IEICE TRANS. FUNDAMENTALS, VOL.E88–A, NO.12 DECEMBER 2005

LETTER

Complexity of Differential Attacks on SHA-0 with Various Message Schedules

Mitsuhiro HATTORI[†*], **Shoichi HIROSE**[††a)], ***Members, and*** **Susumu YOSHIDA**[†], ***Fellow***

SUMMARY The security of SHA-0 with various message schedules is discussed in this letter. SHA-0 employs a primitive polynomial of degree 16 over GF(2) in its message schedule. For each primitive polynomial, a SHA-0 variant can be constructed. The collision resistance and the near-collision resistance of SHA-0 variants to the Chabaud-Joux attack are evaluated. Moreover, the near-collision resistance of a variant to the Biham-Chen attack is evaluated. It is shown that the selection of primitive polynomials highly affects the resistance. However, it is concluded that these SHA-0 variants are not appropriate for making SHA-0 secure.

key words: *hash function, SHA-0, Chabaud-Joux attack, Biham-Chen attack*

1. Introduction

SHA-0 is a hash function proposed by NIST (National Institute for Standards and Technology) in 1993 [1]. It produces a 160-bit hash value, thus the brute-force collision or near-collision attack requires $O(2^{80})$ computation. The first effective collision attack was proposed by Chabaud and Joux in 1998 [2]. The complexity of their attack is 2^{61}. In 2004, Biham and Chen proposed a method to reduce the complexity of the Chabaud-Joux attack using "2-neutral set" as a tool [3]. They reduced the complexity to 2^{56}. They also showed that near-collisions can be used for finding collisions. They found near-collisions of 18-bit differences with complexity 2^{40}. Wang et al. found a collision attack on SHA-0 whose complexity is 2^{40} [4]. Collisions in SHA-0 have finally been found in [5]. In this letter, the Chabaud-Joux attack and the Biham-Chen attack are discussed.

SHA-1, which is a replacement of SHA-0, was proposed in 1995 [6]. Recently, however, an effective collision attack is proposed by Wang et al. which can find a collision with complexity 2^{69} [7]. Therefore, we need other approaches for making SHA-0 secure. This letter considers one of them.

SHA-0 has a process called a message schedule in its compression function. This process is defined as a recurrence formula. This formula is derived from a primitive polynomial of degree 16 over GF(2). In fact, there are 2048 primitive polynomials and NIST adopted one of them. It has not been made public why this polynomial is adopted and it is still unclear how appropriate the selection is. We analyze it in the first part of this letter.

We deal with 2048 SHA-0 variants each of which employs one of 2048 primitive polynomials. We estimate their collision resistance and near-collision resistance to the Chabaud-Joux attack. We first estimate collision resistance and reveal 412 variants are totally resistant to the Chabaud-Joux attack. We also find a collision in the most vulnerable variant using the original Chabaud-Joux attack without any improvements. The complexity is estimated at about 2^{45}. We then estimate near-collision resistance of the SHA-0 variants, where we deal only with the near-collisions which can be used as a tool to find collisions. We find no variants are totally resistant to the Chabaud-Joux near-collision attack. Therefore, all the variants could be collision-attacked by using such near-collisions as a tool. The complexity of the most resistant one is 2^{69}, while that of the most vulnerable one is 2^{35}. Accordingly, it is concluded that the replacement of the primitive polynomial is not a proper way to make SHA-0 secure.

Then, we consider the Biham-Chen attack. We apply the attack to one of the variants which are the most near-collision resistant ones. Our result shows that the attack does not work well on such variants.

2. The Chabaud-Joux Attack and the Biham-Chen Attack

This section briefly describes the algorithm of the Chabaud-Joux attack and the Biham-Chen attack.

2.1 The Chabaud-Joux Attack

The main idea of this attack is to flip several bits in a message and adjust the influences caused by the flips. Let $m = (m_0, \ldots, m_{79})$ be an 80-bit sequence. m is called a *mask base*. Let W_i be the i-th word of the expanded message. If m_i is 1, bit 1 of W_i is flipped, where bit 1 is the bit to the left of LSB. The influences of this flip is adjusted within the succeeding 5 rounds with some probabilities. Therefore, $m_{75}, \ldots, m_{79}$ must be 0, otherwise the influences cannot be adjusted completely. The probability of succeeding in the adjustment is calculated based on the bit-pattern of m. Refer to [2] for details.

Manuscript received April 4, 2005.
Manuscript revised August 15, 2005.
Final manuscript received August 23, 2005.

[†]The authors are with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 606-8501 Japan.
[††]The author is with the Faculty of Engineering, The University of Fukui, Fukui-shi, 910-8507 Japan.
[*]Presently, with Information Technology R&D Center, Mitsubishi Electric Corporation.
a) E-mail: hirose@fuee.fukui-u.ac.jp

DOI: 10.1093/ietfec/e88–a.12.3668

2.2 The Biham-Chen Attack

This attack reduces the complexity needed for the Chabaud-Joux attack by using "2-neutral set." Suppose that we have a message pair (M, M') whose adjustments are succeeding until r rounds (i.e. a message pair such that the differences of the five registers are as expected until r rounds). In this case, it is said that the *conforming round* of the message pair is r. Biham and Chen found that even if several bits of both M and M' are flipped, the conforming round of them is still r in some cases. The set of such bits of a message pair whose conforming round is r is called a *2-neutral set* N_r (in fact, more conditions are needed). Let $p(r_1, r_2)$ be the probability for adjustments to succeed from r_1-th round to r_2-th round. If we can find a 2-neutral set such that $2^{|N_r|} > 1/p(r, r_2)$, we can start the attack from r-th round and thus disregard the prior r rounds. Refer to [3] for details.

3. The Chabaud-Joux Attack on SHA-0 Variants

3.1 Definition of the SHA-0 Variants

We define the SHA-0 variants. The original SHA-0 employs a message schedule such as

$$W_i = W_{i-3} \oplus W_{i-8} \oplus W_{i-14} \oplus W_{i-16}, \quad i = 16, \ldots, 79. \quad (1)$$

Equation (1) is represented by the following polynomial of degree 16 over GF(2):

$$p(x) = x^{16} + x^{13} + x^8 + x^2 + 1\ . \quad (2)$$

$p(x)$ is, in fact, one of the primitive polynomials. The number of primitive polynomials of degree 16 over GF(2) is $\phi(2^{16} - 1)/16 = 2048$, where ϕ is Euler's totient function. We define the SHA-0 variants as the 2048 SHA-0 instances each of which employs one of the 2048 primitive polynomials in its message schedule. These include the original SHA-0.

3.2 Complexity of the Collision Attack

We evaluate the resistance of the SHA-0 variants to the Chabaud-Joux attack in this section.

3.2.1 Method

For each of 2048 SHA-0 variants, the complexity of the attack is calculated as follows.

1. Calculate all the mask base candidates. These candidates must satisfy $m_i = 0$ for $i \in \{75, \ldots, 79\}$.
2. For each mask candidate, compute the probability of succeeding in the attack.
3. Select a mask base which has the maximum success probability and compute the complexity of the attack.

In Step 2 of the above procedure, each success probability is calculated simply by multiplying the success probabilities of the adjustments of influences caused by the flips following the mask base. This is the same way that is presented in [2]. In Step 3, the complexity is simply calculated by taking the inverse number of the probability.

3.2.2 Results

We calculated the computational complexity of the attack for all the 2048 SHA-0 variants. The result is shown in Table 1.

The complexity of the attack on the original SHA-0 is 2^{68}. Actually, in [2], it is stated that the complexity is reduced to 2^{61} for the original SHA-0 by some improvements. However, the result in Table 1 shows computational complexity of the attack without any improvements for each SHA-0 variant.

The maximum complexity represented by 2^{∞} in Table 1 means that no mask base is usable for the attack. This is because each mask base has a pair $(m_i, m_{i+1}) = (1, 1)$ for some $i \in \{0, \ldots, 15\}$. 412 variants are totally resistant to the attack.

The number of variants which are not totally resistant but more resistant to the attack than the original one is 817,

Table 1 Computational complexity needed for the Chabaud-Joux collision attack on the SHA-0 variants.

Complexity	Number of variants	Sum
2^{∞}	412	412
2^{101}	1	413
2^{100}	0	413
2^{99}	0	413
2^{98}	0	413
2^{97}	0	413
2^{96}	0	413
2^{95}	0	413
2^{94}	0	413
2^{93}	1	414
2^{92}	1	415
2^{91}	1	416
2^{90}	3	419
2^{89}	4	423
2^{88}	7	430
2^{87}	7	437
2^{86}	9	446
2^{85}	10	456
2^{84}	12	468
2^{83}	12	480
2^{82}	20	500
2^{81}	22	522
2^{80}	29	551
2^{79}	34	585
2^{78}	36	621
2^{77}	47	668
2^{76}	44	712
2^{75}	70	782
2^{74}	57	839
2^{73}	65	904
2^{72}	78	982
2^{71}	90	1072
2^{70}	79	1151
2^{69}	79	1230
2^{68}	84	1314
2^{67}	81	1395
2^{66}	79	1474
2^{65}	80	1554
2^{64}	68	1622
2^{63}	73	1695
2^{62}	59	1754
2^{61}	52	1806
2^{60}	49	1855
2^{59}	41	1896
2^{58}	31	1927
2^{57}	27	1954
2^{56}	24	1978
2^{55}	16	1994
2^{54}	11	2005
2^{53}	11	2016
2^{52}	11	2027
2^{51}	5	2032
2^{50}	3	2035
2^{49}	5	2040
2^{48}	5	2045
2^{47}	1	2046
2^{46}	1	2047
2^{45}	1	2048

and the number of variants which are as resistant as or more vulnerable to the attack is 818. The original one is therefore weaker one among 2048 variants. The complexity distributes from 2^{45} to 2^{∞}. Thus, the selection of primitive polynomials highly affects the resistance.

The most vulnerable variant employs the polynomial $x^{16}+x^{14}+x^{12}+x^{10}+x^8+x^7+x^6+x^5+x^4+x^2+1$. The complexity is 2^{45}. We implemented the collision attack on this variant. We used 16 computers each of which has a Pentium4 3.06 GHz CPU. We found a collision of this variant after two days of computation;

```
af3bd4fd ada5ead4 5d1da03d ccf3db2e
b3887606 5e068898 40c134dc 263294bf
cbf91ff0 1b997ed2 92118e2f fae0fd89
846f3d48 a53fafd5 cb3a1deb 5f46b7b0
```

and

```
af3bd4fd ada5ead4 5d1da03d ccf3db2e
b3887606 5e06889a 40c1349c 263294bd
4bf91ff0 9b997ed0 12118e6f fae0fd8b
046f3d48 253fafd5 4b3a1de9 5f46b7f0
```

give the same hash value

```
e930e9be 464527bf e489b780 a3314c11 fcb6ea88.
```

3.3 Complexity of the Near-Collision Attack

In this section, we evaluate the resistance of SHA-0 variants to the Chabaud-Joux near-collision attack.

3.3.1 The Near-Collisions in This Letter

In this letter, the restricted near-collisions are treated. This is because we are concerned with the near-collisions which can be used for finding collisions.

In [8], Biham and Chen stated that a collision is obtained by using near-collisions. They considered an attack to the cascade of compression functions. For example, suppose that a collision attack is made on the cascade of N compression functions. At the first compression function, an adversary finds a near-collision. Then, at the second compression function, the adversary finds another near-collision. In this way, the adversary finds a near-collision at each compression function from the first to the $(N-1)$-th. Finally, at the N-th compression function, the adversary obtains a collision.

Biham and Chen showed several collisions of 50-round-reduced SHA-0 which are obtained by this attack. However, the details have been unpublished. We examined them and found that only such near-collisions were used for finding collisions as the differences existed only on bit 1 of A_{80}, B_{80} and C_{80} and bit 31 of D_{80} and E_{80}. These are the bits on which adjustments are made. These differences can be adjusted within several steps at the next compression function in the scheme of the Chabaud-Joux and Biham-Chen attack. If there exists any difference on other than these bits, it cannot be adjusted at the next compression function and a collision cannot be obtained. Our consideration is supported by the fact that the one-bit difference of the near-collision shown as an example in [3], [8] is on bit 1 of B_{80}.

Thus, in the remaining part, we are concerned with 155-bit near-collisions except for the 5-bit differences in bit 1 of A_{80}, B_{80}, C_{80} and bit 31 of D_{80} and E_{80}.

3.3.2 Method

The complexity of the Chabaud-Joux near-collision attack is calculated in the same way as that of the collision attack. In the near-collision attack, the constraint on a mask m is cancelled that $m_{75},\ldots,m_{79}$ must be 0. If m_i $(i=75,\ldots,79)$ is 1, the adjustment of the influences of the flip will not be completed, which results in one-bit difference in one of the 5 bits we mentioned. Thus, this near-collision can be used for the collision attack.

3.3.3 Results

We calculated the computational complexity of the near-collision attack for all the 2048 SHA-0 variants. The result is shown in Table 2. The complexity of the birthday attack is $2^{77.5}$ because 155 bits out of 160 bits must collide. Therefore, all the variants are attacked with complexity lower than the birthday attack.

As seen from Table 2, all the variants could be collision-attacked by using the near-collisions as a tool and thus replacing the primitive polynomial is not a proper way to make SHA-0 secure.

Table 2 Computational complexity needed for the Chabaud-Joux near-collision attack on the SHA-0 variants.

Complexity	Number of variants	Sum
2^{69}	3	3
2^{68}	4	7
2^{67}	8	15
2^{66}	35	50
2^{65}	55	105
2^{64}	71	176
2^{63}	121	297
2^{62}	167	464
2^{61}	192	656
2^{60}	188	844
2^{59}	198	1042
2^{58}	189	1231
2^{57}	189	1420
2^{56}	132	1552
2^{55}	125	1677
2^{54}	87	1764
2^{53}	84	1848

Complexity	Number of variants	Sum
2^{52}	57	1905
2^{51}	38	1943
2^{50}	33	1976
2^{49}	22	1998
2^{48}	20	2018
2^{47}	11	2029
2^{46}	5	2034
2^{45}	5	2039
2^{44}	5	2044
2^{43}	1	2045
2^{42}	1	2046
2^{41}	0	2046
2^{40}	0	2046
2^{39}	0	2046
2^{38}	0	2046
2^{37}	1	2047
2^{36}	0	2047
2^{35}	1	2048

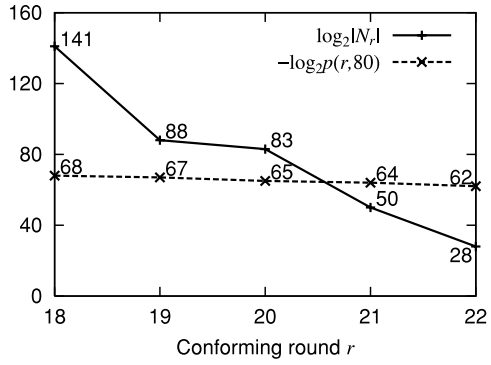

Fig. 1 Relationship between $|N_r|$ and $p(r, 80)$.

4. The Biham-Chen Attack on a Variant

The Biham-Chen attack is an improved version of the Chabaud-Joux attack. It can reduce the complexity needed for the Chabaud-Joux attack. For the Chabaud-Joux near-collision attack on the most near-collision-resistant SHA-0 variants, we need 2^{69} computation. We apply the Biham-Chen attack to the variant and evaluate how much the complexity is reduced.

We consider the attack on the SHA-0 variant which employs $x^{16} + x^{14} + x^{13} + x^{12} + x^{10} + x^9 + x^8 + x^6 + x^3 + x + 1$. This is one of the most near-collision-resistant variant. We use a mask base m such as

$$
\begin{aligned}
m = {} & \texttt{0010 1010 0100 0000 0110} \\
& \texttt{1101 1100 1111 0001 0001} \\
& \texttt{1100 1110 0000 0100 0110} \\
& \texttt{1110 1010 1000 1000 1101}.
\end{aligned}
$$

Let $p(r_1, r_2)$ be the probability for adjustments to succeed from r_1-th round to r_2-th round and N_r be a 2-neutral set at r-th round. The complexity of the Biham-Chen attack is represented by $1/p(r, 80)$. Therefore, the complexity depends on r. In order to find an appropriate r, we generated 20 message pairs for each $r \in \{18, 19, 20, 21, 22\}$ and computed 2-neutral sets. For each r, we selected the maximum 2-neutral set. Figure 1 shows the relationship between $|N_r|$ and $p(r, 80)$. From this figure, we deduce $r = 20$. $p(0, 20) = 2^{-24}$ and $p(20, 80) = 2^{-65}$. Thus, the complexity of the Biham-Chen attack is 2^{65}. The complexity is reduced from 2^{69} to 2^{65}.

As seen from this result, for the SHA-0 variants which have high near-collision resistance, the Biham-Chen attack does not reduce the complexity well. We analyze the reason. The conforming round r is determined as the maximum r which satisfies

$$\log_2 |N_r| \geq -\log_2 p(r, r'). \tag{3}$$

Referring to Fig. 1, if $-\log_2 p(r, r')$ is large, Equation (3) is satisfied only by small r. Thus, the complexity reduction is small since the Biham-Chen attack reduces the complexity from the round 0 to the round $(r - 1)$.

5. Conclusion

In this letter, we evaluated the collision resistance and the near-collision resistance of 2048 SHA-0 variants to the Chabaud-Joux attack and found that none of these variants are resistant to the attack. We also evaluated the near-collision resistance of a variant to the Biham-Chen attack and found that this attack does not work well on variants which exhibit high near-collision resistance.

Some of the future topics are

- vulnerability of SHA-0 variants against the attack by Wang et al.,
- near-collisions other than discussed in this letter, and
- theoretical analysis of the security against the attacks.

References

[1] National Institute of Standards and Technology, Secure hash standard, no.180, FIPS Publication, 1993.
[2] F. Chabaud and A. Joux, "Differential collisions in SHA-0," Crypto'98, Lect. Notes in Comput. Sci., vol.1462, pp.56–71, 1998.
[3] E. Biham and R. Chen, "Near-collisions of SHA-0," Cryptology ePrint Archive, Report 2004/146, 2004. http://eprint.iacr.org/2004/146
[4] X. Wang, D. Feng, X. Lai, and H. Yu, "Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD," Cryptology ePrint Archive, Report 2004/199, 2004. http://eprint.iacr.org/2004/199
[5] A. Joux, "Collisions in SHA-0," Short talk presented at CRYPTO 2004 Rump Session, 2004.
[6] National Institute of Standards and Technology, Secure hash standard, no.180-1, FIPS Publication, 1995.
[7] X. Wang, Y. Yin, and H. Yu, "Collision search attacks on SHA-1," 2005. http://www.infosec.sdu.edu.cn/paper/sha-attack-note.pdf
[8] E. Biham and R. Chen, "New results on SHA-0 and SHA-1," Short talk presented at CRYPTO 2004 Rump Session, 2004.