

New Local Collisions for the SHA-2 Hash Family

Somitra Kumar Sanadhya * and Palash Sarkar

Applied Statistics Unit,
Indian Statistical Institute,
203, B.T. Road, Kolkata,
India 700108.
{somitra_r, palash}@isical.ac.in

Abstract. The starting point for collision attacks on practical hash functions is a local collision. In this paper, we make a systematic study of local collisions for the SHA-2 family. The possible linear approximations of the constituent Boolean functions are considered and certain impossible conditions for such approximations are identified. Based on appropriate approximations, we describe a general method for finding local collisions. Applying this method, we obtain several local collisions and compute the probabilities of the various differential paths. Previously, only one local collision due to Gilbert-Hands Schuh was known. We point out two impossible conditions in the GH local collision and provide an example of an impossible differential path for linearized SHA-2 using this local collision. Sixteen new local collisions are obtained none of which have any impossible conditions. The probabilities of these local collisions are a little less than the GH local collision. On the other hand, the absence of impossible conditions may make them more suitable for (reduced round) collision search attacks on the SHA-2 family.

1 Introduction

Study of collision search attacks on practical hash functions is a topic of intense interest in recent times. Some spectacular successes have been reported for concrete and widely used proposals such as MD5 [12] and SHA-1 [11,1]. Other less popular hash functions such as RIPEMD and HAVAL have also been successfully attacked.

Currently, the two commonly used hash functions are MD5 and SHA-1. In view of the attacks on these functions, there seems to be a tendency to move to the more complicated SHA-2 family. As a result, these hash functions will receive much more attention from the research community.

Usually, the first step in a collision search attack is to find a local collision. This is a collision for a fixed number of steps of the round function. Details about the message expansion are ignored. Further, all nonlinear components of the hash design are approximated by some suitable linear functions. Once a local collision is obtained, one attempts to find a collision for the full hash function by taking into account the message expansion and the nonlinear behaviour of the hash design. For example, Wang et al.'s attack on the SHA-1 hash function [11] uses the local collision obtained by Chabaud and Joux [2]. For details about this approach one may refer to [2].

Known Results for the SHA-2 Family: Gilbert and Handschuh (GH) [4] were the first to study local collisions in the SHA-2 family. They reported a 9-round local collision and estimated the probability of the differential path to be 2^{-66} . The message expansion of the SHA-256 was studied by Mendel et al [7], who reported reduced round (near) collisions. The work [7] remarked that the probability of the GH local collision is 2^{-39} . This value of the probability was also obtained in [5] when modular differences are considered. An earlier work [6] studied a very simplified variant of SHA-256. The encryption mode of SHA-256 is analyzed in [14] and is not relevant to collision search attacks.

* This author is supported by the Department of Information Technology, Govt. of India.

Our Contributions: All previous works have considered only the GH local collision. In this paper, we revisit the problem of obtaining a local collision for the SHA-2 family of functions. Local collisions are found by forming linear approximations of the Boolean functions f_{IF} and f_{MAJ} involved in round function of SHA-2. We make a systematic analysis of the linear approximations of the two Boolean functions. The differential analysis shows that certain kinds of linear approximations give rise to impossible conditions. Given any linear approximations for f_{IF} and f_{MAJ} , we describe a step-by-step method for finding a 9-step local collision for the corresponding linearized round function. This method has been applied on all feasible linear approximations. Two of the cases have been described in details. We also show how to extend the presented local collisions into 17 and 18 step collisions for SHA-2.

The GH local collision was obtained by approximating both f_{IF} and f_{MAJ} by 0. We show that both the approximations have one impossible condition each and this can lead to an impossible differential path. Note that the differential path is impossible for the linearized version of the hash function. It is not impossible for the actual design. An example is provided of an 12-step impossible differential path for the GH local collision. This path is impossible due to the impossible condition on the approximation of f_{IF} by 0. Mendel et al [7] circumvent the impossible conditions of the Boolean functions by using carry propagation in addition. However, this puts extra conditions on message bits reducing the freedom and thereby reducing the probability of the attack. We hope that the new local collisions will help carry out longer round attacks on SHA-2 family.

There are four linear approximations each of f_{MAJ} and f_{IF} which do not have any impossible conditions. These give rise to a total of 16 different linear approximations without any impossible conditions. We develop all these approximations to obtain 16 new local collisions without any impossible conditions. Also, we describe four other local collisions which have one impossible condition for f_{MAJ} and none for f_{IF} .

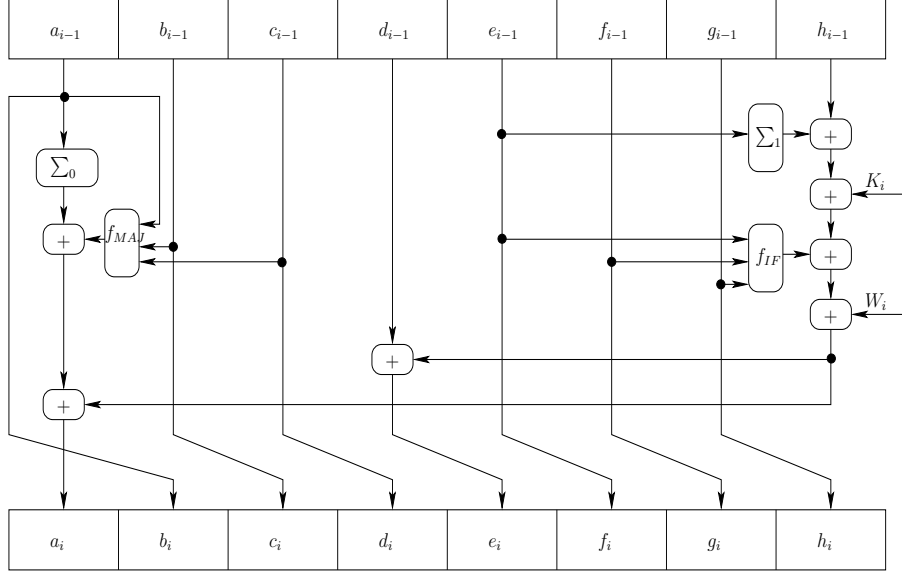
Probabilities of all the local collisions are computed. For the GH local collision we obtain a probability of 2^{-42} . The previous estimate by GH was 2^{-66} . The probabilities of the other local collisions are found to be between 2^{-45} to 2^{-54} . In [5], the probability of the GH local collision was computed to be 2^{-39} using modular differences and in [7] it was remarked (without providing details) that this can be higher than 2^{-39} even with XOR differences. We note that whatever be the method for computing probability estimates, the relative probabilities of the different local collisions will probably remain the same. Further, even though the probabilities of the new local collisions are lower than the GH local collision, the absence of impossible conditions may offset this disadvantage when they are used to find actual (reduced round) collisions for the SHA-2 family.

2 SHA-2 Family of Hash Functions

The round function of the SHA-2 family is shown in Figure 1. In this article, we analyze only the round function. For the complete description of the SHA-2 family see [9]. The 8 registers are updated in each step according to the following equations (all additions are modulo 2^{32}):

$$\left. \begin{aligned} a_i &= \Sigma_0(a_{i-1}) + f_{MAJ}(a_{i-1}, b_{i-1}, c_{i-1}) + \Sigma_1(e_{i-1}) \\ &\quad + f_{IF}(e_{i-1}, f_{i-1}, g_{i-1}) + h_{i-1} + K_i + W_i \\ b_i &= a_{i-1} \\ c_i &= b_{i-1} \\ d_i &= c_{i-1} \\ e_i &= d_{i-1} + \Sigma_1(e_{i-1}) + f_{IF}(e_{i-1}, f_{i-1}, g_{i-1}) \\ &\quad + h_{i-1} + K_i + W_i \\ f_i &= e_{i-1} \\ g_i &= f_{i-1} \\ h_i &= g_{i-1} \end{aligned} \right\} \quad (1)$$

Fig. 1. Round function of SHA-2 family



The f_{IF} and the f_{MAJ} are three variable Boolean functions defined as:

$$\begin{aligned} f_{IF}(x, y, z) &= (x \wedge y) \oplus (\neg x \wedge z) \\ f_{MAJ}(x, y, z) &= (x \wedge y) \oplus (y \wedge z) \oplus (z \wedge x) \end{aligned}$$

The functions Σ_0 and Σ_1 are defined differently for SHA-256 and SHA-512. For SHA-256, these functions are defined as:

$$\begin{aligned} \Sigma_0(x) &= ROTR^2(x) \oplus ROTR^{13}(x) \oplus ROTR^{22}(x) \\ \Sigma_1(x) &= ROTR^6(x) \oplus ROTR^{11}(x) \oplus ROTR^{25}(x) \end{aligned}$$

And for SHA-512, they are defined as:

$$\begin{aligned} \Sigma_0(x) &= ROTR^{28}(x) \oplus ROTR^{34}(x) \oplus ROTR^{39}(x) \\ \Sigma_1(x) &= ROTR^{14}(x) \oplus ROTR^{18}(x) \oplus ROTR^{41}(x) \end{aligned}$$

Our analysis treats Σ_0 and Σ_1 as operators, hence the discussion that follows holds for both SHA-256 and SHA-512 (In the following, we will interchangeably use $\Sigma_i(X)$ and $\Sigma_i X$). Since SHA-384 is just a truncated version of SHA-512, we refer to all the three hash functions as SHA-2 family. Although the word size in SHA-256 is 32 bits and that in SHA-512 / SHA-384 is 64 bits, our analysis remains the same for all these hash functions. There is a minor difference in the probability calculations for the differential paths. This issue is discussed later.

3 Differential Properties of Boolean Functions

Let $f(x)$ be a Boolean function on n variables. By Δx we denote the XOR difference in the input of f , i.e., $\Delta x = x \oplus x'$ for two n -bit strings x and x' . The value of Δx can be any 2^n bit string. Given Δx , define

$\Delta f = f(x \oplus \Delta x) \oplus f(x)$. The value of Δf is either 0 or 1 but is not uniquely determined by the value of Δx . Assuming that x is uniformly distributed over $\{0, 1\}^n$, the value of Δf is 0 or 1 with certain probabilities.

There are two Boolean functions used in SHA-2, namely the f_{IF} and the f_{MAJ} , which are 3-input bit-wise ‘If’ and the ‘Majority’ functions respectively. The three inputs to the functions can have XOR differences of 0 or 1. Depending on their positions, the Boolean functions propagate the differences or absorb them. The differential properties are shown in Table 1. The first 3 columns in this table are the input differences to the Boolean functions, whose output differences are listed in next 2 columns. An entry of 0 (resp. 1) in a Boolean function column means that Δf is 0 (resp. 1) with probability 1. An entry (0, 1) denotes that Δf is 0 with probability half. We will use this table to compute the probabilities of the differential paths that we show later. Note that the differential properties of Boolean function f_{IF} and f_{MAJ} are also considered in [8] but our presentation is different.

Impossible Conditions: Suppose we approximate $f(x)$ by a linear function $l(x)$. Note that Δx fixes the value of Δl with probability one. Now suppose that for some Δx , the value of Δf is also determined with probability one and that $\Delta f \neq \Delta l$ for this value of Δx . Then the particular value of Δx for which this occurs is said to be an **impossible condition** for the approximation of f by l . The complete list of impossible conditions which arise when f_{IF} and f_{MAJ} are approximated by different linear functions is given in Table 2.

Table 1. Differential properties of f_{IF} and f_{MAJ} . A single 1 (0) in the last 2 columns means that this value holds with probability 1. The entry (0,1) implies that both the values are possible with probability $\frac{1}{2}$ each.

Δa	Δb	Δc	$\Delta f_{IF}(a, b, c)$	$\Delta f_{MAJ}(a, b, c)$
0	0	0	0	0
0	0	1	(0,1)	(0,1)
0	1	0	(0,1)	(0,1)
0	1	1	1	(0,1)
1	0	0	(0,1)	(0,1)
1	0	1	(0,1)	(0,1)
1	1	0	(0,1)	(0,1)
1	1	1	(0,1)	1

Table 2. Impossible conditions for the different linear approximations of $f_{IF}(a, b, c)$ and $f_{MAJ}(a, b, c)$. The entries in the table provide the values of $(\Delta a, \Delta b, \Delta c)$ which are the impossible conditions for the corresponding approximation.

	0	a	b	c	$a \oplus b$	$a \oplus c$	$b \oplus c$	$a \oplus b \oplus c$
f_{IF}	(0, 1, 1)	(0, 1, 1)	none	none	none	none	(0, 1, 1)	(0, 1, 1)
f_{MAJ}	(1, 1, 1)	none	none	none	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	none

The probability that $f_{IF}(a, b, c) = 0$ is $1/2$ and the probability that $f_{IF}(a, b, c) = c$ (or b) is $3/4$. This suggests that approximating f_{IF} by c (or b) should be better than approximating f_{IF} by 0. From Table 1, the probability that $\Delta f_{IF} = \Delta c$ is $5/8$, where as the probability for $\Delta f_{IF} = 0$ is still $1/2$. Thus, on an average, the approximation of f_{IF} by c should be better than that by 0 even for a differential analysis.

Remark: It has been mentioned in [7, Page 130, Lines 4–5] that several approximations for f_{IF} and f_{MAJ} are possible and all of these hold with probability 0.5. Table 1 and the discussion above shows that this is not the case. Specifically, the approximation c (or b) is better than the approximation 0 for $f_{IF}(a, b, c)$.

Explanations of two observations on Page 135 of [7]. These observations were made regarding the presence of impossible characteristics in the GH local collision where both f_{IF} and f_{MAJ} are approximated by 0.

1. If there are three consecutive steps in the differential path, such that Δa is 1 in the same bit position, then the resulting characteristics is impossible.
2. If there are three consecutive steps in the differential path, such that there is a bit position where Δe is 1 for the first two steps and 0 for the third step, then the resulting characteristics is impossible.

The first observation is explained by the fact that the condition $(1, 1, 1)$ is an impossible condition for the approximation of f_{MAJ} by 0. The second observation is explained by the fact that the condition $(0, 1, 1)$ is an impossible condition for the approximation of f_{IF} by 0. Note that Mendel et al [7] also explain these observations on the basis of probability of approximations of f_{IF} and F_{MAJ} being 0 in certain cases, without explicitly mentioning the conditions as presented here. We discuss these observations here since they fit with our unified way of considering the impossible conditions in the two Boolean functions.

4 Linear Approximation of SHA-2 Round Function

Local collisions are usually found for the linearized version of the hash function concerned [2,10]. Once it is found for the simple case, the probability for this local collision to hold for the actual hash function is computed. We proceed along similar lines and approximate all additions in SHA-2 by bit-wise XOR. There are many possibilities for the linear approximations of f_{IF} and f_{MAJ} functions. A general form of expressing these approximations is the following

$$\left. \begin{aligned} f_{MAJ}(a, b, c) &= x_1a \oplus x_2b \oplus x_3c \\ f_{IF}(e, f, g) &= y_1e \oplus y_2f \oplus y_3g \end{aligned} \right\} \quad (2)$$

where (x_1, x_2, x_3) and (y_1, y_2, y_3) are 3-bit strings. Thus, the linear approximations are completely specified by these two strings. Let $\Delta \text{reg}_i = (\Delta a_i, \Delta b_i, \Delta c_i, \Delta d_i, \Delta e_i, \Delta f_i, \Delta g_i, \Delta h_i)$. Then the linearized version of the SHA-2 round function can be expressed by an equation of the form

$$(\Delta \text{reg}_i)^t = A(\Delta \text{reg}_{i-1}, \Delta W_i)^t \quad (3)$$

where $()^t$ denotes transpose and A is a suitable matrix which is constructed depending upon the particular linear approximation being used. The form of A in terms of (x_1, x_2, x_3) and (y_1, y_2, y_3) is given by (4).

$$A = \begin{bmatrix} p_1 & x_2 & x_3 & 0 & p_2 & y_2 & y_3 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & p_2 & y_2 & y_3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (4)$$

where $p_1 = (x_1 \oplus \Sigma_0)$ and $p_2 = (y_1 \oplus \Sigma_1)$.

The simplest is to approximate both f_{MAJ} and f_{IF} by the constant function 0 (i.e., $(x_1, x_2, x_3) = 0$ and $(y_1, y_2, y_3) = 0$) as has been done by GH [4]. These approximations, however, give rise to two impossible conditions as has been discussed in Section 3. There are four linear approximations of f_{IF} which do not have any impossible conditions. In Table 3 we consider the situation where f_{MAJ} is approximated by zero and f_{IF} is approximated by zero and the four other linear functions which do not have impossible conditions. From Table 2, we find that there are 16 possible combinations of linear approximations of f_{MAJ} and f_{IF} which do not have any impossible conditions. These are listed in Table 4.

5 Technique for Finding Local Collisions

We describe the method for finding a local collision spanning k steps. For the local collision to exist, the difference of registers at the start and at the end must be zero. Besides, the first and the last message differences must not be zero, to make it exactly a k -step collision.

Table 3. Linear approximations for $f_{MAJ}(a, b, c)$, $f_{IF}(e, f, g)$ and the corresponding $(x_1, x_2, x_3, y_1, y_2, y_3)$. Case A has been considered by Gilbert-Handschuh. It has one impossible condition each for both f_{MAJ} and f_{IF} . Cases B to E have one impossible condition for f_{MAJ} and none for f_{IF} .

Case	$f_{MAJ}(a, b, c)$	$f_{IF}(e, f, g)$	(x_1, x_2, x_3)	(y_1, y_2, y_3)
A	0	0	(0,0,0)	(0,0,0)
B	0	g_{i-1}	(0,0,0)	(0,0,1)
C	0	f_{i-1}	(0,0,0)	(0,1,0)
D	0	$e_{i-1} \oplus g_{i-1}$	(0,0,0)	(1,0,1)
E	0	$e_{i-1} \oplus f_{i-1}$	(0,0,0)	(1,1,0)

Table 4. Linear approximations for $f_{MAJ}(a, b, c)$ and $f_{IF}(e, f, g)$ and corresponding $(x_1, x_2, x_3, y_1, y_2, y_3)$. These approximations do not have any impossible conditions for either f_{MAJ} or f_{IF} .

Case	$f_{MAJ}(a, b, c)$	$f_{IF}(e, f, g)$	(x_1, x_2, x_3)	(y_1, y_2, y_3)
1	a	f	(1,0,0)	(0,1,0)
2	a	g	(1,0,0)	(0,0,1)
3	a	$e \oplus f$	(1,0,0)	(1,1,0)
4	a	$e \oplus g$	(1,0,0)	(1,0,1)
5	b	f	(0,1,0)	(0,1,0)
6	b	g	(0,1,0)	(0,0,1)
7	b	$e \oplus f$	(0,1,0)	(1,1,0)
8	b	$e \oplus g$	(0,1,0)	(1,0,1)
9	c	f	(0,0,1)	(0,1,0)
10	c	g	(0,0,1)	(0,0,1)
11	c	$e \oplus f$	(0,0,1)	(1,1,0)
12	c	$e \oplus g$	(0,0,1)	(1,0,1)
13	$a \oplus b \oplus c$	f	(1,1,1)	(0,1,0)
14	$a \oplus b \oplus c$	g	(1,1,1)	(0,0,1)
15	$a \oplus b \oplus c$	$e \oplus f$	(1,1,1)	(1,1,0)
16	$a \oplus b \oplus c$	$e \oplus g$	(1,1,1)	(1,0,1)

The basic idea is to iterate the linear system in the forward direction; equate the register values to 0 after k steps and then solve the resulting equations. The forward iteration is done in the following manner.

1. $\Delta\text{reg}_0 = (0, 0, 0, 0, 0, 0, 0, 0)$.
2. For $i = 1$ to k do
3. $(\Delta\text{reg}_i)^t = A(\Delta\text{reg}_{i-1}, \Delta W_i)^t$;
4. end do.

The procedure provides Δreg_k in terms of $\Delta W_1, \dots, \Delta W_k$. We now have to set $\Delta\text{reg}_k = 0$ and solve for $\Delta W_1, \dots, \Delta W_k$. Since the expressions for Δreg_k are quite complicated, there does not seem to be any general method for solving these equations. On the other hand, the equations do have a pattern, which we have exploited to obtain solutions. We explain our method for $k = 9$ for Case B of Table 3. Similar methods have been applied to the other two cases. All our computations have been carried out using the symbolic computation package Mathematica [13].

5.1 Case B of Table 3

The actual values of Δreg_9 in this case is given in Section A. Below we show how to solve for $\Delta W_1, \dots, \Delta W_9$ under the condition $\Delta\text{reg}_9 = 0$.

Step 1: The expression for Δh_9 is of the form

$$\begin{aligned} \Delta h_9 = & \Delta W_6 \oplus \Sigma_1(\Delta W_5) \oplus \Sigma_1^2(\Delta W_4) \oplus \Delta W_3 \oplus \Sigma_1^3(\Delta W_3) \oplus \Sigma_1^4(\Delta W_2) \oplus \Sigma_0(\Delta W_1) \\ & \oplus \Sigma_1^2(\Delta W_1) \oplus \Sigma_1^5(\Delta W_1). \end{aligned}$$

Setting $\Delta h_9 = 0$ provides

$$\begin{aligned} \Delta W_6 = & \Sigma_1(\Delta W_5) \oplus \Sigma_1^2(\Delta W_4) \oplus \Delta W_3 \oplus \Sigma_1^3(\Delta W_3) \oplus \Sigma_1^4(\Delta W_2) \oplus \Sigma_0(\Delta W_1) \\ & \oplus \Sigma_1^2(\Delta W_1) \oplus \Sigma_1^5(\Delta W_1). \end{aligned} \tag{5}$$

Step 2: Eliminating ΔW_6 from $(\Delta a_9, \dots, \Delta g_9)$ using (5), we obtain

$$\begin{aligned} \Delta g_9 = & \Delta W_7 \oplus \Delta W_4 \oplus \Sigma_1(\Delta W_3) \oplus \Sigma_0(\Delta W_2) \oplus \Sigma_1^2(\Delta W_2) \oplus \Delta W_1 \oplus \Sigma_0^2(\Delta W_1) \\ & \oplus \Sigma_0(\Sigma_1(\Delta W_1)) \oplus \Sigma_1^3(\Delta W_1). \end{aligned}$$

Setting $\Delta g_9 = 0$ provides

$$\begin{aligned} \Delta W_7 = & W_4 \oplus \Sigma_1(\Delta W_3) \oplus \Sigma_0(\Delta W_2) \oplus \Sigma_1^2(\Delta W_2) \oplus \Delta W_1 \oplus \Sigma_0^2(\Delta W_1) \\ & \oplus \Sigma_0(\Sigma_1(\Delta W_1)) \oplus \Sigma_1^3(\Delta W_1). \end{aligned} \tag{6}$$

Step 3: Eliminating ΔW_7 from $(\Delta a_9, \dots, \Delta f_9)$ using (6), we obtain

$$\begin{aligned} \Delta f_9 = & \Delta W_8 \oplus \Delta W_5 \oplus \Sigma_1(\Delta W_4) \oplus \Sigma_0(\Delta W_3) \oplus \Sigma_1^2(\Delta W_3) \oplus \Delta W_2 \oplus \Sigma_0^2(\Delta W_2) \\ & \oplus \Sigma_0(\Sigma_1(\Delta W_2)) \oplus \Sigma_1^3(\Delta W_2) \oplus \Sigma_0^3(\Delta W_1) \oplus \Sigma_0^2(\Sigma_1(\Delta W_1)) \\ & \oplus \Sigma_0(\Sigma_1^2(\Delta W_1)) \oplus \Sigma_1^4(\Delta W_1). \end{aligned}$$

Setting $\Delta f_9 = 0$ provides

$$\begin{aligned} \Delta W_8 = & \Delta W_5 \oplus \Sigma_1(\Delta W_4) \oplus \Sigma_0(\Delta W_3) \oplus \Sigma_1^2(\Delta W_3) \oplus W_2 \oplus \Sigma_0^2(\Delta W_2) \oplus \Sigma_0(\Sigma_1(\Delta W_2)) \\ & \oplus \Sigma_1^3(\Delta W_2) \oplus \Sigma_0^3(\Delta W_1) \oplus \Sigma_0^2(\Sigma_1(\Delta W_1)) \oplus \Sigma_0(\Sigma_1^2(\Delta W_1)) \oplus \Sigma_1^4(\Delta W_1). \end{aligned} \tag{7}$$

Step 4: Eliminating ΔW_8 in $(\Delta a_9, \dots, \Delta e_9)$ using (7) we obtain

$$\begin{aligned}\Delta e_9 = & \Delta W_9 \oplus \Sigma_0(\Delta W_4) \oplus \Sigma_0^2(\Delta W_3) \oplus \Sigma_0(\Sigma_1(\Delta W_3)) \oplus \Sigma_0^3(\Delta W_2) \oplus \Sigma_0^2(\Sigma_1(\Delta W_2)) \\ & \oplus \Sigma_0(\Sigma_1^2(\Delta W_2)) \oplus \Delta W_9 \oplus \Sigma_0(\Delta W_1) \oplus \Sigma_0^4(\Delta W_1) \oplus \Sigma_0^3(\Sigma_1(\Delta W_1)) \\ & \oplus \Sigma_0^2(\Sigma_1^2(\Delta W_1)) \oplus \Sigma_0(\Sigma_1^3(\Delta W_1)).\end{aligned}$$

Setting $\Delta e_9 = 0$ provides

$$\begin{aligned}\Delta W_9 = & \Sigma_0(\Delta W_4) \oplus \Sigma_0^2(\Delta W_3) \oplus \Sigma_0(\Sigma_1(\Delta W_3)) \oplus \Sigma_0^3(\Delta W_2) \oplus \Sigma_0^2(\Sigma_1(\Delta W_2)) \\ & \oplus \Sigma_0(\Sigma_1^2(\Delta W_2)) \oplus \Delta W_1 \oplus \Sigma_0(\Delta W_1) \oplus \Sigma_0^4(\Delta W_1) \oplus \Sigma_0^3(\Sigma_1(\Delta W_1)) \\ & \oplus \Sigma_0^2(\Sigma_1^2(\Delta W_1)) \oplus \Sigma_0(\Sigma_1^3(\Delta W_1)).\end{aligned}\tag{8}$$

Step 5: Eliminating ΔW_9 in $(\Delta a_9, \dots, \Delta d_9)$ using (7) we obtain

$$\begin{aligned}\Delta d_9 = & \Sigma_0(\Delta W_5) \oplus \Sigma_0^2(\Delta W_4) \oplus \Sigma_0(\Sigma_1(\Delta W_4)) \oplus \Sigma_0^3(\Delta W_3) \oplus \Sigma_0^2(\Sigma_1(\Delta W_3)) \oplus \\ & \Sigma_0(\Sigma_1^2(\Delta W_3)) \oplus \Delta W_2 \oplus \Sigma_0(\Delta W_2) \oplus \Sigma_0^4(\Delta W_2) \oplus \Sigma_0^3(\Sigma_1(\Delta W_2)) \oplus \\ & \Sigma_0^2(\Sigma_1^2(\Delta W_2)) \oplus \Sigma_0(\Sigma_1^3(\Delta W_2)) \oplus \Sigma_0^2(\Delta W_1) \oplus \Sigma_0^5(\Delta W_1) \oplus \Sigma_1(\Delta W_1) \oplus \\ & \Sigma_0^4(\Sigma_1(\Delta W_1)) \oplus \Sigma_0^3(\Sigma_1^2(\Delta W_1)) \oplus \Sigma_0^2(\Sigma_1^3(\Delta W_1)) \oplus \Sigma_0(\Sigma_1^4(\Delta W_1)).\end{aligned}$$

Now the situation is different from the previous 4 steps. In the expression for Δd_9 we do not have any ΔW_i whose “coefficient” is 1. Only ΔW_5 occurs once with a “coefficient” of Σ_0 . We solve for ΔW_5 in the following manner. Set

$$\Delta W_2 = \Sigma_0(x) \oplus \Sigma_1(\Delta W_1)\tag{9}$$

where x is a variable to be determined later. With this substitution, we have $\Delta d_9 = \Sigma_0(\Delta W_5 \oplus X)$, for some expression X which we provide shortly. Now setting $\Delta d_9 = 0$, provides one solution to be $\Delta W_5 = X$, where the value of X is given by the right side of the following expression.

$$\begin{aligned}\Delta W_5 = & (1 \oplus \Sigma_0 \oplus \Sigma_0^4 \Sigma_0^3 \Sigma_1 \oplus \Sigma_0^2 \Sigma_1^2 \oplus \Sigma_0 \Sigma_1^3)(x) \oplus \Sigma_0(\Delta W_4) \oplus \Sigma_1(\Delta W_4) \oplus \Sigma_0^2(\Delta W_3) \\ & \oplus \Sigma_0(\Sigma_1(\Delta W_3)) \oplus \Sigma_1^2(\Delta W_3) \oplus \Sigma_0(\Delta W_1) \oplus \Sigma_0^4(\Delta W_1) \oplus \Sigma_1(\Delta W_1).\end{aligned}\tag{10}$$

Step 6: Eliminating ΔW_5 in $(\Delta a_9, \Delta b_9, \Delta c_9)$ using (10) we obtain

$$\Delta c_9 = \Sigma_0^2(x) \oplus \Sigma_0(\Sigma_1(x)) \oplus \Delta W_3 \oplus \Sigma_0^2(\Delta W_1).$$

Setting $\Delta c_9 = 0$ provides

$$\Delta W_3 = \Sigma_0^2(x) \oplus \Sigma_0(\Sigma_1(x)) \oplus \Sigma_0^2(\Delta W_1).\tag{11}$$

Step 7: Eliminating ΔW_3 in $(\Delta a_9, \Delta b_9)$ using (11) we obtain

$$\Delta b_9 = \Sigma_0^2(\Sigma_1(x)) \oplus \Delta W_4 \oplus \Delta W_1 \oplus \Sigma_0^2(\Sigma_1(\Delta W_1)).$$

Setting $\Delta b_9 = 0$, provides

$$\Delta W_4 = \Sigma_0^2(\Sigma_1(x)) \oplus \Delta W_1 \oplus \Sigma_0^2(\Sigma_1(\Delta W_1)).\tag{12}$$

Step 8: Eliminating ΔW_4 from Δa_9 using (12), we obtain

$$\Delta a_9 = x \oplus \Delta W_1.$$

Setting $\Delta a_9 = 0$ provides

$$\Delta W_1 = x. \quad (13)$$

Equations (5), (6), (7), (8), (9), (10), (11), (12), and (13) form a solution to the problem of finding a local collision for the linearized round function. In this form, the equations are not easy to handle. But, if we start the process of back substitution, i.e., use $\Delta W_1 = x$ in (12) and then use the values of ΔW_1 and ΔW_4 in (11) and so on, then the solution is substantially simplified and we finally obtain

$$(\Delta W_1, \dots, \Delta W_9) = (x, \Sigma_0(x) \oplus \Sigma_1(x), \Sigma_0(\Sigma_1(x)), x, \Sigma_0(x) \oplus x, \Sigma_0(x) \oplus \Sigma_1(x), 0, x, x).$$

5.2 A Difficult Example: Case 3 of Table 4

The technique described in the previous subsection does not work always. There are cases when we cannot solve the equations in the manner described earlier. A slightly modified method is used for such cases. We briefly describe this procedure for the Case 3 of Table 4. The actual values of Δreg_9 are given in Section B for this case.

Steps 1 to 4: Using the method described earlier, we can obtain

$$\begin{aligned} \Delta W_9 = & \Sigma_0^2 \Delta W_1 \oplus \Sigma_0^3 \Delta W_1 \oplus \Sigma_0^4 \Delta W_1 \oplus \Sigma_0^2 \Sigma_1 \Delta W_1 \oplus \Sigma_0^3 \Sigma_1 \Delta W_1 \oplus \Sigma_0 \Sigma_1^2 \Delta W_1 \\ & \oplus \Sigma_0^2 \Sigma_1^2 \Delta W_1 \oplus \Sigma_1^3 \Delta W_1 \oplus \Sigma_0 \Sigma_1^3 \Delta W_1 \oplus \Sigma_0 \Delta W_2 \oplus \Sigma_0^3 \Delta W_2 \oplus \Sigma_1 \Delta W_2 \\ & \oplus \Sigma_0^2 \Sigma_1 \Delta W_2 \oplus \Sigma_1^2 \Delta W_2 \oplus \Sigma_0 \Sigma_1^2 \Delta W_2 \oplus \Sigma_0 \Delta W_3 \oplus \Sigma_0^2 \Delta W_3 \oplus \Sigma_1 \Delta W_3 \\ & \oplus \Sigma_0 \Sigma_1 \Delta W_3 \oplus \Delta W_4 \oplus \Sigma_0 \Delta W_4 \end{aligned} \quad (14)$$

$$\begin{aligned} \Delta W_8 = & \Sigma_0 \Delta W_1 \oplus \Sigma_0^3 \Delta W_1 \oplus \Sigma_1 \Delta W_1 \oplus \Sigma_0^2 \Sigma_1 \Delta W_1 \oplus \Sigma_1^2 \Delta W_1 \oplus \Sigma_0 \Sigma_1^2 \Delta W_1 \\ & \oplus \Sigma_0 \Delta W_2 \oplus \Sigma_0^2 \Delta W_2 \oplus \Sigma_1 \Delta W_2 \oplus \Sigma_0 \Sigma_1 \Delta W_2 \oplus \Delta W_3 \oplus \Sigma_0 \Delta W_3 \end{aligned} \quad (15)$$

$$\begin{aligned} \Delta W_7 = & \Delta W_1 \oplus \Sigma_0 \Delta W_1 \oplus \Sigma_0^2 \Delta W_1 \oplus \Sigma_1 \Delta W_1 \oplus \Sigma_0 \Sigma_1 \Delta W_1 \oplus \Sigma_1^2 \Delta W_1 \\ & \oplus \Sigma_1^4 \Delta W_1 \oplus \Sigma_0 \Delta W_2 \oplus \Sigma_1 \Delta W_2 \oplus \Sigma_1^2 \Delta W_2 \oplus \Sigma_1^3 \Delta W_2 \oplus \Sigma_1^2 \Delta W_3 \\ & \oplus \Delta W_4 \oplus \Sigma_1 \Delta W_4 \oplus \Delta W_5 \end{aligned} \quad (16)$$

$$\begin{aligned} \Delta W_6 = & \Delta W_1 \oplus \Sigma_0 \Delta W_1 \oplus \Sigma_1^4 \Delta W_1 \oplus \Sigma_1^5 \Delta W_1 \oplus \Delta W_2 \oplus \Sigma_1^2 \Delta W_2 \oplus \Sigma_1^4 \Delta W_2 \\ & \oplus \Delta W_3 \oplus \Sigma_1 \Delta W_3 \oplus \Sigma_1^2 \Delta W_3 \oplus \Sigma_1^3 \Delta W_3 \oplus \Sigma_1^2 \Delta W_4 \oplus \Delta W_5 \oplus \Sigma_1 \Delta W_5 \end{aligned} \quad (17)$$

Now in the expression for Δd_9 , we do not have any ΔW_i with coefficient “1”. Therefore, we let the sum of all the terms which do not have Σ_0 in their coefficients be $\Sigma_0 X$. This substitution results in

$$\begin{aligned} \Delta W_5 = & \Sigma_0 X \oplus \Delta W_1 \oplus \Sigma_1 \Delta W_1 \oplus \Sigma_1^2 \Delta W_1 \oplus \Sigma_1^3 \Delta W_1 \oplus \Sigma_1^4 \Delta W_1 \oplus \Sigma_1^3 \Delta W_2 \\ & \oplus \Sigma_1 \Delta W_3 \oplus \Sigma_1^2 \Delta W_3 \oplus \Sigma_1 \Delta W_4 \end{aligned} \quad (18)$$

where X is a variable whose value is not yet known.

Substituting this expression for ΔW_5 in $\Delta d_9 = 0$, we still get an equation in which none of the variables has a coefficient of Σ_0 only. To get such a variable, we sum all the terms which have no Σ_0 coefficient and equate this to $\Sigma_0 Y$, where Y is another variable. This results in the following substitution

$$\Delta W_4 = \Sigma_0 Y \oplus X \oplus \Delta W_1 \oplus \Sigma_1 \Delta W_1 \oplus \Sigma_1^2 \Delta W_1 \oplus \Sigma_1^3 \Delta W_1 \oplus \Sigma_1^2 \Delta W_2 \oplus \Delta W_3 \oplus \Sigma_1 \Delta W_3 \quad (19)$$

Now we need to solve $\Delta d_9 = 0$. Still the form of this equation is

$$\begin{aligned} \Delta d_9 = & \Sigma_0^5 \Delta W_1 \oplus \Sigma_0^4 \Sigma_1 \Delta W_1 \oplus \Sigma_0^2 \Sigma_1^2 \Delta W_1 \oplus \Sigma_0^3 \Sigma_1^2 \Delta W_1 \oplus \Sigma_0^2 \Delta W_2 \oplus \Sigma_0^3 \Delta W_2 \oplus \\ & \Sigma_0^4 \Delta W_2 \oplus \Sigma_0^2 \Sigma_1 \Delta W_2 \oplus \Sigma_0^3 \Sigma_1 \Delta W_2 \oplus \Sigma_0^2 \Delta W_3 \oplus \Sigma_0^3 \Delta W_3 \oplus \Sigma_0^2 Y \oplus \Sigma_0^3 Y \end{aligned} \quad (20)$$

In this expression for Δd_9 , we note that the coefficient of ΔW_3 is $\Sigma_0^2(1 + \Sigma_0)$. To solve for ΔW_3 we try to generate the same coefficient in other terms too. This can be done if we substitute

$$\Delta W_2 = \Sigma_0 \Delta W_1 \oplus c1 \Delta W_1 \oplus (1 \oplus \Sigma_0)Z \quad (21)$$

where Z is another variable unknown as of now. With these substitutions, $\Delta d_9 = 0$ gives

$$\Delta W_3 = \Sigma_0 \Delta W_1 \oplus \Sigma_1 \Delta W_1 \oplus \Sigma_0 \Sigma_1 \Delta W_1 \oplus Y \oplus Z \oplus \Sigma_0^2 Z \quad (22)$$

Now solving for $\Delta c_9 = 0, \Delta b_9 = 0$ and $\Delta a_9 = 0$ with these values of $\Delta W_9 \dots \Delta W_3$ substituted, we get

$$Z = 0 \quad (23)$$

$$Y = 0 \quad (24)$$

$$X = 0 \quad (25)$$

Taking ΔW_1 to be x and then back substituting all the variables results in the solution

$$\begin{aligned} (\Delta W_1, \dots, \Delta W_9) = \\ (x, \Sigma_0(x) \oplus \Sigma_1(x), \Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x)), x \oplus \Sigma_0(x), x, \Sigma_0(x) \oplus \Sigma_1(x), x, 0, x). \end{aligned}$$

6 Differential Path

The values of the XOR differences of the registers at each step constitute a differential path. For a local collision, the initial and final XOR differences should be zero.

Probability of Differential Path: A differential path holds with probability one for the linearized version of the round function. However, when we move to the actual round function, then it holds with lesser probability which in some cases may even be zero. If the differential path holds with probability zero for the actual round function, then we call it to be an impossible differential path. Such impossible differential paths arise due to the impossible conditions in the approximations of the constituent functions by linear functions. Later we will show examples of such differential paths including one obtained from the Gilbert-Handschuh local collision.

We next discuss how to compute the probability for a differential path. This computation is based on the following two points.

1. If a and b differ in one bit position, then $a + c$ and $b + c$ also differ in one bit position with probability one if the differing bit is the most significant bit, else with probability half. (This was also mentioned in [4].) We also assume that if a and b differ in k different bit positions none of which is the most significant bit, then $a + c$ and $b + c$ differ in these k positions with probability $1/2^k$.
2. Table 1 is used to determine the differential probabilities for the approximations of f_{IF} and f_{MAJ} .

Since the XOR and additive differences coincide for the most significant bit, to achieve higher probability, it is advantageous to ensure that many bits in the differential path are MSBs. Based on this observation, we choose $x = 2^{31}$ for SHA-256 and $x = 2^{512}$ for SHA-512 in Table 6 and compute the resulting probabilities. An example of illustration of probability calculations is given in Section D.

7 Reduced round collisions for the SHA-2 family

In this section, we show that it is possible to combine the presented local collisions for getting upto 18 step reduced round collisions for the SHA-2 family. We specify the first step in SHA-2 by Step 0.

First of all, note that all the local collisions discussed in the present work span 9 steps and the message expansion of SHA-2 does not play any role in first 16 steps. Therefore if a local collision spans from Step i to Step $(i + 9)$, and if we take $W_0 = W_1 = \dots = W_{i-1} = W_{i+10} = W_{i+11} = \dots W_{16} = 0$, we get a collision for first 16 steps of SHA-2. All the 16 local collisions described in this work can be used to generate 16 step collisions for the SHA-2 family in this manner.

The 16 step collisions described above are not very interesting since we have completely by-passed the issue of message expansion in obtaining them. Now we tackle the first step of message expansion. Message expansion rule for W_{16} is given by :

$$W_{16} = \sigma_1(W_{14}) + W_9 + \sigma_0(W_1) + W_0 \quad (26)$$

A local collision which starts at Step 2 will end at Step 10. The differential path for such a local collision will have $\Delta W_0 = \Delta W_1 = \Delta W_{14} = 0$ (If we choose the differentials of all the message words outside the span of the local collision to be zero). The local collisions described by Cases 1, 3, 5, 7, 10, 12, 14 and 16 of Table 5 are such that ΔW_9 will be zero for them (Refer the differential paths for the local collisions in tables in the appendix). Thus message expansion yields that $\Delta W_{16} = 0$. Hence we have many 17 step collisions for SHA-2 using a single local collision.

In the same manner it can be shown that starting a single local collision described by Cases 2, 4, 5, 7, 13, 14, 15 or 16 of Table 5 at Step 3, we can get a 17-step collision for SHA-2. Similarly, starting a single local collision described by Cases 5, 7, 14 or 16 of Table 5 will yield $\Delta W_{16} = \Delta W_{17} = 0$. Thus we can get 4 different 18 step collisions for SHA-2.

Previously only one 18 step collision for SHA-256 has been reported in the literature [7]. In the present work we have shown that it is possible to obtain many 17 and 18 step collisions for both SHA-256 and SHA-512. We have explained the procedure for generating such collisions without actually exhibiting them. We hope to exhibit these collisions in a forthcoming paper.

To go beyond 18 steps will require combining several local collisions and it is currently being investigated. For example in [7] a 19 Step 1 bit near collision for SHA-256 is reported which is obtained by using 23 GH local collisions.

8 Results

The detailed differential paths for the cases of Table 3 are shown in Table 6. The differential paths for the cases in Table 4 are shown in Tables 8 to 11 in Section C. Each case has two columns. The first of these provide the message difference for the different steps and the second one provides the probability with which this particular step of the linearized round function behaves as a step of the actual round function. Finally, the product of all the probabilities in one column is listed as the total probability for the corresponding differential path.

From Tables 8 to 11, it is interesting to note that all approximations of f_{MAJ} by the same linear function have the same differential path. The weight of the differential path increases with the increase in the number of variables in the linear approximation of f_{MAJ} . A summary of various features of the different local collisions are given in Table 5.

Table 5. Summary of the different properties of the local collisions. Wt(DP) provides the weight of the differential path; Wt(MD) provides the weight of the message difference; Pr. provides the probability of the differential path; and NIC provides the number of impossible conditions. The cases are from Table 3 and 4. Case A is the GH local collision, rest are new local collisions.

Case	A	B	C	D	E	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Wt(DP)	24	24	24	24	24	28	28	28	28	28	28	28	28	28	28	28	28	36	36	36	36
Wt(MD)	24	29	29	34	34	35	33	35	35	29	35	35	37	35	31	37	35	37	37	43	41
Pr.	$\frac{1}{2^{42}}$	$\frac{1}{2^{45}}$	$\frac{1}{2^{45}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{51}}$	$\frac{1}{2^{51}}$	$\frac{1}{2^{49}}$	$\frac{1}{2^{49}}$	$\frac{1}{2^{52}}$	$\frac{1}{2^{52}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{48}}$	$\frac{1}{2^{51}}$	$\frac{1}{2^{51}}$	$\frac{1}{2^{54}}$	$\frac{1}{2^{54}}$	$\frac{1}{2^{57}}$	$\frac{1}{2^{57}}$
NIC	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Remark: The probability of the differential path of Case A was estimated by GH to be 2^{-66} . Our calculations show this to be 2^{-42} . In [5] this probability was computed to be 2^{-39} when using modular differences (as opposed to XOR differences considered here). Mendel et al [7] remarked (without providing details) that the probability can be higher than 2^{-39} even when considering XOR differences. We think that the relative probabilities of the different local collisions will remain the same irrespective of which method is applied to compute the probabilities.

The GH local collision (Case A) has the highest probability. It is, however, not necessarily the best possible local collision. This is due to the fact that it has two impossible conditions and may result in an impossible differential path. We illustrate this point using the impossible condition for f_{IF} . A 12-step impossible differential path for the GH local collision is shown in Table 7. This is obtained by interleaving two GH local collisions with the second one starting at the fourth step of the first one. In terms of the Chabaud-Joux [2] type disturbance vector, the 12-step differential path is given by the vector 1001. Here, $\Delta e_6 = 0$, $\Delta f_6 = x \oplus \Sigma_0(x)$ and $\Delta g_6 = x$. This shows that whatever be the value of x , there will be one bit position where the differential input to f_{IF} is $(0, 1, 1)$. From Table 1 we have Δf_{IF} to be 1 with probability 1, whereas the approximation of f_{IF} by $l = 0$ will have $\Delta l = 0$. This shows that although the differential path is valid for the linearized version with f_{IF} approximated by $l = 0$, it fails for the actual round function.

As mentioned earlier, the issue of impossible differential paths was also observed in [7]. They developed techniques for circumventing such impossible paths in their collision search attacks on reduced round SHA-2. On the other hand, if we use a local collision such as Case 1, then there are no impossible conditions. Consequently, no circumvention techniques will be required in collision search attacks. The probability of this local collision is a little lower than the GH local collision, but this may be offset by absence of impossible conditions. Further work on this topic can settle this point.

9 Conclusion

In this paper, we have made a systematic study of the local collisions for the SHA-2 family of hash functions. Impossible conditions have been identified in the various approximations of the constituent Boolean functions. In particular, we have shown that the previous local collision by Gilbert and Handschuh [4] has one impossible condition each in the approximation of f_{IF} and f_{MAJ} . We have presented 16 new local collisions with no impossible conditions though the probabilities are a little lower than the GH local collision. In this paper, we have not considered the issue of message expansion. Combining message expansion with the new local collisions to obtain (reduced round) collisions for the SHA-2 family is a topic of future research.

Table 6. Differential paths for the cases of Table 3. Probability calculations are done taking x to be 2^{31} for SHA-256 and 2^{63} for SHA-512.

Step	Registers								Case A		Case B		Case C		Case D		Case E	
i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr
1	x	0	0	0	x	0	0	0	x	1	x	1	x	1	x	1	x	1
2	0	x	0	0	$\Sigma_0(x)$	x	0	0	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$
3	0	0	x	0	0	$\Sigma_0(x)$	x	0	$\Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$\Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$x \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$\Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$
4	0	0	0	x	0	0	$\Sigma_0(x)$	x	0	$\frac{1}{2^5}$	x	$\frac{1}{2^5}$	$\Sigma_0(x)$	$\frac{1}{2^8}$	x	$\frac{1}{2^5}$	$\Sigma_0(x)$	$\frac{1}{2^8}$
5	0	0	0	0	x	0	0	$\Sigma_0(x)$	x	$\frac{1}{2^3}$	$\Sigma_0(x) \oplus x$	$\frac{1}{2^8}$	x	$\frac{1}{2^3}$	$x \oplus \Sigma_0(x)$	$\frac{1}{2^8}$	x	$\frac{1}{2^3}$
6	0	0	0	0	0	x	0	0	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$
7	0	0	0	0	0	0	x	0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$
8	0	0	0	0	0	0	0	x	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$
9	0	0	0	0	0	0	0	0	x	1	x	1	x	1	x	1	x	1
Total Probability									$\frac{1}{2^{42}}$		$\frac{1}{2^{45}}$		$\frac{1}{2^{45}}$		$\frac{1}{2^{48}}$		$\frac{1}{2^{48}}$	

Table 7. A 12-step impossible differential path for the Gilbert-Hands Schuh local collision. In Step 6, the registers (e, f, g) have values $(0, 1, 1)$ for some bit, for any non-zero x . These registers are input to f_{IF} which outputs bit 1 with probability 1 in this case. This shows that it can't be approximated by 0 function.

Step i	ΔW_i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i
1	x	x	0	0	0	x	0	0	0
2	$\Sigma_0(x) \oplus \Sigma_1(x)$	0	x	0	0	$\Sigma_0(x)$	x	0	0
3	$\Sigma_0(x) \oplus \Sigma_1(x)$	0	0	x	0	0	$\Sigma_0(x)$	x	0
4	x	x	0	0	x	x	0	$\Sigma_0(x)$	x
5	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	0	x	0	0	$x \oplus \Sigma_0(x)$	x	0	$\Sigma_0(x)$
6	$\Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	0	0	x	0	0	$x \oplus \Sigma_0(x)$	x	0
7	0	0	0	0	x	0	0	$x \oplus \Sigma_0(x)$	x
8	x	0	0	0	0	x	0	0	$x \oplus \Sigma_0(x)$
9	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	0	0	0	0	0	x	0	0
10	0	0	0	0	0	0	0	x	0
11	0	0	0	0	0	0	0	0	x
12	x	0	0	0	0	0	0	0	0

References

1. Eli Biham, Rafi Chen, Antoine Joux, Patrick Carribault, Christophe Lemuet, and William Jalby. Collisions of SHA-0 and reduced SHA-1. In Cramer [3], pages 36–57.
2. Florent Chabaud and Antoine Joux. Differential collisions in SHA-0. In Hugo Krawczyk, editor, *Advances in Cryptology, CRYPTO '98*, volume 1462 of *Lecture Notes in Computer Science*, pages 56–71. Springer, 1998.
3. Ronald Cramer, editor. *Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings*, volume 3494 of *Lecture Notes in Computer Science*. Springer, 2005.
4. Henri Gilbert and Helena Handschuh. Security analysis of SHA-256 and sisters. In Mitsuru Matsui and Robert J. Zuccherato, editors, *Selected Areas in Cryptography, 2003*, volume 3006 of *Lecture Notes in Computer Science*, pages 175–193. Springer, 2003.
5. Philip Hawkes, Michael Paddon, and Gregory G. Rose. On corrective patterns for the SHA-2 family. *Cryptology eprint Archive*, August 2004. Available at <http://eprint.iacr.org/2004/207>.
6. Krystian Matusiewicz, Josef Pieprzyk, Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Analysis of simplified variants of SHA-256. In Christopher Wolf, Stefan Lucks, and Po-Wah Yau, editors, *WEWoRC*, volume 74 of *LNI*, pages 123–134. GI, 2005.
7. Florian Mendel, Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Analysis of step-reduced SHA-256. In Matthew J. B. Robshaw, editor, *Fast Software Encryption, 2006*, volume 4047 of *Lecture Notes in Computer Science*, pages 126–143. Springer, 2006.
8. Vincent Rijmen and Elisabeth Oswald. Update on SHA-1. In Alfred Menezes, editor, *CT-RSA*, volume 3376 of *Lecture Notes in Computer Science*, pages 58–71. Springer, 2005.
9. Secure Hash Standard. *Federal Information Processing Standard Publication 180-2*. U.S. Department of Commerce, National Institute of Standards and Technology(NIST), 2002. Available at : <http://csrc.nist.gov/encryption/tkhash.html>.
10. Xiaoyun Wang, Xuejia Lai, Dengguo Feng, Hui Chen, and Xiuyuan Yu. Cryptanalysis of the hash functions MD4 and RIPEMD. In Cramer [3], pages 1–18.
11. Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full SHA-1. In Victor Shoup, editor, *Advances in Cryptology, CRYPTO 2005*, volume 3621 of *Lecture Notes in Computer Science*, pages 17–36. Springer, 2005.
12. Xiaoyun Wang and Hongbo Yu. How to break MD5 and other hash functions. In Cramer [3], pages 19–35.
13. Stephen Wolfram. *The Mathematica Book*. Wolfram Media, 5th edition, 2003. <http://www.wolfram.com>.
14. Hirotaka Yoshida and Alex Biryukov. Analysis of a SHA-256 variant. In Bart Preneel and Stafford E. Tavares, editors, *Selected Areas in Cryptography, 2005*, volume 3897 of *Lecture Notes in Computer Science*, pages 245–260. Springer, 2005.

A The Values of Δ_{reg_9} for Case B of Table 3

In Section 5.1, we show how to solve for $\Delta W_1, \dots, \Delta W_9$ by setting $\Delta a_9 = \dots = \Delta h_9 = 0$.

Table 8. Differential paths for the cases of Table 4. Probability calculations are done taking x to be 2^{31} for SHA-256 and 2^{63} for SHA-512.

Step	Registers								Case 1		Case 2		Case 3		Case 4	
i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr
1	x	0	0	0	x	0	0	0	x	1	x	1	x	1	x	1
2	0	x	0	0	$x \oplus \Sigma_0(x)$	x	0	0	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$
3	0	0	x	0	0	$x \oplus \Sigma_0(x)$	x	0	$x \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$\Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$\Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{20}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{20}}$
4	0	0	0	x	0	0	$x \oplus \Sigma_0(x)$	x	$x \oplus \Sigma_0(x)$	$\frac{1}{2^7}$	x	$\frac{1}{2^4}$	$x \oplus \Sigma_0(x)$	$\frac{1}{2^7}$	x	$\frac{1}{2^4}$
5	0	0	0	0	x	0	0	$x \oplus \Sigma_0(x)$	x	$\frac{1}{2^4}$	$\Sigma_0(x)$	$\frac{1}{2^7}$	x	$\frac{1}{2^4}$	$\Sigma_0(x)$	$\frac{1}{2^7}$
6	0	0	0	0	0	x	0	0	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$
7	0	0	0	0	0	0	x	0	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$
8	0	0	0	0	0	0	0	x	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$
9	0	0	0	0	0	0	0	0	x	1	x	1	x	1	x	1
Total Probability									$\frac{1}{2^{48}}$		$\frac{1}{2^{48}}$		$\frac{1}{2^{51}}$		$\frac{1}{2^{51}}$	

Table 9. Differential paths for the cases of Table 4. Probability calculations are done taking x to be 2^{31} for SHA-256 and 2^{63} for SHA-512.

Step	Registers								Case 5		Case 6		Case 7		Case 8	
i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr
1	x	0	0	0	x	0	0	0	x	1	x	1	x	1	x	1
2	0	x	0	0	$\Sigma_0(x)$	x	0	0	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$
3	0	0	x	0	x	$\Sigma_0(x)$	x	0	$\Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$x \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$\Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$
4	0	0	0	x	0	x	$\Sigma_0(x)$	x	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_1(x)$	$\frac{1}{2^8}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_1(x)$	$\frac{1}{2^8}$
5	0	0	0	0	x	0	x	$\Sigma_0(x)$	0	$\frac{1}{2^4}$	$x \oplus \Sigma_0(x)$	$\frac{1}{2^7}$	0	$\frac{1}{2^4}$	$x \oplus \Sigma_0(x)$	$\frac{1}{2^7}$
6	0	0	0	0	0	x	0	x	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$
7	0	0	0	0	0	0	x	0	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$
8	0	0	0	0	0	0	0	x	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$
9	0	0	0	0	0	0	0	0	x	1	x	1	x	1	x	1
Total Probability									$\frac{1}{2^{49}}$		$\frac{1}{2^{49}}$		$\frac{1}{2^{52}}$		$\frac{1}{2^{52}}$	

Table 10. Differential paths for the cases of Table 4. Probability calculations are done taking x to be 2^{31} for SHA-256 and 2^{63} for SHA-512.

Step	Registers								Case 9		Case 10		Case 11		Case 12	
i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr
1	x	0	0	0	x	0	0	0	x	1	x	1	x	1	x	1
2	0	x	0	0	$\Sigma_0(x)$	x	0	0	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$
3	0	0	x	0	0	$\Sigma_0(x)$	x	0	$x \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$\Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{14}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$\Sigma_0(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$
4	0	0	0	x	x	0	$\Sigma_0(x)$	x	$x \oplus \Sigma_0(x)$	$\frac{1}{2^8}$	0	$\frac{1}{2^5}$	$x \oplus \Sigma_0(x)$	$\frac{1}{2^8}$	0	$\frac{1}{2^5}$
5	0	0	0	0	x	x	0	$\Sigma_0(x)$	$x \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{10}}$	$\Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{10}}$
6	0	0	0	0	0	x	x	0	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$
7	0	0	0	0	0	0	x	x	x	1	x	1	x	1	x	1
8	0	0	0	0	0	0	0	x	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$
9	0	0	0	0	0	0	0	0	x	1	x	1	x	1	x	1
Total Probability									$\frac{1}{2^{48}}$		$\frac{1}{2^{48}}$		$\frac{1}{2^{51}}$		$\frac{1}{2^{51}}$	

Table 11. Differential paths for the cases of Table 4. Probability calculations are done taking x to be 2^{31} for SHA-256 and 2^{63} for SHA-512.

Step	Registers								Case 13		Case 14		Case 15		Case 16	
i	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i	Δf_i	Δg_i	Δh_i	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr	ΔW_i	Pr
1	x	0	0	0	x	0	0	0	x	1	x	1	x	1	x	1
2	0	x	0	0	$x \oplus \Sigma_0(x)$	x	0	0	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$
3	0	0	x	0	x	$x \oplus \Sigma_0(x)$	x	0	$\Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$x \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{17}}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{20}}$	$\Sigma_0(x) \oplus \Sigma_1(x) \oplus \Sigma_0(\Sigma_1(x))$	$\frac{1}{2^{20}}$
4	0	0	0	x	x	x	$x \oplus \Sigma_0(x)$	x	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$\Sigma_1(x)$	$\frac{1}{2^8}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{11}}$	$x \oplus \Sigma_1(x)$	$\frac{1}{2^8}$
5	0	0	0	0	x	x	x	$x \oplus \Sigma_0(x)$	$\Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{10}}$	$x \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^{10}}$
6	0	0	0	0	0	x	x	x	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$\Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$	$x \oplus \Sigma_0(x) \oplus \Sigma_1(x)$	$\frac{1}{2^7}$
7	0	0	0	0	0	0	x	x	0	1	0	1	0	1	0	1
8	0	0	0	0	0	0	0	x	x	$\frac{1}{2}$	0	$\frac{1}{2}$	x	$\frac{1}{2}$	0	$\frac{1}{2}$
9	0	0	0	0	0	0	0	0	x	1	x	1	x	1	x	1
Total Probability									$\frac{1}{2^{54}}$		$\frac{1}{2^{54}}$		$\frac{1}{2^{57}}$		$\frac{1}{2^{57}}$	

D Illustration of Probability Calculation

First we calculate the probability for Step 2 of Case 1 in Table 8. For this step the f_{MAJ} function's inputs are registers a_1 , b_1 and c_1 . Differential value of the three registers is (1,0,0). From Table 1, we know that the probability that f_{MAJ} will behave as its first argument, is $\frac{1}{2}$. The f_{IF} function takes as inputs the registers e_1 , f_1 and g_1 in this step. The second and the third inputs to f_{IF} have zero differences while the first input has a 1-bit difference. In this table, f_{IF} is being approximated by the middle argument therefore the desired output difference from f_{IF} is 0. This is the case (1,0,0) of Table 1 for the f_{IF} function and in this case it will not propagate the difference with probability $\frac{1}{2}$.

In computing Δa_2 , there are 6 bits $\Sigma_0(a_1) + \Sigma_1(e_1)$ to be canceled with the input 6-bit message word difference. The probability for this to happen is $\frac{1}{2^6}$. For calculating Δe_2 , there are 3 bits in $\Sigma_1(e_1)$ to be canceled with the input message word difference and 3 bits $\Sigma_0(x)$ to be propagated into Δe_2 from input. The cancellation part's probability has already been taken care of while considering cancellation of difference in register a_2 , and the propagation part's probability is $\frac{1}{2^3}$. The combined probability due to approximations in a_2 and e_2 calculations is $\frac{1}{2^6} \times \frac{1}{2^3}$. Thus, the probability for the differential path to hold for this step is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2^6} \times \frac{1}{2^3} = \frac{1}{2^{11}}$. Probabilities for other steps can be computed similarly. Table 12 shows the calculation of probabilities for Case 1 of Table 8. In this table, p_1^i (resp. p_2^i) is the probability for f_{MAJ} (resp. f_{IF}) to behave as it's first (resp. second) argument in step i given that the differential path holds for the $(i-1)^{th}$ step and p_3^i is the probability for differences in registers a_i and e_i to follow the differential path given that f_{MAJ} and f_{IF} behave correctly in this step and the previous steps follow the differential path. The last column in this table is the product of previous 3 column entries and it is the probability for this step.

Table 12. Probability calculations for Case 1 of Table 8.

Step i	p_1^i	p_2^i	p_3^i	Pr.
1	1	1	1×1	1
2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2^6} \times \frac{1}{2^3}$	$\frac{1}{2^{11}}$
3	$\frac{1}{2}$	$\frac{1}{2^4}$	$\frac{1}{2^{12}} \times 1$	$\frac{1}{2^{17}}$
4	$\frac{1}{2}$	$\frac{1}{2^3}$	$\frac{1}{2^3} \times 1$	$\frac{1}{2^{20}}$
5	1	$\frac{1}{2^4}$	1×1	$\frac{1}{2^{24}}$
6	1	$\frac{1}{2}$	$\frac{1}{2^6} \times 1$	$\frac{1}{2^{27}}$
7	1	$\frac{1}{2}$	1×1	$\frac{1}{2}$
8	1	$\frac{1}{2}$	1×1	$\frac{1}{2}$
9	1	1	1×1	1
Total probability =				$\frac{1}{2^{48}}$

The overall probability for the differential path to hold is the product of the probabilities for each of the individual steps computed as above. This can be seen as follows: Let A_i denote the event that the differential path holds for step i . Then the probability for the differential path to hold till the 9 steps is given by:

$$\begin{aligned} \Pr(\text{Diff Path holds}) &= \Pr(A_1 \wedge A_2 \wedge \dots \wedge A_9) \\ &= \Pr(A_1) \times \Pr(A_2|A_1) \times \dots \times \Pr(A_9|(A_1 \wedge A_2 \dots \wedge A_8)). \end{aligned}$$

The value in the probability column for the i^{th} row in Table 12 is equal to $\Pr(A_i|(A_1 \wedge A_2 \wedge \dots \wedge A_{i-1}))$. Clearly, each term in the product above is the probability of a step in Table 12. The probability for the differential path to hold till 9 steps is the product of the probabilities of individual steps. Thus, the probability for Case 1 of Table 8 is $\frac{1}{2^{48}}$. The probabilities for other differential paths have been computed similarly.