General Distinguishing Attack on NMAC and HMAC with Birthday Attack Complexity

Donghoon Chang[1] and Mridul Nandi[2]

[1] Center for Information Security Technologies(CIST), Korea University, Korea
dhchang@cist.korea.ac.kr
[2] David R. Cheriton School of Computer Science, University of Waterloo, Canada
m2nandi@cs.uwaterloo.ca

Abstract. Kim *et al.* [4] and Contini *et al.* [3] studied on the security of HMAC and NMAC based on HAVAL, MD4, MD5, SHA-0 and SHA-1. Especially, they considered the distinguishing attacks. However, they did not describe a generic distinguishing attack on NMAC and HMAC. In this paper, we describe the generic distinguisher to distinguish NMAC and HMAC when the underlying compression function is the random oracle with the birthday attack complexity.

Keywords : NMAC, HMAC, Distinguishing Attack, Birthday Attack.

1 Introduction.

Since MD4-style hash functions were broken, evaluations on the security of HMAC and NMAC have been required. Kim *et al.* [4] and Contini *et al.* [3] showed the security analyses on them. However, Kim *et al.*' distinguishing attack complexity is far from the birthday attack complexity. Contini *et al.* also suggested 2^{84} as the distinguishing attack complexity of NMAC and HMAC on the reduced SHA-1, which is bigger than the birthday attack complexity. In this paper, we describe the generic distinguisher to distinguish NMAC and HMAC based on the random oracle with the birthday attack complexity.

2 NMAC and HMAC

Fig. 1 and 2 show NMAC and HMAC based on a compression function f from $\{0,1\}^n \times \{0,1\}^b$ to $\{0,1\}^n$. K_1 and K_2 are n bits. $\overline{K} = K||0^{b-n}$ where K is n bits. 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'. $H : \{IV\} \times (\{0,1\}^b)^* \to \{0,1\}^n$ is the iterated hash function. H is defined as follows : $H(IV, x_1||x_2||\cdots||x_t) = f(\cdots f(f(IV, x_1), x_2)\cdots, x_t)$ where x_i is b bits. Let g be a padding method. $g(x) = x||10^t||\mathsf{bin}_{64}(x)$ where t is smallest non-negative integer such that $g(x)$ is a multiple of b. Then, $\mathrm{NMAC}_{K_1,K_2}(x) = H(K_2, g(H(K_1, g(x))))$ and $\mathrm{HMAC}_K(x) = H(IV, g(\overline{K}\oplus\mathsf{opad}||H(IV, g(\overline{K}\oplus\mathsf{ipad}||x))))$.

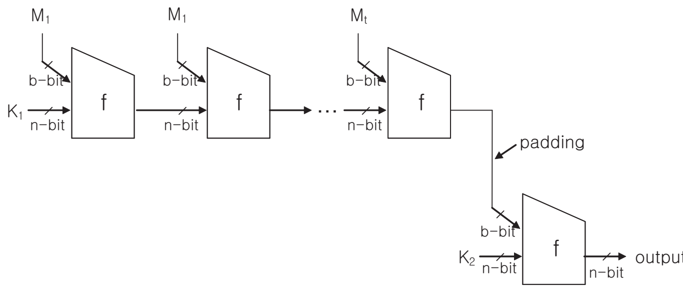

Fig. 1. NMAC

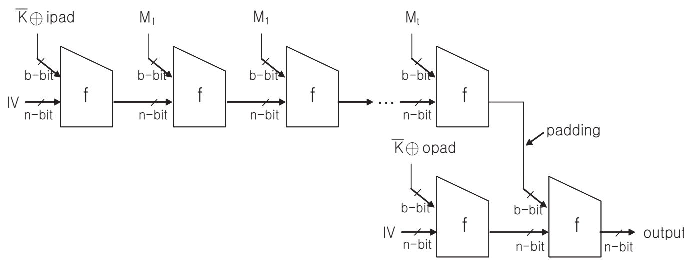

Fig. 2. HMAC

3 General Distinguishing Attack On NMAC and HMAC

Here, we describe two types of distinguishers A_1 and A_2. In case of A_1, we will prove the lower bound of A_1's advantage. On the other hand, A_2 distinguishes heuristically without proving exact proof of security bound. Practically, A_2 is reasonable. For both of distinguishers, queries are same as follows. Let q is the number of queries whose length is fixed and its padded message is l-block ($l \geqslant 3$). Each block is b bits such that $b \geqslant c + 65$ and $c = \lceil \log_2 l \rceil$. $\mathsf{bin}_i(x)$ is the i-bit binary representation of x. q queries are denoted by $M_1, M_2, \cdots, M_q$ such that $g(M_i) = X_i || \mathsf{bin}_c(1) || 0^{b-c} || \mathsf{bin}_c(2) || 0^{b-c} || \cdots || \mathsf{bin}_c(l-1) || 10^{b-c-64} || \mathsf{bin}_{64}(M_i)$ and X_i is b bits and $\{X_1, X_2, \cdots, X_q\} \cap \{1, 2, \cdots, l-1\} = \emptyset$. $\mathsf{Pr}[C_i]$ denotes the probability that for q queries there exist a internal output collision of compression function where i-th block of each query is applied.

Distinguisher A_1

A_1 has an access to oracle $\mathcal{O}$ which is NMAC (or HMAC) or the random function from $\{0,1\}^* \rightarrow \{0,1\}^n$. A_1 makes q queries as described above. Then A_1 outputs '1' if there is a collision among q queries, otherwise outputs '0'. We want to compute the bound of the advantage of A_1. For this, we compute the

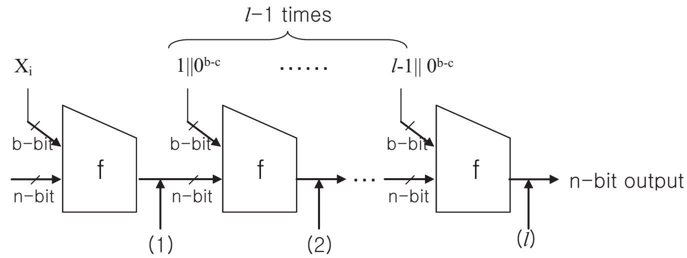

Fig. 3. queries for Distinguishing Attack

probability that there is a collision for both NMAC (or HMAC) and the random function. We denote $\mathsf{Pr}_0[C_i]$ for NMAC or HMAC and $\mathsf{Pr}_1[C]$ for the random function. Let $N = 2^n$. Then $\mathsf{Pr}_0[\neg C_1] = \frac{N(N-1)\cdots(N-q+1)}{N^q}$ because all X_i $(1 \leqslant i \leqslant q)$ are different. Since the first message blocks for q queries are different from the second message blocks for q queries, if there is no collision in (1) of Fig. 3, $\mathsf{Pr}_0[\neg C_1 \wedge \neg C_2] = \mathsf{Pr}_0[\neg C_2|\neg C_2]\mathsf{Pr}_0[\neg C_1] = \frac{N(N-1)\cdots(N-q+1)}{N^q} \cdot \frac{N(N-1)\cdots(N-q+1)}{N^q} = (\frac{N(N-1)\cdots(N-q+1)}{N^q})^2$. Similarly, $\mathsf{Pr}_0[\neg C_1 \wedge \neg C_2 \wedge \cdots \wedge \neg C_l \wedge \neg C_{l+1}] = (\frac{N(N-1)\cdots(N-q+1)}{N^q})^{l+1}$. Since $\mathsf{Pr}_0[C_{l+1}] = 1 - \mathsf{Pr}_0[\neg C_1 \wedge \neg C_2 \wedge \cdots \wedge \neg C_l \wedge \neg C_{l+1}]$, $\mathsf{Pr}_0[C_{l+1}] = 1 - (\frac{N(N-1)\cdots(N-q+1)}{N^q})^{l+1}$. On the other hand, in case of the random function, $\mathsf{Pr}_1[C] = 1 - \frac{N(N-1)\cdots(N-q+1)}{N^q}$. With using $1 - x \leqslant e^{-x}$ for $x \leqslant 1$, $\frac{N(N-1)\cdots(N-q+1)}{N^q} = (1 - \frac{1}{N})(1 - \frac{2}{N}) \cdots (1 - \frac{q-1}{N}) \leqslant e^{\frac{1}{N} + \frac{2}{N} + \cdots + \frac{q-1}{N}} = e^{-\frac{q(q-1)}{2N}}$. If $q \leqslant \sqrt{2N}$ then $\frac{q(q-1)}{2N} \leqslant 1$ [1]. With using $e^{-x} \leqslant 1 - (1 - e^{-1})x$ for $x \leqslant 1$, we know that $e^{-\frac{q(q-1)}{2N}} \leqslant 1 - (1 - e^{-1})\frac{q(q-1)}{2N}$. Since $1 - e^{-1} > 0.632$, $e^{-\frac{q(q-1)}{2N}} < 1 - 0.632 \cdot \frac{q(q-1)}{2N}$. And $\frac{N(N-1)\cdots(N-q+1)}{N^q} \geqslant 1 - \frac{q(q-1)}{2N}$ by the result of [1]. Therefore, $1 - \frac{q(q-1)}{2N} \leqslant \frac{N(N-1)\cdots(N-q+1)}{N^q} < 1 - 0.632 \cdot \frac{q(q-1)}{2N}$.

$$
\begin{aligned}
\mathsf{Adv}_{A_1}(q) &= |\mathsf{Pr}[A_1^{\text{HMAC or NMAC}} = 1] - \mathsf{Pr}[A_1^{\text{Rand}} = 1]| \\
&= |\frac{N(N-1)\cdots(N-q+1)}{N^q} - (\frac{N(N-1)\cdots(N-q+1)}{N^q})^{l+1}| \\
&\geqslant |(1 - \frac{q(q-1)}{2N}) - (1 - 0.632 \cdot \frac{q(q-1)}{2N})^{l+1}|
\end{aligned}
$$

In case of $q = \sqrt{N}$, $\mathsf{Adv}_{A_1}(q) \approx |\frac{1}{2} - 0.684^{l+1}|$. And in case of $l = 11$, $\mathsf{Adv}_{A_1}(q) \approx 0.49$.

Distinguisher A_2

See Fig. 3. We know that there is an internal collision pair in (1) with the following probability.

$$\binom{2^{n/2}}{2} \cdot 2^{-n} = \frac{1}{2} - 2^{(2-n)/2}$$

Then automatically the pair becomes also an internal collision pair in from (2) to (l) in Fig. 3. Except the pair, we also know that there exist an internal collision pair which is collided in (2) with above probability. By this logic, we can get l internal collision pairs in (l). In case of NMAC and HMAC, since the value in (l) is applied to f once more, we can get $(l+1)\cdot(\frac{1}{2}-2^{(2-n)/2})$ collision pairs of NMAC and HMAC on average. On the other hand, in case of random function, we can get only $(\frac{1}{2}-2^{(2-n)/2})$ collision pair on average.

	NMAC or HMAC	Random Function
Average	$(l+1)\cdot(\frac{1}{2}-2^{(2-n)/2})\approx\frac{l+1}{2}$	$(\frac{1}{2}-2^{(2-n)/2})\approx\frac{1}{2}$
Standard Deviation	$\approx\sqrt{2}/2$	$\approx\sqrt{2\cdot(l+1)}/2$

Then, distinguisher A says '1' (NMAC or HMAC) if there are $\frac{l+1}{2}-\sqrt{2(l+1)}$ collision pairs at least. Otherwise A says '0' (random function). So, with high probability A can distinguish NMAC and HMAC from the random function. In case $l=31$, Advantage of A is

$$\begin{aligned}\mathsf{Adv}_A(2^{n/2}) &= |\mathsf{Pr}[A^{\text{NMAC or HMAC}}=1]-\mathsf{Pr}[A^{Rand}=1]| \\ &\approx |0.977-0| = 0.977.\end{aligned}$$

4 Conclusion

In this paper, we described a generic distinguishing attack on NMAC and HMAC where a compression function f is used iteratively and the size of the internal state is same as that of the hash output. Therefore, we can know that the security bound of NMAC and HMAC is the birthday attack complexity in case that the size of the internal state is same as that of the hash output.

References

1. M. Bellare, J. Kilian, and P. Rogaway, *The Security of the Cipher Block Chaining Message Authentication Code*, Appears in Journal of Computer and System Sciences, Vol. 61, No. 3, Dec 2000, pp. 362-399.
2. M. Bellare, *New Proofs for NMAC and HMAC: Security without Collision-Resistance*, Advances in Cryptology - CRYPTO'06, LNCS ??, Springer-Verlag, pp. ??-??, ??.
3. S. Contini and Y. L. Yin, *Forgery and Partial Key-Recovery Attacks on HMAC and NMAC Using Hash Collisions*, Advances in Cryptology - Asiacrypt'06, LNCS 4284, Springer-Verlag, pp. 37-53, 2006.
4. J. Kim, A. Biryukov, B. Preneel, and S. Hong, *On the Security of HMAC and NMAC Based on HAVAL, MD4, MD5, SHA-0 and SHA-1*, SCN'06, to appear. (http://eprint.iacr.org/2006/187).