\bar{\mu}$ and returns $\bar{\mu}(M)$ to $\mathbf{A}$. Suppose that $m > k$. Let $\mathcal{J}$ be a set of integers such that

$$\mathcal{J} = \{j' \mid \text{The } j'(< j)\text{-th query } M' \text{ of } \mathbf{A} \text{ satisfies } m' > k \text{ and } M'_{[1,k]} = M_{[1,k]}\} \ ,$$

where $\mathsf{pad}(M') = M'_1 \| \cdots \| M'_{m'}$. Let $j^* \leftarrow j$ if $\mathcal{J} = \emptyset$ and $j^* \leftarrow \min \mathcal{J}$ otherwise. For the $j$-th query $M$ of $\mathbf{A}$, $\mathbf{D}_k$ acts as follows:

– If $m = k+1$, then $\mathbf{D}_k$ returns to $\mathbf{A}$

$$\begin{cases} G_{j^*}(K_{j^*} \oplus (0^{n/2} \| c_0), M_{k+1}) & \text{if } |M| > 0 \wedge |M| \equiv_w 0, \\ G_{j^*}(K_{j^*} \oplus (0^{n/2} \| c_1), M_{k+1}) & \text{if } |M| = 0 \vee |M| \not\equiv_w 0. \end{cases}$$

$\mathbf{D}_k$ gets $G_{j^*}(K_{j^*} \oplus (0^{n/2} \| c_b), M_{k+1})$ by asking $(\mathsf{x}_{c_b}, M_{k+1})$ to $G_{j^*}[K_{j^*}]$ for $b \in \Sigma$.

– If $m \geq k+2$, then

$$\begin{cases} \mathsf{I}^{c_0}(G_{j^*}(K_{j^*}, M_{k+1}), M_{[k+2,m]}) & \text{if } |M| > 0 \wedge |M| \equiv_w 0, \\ \mathsf{I}^{c_1}(G_{j^*}(K_{j^*}, M_{k+1}), M_{[k+2,m]}) & \text{if } |M| = 0 \vee |M| \not\equiv_w 0. \end{cases}$$

$\mathbf{D}_k$ gets $G_{j^*}(K_{j^*}, M_{k+1})$ by asking $(\mathsf{id}, M_{k+1})$ to $G_{j^*}[K_{j^*}]$.

In the process above, for the $j$-th query $M$, if $M_{[1,k]}$ is new, that is, $\mathcal{J} = \emptyset$, then $\mathbf{D}_k$ uses the new oracle $G_j[K_j]$ to compute the answer to the query. $\mathbf{D}_k$ implements $\mathsf{Hy}[k]^{\mu,\bar{\mu}}$ as the oracle of $\mathbf{A}$ if its oracles are $F[K_1], \ldots, F[K_q]$ since new $K_j$, which is chosen uniformly at random, is assigned to new $M_{[1,k]}$. $\mathbf{D}_k$ implements $\mathsf{Hy}[k+1]^{\mu,\bar{\mu}}$ if its oracles are $\rho_1[K_1], \ldots, \rho_q[K_q]$ since new $\rho_j[K_j]$ is assigned to new $M_{[1,k]}$ and $\rho_j(K_j, \cdot), \rho_j(K_j \oplus (0^{n/2} \| c_0), \cdot), \rho_j(K_j \oplus (0^{n/2} \| c_1), \cdot)$ are independent. Thus,

$$\left| \Pr[\mathbf{A}^{\mathsf{Hy}[k]^{\mu,\bar{\mu}}} = 1] - \Pr[\mathbf{A}^{\mathsf{Hy}[k+1]^{\mu,\bar{\mu}}} = 1] \right| = \mathrm{Adv}^{q\text{-prf-rka}}_{F,\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{D}_k) \ . \tag{3}$$

$\mathbf{D}_k$ takes at most about $t + O(\ell q \tau_F)$ time and makes at most $q$ queries.

From Inequality (1), Equalities (2) and (3), and Lemma 1, there exist adversaries $\mathbf{A}_1$ and $\mathbf{A}_2$ such that

$$\mathrm{Adv}^{\mathrm{prf}}_{\mathsf{H}^F}(\mathbf{A}) \leq \mathrm{Adv}^{\text{prf-rka}}_{\tilde{F},\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{A}_1) + (\ell - 1)q \cdot \mathrm{Adv}^{\text{prf-rka}}_{F,\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{A}_2) \ .$$

Both $\mathbf{A}_1$ and $\mathbf{A}_2$ take at most about $t + O(\ell q \tau_F)$ time and make at most $q$ queries. □

**Algorithm 2:** $\mathsf{I}^c : \Sigma^n \times (\Sigma^w)^+ \to \Sigma^n$

**input :** $(U, X_1 \| X_2 \| \cdots \| X_x)$
**output:** $\mathsf{I}^c(U, X_1 \| X_2 \| \cdots \| X_x)$
$V_0 \leftarrow U$;
**for** $i = 1$ **to** $x - 1$ **do** $V_i \leftarrow F(V_{i-1}, X_i)$; /* $|X_i| = w$ for $1 \leq i \leq x$ */
$V_x \leftarrow F(V_{x-1} \oplus (0^{n/2} \| c), X_x)$;
**return** $V_x$;

*Remark 1.* Even if $\tilde{F}$ is a secure PRF, $F$ is not necessarily a secure PRF. For example, suppose that $F(K_0\|K_1, X) = 0^n$ for any $K_0$ and $X$ if $K_1 \neq IV$, where $|K_0| = |K_1| = n/2$. Then, $F$ cannot be a secure PRF, while $\tilde{F}$ can be a secure PRF.

Even if $F$ is a secure PRF, $\tilde{F}$ is not necessarily a secure PRF. For example, suppose that $F(K_0\|K_1, X) = 0^n$ for any $K_0$ and $X$ if $K_1 = IV$. Then, $\tilde{F}$ cannot be a secure PRF, while $F$ can be a secure PRF.

*Remark 2.* The actual key length of $\tilde{F}$ is $(n/2)$-bits. $\tilde{F}$ may be viewed as a tweakable keyed function with its key space $\Sigma^{n/2}$ and its tweak space $\Sigma^{n/2}$. A proof applying a hybrid argument under the sole assumption that $\tilde{F}$ is a secure tweakable PRF would give an upper bound containing $\ell q \cdot \mathrm{Adv}^{\text{prf-rka}}_{\tilde{F},\{\mathsf{id},\mathsf{x}_{c_0},\mathsf{x}_{c_1}\}}(\mathbf{A}_1)$. It guarantees only $(n/4)$-bit security due to the simple guessing-key attack on $\tilde{F}$.

# 6 Discussion

## 6.1 Instantiation

KMDP$^+$ can be instantiated with the SHA-256 compression function together with, for example, the following constants:

$$\begin{aligned} IV &= \texttt{510e527f 9b05688c 1f83d9ab 5be0cd19}\ , \\ c_0 &= \texttt{36363636 36363636 36363636 36363636}\ , \\ c_1 &= \texttt{5c5c5c5c 5c5c5c5c 5c5c5c5c 5c5c5c5c}\ . \end{aligned}$$

$IV$ is the second half of the initial hash value of the SHA-256 hash function [12]. $c_0$ and $c_1$ are taken from the constants `ipad` and `opad` of HMAC (Fig. 3) [13], respectively. For such $c_0$ and $c_1$,

$$c_0 \oplus c_1 = \texttt{6a6a6a6a 6a6a6a6a 6a6a6a6a 6a6a6a6a}\ .$$

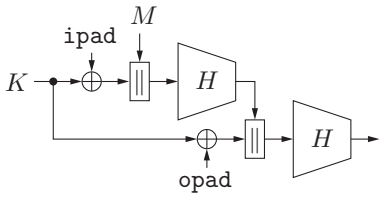



Fig. 3: HMAC using a hash function $H$. Both `ipad` and `opad` are fixed constants.

### 6.2 Efficiency

The SHA-256 compression function accepts a 512-bit message block. Thus, for KMDP$^+$ instantiated with the SHA-256 compression function, the number of calls to the compression function required to process an input message $M$ is 1 if $M$ is the empty sequence and $\lceil |M|/512 \rceil$ otherwise.

For HMAC [13], the amount of computation required to process an input message depends on the amount of computation required by its hash function. Here, we assume HMAC using SHA-256 [12]. The padding scheme of SHA-256 appends $10^*\|\|M|_{64}$ to an input message $M$, where $|M|_{64}$ is the 64-bit binary representation of $|M|$. Thus, the number of calls to its compression function required to process $M$ is $\lceil (M - 447)/512 \rceil + 4$.

$\mathsf{I}^c(K, \mathsf{mdspad}(M))$ is an implementation of Keyed-MDP [18], where $\mathsf{mdspad}$ is padding with Merkle-Damgård strengthening. If $\mathsf{I}^c(K, \mathsf{mdspad}(M))$ is instantiated with the SHA-256 compression function and $\mathsf{mdspad}$ is the padding scheme of SHA-256, then the number of calls to the compression function is $\lceil (M - 447)/512 \rceil + 1$. Thus, Keyed-MDP is very competitive with KMDP$^+$:

$$\lceil (M - 447)/512 \rceil + 1 = \begin{cases} \max\{1, \lceil |M|/512 \rceil\} & \text{if } 0 \leq |M| \bmod 512 \leq 447, \\ \lceil |M|/512 \rceil + 1 & \text{otherwise.} \end{cases}$$

## 7 Concluding Remark

We have proposed a collision-resistant and pseudorandom hash function based on Merkle-Damgård hashing. It achieves the minimum number of calls to its underlying compression function for any input. It can be instantiated with the SHA-256 compression function. Future work is to explore the PRF property of the SHA-256 compression function keyed via the chaining value against related-key attacks assumed for KMDP$^+$.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP21K11885.

## References

1. Andreeva, E., Bhattacharyya, R., Roy, A.: Compactness of hashing modes and efficiency beyond Merkle tree. In: Canteaut, A., Standaert, F. (eds.) EUROCRYPT 2021, Proceedings, Part II. Lecture Notes in Computer Science, vol. 12697, pp. 92–123. Springer (2021). https://doi.org/10.1007/978-3-030-77886-6_4
2. Bellare, M.: New proofs for NMAC and HMAC: security without collision-resistance. In: Dwork, C. (ed.) CRYPTO 2006, Proceedings. Lecture Notes in Computer Science, vol. 4117, pp. 602–619. Springer (2006). https://doi.org/10.1007/11818175_36

3. Bellare, M., Bernstein, D.J., Tessaro, S.: Hash-function based PRFs: AMAC and its multi-user security. In: Fischlin, M., Coron, J. (eds.) EUROCRYPT 2016, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9665, pp. 566–595. Springer (2016). https://doi.org/10.1007/978-3-662-49890-3_22
4. Bellare, M., Canetti, R., Krawczyk, H.: Keying hash functions for message authentication. In: Koblitz, N. (ed.) CRYPTO '96, Proceedings. Lecture Notes in Computer Science, vol. 1109, pp. 1–15. Springer (1996). https://doi.org/10.1007/3-540-68697-5_1
5. Bellare, M., Canetti, R., Krawczyk, H.: Pseudorandom functions revisited: The cascade construction and its concrete security. In: Proceedings of the 37th IEEE Symposium on Foundations of Computer Science. pp. 514–523 (1996)
6. Bellare, M., Kohno, T.: A theoretical treatment of related-key attacks: RKA-PRPs, RKA-PRFs, and applications. In: Biham, E. (ed.) EUROCRYPT 2003, Proceedings. Lecture Notes in Computer Science, vol. 2656, pp. 491–506. Springer (2003). https://doi.org/10.1007/3-540-39200-9_31
7. Bellare, M., Ristenpart, T.: Multi-property-preserving hash domain extension and the EMD transform. In: Lai, X., Chen, K. (eds.) ASIACRYPT 2006, Proceedings. Lecture Notes in Computer Science, vol. 4284, pp. 299–314. Springer (2006). https://doi.org/10.1007/11935230_20
8. Boneh, D., Eskandarian, S., Fisch, B.: Post-quantum EPID signatures from symmetric primitives. In: Matsui, M. (ed.) CT-RSA 2019, Proceedings. Lecture Notes in Computer Science, vol. 11405, pp. 251–271. Springer (2019). https://doi.org/10.1007/978-3-030-12612-4_13
9. Damgård, I.: A design principle for hash functions. In: Brassard, G. (ed.) CRYPTO '89, Proceedings. Lecture Notes in Computer Science, vol. 435, pp. 416–427. Springer (1989). https://doi.org/10.1007/0-387-34805-0_39
10. Dodis, Y., Grubbs, P., Ristenpart, T., Woodage, J.: Fast message franking: From invisible salamanders to encryptment. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018, Proceedings, Part I. Lecture Notes in Computer Science, vol. 10991, pp. 155–186. Springer (2018). https://doi.org/10.1007/978-3-319-96884-1_6
11. Dodis, Y., Khovratovich, D., Mouha, N., Nandi, M.: T5: Hashing five inputs with three compression calls. Cryptology ePrint Archive, Report 2021/373 (2021), https://ia.cr/2021/373
12. FIPS PUB 180-4: Secure hash standard (SHS) (Aug 2015)
13. FIPS PUB 198-1: The keyed-hash message authentication code (HMAC) (2008)
14. Goldreich, O., Goldwasser, S., Micali, S.: How to construct random functions. Journal of the ACM **33**(4), 792–807 (1986). https://doi.org/10.1145/6490.6503
15. Goldwasser, S., Micali, S.: Probabilistic encryption. Journal of Computer and System Sciences **28**(2), 270–299 (1984). https://doi.org/10.1016/0022-0000(84)90070-9
16. Grubbs, P., Lu, J., Ristenpart, T.: Message franking via committing authenticated encryption. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Proceedings, Part III. Lecture Notes in Computer Science, vol. 10403, pp. 66–97. Springer (2017). https://doi.org/10.1007/978-3-319-63697-9_3
17. Hirose, S.: Sequential hashing with minimum padding. Cryptography **2**(2), 11 (2018). https://doi.org/10.3390/cryptography2020011
18. Hirose, S., Park, J.H., Yun, A.: A simple variant of the Merkle-Damgård scheme with a permutation. In: Kurosawa, K. (ed.) ASIACRYPT 2007, Proceedings. Lecture Notes in Computer Science, vol. 4833, pp. 113–129. Springer (2007). https://doi.org/10.1007/978-3-540-76900-2_7

19. Hirose, S., Yabumoto, A.: A tweak for a PRF mode of a compression function and its applications. In: Bica, I., Reyhanitabar, R. (eds.) SECITC 2016, Revised Selected Papers. Lecture Notes in Computer Science, vol. 10006, pp. 103–114 (2016). https://doi.org/10.1007/978-3-319-47238-6_7
20. ISO/IEC 9797-2:2021: Information security – Message authentication codes (MACs) – Part 2: Mechanisms using a dedicated hash-function (2021)
21. Iwata, T., Kurosawa, K.: OMAC: One-key CBC MAC. Cryptology ePrint Archive, Report 2002/180 (2002), https://ia.cr/2002/180
22. Iwata, T., Kurosawa, K.: OMAC: One-key CBC MAC. In: Johansson, T. (ed.) FSE 2003, Revised Papers. Lecture Notes in Computer Science, vol. 2887, pp. 129–153. Springer (2003). https://doi.org/10.1007/978-3-540-39887-5_11
23. Kuwakado, H., Hirose, S.: Pseudorandom-function property of the step-reduced compression functions of SHA-256 and SHA-512. In: Chung, K., Sohn, K., Yung, M. (eds.) WISA 2008, Revised Selected Papers. Lecture Notes in Computer Science, vol. 5379, pp. 174–189. Springer (2008). https://doi.org/10.1007/978-3-642-00306-6_13
24. Merkle, R.C.: One way hash functions and DES. In: Brassard, G. (ed.) CRYPTO '89, Proceedings. Lecture Notes in Computer Science, vol. 435, pp. 428–446. Springer (1989). https://doi.org/10.1007/0-387-34805-0_40
25. NIST Special Publication 800-38B: Recommendation for block cipher modes of operation: The CMAC mode for authentication (2005)