Collisions for Step-Reduced SHA-256

Ivica Nikolić* and Alex Biryukov

University of Luxembourg
{ivica.nikolic,alex.biryukov}@uni.lu

Abstract. In this article we find collisions for step-reduced SHA-256. We develop a differential that holds with high probability if the message satisfies certain conditions. We solve the equations that arise from the conditions. Due to the carefully chosen differential and word differences, the message expansion of SHA-256 has little effect on spreading the differences in the words. This helps us to find full collision for 21-step reduced SHA-256, semi-free start collision, i.e. collision for a different initial value, for 23-step reduced SHA-256, and semi-free start near collision (with only 15 bit difference out of 256 bits) for 25-step reduced SHA-256.

1 Introduction

The SHA-2 family of hash functions was introduced to the cryptographic community as a new, more complex, and hopefully, more secure variant of MD4-family of hash functions. The recent results on the widely used MD4-family hash functions SHA-1 and MD5 [6],[7] show flaws in the security of these functions, with respect to collision attacks. The question arises, if the most complex member of MD4-family, the SHA-2 family, is also vulnerable to collision attacks.

Known Results for the SHA-2 Family. Research has been made on finding a local collisions for the SHA-2 family. Gilbert and Handschuh [2] reported a 9-step local collision with probability of the differential path of 2^{-66}. Later, Mendel et al [4] estimated the probability of this local collision to be 2^{-39}. Somitra and Palash obtained a local collision with probability 2^{-42}. Using modular differences Hawkes, Paddon and Rose [3] were able to find a local collision with probability 2^{-39}. As far as we know, the only work on finding a real collision for SHA-2 was made by Mendel et al[4]. They studied message expansion of the SHA-256 and reported a 19-step near collision.

Our Contributions. We find a 9-step differential that holds with probability of $\frac{1}{3}$ by fixing some of the intermediate values and solving the equations that arise. We show that it is not necessary to introduce differences in message words on each step of the differential. This helps us to overcome the message expansion. We use modular substraction differences. Using only one instance of this differential we find 20 and 21-step collisions (collisions for the original initial value) with

* The work of this author was supported by the BFR grant 07/031 of the FNR.

K. Nyberg (Ed.): FSE 2008, LNCS 5086, pp. 1–15, 2008.

probabilities $\frac{1}{3}$ and 2^{-19} respectively. Also, using slightly different differential we were able to find a 23-step semi-free start collision (collisions for a specific initial value) with probability 2^{-21}. Our final result is a 25-step semi-free start near collision with Hamming distance of 15 bits and probability 2^{-34}.

Let $H(M, h_0)$ be a hash function, where M is the input message, and h_0 is the initial chaining value. The following attacks are considered in the paper:

Collision attack: Find messages M_1 and M_2 such that $M_1 \neq M_2$ and $H(M_1, h_0) = H(M_2, h_0)$.

Semi-free start collision attack: Find messages M_1,M_2 and hash value h_0^* such that $M_1 \neq M_2$ and $H(M_1, h_0^*) = H(M_2, h_0^*)$.

Near collision attack: Find messages M_1 and M_2 such that $M_1 \neq M_2$ and Hamming distance between $H(M_1, h_0)$ and $H(M_2, h_0)$ is small compared to the output size n of the hash function.

2 Description of SHA-2

SHA-2 family consists of iterative hash functions SHA-224, SHA-256, SHA-384, and SHA-512. For our purposes, we will describe only SHA-256. The definitions of the rest of the functions can be found in [1]. The SHA-256 takes a message of length less than 2^{64} and produces a 256-bit hash value. First, the input message is padded so the length becomes a multiple of 512, and afterwards each 512-bit message block is processed as an input in the Damgard-Merkle iterative structure. Each iteration calls a compression function which takes for an input a 256-bit chaining value and a 512-bit message block and produces an output 256-bit chaining value. The output chaining value of the previous iteration is an input chaining value for the following iteration. The initial chaining value, i.e. the value for the first iteration, is fixed, and the chaining value produced after the last message block is proceeded is the hash value of the whole message. Internal state of SHA-256 compression function consists of 8 32-bit variables A, B, C, D, E, F, G, and H, each of which is updated on every of the 64 steps. These variables are updated according to the following equations:

$$
\begin{aligned}
A_{i+1} &= \Sigma_0(A_i) + Maj(A_i, B_i, C_i) + \Sigma_1(E_i) + Ch(E_i, F_i, G_i) + H_i + K_i + W_i \\
B_{i+1} &= A_i \\
C_{i+1} &= B_i \\
D_{i+1} &= C_i \\
E_{i+1} &= \Sigma_1(E_i) + Ch(E_i, F_i, G_i) + H_i + K_i + W_i + D_i \\
F_{i+1} &= E_i \\
G_{i+1} &= F_i \\
H_{i+1} &= G_i
\end{aligned}
$$

The $Maj(X,Y,Z)$ and $Ch(X,Y,Z)$ are bitwise boolean functions defined as:

$$Ch(X,Y,Z) = (X \wedge Y) \vee (\neg X \wedge Z)$$
$$Maj(X,Y,Z) = (X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z)$$

For SHA-256 $\Sigma_0(X)$ and $\Sigma_1(X)$ are defined as:

$$\Sigma_0(X) = ROTR^2(X) \oplus ROTR^{13}(X) \oplus ROTR^{22}(X)$$
$$\Sigma_1(X) = ROTR^6(X) \oplus ROTR^{11}(X) \oplus ROTR^{25}(X)$$

State update function uses constants K_i, which are different for every step. The 512-bit message block itself is divided in 16 32-bit bit words: $m_0, m_1, \ldots, m_{16}$. Afterwards, the message block is expanded to 64 32-bit words according to the following rule:

$$W_i = \begin{cases} m_i, & 0 \leq i \leq 15 \\ \sigma_1(W_{i-2}) + W_{i-7} + \sigma_0(W_{i-15}) + W_{i-16}, & i > 15 \end{cases}$$

For SHA-256 $\sigma_0(X)$ and $\sigma_1(X)$ are defined as:

$$\sigma_0(X) = \ ROTR^7(X) \oplus ROTR^{18}(X) \oplus SHR^3(X)$$
$$\sigma_1(X) = ROTR^{17}(X) \oplus ROTR^{19}(X) \oplus SHR^{10}(X)$$

The compression function after the 64-th step adds the initial values to the chaining variables, i.e. the hash result of the compression function is:

$$h(M) = (A_{64}+A_0, B_{64}+B_0, C_{64}+C_0, D_{64}+D_0, E_{64}+E_0, F_{64}+F_0, G_{64}+G_0, H_{64}+H_0).$$

These values become the initial chaining value for the next compression function.

3 Technique for Creating Collisions

Differences used in this paper are subtractions mod 2^{32} differences.

We use the following notation:

$$\Delta X = X' - X, \ X \in \{A, B, D, D, E, F, G, H, W, m\},$$
$$\Delta Maj^i(\Delta a, \Delta b, \Delta c) = Maj(A_i + \Delta a, B_i + \Delta b, C_i + \Delta c) - Maj(A_i, B_i, C_i),$$
$$\Delta Ch^i(\Delta e, \Delta f, \Delta g) = Ch(E_i + \Delta e, F_i + \Delta f, G_i + \Delta g) - Ch(E_i, F_i, G_i).$$
$$\Delta\Sigma_0(A_i) = \Sigma_0(A_i') - \Sigma_0(A_i)$$
$$\Delta\Sigma_1(E_i) = \Sigma_1(E_i') - \Sigma_1(E_i)$$
$$\Delta\sigma_0(m_i) = \sigma_0(m_i') - \sigma_0(m_i)$$
$$\Delta\sigma_1(m_i) = \sigma_1(m_i') - \sigma_1(m_i)$$

We introduce perturbation on step i and in the following 8 steps we try to correct the differences in the internal variables. We use the following differential:

Table 1. A 9 step differential for SHA-2 family. Notice that only 5 differences are introduced, i.e. in steps $i, i+1, i+2, i+3$, and $i+8$.

step	ΔA	ΔB	ΔC	ΔD	ΔE	ΔF	ΔG	ΔH	ΔW
i	0	0	0	0	0	0	0	0	1
i+1	1	0	0	0	1	0	0	0	δ_1
i+2	0	1	0	0	-1	1	0	0	δ_2
i+3	0	0	1	0	0	-1	1	0	δ_3
i+4	0	0	0	1	0	0	-1	1	0
i+5	0	0	0	0	1	0	0	-1	0
i+6	0	0	0	0	0	1	0	0	0
i+7	0	0	0	0	0	0	1	0	0
i+8	0	0	0	0	0	0	0	1	δ_4
i+9	0	0	0	0	0	0	0	0	0

As you can see from the table (column ΔW), only the perturbation has been fixed. All the other differences are to be determined.

3.1 Conditions for the Local Collision

From the definition of SHA-2, focusing on registers A_{i+1} and E_{i+1}, we get:

$$\Delta A_{i+1} - \Delta E_{i+1} = \Delta\Sigma_0(A_i) + \Delta Maj^i(\Delta A_i, \Delta B_i, \Delta C_i) - \Delta D_i,$$
$$\Delta E_{i+1} = \Delta\Sigma_1(E_i) + \Delta Ch^i(\Delta E_i, \Delta F_i, \Delta G_i) + \Delta H_i + \Delta D_i + \Delta W_i.$$

We will keep in mind that if $\Delta A_i = \Delta B_i = \Delta C_i = 0$ then $\Delta Maj^i(0,0,0) = 0$. Also if $\Delta E_i = \Delta F_i = \Delta G_i = 0$ then $\Delta Ch^i(0,0,0) = 0$.

We fix the differences for the registers A and E (as shown in the table). The variables B, C, D, F, G, H can only inherit the values from A and E. So, for each step we get some equations with respect to δ_i and A_i or E_i.

Step i+1. We have that $\Delta D_i = 0$, $\Delta H_i = 0$, $\Delta\Sigma_0(A_i) = 0$, $\Delta\Sigma_1(E_i) = 0$. We require $\Delta A_{i+1} = 1$, $\Delta E_{i+1} = 1$. So we deduce:

$$\Delta W_i = 1 \tag{1}$$

Step i+2. We have that $\Delta D_{i+1} = 0$, $\Delta H_{i+1} = 0$. We require $\Delta A_{i+2} = 0$, $\Delta E_{i+2} = -1$. We want also $\Delta\Sigma_0(A_{i+1}) = 1$ to be satisfied. So we deduce:

$$\Delta Maj^{i+1}(1,0,0) = 0, \tag{2}$$
$$\Delta W_{i+1} = -1 - \Delta Ch^{i+1}(1,0,0) - \Delta\Sigma_1(E_{i+1}). \tag{3}$$
$$\Delta\Sigma_0(A_{i+1}) = 1 \tag{4}$$

Step i+3. We have that $\Delta D_{i+2} = 0$, $\Delta H_{i+2} = 0$, $\Delta\Sigma_0(A_{i+2}) = 0$. We require $\Delta A_{i+3} = 0$, $\Delta E_{i+3} = 0$. So we deduce:

$$\Delta Maj^{i+2}(0,1,0) = 0, \tag{5}$$
$$\Delta W_{i+2} = -\Delta\Sigma_1(E_{i+2}) - \Delta Ch^{i+2}(-1,1,0). \tag{6}$$

Step i+4. We have that $\Delta D_{i+3} = 0$, $\Delta H_{i+3} = 0$, $\Delta\Sigma_0(A_{i+3}) = 0$, $\Delta\Sigma_1(E_{i+3}) = 0$. We require $\Delta A_{i+4} = 0$, $\Delta E_{i+4} = 0$. So we deduce:

$$\Delta Maj^{i+3}(0,0,1) = 0, \tag{7}$$
$$\Delta W_{i+3} = -\Delta Ch^{i+3}(0,-1,1). \tag{8}$$

Step i+5. We have that $\Delta D_{i+4} = 1$, $\Delta H_{i+4} = 1$, $\Delta\Sigma_0(A_{i+4}) = 0$, $\Delta\Sigma_1(E_{i+4}) = 0$. We require $\Delta A_{i+5} = 0$, $\Delta E_{i+5} = 1$. So we deduce:

$$\Delta Ch^{i+4}(0,0,-1) = -1. \tag{9}$$

Step i+6. We have that $\Delta D_{i+5} = 0$, $\Delta H_{i+5} = -1$, $\Delta\Sigma_0(A_{i+5}) = 0$. We require $\Delta A_{i+6} = 0$, $\Delta E_{i+6} = 0$. We want also $\Delta\Sigma_0(E_{i+5}) = 1$ to be satisfied. So we deduce:

$$\Delta Ch^{i+5}(1,0,0) = 0. \tag{10}$$
$$\Delta\Sigma_1(E_{i+5}) = 1 \tag{11}$$

Step i+7. We have that $\Delta D_{i+6} = 0$, $\Delta H_{i+6} = 0$, $\Delta\Sigma_0(A_{i+6}) = 0$, $\Delta\Sigma_1(E_{i+6}) = 0$. We require $\Delta A_{i+7} = 0$, $\Delta E_{i+7} = 0$. So we deduce:

$$\Delta Ch^{i+6}(0,1,0) = 0. \tag{12}$$

Step i+8. We have that $\Delta D_{i+7} = 0$, $\Delta H_{i+7} = 0$, $\Delta\Sigma_0(A_{i+7}) = 0$, $\Delta\Sigma_1(E_{i+7}) = 0$. We require $\Delta A_{i+8} = 0$, $\Delta E_{i+8} = 0$. So we deduce:

$$\Delta Ch^{i+7}(0,0,1) = 0. \tag{13}$$

Step i+9. We have that $\Delta D_{i+8} = 0$, $\Delta H_{i+8} = 1$, $\Delta\Sigma_0(A_{i+8}) = 0$, $\Delta\Sigma_1(E_{i+8}) = 0$. We require $\Delta A_{i+9} = 0$, $\Delta E_{i+9} = 0$. So we deduce:

$$\Delta W_{i+8} = -1. \tag{14}$$

3.2 Solution of the System of Equations

Let's first observe (4) and (11). From the differential we can see that $\Delta A_{i+1} = \Delta E_{i+5} = 1$. It means that we want the functions $\Delta\Sigma_0(A_{i+1}), \Delta\Sigma_1(E_{i+5})$ to preserve the difference 1, in other words:

$$\Sigma_0(A_{i+1} + 1) - \Sigma_0(A_{i+1}) = 1,$$
$$\Sigma_1(E_{i+5} + 1) - \Sigma_1(E_{i+5}) = 1.$$

The only solution to these equations is $A_{i+1} = E_{i+5} = -1$, so we get:

$$A_{i+1} = -1, \qquad A'_{i+1} = 0, \tag{15}$$
$$E_{i+5} = -1, \qquad E'_{i+5} = 0. \tag{16}$$

Now let's consider the function $\Delta Maj^i = Maj(A_i^{'}, B_i^{'}, C_i^{'}) - Maj(A_i, B_i, C_i)$. Let's suppose that $B_i^{'} = B_i$, $C_i^{'} = C_i$ and A_i and $A_i^{'}$ differ in every single bit, i.e. $A_i \oplus A_i^{'} =$`0xffffffff`. Then:

$$\Delta Maj^i = 0 \Leftrightarrow B_i = C_i$$

Therefore (2) gives us $B_{i+1} = C_{i+1}$, which is $A_i = A_{i-1}$. With the same reasoning we can deduce from (5) that $A_{i+2} = A_i$, and from (7) that $A_{i+3} = A_{i+2}$. So, from (2),(5) and (7) we get that

$$A_{i-1} = A_i = A_{i+2} = A_{i+3} \tag{17}$$

Similarly to what we have done with Maj, now let's consider ΔCh^i and suppose that $F_i^{'} = F_i$, $G_i^{'} = G_i$ and E_i and $E_i^{'}$ differ in every single bit. Then:

$$\Delta Ch^i = 0 \Leftrightarrow F_i = G_i$$

Therefore (10) and the result (16) gives us $F_{i+5} = G_{i+5}$, which is:

$$E_{i+4} = E_{i+3} \tag{18}$$

Solving (12) requires slightly different reasoning; if we have $E_{i+6} = E_{i+6}^{'}$, $G_{i+6} = G_{i+6}^{'}$ and F_{i+6} and $F_{i+6}^{'}$ would differ in every bit (and they do, see (16)) then :

$$\Delta Ch^{i+6} = 0 \Leftrightarrow E_{i+6} = 0. \tag{19}$$

Analogously, from (13) we get:

$$E_{i+7} = -1 \tag{20}$$

The only remaining condition is (9):

$$\Delta Ch^{i+4} = Ch(E_{i+4}, F_{i+4}, G_{i+4}^{'}) - Ch(E_{i+4}, F_{i+4}, G_{i+4}) = -1,\ G_{i+4}^{'} - G_{i+4} = -1.$$

The words $E_{i+4}, F_{i+4}, G_{i+4}$ are already determined to satisfy the previous conditions. So, we don't have any degrees of freedom left to control precisely the solution of this equation. Therefore we will try to find the probability that this condition holds. We can see that it holds if and only if register E_{i+4} has 0's in the bits where $G_{i+4}^{'}$ and G_{i+4} are different. The $G_{i+4}^{'}$ and G_{i+4} can differ in the last i bits, where $1 \leq i \leq 32$., and these bits are uniquely determined. So, for the probability we get:

$$\sum_{i=1}^{i=32} \text{P\{Last i bits of } E_{i+4} \text{ are zero\}} \times \text{ P\{Difference in the exactly i last bits\}} =$$

$$= \sum_{i=1}^{i=32} \frac{1}{2^i} \frac{1}{2^i} \approx \frac{1}{3}.$$

So, the overall probability of our differential is $\frac{1}{3} = 2^{-1.58}$.

The differences in message words of the differential as in Table 1 are the following:

$$\delta_1 = -1 - \Delta Ch^{i+1}(1,0,0) - \Delta\Sigma_1(E_{i+1}),$$
$$\delta_2 = -\Delta\Sigma_1(E_{i+2}) - \Delta Ch^{i+2}(-1,1,0),$$
$$\delta_3 = -\Delta Ch^{i+3}(0,-1,1)$$
$$\delta_4 = -1$$

Notice that the condition (17) shows us that A_i=B_i has to hold.

4 Full, Semi-free and Near Collisions for Step-Reduced SHA-256

Our attack technique is the following:

1. Introduce perturbation at step i;
2. Correct the differences in the following 8 steps (probability of success is the probability of our differential, i.e. $\frac{1}{3}$). After the last step of the differential, the differences in the internal variables are zero;
3. All the message words that follow the last step of the differential have to have zero differences;

4.1 20-Step Collision

From the Table 3 of Appendix A we can see that the words m_5, m_6, m_7, m_8, and m_{13} are used only once in the first 20 steps of SHA-2, i.e. they are not used to compute the values of expanded words W_{16}, W_{17}, W_{18}, and W_{19}. This means that message expansion doesn't introduce any difference after the last step of the differential. So, we get collision for 20 step reduced SHA-2, and the collisions can be found practically by hand. The probability of collision is $2^{-1.58}$.

4.2 21-Step Collision

From the Table 3 of Appendix A we can easily see that we have to consider message expansion since there are no message words that are used only once in the first 21 steps and that have the proper indexes for the differential.

We will introduce differences in the words m_6, m_7, m_8, m_9, and m_{14}. The words m_6, m_7, m_8 are used only once in the first 21 steps. Therefore the message expansion in the first 21 steps is irrelevant with respect to these words, i.e. differences in these words don't introduce any other new differences, after the last step of the differential(step 14). Now, we want to find words $m_9, m_9', m_{14}, m_{14}'$ such that after the 14-th step, the message expansion will not introduce any difference in the following steps. From the Table 3 of Appendix A we can see that

the words m_9 and m_{14} are used in W_{16}, W_{18}, and W_{20}. So, from the definition of W_i we get the equations:

$$\Delta W_{16} = \Delta\sigma_1(m_{14}) + \Delta m_9 + \Delta\sigma_0(m_1) + \Delta m_0 = 0 \quad (21)$$
$$\Delta W_{17} = \Delta\sigma_1(m_{15}) + \Delta m_{10} + \Delta\sigma_0(m_2) + \Delta m_1 = 0 \quad (22)$$
$$\Delta W_{18} = \Delta\sigma_1(W_{16}) + \Delta m_{11} + \Delta\sigma_0(m_3) + \Delta m_2 = 0 \quad (23)$$
$$\Delta W_{19} = \Delta\sigma_1(W_{17}) + \Delta m_{12} + \Delta\sigma_0(m_4) + \Delta m_3 = 0 \quad (24)$$
$$\Delta W_{20} = \Delta\sigma_1(W_{18}) + \Delta m_{13} + \Delta\sigma_0(m_5) + \Delta m_4 = 0 \quad (25)$$

Obviously if $m_i^{'} = m_i$ ($W_i^{'} = W_i$) then $\Delta\sigma_0(m_i) = 0$ ($\Delta\sigma_0(W_i) = 0$). This means that $\Delta W_{17} = \Delta W_{19} = 0$. If we can make so that $\Delta W_{16} = 0$ then $\Delta W_{18} = \Delta W_{20} = 0$. So, we get the equation:

$$\Delta\sigma_1(m_{14}) + \Delta m_9 = 0 \quad (26)$$

Considering that $\Delta m_{14} = \delta_4 = -1$, and m_9 can take any value, our experimental results (Monte Carlo method with 2^{32} trials) give us a probability of $2^{-17.5}$ that Δm_{14} and Δm_9 satisfy this equation. Therefore, the overall probability of 21 step collision is around 2^{-19}.

4.3 23-Step Semi-free Start Collision

For 23 step collision we introduce differences in the words m_9, m_{10}, m_{11}, and m_{12}.

If we would follow our differential, we are supposed to introduce difference in the message word W_{17}. We can not control W_{17} directly because it is an expanded word. From the condition $W_{17} = \delta_4 = -1$ (differential) and the message expansion, we get:

$$\Delta W_{17} = \Delta\sigma_1(m_{15}) + \Delta m_{10} + \Delta\sigma_0(m_2) + \Delta m_1 = -1.$$

Since $\Delta m_{15} = \Delta m_2 = \Delta m_1 = 0$, we get:

$$\Delta m_{10} = -1. \quad (27)$$

In our original differential there are no message differences in the word W_{16}. But for W_{16} we have:

$$\Delta W_{16} = \Delta\sigma_1(m_{14}) + \Delta m_9 + \Delta\sigma_0(m_1) + \Delta m_0.$$

Obviously only $\Delta m_9 \neq 0$ and therefore $\Delta W_{16} = \Delta m_9 = 1 \neq 0$. Therefore we shall use slightly different differential:one were there is a difference in the word W_{16}. To keep everything else unchanged, the equations for the step 17 become the following:

$$\Delta E_{17} = \Delta\Sigma_1(E_{16}) + \Delta Ch^{16}(0,0,1) + \Delta D_{16} + \Delta H_{16} + \Delta W_{16}.$$

From the differential we can see that: $\Delta E_{17} = \Delta\Sigma_1(E_{16}) = \Delta D_{16} = \Delta H_{16} = 0$. Therefore we get:

$$\Delta Ch^{16}(0,0,1) + \Delta W_{16} = 0. \quad (28)$$

Now, let's observe the other words of the message expansion.

For W_{18} we have:

$$W_{18} = \Delta\sigma_1(W_{16}) + \Delta m_{11} + \Delta\sigma_0(m_3) + \Delta m_2 = 0$$

Since $\Delta m_3 = \Delta m_2 = 0, \Delta W_{16} = 1$ we get the equation:

$$\Delta\sigma_1(W_{16}) + \Delta m_{11} = 0. \tag{29}$$

For W_{19} we have:

$$W_{19} = \Delta\sigma_1(W_{17}) + \Delta m_{12} + \Delta\sigma_0(m_4) + \Delta m_3 = 0$$

Since $\Delta m_4 = \Delta m_3 = 0, \Delta W_{17} = -1$ we get the equation:

$$\Delta\sigma_1(W_{17}) + \Delta m_{12} = 0. \tag{30}$$

For W_{20} we have:

$$W_{20} = \Delta\sigma_1(W_{18}) + \Delta m_{13} + \Delta\sigma_0(m_5) + \Delta m_4 = 0$$

Since $\Delta W_{18} = \Delta m_{13} = \Delta m_5 = \Delta m_4 = 0$ we get that this equation is satisfied for all values of W_{18}, m_{13}, m_5, m_4.

For W_{21} we have:

$$W_{21} = \Delta\sigma_1(W_{19}) + \Delta m_{14} + \Delta\sigma_0(m_6) + \Delta m_5 = 0$$

Since $\Delta W_{19} = \Delta m_{14} = \Delta m_6 = \Delta m_5 = 0$ we get that this equation is satisfied for all values of W_{19}, m_{14}, m_6, m_5.

For W_{22} we have:

$$W_{22} = \Delta\sigma_1(W_{20}) + \Delta m_{15} + \Delta\sigma_0(m_7) + \Delta m_6 = 0$$

Since $\Delta W_{20} = \Delta m_{15} = \Delta m_7 = \Delta m_6 = 0$ we get that this equation is satisfied for all values of W_{20}, m_{15}, m_7, m_6.

For W_{23} we have:

$$W_{23} = \Delta\sigma_1(W_{21}) + \Delta W_{16} + \Delta\sigma_0(m_8) + \Delta m_7 = 0$$

Since $\Delta W_{21} = \Delta m_8 = \Delta m_7$ and $\Delta W_{16} \neq 0$ we get that this equation has no solution. That is why we can not get more than 23 step collision.

Let's try to solve (27), (28), (29) and (30).

For (27) and the value of the register E_{11} from the differential's conditions we have:

$$\Delta E_{11} = \Delta\Sigma_1(E_{10})) + \Delta Ch^{10}(1,0,0) + \Delta m_{10}.$$

Since $\Delta E_{11} = m_{10} = -1$ we get:

$$\Delta\Sigma_1(E_{10}) + \Delta Ch^{10}(1,0,0) = 0.$$

We solve this equation by setting $\Delta\Sigma_1(E_{10}) = 1$ and $\Delta Ch^{10}(1,0,0) = -1$. The first one has solution:

$$E_{10} = -1,\ E'_{10} = 0. \tag{31}$$

The second equation holds for the values:

$$F_{10} = G_{10} + 1. \tag{32}$$

Now let's turn to the solution of (28). Using the fact that $G_{16} = -1$ and $G'_{16} = 0$, we get that this equation is satisfied if:

$$E_{16} = \texttt{0xffffffe} \tag{33}$$

Let's observe the equation (30). From the conditions of the differential we have:

$$\Delta E_{13} = \Delta\Sigma_1(E_{12}) + \Delta Ch^{12}(0,-1,1) + \Delta H_{12} + \Delta D_{12} + \Delta m_{12}$$

Since $\Delta E_{13} = \Delta E_{12} = \Delta H_{12} = \Delta D_{12} = 0$ we get:

$$\Delta Ch^{12}(0,-1,1) + \Delta m_{12} = 0.$$

If we substitute m_{12} from (30) we can get:

$$\Delta Ch^{12}(0,-1,1) = \Delta\sigma_1(-1).$$

This equation can be satisfied if we can control E_{12} and F_{12}.

For E_{12}, from the definition of A_{12} and E_{12} we have:

$$A_{12} - E_{12} = \Sigma_1(A_{11}) + Ch(A_{11}, B_{11}, C_{11}) - D_{11}$$

Considering that $A_{12} = A_{11} = C_{11} = D_{11}$ from the differential's conditions, we get:

$$E_{12} = A_9 - \Sigma_1(A_9)$$

Since A_9 can take any value (we consider semi-free start collision) we deduce that E_{12} can take any value.

The F_{12} value, which is E_{11} can be controlled through H_{10}. Notice that changing H_{10}, which is G_9, doesn't effect E_{10}, because from (31) we can see that E_{10} always takes the arranged value.

We proved that we can fully control E_{12} and F_{12}. We can choose some specific value for $\Delta\sigma_1(-1)$ which is possible to get from $\Delta Ch^{12}(0,-1,1)$, and set the A_9 and G_9 so that the equation (30) will hold.

The last equation, i.e. (29), is satisfied for some specific values of W_{16} and m_{11}. Our experimental results show that with probability $2^{-19.5}$ W_{16} and m_{11} satisfy (29). Therefore the overall probability of semi-free start collision for 23-step reduced SHA-256 is around 2^{-21}.

4.4 25-Steps Semi-free Start Near Collision

Let's suppose we have a semi-free start collision on the 23-rd step. Each following step introduces differences in the chaining variables A and E. The variables B, C, D, F, G, H can only inherit differences from A and E. Therefore, for each step, we should try to minimize the differences in A and E. When we say to minimize the differences we mean to minimize the Hamming distances between $A^{'}$ and A, and between $E^{'}$ and E.

Step 24

$$\min_{W'_{23}-W_{23}=1} h_d(E^{'}_{24}, E_{24}) = \min_{W'_{23}-W_{23}=1} h_d(C_1 + 1, C_1) = 1,$$

where $C_1 = \Sigma_1(E_{23}) + Ch(E_{23}, F_{23}, G_{23}) + H_{23} + D_{23} + K_{23} + W_{23}$.

$$\min_{W'_{23}-W_{23}=1} h_d(A^{'}_{24}, A_{24}) = \min_{W'_{23}-W_{23}=1} h_d(C_2 + 1, C_2) = 1,$$

where $C_2 = \Sigma_0(A_{23}) + Maj(A_{23}, B_{23}, C_{23}) + \Sigma_1(E_{23}) + Ch(E_{23}, F_{23}, G_{23}) + H_{23} + K_{23} + W_{23}$.

We have the minimal Hamming distances when $C_1^{32} = C_2^{32} = 0$, which means with probability 2^{-2}.

Step 25

$$\min_{W'_{24}-W_{24}=-1+\Delta\sigma_0(1)} h_d(E^{'}_{25}, E_{25}) =$$

$$= \min h_d(\Sigma_1(E^{'}_{24}) + Ch(E^{'}_{24}, F_{24}, G_{24}) - 1 + \sigma_0(m_9+1) + C_1, \Sigma_1(E_{24}) + Ch(E_{24}, F_{24}, G_{24}) + \sigma_0(m_9) + C_1),$$

where $C_1 = H_{24} + D_{24} + K_{24} + \sigma_1(W_{22}) + m_8$. If $F_{24}^{32} = 1$ and $G_{24}^{32} = 0$ (probability 2^{-2}) then, considering that $E_{24}^{'32} = 1, E_{24}^{32} = 0$, we have $Ch(E^{'}_{24}, F_{24}, G_{24}) - 1 = Ch(E_{24}, F_{24}, G_{24})$, and we can rewrite the last expression as:

$$\min h_d(\Sigma_1(E^{'}_{24}) + \sigma_0(m_9 + 1) + C_2, \Sigma_1(E_{24}) + \sigma_0(m_9) + C_2),$$

where $C_2 = C_1 + Ch(E_{24}, F_{24}, G_{24})$.

If no carry occurs due to the differences, then the above minimum is:

$$\min h_d(\Sigma_1(E^{'}_{24}) + \sigma_0(m_9 + 1) + C_2, \Sigma_1(E_{24}) + \sigma_0(m_9) + C_2) = 5.$$

For $\Sigma_1(E^{'}_{24})$ (difference in three bits) there are no carries with probability 2^{-3}. For $\sigma_0(m_9 + 1)$ (two differences if $m_9^{32} = 0$) with probability 2^{-3}. Therefore the minimum is 5 with probability 2^{-8}.

Using the same methods we can get:

$$\min_{W'_{24}-W_{24}=-1+\Delta\sigma_0(1)} h_d(A^{'}_{25}, A_{25}) = 8,$$

with probability 2^{-11}. Notice that if minimum holds for A_{25} then it holds for E_{25}.

So, for the whole hash value, we have:

$$h_d((A'_{25}, B'_{25}, C'_{25}, D'_{25}, E'_{25}, F'_{25}, G'_{25}, H'_{25})), (A_{25}, B_{25}, C_{25}, D_{25}, E_{25}, F_{25}, G_{25}, H_{25})) =$$
$$= h_d((A'_{25}, A'_{24}, C_{25}, D_{25}, E'_{25}, E'_{24}, G_{25}, H_{25}), (A_{25}, A_{24}, C_{25}, D_{25}, E_{25}, E_{24}, G_{25}, H_{25})) =$$
$$= h_d(A'_{25}, A_{25}) + h_d(E'_{25}, E_{25}) + h_d(A'_{24}, A_{24}) + h_d(E'_{24}, E_{24}) =$$
$$= 8 + 5 + 1 + 1 = 15$$

Therefore we get a 25-step semi-free start near collision with the Hamming weight of 15 bits and probability 2^{-34}. Notice that we haven't investigated all the possible outcomes of the carry effects. Therefore, it is possible that the real probability is higher.

Table 2. Collision search attacks for SHA-256

# of steps	Type of collision	Complexity(*)	Paper
19	Near collision	(**)	[4]
20	Collision	$2^{1.58}$	This paper
21	Collision	2^{19}	This paper
22	Pseudo-collision	(**)	[4]
23	Semi-free start collision	2^{21}	This paper
25	Semi-free start near collision	2^{34}	This paper

(*) Complexity is measured in reduced SHA-256 calls
(**) Complexity not mentioned in the paper

5 Conclusion

We created a 9-step differential for SHA-256 that holds with high probability. Using the characteristics of this differential, precisely, the fact that not all of the input message words have differences, we were able to overcome the beginning steps of the message expansion. We created a full collisions for 20 and 21-step reduced SHA-256. Also, we found a 23-step reduced semi-free start collision, and 25-step reduced near collision with Hamming distance of 17 out of 256 bits. The complexities of these collisions search attacks are showed in Table 2. Obviously, our results hold for SHA-224 too. For SHA-384 and SHA-512 different equations arise. We have not analyzed them, but our guess is that complexities of the attacks should stay the same.

References

1. Secure Hash Standard. Federal Information Processing Starndard Publication 180-2. U.S. Department of Commerce, National Institute of Standards and Technology (NIST) (2004)
2. Gilbert, H., Handschuh, H.: Security analysis of SHA-256 and sisters. In: Matsui, M., Zuccherato, R.J. (eds.) Selected Areas in Cryptography, 2003. LNCS, vol. 3006, pp. 175–193. Springer, Heidelberg (2003)

3. Hawkes, P., Paddon, M., Rose, G.G.: On Corrective Patterns for the SHA-2 Family. Cryptology eprint Archive (August 2004), http://eprint.iacr.org/2004/207
4. Mendel, F., Pramstaller, N., Rechberger, C., Rijmen, V.: Analysis of step-reduced SHA-256. In: Robshaw, M.J.B. (ed.) FSE 2006. LNCS, vol. 4047, pp. 126–143. Springer, Heidelberg (2006)
5. Sanadhya, S.K., Sarkar, P.: New Local Collision for the SHA-2 Hash Family.Cryptology eprint Archive (2007), http://eprint.iacr.org/2007/352
6. Wang, X., Yin, Y.L., Yu, H.: Finding collisions in the full SHA-1. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 17–36. Springer, Heidelberg (2005)
7. Wang, X., Yu, H.: How to break MD5 and other hash functions. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 19–35. Springer, Heidelberg (2005)

A Message Expansion

Table 3. Message expansion of SHA-2. There is 'x' in the intersection of row with index i and column with index j if W_i uses m_j.

W	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	x															
1		x														
2			x													
3				x												
4					x											
5						x										
6							x									
7								x								
8									x							
9										x						
10											x					
11												x				
12													x			
13														x		
14															x	
15																x
16	x	x								x					x	
17		x	x								x					x
18	x	x	x	x						x		x			x	
19		x	x	x	x						x		x			x
20	x	x	x	x	x	x				x		x		x	x	
21		x	x	x	x	x	x				x		x		x	x
22	x	x	x	x	x	x	x	x		x		x		x	x	x

B Conditions for Collision

Table 4. The differences propagation for 20, 21, and 23-step collisions for SHA-256. Notice that for each collision initial difference is introduced in different steps (steps 5,6,9 respectively).

20 step	21 step	23 step	ΔA	ΔB	ΔC	ΔD	ΔE	ΔF	ΔG	ΔH	ΔW
5	6	9	0	0	0	0	0	0	0	0	1
6	7	10	1	0	0	0	1	0	0	0	δ_1
7	8	11	0	1	0	0	-1	1	0	0	δ_2
8	9	12	0	0	1	0	0	-1	1	0	δ_3
9	10	13	0	0	0	1	0	0	-1	1	0
10	11	14	0	0	0	0	1	0	0	-1	0
11	12	15	0	0	0	0	0	1	0	0	0
12	13	16	0	0	0	0	0	0	1	0	δ_5
13	14	17	0	0	0	0	0	0	0	1	-1
14	15	18	0	0	0	0	0	0	0	0	0

Table 5. The values of the word differences in 20, 21, and 23-step collisions for SHA-256. Notice that 23-step semi-free start collision has a word difference in δ_5. That is why its collision path is slightly different than the one used for 20 and 21-step collision.

	δ_1	δ_2	δ_3	δ_5
20-step	$-1-\Delta Ch^6(1,0,0)-\Delta\Sigma_1(E_6)$	$-\Delta\Sigma_1(E_7)-\Delta Ch^7(-1,1,0)$	$-\Delta Ch^8(0,-1,1)$	0
21-step	$-1-\Delta Ch^7(1,0,0)-\Delta\Sigma_1(E_7)$	$-\Delta\Sigma_1(E_8)-\Delta Ch^8(-1,1,0)$	$-\Delta Ch^9(0,-1,1)$	0
23-step	-1	$-\Delta\Sigma_1(E_{11})-\Delta Ch^{11}(-1,1,0)$	$-\Delta Ch^{12}(0,-1,1)$	1

Table 6. The additional conditions that have to hold in order to get a 20, 21, and 23-step collisions for SHA-256

20-step	$A_4=A_5=A_7=A_8$ $A_6=-1, A_6^{'}=0$	$E_9=E_8, E_{10}=-1, E_{10}^{'}=0$ $E_{11}=0, E_{12}=-1$	$\Delta Ch^9(0,0,-1)=-1$
21-step	$A_5=A_6=A_8=A_9$ $A_7=-1, A_7^{'}=0$	$E_{10}=E_9, E_{11}=-1, E_{11}^{'}=0$ $E_{12}=0, E_{13}=-1$	$\Delta Ch^{10}(0,0,-1)=-1$ $\Delta\sigma_1(-1)+\delta_3=0$
23-step	$A_8=A_6=A_9=A_{10}$ $A_{10}=-1, A_{10}^{'}=0$	$E_{13}=E_{12}, E_{14}=-1, E_{14}^{'}=0$ $E_{15}=0, E_{16}=$`0xfffffffe` $E_9=E_8+1, E_{10}=-1, E_{10}^{'}=0$	$\Delta Ch^{13}(0,0,-1)=-1$ $\Delta\sigma_1(-1)+\delta_3=0$ $\Delta\sigma_1(1)+\delta_2=0$

C Collision Examples

Table 7. A 21-step collision for SHA-256

M_0	0004024f	00000000	00000000	00000000	00000000	2c51fd8d	b83daf3**c**	bc852709
	ae18a3e7	1d11dbc7	21d06175	ab551b5f	a48e9a8b	00000000	19000000	00000000
M_0'	0004024f	00000000	00000000	00000000	00000000	2c51fd8d	b83daf3d	7c652ab7
	b238a344	1d11dac8	21d06175	ab551b5f	a48e9a8b	00000000	18fffff	00000000
H	73f5fcd2	682f578e	8d9c3d05	f93ad865	662b0636	a5a5d4c2	32091775	04ac6dae

Table 8. A 23-step semi-free start collision for SHA-256

H_0	cb518aaa	55d8f4ad	231e476a	89ac8889	f29c30cc	2e1f63c5	cf4f2366	75367200
M_0	b5c16a2d	6da1708b	00000000	00000000	00000000	00000000	00000000	00000000
	00000000	a9d5faeb	54eb8149	085be1ce	b9e61e60	9380ae01	efa5a517	cdc5da00
M_0'	b5c16a2d	6da1708b	00000000	00000000	00000000	00000000	00000000	00000000
	00000000	a9d5faec	54eb8148	085c0205	b9e61d61	9380ae01	efa5a517	cdc5da00
H	6682cc14	9c825293	bc17ea6d	d89770cf	a69ac7ed	cfa5ee3e	e35c0091	7249d71e

Table 9. A 25-step semi-free start near collision with Hamming distance of 17 bits for SHA-256

H_0	8e204f9e	bca27aea	42da63d7	00f2f219	fd1db715	6389ae13	c6f57538	de4e655c
M_0	c63714eb	13d5fa9c	00000000	00000000	00000000	00000000	00000000	00000000
	00000000	d51b4dba	aeb6f738	61dce9b7	0ab5c01a	83406f01	df65666b	cdc5da00
M_0'	c63714eb	13d5fa9c	00000000	00000000	00000000	00000000	00000000	00000000
	00000000	d51b4dbb	aeb6f737	71dd499a	0ab5bf1b	83406f01	df65666b	cdc5da00
H	2e2fcb73	8192d3a4	f85b5a7d	801c4583	9307e51c	cf57fb61	11c48b0d	7131ccd2
H'	6c478ef3	8192d3a5	f85b5a7d	801c4583	9127a49c	cf57fb62	11c48b0d	7131ccd2