Various Security Analysis of a pfCM-MD Hash Domain Extension and Applications based on the Extension

Donghoon Chang[1], Seokhie Hong[1], Jaechul Sung[2], and Sangjin Lee[1]

[1] Center for Information Security Technologies(CIST), Korea University, Korea
pointchang@gmail.com
{hsh,sangjin}@cist.korea.ac.kr
[2] Department of Mathematics, University of Seoul, Korea
jcsung@uos.ac.kr

Abstract. We propose a new hash domain extension *a prefix-free-Counter-Masking-MD (pfCM-MD).* And, among security notions for the hash function, we focus on the indifferentiable security notion by which we can check whether the structure of a given hash function has any weakness or not. Next, we consider the security of HMAC, two new prf constructions, NIST SP 800-56A key derivation function, and the randomized hashing in NIST SP 800-106, where all of them are based on the pfCM-MD. Especially, due to the counter of the pfCM-MD, the pfCM-MD are secure against all of generic second-preimage attacks such as Kelsey-Schneier attack [20] and Elena *et al.*' attck [1]. Our proof technique and most of notations follow those in [6, 3, 4].

1 Introduction

Since a standard hash function may be used in various areas, it is very important to identify security requirements of the hash function for the implementation of secure cryptosystems in each area. Based on such information, designers of hash functions do the best so that a developed hash algorithm may satisfy all of the security requirements. Usually, the security requirements are concentrated on the underlying compression function because most of hash functions are designed with a domain extension and an underlying compression function. Therefore, we have to know what kinds of security requirements are needed for the underlying compression function.

For development of SHA-3, NIST [28] recently announced that HMAC [5], alternative pseudorandom function (in short, prf) constructions (which are not fixed and will be proposed by designers of SHA-3 candidate), NIST SP 800-56A key derivation function [25], the randomized hashing in NIST SP 800-106 [27] and pseudorandom-bit generator [26] based on a new hash function should be secure. In this paper, except for pseudorandom-bit generator [26], we consider the security requirements of the underlying compression function of our new domain extension "pfCM-MD" for their securities. In the case of pseudorandom-bit generator [26], there are two constructions : HMAC_DRBG and Hash_DRBG. The security of HMAC_DRBG depends on the prf security of HMAC based on a underlying hash function [19]. Since we prove the prf security of HMAC based on pfCM-MD in Sect. 4, if the compression function of pfCM-MD satisfies some security requirements described in Sect. 4, the security of HMAC_DRBG based on pfCM-MD are guaranteed. In the case of Hash_DRBG, $T = H(Z)||H(Z+1)||\cdots||H(Z+i)$ is used as a pseudorandom bit string where H is a hash function, Z is a secret value, and Z is newly updated whenever the bit length of T is larger than $2^{19}-1$. When the bit-size of Z is less than the block size b of the compression function (see Sect. 2), it can be easily shown that the security of Hash_DRBG depends on the rka-prf of the compression function of a hash function in the related-key attack model.

Addition to above applications, a standard hash function may be used in other applications so that we may need new security requirements. However, we cannot define any security requirement because new applications are not defined. Fortunately, due to Maurer *et al.*'s work [22], where the new security notion *Indifferentiability* is introduced, we can measure the security of a given domain extension against any adversary, under the assumption that the underlying compression function is ideal such as the ideal cipher and the random oracle models. So we give a simple indifferentiable security analysis on pfCM-MD. Our new domain extension has several advantages when compared with other domain extensions.

- Use of counter : During the computation of a hash value for a given message, each compression function uses a different counter. So all of generic second-preimage attacks such as Kelsey-Schneier attack [20] and Elena *et al.*' attck [1] cannot be applied to pfCM-MD. On the other hand, in cases of domain extensions without any counter such as MDP [18], which was proposed by Hirose, Park and Yun, does not guarantee the full security against them.
- Characteristic of counter : The pfCM-MD domain extension XORs (where the operation is $\oplus$) a counter with the input chaining variables of each compression function during the computation of a hash value. Since a counter is just XORed with the input chaining variables of the compression function, we do not need to make the input size of the compression function large. More precisely, in the case of pfCM-MD, $f(c \oplus i, m)$ is used where i is the counter, and f is the underlying compression function. On the other hand, for example, in the case of HAIFA domain extension [11], which was proposed by Biham and Dunkelman, a counter should be a part of the input string of the compression function. That is, if the bit-size of the counter is larger, then the bit-size of an input message block per the compression function is reduced, because the total input size of the underlying compression function is already fixed. More precisely, in the case of HAIFA, $f(c, m||i)$ is used where i is the counter, and f is the underlying compression function.

Organization: The organization of this paper is as follows. In Sect. 2, we introduce notations, definitions, and known results for security proofs. In Sect. 3, we give the indifferentiable security proof on the *pfCM-MD*. In Sect. 4, we provide a prf security of HMAC based on the *pfCM-MD*. In Sect. 5, we define two prf constructions based on the *pfCM-MD* and prove the prf security of them. In Sect. 6, we provide a prf security of NIST SP 800-56A key derivation function based on the *pfCM-MD*. In Sect. 7, we provide eTCR security analysis of *pfCM-MD* with the message randomization (in short, mr) in NIST SP 800-106.

2 Notations, Definitions and Known Results

Here we consider the compression function $f : \{0,1\}^n \times \{0,1\}^b \to \{0,1\}^n$. We write $||m||_b = k$ if $m \in \{0,1\}^{kb}$. That is, m is a message of k b-bit blocks. We denote the set of all functions from the domain $\mathcal{C}$ to the codomain $\mathcal{D}$ by $\mathsf{Maps}(\mathcal{C}, \mathcal{D})$.

Padding. We say any injective and length-consistent function $\mathsf{pad} : \{0,1\}^* \to (\{0,1\}^b)^*$ as a padding rule.

MD [24, 16]. The traditional Merkle-Damgård extension (MD) works as follow: for a message M, $\mathsf{pad}(M) = m_1|| \cdots ||m_t$ and $\mathrm{MD}^f_{\mathsf{pad}}(IV, M) = f(\cdots f(f(IV, m_1), m_2) \cdots, m_t)$, where f is a compression function and IV is the initial value.

pfCM-MD. CM-MD (MD with a counter-masking) works similar to MD as follow : for given a message M, $\mathsf{pad}(M) = m_1||\cdots||m_t$ and CM-MD$^f_{\mathsf{pad}}(IV, M)$=CM-MD$^f(IV,\mathsf{pad}(M)) = f(\cdots f(\, f(\, IV \oplus c_0, m_1) \oplus c_1, m_2) \oplus c_3, \cdots, m_t)$. For any two $c = c_0||\cdots||c_{t-1}$ and $c' = c'_0||\cdots||c'_{t'-1}$, if c is not a prefix of c', then we say its counter-masking is prefix-free. So, pfCM-MD means prefix-free-Counter-Masking-MD. One example is a case that for any $c = c_0||\cdots||c_{t-1}$, where $c_0 = 0$ and $c_{i+1} = c_i + 1$ for $0 \le i \le t-3$ and $c_{t-1} = P$, where P is a fixed value bigger than other counter c_j's. For example, when the maximum bit-size of an input message is $2^{64} - 1$, P can be any value larger than or equal to 2^{64}. When the maximum bit-size of an input message is $2^{128} - 1$, any value P can be any value larger than or equal to 2^{128}. In this document, in the case that $c_0 = d$ and $c_{i+1} = c_i+1$ for $0 \le i \le t-3$ and $c_{t-1} = P$, we denote it by pfCM^d-MD.

chop. For $0 \le s \le n$ we define $\mathrm{chop}_s(x) = x_L$ where $x = x_L \parallel x_R$ and $|x_R| = s$.

pfCM-chopMD. pfCM-chopMD$^f_{\mathsf{pad}}(IV, M) = \mathrm{chop}_s(\text{pfCM-MD}^f_{\mathsf{pad}}(IV, M))$. Note that pfCM-chopMD with $s = 0$ is the same as pfCM-MD. That is, pfCM-MD is a special case of pfCM-chopMD. So, in the Appendix A.2, we focus on providing an indifferentiable security proof of pfCM-chopMD with any s.

NMAC and HMAC [5]. Let K_1 and K_2 be n bits. $\overline{K} = K||0^{b-n}$. opad is formed by repeating the byte '0x36' as many times as needed to get a b-bit block, and ipad is defined similarly using the byte '0x5c'. Then, NMAC and HMAC are defined as follows, where H is any hash function.

$$\mathrm{NMAC}^H(K_2||K_1, M) = H(K_2, H(K_1, M))$$
$$\mathrm{HMAC}^H_{IV}(K, M) = H(IV, \overline{K} \oplus \mathsf{opad}||H(IV, \overline{K} \oplus \mathsf{ipad}||M))).$$

In this document, we consider the case that H is pfCM^0-MD$^f_{\mathsf{pad}}(\star, \star)$. And it is clear that for any pad, there exists pad_1 such that $\mathrm{NMAC}^{\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathsf{pad}_1}}(K_2||K_1, M)$= $\mathrm{HMAC}^{\mathrm{pfCM}^0-\mathrm{MD}^f_{\mathsf{pad}}}_{IV}(K, M)$, where K_2=$f(IV, \overline{K}\oplus\mathsf{opad})$ and K_1=$f(IV, \overline{K}\oplus\mathsf{ipad})$. *And we assume that in the case of NMAC, the outer hash function uses the compression function one time, and in the case of HMAC, the outer hash function uses the compression function two times.*

Two PRF Constructions based on a pfCM-MD. We propose new two prf constructions as follows.

1. $\mathrm{pfCM^i} - \mathrm{MD}^f_{\mathsf{pad}}(K, \star)$, where $K \xleftarrow{\$} \{0,1\}^n$.
2. $\mathrm{pfCM^i} - \mathrm{MD}^f_{\mathsf{pad}}(IV, K||0^{b-n}||\,\star)$, where $K \xleftarrow{\$} \{0,1\}^n$.

It is clear that for any pad, K, and any M, there exists pad_1 such that $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathsf{pad}_1}\,(K', M)$ $= \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, K||0^{b-n}||M)$, where $K' = f(IV, K||0^{b-n})$.

Inequality. The following inequality will be used to prove Theorem 2.

Ineq 1. For any $0 \le a_i \le 1$, $\prod_{i=1}^q (1 - a_i) \ge 1 - \sum_{i=1}^q a_i$. One can prove it by induction on q.

Random Oracle Model : f is said to be *a random oracle* from X to Y if for each $x \in X$ the value of $f(x)$ is chosen randomly from Y [9]. More precisely, $\Pr[f(x) = y \mid f(x_1) = y_1, f(x_2) = y_2, \ldots f(x_q) = y_q] = \frac{1}{T}$, where $x \notin \{x_1, \ldots, x_q\}$, $y, y_1, \cdots, y_q \in Y$ and $|Y| = T$. In the case that $X = \{0,1\}^d$ for a fixed value d, we say f is a FIL (Fixed Input Length) random oracle. In the case that $X = \{0,1\}^*$, we say f is a VIL (Variable Input Length) random oracle. A VIL random oracle is usually denoted by R.

The cost of Queries. The security bound of a scheme is usually described using the number q of queries and the maximum length l of each queries. On the other hand, in [6], the notion *cost* is used to describe the security bound of sponge construction. The notion *cost* denotes the total block length of q queries. The notion *cost* is significant because the unit of time complexity corresponds to the time of an underlying function call and the total time complexity depends on how many the underlying function is called. The notion *cost* exactly reflects how many the underlying function is called. So, we can consider two cases. The first case is that the number of queries is bounded by q. The second case is that the cost of queries is bounded by q. Without loss of generality, for describing notions and some results in this section, we assume that the number of queries is bounded by q.

Computational Distance. Let $F = (F_1, F_2, \cdots, F_t)$ and $G = (G_1, G_2, \cdots, G_t)$ be tuples of probabilistic oracle algorithms. We define the computational distance of a probabilistic attacker A distinguishing F from G as

$$\mathbf{Adv}_A(F, G) = |\Pr[A^F = 1] - \Pr[A^G = 1]|.$$

Statistical Distance. Let $F = (F_1, F_2, \cdots, F_t)$ and $G = (G_1, G_2, \cdots, G_t)$ be tuples of probabilistic oracle algorithms. We define the statistical distance of a deterministic attacker A distinguishing F from G as

$$\mathbf{Stat}_A(F, G) = \frac{1}{2} \sum_{v \in V_A} |\Pr[F = v] - \Pr[G = v]|,$$

where $\Pr[O = v]$ denotes $\Pr[O(c_i, x_i) = y_i, 1 \leq i \leq q, v = ((c_1, x_1, y_1), \cdots, (c_q, x_q, y_q))]$, where $O(c_i, x_i) = O_{c_i}(x_i)$. And we let the maximum statistical distance of F and G against any deterministic algorithm A be $\mathbf{Stat}(F, G)$, where the number of queries of A is bounded by q.

Computational Distance vs. Statistical Distance

Lemma 1. *Let $F = (F_1, F_2, \cdots, F_t)$ and $G = (G_1, G_2, \cdots, G_t)$ be tuples of probabilistic oracle algorithms. For any probabilistic algorithm A which can make at most q queries*

$$\mathbf{Adv}_A(F, G) \leq \mathbf{Stat}(F, G).$$

Proof. See [14]. ■

Indifferentiability

We give a brief introduction of the indifferentiable security notion.

Definition 1. *Indifferentiability. [22] A Turing machine H with oracle access to an ideal primitive f is said to be $(t_D, t_S, q, \varepsilon)$ indifferentiable from an ideal primitive R if there exists a simulator S such that for any distinguisher D it holds that :*

$$|\Pr[D^{H,f} = 1] - \Pr[D^{R,S} = 1] < \varepsilon$$

The simulator has oracle access to R and runs in time at most t_S. The distinguisher runs in time at most t_D and makes at most q queries. Similarly, H^f is said to be (computationally) indifferentiable from R if ε is a negligible function of the security parameter k (for polynomially bounded by t_D and t_S).

The following Theorem [22] shows the relation between indifferentiable security notion and the security of a cryptosystem.

Theorem 1. *[22] Let $\mathcal{P}$ be a cryptosystem with oracle access to an ideal primitive R. Let H be an algorithm such that H^f is indifferentiable from R. Then cryptosystem $\mathcal{P}$ is at least as secure in the f model with algorithm H as in the R model.*

Above theorem says that if a domain extension (with a padding rule) based on a FIL random oracle f is indifferentiable from a VIL random oracle R, then a cryptosystem, which is proved in the VIL random oracle model, can use the domain extension (with a padding rule) based on a FIL random oracle f instead of R with negligible loss of security.

Definition 2 (prf-advantage). *The prf-advantage of A on $f : \{0,1\}^n \times \{0,1\}^b \to \{0,1\}^n$ is defined by*

$$\boldsymbol{Adv}^{\mathsf{prf}}_{f(K,\star)}(A) = |\Pr[K \xleftarrow{\$} \{0,1\}^n : A^{f(K,\star)} = 1] - \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^b, \{0,1\}^n) : A^{g(\star)} = 1]|.$$

$$\boldsymbol{Adv}^{\mathsf{prf}}_{f(\star,K||0^{b-n})}(A) = |\Pr[K \xleftarrow{\$} \{0,1\}^n : A^{f(\star,K||0^{b-n})} = 1] - \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^n, \{0,1\}^n) : A^{g(\star)} = 1]|,$$

For any function, its prf-advantage can be similarly defined.

Definition 3 (rka-prf-advantage [7]). *Let Φ_1 be a set of functions mapping $\{0,1\}^b$ to $\{0,1\}^b$ and let Φ_2 be a set of functions mapping $\{0,1\}^n$ to $\{0,1\}^n$. Let A be an adversary whose queries have the form (X, ϕ) where $X \in \{0,1\}^n$ and $\phi \in \Phi_1$, or the form (ϕ, X) where $X \in \{0,1\}^b$ and $\phi \in \Phi_2$. For $i = 1$ or 2, the rka-prf-advantage of A in a Φ_i-restricted related-key attack (RKA) on $f : \{0,1\}^n \times \{0,1\}^b \to \{0,1\}^n$ is defined by*

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{rka\text{-}prf}}_{f(\star,RK(\star,K||0^{b-n})),\Phi_1}(A) = |&\Pr[K \xleftarrow{\$} \{0,1\}^n : A^{f(\star,RK(\star,K||0^{b-n}))} = 1] \\ &- \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : A^{g(\star,RK(\star,K||0^{b-n}))} = 1]|,\end{aligned}$$

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{rka\text{-}prf}}_{f(RK(\star,K),\star),\Phi_2}(A) = |&\Pr[K \xleftarrow{\$} \{0,1\}^n : A^{f(RK(\star,K),\star)} = 1] \\ &- \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : A^{g(RK(\star,K),\star)} = 1]|,\end{aligned}$$

where in the first case, on query (X, ϕ) of A, the oracle $O(\star, RK(\star, K||0^{b-n}))$ returns the value of $O(X, \phi(K||0^{b-n}))$ to A, and in the second case, on query (ϕ, X) of A, the oracle $O(RK(\star, K), \star)$ returns the value of $O(\phi(K), X)$ to A.

Definition 4 (multi-rka-prf-advantage). *Let A be an adversary whose queries have the form (i, X, ϕ) where $X \in \{0,1\}^n$ and $\phi \in \Phi_1$, or the form (i, ϕ, X) where $1 \le i \le q$ and $X \in \{0,1\}^b$ and $\phi \in \Phi_2$. For $i = 1$ and 2, the multi-rka-prf-advantage of A in a Φ_i-restricted related-key attack (RKA) on $f : \{0,1\}^n \times \{0,1\}^b \to \{0,1\}^n$ is defined by*

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{multi\text{-}rka\text{-}prf}}_{f(\star,RK(\star,K_\star||0^{b-n})),\Phi_1}(A) = |&\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : A^{f(\star,RK(\star,K_\star||0^{b-n}))} = 1] \\ &- \Pr[g_1, \cdots, g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : A^{g_\star(\star,RK(\star,K||0^{b-n}))} = 1]|,\end{aligned}$$

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{multi\text{-}rka\text{-}prf}}_{f(RK(\star,K_\star),\star),\Phi_2}(A) = |&\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : A^{f(RK(\star,K_\star),\star)} = 1] \\ &- \Pr[g_1, \cdots, g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : A^{g_\star(RK(\star,K),\star)} = 1]|,\end{aligned}$$

where in the first case, on query (i, X, ϕ) of A, $f(\star, RK(\star, K_\star||0^{b-n}))$ returns $f(X, \phi(K_i||0^{b-n}))$ to A, and $g_\star(\star, RK(\star, K||0^{b-n}))$ returns $g_i(X, \phi(K||0^{b-n}))$ to A. The second case is also similarly defined.

Definition 5 (au-advantage [3]). *For any almost universal (au) adversary A, the au-advantage of A on $F(K, \star)$ is defined as follows, where $F : \{0,1\}^n \times \{0,1\}^* \to \{0,1\}^n$.*

$$\boldsymbol{Adv}^{\mathsf{au}}_{F(K,\star)}(A) = \Pr[K \xleftarrow{\$} \{0,1\}^n; (M \neq M') \xleftarrow{\$} A : F(K, M) = F(K, M')].$$

Definition 6 (eTCR-advantage [17]). *For any eTCR-adversary A, the eTCR-advantage of A on a hash family $\boldsymbol{H} = \{H_r(IV, \star)\}_{r\in\mathcal{R}}$ is as follows,*

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{eTCR}}_{\boldsymbol{H}}(A) = \Pr[&(M, \texttt{State}) \xleftarrow{\$} A; r \xleftarrow{\$} \mathcal{R}; (r', M') \xleftarrow{\$} A(r, M, \texttt{State}) \\ &: (r, M) \neq (r', M') \text{ and } H_r(IV, M) = H_{r'}(IV, M')].\end{aligned}$$

Definition 7 (eSPR†-advantage). *Given a hash family $\boldsymbol{H} = \{H_r(IV, \star)\}_{r\in\mathcal{R}}$, for each r we let $H_r(IV, M)[i]$ be the input value of i-th compression function during the computation of $H_r(IV, M)$, that is, $H_r(IV, M)[i] = (c, m)$, where $c \in \{0,1\}^n$, $m \in \{0,1\}^b$, $M \in \{0,1\}^*$, and $H_r : \{IV\} \times \{0,1\}^* \to \{0,1\}^n$ is based on a compression function $f : \{0,1\}^n \times \{0,1\}^b \to \{0,1\}^n$. Then, for any eSPR†-adversary A, the eSPR†[1]-advantage of A on a hash family $\boldsymbol{H}$ is defined as follows,*

$$\begin{aligned}\boldsymbol{Adv}^{\mathsf{eSPR}^\dagger}_{\boldsymbol{H}}(A) = \Pr[&(M, \texttt{State}) \xleftarrow{\$} A; r \xleftarrow{\$} \mathcal{R}; i \xleftarrow{\$} [1, l]; (c', m') \xleftarrow{\$} A(i, r, M, \texttt{State}) \\ &: (c, m) = H_r(IV, M)[i] \text{ and } (c, m) \neq (c', m') \text{ and } f(c, m) = f(c', m')],\end{aligned}$$

where $l = \mathsf{Len}_f(H_r(IV, M))$ is the number of computations of the compression function f when computing $H_r(IV, M)$ for any r, where M is generated by the adversary A.

Relation between a SPR Security of Compression function of H and eSPR† Security of H. In the definition of eSPR†-advantage, the eSPR† security of $\mathbf{H}$ is very similar to the second preimage resistance (SPR) security of f. In the case of SPR security of f, given a random input (c, m), it should be difficult for any adversary to find a different (c', m') such that $f(c, m) = f(c', m')$. Here, (c, m) has $(n + b)$-bit entropy. On the other hand, in the case of eSPR† security of $\mathbf{H}$, given an input (c, m) (which is generated from a random string M and r, where r has $|r|$-bit entropy), it should be difficult for any adversary to find a different (c', m') such that $f(c, m) = f(c', m')$.

3 Indifferentiable Security Analysis of a pfCM-chopMD Domain Extension

The security notion *Indifferentiability* was introduced by Maurer *et al.* in TCC 2004 [22]. Since the concept indifferentiability makes it possible to evaluate the security of domain extensions against all possible generic attackers, under the assumption that the underlying function is a random oracle or an ideal cipher, it is considered one of the significant notions of provable security. In Crypto 2005, Coron *et al.* [15] proved that the classical MD iteration is not indifferentiable with random oracle model even if we assume that the underlying compression function is a random oracle. But they have shown indifferentiability for prefix-free MD hash functions

[1] eSPR† is similar to eSPR defined in [17].

or some other definitions of hash functions like HMAC construction, NMAC construction and chopMD hash function. Since then, several works [8, 12, 23, 18, 13, 6] have been published.

In this section, we provide an indifferentiable security analysis of pfCM^0-chopMD. For any i, the indifferentiable security analysis of pfCM^i-chopMD can be also similarly done. Our proof follows the proof technique in [6, 14].

Construction of the Simulator Here, we define simulators as follows. the simulator S_{pfCM} will be used in order to prove the indifferentiable security of pfCM^0-chopMD. For defining the simulator, we follow the style of construction of the simulator in [13], where $R : \{0,1\}^* \to \{0,1\}^{n-s}$ is a VIL random oracle.

Definition of Simulator S_{pfCM}

INITIALIZATION :

1. A partial function $e : \{0,1\}^{n+b} \to \{0,1\}^n$ initialized as empty,
2. a partial function $e^* = \text{CM-MD}^e : (\{0,1\}^b)^* \to \{0,1\}^n$ initialized as $e^*(\mathsf{null}) = \text{IV}$.
3. a set $I = \{\text{IV}\}$ and a set $U = \{\mathsf{null}\}$.

$\boxed{\text{On query } S^R_{pfCM}(x, m) :}$

001 `if` $(e(x, m) = x')$
    `return` x';

002 `else if` $(\exists\ M'\ \textit{and}\ M, e^*(M') = x \oplus P, ||M'||_b = i, \mathsf{pad}(M) = M'||m))$
    $y = R(M)$;
    `choose` $w \in_R \{0,1\}^s$;
    `define` $e(x, m) = z := y \parallel w$;
    `return` z;

003 `else if` $(\exists\ M', e^*(M') = x \oplus i, ||M'||_b = i)$
    `choose` $z \in_R \{0,1\}^n \setminus \{c \oplus (i+1) : c \in I\} \cup \{c \oplus P : c \in I\} \cup \{a : (i_a, a) \in U\}$
    $\cup \{a \oplus P \oplus (i+1) : (i_a, a) \in U\} \cup \{a \oplus i_a \oplus (i+1) : (i_a, a) \in U\}$
    $\cup \{a \oplus i_a \oplus P : (i_a, a) \in U\}$;
    `define` $e(x, m) = z$;
    `define` $U = U \cup \{(i+1, z)\}$;
    `define` $e^*(M'||m) = z$;
    `return` z;

004 `else`
    $z \in_R \{0,1\}^n$;
    `define` $e(x, m) = z$;
    `define` $I = I \cup \{x\}$;
    `return` z;

Some Important Observations on the Simulator S_{pfCM}

THE BOUND OF THE NUMBER OF QUERIES. In line 003, the number q of queries of S should be bounded by $q < 2^n/6$ in order to choose z. If $q \geq 2^n/6$, the simulator may not work. So, we assume that $q < 2^n/6$.

The bound of the number of possible input message. Firstly, in 002 and 003, there exists at most one M' such that $e^*(M') = x \oplus i$ or $e^*(M') = x \oplus P$ by the process of selecting z unrelated to the set U in line 003. This first observation corresponds to Lemma 1 in [6]. Secondly, in line 002 and 003, by the process of selecting z unrelated to the set I in line 003, the following holds : if $e(x, m)$ is already defined under the assumption that $e^*(M'||m)$ is not defined for all M' previously defined on e^*, where $||M'||_b = i - 1$, then no $M(= M'||m)$ can be newly defined such that $e^*(M) = x \oplus i$ or $e^*(M) = x \oplus P$, where where $||M||_b = i$. This second observation corresponds to the second part of proof of Lemma 2 in [6].

Indifferentiable Security Analysis of pfCM^0-chopMD Hash Domain Extension

We will describe the indifferentiable security bound of each domain extension using the notion *cost* of queries. We let the cost be q. For example, with the cost q of queries, A can have access to O_2 q times and no access to O_1, where O_1 corresponds to a hash function or a VIL random oracle, and O_2 corresponds to a compression function or a FIL random oracle. By observations of simulators described above, the following Lemma holds.

Lemma 2. *Let $q < 2^n/6$. When the total cost of queries to O_1 is t less than or equal to q, the queries to O_1 can be converted to t queries to O_2, where O_2 gives at least the same amount of information to an attacker A and has no higher cost than O_1.*

Proof. The proof is the same as that of Lemma 3 in [6]. ∎

The Lemma 2 says that to give all queries to O_2 and no query to O_1 is the best strategy to obtain better computational distance. That is, when the cost of queries is bound by q, for any A there is an attacker B such that the following holds :

$$\mathbf{Adv}_A((H^f, f), (R, S)) \leq \mathbf{Adv}_B(f, S),$$

where H^f=pfCM^0 − chopMD^f_g, and S=S_{pfCM}. Therefore, we focus on computing the upper bound of the computational distance between f and S as shown in the following theorems.

Theorem 2. *Let $q < (2^n-1)/6$ be the number of queries and $0 \leq s < n$. $f : \{0,1\}^{n+b} \to \{0,1\}^n$ is a FIL random oracle. S_{pfCM} is the simulator defined in the previous section. Then for any (deterministic or probabilistic) algorithm A*

$$\mathbf{Adv}_A(f, S) \leq \tfrac{q(3q-1)}{2^n}.$$

Proof. Let S be S_{pfCM}. By Lemma 1, we only focus on computing an upper bound of $\mathbf{Stat}(f, S)$. Note that $\mathbf{Stat}(f, S)$ is defined over all deterministic algorithms. So when the oracle is f, the number of possible views is 2^{nq}. And for any deterministic algorithm A, each view occurs with probability $1/2^{nq}$. We let the set of 2^{nq} possible views be V_A. On the other hand, when the oracle is S, the number of possible views is at least $(2^n - 2)(2^n - 8) \cdots (2^n - 6q + 4)$. We let the set of the smallest possible views be T_S and the size of T_S be r_q. Assume that each of T_S views occurs with probability $1/r_q$. Therefore,

$$\begin{aligned}
\mathbf{Stat}_A(f, S) &= \tfrac{1}{2} \textstyle\sum_{v \in V_A} |\Pr[f = v] - \Pr[S = v]| \\
&= \tfrac{1}{2} \textstyle\sum_{v \in V_A \setminus T_S} |\Pr[f = v] - \Pr[S = v]| + \tfrac{1}{2} \textstyle\sum_{v \in T_S} |\Pr[f = v] - \Pr[S = v]| \\
&\leq \tfrac{1}{2} \textstyle\sum_{v \in V_A \setminus T_S} |\tfrac{1}{2^{nq}} - 0| + \tfrac{1}{2} \textstyle\sum_{v \in T_S} |\tfrac{1}{2^{nq}} - \tfrac{1}{r_q}| \\
&= \tfrac{1}{2} \cdot \tfrac{2^{nq} - r_q}{2^{nq}} + \tfrac{1}{2} \cdot |\tfrac{r_q}{2^{nq}} - \tfrac{r_q}{r_q}| \\
&= \tfrac{1}{2} \cdot (1 - \tfrac{r_q}{2^{nq}}) + \tfrac{1}{2} \cdot (1 - \tfrac{r_q}{2^{nq}})
\end{aligned}$$

$= 1 - \frac{r_q}{2^{nq}}$
$= 1 - \prod_{i=1}^{q}(1 - \frac{6i-4}{2^n})$
$\leq \sum_{i=1}^{q}(\frac{6i-4}{2^n})$ (by Ineq 1.)
$= \frac{q(3q-1)}{2^n}$. ■

From Lemma 2 and Theorem 2, we can get indifferentiable security bound of pfCM^0-chopMD as the following corollary.

Corollary 1. *Let $q < (2^n - 1)/6$ be the cost of queries and $0 \leq s < n$. $f : \{0,1\}^{n+b} \rightarrow \{0,1\}^n$ is a FIL random oracle. S_{pfCM} is the simulator defined in the previous section. Then for any attacker A*

$$\mathbf{Adv}_A((\mathrm{pfCM}^0 - \mathrm{chopMD}^f_{\mathit{pad}}, f), (R, S_{pfCM})) \leq \frac{q(3q-1)}{2^n}.$$

4 PRF Security Analysis of HMAC based on a pfCM-MD Domain Extension

In this section, with game-based proof technique, we provide a prf security analysis of HMAC based on a pfCM^0-MD domain extension. Our proof follows the proof technique for HMAC by Bellare [3]. For any i, HMAC based on a pfCM^i-MD domain extension can be also proved in the similar way.

Lemma 3. *For any rka-prf-adversary A with q queries, there exists an adversary B_A such that*

$$|\Pr[A^{G_7} = 1] - \Pr[A^{G_6} = 1]| = \boldsymbol{Adv}^{\mathit{rka\text{-}prf}}_{f(\star, RK(\star, K||0^{b-n})), \Phi_1}(B_A),$$

where G_7 and G_6 are games defined in Fig. 1, B_A is defined in Fig. 2. B_A can only make two $(IV, \phi_{\mathit{ipad}})$ and $(IV, \phi_{\mathit{opad}})$ queries. $\Phi_1 = \{\phi_{\mathit{ipad}}, \phi_{\mathit{opad}}\}$ where $\phi_{\mathit{ipad}}(x) = x \oplus \mathit{ipad}$ and $\phi_{\mathit{opad}}(x) = x \oplus \mathit{opad}$.

Proof. Since $\Pr[A^{G_7} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : B_A^{f(\star, RK(\star, K||0^{b-n}))} = 1]$ and $\Pr[A^{G_6} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : B_A^{g(\star, RK(\star, K||0^{b-n}))} = 1]$, this lemma holds. ■

Lemma 4. *For any prf-adversary A, the following equality holds :*

$$\Pr[A^{G_6} = 1] = \Pr[A^{G_5} = 1],$$

where G_6 and G_5 are games defined in Fig. 1.

Proof. We already assumed that in the case of NMAC, the outer hash function uses the compression function one time, and in the case of HMAC, the outer hash function uses the compression function two times. So, this lemma is clear. ■

Lemma 5. *For any prf-adversary A with q queries, there exists a prf-adversary C_A such that*

$$|\Pr[A^{G_5} = 1] - \Pr[A^{G_4} = 1]| = \boldsymbol{Adv}^{\mathit{prf}}_{f(K, \star)}(C_A),$$

where G_5 and G_4 are games defined in Fig. 1, and C_A is defined in Fig. 2. C_A can make at most q queries.

Game G_1

On query M
    $Z \xleftarrow{\$} \{0,1\}^n$
    Return Z

Game G_2

$K_1 \xleftarrow{\$} \{0,1\}^n$; $s \leftarrow 0$
$Z_1, \cdots, Z_q \xleftarrow{\$} \{0,1\}^n$
On query M
    $s \leftarrow s+1$; $M_s \leftarrow M$
    $Y_s \leftarrow \text{pfCM}^1\text{-MD}^f_{\mathsf{pad}_1}(K_1, M_s)$
    If $(\exists\, r < s : Y_r = Y_s)$ then
        $\mathsf{bad} \leftarrow \mathsf{true}$;
    Return Z_s

Game G_3

$K_1 \xleftarrow{\$} \{0,1\}^n$; $s \leftarrow 0$
$Z_1, \cdots, Z_q \xleftarrow{\$} \{0,1\}^n$
On query M
    $s \leftarrow s+1$; $M_s \leftarrow M$
    $Y_s \leftarrow \text{pfCM}^1\text{-MD}^f_{\mathsf{pad}_1}(K_1, M_s)$
    If $(\exists\, r < s : Y_r = Y_s)$ then
        $\mathsf{bad} \leftarrow \mathsf{true}$; $Z_s \leftarrow Z_r$
    Return Z_s

Game G_4

$K_1 \xleftarrow{\$} \{0,1\}^n$
$g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^b, \{0,1\}^n)$
On query M
    Return $g(\mathsf{pad}_1(\text{pfCM}^1 - \text{MD}^f_{\mathsf{pad}_1}(K_1, M_s)))$

Game G_5

$K_2, K_1 \xleftarrow{\$} \{0,1\}^n$
On query M
    Return $f((K_2 \oplus P), \mathsf{pad}_1(\text{pfCM}^1 - \text{MD}^f_{\mathsf{pad}_1}(K_1, M_s)))$

Game G_6

$K_2, K_1 \xleftarrow{\$} \{0,1\}^n$
On query M
    Return $\text{NMAC}^{\text{pfCM}^1 - \text{MD}^f_{\mathsf{pad}_1}}(K_2 || K_1, M)$

Game G_7

$K \xleftarrow{\$} \{0,1\}^n$
$\overline{K} \leftarrow K || 0^{b-n}$
$K_1 \leftarrow f(IV, \overline{K} \oplus \mathsf{ipad})$
$K_2 \leftarrow f(IV, \overline{K} \oplus \mathsf{opad})$
On query M
    Return $\text{NMAC}^{\text{pfCM}^1 - \text{MD}^f_{\mathsf{pad}_1}}(K_2 || K_1, M)$

Fig. 1. Game $G_1 \sim G_7$

Adversary $B_A^{O(\star, RK(\star, K||0^{b-n}))}$, where O is $f(\star, K||0^{b-n})$ or $g(\star, K||0^{b-n})$.

100 $K_1 \leftarrow O(IV, RK(\phi_{\mathsf{ipad}}, K||0^{b-n}))$
200 $K_2 \leftarrow O(IV, RK(\phi_{\mathsf{opad}}, K||0^{b-n}))$
300 Run A as follows:
301 On query M of A, reply $\mathrm{NMAC}^{\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathsf{pad}_1}(\star,\star)}(K_2||K_1, M)$ to A
302 Let T be the final output of A
400 Return T

Adversary C_A^O, where O is $f(K, \star)$ or $g(\star)$.

100 $K_1 \xleftarrow{\$} \{0,1\}^n$
200 Run A as follows:
201 On query M of A, reply $O(\mathsf{pad}_1(\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathsf{pad}_1}(K_1, M))))$ to A
202 Let T be the final output of A
300 Return T

Adversary D_A

100 $s \leftarrow 0$ and $Z_1, \cdots, Z_q \xleftarrow{\$} \{0,1\}^n$
200 Run A as follows:
201 On query M of A, $s \leftarrow s+1$ and $M_s \leftarrow M$ and reply Z_s to A
300 $i, j \xleftarrow{\$} [1, q]$ with $i \neq j$
400 Return M_i and M_j

Fig. 2. Adversary B_A, C_A, D_A

Adversary $E_A^{O(RK(\star,K),\star)}$, where O is $f(K, \star)$ or $g(K, \star)$.

100 Run A, and obtain M, M' from A, and let $m = ||\mathsf{pad}_1(M)||_b$, $m' = ||\mathsf{pad}_1(M')||_b$
200 Let $\mathsf{pad}_1(M) = M_1|| \cdots ||M_m$ and $\mathsf{pad}_1(M') = M'_1|| \cdots ||M'_{m'}$ and $r = LCP(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$
/* r is the b-bit block length of the largest common prefix of $\mathsf{pad}_1(M)$ and $\mathsf{pad}_1(M')$ */
300 Randomly choose (l, l') from $I(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$
/* total number of cases is at most $m + m' - 1$.
$I(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$ is a sequence of $(1,1)|| \cdots ||(r,r)||(r+1, r+1)$
$||(r+2, r+1)|| \cdots ||(m, r+1)||(m, r+2)|| \cdots ||(m, m')$. */
400 If $(l, l') \in I_1(\mathsf{pad}_1(M), \mathsf{pad}_1(M')) \cup \{(r+1, r+1)\} \cup I_2(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$
/* $I_1(\mathsf{pad}_1(M), \mathsf{pad}_1(M')) = \{(1,1), \cdots, (r,r)\}$ and $I_2 = \{(r+2, r+1), \cdots, (m, r+1)\}$ */
401 then if $l = m$ then $a_l \leftarrow O(\phi_P, M_l)$ else $a_l \leftarrow O(\phi_l, M_l)$
402 else $a_l \xleftarrow{\$} \{0,1\}^n$
500 If $(l, l') \in I_1(\mathsf{pad}_1(M), \mathsf{pad}_1(M')) \cup \{(r+1, r+1)\} \cup I_3(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$
/* $I_3 = \{(m, r+2), \cdots, (m, m')\}$ */
501 then if $l' = m'$ then $a'_{l'} \leftarrow O(\phi_P, M'_{l'})$ else $a'_{l'} \leftarrow O(\phi_{l'}, M'_{l'})$
502 else $a'_{l'} \xleftarrow{\$} \{0,1\}^n$
600 For $i = l+1$ to m do
601 if $i < m$ then $a_i \leftarrow f(a_{i-1} \oplus i, M_i)$
602 if $i = m$ then $a_i \leftarrow f(a_{i-1} \oplus P, M_i)$
700 For $i = l'+1$ to m' do
701 if $i < m'$ then $a'_i \leftarrow f(a'_{i-1} \oplus i, M'_i)$
702 if $i = m'$ then $a'_i \leftarrow f(a'_{i-1} \oplus P, M'_i)$
800 If $a_m = a'_{m'}$ then return 1 else return 0.

Fig. 3. Adversary E_A: P is the last counter value of pfCM[1]-MD.

Proof. Since $\Pr[A^{G_5} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : C_A^{f(K,\star)} = 1]$ and $\Pr[A^{G_4} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^b, \{0,1\}^n) : C_A^{g(\star)} = 1]$, this lemma holds. ∎

Lemma 6. *For any prf-adversary A with q queries, the following equality holds :*

$$\Pr[A^{G_4} = 1] = \Pr[A^{G_3} = 1],$$

where G_4 and G_3 are games defined in Fig. 1.

Proof. By the definitions of G_3 and G_4, it is clear. ∎

Lemma 7. *For any prf-adversary A with q queries, the following inequality holds :*

$$|\Pr[A^{G_3} = 1] - \Pr[A^{G_2} = 1]| \leq \Pr[A^{G_2} \text{ sets } \textsf{bad}],$$

where G_3 and G_2 are games defined in Fig. 1.

Proof. As described in [10], this lemma follows from the Fundamental Lemma of Game Playing. ∎

Lemma 8. *For any prf-adversary A with q queries, the following equality holds :*

$$\Pr[A^{G_2} = 1] = \Pr[A^{G_1} = 1],$$

where G_2 and G_1 are games defined in Fig. 1.

Proof. By the definitions of G_1 and G_2, it is clear. ∎

Lemma 9. *For any prf-adversary A with q queries, there exists an au-adversary D_A such that*

$$\Pr[A^{G_2} \text{ sets } \textsf{bad}] \leq \frac{q(q-1)}{2} \boldsymbol{Adv}^{\textsf{au}}_{\mathrm{pfCM}^1 - \mathrm{MD}^f_{\textsf{pad}_1}(K,\star)}(D_A),$$

where G_2 is a game defined in Fig. 1, and D_A is defined in Fig. 2.

Proof. We let $F(K,\star)$ be $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\textsf{pad}_1}(K,\star)$. Without loss of generality, we assume that A makes q different queries.

$$\begin{aligned}
&\mathbf{Adv}^{\textsf{au}}_{F(K,\star)}(D_A) \\
&\quad = \textstyle\sum_{i<j} \Pr[K \xleftarrow{\$} \{0,1\}^n; M_1, \cdots, M_q \xleftarrow{\$} A^D : F(K, M_i) = F(K, M_j)] \Pr[M_i, M_j \xleftarrow{\$} D_A] \\
&\quad = \textstyle\sum_{i<j} \Pr[K \xleftarrow{\$} \{0,1\}^n; M_1, \cdots, M_q \xleftarrow{\$} A^{G_2} : F(K, M_i) = F(K, M_j)] \frac{2}{q(q-1)} \\
&\quad \geq \Pr[K \xleftarrow{\$} \{0,1\}^n; M_1, \cdots, M_q \xleftarrow{\$} A^{G_2} : \exists M_i, M_j \text{ s.t. } F(K, M_i) = F(K, M_j)] \frac{2}{q(q-1)} \\
&\quad = \Pr[A^{G_2} \text{ sets } \textsf{bad}] \frac{2}{q(q-1)}.
\end{aligned}$$

∎

Lemma 10. *For given M and M', where $||\mathsf{pad}_1(M)||_b = m \leq t$ and $||\mathsf{pad}_1(M')||_b = m' \leq t'$, if (α', β') is the predecessor of (α, β) in the sequence of $I(\mathsf{pad}_1(M), \mathsf{pad}_1(M'))$, then the following holds.*

$$\begin{aligned}&\Pr[K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{f(RK(\star,K),\star)} = 1 | (l,l') = (\alpha,\beta) \leftarrow E_{A(M,M')}^{f(RK(\star,K),\star)}]\\ =&\Pr[g \xleftarrow{\$} \mathit{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{g(RK(\star,K),\star)} = 1 | (l,l') = (\alpha',\beta') \leftarrow E_{A(M,M')}^{g(RK(\star,K),\star)}],\end{aligned}$$

Here, $E_A^{O(RK(\star,K),\star)}$, I_1, I_2, I_3 and I are defined in Fig. 3. In a sequence $((\alpha_1, \beta_1), \cdots, (\alpha_n, \beta_n))$, (α_i, β_i) is called the predecessor of $(\alpha_{i+1}, \beta_{i+1})$. For example, in the sequence I, the predecessor of $(r+2, r+1)$ is $(r+1, r+1)$ and the predecessor of $(m, r+2)$ is $(m, r+1)$.

Proof. It follows from the definition of E_A in Fig. 3. ■

Lemma 11. *For any au-adversary A, the following holds.*

$$\begin{aligned}&\Pr[K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{f(RK(\star,K),\star)} = 1 | (l,l') = (1,1) \leftarrow E_{A(M,M')}^{f(RK(\star,K),\star)}]\\ &\qquad = \Pr[K \xleftarrow{\$} \{0,1\}^n : F(K,M) = F(K,M')],\\ &\Pr[g \xleftarrow{\$} \mathit{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{g(RK(\star,K),\star)} = 1 | (l,l') = (m,m') \leftarrow E_{A(M,M')}^{g(RK(\star,K),\star)}]\\ &\qquad = 2^{-n},\end{aligned}$$

where $F(K,\star)$ denotes $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathsf{pad}_1}(K,\star)$.

Proof. It is clear by the construction of E_A in Fig. 3. ■

Lemma 12. *For any au-adversary A, there exists a rka-prf-adversary E_A such that*

$$\boldsymbol{Adv}^{au}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathsf{pad}_1}(K,\star)}(A) \leq (t + t' - 1)\boldsymbol{Adv}^{\mathsf{rka\text{-}prf}}_{f(K,\star),\Phi_2}(E_A) + 2^{-n},$$

where E_A is defined in Fig. 3.For any output (M, M') of A, $||\mathsf{pad}_1(M)||_b \leq t$ and $||\mathsf{pad}_1(M')||_b \leq t'$. When $t^* = max(t, t')$, $\Phi_2 = \{\phi_1, \cdots, \phi_{t^*}, \phi_P\}$ where $\phi_i(x) = x \oplus i$. E_A can only make at most two (M_i, ϕ) and (M'_j, ϕ') queries, where M_i and M'_j are any value of b-bit, and $\phi, \phi' \in \Phi_2$.

Proof. We let $F(K,\star)$ be $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathsf{pad}_1}(K,\star)$.

$$\begin{aligned}&\mathbf{Adv}^{\mathsf{rka\text{-}prf}}_{f(RK(\star,K),\star),\Phi_2}(E_A)\\ &\quad= |\Pr[K \xleftarrow{\$} \{0,1\}^n : E_A^{f(RK(\star,K),\star)} = 1]\\ &\qquad - \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : E_A^{g(RK(\star,K),\star)} = 1]|\\ &\quad= |\textstyle\sum_{M \neq M'} \Pr[K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{f(RK(\star,K),\star)} = 1]\Pr[(M,M') \leftarrow A]\\ &\quad-\textstyle\sum_{M \neq M'} \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : E_{A(M,M')}^{g(RK(\star,K),\star)} = 1]\Pr[(M,M') \leftarrow A]|\\ &\quad\geq |\textstyle\sum_{M \neq M'} \frac{\Pr[K \xleftarrow{\$} \{0,1\}^n : F(K,M) = F(K,M')] - 2^{-n}}{t+t'-1}\Pr[M,M' \leftarrow A]| \text{ by Lemma 10, 11}\\ &\quad= |\tfrac{1}{t+t'-1}[(\textstyle\sum_{M \neq M'} \Pr[K \xleftarrow{\$} \{0,1\}^n : F(K,M) = F(K,M')]\Pr[M,M' \leftarrow A]) - 2^{-n}]|\\ &\quad= |\tfrac{1}{t+t'-1}(\mathbf{Adv}^{\mathsf{au}}_{F(K,\star)}(A) - 2^{-n})|\\ &\quad\geq \tfrac{1}{t+t'-1}(\mathbf{Adv}^{\mathsf{au}}_{F(K,\star)}(A) - 2^{-n}).\end{aligned}$$

■

Theorem 3. *For any prf-adversary A, there exist adversaries B_A, C_A, D_A, E_{D_A} such that*

$$\boldsymbol{Adv}^{\textit{prf}}_{\mathrm{HMAC}_{IV}^{\mathrm{pfCM}^0-\mathrm{MD}^f_{\textsf{pad}}}}(A) \le \boldsymbol{Adv}^{\textit{rka-prf}}_{f(\star, RK(\star, K||0^{b-n})),\Phi_1}(B_A) + \boldsymbol{Adv}^{\textit{prf}}_{f(K,\star)}(C_A) + \frac{q(q-1)(t+t'-1)}{2}\boldsymbol{Adv}^{\textit{rka-prf}}_{f(RK(\star,K),\star),\Phi_2}(E_{D_A}) + \frac{q(q-1)}{2^{n+1}},$$

where B_A, C_A, D_A, E_{D_A}, Φ_1, and Φ_2 are defined as before.

Proof. By the definition of the prf-advantage, $\mathbf{Adv}^{\mathsf{prf}}_{\mathrm{HMAC}_{IV}^{\mathrm{pfCM-MD}^f_{\mathsf{pad}}}}(A) = |\Pr[A^{G_7} = 1] - \Pr[A^{G_1} = 1]|$. So, we can get the above theorem with Lemma 3 $\sim$ Lemma 12. ∎

5 Security Analysis of Two PRF Constructions based on a pfCM-MD Domain Extension

In this section, we provide prf security analysis of $\mathrm{pfCM}^0\text{-}\mathrm{MD}^f_{\mathsf{pad}}(IV, K||0^{b-n}||\star)$ and $\mathrm{pfCM}^1\text{-}\mathrm{MD}^f_{\mathsf{pad}}(K,\star)$, where $K \xleftarrow{\$} \{0,1\}^n$. Our analysis follows the analysis technique of Bellare *et al.*' paper [4]. For any d and d', $\mathrm{pfCM}^d\text{-}\mathrm{MD}^f_{\mathsf{pad}}(IV, K||0^{b-n}||\star)$ and $\mathrm{pfCM}^{d'}\text{-}\mathrm{MD}^f_{\mathsf{pad}}(K,\star)$ can be also proved in the similar way.

Lemma 13. *For any prf-adversary A with q queries, there exists a prf-adversary F_A such that*

$$|\Pr[A^{G'_3} = 1] - \Pr[A^{G'_2} = 1]| = \boldsymbol{Adv}^{\textit{prf}}_{f(\star, K||0^{b-n})}(F_A)$$

where G'_3 and G'_2 are games defined in Fig. 4, and F_A is defined in Fig. 5. F_A can only make the query IV.

Proof. Since $\Pr[A^{G'_3} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : F_A^{f(\star,K||0^{b-n})} = 1]$ and $\Pr[A^{G'_2} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^n, \{0,1\}^n) : F_A^{g(\star)} = 1]$, the lemma holds. ∎

Lemma 14. *For any prf-adversary A, the following equality holds :*

$$|\Pr[A^{G'_2} = 1] - \Pr[A^{G'_1} = 1]| = \boldsymbol{Adv}^{\textit{prf}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{\textsf{pad}_1}(K,\star)}(A)$$

where G'_2 and G'_1 are games defined in Fig. 4.

Proof. By the definition of the prf-advantage, the lemma holds. ∎

Lemma 15. *For any $2 \le j \le l$, the following holds.*

$$\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : H_{A,i\leftarrow j}^{f(RK(\star,K_\star),\star)} = 1]$$
$$= \Pr[g_1, \cdots, g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : H_{A,i\leftarrow j-1}^{g_\star(RK(\star,K),\star)} = 1],$$

where H_A is defined in Fig. 5, and $i \leftarrow j$ is described in line 10000 in Fig. 5. If A makes q queries, then H_A can make at most q queries. We assume that for each query M of A, the b-bit block length of $\mathsf{pad}_1(M)$ is at most l. $\Phi_3 = \{\phi_1, \cdots, \phi_l, \phi_P\}$, where $\phi_i(X) = X \oplus i$ and P

is the last counter of pfCM-MD. When we denote t-th query of H_A by (i^t, ϕ^t, X^t), we assume that $\{\phi^1, \cdots, \phi^q\} \subset \{\phi_P, \phi_j\}$ for some j. In other words, even though H_A can make queries to any one of $\{O_1, O_2, \cdots, O_q\}$, H_A can use at most two related-key-deriving (RKD) functions ϕ's from Φ_3.

Proof. It follows from the definition of $H_A^{O_1,\cdots,O_q}$ in Fig. 5. ■

Game G_1'	Game G_2'
100 On query M 101 $\quad Z \xleftarrow{\$} \{0,1\}^n$ 102 $\quad$ Return Z	100 $K' \xleftarrow{\$} \{0,1\}^n$ 200 On query M 201 $\quad$ Return $\mathrm{pfCM}^1\text{-}\mathrm{MD}^f_{\mathsf{pad}_1}(K', M)$

Game G_3'
100 $K \xleftarrow{\$} \{0,1\}^n$ 200 $K' \leftarrow f(IV, K\|\|0^{b-n})$ 300 On query M 301 $\quad$ Return $\mathrm{pfCM}^1\text{-}\mathrm{MD}^f_{\mathsf{pad}_1}(K', M)$

Fig. 4. Game $G_1' \sim G_3'$

Lemma 16. *For any prf-adversary A with q queries, the following holds.*

$$\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : H_{A,i\leftarrow 1}^{f(RK(\star,K_\star),\star)} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : A^{F(K,\star)} = 1],$$
$$\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : H_{A,i\leftarrow l}^{f(RK(\star,K_\star),\star)} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^*, \{0,1\}^n) : A^{g(\star)} = 1],$$

where $F(K,\star)$ denotes $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathit{pad}_1}(K,\star)$.

Proof. It is clear by the construction of H_A in Fig. 5. ■

Theorem 4. *For any prf-adversary A with q queries, there exists a multi-rka-prf-adversary H_A such that*

$$\boldsymbol{Adv}^{\mathit{prf}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathit{pad}_1}(K,\star)}(A) = l \cdot \boldsymbol{Adv}^{\mathit{multi\text{-}rka\text{-}prf}}_{f(RK(\star,K_\star),\star),\Phi_3}(H_A),$$

where H_A is defined as before.

Proof. We let $F(K,\star)$ be $\mathrm{pfCM}^1 - \mathrm{MD}^f_{\mathsf{pad}_1}(K,\star)$.

$$\begin{aligned}
&\mathbf{Adv}^{\mathsf{multi\text{-}rka\text{-}prf}}_{f(RK(\star,K_\star),\star),\Phi_3}(H_A)\\
&\quad = |\Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : H_A^{f(RK(\star,K_\star),\star)} = 1]\\
&\qquad - \Pr[g_1, \cdots, g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b}, \{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : H_A^{g_\star(RK(\star,K),\star)} = 1]|\\
&\quad = |\textstyle\sum_{j=1}^{l} \Pr[K_1, \cdots, K_q \xleftarrow{\$} \{0,1\}^n : H_{A,i=j}^{f(RK(\star,K_\star),\star)} = 1] \cdot \frac{1}{l}
\end{aligned}$$

Adversary $F_A^{O(\star,K||0^{b-n})}$, where O is $f(\star,K||0^{b-n})$ or $g(\star)$.

100 $K' \leftarrow O(IV)$
200 Run A as follows:
201   On query M of A, reply $\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathsf{pad}_1}(K',M)$ to A
202   Let T be the final output of A
300 Return T

Adversary $H_A^{O_1,\cdots,O_q}$, where O_i is $f(RK(\star,K_i),\star)$ or $g_i(RK(\star,K),\star)$.

10000 Randomly choose j from $[1,l]$ and $i \leftarrow j$ and $s \leftarrow 0$
20000 Run A as follows:
21000   On query t-th query M^t of A,   // $1 \le t \le q$
21100 $m \leftarrow ||\mathsf{pad}_1(M^t)||_b$ and let $\mathsf{pad}_1(M^t)=M_1^t||\cdots||M_m^t$ // $||\mathsf{pad}_1(M^t)||_b \le l$
21200   if $m \le i-1$ then pick at random an n-bit string a^t and return a^t to A
21300   else (namely $m \ge i$),
21310   if $(M_1^t,\cdots,M_{i-1}^t)\neq(M_1^r,\cdots,M_{i-1}^r)$ for all $r<t$
21320   then $s \leftarrow s+1$ and let $c^t=s$
21330   else if $(M_1^t,\cdots,M_{i-1}^t)=(M_1^r,\cdots,M_{i-1}^r)$ & $||\mathsf{pad}_1(M^r)||_b \neq i-1$ for a r s.t. $r<t$
21331   then let $c^t=c^r$
21332   else $s \leftarrow s+1$ and let $c^t=s$
21340   if $m>i$ then $a^t=O_{c^t}(\phi_i,M_i^t)$ else $a^t=O_{c^t}(\phi_P,M_i^t)$
21350   return $\mathrm{pfCM}^{\mathrm{i+1}}-\mathrm{MD}^f(a^t,M_{i+1}^t||\cdots||M_m^t)$ to A
30000 Let T be the final output of A
40000 Return T

Fig. 5. Adversary F_A and H_A: P is the last counter value of pfCM-MD.

$$- \sum_{j=1}^{l} \Pr[g_1,\cdots,g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : H_{A,i=j}^{g_\star(RK(\star,K),\star)} = 1] \cdot \frac{1}{l}|$$
$$= \frac{1}{l}|\Pr[K \xleftarrow{\$} \{0,1\}^n : A^{F(K,\star)} = 1] - \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^*,\{0,1\}^n) : A^{g(\star)} = 1]|$$
$$= \frac{1}{l}\mathbf{Adv}^{\mathsf{prf}}_{F(K,\star)}(A). \qquad \blacksquare$$

The second equality follows from the definition of H_A in Fig. 5 and the third equality follows from Lemma 15 and Lemma 19.

Theorem 5. *For any prf-adversary A with q queries, there exists a prf-adversary F_A such that*

$$\boldsymbol{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{pad}(IV,K||\star)}(A) \leq \boldsymbol{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{pad_1}(K,\star)}(A) + \boldsymbol{Adv}^{\mathsf{prf}}_{f(\star,K||0^{b-n})}(F_A),$$

where F_A can only make the query IV and is defined in Fig. 5.

Proof. By the definition of the prf-advantage, $\mathbf{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{\mathsf{pad}}(IV,K||\star)}(A) = |\Pr[A^{G'_3} = 1] - \Pr[A^{G'_1} = 1]|$. So, we can get above theorem with Lemma 13 $\sim$ Lemma 14. $\blacksquare$

Corollary 2. *For any prf-adversary A with q queries, there exist adversaries F_A and H_A such that*

$$\boldsymbol{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{pad}(IV,K||\star)}(A) \leq l \cdot \boldsymbol{Adv}^{\mathsf{multi\text{-}rka\text{-}prf}}_{f(RK(\star,K_\star),\star),\Phi_3}(H_A) + \boldsymbol{Adv}^{\mathsf{prf}}_{f(\star,K||0^{b-n})}(F_A),$$

where F_A, H_A and Φ_3 are defined as before.

Proof. This holds by Theorem 4 and 5. $\blacksquare$

6 PRF Security Analysis of NIST SP 800-56A Key Derivation Function based on a pfCM-MD Domain Extension

NIST special publication 800-56A [25] describes key derivation functions (KDF) based on a hash function. Any key derivation function is used to derive secret keying material from a shared secret. Secret keying material means a symmetric key, a secret initialization vector, or a master key which is used to generate other keys. The process of KDF in the document is as follows: (See the page 49 of NIST SP 800-56A for details.)

1. $reps = \lceil keydatalen/hashlen \rceil$.
2. If $reps > (2^{32} - 1)$, then ABORT : output an error indicator and stop.
3. Initialize a 32-bit, big-endian bit string $counter$ as 00000001_{16}.
4. If $counter||Z||OtherInfo$ is more than $max\_hash\_inputlen$ bits long, then ABORT : output an error indicator and stop.
5. For $i = 1$ to $reps$ by 1, do the followings:
   (a) Compute $Hash_i$=H($counter||Z||OtherInfo$).
   (b) Increment $counter$ (modulo 2^{32}), treating it as an unsigned 32-bit integer.
6. Let $Hhash$ be set to $Hash_{reps}$ if ($keydatalen/hashlen$) is an integer, otherwise, let $Hhash$ be set to the ($keydatalen$ mod $hashlen$) leftmost bits of $Hash_{reps}$.
7. Set $DerivedKeyingMaterial = Hash_1||Hash_2||\cdots||Hash_{reps-1}||Hhash$.

In the above process, H is a hash function, Z is a shared secret, and $OtherInfo$ is known fixed value. $Counter$ is a changeable input variable. Then, the concatenation of hash outputs is used as secret keying material. In this section, it is shown that the pseudorandomness of KDF-pfCM-MD is reduced to the RKA-pseudorandomness and pseudorandomness of the compression function f. More precisely, we provide prf security analysis of pfCM^0-MD$^f_{\mathsf{pad}}(IV,\ \star_{32}\,||K||\,\star)$, where $\star_{32}$ is any 32-bit string, and $K \xleftarrow{\$} \{0,1\}^n$. pfCM^0-MD$^f_{\mathsf{pad}}(IV,\ \star_{32}\,||K||\,\star)$ corresponds to NIST SP 800-56A key derivation function based on pfCM^0-MD. Our analysis follows the analysis technique of Bellare *et al.*' paper [4]. For any i, the prf security of pfCM^i-MD$^f_{\mathsf{pad}}(IV,\ \star_{32}\,||K||\,\star)$ can be also proved in the similar way.

Lemma 17. *For any prf-adversary A with q queries, there exists a prf-adversary Q_A such that*

$$|\Pr[A^{G_3''} = 1] - \Pr[A^{G_2''} = 1]| = \boldsymbol{Adv}^{\mathsf{prf}}_{f(\star,\star_{32}||K||\star_{b-n-32})}(Q_A)$$

where G_3'' and G_2'' are games defined in Fig. 6, and Q_A is defined in Fig. 7. And Q_A can make q queries of the form $(IV||\,\star_{32}\,||\star_{b-n-32})$, and $\star_i$ means any i-bit string.

Proof. Since $\Pr[A^{G_3''} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : Q_A^{f(\star_n,\star_{32}||K||\star_{b-n-32})} = 1]$ and $\Pr[A^{G_2'} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^n, \{0,1\}^n) : Q_A^{g(\star_n||\star_{32}||\star_{b-n-32})} = 1]$, the lemma holds. ∎

Lemma 18. *For any prf-adversary A, the following equality holds :*

$$|\Pr[A^{G_2''} = 1] - \Pr[A^{G_1''} = 1]| = \boldsymbol{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{pad_1}(g(IV,\star_{32}||K||\star_{b-n-32}),\ \star)}(A)$$

where G_2'' and G_1'' are games defined in Fig. 6, and $g \xleftarrow{\$}$ $\mathsf{Maps}(\{0,1\}^b, \{0,1\}^n)$.

Proof. By the definition of the prf-advantage, the lemma holds. ∎

Lemma 19. *For any $2 \le j \le l-1$, the following holds.*

$$\begin{aligned}&\Pr[K_1,\cdots,K_q \xleftarrow{\$} \{0,1\}^n : V_{A,i=j}^{f(RK(\star,K_\star),\star)} = 1]\\ &=\Pr[g_1,\cdots,g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : V_{A,i=j-1}^{g_\star(RK(\star,K),\star)} = 1],\end{aligned}$$

where V_A is defined in Fig. 7.

Proof. It is clear by the definition of V_A. ∎

Lemma 20. *For any prf-adversary A of q queries, the following holds.*

$$\begin{aligned}&\Pr[K_1,\cdots,K_q \xleftarrow{\$} \{0,1\}^n : V_{A,i=1}^{f(RK(\star,K_\star),\star)} = 1] = \Pr[K \xleftarrow{\$} \{0,1\}^n : A^{F(K,\star)} = 1],\\ &\Pr[K_1,\cdots,K_q \xleftarrow{\$} \{0,1\}^n : V_{A,i=l-1}^{f(RK(\star,K_\star),\star)} = 1] = \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^*,\{0,1\}^n) : A^{g(\star)} = 1],\end{aligned}$$

where $F(K,\star)$ denotes $\mathrm{pfCM}^1 - \mathrm{MD}^f_{pad_1}(g(IV,\star_{32}||K||\star_{b-n-32}),\ \star)$.

Proof. It is clear by the construction of V_A in Fig. 7. ∎

Game G_1''

100 On query M
101 $\quad Z \xleftarrow{\$} \{0,1\}^n$
102 $\quad$ Return Z

Game G_2''

100 $g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^b, \{0,1\}^n)$
300 On query $M = M_1||M_2$ // $|M_1| = 32$ and $|M_2| = t$ where t is any value.
200 $\quad K' \leftarrow g(IV, M_1||M_2[1, b-n-32])$ // $M_2[1, x]$ denotes the first x-bit of M_2
301 $\quad$ Return $\text{pfCM}^1\text{-MD}^f_{\mathsf{pad}_1}(K', M[b-n-31, t])$

Game G_3''

100 $K \xleftarrow{\$} \{0,1\}^n$
300 On query $M = M_1||M_2$ // $|M_1| = 32$ and $|M_2| = t$ where t is any value.
200 $\quad K' \leftarrow f(IV, M_1||K||M_2[1, b-n-32])$ // $M_2[1, x]$ denotes the first x-bit of M_2
301 $\quad$ Return $\text{pfCM}^1\text{-MD}^f_{\mathsf{pad}_1}(K', M[b-n-31, t])$

Fig. 6. Game $G_1'' \sim G_3''$

Adversary $Q_A^{O(\star_n, \star_{32}||\star_{b-n-32})}$, where O is $f(\star_n, \star_{32}||K||\star_{b-n-32})$ or $g(\star_n|| \star_{32} ||\star_{b-n-32})$.

100 Run A as follows:
200 $\quad$ On query M of A, $K' \leftarrow O(IV, M[1, b-n])$ // $|M| = t$
201 $\quad$ Reply $\text{pfCM}^1 - \text{MD}^f_{\mathsf{pad}_1}(K', M[b-n, t])$ to A
202 $\quad$ Let T be the final output of A
300 Return T

Adversary $V_A^{O_1, \cdots, O_q}$, where O_i is $f(RK(\star, K_i)||\star)$ or $g_i(RK(\star, K)||\star)$.

100000 Randomly choose j from $[1, l-1]$ and $i \leftarrow j$ and $s \leftarrow 0$
200000 Run A as follows:
210000 $\quad$ On query t-th query M^t of A, // $1 \le t \le q$
211000 $\quad\quad$ Let $\mathsf{pad}(M^t) = M_1^t|| \cdots ||M_{m_t}^t$ where $|M_1^t| = b-n$, $|M_j^t| = b$ for $2 \le j \le m_t$ // $m_t \le l$
212000 $\quad\quad$ if $m_t \le i-1$ then $a^t \xleftarrow{\$} \{0,1\}^n$ and return a^t to A
213000 $\quad\quad$ else (namely $m_t \ge i$),
213100 $\quad\quad\quad$ if $(M_1^t, \cdots, M_i^t) \neq (M_1^r, \cdots, M_i^r)$ for all $r < t$
213200 $\quad\quad\quad$ then $s \leftarrow s+1$ and let $c^t = s$ and $a^t \xleftarrow{\$} \{0,1\}^n$
213300 $\quad\quad\quad$ else if $(M_1^t, \cdots, M_i^t) = (M_1^r, \cdots, M_i^r)$ & $((m_t = i$ & $m_r = i)$ or $(m_t \neq i$ & $m_r \neq i))$ for some r with $r < t$
213310 $\quad\quad\quad\quad$ then let $c^t \leftarrow c^r$ and $a^t \leftarrow a^r$
213320 $\quad\quad\quad\quad$ else if $(M_1^t, \cdots, M_i^t) = (M_1^r, \cdots, M_i^r)$ & $(m_t = i$ or $m_r = i)$ for some r with $r < t$
213321 $\quad\quad\quad\quad\quad$ then $s \leftarrow s+1$ and let $c^t = s$ and $a^t \xleftarrow{\$} \{0,1\}^n$
213400 $\quad\quad\quad$ if $m^t > i+1$ then $a^t \leftarrow O_{c^t}(\phi_i, M_{i+1}^t)$
213500 $\quad\quad\quad$ if $m^t = i+1$ then $a^t \leftarrow O_{c^t}(\phi_P, M_{i+1}^t)$
213600 $\quad\quad\quad$ return $\text{pfCM}^{i+1} - \text{MD}^f(a^t, M_{i+2}^t|| \cdots ||M_{m_t}^t)$ to A // $\text{pfCM}^{i+1} - \text{MD}^f(a^t, \mathsf{null}) = a^t$
300000 Let T be the final output of A
400000 Return T

Fig. 7. Adversary Q_A and V_A: P is the last counter value of pfCM-MD.

Theorem 6. *For any prf-adversary A with q queries, there exist adversaries V_A such that*

$$\boldsymbol{Adv}^{\textsf{\textit{prf}}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{\textsf{\textit{pad}}_1}(g(IV,\star_{32}||K||\star_{b-n-32}),\ \star)}(A) = (l-1)\cdot \boldsymbol{Adv}^{\textsf{\textit{multi-rka-prf}}}_{f(RK(\star,K_\star),\star),\Phi_4}(V_A),$$

where V_A is defined in Fig. 7 and V_A can make at most q queries. $g \xleftarrow{\$} \textsf{\textit{Maps}}(\{0,1\}^b,\{0,1\}^n)$. We assume that for each query M of A, the b-bit block length of $\textsf{\textit{pad}}_1(M)$ is at most l. $\Phi_4 = \{\phi_1,\cdots,\phi_l,\phi_P\}$, where $\phi_i(X) = X \oplus i$ and P is the last counter of pfCM-MD. We assume that $\{\phi^1,\cdots,\phi^q\} \subset \{\phi_P,\phi_j\}$ for some j, where (i^t,ϕ^t,X^t) is t-th query of V_A. In other words, even though V_A can make queries to any one of $\{O_1,O_2,\cdots,O_q\}$, V_A can use at most two related-key-deriving (RKD) functions ϕ's from Φ_4.

Proof. We let $F(K,\star)$ be $\mathrm{pfCM}^1-\mathrm{MD}^f_{\mathsf{pad}_1}(g(IV,\star_{32}||K||\star_{b-n-32}),\ \star)$.

$$\begin{aligned}
&\mathbf{Adv}^{\textsf{multi-rka-prf}}_{f(RK(\star,K_\star),\star),\Phi_4}(V_A)\\
&\quad= |\Pr[K_1,\cdots,K_q \xleftarrow{\$} \{0,1\}^n : V_A^{f(RK(\star,K_\star),\star)} = 1]\\
&\qquad - \Pr[g_1,\cdots,g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : V_A^{g_\star(RK(\star,K),\star)} = 1]|\\
&\quad= |\textstyle\sum_{j=1}^{l-1} \Pr[K_1,\cdots,K_q \xleftarrow{\$} \{0,1\}^n : V_{A,i=j}^{f(RK(\star,K_\star),\star)} = 1]\cdot \frac{1}{l-1}\\
&\qquad - \textstyle\sum_{j=1}^{l-1} \Pr[g_1,\cdots,g_q \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^{n+b},\{0,1\}^n); K \xleftarrow{\$} \{0,1\}^n : V_{A,i=j}^{g_\star(RK(\star,K),\star)} = 1]\cdot \frac{1}{l-1}|\\
&\quad= \tfrac{1}{l-1}|\Pr[K \xleftarrow{\$} \{0,1\}^n : V^{F(K,\star)} = 1] - \Pr[g \xleftarrow{\$} \mathsf{Maps}(\{0,1\}^*,\{0,1\}^n) : V^{g(\star)} = 1]|\\
&\quad= \tfrac{1}{l-1}\mathbf{Adv}^{\mathsf{prf}}_{F(K,\star)}(A).
\end{aligned}$$

∎

The second equality follows from the definition of V_A in Fig. 7 and the third equality follows from Lemma 19 and Lemma 20.

Theorem 7. *For any prf-adversary A with q queries, there exist a prf-adversary Q_A such that*

$$\begin{aligned}\boldsymbol{Adv}^{\textsf{\textit{prf}}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{\textsf{\textit{pad}}}(IV,\ \star_{32}||K||\ \star)}(A) \le\ & \boldsymbol{Adv}^{\textsf{\textit{prf}}}_{\mathrm{pfCM}^1-\mathrm{MD}^f_{\textsf{\textit{pad}}_1}(g(IV,\star_{32}||K||\star_{b-n-32}),\ \star)}(A)\\ &+ \boldsymbol{Adv}^{\textsf{\textit{prf}}}_{f(\star,\star_{32}||K||\star_{b-n-32})}(Q_A),\end{aligned}$$

where Q_A can make q queries of the form $(IV||\star_{32}||\star_{b-n-32})$ and is defined in Fig. 7, $\star_i$ means any i-bit string, and $g \xleftarrow{\$} \textsf{\textit{Maps}}(\{0,1\}^b,\{0,1\}^n)$.

Proof. By the definition of the prf-advantage, $\mathbf{Adv}^{\mathsf{prf}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{\mathsf{pad}}(IV,\ \star_{32}||K||\ \star)}(A) = |\Pr[A^{G_3''} = 1] - \Pr[A^{G_1''} = 1]|$. So, we can get above theorem with Lemma 17 $\sim$ Lemma 18. ∎

Corollary 3. *For any adversary A with q queries, there exist adversaries Q_A and V_A such that*

$$\boldsymbol{Adv}^{\textsf{\textit{prf}}}_{\mathrm{pfCM}^0-\mathrm{MD}^f_{\textsf{\textit{pad}}}(IV,\ \star_{32}||K||\ \star)}(A) \le (l-1)\cdot \boldsymbol{Adv}^{\textsf{\textit{multi-rka-prf}}}_{f(RK(\star,K_\star),\star),\Phi_4}(V_A) + \boldsymbol{Adv}^{\textsf{\textit{prf}}}_{f(\star,\star_{32}||K||\star_{b-n-32})}(Q_A),$$

where Q_A, V_A and Φ_4 are defined as before.

Proof. This holds by Theorem 6 and 7. ∎

7 eTCR Security Analysis of a pfCM-MD Domain Extension with the message randomization in NIST SP 800-106

Draft NIST SP 800-106 [27] describes a randomizing hashing for digital signatures [17]. More precisely, Draft NIST SP 800-106 defines a randomization method for randomizing messages prior to hashing. That is, the randomized method works independently from a hash function. There is only a restriction on the hash function, which should process messages in the usual left-to-right order. pfCM-MD is such an example. When $\mathbf{H} = \{H_r(IV, \star)\}_{r\in\mathcal{R}}$ is a hash family, the security of the randomized hashing is measured by the following game : an adversary A chooses M in advance, then a random string r is given to A, and A tries to find (r', M') such that $H_r(IV, M) = H_{r'}(IV, M')$ and $(r, M) \neq (r', M')$. The measurement of this game is formally defined by the definition of eTCR (which is described in the section 2). In this Section, we show that pfCM-MD with the randomizing hashing in the Draft NIST SP 800-106 is secure if the compression function meets a security assumption. More precisely, we provide eTCR security analysis of pfCM^0-MD with the message randomization (in short, mr) in NIST SP 800-106. And we define a hash family $\mathbf{H} = \{\text{pfCM}^0 - \text{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r, M))\}_{r\in\cup_{80\le i\le 1024}\{0,1\}^i}$, where mr is the message randomization in NIST SP 800-106, and $M \in \{0,1\}^*$. And we let $\mathsf{pad}(M) = M||10^t||\mathsf{bin}_d(|M|)$, where $\mathsf{bin}_d(|M|)$ is the d-bit representation of the bit-length of M and t is the smallest non-negative integer such that $\mathsf{pad}(M)$ is a multiple of b-bit block.

Message Randomization (mr) in NIST SP 800-106

$\boxed{\mathsf{mr}(r, M) = M' :}$

1 `If` $(|M| \geq |r| - 1)$ `then` $padding = 1$ `else` $padding = 1||0^{|r|-|M|-1}$
2 $m = M||padding$
3 Let $n = |r|$
4 `If` $(n > 1024)$ `then` stop and output an error indicator
5 $counter = \lfloor |m|/n \rfloor$
6 $remainder = (|m| \ mod \ n)$
7 Concatenate *counter* copies of the r to the *remainder* left-most bits of the r to get R such that $|R| = |m|$

$$R = r||r||\cdots||r||r[0\ldots(remainder-1)]$$

8 $r\_length\_indicator = r\_length\_indicator\_generation(n)$
9 $M' = r||(m \oplus R)||r\_length\_indicator$
10 Return M';

$\boxed{r\_length\_indicator\_generation(n) :}$ // 80≤ n ≤1024 and the output is 16-bit.

1 $A = n$ and $B = A \ mod \ 2$
2 `If` $B = 0$ `then` $b_{15} = 0$ `else` $b_{15} = 1$
3 `For` $i = 14$ to 0
  3.1 $A = \lfloor A/2 \rfloor$ and $B = A \ mod \ 2$
  3.2 `If` $B = 0$ `then` $b_i = 0$ `else` $b_i = 1$
4 $r\_length\_indicator = b_0||b_1||\cdots||b_{15}$
5 Return $r\_length\_indicator$;

eTCR Security Analysis of pfCM^0-MD with mr in NIST SP 800-106

Lemma 21. *For any* $(r, M) \neq (r', M')$*,* $\mathsf{mr}(r, M) \neq \mathsf{mr}(r', M')$*,*

where mr *is the message randomization in NIST SP 800-106.*

Proof. If $\mathsf{mr}(r, M) = \mathsf{mr}(r', M')$, then by the definition of mr the following equality hold.

$$r||(m \oplus R)||r\_length\_indicator = r'||(m' \oplus R')||r'\_length\_indicator. \qquad (1)$$

Since $|r\_length\_indicator| = |r'\_length\_indicator| = 16$ by the definition of mr, $r\_length\_indicator$ should be equal to $r'\_length\_indicator$, which means that $|r| = |r'|$. And since r and r' are located in the first some bits in the equality (1), we know that $r = r'$, which means also that $m = m'$ and $R = R'$, where R and R' are generated from the identical r $(=r')$. Finally, by the padding method defined in line 1 and 2 of mr, $m = m'$ means that $M = M'$. Therefore, the lemma holds. ■

In the following theorem, it is shown that the eTCR-advantage of A on the pfCM^0-MD with mr is bounded by the eSPR†-advantage of A on the pfCM^0-MD with mr.

Theorem 8. *For any eTCR-adversary A, there exists a SPR†-adversary B_A such that*

$$\boldsymbol{Adv}_{\boldsymbol{H}}^{eTCR}(A) \leq l \cdot \boldsymbol{Adv}_{\boldsymbol{H}}^{eSPR^\dagger}(B_A),$$

where $\boldsymbol{H} = \{\mathrm{pfCM}^0 - \mathrm{MD}^f_{pad}(IV, \mathsf{mr}(r, \star))\}_{r \in \cup_{80 \leq i \leq 1024}\{0,1\}^i}$*, and* mr *is the message randomization in NIST SP 800-106. B_A is defined in Fig. 8. l is defined in Fig. 8.*

Proof. Let $H_r(IV, \star)$ be $\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r, \star))$. Δ is the statement that "$(M, \mathtt{State}) \overset{\$}{\leftarrow} A; r \overset{\$}{\leftarrow} \mathcal{R}; (r', M') \overset{\$}{\leftarrow} A(r, M, \mathtt{State}) : (r, M) \neq (r', M')$ and $H_r(IV, M) = H_{r'}(IV, M')$". Υ is the statement that "$(M, \mathtt{State}) \overset{\$}{\leftarrow} B_A; r \overset{\$}{\leftarrow} \cup_{80 \leq j \leq 1024}\{0,1\}^j; i \overset{\$}{\leftarrow} [1, l]; (c', m') \overset{\$}{\leftarrow} B_A(i, r, M, \mathtt{State}) : (c, m) = H_r(IV, M)[i]$ and $(c, m) \neq (c', m')$ and $f(c, m) = f(c', m')$".

$$\begin{aligned}\mathbf{Adv}_{\mathbf{H}}^{\mathsf{eTCR}}(A) &= \Pr[\Delta] = \Pr[\Delta \wedge (|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|)] + \Pr[\Delta \wedge (|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|)] \\ &\leq l \cdot \Pr[\Upsilon \wedge (|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|)] + l \cdot \Pr[\Upsilon \wedge (|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|)] \\ &= l \cdot \Pr[\Upsilon] = l \cdot \mathbf{Adv}_{\mathbf{H}}^{\mathsf{eSPR}^\dagger}(B_A).\end{aligned}$$

The equality of the second line is guaranteed by Claim 1 and Claim 2.

Claim 1. $\Pr[\Delta \wedge (|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|)] \leq l \cdot \Pr[\Upsilon \wedge (|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|)]$.

Proof. Since $\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \star)$ preserves the collision-resistance of f and $|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|$, if $(\mathsf{mr}(r, M), \mathsf{mr}(r', M'))$ is a collision pair of $\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \star)$, there exists a i such that $f(c, x) = f(c', x')$, where $(c, x) = \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r, M))[i]$, $(c', x') = \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M'))[i]$, and $(c, x) \neq (c', x')$. In the definition of B_A in Fig. 8, the probability that i is correctly guessed is $1/l$. So, the Claim 1 holds.

Claim 2. $\Pr[\Delta \wedge (|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|)] = l \cdot \Pr[\Upsilon \wedge (|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|)]$.

Proof. Since $\mathsf{pad}(M) = M||10^t||\mathsf{bin}_d(|M|)$, if $|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|$, and $(\mathsf{mr}(r, M), \mathsf{mr}(r', M'))$ is a collision pair of $\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \star)$, then $f(c, x) = f(c', x')$, where $(c, x) = \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r, M))[l]$, $(c', x') = \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M'))[l']$, and $(c, x) \neq (c', x')$. In the definition of B_A in Fig. 8, the probability that $i = l$ is $1/l$. So, the Claim 2 holds. ■

Adversary B_A.

Run A and obtain M from A and Choose M as a target message.
Given $r \xleftarrow{\$} \cup_{80 \leq i \leq 1024}\{0, 1\}^i$
Given $i \xleftarrow{\$} [1, l]$ // $l = \mathsf{Len}_f(\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r, M)))$
Forward r to A.
Obtain (r', M') from A and let $l' = \mathsf{Len}_f(\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M')))$.
if $|\mathsf{mr}(r, M)| = |\mathsf{mr}(r', M')|$ then $(c', m') \leftarrow \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M'))[i]$
if $|\mathsf{mr}(r, M)| \neq |\mathsf{mr}(r', M')|$ then $(c', m') \leftarrow \mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M'))[l']$
Return (c', m')

Fig. 8. Adversary B_A: $l' = \mathsf{Len}_f(\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M')))$ is the number of computations of the compression function f when computing $\mathrm{pfCM}^0 - \mathrm{MD}^f_{\mathsf{pad}}(IV, \mathsf{mr}(r', M'))$ for any r, where M is generated by the adversary A. mr is the message randomization in NIST SP 800-106.

8 Conclusion

In this paper, we have provided the security requirements of the compression function of pfCM-MD, so that several schemes based on pfCM-MD become secure. That is, if a designer want to develop new hash function based on pfCM-MD, out results can be the guideline for the measurement of the security of the underlying compression function. And we also give a simple indifferentiable security analysis on pfCM-chopMD. Till now, there are many domain extensions which are required to be evaluated as shown in this paper. These kinds of research may be future works.

References

1. E. Andreeva, C. Bouillaguet, P. Fouque, J. J. Hoch, J. Kelsey, A. Shamir and S. Zimmer, *Second Preimage Attacks on Dithered Hash Functions*, Advances in Cryptology – EUROCRYPT'08, LNCS 4965, Springer-Verlag, pp. 270–288, 2008.
2. G. Bertoni, J. Daemen, M. Peeters and G. V. Assche, *On the Indifferentiability of the Sponge Construction*, Advances in Cryptology – EUROCRYPT'08, LNCS 4965, Springer-Verlag, pp. 181–197, 2008.
3. M. Bellare, *N*ew Proofs for NMAC and HMAC: Security without Collision-Resistance, Advances in Cryptology – CRYPTO'06, LNCS 4117, Springer-Verlag, pp. 602–619, 2006.
4. M. Bellare, R. Canetti and H. Krawczyk, *Pseudorandom Functions Revisited: The Cascade Construction and its Concrete Security*, In the preceedings of the 37th Symposium on Foundations of Computer Science, IEEE, 1996.

5. M. Bellare, R. Canetti and H. Krawczyk, *Keying Hash Functions for Message Authentication*, Advances in Cryptology – CRYPTO'96, LNCS 1109, Springer-Verlag, pp. 1–15, 1996.
6. G. Bertoni, J. Daemen, M. Peeters, and G. V. Assche, *On the Indifferentiability of the Sponge Construction*, Advances in Cryptology – EUROCRYPT'08, LNCS 4965, Springer-Verlag, pp. 181–197, 2008.
7. M. Bellare and T. Kohno, *A Theoretical Treatment of Related-Key Attacks: RKA-PRPs, RKA-PRFs, and Applications*, Advances in Cryptology – EUROCRYPT'2003, LNCS 2656, Springer-Verlag, pp. 491–506, 2003.
8. M. Bellare and T. Ristenpart, *Multi-Property-Preserving Hash Domain Extension and the EMD Transform*, Asiacrypt'2006, LNCS 4284, pp. 299–314, 2006.
9. M. Bellare and P. Rogaway, *Random Oracles Are Practical : A Paradigm for Designing Efficient Protocols*, In 1st Conference on Computing and Communications Security, ACM, pages 62–73, 1993.
10. M. Bellare and P. Rogaway, *The game-playing technique and its application to triple encryption*, Cryptology ePrint Archive: Report 2004/331, 2004.
11. E. Biham and O. Dunkelman, *A Framework for Iterative Hash Functions: HAIFA*, In the second NIST Hash Workshop, 2006.
12. D. Chang, S. Lee, M. Nandi and M. Yung, *Indifferentiable Security Analysis of Popular Hash Functions with Prefix-Free Padding.* Advances in Cryptology – ASIACRYPT'06, LNCS 4284, pp. 283–298, 2006.
13. D. Chang and M. Nandi, *Improved Indifferentiability Security Proof of chopMD Hash Function*, FSE'2008, LNCS 5086, Springer-Verlag, pp. 429–443, 2008.
14. D. Chang, J. Sung, S. Hong and S. Lee, *Indifferentiable Security Analysis of choppfMD, chopMD, a chopMDP, chopWPH, chopNI, chopEMD, chopCS, chopESh Hash Domain Extensions*, Cryptology ePrint Archive: Report 2008/407, 2008.
15. J. S. Coron, Y. Dodis, C. Malinaud and P. Puniya, *Merkle-Damgård Revisited: How to Construct a Hash Function*, Advances in Cryptology – CRYPTO'05, LNCS 3621, Springer-Verlag, pp. 430–448, 2005.
16. I. B. Damgard, *A design principle for hash functions*, Advances in Cryptology – CRYPTO'89, LNCS 435, Springer-Verlag, pp. 416–427, 1990.
17. S. Halevi and H. Krawczyk, *Strengthening Digital Signatures via Randomized Hashing*, Advances in Cryptology – CRYPTO'06, LNCS 4117, Springer-Verlag, pp. 41–59, 1996.
18. S. Hirose, J. H. Park and A. Yun, *A Simple Variant of the Merkle-Damgård Scheme with a Permutation*, Advances in Cryptology – ASIACRYPT'07, LNCS 4833, Springer-Verlag, pp. 113–129, 2007.
19. S. Hirose, *Security Analysis of DRBG Using HMAC in NIST SP 800-90*, WISA'08, to appear.
20. J. Kelsey and B. Schneier, *Second preimages on n-bit hash functions for much less than 2^n work*, Advances in Cryptology – EUROCRYPT'05, LNCS 3494, Springer-Verlag, pp. 474–490, 2005.
21. X. Lai and J. L. Massey, *Hash Function Based on Block Ciphers*, Advances in Cryptology – EUROCRYPT'92, LNCS 658, Springer-Verlag, pp. 55–70, 1993.
22. U. Maurer, R. Renner and C. Holenstein, *Indifferentiability, Impossibility Results on Reductions, and Applications to the Random Oracle Methodology*, TCC'04, LNCS 2951, Springer-Verlag, pp. 21–39, 2004.
23. Ueli Maurer and Stefano Tessaro, *Domain Extension of Public Random Functions: Beyond the Birthday Barrier*, CRYPTO'2007, LNCS 4622, pp. 187–204, 2007.
24. R. C. Merkle, *One way hash functions and DES*, Advances in Cryptology – CRYPTO'89, LNCS 435, Springer-Verlag, pp. 428–446, 1990.

25. NIST SP 800-56A, *Recommendation for Pair-Wise Key Establishment Schemes Using Discrete Logarithm Cryptography*, http://csrc.nist.gov/publications/nistpubs/800-56A/SP800-56A_Revision1_Mar08-2007.pdf.
26. NIST SP 800-90, *Recommendation for Random Number Generation Using Deterministic Random Bit Generators*, http://csrc.nist.gov/publications/nistpubs/800-90/SP800-90revised_March2007.pdf.
27. NIST SP 800-106, *DRAFT Randomized Hashing Digital Signatures (2nd draft)*, http://csrc.nist.gov/publications/drafts/800-106/2nd-Draft_SP800-106_July2008.pdf.
28. NIST Hash Project, *Announcing Request for Candidate Algorithm Nominations for a New Cryptographic Hash Algorithm (SHA-3) Family*, http://csrc.nist.gov/groups/ST/hash/documents/SHA-3_FR_Notice_Nov02_2007%20-%20more%20readable%20version.pdf.