한국항행학회 논문지 제13권 제4호 2009년 8월

여러 가지 해쉬 함수 도메인 확장 방법에 대한 Indifferentiability 관점에서의 안전성 분석

Indifferentiable Security Analysis of Several Hash Domain Extensions

장동훈*, 성재철**, 홍석희***, 이상진***

Dong-Hoon Chang*, Jae-Chul Sung**, Seok-Hie Hong*** and Sang-Jin Lee***

요 약

본 논문에서는 pfMD, MDP, WPH, EMD, NI, CS 해쉬 도메인 확장 방법과 각각에 구조에서 출력 값 일부를 이용하지 않는 truncated 버전에 대한 Indifferentiability 관점에서의 안전성 분석 결과를 제시한다. 본 논문에서 고려한 분석 방법은 기존의 분석 방법과는 달리 단순하고 쉽다는 특징을 갖는다. 뿐만 아니라 본 분석 방법은 해쉬 함수의 임의의 구조에 대해 쉽게 적용이 된다는 특징을 지니고 있기에, 신규 해쉬 함수 개발 시 안전성 분석 도구로 사용될 수 있다.

Abstract

We provide indifferentiable security analyses of pfMD, MDP, WPH, EMD, NI and CS hash domain extensions and their truncated versions. Unlike previous analytic techniques, the analytic technique considered in this paper is simple and easy. Moreover, the analytic technique can be generally applied to any types of hash domain extensions. That means that the technique can be used as an analyzing tool for any new developed hash function.

Key words : Indifferentiability, Hash Function, Domain Extension

I. 서 론

암호학적 해쉬 함수 (cryptographic hash function)는 임의의 길이의 비트열을 입력으로 받아 고정된 길이의 비트열을 출력하는 함수를 의미하며, 기본적으로 다음의 성질을 만족시켜야 한다. (여기서 H는 해쉬 함수를 의미한다.)

○ 역상 저항성 : 랜덤한 출력값 y가 주어졌을 경우, H(M)=y인 M을 찾는 것이 어려워야 한다.

○ 제 2 역상 저항성 : 랜덤한 메시지 M이 주어졌을 경우, H(M)=H(M')인 M'(≠M)을 찾는 것이 어려워야 한다.

○ 충돌 저항성 : H(M)=H(M')인 M과 M' (M≠M')을 찾는 것이 어려워야 한다.

* 컬럼비아대학 포닥 연구원
** 서울시립대 수학과
*** 고려대 정보보호기술연구원
· 제1저자 (First Author) : 장동훈
· 투고일자 : 2009년 7월 24일
· 심사(수정)일자 : 2009년 7월 27일 (수정일자 : 2009년 8월 17일)
· 게재일자 : 2009년 8월 30일

최근 2005년에는 MD5 [26]과 SHA-1 [18]에 대한 충돌 쌍 탐색 공격 [27,28]이 발표가 되어, NIST에서는 2007년부터 SHA-3 해쉬 알고리즘 개발을 위한 프로젝트를 진행 중에 있다 [17]. 이는 2001년에 개발이 완료된 블록 암호 AES [21]의 선정 과정과 비슷하게 진행이 되고 있다. 이처럼 SHA-3의 개발의 필요성이 대두되게 된 근본적인 이유 중에 하나는 SHA-2 [19]가 SHA-1의 설계 논리와 비슷하여 SHA-2의 안전성을 보장할 수 없는 상황에 이르렀기 때문이다. 해쉬 함수는 위의 세 가지 성질 외에도 의사 충돌 저항성 (Pseudo-Collision Resistance), 유사 충돌 저항성 (Near-Collision Resistance) 등 다양한 암호학적 성질을 충족시켜야 하는데 이는 해쉬 함수가 전자서명 [20], 메시지인증코드 [2,22], 의사난수생성기 [24], 키 생성 함수 [23], 랜덤마이징 해싱 [11] 등 여러 분야에 사용되고 있기 때문이다. 무엇보다 해쉬 함수는 랜덤 오라클 [4]을 구현할 경우에도 사용되기 때문에, 랜덤 오라클을 구현할 필요가 있는 경우 해쉬 함수는 랜덤 오라클에서 발생하지 않는 그 어떠한 취약성도 발견되면 안 된다. 그러나 해쉬 함수는 랜덤 오라클과 달리, 하나의 공개된 함수이기 때문에 엄밀히 말해서 랜덤 오라클을 대체할 수 없다. 설사 대체할 수는 없다 할지라도, 최소한 다음의 말은 할 수 있어야 바람직한 해쉬 함수라고 말을 할 수는 있다 : 가령, 압축 함수 대신 고정된 입출력 크기를 갖는 랜덤 오라클 (Fixed Input Length random oracle or FIL random oracle)을 사용할 경우, 그러한 해쉬 함수 구조가 과연 임의 길이의 입력 값을 갖는 랜덤 오라클을 과연 대체할 수 있을 것인가? 만약 해쉬 함수의 기반이 되는 해쉬 구조가 잘 설계된 구조라면, 랜덤 오라클을 대체할 수 있어야 한다. 거꾸로 랜덤 오라클을 대신할 수 없는 해쉬 구조라면, 아무리 압축 함수를 잘 설계한다고 하더라도, 취약한 해쉬 구조에 기반하여 설계된 해쉬 함수는 여전히 취약할 수밖에 없기 때문이다. 그렇다면, 주어진 해쉬 구조에 대해, 압축 함수 대신 FIL 랜덤 오라클을 사용할 경우, 그러한 해쉬 구조가 과연 랜덤 오라클을 대신할 수 있는지에 대한 여부를 어떻게 판단할 것인가의 문제가 남게 된다. 이러한 문제에 방향을 제시하는 논문을 Maurer 등 [14]이 TCC 2004에서 발표하였는데, 제안된 신규 안전성 개념이 바로 Indifferentiability이다. 그리고 이듬해 CRYPTO 2005에서 Coron 등 [9]은 Indifferentiability의 이론을 해쉬 함수 구성 방법에 적용한 결과를 발표하였다. Coron 등은 압축 함수가 고정된 입력 크기를 갖는 랜덤 오라클일 경우, prefix-free MD, chop MD, NMAC 구성 방법, HMAC 구성 방법이 랜덤 오라클과 구별 불가능하게 됨을 증명하였다. 즉, 랜덤 오라클을 대신할 수 있다는 뜻이다. 사실, Coron의 결과는 압축 함수가 FIL 랜덤 오라클이라는 가정을 기반으로 하고 있기에 실제 해쉬 함수와는 맞지는 않지만, 최소한 구조적인 측면에서 취약성이 발견되지 않는다는 중요한 정보를 제공하기에 실제적인 측면에서 중요한 의미를 갖는다. 이어서 ASIACRYPT 2006에서는 장동훈 등 [7], Bellare 등 [3]이 해쉬 함수에 적용된 Indifferentiability 이론을 발전시켰고, 동시에 신규 구조를 제안하고 이에 대한 안전성 분석을 제시하였다. ASIACRYPT 2007에서는 Hirose 등 [12]이 신규 구조인 MDP 구조를 제안하고 동시에 안전성 증명을 수행하였다. 그 이후에도 Indifferentiability 개념과 관련된 여러 연구 결과들이 발표되었다. 그런데, 위의 안전성 분석들은 복잡한데다가 내용이 길기 때문에, 다른 신규 구조의 안전성 증명 시 적용하는 데에 한계를 갖고 있다. 최근, EUROCRYPT 2008에서 Bertoni 등 [5]은 해쉬 함수의 Indifferentiability 안전성 증명과 관련하여 일반적으로 해쉬 구조의 안전성을 증명할 수 있는 방향을 제시하였다.

본 논문은 EUROCRYPT 2008의 논문 [5]의 증명 방식을 개선하고, 수학적으로 엄밀하게 다루어, 여러 해쉬 함수 구조의 안전성 증명을 제시하고자 한다. 본 논문에서의 증명은 매우 단순하여 임의의 구조에 대해 쉽게 적용이 가능하다는 특징을 갖는다.

본 논문은 다음과 같이 구성된다. 2 절에서는 여러 해쉬 함수 구조에 대한 설명 및 일반적인 개념과 용어 그리고 정리들을 기술한다. 3 절에서는 2 절에서 제시한 여러 해쉬 함수 구조에 대하여 Indifferentiability 분석 결과를 제시한다. 마지막으로 4 절에서는 결론으로 끝을 맺는다.

Ⅱ. 용어, 정의 및 정리

2-1. 용어 정의

본 논문에서는 $f:\{0,1\}^n\times\{0,1\}^b\rightarrow\{0,1\}^n$로 정의하며, f를 압축함수라고 한다. $m\in\{0,1\}^{kb}$ 인 경우, $\|m\|_b=k$ 라고 표현하고, m을 k 블록 메시지라고 말한다.

MD. Merkle-Damgård (MD) 구성 방법 [10,16]은 다음과 같이 동작한다. 임의의 t 블록 메시지 $M=m_1\|\cdots\|m_t$에 대해,

$$MD^f(M)=f(\cdots f(f(IV,m_1),m_2)\cdots,m_t)$$

로 정의하며, 이때, f는 압축함수, IV는 초기치를 의미한다.

패딩 룰. EMD [3]와 CS [15]를 제외한 경우를 먼저 살펴본다. 함수 $g:\{0,1\}^*\rightarrow(\{0,1\}^b)^*$가 단사(injective) 함수이면서, 동일한 비트 길이를 갖는 입력 값은 모두 동일한 비트 길이의 출력 값에 대응된다는 특징(length-preserving)을 지닐 때, 그러한 함수 g를 패딩 룰(padding rule)이라고 말한다. EMD와 CS의 경우에는 패딩 룰

$$g:\{0,1\}^*\rightarrow(\{0,1\}^b)^*\times\{0,1\}^{b-n}$$

를 이용한다. 특히 임의의 $M\neq M'$에 대해, $g(M)$이 $g(M')$의 prefix가 아닐 경우, 이 때 g를 prefix-free 패딩 룰이라고 말한다.

chop. $0\le s\le n$에 대해, $chop_s(x)=x_L$로 정의한다. 여기서 $x=x_L\|x_R$이고 $|x_R|=s$이다.

last. $0\le s\le n$에 대해, $last_s(x)=x_R$로 정의한다. 여기서 $x=x_L\|x_R$이고 $|x_R|=s$이다.

pfMD. prefix-free MD (pfMD) [9]는 다음과 같이 정의된다. 임의의 메시지 M에 대해,

$$pfMD_g^f(M)=MD^f(g(M))$$

로 정의되며, 이때 g는 prefix-free 패딩 룰이다.

MDP. MDP [12]는 MD 구조에다가 한 개의 치환 P가 추가된 구조이며, 이때

$$MDP_g^f(M)=f(P(MD^f(chop_b(g(M)))),last_b(g(M)))$$

으로 정의한다. 여기서 g는 임의의 패딩 룰이며 P와 P^{-1}는 효율적으로 계산이 가능해야 하며, $P\circ P=id$ (항등 함수) 이어야 한다. 본 논문에서는 0이 아닌 상수 값 c을 XOR하는 연산, 즉 $P(x)=x\oplus c$ 인 경우를 살펴볼 것이다.

WPH. Wide-Pipe Hash (WPH) [13]는 Lucks에 의해 제안된 해쉬 구조로, 두 개의 함수, f_1과 f_2에 기반하여 정의된다. 여기서

$$f_1:\{0,1\}^w\times\{0,1\}^b\rightarrow\{0,1\}^w,$$

$$f_2:\{0,1\}^w\rightarrow\{0,1\}^n$$

으로 정의하며, WPH는 임의의 패딩 룰 g에 대해 다음과 같이 동작한다.

$$WPH_g^{f_1,f_2}(M)=f_2(MD^{f_1}(g(M)))$$

로 정의되며, IV는 고정된 w 비트 초기치를 뜻한다. Wide-Pipe Hash라고 이름을 붙은 이유는, w 값이 내부 상태 비트 크기를 결정하기 때문에, w를 조정하여 내부 상태 값을 크게 늘릴 수가 있기 때문이다.

EMD. EMD [3]는 다음과 같이 정의된다.

$$EMD^f(M)=f(IV_2,MD^f(Q)\|M_t)$$

로 정의하며, 여기서 IV_2는 IV와 다른 고정된 초기치를 의미하며, $Q\|M_t=M\|10^r\|bin_{64}(|M|)$을 뜻한다. 이때, $|M_t|=b-n$, $bin_i(x)$는 x를 i 비트로 표현한 값을 의미하며, r은 $|g(x)|-(b-n)$이 b의 배수가 되도록 하는 가장 작은 음이 아닌 정수를 뜻한다.

NI. Nested Iteration (NI) [1]는 다음과 같이 정의된다.

$$NI_g^f(K_1,K_2,M)=f(K_2,MD^{f_{K_1}}(chop_b(g(M))),last_b(g(M)))$$

로 정의하며, g는 임의의 패딩 룰을 뜻한다.

CS. Chain Shift (CS) [15]는 다음과 같이 정의된다.

$$CS_g^f(K,M)=f(K,IV_2,MD^{f_K}(chop_{b-n}(g(M))),last_{b-n}(g(M)))$$

로 정의하며, g는 임의의 패딩 룰을 뜻한다.

Truncated Versions. pfMD, MDP, WPH, EMD, NI,

CS의 truncated 버전은 chop 함수를 이용하여 정의한다. 가령, chop-prefix-free MD (choppfMD)는 다음과 같이 정의된다. 임의의 메시지 M에 대해, $\mathrm{choppfMD}_g^f(M) = \mathrm{chop}_s(\mathrm{MD}_g^f(M))$로 정의되며, 이때 g는 prefix-free 패딩 룰이다. pfMD는 choppfMD의 하나의 예로, $s=0$인 경우이다. HAIFA [6] 역시 choppfMD의 한 가지 예이다. 나머지 경우에 대한 truncated 버전은 동일하게 정의되며, chopMDP, chopWPH, chopEMD, chopNI, chopCS로 표현한다. 본 논문에서는 임의의 s 값을 갖는 truncated 버전에 대한 안전성 분석 결과를 제시한다.

부등식 1. 다음의 부등식은 정리 2를 증명하는 데에 사용된다. $0 \le a_i \le 1$을 만족시키는 임의의 a_i들에 대하여, $\Pi_{i=1}^{q}(1-a_i) \ge 1-\sum_{i=1}^{q}a_i$ 가 성립한다. 본 부등식은 수학적 귀납법에 의해 쉽게 증명된다.

랜덤오라클 모델 (Random Oracle Model) [4]. f의 정의역이 X이고 공역이 Y일 경우, 각각의 $x \in X$에 대해, f(x)의 값이 Y에서 랜덤하게 선택되어진다면, 그때 f를 랜덤 오라클이라고 말한다. 좀더 정확하게 말한다면, $\Pr[f(x)=y \mid f(x_1)=y_1,\cdots,f(x_q)=y_q] = \frac{1}{|Y|}$이 된다. 여기서 $x \notin \{x_1,\cdots,x_q\}$이고 $y_1,\cdots,y_q \in Y$이다. 고정된 d에 대해 $X=\{0,1\}^d$ 라면, 그때의 f를 FIL (Fixed Input Length) 랜덤 오라클이라고 말하고, $X=\{0,1\}^*$ 라면, 그때의 f를 VIL (Variable Input Length) 랜덤 오라클이라고 말한다. 본 논문에서는 VIL 랜덤오라클을 R로 표기한다.

질문의 비용 (Cost). 임의의 주어진 해쉬 구조의 안전도는 질문의 수 q와 각 질문의 최대 길이 l로 표현될 수 있다. 그런데, 최근 [5] 논문에서, sponge construction에 대한 안전도를 기술하기 위하여, 비용 (cost) 라는 용어가 도입하였다. 비용 (cost)는 q 개의 전체 질문의 블록 크기를 뜻한다. 이러한 비용 (cost)의 개념은 보통 수행시간이 압축함수의 수행 횟수에 의해 결정된다는 점에 있어서 의미가 있다. 즉, 질문의 수 q와 각 질문의 최대 길이 l을 정의하지 않고, 단순히 cost가 얼마라고 제한을 두고 안전도를 측정하는 것이다. 비용의 개념은 본 논문에서 각 구조에 대한 indifferentiability 관점에서의 안전도를 측정할 때 사용될 것이다.

Computational Distance. $F=(F_1,\cdots,F_t)$과 $G=(G_1,\cdots,G_t)$를 확률적 오라클 알고리즘들의 모임이라고 하자. 그리고 A는 확률적 알고리즘이라고 하자. 이때, A가 F와 G를 구별하고자 하는 경우에 대한 computational distance는 다음과 같이 정의된다.

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

Statistical Distance. $F=(F_1,\cdots,F_t)$과 $G=(G_1,\cdots,G_t)$를 확률적 오라클 알고리즘들의 모임이라고 하자. 그리고 A는 결정적 알고리즘이라고 하자. 이때, A가 F와 G를 구별하고자 하는 경우에 대한 statistical distance는 다음과 같이 정의된다.

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

이때 $\Pr[O=v]$는 다음을 뜻한다.

$$\Pr[O_{c_i}(x_i)=y_i, 1 \le i \le q, v=((c_1,x_1,y_1),\cdots,(c_q,x_q,y_q))$$

를 가리킨다. 그리고 최대 q개의 질문을 던지는 임의의 공격자 A에 대해 F와 G의 최대 statistical distance를 Stat(F,G)라고 표현한다.

Computational Distance vs. Statistical Distance.

[소정리 1] $F=(F_1,\cdots,F_t)$과 $G=(G_1,\cdots,G_t)$를 확률적 오라클 알고리즘들의 모임이라고 하자. 최대 q개의 질문을 던지는 임의의 확률적 알고리즘 A에 대하여 다음이 성립한다.

$$\mathrm{Adv}_A(F,G) \le \mathrm{Stat}(F,G).$$

2-2. Indifferentiability 소개

우선 indifferentiability 안전성 개념을 소개한다.

[정의 1] Indifferentiability. [14] 튜어링 머신 H가 이상적인 프리미티브인 f에 접근이 가능할 경우에 대하여, 모든 임의의 공격자 D에 대해 다음을 만족할 경우, H^f는 또다른 이상적인 프리미티브 R과

(t_D, t_S, q, ϵ)-indifferentiable 하다고 정의한다.

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

여기서 시뮬레이터 S는 R에 접근이 가능하며, S의 최대 수행시간은 최대 t_S가 된다. 구별하고자 하는 공격자 D의 최대 수행시간은 t_D로 정의하며, 최대 q번 질문을 하거나, 또는 비용(cost)이 최대 q인 경우를 가리킬 수도 있다. 위의 부등식에서 ϵ의 값이 무시할 정도로 매우 작을 경우, H^f는 이상적인 프리미티브 R과 indifferentiable이라고 말을 한다.

[정리 1] [14] P를 이상적인 프리미티브 R에 기반한 임의의 암호시스템이라고 하자. 그리고 H^f는 이상적인 프리미티브 R과 indifferentiable이라고 하자. 이때, R에 기반한 P는 거의 비슷한 안전도로 R 대신 H^f를 대체하여 사용할 수 있다.

위의 정리 1이 의미하는 바를 해쉬 알고리즘에 적용하면 다음과 같이 설명할 수 있다. f는 FIL 랜덤 오라클에 대응되고, H는 f와 특정 해쉬 도메인 확장 방법에 기반한 해쉬 알고리즘으로 볼 수 있다. 그리고 R은 VIL 랜덤 오라클이다. 본 논문에서는 각각의 구조에 대하여 해쉬 알고리즘이 설계될 경우, 압축함수가 FIL 랜덤 오라클이라면, 각 구조가 VIL 랜덤 오라클과 Indifferentiable 함을 증명하고자 한다. 이는 VIL 랜덤 오라클 하에서 안전성이 증명된 암호 시스템에 대하여, 랜덤 오라클에 기반하여 설계된 각 구조를 R 대신 적용하여 사용할 수 있음을 보여준다.

Indifferentiability 개념의 의미. MD 구성 방법의 경우, 압축함수가 FIL 랜덤 오라클이라 할지라도, 확장(extension) 공격이 쉽게 가능하다. 확장 공격이란, H가 해쉬 알고리즘이고 K가 비밀 키인 경우, H(K,M)의 값으로부터, K를 알지 않고서도 H(K,M‖M′)의 값을 계산할 수 있다. 이는 VIL 랜덤 오라클 하에서는 발생하지 않는 취약점이라고 볼 수 있다. 이처럼 Indifferentiability는 해쉬 알고리즘의 구조에 대한 안전도를 측정할 수 있다는 점에 있어서 의미가 있다.

Indifferentiability 개념의 한계. 현실에서는 랜덤 오라클과 같이 이상적인 모델은 존재하지 않기 때문에, 말 그대로 이상적인 모델에 지나지 않는다. 또한 VIL 랜덤 오라클과 Indifferentiable한 것을 구성하기 위해서도 또 다른 이상적인 모델 FIL 랜덤 오라클이 요구된다. 이처럼, Indifferentiability 개념은 이상적인 모델 사이에서의 관계를 살펴보는 것이기 때문에, 실제 사용되는 알고리즘의 안전성을 보장해주지는 않는다.

Indistinguishability vs. Indifferentiability. Indistinguishability 개념에서는 공격자에게 오라클에 대한 질문에 대한 답을 받을 뿐, 답이 나오기까지 내부 상태 정보에 대한 어떠한 정보도 제공하지 않는 두 상황을 구별하는 데에 초점을 맞춘다. 반면, Indifferentiability 개념은 암호 시스템의 입출력 값 이외에, 내부 상태 값에 대한 정보가 주어지는 경우에 대해 안전도를 측정하는 개념이라고 볼 수 있다.

Ⅲ. Indifferentiability 관점에서의 안전성 분석

Indifferentiability 안전성 개념은 TCC 2004에서 Maurer 등에 의해 소개되었다 [14]. Indifferentiability 안전성 개념은, 압축 함수의 기반이 되는 함수가 FIL 랜덤 오라클 또는 이상적인 블록암호 (ideal cipher)라는 가정 하에서, 모든 가능한 공격자에 대한 해쉬 함수의 안전성을 평가할 수 있는 방법을 제시하고 있다. CRYPTO 2005에서는 Coron 등 [9]은, 압축함수가 FIL 랜덤 오라클이라는 가정 하에, MD 구성 방법이 VIL 랜덤 오라클과 indifferentiable 하지 않음을 보이고, VIL 랜덤 오라클과 indifferentiable한 네 가지 구성 방법, 즉 prefix-free MD, chopMD, NMAC construction, HMAC construction 을 제안하였다. 이 논문을 필두로 하여, 여러 결과가 지금까지 발표되어왔다 [3,5,7,8,12]. 본 절에서는 chopEMD에 대한 indifferentiable 안전성 분석 결과를 제시한다. 이때, $chop_s$ 함수에서 s=0이라면, chopEMD는 EMD와 동일하므로, EMD는 chopEMD의 한 예라고 볼 수 있다. 따라서 본 절에서는 임의의 s에 대해서 안전성 분석 결과를 제시한다. 그 밖의 해쉬 구조에 대해서도 동

일한 방법으로 분석되기 때문에, 증명 결과만 제시하고자 한다.

3-1. 시뮬레이터 구성

먼저 chopEMD에 대한 indifferentiable 안전성 분석을 하기에 앞서, 시뮬레이터 $S_{chopEMD}$를 먼저 구성한다. 시뮬레이터을 기술하는 방식은 [8] 논문을 참조하였다. 시뮬레이터에서 $R:\{0,1\}^{*}\rightarrow\{0,1\}^{n}$ 은 VIL 랜덤 오라클을 가리킨다.

시뮬레이터 $S_{chopEMD}$ 의 정의.

① 초기화.

1-1. 부분 함수(partial function) $e:\{0,1\}^{n+b}\rightarrow\{0,1\}^{n}$ 는 공집합으로 시작한다.

1-2. 부분 함수 $e^{*}=CM-MD^{e}:(\{0,1\}^{b})^{*}\rightarrow\{0,1\}^{n}$ 는 처음에 $e^{*}(\lambda)=IV$ 으로 시작한다.

1-3. 집합 $C=\{IV,IV_2\}$ 와 집합 $I=\{\lambda\}$ 를 정의한다.

② $S^{R}_{chopEMD}(x,m)$ 질문에 대하여 다음과 같이 동작한다.

001 if ($e(x,m)=x'$)
  return x';
002 else if (x=IV_2 and $\exists\, M'$ and M, $e^{*}(M')=chop_{b-n}(m), g(M)=M'\|last_{b-n}(m)$))
  $y=R(M)$;
  choose $w\in_{R}\{0,1\}^{s}$;
  define $e(x,m)=z:=y\|w$;
  return z ;
003 else if ($\exists\, M'$, $e^{*}(M')=x$)
  choose $z\in_{R}\{0,1\}^{n}\setminus C\cup I$;
  define $e(x,m)=z$;
  define $C=C\cup\{z\}$;
  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 ;

3-2. 시뮬레이터 $S_{chopEMD}$에 대한 두 가지 중요한 사항들

질문 횟수의 상한. 라인 003에서 z가 선택되기 위해서는, 시뮬레이터의 총 질문 횟수 q는 $q<2^{n}-2$ 으로 바운드되어야 한다. 만약 $q\geq 2^{n}-2$ 라면, 시뮬레이터가 동작하지 않을 수도 있기 때문이다.

정합성 (Consistency). 라인 003을 보면, z를 선택할 때, 집합 C와 I를 제외한 값으로부터 선택하는 것을 볼 수 있다. C에 포함되지 않는다는 뜻은 부분 함수 e^{*}에 대하여 절대로 충돌쌍이 존재하지 않음을 뜻하고, I에 포함되지 않는다는 뜻은 절대로 기존에 얻은 입출력 값 쌍에서의 그 어떤 입력 값에도 연결이 되지 않음을 의미한다. 이는 chopEMD의 구조적인 특징을 이용하여 (chopEMD,f) 와 (R, $S_{chopEMD}$)를 절대로 구별할 수 없음을 뜻한다. 즉, 둘 사이를 구별하기 위해서는 두 가지 경우에 대해, 각각의 분포 상의 차이점을 보고 구별하는 수밖에 없다.

3-3. chopEMD 해쉬 도메인 확장 방법에 대한 Indifferentiable 안전성 분석

chopEMD의 Indifferentiable 안전도를 기술하기 위하여, 질문의 비용 (cost) 개념을 이용하고자 한다. cost가 q라고 하자. 예를 들어, cost가 q라는 범위 하에서 공격자 A가 질문을 할 경우, O_2에는 q 번의 질문을 하고, O_2에게는 질문을 아예 하지 않을 수도 있다 (여기서 O_1은 해쉬 함수 또는 VIL 랜덤 오라클에 대응되고, O_2은 압축 함수 또는 FIL 랜덤 오라클에 대응된다.). 앞에서 설명한 시뮬레이터 $S_{chopEMD}$에 대한 두 가지 사항들로 인하여, 다음의 소정리가 성립한다.

[소정리 2] $q<2^{n}-2$ 이라 하자. O_1에 대한 질문의 전체 비용 (cost)을 t라고 할 때, $t\leq q$ 라면, O_1의 질문들은 t 개의 O_2 질문으로 변환할 수 있으며, 변

환 시 공격자 A에게 변환하기 전보다, 더 많은 정보를 제공하게 된다.

증명. $S_{chopEMD}$의 정합성에 의하여, O_2과 O_1에 질문을 했을 때, 각각에 대한 답변 사이에 그 어떠한 불일치도 존재하지 않는다. 즉, 구조적인 어떠한 불일치를 발견할 수 없게 된다. 따라서 공격자는 동일한 비용을 가지고 O_1에 질문을 하여 해쉬 함수의 입출력 값만을 얻을 바에야, 차라리 O_2에만 질문을 함으로써 해쉬 함수의 입출력 값 정보 외에 내부 상태 정보까지 얻는 것이 유리하다. 다시 말해, O_1과 O_2를 모두 이용할 때 보다는, O_2에만 질문을 하고 이로부터 얻은 출력 정보만을 가지고 (chopEMD,f)와 (R, $S_{chopEMD}$)를 구별하는 것이 유리함을 뜻한다.

위의 소정리 2는 O_1에는 질문을 하지 않고, O_2에만 질문을 하는 것이 가장 좋은 공격 전략임을 가르쳐 주고 있다. 따라서 임의의 A에 대하여 다음이 성립하는 B가 존재하게 된다.

$$Adv_A((H^f,f),(R,S)) \le Adv_B(f,S),$$

여기서 H^f는 chopEMD를, S는 $S_{chopEMD}$를 가리킨다. 따라서 chopEMD의 indifferentiable 안전성을 분석하기 위하여, 다음의 정리에서 보여주는 것과 같이, f와 $S_{chopEMD}$의 computational distance를 계산하는 데에 초점을 맞추고자 한다.

[정리 2] $q < 2^n - 2$ 는 전체 질문의 횟수, $0 \le s < n$, $f:\{0,1\}^{n+b} \to \{0,1\}^n$일 때, 임의의 A에 대하여 다음이 성립한다. (s 값과 상관없이 안전도의 상한은 동일하다.)

$$Adv_A(f,S_{chopEMD}) \le \frac{q(q+3)}{2^n}.$$

증명. 정리 1에 의해, 우리는 $Stat(f,S_{chopEMD})$의 상한을 계산하는 데에 초점을 맞추면 된다. $Stat(f,S_{chopEMD})$는 결정적 알고리즘 위에서 정의되기 때문에, 오라클이 f인 경우에는 총 가능한 view의 수는 정확하게 2^{nq}가 된다. 그리고 각 view가 발생할 확률은 모두 $1/2^{nq}$가 된다. 이러한 2^{nq} 개의 모든 가능한 view들의 집합을 V_A로 놓는다. 이번엔, 오라클이 $S_{chopEMD}$인 경우를 살펴보자. $S_{chopEMD}$의 정의에 의하여, 가능한 view의 수는 적어도 $(2^n-2)(2^n-3)\cdots(2^n-q)$가 된다. 이러한 최소의 가능한 view들의 집합을 T_S라고 하자. 그리고 T_S의 크기를 r_q라고 쓰자. 증명의 목표는 상한을 구하는 것이기 때문에, T_S 각각의 view이 발생할 확률을 $1/r_q$라고 가정한다. 따라서

$$\begin{aligned}
Stat_A(f,S_{chopEMD}) &= \frac{1}{2}\Sigma_{v\in V_A}|\Pr[f=v]-\Pr[S_{chopEMD}=v]| \\
&= \frac{1}{2}\Sigma_{v\in V_A\setminus T_S}|\Pr[f=v]-\Pr[S_{chopEMD}=v]| \\
&\quad + \frac{1}{2}\Sigma_{v\in T_S}|\Pr[f=v]-\Pr[S_{chopEMD}=v]| \\
&\le \frac{1}{2}\cdot\frac{2^{nq}-r_q}{2^{nq}} + \frac{1}{2}|\frac{r_q}{2^{nq}}-\frac{r_q}{r_q}| \\
&= \frac{1}{2}\cdot(1-\frac{r_q}{2^{nq}}) + \frac{1}{2}\cdot(1-\frac{r_q}{2^{nq}}) \\
&= 1-\frac{r_q}{2^{nq}} = 1-\Pi_{i=1}^{q}(1-\frac{i+1}{2^n}) \\
&\le \Sigma_{i=1}^{q}(\frac{i+1}{2^n}) \\
&= \frac{q(q+3)}{2^n}
\end{aligned}$$

소정리 2와 정리 2로부터, 다음과 같이 chopEMD에 대한 indifferentiable 안전성을 얻을 수 있다.

[따름 정리 1] $q < 2^n - 2$ 는 전체 질문의 비용, $0 \le s < n$, $f:\{0,1\}^{n+b} \to \{0,1\}^n$일 때, 임의의 A에 대하여 다음이 성립한다.

$$Adv_A((chopEMD,f),(R,S_{chopEMD})) \le \frac{q(q+3)}{2^n}.$$

위와 동일한 분석 방법을 이용하여, 다음의 표 1과 같은 결과를 얻을 수 있다. 본 결과를 통해 chop 함수에서의 s 값과 상관없이 각각의 해쉬 구조에 대한 Indifferentiability 관점에서의 안전도의 상한 값이 결정됨을 알 수 있다.

Ⅳ. 결 론

본 논문에서는 Indifferentiability 관점에서 여러 가지 해쉬 구조에 대한 안전성 분석 결과를 제시하였다. 본 논문의 분석 방법은 다양한 해쉬 구조에 대해 일반적으로 적용이 가능하며, SHA-3 후보들에 대한 안전성 분석 시 본 논문에서 기술한 방법을 적용할 수 있다. 한편, 본 논문에서는 랜덤 오라클 모델만 고려를 하였기 때문에, 압축 함수가 이상 블록 암호 모델 (ideal cipher model)에 기반 하여 설계되었을 경우에, 추가적으로 이에 대한 안전성 분석 연구가 요구된다.

표 1. Indifferentiability 안전도 상한 (chop 함수에서 임의의 s 값에 대해 성립함)
Table 1. Indifferentiable Security Upper Bound (for any value of s in chop function, each bound holds.)

해쉬 구조	Indifferentiability 안전도의 상한
choppfMD	$\frac{q(q+1)}{2^n}$
chopMDP	$\frac{q^2}{2^n}$
chopWPH	$\frac{q(q+1)}{2^w}$
chopEMD	$\frac{q(q+3)}{2^n}$
chopNI	$\frac{q(q+1)}{2^n}$
chopCS	$\frac{q(q+3)}{2^n}$

감사의 글

이 논문은 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임. [KRF-2008-357-C00010]

참 고 문 헌

[1] J. H. An and M. Bellare, "Constructing VIL-MACs from FIL-MACs: Message authentication under weakened assumptions", *Crypto 1999, LNCS* 1666, pp. 252-69, 1999.

[2] M. Bellare, R. Canetti and H. Krawczyk, "Keying Hash Functions for Message Authentication", *Crypto 1996,* LNCS 1109, pp. 1-15, 1996.

[3] M. Bellare and T. Ristenpart, "Multi-Property -Preserving Hash Domain Extension and the EMD Transform", *Asiacrypt 2006,* LNCS 4284, pp. 299-14, 2006.

[4] M. Bellare and P. Rogaway, "Random Oracles Are Practical : A Paradigm for Designing Efficient Protocols". *In 1st Conference on Computing and Communications Security,* ACM, pp. 62-73, 1993.

[5] G. Bertoni, J. Daemen, M. Peeters, and G. V. Assche. "On the Indifferentiability of the Sponge Construction", *Eurocrypt 2008*, LNCS 4965, pp. 181-97, 2008.

[6] E. Biham and O. Dunkelman, "A Framework for Iterative Hash Functions - HAIFA", *In The second NIST Hash Workshop,* 2006.

[7] D. Chang, S. Lee, M. Nandi and M. Yung, "Indifferentiable Security Analysis of Popular Hash Functions with Prefix-Free Padding", *Asiacrypt 2006,* LNCS 4284, pp. 283-98, 2006.

[8] D. Chang and M. Nandi, "Improved Indifferentiability Security Proof of chopMD Hash Function", *FSE 2008,* LNCS 5086, pp. 429-43, 2008.

[9] J. S. Coron, Y. Dodis, C. Malinaud and P. Puniya, “Merkle-Damgård Revisited: How to Construct a Hash Function”, *Crypto 2005,* LNCS 3621, pp. 430-448, 2005.

[10] I. B. Damgård, "A Design Principle for Hash Functions," *Crypto 1989*, LNCS 435, pp. 416-427, 1989.

[11] S. Halevi and H. Krawczyk, “Strengthening Digital Signatures via Randomized Hashing”, *Crypto 2006,* LNCS 4117, pp. 41-59, 2006.

[12] S. Hirose, J. H. Park and A. Yun, “A Simple Variant of the Merkle-Damård Scheme with a Permutation”, *Asiacrypt 2007*, LNCS 4833, pp. 113-129, 2007.

[13] S. Lucks, "Design principles for iterated hash functions", *Cryptology ePrint Archive,* Report

2004/253, 2004. http://eprint.iacr.org/.
[14] U. Maurer, R. Renner and C. Holenstein, “Indifferentiability, Impossibility Results on Reductions, and Applications to the Random Oracle Methodology”, *TCC 2004*, LNCS 2951, pp. 21-39, 2004.
[15] U. Maurer and J. Sj¨odin, "Single-key AIL-MACs from any FIL-MAC", *ICALP 2005,* LNCS 3580, pp. 472−84, 2005.
[16] R. C. Merkle, "One Way Hash Functions and DES," *Crypto 1989*, LNCS 435, Springer-Verlag 1989.
[17] NIST Homepage for Hash Project : http://csrc.nist.gov /groups /ST /hash /sha-3/.
[18] NIST, "FIPS 180-1" (superseded by FIPS 180-2). See also NIST's Secure Hashing site.
[19] NIST, "FIPS 180-2: Secure Hash Standard (SHS)", August 2002 (change notice: February 2004). See also NIST's Secure Hashing site.
[20] NIST, "FIPS PUB 186-2: DIGITAL SIGNATURE STANDARD (DSS)", 27 January 2000.
[21] NIST, "FIPS PUB 197: Announcing the ADVANCED ENCRYPTION STANDARD (AES)", 26 November 2001.
[22] NIST, "FIPS PUB 198: The Keyed-Hash Message Authentication Code (HMAC)", 6 March 2002.
[23] 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.
[24] 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.
[25] 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.
[26] R. L. Rivest, "The MD5 Message Digest Algorithm", *RFC 1321* (1992).
[27] X. Wang, H. Yu, "How to Break MD5 and Other Hash Functions", *Eurocrypt 2005,* LNCS 3494, pp. 19-35, 2005.
[28] X. Wang, Y. L. Yin and H. Yu, "Finding Collisions in the Full SHA-1", *Crypto 2005*, LNCS 3621, pp. 17-36, 2005.

장 동 훈 (張東勳)

2001년 2월 : 고려대학교 수학과(이학사)
2003년 2월 : 고려대학교 정보보호대학원 정보보호전공 (공학석사)
2008년 8월 : 고려대학교 정보경영공학전문대학원 정보보호전공 (공학박사)
2008년 11월 : 고려대학교 정보경영공학전문대학원 포닥연구원
2008년 12월 ~ 현재 : 미국 컬럼비아대학 컴퓨터학과 포닥연구원
관심분야 : 대칭키 암호 시스템 및 해쉬 함수에 관한 분석, 설계, 이론 연구.

성 재 철 (成在喆)

1997년 8월 : 고려대학교 수학과 학사
1999년 8월 : 고려대학교 수학과 석사
2002년 8월 : 고려대학교 수학과 박사
2002년 8월 ~ 2004년 1월 : 한국정보보호진흥원 선임연구원
2004년 2월 ~ 현재 : 서울시립대학교 수학과 조교수
관심분야 : 암호 알고리즘 설계 및 분석

홍 석 희 (洪錫喜)

1995년 2월 : 고려대학교 수학과 학사
1997년 2월 : 고려대학교 수학과 석사
2001년 2월 : 고려대학교 수학과 박사
1999년 8월 ~ 2004년 2월 : (주) 시큐리티 테크놀로지스 선임연구원

2004년 4월 ~ 2005년 2월 : K.U.Leuven, ESAT/SCD -COSIC 박사후연구원
2005년 3월 ~ 2008년 8월 : 고려대학교 정보보호대학원 조교수
2008년 9월 ~ 현재 : 고려대학교 정보경영공학전문대학원 부교수
관심분야 : 대칭키 암호 시스템 분석 및 설계, 컴퓨터 포렌식

이 상 진 (李相珍)

1987년 2월 : 고려대학교 수학과 (이학사)
1989년 2월 : 고려대학교 수학과 (석사)
1994년 8월 : 고려대학교 수학과 (박사)
1989년 10월~1999년 2월 : ETRI 연구원 역임
1999년 3월 ~ 현재 : 고려대학교 정교수
1997년 12월 : 국가안전기획부장 표창
관심분야 : 디지털 포렌식, 모바일 포렌식, 심층암호, 해쉬 함수.