Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.26, NO.2, Apr. 2016 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2016.26.2.315

MJH 해쉬 함수 역상 공격*

이 주 영,[1†] 김 종 성[2,3‡]
[1]세종대학교 수학통계학부, [2]국민대학교 수학과, [3]국민대학교 금융정보보안학과

A Preimage Attack on the MJH Hash Function*

Jooyoung Lee,[1†] Jongsung Kim[2,3‡]
[1]Dept. of Mathematics - Applied Statistics, Sejong University,
[2]Dept. of Mathematics, Kookmin University,
[3]Dept. of Financial Information Security , Kookmin University

요 약

본 논문에서 우리는 이중 블록 길이 블록 암호 기반 해쉬 함수 MJH에 대한 새로운 역상 공격을 제안한다. MJH 해쉬 함수가 n 비트 블록 암호에 기반하여 $2n$ 비트를 출력할 때, 기존 공격은 $O(2^{3n/2})$회의 질의를 요구하였으나, 본 논문에서는 이를 더욱 개선시켜 $O(n\,2^n)$의 계산량 및 같은 수준의 메모리를 사용하여 역상을 찾을 수 있음을 보인다.

ABSTRACT

In this paper, we present a new preimage attack on MJH, a double-block-length block cipher-based hash function. Currently, the best attack requires $O(2^{3n/2})$ queries for the $2n$-bit MJH hash function based on an n-bit block cipher, while our attack requires $O(n\,2^n)$ queries and the same amount of memory, significantly improving the query complexity compared to the existing attack.

Keywords: Hash function, Preimage attack, Block ciphers

I. 서 론

MJH 해쉬 함수는 임의의 $\varepsilon > 0$에 대하여, $O(2^{(1-\varepsilon)n})$회의 질의까지 충돌 안전성이 보장된다는 것이 알려져 있다[3]. 따라서 거의 최적의 안전성을 제공하지만, 실제 사용되는 패러미터(예를 들어, $n = 128$)의 경우에는 안전성 상계에 곱해지는 상수의 영향을 무시하지 못하여 만족할 만한 수치를 제공하지는 못할 것으로 보인다.

역상 안전성의 경우, 한 개의 블록암호를 놓고 봤을 때, $O(2^n)$회의 질의까지 역상 안전성이 보장된다는 것은 자명하다. 그러나 그 이상의 안전성에 대해서는 알려진 바가 없다.

한편, MDC-2에 대한 포괄적 충돌쌍 공격[3]이 MJH에 유사하게 적용된다는 것이 알려져 있으며[2], 예를 들어 $n = 128$ 인 경우, 123.81 비트 수준의 공격이 가능하므로, 최적의 안전성 수준과 다소 격차가 있다고 볼 수 있다. 또한 $O(2^{3n/2})$회의 질의로 역상을 찾을 수 있다는 것이 알려져 있으며, 이는 이상적인 $2n$ 비트 해쉬 함수가 $2n$ 비트의 역상 안전성을 제공해야 한다는 점을 감안할 때, 만족스럽지 못한 수준이라 하겠다. 본 과제에서는 이 공격을 더욱 개선

Received(11. 20. 2015), Modified(03. 08. 2016), Accepted(03. 16. 2016)
* 이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2013R1A1A2059864).
† 주저자, jlee05@sejong.ac.kr
‡ 교신저자, jskim@kookmin.ac.kr(Corresponding author)

시켜 $O(n\,2^n)$의 계산량과 같은 수준의 메모리를 요구하는 새로운 역상 공격을 제안한다.

II. MJH 압축 함수

먼저 MJH 해쉬 함수를 일반적인 블록 암호에 기반하는 운용 모드로서 정의하자. 이 운용 모드는 기반 블록 암호의 키 길이에 대하여 유연성을 제공한다. 구체적으로, 임의의 $c \geq 0$ 에 대하여, $n+c$ 비트 키를 사용하는 n 비트 블록 암호 E를 사용할 수 있다. 또한 MJH 해쉬 함수를 정의하기 위해서는 고정점을 갖지 않는 인볼루션(involution) σ 와 2^n개의 원소를 갖는 유한체 $GF(2^n)$의 원소로서 0 또는 1 이 아닌 상수 θ 가 주어져야 한다. 그러면 $2n$ 비트 값을 연쇄 변수로 하여 $n+c$ 비트 블록을 압축하는 함수 $G[\sigma,\theta]$는 다음과 같이 정의된다. u_L, u_R, z_1이 n 비트, z_2가 c 비트 입력값일 때,

$$G[\sigma,\theta]:\{0,1\}^{2n}\times\{0,1\}^{n+c}\rightarrow\{0,1\}^{2n}$$
$$(u_L\|u_R,\ z_1\|z_2)\mapsto(v_L\|v_R)$$

이고, 여기서

$$v_L = E(K,X)+X,$$
$$v_R = \theta\cdot(E(K,\sigma(X))+\sigma(X))+X+z_1,$$

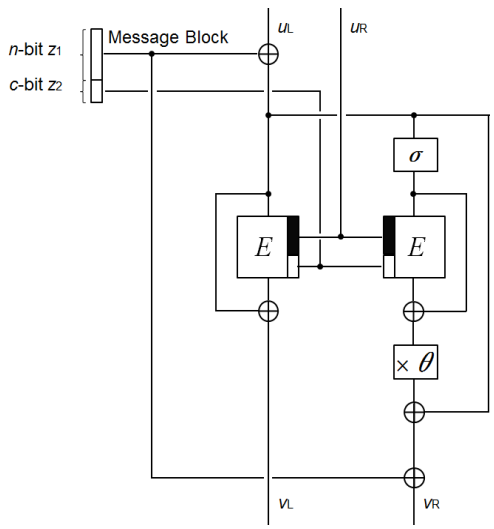

Fig. 1. MJH Compression Function

다시 $X=u_L+z_1$, $K=u_R\|z_2$ 로 정의된다. Fig. 1을 참고하자.

안전성 증명을 위해서는 σ 가 고정점을 갖지 않는 인볼루션이라는 조건으로 충분하지만, 실제 구현 상에서는 0 아닌 상수의 덧셈(xor)로 충분할 것이다. 또한 상수 θ 역시 최대한 간단한 값으로 정하면 되는데, 예를 들어 ($GF(2^n)$를 $GF(2)$ 상에서의 $n-1$ 차 다항식으로 표현하고) $\theta=x$ 로 놓으면 이 상수에 의한 곱셈은 좌측 순환과 조건부 덧셈으로 표현될 것이다. 마지막으로 이 압축 함수는 Merkle-Damgard 방식으로 변환되어, 임의의 길이의 메시지를 입력으로 받는 해쉬 함수 $MJH[\sigma,\theta]$가 된다.

III. 새로운 MJH 해쉬 함수 역상 공격

본 절에서는 MJH 해쉬함수에 대한 새로운 역상 공격을 제안한다. 공격은 다음과 같이 구성된다.

1) 블록암호 E에 $q=O(n\,2^n)$회 질의하여 $i=1,\ldots,q$에 대하여 $E_{k_i}(x_i)=y_i$을 만족하는 질의-응답 쌍

$$Q=((x_1,k_1,y_1),\ldots,(x_q,k_q,y_q))$$

를 얻고, $x_i+y_i=v$를 만족하는 질의를 두 개 찾는다.

2) 목표 해쉬값 $v^*=v_L^*\|v_R^*$이 주어지면,

$$G[\sigma,\theta](u_L^*\|u_R^*,\ z_1\|z_2)=v^*$$

을 만족하는 $u_L^*\|u_R^*$와 $z_1\|z_2$를 찾는다.

3) 중간 일치 기법을 적용하여, 고정된 IV를 사용하는 전체 해쉬함수에 대하여 목표 해쉬값의 역상을 구한다.

이제 이 공격을 분석하기 위하여, 공격의 첫 번째 단계가 성공적으로 수행됨을 증명하겠다. 즉, 이상적 블록암호(ideal block cipher) E에 $q=O(n\,2^n)$ 회 질의하여 질의-응답 쌍 $Q=((x_1,k_1,y_1),\ldots,(x_q,k_q,y_q))$를 얻으면, 임의의 $v\in\{0,1\}^n$에 대하여, $x_i+y_i=v$를 만족하는 질의의 개수가 언제나 두 개 이상임을 보이는 것이다.

먼저 $N=2^n$이라 하고, $m>1$에 대하여 $q=mN$으로 쓰자. 또한 질의 방식을 보다 구체적으로 정의하자. 즉, m개의 서로 다른 키에 대하여, 모든 가능한 평문 $x\in\{0,1\}^n$를 질의한다. 키가 고정된 블록암호는 랜덤 치환함수로 가정할 수 있는데, 고정된 키 k와 고정된 $v\in\{0,1\}^n$에 대하여, $x+E_k(x)=v$를 만족하는 질의가 존재하지 않을 확률은 대략 $1/e$로 근사된다. 이렇게 되는 이유는, 키가 박혀있는 블록암호 E_k를 랜덤 치환함수 P로 놓고, $Q(x)=v+P(x)$라 하면, 위의 확률을 랜덤 치환함수 Q가 고정점을 갖는 확률로 이해할 수 있기 때문이다[1]. 따라서 m개의 서로 다른 키를 모두 사용한 경우에도 $x+E_k(x)=v$를 만족하는 질의가 존재하지 않을 확률은 $1/e^m$으로 근사된다. 비슷하게, $x+E_k(x)=v$를 만족하는 질의가 단 한 개 있을 확률은 m/e^{m-1}보다 크지 않다. 실제로 이 상계는 $x+E_k(x)=v$를 만족하는 질의가 단 한 개의 키에 대해서만 허용되는 경우의 확률에 대한 상계이다. 마지막으로 모든 가능한 $v\in\{0,1\}^n$의 개수는 N이므로, 어떤 $v\in\{0,1\}^n$에 대하여, $x_i+y_i=v$를 만족하는 질의의 개수가 한 개 이하인 확률은

$$N\cdot\left(\frac{1}{e^m}+\frac{m}{e^{m-1}}\right)$$

보다 크지 않다. 이때 $m=2n$이라 놓으면, 위 값은 $O(n/N)$이 되어 무시할 수 있을만큼 작아지게 된다. 다시 얘기하면, 이 확률을 제외하면, 임의의 $v\in\{0,1\}^n$에 대하여, $x_i+y_i=v$를 만족하는 질의의 개수가 언제나 두 개 이상이 된다.

다음으로 공격의 두번째 단계가 구체적으로 수행되는지 살펴보자. 목표 해쉬값 $v^*=v_L^*\|v_R^*$에 대하여

$$G[\sigma,\theta](u_L^*\|u_R^*, z_1\|z_2)=v^*$$

을 만족하는 $u_L^*\|u_R^*$와 $z_1\|z_2$를 찾기 위하여 $K=u_R^*\|z_2$를 랜덤 값으로 고정하고, $v_L^*=E(K,X)+X$을 만족하는 X를 찾자. 그러면, $X=u_L^*+z_1$이고

$$v_R^*=\theta\cdot(E(K,\sigma(X))+\sigma(X))+X+z_1$$

이 되는 u_L^*, z_1을 쉽게 찾을 수 있다. 이 작업에 드는 계산량은 대략 $O(2^n)$이 된다. 그러나 만약 전체 해쉬 함수가 마지막 블록에 메시지의 길이 정보를 포함하는 패딩을 사용한다면, 이 계산량은 다소 늘어나게 된다. 예를 들어, $c=0$, $n=128$로서 각 메시지 블록이 128 비트라 하자. 그리고 마지막 메시지 블록의 끝에 64 비트의 메시지의 비트 길이 정보가 들어가고, 메시지의 끝과 길이 정보를 표현하는 64 비트 값 사이에 이진수열 $10\cdots0$이 채워진다고 하자. 랜덤한 128 비트 블록 M을 64 비트 값 L을 사용하여 $M=M'10\cdots0L$으로 표현할 때, 이 블록이 패딩된 메시지의 맨 마지막 블록이 될 수 있으려면, L이 결정하는 메시지의 길이를 128로 나눈 나머지가 M'의 길이가 되어야 한다. 이 확률은 1/128이 될 것이다. 추가적으로, L이 표현하는 길이가 $n(n+c)$보다 커야 하는데(이유는 뒤에 밝혀지겠지만), 대부분의 패러미터에서 랜덤한 64 비트 값 L이 $n(n+c)$보다 작게 될 확률은 미미하므로 이 확률은 무시해도 좋다. 결론적으로, 목표 해쉬값의 역상으로 맨 마지막 블록을 찾을 때에, 이 블록이 길이 $n(n+c)$ 이상의 메시지의 패딩된 마지막 블록이 될 수 있도록 하려면, 대략 $O((n+c)2^n)$의 계산량이 요구된다 하겠다.

이제 공격의 마지막 단계를 구체적으로 분석하자. 두번째 단계에서 찾은 $u^*=u_L^*\|u_R^*$과 첫 번째 단계에서 기록 및 정렬해 둔 질의-응답 목록 Q를 이용하여, u^*의 역상을 두 개 찾고, 다시 역상의 역상을 두 개씩 찾아서, 결과적으로 (1) u^*를 조상 노드로 하고, (2) 깊이가 n이고, 잎사귀 노드의 개수가 2^n이 되는 이진 균형 트리를 구성한다. Fig. 2를 참조하자.

이제 초기값 $IV=IV_L\|IV_R$을 고정하고 마지막 블록에 포함된 길이 정보에 따라 적당한 개수의 메시지 블록을 임의로 적용하여, 중간 변수값 $u_L\|u_R$을 계산

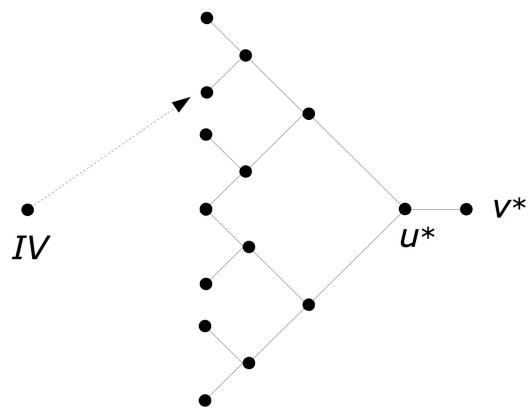

Fig. 2. Binary Tree for Pre-images

한 후에, 여기에 다시 랜덤한 $z_1\|z_2$값을 선택하여 $G[\sigma,\theta](u_L\|u_R, z_1\|z_2)$의 값이 잎사귀 노드 가운데 하나와 같아지도록 한다. 중간 변수의 크기가 $2n$ 비트이므로, 이 단계에서 요구되는 계산량은 $O(2^n)$이 된다.

전체 계산량을 분석하면, 우리는 질의-응답 목록의 생성에 계산량 $O(n\,2^n)$, 저장에 메모리 $O(n\,2^n)$을 사용하여 MJH의 목표 해쉬값에 대한 역상을 찾게 된다.

IV. 결 론

본 논문에서 $O(2^{3n/2})$회의 질의를 사용하는 기존 MJH 해쉬 함수 역상 공격을 개선하여, $O(n\,2^n)$의 계산량 및 이와 같은 수준의 메모리를 요구하는 새로운 역상 공격을 제안하였다. 본고에서 사용한 바와 같이 이진트리를 구성하여 역상을 찾는 공격이 다른 종류의 블록 암호 기반 해쉬 함수에도 적용되는지 살펴보는 것이 향후 흥미로운 연구 주제가 될 것으로 보인다.

References

〔1〕 M. Bóna, "A Walk Through Combinatorics: An Introduction to Enumeration and Graph Theory," 3rd Edition, pp. 134-136, May 2006.

〔2〕 D. Hong and D. Kwon, "Cryptanalysis of Double-Block-Length Hash Modes MDC-4 and MJH," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E97-A, No. 8, pp. 1747-1753, August 2014.

〔3〕 L. R. Knudsen, F. Mendel, C. Rechberger and S. S. Thomsen, "Cryptanalysis of MDC-2," Eurocrypt 2009, LNCS 5479, pp. 106-120, April 2009.

〔4〕 J. Lee and M. Stam, "MJH: a faster alternative to MDC-2," Designs, Codes and Cryptography, Vol. 76. Issue 2. pp. 179-205, August 2015.

〈저 자 소 개〉

이 주 영 (Jooyoung Lee) 정회원
1996년 2월: 서울대학교 수학과 졸업
1998년 2월: 서울대학교 수학과 석사
2005년 10월: University of Waterloo, Combinatorics and Optimization 박사
2005년 11월~2011년 2월: 국가보안기술연구소 선임연구원
2011년 3월~현재: 세종대학교 수학통계학부 교수
〈관심분야〉 암호학

김 종 성 (Jongsung Kim) 종신회원
2000년 8월/2002년 8월: 고려대학교 수학 학사/이학석사
2006년 11월: K.U.Leuven, ESAT/SCD-COSIC 정보보호 공학박사
2007년 2월: 고려대학교 정보보호대학원 공학박사
2007년 3월~2009년 8월: 고려대학교 정보보호기술연구센터 연구교수
2009년 9월~2013년 2월: 경남대학교 e-비즈니스학과 조교수
2013년 3월~현재: 국민대학교 수학과 부교수
2014년 3월~현재: 국민대학교 일반대학원 금융정보보안학과 부교수
〈관심분야〉 정보보호, 암호 알고리즘, 디지털 포렌식