Attacks On a Double Length Blockcipher-based Hash Proposal

Yiyuan Luo, Xuejia Lai

Department of Computer Science and Engineering,
Shanghai Jiao Tong University
luoyiyuan@sjtu.edu.cn

Abstract. In this paper we attack a $2n$-bit double length hash function proposed by Lee *et al.* This proposal is a blockcipher-based hash function with hash rate 2/3. The designers claimed that it could achieve ideal collision resistance and gave a security proof. However, we find a collision attack with complexity of $\Omega(2^{3n/4})$ and a preimage attack with complexity of $\Omega(2^n)$. Our result shows this construction is much worse than an ideal $2n$-bit hash function.

1 Introduction

Cryptographic hash functions are one of the most important primitives in cryptography [16]. A hash function maps from inputs of arbitrary length to a binary sequence of some fixed length. A hash function usually consists of iteration of a compression function with fixed input and output length. One first designs a fixed domain compression function and then extends the domain to an arbitrary domain by iterating that function.

As flaws in popular classic hash functions MD5 [24] and SHA-1 [1] have been discovered [30,29], NIST has launched a competition for a new hash function standard SHA-3. Many of the popular ideas in the design of hash functions come from the design of block ciphers, either explicitly as for MDC-2 [8] and other schemes [17] or implicity as for MD5. Of the five finalists in the SHA-3 competition, two of them (BLAKE and Skein) are blockcipher-based designs and the other three are permutation-based designs, which are related to blockciphers [20]. Thus, hash functions composed of blockciphers are worthy of study.

We say a compression function is single call or double call depending how many calls it makes to the underlying blockcipher. A blockcipher-based hash function may be a single block length (SBL) function, where the length of the output is equal to that of the blockcipher, or a double block length (DBL) function. where the length of the output is twice that of the blockcipher.

For a typical blockcipher such as AES, the block length is 128 bits, and a hash function with 128-bit output is no longer secure against the birthday attack. Thus, more and more works start to focus on blockcipher-based functions with longer output length [3,5,6,15,18,19,21,27].

For single call DBL blockcipher-based hash functions, Lucks [15] first proposed a collision resistant single call DBL blockcipher-based hash function in the iteration. Later, Stam [26] proposed a single call rate-1 DBL blockcipher-based supercharged compression that is opimally collision resistant up to a logarithmic factor. Their construction give ideal collision resistance but not ideal preimage resistance. Although Lucks and Stam claimed their construction has rate-1, their constructions are much slower than the real rate-1 compression functions in practice due to the computation of polynomial multiplication.

For double call DBL hash functions, Knudsen *et al.* [9] discussed the security of DBL hash functions with rate 1 based on (n, n) blockciphers. Hohl *et al.* [7] discussed the security of compression functions of DBL hash functions with rate 1/2. Satoh *et al.* [25] and Hattori *et al.* [4] and Hirose [5,6] discussed DBL hash functions with rate 1 based on $(2n, n)$ blockciphers.

Nandi *et al.* proposed a rate-2/3 DBL compression function which later was attacked by Knudsen *et al.* [10]. In [22], Peyrin *et al.* gave a general analysis of combining smaller compression functions to build a larger compression function. Fleischmann *et al.* [3,2] address the collision resistance of two old DBL constructions known as Abreast-DM and Tandem-DM [12,11], later their proof of Tandem-DM was revised by Lee *et al.* [14]. In [21], Özen and Stam proposed a novel framework for DBL blockcipher-based hash functions.

In [13], Lee *et al.* proposed another rate-2/3 DBL construction using a Feistel structure. They build a $(2n, 2n)$-blockcipher E^* with 3-round Feistel structure from a $(2n, n)$-blockcipher E, and then embed E^* in PGV compression function, such as the Davies-Meyer structure. They proved the ideal collision resistance in the ideal cipher model, that is, the advantage of a adversary makes q queries to the underlying blockcipher is upper bounded by $\Omega(q^2/2^{2n})$. Thus, the strength bound of this proposal against a collision-finding attack is $\Omega(2^n)$. Compare with other proposals, the authors claimed that it is the most efficient DBL compression function with ideal collision resistance.

However, in this paper, we find a $2^{3n/4}$ collision attack and a 2^n preimage attack on this construction. Thus it contradicts Lee *et al.*'s security proof. Our result shows that it is still an open problem to build ideal collision and preimage resistant DBL blockcipher-based hash functions with rate larger than 1/2.

2 Preliminaries

2.1 Iterated Hash Functions

A hash function $H : \{0,1\}^* \to \{0,1\}^a$ usually consists of a compression function $F : \{0,1\}^a \times \{0,1\}^b \to \{0,1\}^a$ and an initial value $IV \in \{0,1\}^a$. An input M is divided into the b-bit blocks $m_1, m_2, \ldots, m_l$, if the length of M is not a multiple of b, M is padded using an unambiguous padding rule. Then, $h_i = F(h_{i-1}, m_i)$ is computed successively for $1 \leq i \leq l$ and $h_l = H(M)$. Thus H is called an iterated hash function. We use Merkle-Damgård padding in this paper. The hash function H should have the following properties:

Preimage resistance For a given output y, it is intractable to find an input x such that $y = H(x)$.

Second-preimage resistance For a given input x, it is intractable to find an input $x' \neq x$ such that $H(x) = H(x')$.
Collision resistance It is intractable to find a pair of inputs x and x' such that $H(x) = H(x')$ and $x \neq x'$.

2.2 Ideal Cipher Model.

The ideal cipher model, also called the black box model, is a formal model for the security analysis of blockcipher-based hash functions. An ideal cipher is an ideal primitive that models a random block-cipher $E : \{0,1\}^k \times \{0,1\}^n \mapsto \{0,1\}^n$. Each key $k \in \{0,1\}^k$ defines a random permutation $E_k = E(k, \cdot)$ on $\{0,1\}^n$. An adversary is given forward or inverse queries to oracles E, when he makes a forward query to E with $(+, k, p)$, it returns the point c such that $E_k(p) = c$, when he makes an inverse query to E with $(-, k, c)$, it returns the point p such that $E_k(p) = c$.

Without loss of generality, it is assumed that any adversary with forward and inverse queries asks only once on a triplet of a key, a plaintext and a ciphertext obtained by a query and a corresponding answer and there are no redundant queries.

2.3 Double-Block-Length Hash Function

Definition 1. *Let F be a compression function composed of block ciphers, m the number of message blocks in terms of the block length of the underlying blockcipher, and N the number of cipher calls in F. Then the efficiency rate r defined below is an index of efficiency:*

$$r = \frac{m}{N}.$$

The original definition of hash rate is in [9]. We realized that this definition is only related to the efficiency of the hash. It has no relationship to the key length of the underlying blockcipher. We can modify it to a more accurate definition we called security rate:

Definition 2. *Let F be a compression function composed of blockciphers, m the number of message blocks in terms of the block length of the underlying blockcipher, N the number of cipher calls in F, K the key length of the blockcipher and L the output length of F. Then the security rate R defined below is an index of security:*

$$R = \frac{m \cdot L}{N \cdot K}.$$

The security rate of a compression function F can be seen as an index of the security of the function. Its security is related to the input and output length of F, the key length of the underlying blockciphers and the number of cipher calls.

This definition is more general than the efficiency rate. The security rate of a classical Davies-Meyer compression function [23] based on a (n, n) blockcipher is 1, and the security rate will still be 1 even it is based on a $(2n, n)$ blockcipher. This definition can also be applied to DBL blockcipher-based hash functions and thus reduces the complexity of classification of blockcipher-based hash functions. For DBL hash functions based on $(2n, n)$ blockciphers, the efficiency rate is the same as the security rate since $L = K = 2$ in the definition 2. In the remaining part of this paper we use R to denote the security rate and r to denote the efficiency rate.

3 Lee *et al.*'s Proposal

In [13], Lee *et al.* first designed a DBL cipher with 3-round Feistel structure using a blockcipher, then the cipher is embedded into a PGV-style compression function. Without loss of generality, they first considered the Davies-Meyer construction and proved its collision resistance. Then they claimed this proof can be extended to other constructions in a similar way. Thus we only need to consider the Davies-Meyer construction.

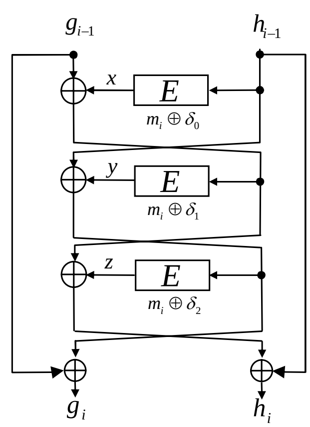

Fig. 1. Lee *et al.*'s Rate$-2/3$ proposal.

Definition 3 (Lee *et al.*'s Proposal). *Let* $E : \{0,1\}^{2n} \times \{0,1\}^n \rightarrow \{0,1\}^n$ *be a blockcipher. Let* $\delta_0, \delta_1, \delta_2$ *are distinct constants in* $\{0,1\}^{2n}$*. The compression function* $F : \{0,1\}^{2n} \times \{0,1\}^{2n} \rightarrow \{0,1\}^{2n}$ *is written as* $(g_i, h_i) = F(g_{i-1}, h_{i-1}, m_i)$*.*

Let x, y, z satisfy the following equations:

$$x = E_{m_i \oplus \delta_0}(h_{i-1})$$
$$y = E_{m_i \oplus \delta_1}(g_{i-1} \oplus x)$$
$$z = E_{m_i \oplus \delta_2}(h_{i-1} \oplus y)$$

Then the output of the compression function (g_i, h_i) is:

$$g_i = g_{i-1} \oplus y \oplus h_{i-1}$$
$$h_i = h_{i-1} \oplus x \oplus z \oplus g_{i-1}$$

The compression function is depicted in Fig. 1

4 The Security of the Construction

Lee *et al.* proved that the collision resistance of this construction can achieve an ideal security bound. That is, to find a collision in F with high probability, the adversary needs almost $\Omega(2^n)$ queries to the underlying blockcipher. They stated the following theorem.

Theorem 1. *Let F be the above compression function, then in the ideal cipher model, for any $q, n \geq 1$, the advantage of an adversary queries q times is*

$$\mathbf{Adv}^F(q) \leq \frac{(q-2)\cdot(q-3)}{2} \cdot (\frac{1}{2^n - 1})^2 \approx \Omega(\frac{q^2}{2^{2n}}).$$

We find a collision attack with complexity about $\Omega(2^{3n/4})$ and a preimage attack with complexity about $\Omega(2^n)$, thus we disprove Lee *et al.*'s conclusion.

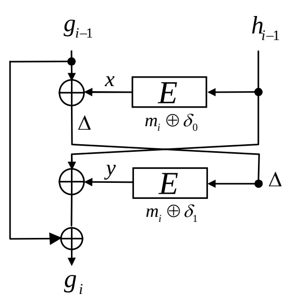

Fig. 2. Attack on the left half.

Theorem 2. *Let F be the above compression function, then there exist a collision attack with complexity of about $2 \times 2^{3n/4}$ queries and a preimage attack with complexity of about 3×2^n queries to the underlying blockcipher E.*

Proof. The idea is to first attack on the left half of the construction, which is shown is Fig. 2. We first construct a set of inputs $\{(m_i, g_{i-1}, h_{i-1})\}$ all hitting the same value of g_i, then we attack the right half of the construction.

– Collision attack:
  1. Set m_i to a constant.
  2. Choose $2^{3n/4}$ random distinct values of h_{i-1} and compute the corresponding ciphertext x. Since E is an ideal cipher, we thus get $2^{3n/4}$ distinct random pairs of (h_{i-1}, x).
  3. Choose $2^{3n/4}$ random distinct values of Δ and compute the corresponding ciphertext y. Since E is an ideal cipher, we thus get $2^{3n/4}$ distinct random pairs of (Δ, y).
  4. $g_{i-1} = x \oplus \Delta$ and $g_i = h_{i-1} \oplus x \oplus \Delta \oplus y$, since there are $2^{3n/4}$ pairs of (h_{i-1}, x) and $2^{3n/4}$ values of (Δ, y). Using Wagner's join technology [28], with complexity $\Omega(2^{3n/4})$ we can find
  $$\frac{2^{3n/4} \times 2^{3n/4}}{2^n} = 2^{n/2}$$
  values of (g_{i-1}, h_{i-1}) all hitting the same value of g_i.
  5. Since $h_i = h_{i-1} \oplus z \oplus \Delta$, according to the birthday paradox, given $2^{n/2}$ random (g_{i-1}, h_{i-1}, m_i), there exists two pairs colliding at h_i with probability 0.39.
  6. The adversary needs $2 \times 2^{3n/4} + 2^{n/2}$ queries to the blockcipher E and the total complexity is about $3 \times 2^{3n/4}$.

– Preimage attack:
  1. Given the image (g_i, h_i), set m_i to a constant.
  2. Choose 2^n random distinct values of h_{i-1} and compute the corresponding ciphertext x. Since E is an ideal cipher, we thus get 2^n distinct random pairs of (h_{i-1}, x).
  3. Choose 2^n random distinct values of Δ and compute the corresponding ciphertext y. Since E is an ideal cipher, we thus get 2^n distinct random pairs of (Δ, y).
  4. $g_i = h_{i-1} \oplus x \oplus \Delta \oplus y$, since there are 2^n pairs of (h_{i-1}, x) and 2^n values of (Δ, y). Using Wagner's join technology, with complexity $\Omega(2^n)$ we can find
  $$\frac{2^n \times 2^n}{2^n} = 2^n$$
  values of (g_{i-1}, h_{i-1}) all hitting the given value g_i.
  5. Since $h_i = h_{i-1} \oplus z \oplus \Delta$, with high probability, there exists a pair (g_{i-1}, h_{i-1}) hitting at the image h_i.
  6. The adversary needs 3×2^n queries to the blockcipher E and the total complexity is about 4×2^n.

□

In the above we show that the collision resistance and preimage resistance of this compression function are much worse than an ideal $2n$-bit compression function. If we iterate this compression function and fix the initial value, we can also give a $\Omega(2^n)$ preimage attack by using the meet-in-the-middle attack. However, currently we cannot find an efficient collision attack using the above technology, thus we leave this as an open problem.

Although we only consider the Davies-Meyer construction, our attack can also be applied when the other 11 PGV-styles are used. We omit the details here since the attacks are similar.

5 Conclusion

In this paper we have investigated the security of a DBL blockcipher-based hash function proposed by Lee *et al.* They first extended an $(2n, n)$ blockcipher to a $(2n, 2n)$ blockcipher by using 3-round Feistel structure, then embedded this blockcipher into a PGV-style construction, such as Davies-Meyer. We find collision attacks and preimage attacks that contradict their security proofs; we show that the security level of this construction is much worse than an ideal $2n$-bit compression function.

Our result shows that it is still an open question whether an ideal collision resistant and preimage resistant DBL blockcipher-based compression function with hash rate larger than 1/2 exists.

References

1. FIPS. FIPS 180-1 Secure Hash Standard. Federal Information Processing Standard (FIPS), Publication 180-1, National Institute of Standards and Technology, US Department of Commerce, Washington D.C, 1995.
2. E. Fleischmann, M. Gorski, and S. Lucks. Security of cyclic double block length hash functions. In *Cryptography and Coding 2009*, volume LNCS 5921, pages 153–175. Springer-Verlag, 2009.
3. Ewan Fleischmann, Michael Gorski, and Stefan Lucks. On the security of tandem-DM. volume 5665 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 84–103, Leuven, Belgium, 2009. Springer Verlag.
4. M. Hattori, S. Hirose, and S. Yoshida. Analysis of double block length hash functions. *Cryptography and Coding, Proceedings*, 2898:290–302, 2003.
5. S. Hirose. A security analysis of double-block-length hash functions with the rate 1. *Ieice Transactions on Fundamentals of Electronics Communications and Computer Sciences*, E89A(10):2575–2582, 2006.
6. S. Hirose. Some plausible constructions of double-block-length hash functions. In *Fast Software Encryption*, volume LNCS 4047, pages 210–225, 2006.
7. Walter Hohl, Xuejia Lai, Thomas Meier, and Christian Waldvogel. Security of iterated hash functions based on block ciphers. In *Advances in Cryptology - CRYPTO'93*, volume LNCS 773, pages 379–379, Santa Barbara, CA, United states, 1994. Springer-Verlag.

8. ISO. ISO/IEC 10118 Information technology - Security techniques - Hash-functions, 1994.
9. L. R. Knudsen, X. J. Lai, and B. Preneel. Attacks on fast double block length hash functions. *Journal of Cryptology*, 11(1):59–72, 1998.
10. Lars R. Knudsen and Frederic Muller. Some attacks against a double length hash proposal. In *ASIACRYPT 2005*, pages 462–473, 2005.
11. X. Lai. *On the design and security of block ciphers*, volume 1 of *ETH Series in Information Processing*. Hartung-Gorre Verlag, Konstanz, 1992.
12. X. Lai and J. L. Massey. Hash functions based on block ciphers. In R. A. Rueppel, editor, *Advances in Cryptography-Eurocrypt'92*, volume LNCS 658, pages 55–70. Springer-Verlag, 1992.
13. J. Lee, S. Hong, J. Sung, and H. Park. A new double-block-length hash function using feistel structure. In J. H. Park et al., editor, *ISA 2009*, volume LNCS 5576, pages 11–20, 2009.
14. J. Lee and J. Steinberger. Multi-property-preserving domain extension using polynomial-based modes of operation. In *Advances in Cryptology - EUROCRYPT'10*, volume LNCS 6110, pages 573–596, French Riviera, France, 2010. Springer-Verlag.
15. S. Lucks. A collision-resistant rate-1 double-block-length hash function. In *Symmetric Cryptography, number 07021 in Dagstuhl Seminar Proceedings*, Dagstuhl, Germany, 2007. Internationales Begegnungs - und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
16. A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1997.
17. R. C. Merkle. One way hash functions and DES. In *Advances in Cryptology - CRYPTO'89*, volume LNCS 435, pages 428–446. Springer-Verlag, 1989.
18. M. Nandi. Towards optimal double-length hash functions. In *INDOCRYPT'05*, volume LNCS 3797, pages 77–89. Springer-Verlag, 2005.
19. M. Nandi, W. Lee, K. Sakurai, and S. Lee. Security analysis of a 2/3-rate double length compression function in the black-box model. In *Fast Software Encryption - FSE'2005*, volume LNCS 3557, pages 243–254. Springer-Verlag, 2005.
20. NIST. Third (final) round candidates, 2010. http://csrc.nist.gov/groups/ST/hash/sha-3/Round3/submissions-rnd3.html.
21. O. Özen and M. Stam. Another glance at double-length hashing. In *Cryptography and Coding, 12th IMA International Conference, Cryptography and Coding 2009*, volume LNCS 5921, pages 176–201. Springer-Verlag, Berlin, 2009.
22. T. Peyrin, H. Gilbert, F. Muller, and M. Robshaw. Combining compression functions and block cipher-based hash functions. *Advances in Cryptology - ASIACRYPT 2006*, 4284:315–331 468, 2006.
23. B. Preneel, R. Govaerts, and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. In D.R. Stinson, editor, *Advances in Cryptology - Proc. Crypto'93*, volume LNCS 773, pages 368–378. Springer-Verlag, Berlin, 1993.
24. R. L. Rivest. The MD5 message digest algorithm. In *Request for Comments (RFC) 1321*. Internet Activities Board, Internet Privacy Task Force, 1992.
25. Takashi Satoh, Mio Haga, and Kaoru Kurosawa. Towards secure and fast hash functions. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E82-A(1):55–62, 1999.
26. M. Stam. Block cipher based hashing revisited. In *Fast Software Encryption 2009*, volume LNCS 5665, pages 67–83. Springer, Berlin, 2009.

27. John P. Steinberger. The collision intractability of MDC-2 in the ideal-cipher model. In *Advances in Cryptology-Proceedings of EUROCRYPT 2007*, volume LNCS 4515 of *Lecture Notes in Computer Science*, pages 34–51, Barcelona, Spain, 2007. Springer Verlag, Berlin.
28. D. Wagner. A generalized birthday problem. In M. Yung, editor, *CRYPTO 2002*, volume LNCS 2442, pages 288–303. Springer, 2002.
29. Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full SHA-1. In Victor Shoup, editor, *Advances in Cryptology - CRPTO'05*, volume LNCS 3621, pages 17–36, Santa Barbara, CA, USA, 2005. Springer-Verlag.
30. Xiaoyun Wang and Hongbo Yu. How to break MD5 and other hash functions. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT'05*, volume LNCS 3494, pages 19–35, Aarhus, Denmark, 2005. Springer-Verlag.