



Collision-Resistant and Pseudorandom Hash Function Using Tweakable Block Cipher

Shoichi Hirose^(✉) 

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

Abstract. This paper presents a method to construct a keyed Merkle-Damgård hash function satisfying collision resistance and the pseudorandom function property using a tweakable block cipher in the TWEAKEY framework. Its compression function adopts double-block construction to achieve sufficient level of collision resistance. Not only does the padding of the proposed keyed hash function not employ Merkle-Damgård strengthening, but it is also not injective. Due to the novel feature, the proposed keyed hash function achieves the minimum number of calls to its compression function for any message input. The proposed keyed hash function is shown to be optimally collision-resistant in the ideal cipher model. It is also shown to be a secure pseudorandom function if the underlying tweakable block cipher in the TWEAKEY framework is a secure tweakable pseudorandom permutation in two tweakable strategies.

Keywords: Hash function · Collision resistance · Pseudorandom function · Tweakable block cipher

1 Introduction

Background. Cryptographic hash functions are an important primitive in cryptography. They are classified into two classes: Unkeyed hash functions and keyed hash functions. The characteristic security requirement of an unkeyed hash function is collision resistance, which is the intractability of finding a pair of distinct inputs mapped to the same output. On the other hand, a keyed hash function is required to be a pseudorandom function (PRF) [10], which is indistinguishable from a uniform random function.

If a keyed hash function is a PRF satisfying collision resistance, then one can use it for computationally hiding and computationally binding string commitment. In addition, it has recently been shown that one can use it to achieve interesting cryptographic schemes such as compactly committing authenticated encryption with associated data (ccAEAD) [8, 11] and hash-based post-quantum EPID signatures [6]. In this paper, a keyed hash function satisfying collision resistance and the PRF property is called a collision-resistant and pseudorandom hash function.

HMAC [1] is a standardized keyed hash function in FIPS PUB 198-1 [9]. It is a collision-resistant and pseudorandom hash function. However, it is not so efficient for short message inputs.

Contribution. We present a method to construct a collision-resistant and pseudorandom hash function using a tweakable block cipher (TBC) in the TWEAKEY framework [19]. It is a kind of Merkle-Damgård iterated hash function [7, 20]. Its compression function adopts the double-block (DBL) construction [12] using a TBC to achieve sufficient level of collision resistance. Its domain extension extends KMDP^+ [13] to achieve a PRF using the DBL compression function.

The proposed construction does not use the Merkle-Damgård strengthening for padding. Due to the feature, it achieves the minimum number of calls to its compression function for any message input under the assumption that the message input is fed only into the message-block input of its compression function.

The proposed construction is shown to be optimally collision-resistant in the ideal cipher model. It is also shown to be a secure PRF if the underlying TBC in the TWEAKEY framework is a secure tweakable pseudorandom permutation (PRP) in two tweakkey strategies. In one tweakkey strategy, the underlying TBC is required to be a secure tweakable PRP against related-key attacks. However, the related-key attacks are not so powerful in that the key-deriving functions are chosen by the designers.

Related Work. There have been proposals of keyed hash functions satisfying the PRF property and collision resistance: HMAC [1], EMD [4], Keyed-MDP [15], and KMDP^+ [13]. All the constructions mentioned above except KMDP^+ use the Merkle-Damgård strengthening for their padding. Thus, in terms of the number of calls to the underlying compression function, our proposed construction is more efficient than these constructions though they are competitive to ours.

The Merkle-Damgård hash function keyed via the initial value with prefix-free padding is shown to be a secure PRF if its compression function is a secure PRF [2]. Our proof on the PRF property is based on this proof.

Iwata and Kurosawa designed a CBC-MAC function called CMAC [21], which achieves the minimum number of calls to its block cipher [17, 18]. Though it is shown to be a secure PRF if its block cipher is a secure PRP, it is not aimed at collision resistance.

Organization. Notations and definitions are given in Sect. 2. The proposed construction is presented in Sect. 3. It is shown to satisfy collision resistance in the ideal cipher model in Sect. 4. It is shown to be a secure PRF if the underlying tweakable block cipher in the TWEAKEY framework is a secure tweakable PRP in two tweakkey strategies in Sect. 5.

2 Preliminaries

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

The length of a sequence $x \in \Sigma^*$ is denoted by $|x|$. The least significant bit of x is denoted by $\text{lsb}(x)$. For sequences $y_i, y_{i+1}, \dots, y_{i+j} \in \Sigma^*$, their concatenation is denoted by $y_i \| y_{i+1} \| \dots \| y_{i+j}$ or $y_{[i, i+j]}$.

For sequences $x, y \in \Sigma^*$, $x \oplus y$ represents bit-wise XOR of x and y . If $|x| > |y|$, then $x \oplus y := x \oplus (0^{|x|-|y|} \| y)$.

Let $s \leftarrow \mathcal{S}$ represent that s is an element chosen uniformly at random from a set \mathcal{S} .

For integers a, b , and d , let $a \equiv_d b$ represent $a \equiv b \pmod{d}$.

2.1 Cryptographic Hash Function

A cryptographic hash function is a function mapping an input of arbitrary length to an output of fixed length. It is often called simply a hash function. The characteristic security requirement of a hash function is collision resistance.

Let H^P be a hash function using a primitive P . In this paper, P is assumed to be an ideal TBC. Namely, P is chosen uniformly at random from the set of all TBCs with the same domain and range.

Let \mathbf{A} be an adversary trying to find a colliding pair of inputs for H^P , which are a pair of distinct inputs mapped to the same output. \mathbf{A} can make encryption and decryption queries to its oracle P . The advantage of \mathbf{A} against H^P for collision resistance is given by

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

It is assumed that \mathbf{A} makes all the queries to P necessary to compute both $H^P(X)$ and $H^P(X')$.

2.2 Pseudorandom Function

Let $f : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ be a keyed function with its key space \mathcal{K} . A security requirement of f is indistinguishability from a uniform random function. The goal of an adversary \mathbf{A} against f is to distinguish between $f_K(\cdot) := f(K, \cdot)$ and a random oracle $\rho : \mathcal{X} \rightarrow \mathcal{Y}$, where $K \leftarrow \mathcal{K}$. \mathbf{A} has either f_K or ρ as an oracle and outputs 0 or 1. The advantage of \mathbf{A} against f as a PRF is defined as

$$\text{Adv}_f^{\text{prf}}(\mathbf{A}) := |\Pr[\mathbf{A}^{f_K} = 1] - \Pr[\mathbf{A}^\rho = 1]|.$$

f is called a secure PRF if no efficient adversary \mathbf{A} has any significant advantage. The advantage can be extended to adversaries with multiple oracles:

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

where $(K_1, \dots, K_p) \leftarrow \mathcal{K}^p$ and ρ_1, \dots, ρ_p are independent random oracles.

2.3 Tweakable Block Cipher in TWEAKEY Framework

A TBC in the TWEAKEY framework is a function $\tilde{e} : \Sigma^\nu \times \Sigma^n \rightarrow \Sigma^n$ with its tweak space Σ^ν such that, for every $Y \in \Sigma^\nu$, $\tilde{e}(Y, \cdot)$ is a permutation. We assume that $\tilde{e} : \Sigma^\nu \times \Sigma^n \rightarrow \Sigma^n$ is a family of TBCs with their key space and tweak space Σ^κ and Σ^τ , respectively, satisfying $\Sigma^\nu = \Sigma^\kappa \times \Sigma^\tau$.

Let $\mathcal{P}_{\tau,n}$ be the set of all tweakable permutations over Σ^n with their tweak space Σ^τ . Namely, for every $\varpi \in \mathcal{P}_{\tau,n}$ and every $T \in \Sigma^\tau$, $\varpi(T, \cdot)$ is a permutation over Σ^n .

A security requirement of a TBC $e : \Sigma^\kappa \times \Sigma^\tau \times \Sigma^n \rightarrow \Sigma^n$ is indistinguishability from a tweakable uniform random permutation. The advantage of an adversary \mathbf{A} against e as a tweakable PRP (TPRP) is defined as

$$\text{Adv}_e^{\text{tprp}}(\mathbf{A}) := |\Pr[\mathbf{A}^{e^\kappa} = 1] - \Pr[\mathbf{A}^\varpi = 1]|,$$

where $K \leftarrow \Sigma^\kappa$ and $\varpi \leftarrow \mathcal{P}_{\tau,n}$. \mathbf{A} is allowed to make queries in $\Sigma^\tau \times \Sigma^n$ adaptively to its oracle e_K or ϖ and outputs 0 or 1.

The following lemma is a kind of PRP/PRF switching lemma [5, 16] for a TBC.

Lemma 1. *For any adversary \mathbf{A} against a TBC $e : \Sigma^\kappa \times \Sigma^\tau \times \Sigma^n \rightarrow \Sigma^n$ taking at most t time and making at most q queries to its oracle, there exists an adversary \mathbf{X} against e such that*

$$\text{Adv}_e^{\text{prf}}(\mathbf{A}) \leq \text{Adv}_e^{\text{tprp}}(\mathbf{X}) + q^2/2^{n+1}$$

and \mathbf{X} takes at most t time and makes at most q queries.

2.4 PRF and TPRP Under Related-Key Attack

Let $f : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$. Let Φ be a set of functions from \mathcal{K} to \mathcal{K} . Let \mathbf{A} be an adversary against f making a related-key attack restricted to Φ (Φ -RKA) [3]: \mathbf{A} is given $g[K] : \Phi \times \mathcal{X} \rightarrow \mathcal{Y}$ such that $g[K](\varphi, X) := g(\varphi(K), X)$ as an oracle, where g is either f or a random oracle $\rho : \mathcal{K} \times \mathcal{X} \rightarrow \mathcal{Y}$ and $K \leftarrow \mathcal{K}$. φ is called a key-deriving function. The advantage of \mathbf{A} against f as a PRF under a Φ -RKA is defined as

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

f is called a secure PRF under Φ -RKAs if no efficient adversary \mathbf{A} has any significant advantage. The advantage of \mathbf{A} with p oracles is defined as

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

where $(K_1, \dots, K_p) \leftarrow \mathcal{K}^p$ and ρ_1, \dots, ρ_p are independent random oracles.

For a TBC e , $\text{Adv}_{e,\Phi}^{\text{tprp-rka}}(\mathbf{A})$ and $\text{Adv}_{e,\Phi}^{p\text{-tprp-rka}}(\mathbf{X})$ are defined similarly.

The following lemma is a kind of PRP/PRF switching lemma against adversaries making related-key attacks [14] for a TBC e .

Lemma 2. *Let \mathbf{A} be any adversary with p oracles against e taking at most t time and making at most q_i queries to its i -th oracle for $1 \leq i \leq p$. Let $q := q_1 + \dots + q_p$. Then, there exists an adversary \mathbf{X} against e such that*

$$\text{Adv}_{e,\Phi}^{p\text{-prf-rka}}(\mathbf{A}) \leq p \cdot \text{Adv}_{e,\Phi}^{t\text{prp-rka}}(\mathbf{X}) + q^2/2^{n+1}$$

and \mathbf{X} takes at most $t + O(qT_e)$ time and makes at most $\max\{q_1, q_2, \dots, q_p\}$ queries, where T_e represents the time required to compute e .

3 Proposed Construction

Let $E : \Sigma^\nu \times \Sigma^n \rightarrow \Sigma^n$ be a TBC in the TWEAKEY framework such that $\nu \geq 2n$. The proposed construction $C^E : \Sigma^n \times \Sigma^* \rightarrow \Sigma^{2n}$ is described in Algorithm 1. It is also depicted in Fig. 1. For the PRF property, C^E is viewed as a keyed function with its key space Σ^n . C^E incorporates constants $IV \in \Sigma^n$ and $c_{00}, c_{01}, c_{10}, c_{11}, \delta \in \Sigma^n \setminus \{0^n\}$. δ is a constant such that $\text{lsb}(\delta) = 1$. $c_{00}, c_{01}, c_{10}, c_{11}$ are distinct from each other. $\text{cf}^E : \Sigma^{2n} \times \Sigma^{\nu-n} \rightarrow \Sigma^{2n}$ is a compression function such that

$$\text{cf}^E(V_{i-1}, M_i) := E(V_{i-1} \| M_{i,0}, M_{i,1}) \| E(V_{i-1} \| M_{i,0}, M_{i,1} \oplus \delta),$$

where $M_i = M_{i,0} \| M_{i,1}$ and $|M_{i,1}| = n$. If $\nu = 2n$, then $M_{i,0} = \varepsilon$. $\text{pad} : \Sigma^* \rightarrow (\Sigma^{\nu-n})^+$ is a padding function such that

$$\text{pad}(M) := \begin{cases} M & \text{if } |M| > 0 \text{ and } |M| \equiv_{\nu-n} 0 \\ M \| 10^a & \text{otherwise,} \end{cases}$$

where a is the non-negative integer such that $|\text{pad}(M)|$ is the smallest multiple of $\nu - n$. Notice that $\text{pad}(\varepsilon) = 10^{\nu-n-1}$.

Remark 1. Let sw be a permutation over $\Sigma^n \times \Sigma^n$ such that $(x_0, x_1) \mapsto (x_1, x_0)$. Then, for any $(V_{i-1}, M_i) \in \Sigma^{2n} \times \Sigma^{\nu-n}$, $\text{cf}^E(V_{i-1}, M_i \oplus \delta) = \text{sw}(\text{cf}^E(V_{i-1}, M_i))$. For the PRF property, C^E is designed so that the *reflectiveness* of cf^E does not appear at the last call to cf^E . Notice that $\text{lsb}(M_m) \neq \text{lsb}(M_m \oplus \delta)$.

Algorithm 1: The proposed construction $C^E : \Sigma^n \times \Sigma^* \rightarrow \Sigma^{2n}$

input : (K, M)
output: $C^E(K, M)$
 $M_1 \| M_2 \| \dots \| M_m \leftarrow \text{pad}(M);$ */* $|M_i| = \nu - n$ for $1 \leq i \leq m$ */*
 $V_0 \leftarrow K \| IV;$
for $i = 1$ **to** $m - 1$ **do** $V_i \leftarrow \text{cf}^E(V_{i-1}, M_i);$
if $|M| > 0 \wedge |M| \equiv_{\nu-n} 0 \wedge \text{lsb}(M_m) = 0$ **then** $c \leftarrow c_{00};$
if $|M| > 0 \wedge |M| \equiv_{\nu-n} 0 \wedge \text{lsb}(M_m) = 1$ **then** $c \leftarrow c_{01};$
if $(|M| = 0 \vee |M| \not\equiv_{\nu-n} 0) \wedge \text{lsb}(M_m) = 0$ **then** $c \leftarrow c_{10};$
if $(|M| = 0 \vee |M| \not\equiv_{\nu-n} 0) \wedge \text{lsb}(M_m) = 1$ **then** $c \leftarrow c_{11};$
 $V_m \leftarrow \text{cf}^E(V_{i-1} \oplus c, M_i);$
return $V_m;$

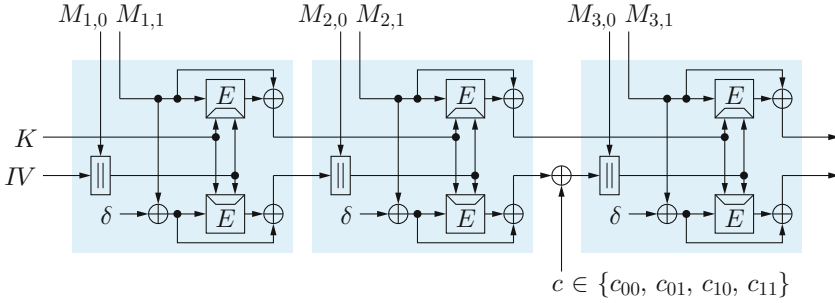


Fig. 1. The proposed construction

4 Collision Resistance

Any adversary needs $\Omega(2^n)$ queries to find a colliding pair of inputs for C^E under the assumption that E is an ideal cipher:

Theorem 1. For any adversary \mathbf{A} making at most q queries,

$$\text{Adv}_{C^E}^{\text{col}}(\mathbf{A}) \leq 7q/(2^n - 2q) + 5q(q - 1)/(2^n - 2q)^2.$$

Proof. Suppose that \mathbf{A} finds a colliding pair of inputs, (K, M) and (K', M') for C^E . Then, $C^E(K, M) = C^E(K', M')$ and $(K, M) \neq (K', M')$. Without loss of generality, suppose that $|M| \leq |M'|$. Let $\text{pad}(M) = M_1 \| M_2 \| \dots \| M_m$ and $\text{pad}(M') = M'_1 \| M'_2 \| \dots \| M'_{m'}$.

- (i) Suppose that $m = m' = 1$. If $K \neq K'$, then \mathbf{A} finds a colliding pair for cf^E . Otherwise, $M \neq M'$.
 - If $|M| = |M'| = \nu - n$, then \mathbf{A} finds a colliding pair for cf^E since $\text{pad}(M) \neq \text{pad}(M')$.
 - If $|M| < \nu - n$ and $|M'| < \nu - n$, then \mathbf{A} finds a colliding pair for cf^E since $\text{pad}(M) \neq \text{pad}(M')$.

- If $|M| < \nu - n$ and $|M'| = \nu - n$, then

$$\text{cf}^E(K\|(IV \oplus c), M_1) = \text{cf}^E(K\|(IV \oplus c'), M'_1),$$

where $c \in \{c_{10}, c_{11}\}$ and $c' \in \{c_{00}, c_{01}\}$. Thus, \mathbf{A} finds a colliding pair for cf^E since $\{c_{10}, c_{11}\} \cap \{c_{00}, c_{01}\} = \emptyset$.

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

$$\text{cf}^E(K\|(IV \oplus c), M_1) = \text{cf}^E(V'_{m'-1} \oplus c', M'_{m'}),$$

where

$$c \in \begin{cases} \{c_{00}, c_{01}\} & \text{if } |M| = \nu - n, \\ \{c_{10}, c_{11}\} & \text{if } |M| < \nu - n, \end{cases} \quad c' \in \begin{cases} \{c_{00}, c_{01}\} & \text{if } |M'| \equiv_{\nu-n} 0, \\ \{c_{10}, c_{11}\} & \text{if } |M'| \not\equiv_{\nu-n} 0. \end{cases}$$

If $(K\|(IV \oplus c), M_1) \neq (V'_{m'-1} \oplus c', M'_{m'})$, then \mathbf{A} finds a colliding pair for cf^E . Otherwise, $M_1 = M'_{m'}$ and the least significant n bits of $V'_{m'-1}$ equals $IV \oplus c \oplus c'$. Thus, \mathbf{A} finds an input for cf^E such that the least significant n bits of the corresponding output equals

- IV if $|M| = \nu - n$ and $|M'| \equiv_{\nu-n} 0$,
- IV if $|M| < \nu - n$ and $|M'| \not\equiv_{\nu-n} 0$, and
- $IV \oplus c_{00} \oplus c_{10}$ or $IV \oplus c_{01} \oplus c_{11}$ otherwise.

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

$$\text{cf}^E(V_{m-1} \oplus c, M_m) = \text{cf}^E(V'_{m'-1} \oplus c', M'_{m'})$$

where

$$c \in \begin{cases} \{c_{00}, c_{01}\} & \text{if } |M| \equiv_{\nu-n} 0, \\ \{c_{10}, c_{11}\} & \text{if } |M| \not\equiv_{\nu-n} 0, \end{cases} \quad c' \in \begin{cases} \{c_{00}, c_{01}\} & \text{if } |M'| \equiv_{\nu-n} 0, \\ \{c_{10}, c_{11}\} & \text{if } |M'| \not\equiv_{\nu-n} 0. \end{cases}$$

If $(V_{m-1} \oplus c, M_m) \neq (V'_{m'-1} \oplus c', M'_{m'})$, then \mathbf{A} finds a colliding pair for cf^E . Otherwise, $V_{m-1} \oplus V'_{m'-1} = 0^n \|(c \oplus c')$ and $M_m = M'_{m'}$.

- If $|M| \equiv_{\nu-n} 0$ and $|M'| \not\equiv_{\nu-n} 0$, or $|M| \not\equiv_{\nu-n} 0$ and $|M'| \equiv_{\nu-n} 0$, then \mathbf{A} finds a colliding pair for cf^E wrt $0^n \|(c_{00} \oplus c_{10})$ or $0^n \|(c_{01} \oplus c_{11})$, that is, a pair of inputs (V_{m-2}, M_{m-1}) and $(V'_{m'-2}, M'_{m'-1})$ such that $\text{cf}^E(V_{m-2}, M_{m-1}) \oplus \text{cf}^E(V'_{m'-2}, M'_{m'-1})$ equals $0^n \|(c_{00} \oplus c_{10})$ or $0^n \|(c_{01} \oplus c_{11})$.
- Suppose that $|M| \equiv_{\nu-n} 0$ and $|M'| \equiv_{\nu-n} 0$. Then, since $M_m = M'_{m'}$, $c = c'$ and $V_{m-1} = V'_{m'-1}$. If $m = m'$, then \mathbf{A} finds a colliding pair for cf^E since $(K, M) \neq (K', M')$. If $m < m'$, then \mathbf{A} finds a colliding pair for cf^E or an input for cf^E such that the least significant n bits of the corresponding output equals IV .
- Suppose that $|M| \not\equiv_{\nu-n} 0$ and $|M'| \not\equiv_{\nu-n} 0$. This case is similar to the case above. Thus, \mathbf{A} finds a colliding pair for cf^E or an input for cf^E such that the least significant n bits of the corresponding output equals IV .

Thus, a colliding pair for C^E implies for cf^E

1. a colliding pair,
2. a colliding pair wrt $0^n \parallel (c_{00} \oplus c_{10})$ or $0^n \parallel (c_{01} \oplus c_{11})$, or
3. an input mapped to an output whose least significant n bits equals IV , $IV \oplus c_{00} \oplus c_{10}$ or $IV \oplus c_{01} \oplus c_{11}$.

Let us consider an adversary $\tilde{\mathbf{A}}$ running \mathbf{A} . For each query \mathbf{A} , $\tilde{\mathbf{A}}$ makes at most 2 queries. Thus, $\tilde{\mathbf{A}}$ makes at most $2q$ queries in total.

For a query of \mathbf{A} , if $\tilde{\mathbf{A}}$ already knows the corresponding answer, then $\tilde{\mathbf{A}}$ simply returns it to \mathbf{A} . Suppose that $\tilde{\mathbf{A}}$ does not know the answer. If the query of \mathbf{A} is an encryption query (TK, PT) , then $\tilde{\mathbf{A}}$ asks (TK, PT) and $(TK, PT \oplus \delta)$ to E and receives replies CT and CT' , respectively. Then, $\tilde{\mathbf{A}}$ returns CT to \mathbf{A} . If the query of \mathbf{A} is a decryption query (TK, CT) , then $\tilde{\mathbf{A}}$ asks (TK, CT) to E^{-1} , receives a reply PT and returns it to \mathbf{A} . Then, $\tilde{\mathbf{A}}$ asks $(TK, PT \oplus \delta)$ to E and receives a reply CT' . In both of the cases, $\tilde{\mathbf{A}}$ gets (TK, PT, CT) and $(TK, PT \oplus \delta, CT')$. Notice that $\text{cf}^E(TK_v, TK_m \parallel PT) = (CT \oplus PT) \parallel (CT' \oplus PT \oplus \delta)$ and $\text{cf}^E(TK_v, TK_m \parallel (PT \oplus \delta)) = (CT' \oplus PT \oplus \delta) \parallel (CT \oplus PT)$, where $TK = TK_v \parallel TK_m$ and $|TK_v| = 2n$. Also notice that, if $CT \oplus CT' = \delta$, then $\text{cf}^E(TK_v, TK_m \parallel PT) = \text{cf}^E(TK_v, TK_m \parallel (PT \oplus \delta))$.

Let (TK_j, PT_j, CT_j) and $(TK_j, PT_j \oplus \delta, CT'_j)$ be the tuples obtained by $\tilde{\mathbf{A}}$ for the j -th query of \mathbf{A} . Let $U_j := CT_j \oplus PT_j$ and $U'_j := CT'_j \oplus PT_j \oplus \delta$. Then, for an execution of $\tilde{\mathbf{A}}$, the j -th query of \mathbf{A} induces 1 or 2 above for cf^E if $U_j = U'_j$ or there exists some $j' < j$ such that

- $U_j \parallel U'_j \in \{U_{j'} \parallel U'_{j'}, U_{j'} \parallel (U'_{j'} \oplus c_{00} \oplus c_{10}), U_{j'} \parallel (U'_{j'} \oplus c_{01} \oplus c_{11})\}$,
- $U'_j \parallel U_j \in \{U'_{j'} \parallel U_{j'}, U'_{j'} \parallel (U_{j'} \oplus c_{00} \oplus c_{10}), U'_{j'} \parallel (U_{j'} \oplus c_{01} \oplus c_{11})\}$,
- $U'_j \parallel U_j \in \{U_{j'} \parallel U'_{j'}, U_{j'} \parallel (U'_{j'} \oplus c_{00} \oplus c_{10}), U_{j'} \parallel (U'_{j'} \oplus c_{01} \oplus c_{11})\}$, or
- $U'_j \parallel U_j \in \{U'_{j'} \parallel U_{j'}, U'_{j'} \parallel (U_{j'} \oplus c_{00} \oplus c_{10}), U'_{j'} \parallel (U_{j'} \oplus c_{01} \oplus c_{11})\}$.

Thus, the probability that the j -th query of \mathbf{A} induces 1 or 2 above for cf^E is at most $10(j-1)/(2^n - 2q)^2 + 1/(2^n - 2q)$. The probability that it induces 3 above for cf^E is at most $6/(2^n - 2q)$. Since \mathbf{A} makes at most q queries,

$$\sum_{j=1}^q (10(j-1)/(2^n - 2q)^2 + 7/(2^n - 2q)) \leq 7q/(2^n - 2q) + 5q(q-1)/(2^n - 2q)^2.$$

It is also an upper bound on $\text{Adv}_{C^E}^{\text{col}}(\mathbf{A})$. □

5 Pseudorandom-Function Property

The proposed construction C^E treats the TBC $E : \Sigma^\nu \times \Sigma^n \rightarrow \Sigma^n$ in the TWEAKEY framework in two tweak strategies: $\Sigma^\nu := \Sigma^n \times \Sigma^{\nu-n}$ in one strategy and $\Sigma^\nu := \Sigma^{2n} \times \Sigma^{\nu-2n}$ in the other strategy. We denote E in the former and the latter tweak strategies by \tilde{E} and $\tilde{\tilde{E}}$, respectively.

For \tilde{E} , we consider related-key attacks with related-key-deriving functions $\tilde{\Phi} := \{\text{id}, \text{sw}, \times_{c_{00}}, \times_{c_{01}}, \times_{c_{10}}, \times_{c_{11}}, \text{sw} \circ \times_{c_{00}}, \text{sw} \circ \times_{c_{01}}, \text{sw} \circ \times_{c_{10}}, \text{sw} \circ \times_{c_{11}}\}$, where

- id is the identity permutation over Σ^{2n} ,
- sw is a permutation over $\Sigma^n \times \Sigma^n$ such that $(x_0, x_1) \mapsto (x_1, x_0)$, and
- for $c \in \Sigma^n$, \times_c is a permutation over Σ^{2n} such that $x \mapsto x \oplus c$.

C^E is a secure PRF if \dot{E} is a secure TPRP and \ddot{E} is a secure TPRP under Φ -related-key attacks:

Theorem 2. *For any adversary \mathbf{A} against C^E taking at most t time and making at most q queries each of which has at most ℓ blocks after padding, there exist adversaries $\dot{\mathbf{A}}$ against \dot{E} and $\ddot{\mathbf{A}}$ against \ddot{E} such that*

$$\text{Adv}_{C^E}^{\text{prf}}(\mathbf{A}) \leq \text{Adv}_{\dot{E}}^{\text{tprp}}(\dot{\mathbf{A}}) + (\ell - 1)q \text{Adv}_{\ddot{E}, \Phi}^{\text{tprp-rka}}(\ddot{\mathbf{A}}) + \ell q^2 / 2^{n+1} + (\ell - 1)q / 2^n.$$

Both $\dot{\mathbf{A}}$ and $\ddot{\mathbf{A}}$ take at most about $t + O(\ell q T_E)$ time and make at most $2q$ queries, where T_E is time required to compute E .

Proof. Let $\text{I}^c : \Sigma^{2n} \times (\Sigma^{\nu-n})^+ \rightarrow \Sigma^{2n}$ be a keyed function specified in Algorithm 2. For an integer $k \geq 0$ and functions $\zeta : \Sigma^* \rightarrow \Sigma^{2n}$ and $\eta : (\Sigma^{\nu-n})^* \rightarrow \Sigma^{2n}$, let $\text{Hy}[k]^{\zeta, \eta} : \Sigma^* \rightarrow \Sigma^{2n}$ be a function specified as follows: For $M \in \Sigma^*$ such that $\text{pad}(M) = M_1 \| M_2 \| \cdots \| M_m$ and $|M_i| = \nu - n$ for $1 \leq i \leq m$,

$$\text{Hy}[k]^{\zeta, \eta}(M) := \begin{cases} \zeta(M) & \text{if } m \leq k, \\ \text{I}^c(\eta(M_{[1, k]}), M_{[k+1, m]}) & \text{if } m > k, \end{cases}$$

and

$$c \leftarrow \begin{cases} c_{00} & \text{if } |M| > 0 \wedge |M| \equiv_{\nu-n} 0 \wedge \text{lsb}(M_m) = 0, \\ c_{01} & \text{if } |M| > 0 \wedge |M| \equiv_{\nu-n} 0 \wedge \text{lsb}(M_m) = 1, \\ c_{10} & \text{if } (|M| = 0 \vee |M| \not\equiv_{\nu-n} 0) \wedge \text{lsb}(M_m) = 0, \\ c_{11} & \text{if } (|M| = 0 \vee |M| \not\equiv_{\nu-n} 0) \wedge \text{lsb}(M_m) = 1. \end{cases} \quad (1)$$

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

Suppose that ζ is a random oracle and η is a random function such that

- $\eta(\varepsilon) \leftarrow \Sigma^n \times \{IV\}$, and
- for every $k \geq 1$ and $M_{[1, k]}$ such that $\text{lsb}(M_k) = 0$, $\eta(M_{[1, k]}) \leftarrow \Sigma^{2n}$ and $\eta(M_{[1, k]} \oplus \delta) \leftarrow \text{sw}(\eta(M_{[1, k]}))$.

Then,

$$\text{Hy}[0]^{\zeta, \eta}(M) = \text{I}^c(\eta(\varepsilon), M_{[1, m]})$$

and c is chosen as specified by Formula (1). Thus, $\text{Hy}[0]^{\zeta, \eta}$ is equivalent to C^E . $\text{Hy}[\ell]^{\zeta, \eta}$ works as a random oracle for any $M \in \Sigma^*$ such that $\text{pad}(M)$ consists of at most ℓ blocks. Since every query made by \mathbf{A} is assumed to consist of at most ℓ blocks after padding,

$$\text{Adv}_{C^E}^{\text{prf}}(\mathbf{A}) = |\Pr[\mathbf{A}^{\text{Hy}[0]^{\zeta, \eta}} = 1] - \Pr[\mathbf{A}^{\text{Hy}[\ell]^{\zeta, \eta}} = 1]|.$$

Let $\Delta_k := |\Pr[\mathbf{A}^{\text{Hy}[k]^{\zeta, \eta}} = 1] - \Pr[\mathbf{A}^{\text{Hy}[k+1]^{\zeta, \eta}} = 1]|$. Then,

$$\text{Adv}_{C^E}^{\text{prf}}(\mathbf{A}) \leq \Delta_0 + \Delta_1 + \cdots + \Delta_{\ell-1}. \quad (2)$$

For Δ_0 , let \mathbf{D}_0 be an adversary against \dot{E} . \mathbf{D}_0 runs \mathbf{A} and produces the same output as \mathbf{A} . It also simulates the oracle of \mathbf{A} using its oracle. Let $\dot{F} : \Sigma^{\nu-n} \times \Sigma^n \rightarrow \Sigma^n$ be the oracle of \mathbf{D}_0 , which is either \dot{E}_K or $\dot{\rho}$, where $K \leftarrow \Sigma^n$ and $\dot{\rho}$ is a random oracle. For each query M of \mathbf{A} such that $\text{pad}(M) = M_1 \| \dots \| M_m$, \mathbf{D}_0 acts as follows: For $1 \leq i \leq m$, $M_i := M_{i,0} \| M_{i,1}$, where $|M_{i,0}| = \nu - 2n$ and $|M_{i,1}| = n$.

- If $m = 1$, then \mathbf{D}_0 asks $((IV \oplus c) \| M_{1,0}, M_{1,1})$ and $((IV \oplus c) \| M_{1,0}, M_{1,1} \oplus \delta)$ to \dot{F} and returns $\text{cf}^{\dot{F}}(K \| (IV \oplus c), M_1)$ to \mathbf{A} .
- If $m \geq 2$, then \mathbf{D}_0 asks $(IV \| M_{1,0}, M_{1,1})$ and $(IV \| M_{1,0}, M_{1,1} \oplus \delta)$ to \dot{F} and returns $\text{l}^c(\text{cf}^{\dot{F}}(K \| IV, M_1), M_{[2,m]})$ to \mathbf{A} .

In both of the cases above, c is chosen as specified by Formula (1). \mathbf{D}_0 implements $\text{Hy}[0]^{\zeta, \eta}$ as the oracle of \mathbf{A} if its oracle is \dot{E}_K . It implements $\text{Hy}[1]^{\zeta, \eta}$ if its oracle is $\dot{\rho}$ since $\dot{\rho}((IV \oplus c_{00}) \| M_{1,0}, M_{1,1})$, $\dot{\rho}((IV \oplus c_{01}) \| M_{1,0}, M_{1,1})$, $\dot{\rho}((IV \oplus c_{10}) \| M_{1,0}, M_{1,1})$, $\dot{\rho}((IV \oplus c_{11}) \| M_{1,0}, M_{1,1})$ and $\dot{\rho}(IV \| M_{1,0}, M_{1,1})$ are independent from each other. Thus,

$$\Delta_0 = |\Pr[\mathbf{D}_0^{\dot{E}_K} = 1] - \Pr[\mathbf{D}_0^{\dot{\rho}} = 1]| = \text{Adv}_{\dot{E}}^{\text{prf}}(\mathbf{D}_0). \quad (3)$$

\mathbf{D}_0 takes at most about $t + O(\ell q T_E)$ time and makes at most $2q$ queries.

Suppose that $1 \leq k \leq \ell - 1$. For Δ_k , let \mathbf{D}_k be an adversary making a Φ -related-key attack on \dot{E} . \mathbf{D}_k runs \mathbf{A} and produces the same output as \mathbf{A} . It also simulates the oracle of \mathbf{A} using its oracle. \mathbf{D}_k has q oracles $\dot{F}_i[K_i] : \Sigma^{\nu-2n} \times \Sigma^n \rightarrow \Sigma^n$ for $1 \leq i \leq q$. They are either $\dot{E}[K_1], \dots, \dot{E}[K_q]$ or $\dot{\rho}_1[K_1], \dots, \dot{\rho}_q[K_q]$, where $K_i \leftarrow \Sigma^{2n}$ and $\dot{\rho}_i$ is a random oracle for $1 \leq i \leq q$. For the j -th query M of \mathbf{A} , let $\text{pad}(M) = M_1 \| \dots \| M_m$, where $M_i := M_{i,0} \| M_{i,1}$, $|M_{i,0}| = \nu - 2n$ and $|M_{i,1}| = n$ for $1 \leq i \leq m$. Suppose that $m \leq k$. Then, \mathbf{D}_k simulates ζ and returns $\zeta(M)$ to \mathbf{A} . Suppose that $m > k$. Let \mathcal{J} be a set of integers j' such that $j' < j$ and the j' -th query M' of \mathbf{A} satisfies $m' > k$ and $M'_{[1,k]} = M_{[1,k]} \vee M'_{[1,k]} = M_{[1,k-1]} \| M_{k,0} \| (M_{k,1} \oplus \delta)$, where $\text{pad}(M') = M'_1 \| \dots \| M'_{m'}$. Let $j^* \leftarrow j$ if $\mathcal{J} = \emptyset$ and $j^* \leftarrow \min \mathcal{J}$ otherwise. Let M^* be the j^* -th query of \mathbf{A} . Then, for the j -th query M of \mathbf{A} , \mathbf{D}_k acts as follows:

- Suppose that $m = k + 1$. Then,
 - \mathbf{D}_k asks $(x_c, M_{k+1,0}, M_{k+1,1})$ and $(x_c, M_{k+1,0}, M_{k+1,1} \oplus \delta)$ to $\dot{F}_{j^*}[K_{j^*}]$ and returns $\text{cf}^{\dot{F}_{j^*}}(K_{j^*} \oplus c, M_{k+1})$ to \mathbf{A} if $j^* = j$ or $j^* < j \wedge M^*_{[1,k]} = M_{[1,k]}$, and
 - \mathbf{D}_k asks $(\text{sw} \circ x_c, M_{k+1,0}, M_{k+1,1})$ and $(\text{sw} \circ x_c, M_{k+1,0}, M_{k+1,1} \oplus \delta)$ to $\dot{F}_{j^*}[K_{j^*}]$ and returns $\text{cf}^{\dot{F}_{j^*}}(\text{sw}(K_{j^*}) \oplus c, M_{k+1})$ to \mathbf{A} otherwise.

In both of the cases above, c is chosen as specified by Formula (1).
- Suppose that $m \geq k + 2$. Then,
 - \mathbf{D}_k asks $(\text{id}, M_{k+1,0}, M_{k+1,1})$ and $(\text{id}, M_{k+1,0}, M_{k+1,1} \oplus \delta)$ to $\dot{F}_{j^*}[K_{j^*}]$ and returns $\text{l}^c(\text{cf}^{\dot{F}_{j^*}}(K_{j^*}, M_{k+1}), M_{[k+2,m]})$ to \mathbf{A} if $j^* = j$ or $j^* < j \wedge M^*_{[1,k]} = M_{[1,k]}$, and

- \mathbf{D}_k asks $(\mathbf{sw}, M_{k+1,0}, M_{k+1,1})$ and $(\mathbf{sw}, M_{k+1,0}, M_{k+1,1} \oplus \delta)$ to $\ddot{F}_{j^*}[K_{j^*}]$ and returns $!^c(\text{cf}^{\ddot{F}_{j^*}}(\mathbf{sw}(K_{j^*}), M_{k+1}), M_{[k+2,m]})$ to \mathbf{A} otherwise.

In both of the cases above, c is chosen as specified by Formula (1).

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 $\ddot{F}_j[K_j]$ to compute the answer to the query. It implies that new K_j , which is chosen uniformly at random, is assigned to new $M_{[1,k]}$. Suppose that the oracles of \mathbf{D}_k are $\ddot{E}[K_1], \dots, \ddot{E}[K_q]$. Then, \mathbf{D}_k implements $\text{Hy}[k]^{\zeta, \eta}$ as the oracle of \mathbf{A} . On the other hand, suppose that the oracles of \mathbf{D}_k are $\ddot{\rho}_1[K_1], \dots, \ddot{\rho}_q[K_q]$. Then, \mathbf{D}_k implements $\text{Hy}[k+1]^{\zeta, \eta}$ if $K_i \neq \mathbf{sw}(K_i)$ for every i such that $1 \leq i \leq q$. Thus,

$$\begin{aligned} \Delta_k &= \text{Adv}_{\ddot{E}, \Phi}^{q\text{-prf-rka}}(\mathbf{D}_k) + |\Pr[\mathbf{D}_k^{\ddot{\rho}[K_1], \dots, \ddot{\rho}[K_q]} = 1] - \Pr[\mathbf{A}^{\text{Hy}[k+1]^{\zeta, \eta}} = 1]| \\ &\leq \text{Adv}_{\ddot{E}, \Phi}^{q\text{-prf-rka}}(\mathbf{D}_k) + q/2^n. \end{aligned} \quad (4)$$

\mathbf{D}_k takes at most about $t + O(\ell q T_E)$ time and makes at most $2q$ queries.

From Inequality (2), Equalities (3) and (4), and Lemmas 1 and 2, there exist adversaries \mathbf{A} and \mathbf{A} such that

$$\text{Adv}_{\mathcal{C}^E}^{\text{prf}}(\mathbf{A}) \leq \text{Adv}_{\ddot{E}}^{\text{tprp}}(\mathbf{A}) + (\ell - 1)q \cdot \text{Adv}_{\ddot{E}, \Phi}^{\text{tprp-rka}}(\mathbf{A}) + \ell q^2/2^{n+1} + (\ell - 1)q/2^n.$$

Both \mathbf{A} and \mathbf{A} take at most about $t + O(\ell q T_E)$ time and make at most $2q$ queries. \square

Algorithm 2: $!^c : \Sigma^{2n} \times (\Sigma^{\nu-n})^+ \rightarrow \Sigma^{2n}$

input : $(W, X_1 \| X_2 \cdots \| X_x)$

output: $!^c(W, X_1 \| X_2 \cdots \| X_x)$

$V_0 \leftarrow W;$

/ $|X_i| = \nu - n$ for $1 \leq i \leq x$ */*

for $i = 1$ **to** $x - 1$ **do** $V_i \leftarrow \text{cf}^E(V_{i-1}, X_i)$ $V_x \leftarrow \text{cf}^E((V_{x-1} \oplus c), X_x);$

return $V_x;$

Acknowledgements. This work was supported by JSPS KAKENHI Grant Number JP21K11885.

References

1. Bellare, M., Canetti, R., Krawczyk, H.: Keying hash functions for message authentication. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 1–15. Springer, Heidelberg (1996). https://doi.org/10.1007/3-540-68697-5_1
2. 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)

3. Bellare, M., Kohno, T.: A theoretical treatment of related-key attacks: RKA-PRPs, RKA-PRFs, and applications. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 491–506. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-39200-9_31
4. Bellare, M., Ristenpart, T.: Multi-property-preserving hash domain extension and the EMD transform. In: Lai, X., Chen, K. (eds.) ASIACRYPT 2006. LNCS, vol. 4284, pp. 299–314. Springer, Heidelberg (2006). https://doi.org/10.1007/11935230_20
5. Bellare, M., Rogaway, P.: Code-based game-playing proofs and the security of triple encryption. Cryptology ePrint Archive, Report 2004/331 (2006). <http://eprint.iacr.org/>
6. Boneh, D., Eskandarian, S., Fisch, B.: Post-quantum EPID signatures from symmetric primitives. In: Matsui, M. (ed.) CT-RSA 2019. LNCS, vol. 11405, pp. 251–271. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12612-4_13
7. Damgård, I.B.: A design principle for hash functions. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 416–427. Springer, New York (1990). https://doi.org/10.1007/0-387-34805-0_39
8. Dodis, Y., Grubbs, P., Ristenpart, T., Woodage, J.: Fast message franking: from invisible salamanders to encryptment. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018. LNCS, vol. 10991, pp. 155–186. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96884-1_6
9. FIPS PUB 198-1: The keyed-hash message authentication code (HMAC) (2008)
10. Goldreich, O., Goldwasser, S., Micali, S.: How to construct random functions. J. ACM **33**(4), 792–807 (1986). <https://doi.org/10.1145/6490.6503>
11. Grubbs, P., Lu, J., Ristenpart, T.: Message franking via committing authenticated encryption. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017. LNCS, vol. 10403, pp. 66–97. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63697-9_3
12. Hirose, S.: Some plausible constructions of double-block-length hash functions. In: Robshaw, M. (ed.) FSE 2006. LNCS, vol. 4047, pp. 210–225. Springer, Heidelberg (2006). https://doi.org/10.1007/11799313_14
13. Hirose, S.: Collision-resistant and pseudorandom function based on Merkle-Damgård hash function. In: Park, J.H., Seo, S. (eds.) ICISC 2021. LNCS, vol. 13218, pp. 325–338. Springer, Cham (2021). https://doi.org/10.1007/978-3-031-08896-4_17
14. Hirose, S., Ideguchi, K., Kuwakado, H., Owada, T., Preneel, B., Yoshida, H.: An AES based 256-bit hash function for lightweight applications: Lesamnta-LW. IEICE Trans. Fundam. **E95-A**(1), 89–99 (2012)
15. Hirose, S., Park, J.H., Yun, A.: A simple variant of the Merkle-Damgård scheme with a permutation. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 113–129. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76900-2_7
16. Impagliazzo, R., Rudich, S.: Limits on the provable consequences of one-way permutations. In: Goldwasser, S. (ed.) CRYPTO 1988. LNCS, vol. 403, pp. 8–26. Springer, New York (1990). https://doi.org/10.1007/0-387-34799-2_2
17. Iwata, T., Kurosawa, K.: OMAC: One-key CBC MAC. Cryptology ePrint Archive, Report 2002/180 (2002). <https://ia.cr/2002/180>
18. Iwata, T., Kurosawa, K.: OMAC: one-key CBC MAC. In: Johansson, T. (ed.) FSE 2003. LNCS, vol. 2887, pp. 129–153. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-39887-5_11

19. Jean, J., Nikolić, I., Peyrin, T.: Tweaks and keys for block ciphers: the **TWEAKEY** framework. In: Sarkar, P., Iwata, T. (eds.) ASIACRYPT 2014. LNCS, vol. 8874, pp. 274–288. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45608-8_15
20. Merkle, R.C.: One way hash functions and DES. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 428–446. Springer, New York (1990). https://doi.org/10.1007/0-387-34805-0_40
21. NIST Special Publication 800-38B: Recommendation for block cipher modes of operation: The CMAC mode for authentication (2005)