Sequential Hashing with Minimum Padding

Shoichi Hirose

University of Fukui

ASK 2016 (2016/09/28-30, Nagoya)

Introduction

Hash function $H : \Sigma^* \to \Sigma^n$

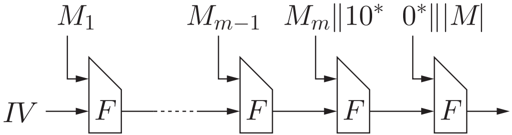
Two popular design strategies:

- Compression-function-based: SHA-2
- Permutation-based: SHA-3

Construction: FIL primitive + domain extension

Strengthened MD

$H^F_{IV}(M)$, where $M = M_1 \| M_2 \| \cdots \| M_m$

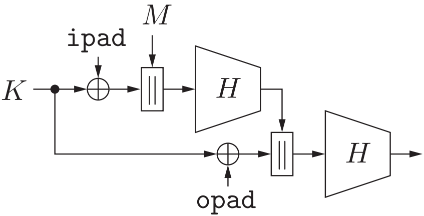


Pros
- Collision resistance is preserved.

Cons
- Length-extension property
- The last message block may consist only of the padding sequence.

Cons degrade efficiency.

HMAC [BCK96]



- Calls H twice to prevent length-extension attacks
- Not efficient for short messages

Overview of the Results

Domain extension scheme for sequential hashing

- with minimum padding
- free from length-extension

Security analysis of the domain extension scheme

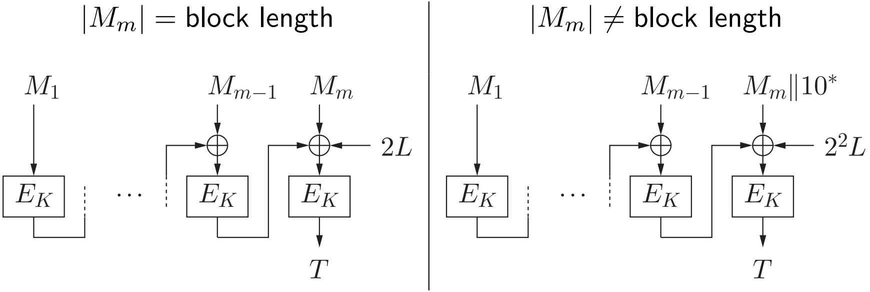
- Collision resistance
- Indifferentiability from a random oracle (IRO)
- pseudorandom function (PRF) of keyed-via-IV mode

Application to sponge construction

- Indifferentiability from a random oracle

Minimum and non-injective padding is common for BC-based MAC

E.g.) CMAC



- $L = E_K(\mathbf{0})$
- $2L$ and 2^2L are used for
  - preventing the length-extension
  - separating the domain (Padding is not injective)

Minimum Padding for Sequential Hashing

For sequential iteration of $F : \Sigma^n \times \Sigma^w \to \Sigma^n$ with IV

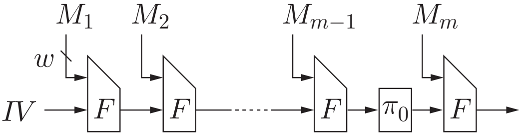
$$\mathtt{pad}(M) = \begin{cases} M & \text{if } |M| > 0 \text{ and } |M| \equiv 0 \pmod{w} \\ M \| 10^* & \text{if } |M| = 0 \text{ or } |M| \not\equiv 0 \pmod{w} \end{cases}$$

- Identical to the padding of CMAC, PMAC, etc.
- Minimum padding sequence
- Not injective

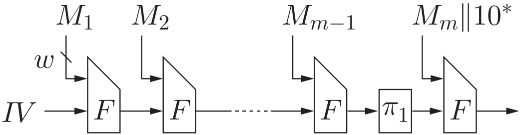
Proposed Domain Extension Scheme

For message $M = M_1 \| M_2 \| \cdots \| M_m$ such that

1. $|M| > 0$ and $|M| \equiv 0 \pmod{w}$,



2. $|M| = 0$ or $|M| \not\equiv 0 \pmod{w}$,

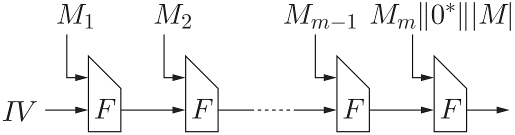


π_0 and π_1 are not cryptographic operations

- Assumption: $\pi_0(v) \neq v \wedge \pi_1(v) \neq v \wedge \pi_0(v) \neq \pi_1(v)$ for any v
- E.g.) XOR with distinct non-zero constants

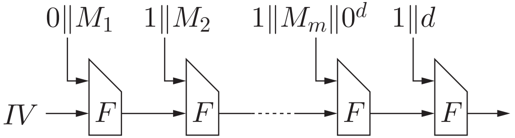
Related Work (CR-Preserving Domain Extension)

Merkle 1989



- Padding-length $\leq$ message-block-length $+ s - 1$ (if $|M|$ is in s-bit)
- Admits M of bounded length, $|M| \leq 2^s - 1$

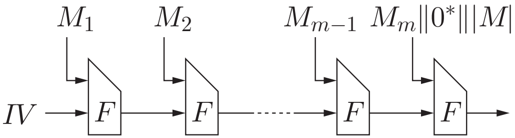
Damgård 1989



- Padding length is $O(|M|)$
- Admits M of arbitrary length

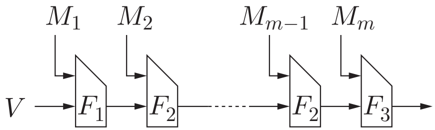
Related Work (CR-Preserving Domain Extension)

Nandi 2009



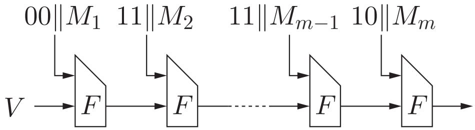
- Admits M of arbitrary length by variable length encoding of $|M|$
- Padding-length $= O(\log |M|)$

Suffix-Free-Prefix-Free Hashing [BGKZ12]



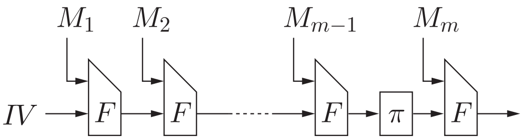
- IV is variable; without MD strengthening
- Needs three CFs
  - F_1 provides prefix-freeness; F_3 provides suffix-freeness
- Satisfies IRO
- Assumes injective padding

Cf.)



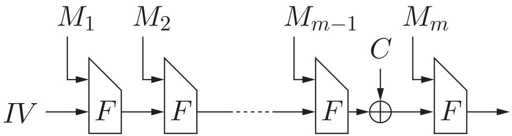
- Padding-length $= O(|M|)$

Merkle-Damgård with Permutation (MDP) [HPY07]

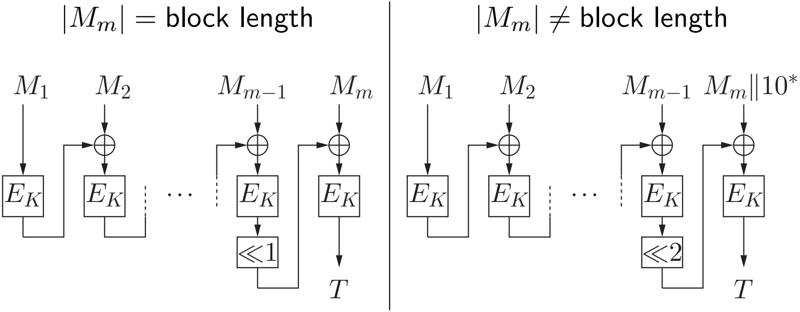


- π is not a cryptographic primitive

Cf.) Ferguson, Kelsey 2001 (Comment on Draft FIPS 180-2)



GCBC1 [Nandi 09]



- XOR with constants does not work
- Requires at least two message blocks

Collision Resistance in the Standard Model

Lemma

Any collision pair for $H_{IV}^{F,\{\pi_0,\pi_1\}}$ implies

- *a collision pair,*
- *a $\{\pi_0, \pi_1\}$-pseudo-collision pair, or*
- *a preimage of IV, $\pi_0^{-1}(\pi_1(IV))$, or $\pi_1^{-1}(\pi_0(IV))$*

for F

Proof: Backward induction

$\{\pi_0, \pi_1\}$-pseudo-collision pair for F:

$$(V, X) \text{ and } (V', X') \text{ s.t. } \pi_0(F(V, X)) = \pi_1(F(V', X'))$$

Collision Resistance in the Standard Model

Theorem

The collision resisntance of $H_{IV}^{F,\{\pi_0,\pi_1\}}$ is reduced to

- *the collision resistance*
- *the $\{\pi_0, \pi_1\}$-pseudo-collision resistance, and*
- *the everywhere preimage resistance*

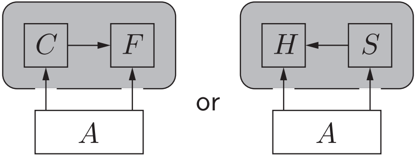
of F.

Everywhere preimage resistance of h:

$$\mathrm{Adv}_h^{\mathrm{epre}}(A) = \max_{Y \in \mathcal{Y}} \{\Pr[M \leftarrow A(h) : h(M) = Y]\}$$

Definition of Indifferentiability from a Random Oracle

[Maurer, Renner, Holenstein 04], [Coron, Dodis, Malinaud, Puniya 05]



- C is hashing mode of F
- F is FIL ideal primitive
  - Random oracle
  - Ideal block cipher
- H is VIL RO
- Simulator S tries to mimic F with access to oracle H

C^F is indiff. from VIL RO (IRO) if no efficient adver A can tell apart

$$(C^F, F) \quad \text{and} \quad (H, S^H)$$

Indifferentiability from a Random Oracle (IRO)

Theorem

Suppose that CF $F : \Sigma^n \times \Sigma^w \to \Sigma^n$ is chosen uniformly at random. Then, for HF $H_{IV}^{F,\{\pi_0,\pi_1\}}$, there exists a simulator S of F s.t., for any adversary A making

- *at most q queries to its FIL oracle*
- *queries to its VIL oracle which cost at most σ message blocks in total,*

$$\mathrm{Adv}^{\mathrm{indiff}}_{H_{IV}^{F,\{\pi_0,\pi_1\}},S}(A) \leq \frac{5(\sigma+q)^2}{2^n} + \frac{3\sigma q}{2^n - 6q + 1} \ ,$$

and S makes at most q queries.

Secure if $\sigma + q = o(2^{n/2})$

IRO in the Ideal Cipher Model

The CF $F : \Sigma^n \times \Sigma^w \to \Sigma^n$ is the Davies-Meyer mode of a BC E

- E is chosen uniformly at random

Theorem

For the hash function $H_{IV}^{F,\{\pi_0,\pi_1\}}$, there exists a simulator S of E s.t., for any adversary A making

- *at most q_{e} queries to its FIL encryption oracle*
- *at most q_{d} queries to its FIL decryption oracle*
- *queries to its VIL oracle which cost at most σ message blocks in total,*

$$\mathrm{Adv}^{\mathrm{indiff}}_{H_{IV}^{F,\{\pi_0,\pi_1\}},S}(A) \leq \frac{12(\sigma + q_{\mathrm{e}} + q_{\mathrm{d}})^2}{2^n} + \frac{3\sigma(q_{\mathrm{e}} + q_{\mathrm{d}})}{2^n - 6(q_{\mathrm{e}} + q_{\mathrm{d}}) - 5} ,$$

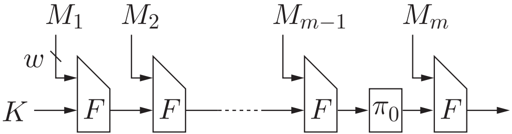
and S makes at most q_{e} queries.

Secure if $\sigma + q_{\mathrm{e}} + q_{\mathrm{d}} = o(2^{n/2})$

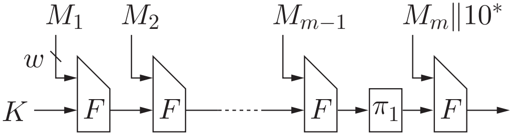
Keyed via IV mode of $H_{IV}^{F,\{\pi_0,\pi_1\}}$

For message M such that

1. $|M| > 0$ and $|M| \equiv 0 \pmod{w}$,



2. $|M| = 0$ or $|M| \not\equiv 0 \pmod{w}$,



PRF Security

Theorem

Let A be any adversary against KIV mode of $H_{IV}^{F,\{\pi_0,\pi_1\}}$:

- *A runs in time at most t and makes at most q queries*
- *The length of each query is at most ℓw*

Then, there exists an adversary B against F such that

$$\mathrm{Adv}^{\mathrm{prf}}_{H_{IV}^{F,\{\pi_0,\pi_1\}}}(A) \leq \ell q \, \mathrm{Adv}^{\mathrm{prf\text{-}rka}}_{\{id,\pi_1,\pi_2\},F}(B) \ .$$

B runs in time at most $t + O(\ell q T_F)$ and makes at most q queries.

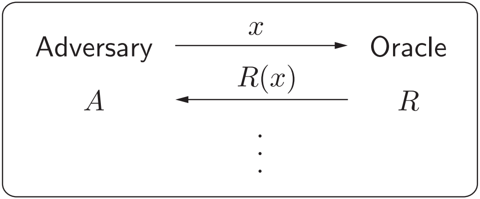
$H^{F,\{\pi_0,\pi_1\}}$ is PRF $\Longleftarrow$ F is PRF against $\{id, \pi_1, \pi_2\}$-restricted RKAs

Definition of PRF

A keyed function $f : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$ is PRF

$\Longleftarrow$ f_K is indistinguishable from uniform random function $\rho : \mathcal{D} \to \mathcal{R}$

- Secret key $K \in \mathcal{K}$ is chosen uniformly at random
- Adversary makes queries to f_K or ρ



R is f_K or ρ

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

PRF against Related Key Attacks

$f : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$ is PRF against Ψ-restricted RKAs if

f is indistinguishable from uniform random keyed function $\rho : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$

- Ψ is a set of related-key deriving functions
- Secret key $K \in \mathcal{K}$ is chosen uniformly at random
- Adversary makes queries to $f_{\psi(K)}$ or $\rho_{\psi(K)}$ for any $\psi \in \Psi$

$$\begin{array}{ccc} \text{Adversary} & \xrightarrow{\quad \psi,\ x \quad} & \text{Oracle} \\ A & \xleftarrow{\quad R_{\psi(K)}(x) \quad} & R,\ K \\ & \vdots & \end{array}$$

$R \in \{f, \rho\}$
$\psi \in \Psi$

$$\mathrm{Adv}^{\text{prf-rka}}_{\Psi,f}(A) = \left| \Pr[A^{(f_{\psi(K)})_{\psi\in\Psi}} = 1] - \Pr[A^{(\rho_{\psi(K)})_{\psi\in\Psi}} = 1] \right|$$

PRF Modes Based on Hash Function

Modes using a hash function

- HMAC [Bellare, Canetti, Krawczyk 1996]
- Sandwich MD [Yasuda 2007]
- HMAC without the second key [Yasuda 2009]
- AMAC (Augmented MAC) [Bellare, Bernstein, Tessaro 2016]

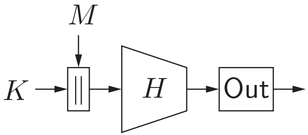
Modes using a compression function

- Plain Merkle-Damgård (MD) with prefix-free encoding [BCK1996]
- EMD (Enveloped MD) [Bellare, Ristenpart 2006]
- MDP (MD with Permutation) [Hirose, Park, Yun 2007]
- Boosting MD [Yasuda 2007]
- OMD MAC function [CMNCRVV2014]

All of the above assume injective padding except for OMD MAC function.

- OMD MAC function uses keyed CF with tweaks.

AMAC (Augmented MAC) [BBT16]

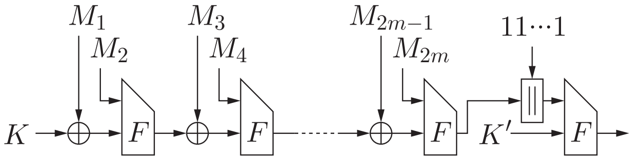


- Used in the Ed25519 signature scheme
- Out is not a cryptographic primitive
  E.g.) truncation or mod function

AMAC^H is PRF $\Longleftarrow$ F is PRF under leakage of the key by Out

BNMAC (Boosted NMAC) [Yas07]

Double-key version (Single-key version is also presented)

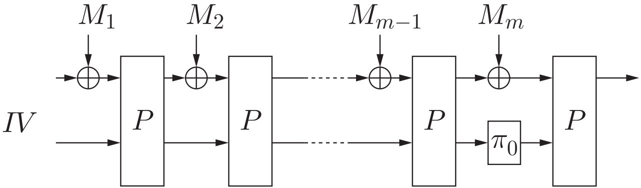


BNMAC^F is PRF $\Longleftarrow$ F is PRF and Δ-2PRF

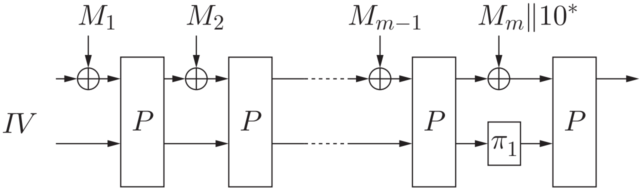
Application to Sponge Construction

For message M such that

1. $|M| > 0$ and $|M| \equiv 0 \pmod{w}$,



2. $|M| = 0$ or $|M| \not\equiv 0 \pmod{w}$,



IRO in the Ideal Permutaton Model

The permutation $P : \Sigma^b \to \Sigma^b$ is chosen uniformly at random

- $b = r + c$ and c is capacity of sponge construction

Theorem

For the hash function $G_{IV}^{P,\{\pi_0,\pi_1\}}$, there exists a simulator S of P s.t., for any adversary A making

- *at most q_f queries to its FIL forward oracle*
- *at most q_b queries to its FIL backward oracle*
- *queries to its VIL oracle which cost at most σ message blocks in total,*

$$\mathrm{Adv}^{\mathrm{indiff}}_{G_{IV}^{P,\{\pi_0,\pi_1\}},S}(A) \leq \frac{12(\sigma + q_\mathrm{f} + q_\mathrm{b})^2}{2^c} + \frac{3\sigma(q_\mathrm{f} + q_\mathrm{b})}{2^c - 6(q_\mathrm{f} + q_\mathrm{b}) - 5} ,$$

and S makes at most q_f queries.

Secure if $\sigma + q_\mathrm{f} + q_\mathrm{b} = o(2^{c/2})$

Conclusion

Domain extension scheme for sequential hashing

- with minimum padding
- free from length-extension

Security analysis of the domain extension scheme

- Collision resistance
- Indifferentiability from a random oracle
  - in the random oracle model
  - in the ideal cipher model with Davies-Meyer CF
- Pseudorandom function by keyed-via-IV

Application to sponge construction

- IRO in the ideal permutation model