The Impact of Carries on the Complexity of Collision Attacks on SHA-1 ⋆

Florian Mendel⋆⋆, Norbert Pramstaller,
Christian Rechberger and Vincent Rijmen

Norbert.Pramstaller@iaik.tugraz.at

Institute for Applied Information Processing and Communications (IAIK)
Graz University of Technology, Austria

Abstract. In this article we present a detailed analysis of the impact of carries on the estimation of the attack complexity for SHA-1. We build up on existing estimates and refine them. We show that the attack complexity is slightly lower than estimated in all published work to date. We point out that it is more accurate to consider probabilities instead of conditions.

1 Introduction

In past years, significant progress has been made in the cryptanalysis of the hash functions MD4, MD5, RIPEMD, SHA-0, and SHA-1 [2,3,5,6,9,10,11,13,14,15]. In 2004 and 2005, Wang *et al.* announced that they had broken the hash functions MD4, MD5, RIPEMD, HAVAL, SHA-0, and SHA-1 [16,17,19,20,21].

SHA-1, a widely used hash function in practice, has attracted most attention over the last years. This year, at the CRYPTO 2005 rump session, Wang *et al.* announced that they have further improved their attack on SHA-1. They updated the attack complexity from 2^{69} to 2^{63} [18].

As it will be explained in Section 2, the attack complexity is mainly determined by the probabilities of so-called 6-step local collisions in a linearized variant of SHA-1. For each local collision, the attacker derives conditions such that the local collision holds for the original SHA-1. Based on the derived conditions the attack complexity is conjectured. The main contribution of this article is that we will show that it is more accurate to look at probabilities instead of estimating the attack complexity based on the number of conditions.

The remainder of this article is structured as following. We start with a short description of the hash function SHA-1 and review the basic attack strategy of Wang *et al.* in Section 2. In Section 3, we perform a detailed analysis of local collisions. Section 3.2 describes how Wang *et al.* derive conditions for local collisions. In Section 3.3 and Section 3.4 we present a more accurate analysis of local collisions and the corresponding probabilities. Based on these results we

⋆ The work in this paper has been supported by CRYPTREC.
⋆⋆ This author is supported by the Austrian Science Fund (FWF) project P18138.

update the complexity of the collision attack on SHA-1 in Section 3.5. Finally, we present conclusions in Section 4.

2 Collision Attacks on SHA-1

In this section we will review the basic attack strategy for collision attacks on SHA-1. We start with a short description of SHA-1, giving only the details we need later in this article.

2.1 Short Description of SHA-1

The input message is split into 512-bit message blocks (after padding). The compression function is then applied to each of these 512-bit message blocks. The compression function basically consists of two parts: the message expansion and the state update. The message expansion expands the 512-bit input message block into 80 32-bit words W_i that are used in each step of the state update. A single step of the state update is shown in Figure 1.

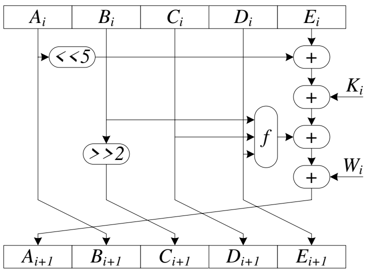

Fig. 1. One step of the state update of SHA-1

As it can be seen in Figure 1, in each step the function f is applied to the inputs B_i, C_i, and D_i. The function f depends on the step number: steps 0 to 19 (round 1) use f_{IF} and steps 40 to 59 (round 3) use f_{MAJ}. f_{XOR} is applied in the remaining steps (round 2 and 4). The functions are defined as:

$$f_{IF}(B,C,D) = B \wedge C \oplus \overline{B} \wedge D \tag{1}$$

$$f_{MAJ}(B,C,D) = B \wedge C \oplus B \wedge D \oplus C \wedge D \tag{2}$$

$$f_{XOR}(B,C,D) = B \oplus C \oplus D \ . \tag{3}$$

For a detailed description of SHA-1 refer to [12].

2.2 The Basic Attack Strategy on SHA-1

In 1998, Chabaud and Joux presented an attack on SHA-0 [3]. They used a linearized variant of SHA-0 to find a characteristic, which we will refer to as *L-characteristic* throughout the remainder of this article. For the linearized variant all modular additions are replaced by XOR and the functions f_{MAJ} and f_{IF} are replaced by f_{XOR}. They observed the following: the probability that the characteristic holds for the original SHA-1 is related to the Hamming weight of the characteristic. In general, the lower the weight the higher the probability.

In 2004 and 2005, Wang *et al.* announced that they have broken the hash functions MD4, MD5, RIPEMD, SHA-0, and SHA-1 [19,20,21]. For the collision attack on SHA-1 they use basically the following strategy, which is also depicted in Figure 2. They search for a low-weight *L-characteristic* that leads to a pseudo collision in the last 60 steps (referred to as P2 in Figure 2). Then by using a nonlinear characteristic (referred to as *NL-characteristic*) in the first 20 steps (referred to as P1 in Figure 2), they are able to turn the pseudo collision into a collision. Furthermore, they improved their attack by searching for an *L-characteristic* that leads to a pseudo-near collision in P2. As before, they turn the pseudo collision into a collision with the *NL-characteristic* and by using two-block messages they construct a collision from the near collision in each block. The fact that it is easier to find a near collision than a collision was observed already by Biham and Chen in [1]. An important property of this attack strategy is that the *NL-characteristic* has no impact on the complexity of the attack since conditions in P1 are fulfilled by using message modification techniques invented by Wang *et al.* Therefore, only the *L-characteristic* determines the attack complexity.

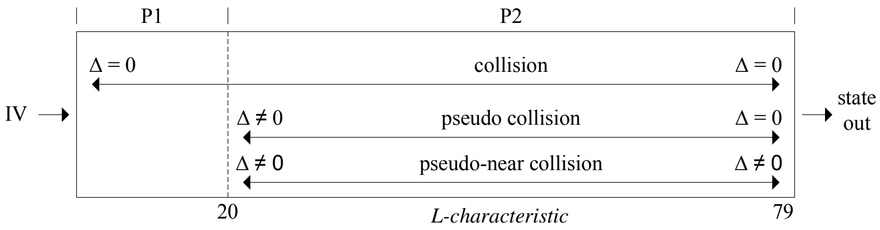

Fig. 2. Attack strategy of Wang *et al.*

The *L-characteristic* consists of overlapped single local collisions as it has been shown in [19]. To determine the attack complexity, Wang *et al.* count the number of conditions for each local collision such that it holds for the original SHA-1. Then they conjecture the attack complexity by assuming that after fulfilling the first 20 steps, random trials are performed to find the colliding messages. The complexity for this random trials is estimated to be $2^{\#\text{ conditions}}$.

Many researchers investigated the *L-characteristic* and tried to find *L-characteristics* with lower weight. A possible approach is to exploit coding theory since finding a low-weight *L-characteristic* in P2 corresponds to finding a low-weight codeword in a linear code describing P2. Results of the coding-theory approach are presented for instance in [8,11,13,15]. In 2005, Jutla and Patthak [7] used a computer aided proof to show that the minimum Hamming weight in the last

Table 1. Notation

notation	description
step	the SHA-1 compression function consists of 80 steps, $0 \leq i \leq 79$
round	the SHA-1 compression function consists of 4 rounds $= 4 \times 20$ steps
$W_{i,j}$	bit j of expanded message word in step i, $0 \leq j \leq 31$
w'_j	sign of bit difference in bit position (j mod 32), $w'_j \in \{-1, 0, +1\}$
$W'_j = w'_j 2^j$	signed bit difference in bit position (j mod 32), $W'_j \in \{-2^j, 0, +2^j\}$
$W'_{i,j}$	signed bit difference in step i, bit position j
$(j + n \bmod 32)$	bit position j rotated to the left by n positions
$(j - n \bmod 32)$	bit position j rotated to the right by n positions

60 steps of the SHA-1 message expansion is 25. This low-weight vector is also referred to as the disturbance vector, since it contains the disturbances for the single local collisions. However, Wang *et al.* use a disturbance vector with higher weight (weight = 27). The reason for this is that the vector with higher weight leads to a smaller number of conditions (see [19]). Since the attack complexity is determined by the number of local collisions and the corresponding probabilities (conditions) we will analyze them accurately in the next section.

3 Detailed Analysis of Local Collisions in SHA-1

In the first part of this section, we start with deriving the conditions and corresponding probabilities for all possible local collisions in the *L-characteristic* of SHA-1. We follow the work of Wang *et al.* in [19] to conjecture the overall probability of a collision attack on SHA-1 based on these local collisions. Note that the *L-characteristic* does not include the first 20 steps of SHA-1 and therefore, we only consider the functions f_{XOR} and f_{MAJ} described in Section 2.1. In the second part of this section, we derive a more accurate estimation of the probabilities for local collisions. With this analysis we update the attack complexity of Wang *et al.* presented in [19].

3.1 Notation and Definitions

For the analysis of local collisions we follow the notation given in Table 1. Throughout the remainder of this article we will use signed bit differences. In the following we describe the basic properties of signed bit differences that we require for our analysis. A detailed discussion of signed bit differences can be found in [4, Chapter 4].

We define the sign of a difference in bit position j as

$$w'_j = w_j - w^*_j, \quad \text{where } w_j, w^*_j \in \{0, 1\} \text{ and } w'_j \in \{-1, 0, +1\} \ . \qquad (4)$$

In particular, if $w'_j = 0$ the difference is zero. The signed bit difference is then defined as $W'_j = w'_j 2^j$. A useful property of signed bit differences is the fact that

Table 2. Addition of signed bit differences

A'_j	B'_j	C'_j	S'_j	C'_{j+1}
0	0	0	0	0
0	0	v	$(-1)^{A_j \oplus B_j} v$	$-v(A_j \oplus B_j)$
0	v	0	$(-1)^{A_j \oplus C_j} v$	$-v(A_j \oplus C_j)$
v	0	0	$(-1)^{B_j \oplus C_j} v$	$-v(B_j \oplus C_j)$
v	v	v	$(-1)^{A_j \oplus B_j \oplus 1} v$	$(-1)^{A_j \oplus B_j} v$

A'_j	B'_j	C'_j	S'_j	C'_{j+1}
0	u	v	0	$\frac{1}{2}(u+v)$
u	0	v	0	$\frac{1}{2}(u+v)$
u	v	0	0	$\frac{1}{2}(u+v)$
v	v	$-v$	$(-1)^{A_j \oplus B_j \oplus 1} v$	$(-1)^{A_j \oplus B_j} v$
v	$-v$	v	$(-1)^{A_j \oplus C_j \oplus 1} v$	$(-1)^{A_j \oplus C_j} v$
$-v$	v	v	$(-1)^{B_j \oplus C_j \oplus 1} v$	$(-1)^{B_j \oplus C_j} v$

Table 3. Differential properties of f_{XOR} and f_{MAJ} for signed bit differences

B'_j	C'_j	D'_j	$f_{XOR}(B'_j, C'_j, D'_j)$	$f_{MAJ}(B'_j, C'_j, D'_j)$
0	0	v	$(-1)^{B_j \oplus C_j} v$	$(B_j \oplus C_j)v$
0	v	0	$(-1)^{B_j \oplus D_j} v$	$(B_j \oplus D_j)v$
v	0	0	$(-1)^{C_j \oplus D_j} v$	$(C_j \oplus D_j)v$

the difference also includes information about the values of w_j and w_j^*. This is shown in (5).

$$W'_j = \begin{cases} +2^j & \text{if } w_j = 1 \text{ and } w_j^* = 0 \\ 0 & \text{if } w_j = w_j^* \\ -2^j & \text{if } w_j = 0 \text{ and } w_j^* = 1 \end{cases} \tag{5}$$

Let us now consider the addition of two signed bit differences. The addition $S = A + B$ is defined as $S_j = A_j \oplus B_j \oplus C_j$ and $C_{j+1} = f_{MAJ}(A_j, B_j, C_j)$ with $C_0 = 0$, where C_{j+1} is the resulting carry of the addition in bit position j. Table 2 lists all possible cases for the output and carry difference of a signed bit addition with $v, u \in \{-1, +1\}$.

To perform the addition of two signed bit differences we can use Table 2 for computing the resulting difference. We know that the output difference is $C'_{j+1}2^{j+1} + S'_j 2^j$. For instance, if there are two non-zero differences at the input with opposite signs, then both C'_{j+1} and S'_j are zero and hence the output difference is zero. If the differences have the same sign, for instance -2^j and -2^j, the output difference is -2^{j+1}, since $C'_{j+1} = -1$ and $S'_j = 0$.

For our analysis we need the differential properties of f_{XOR} and f_{MAJ} with respect to signed bit differences. In Table 3, we list the cases that occur in a local collision (see Figure 3) where $v \in \{-1, +1\}$. As it can be seen in Table 3, for f_{XOR} the sign of the input difference is flipped with probability 1/2 depending on the input values. For f_{MAJ} the sign is preserved but the difference propagates with probability 1/2.

3.2 Considering the Number of Conditions

In [3], Chabaud and Joux showed how the corrections for a single bit disturbance in SHA-0 can be constructed. Since the state update for SHA-0 and SHA-1 is the

Table 4. Local collision (disturbance-corrections) for SHA-1

step	**difference**			**description**
		f_{XOR}	f_{MAJ}	
i	$W'_i =$	$+2^j$	$+2^j$	single bit disturbance at bit position j
$i+1$	$W'_{i+1} =$	-2^{j+5}	-2^{j+5}	correction
$i+2$	$W'_{i+2} =$	$\pm 2^j$	-2^j	correction
$i+3$	$W'_{i+3} =$	$\pm 2^{j-2}$	-2^{j-2}	correction
$i+4$	$W'_{i+5} =$	$\pm 2^{j-2}$	-2^{j-2}	correction
$i+5$	$W'_{i+8} =$	-2^{j-2}	-2^{j-2}	correction

same, this construction is also valid for SHA-1. Table 4 shows a local collision with signed bit differences for f_{XOR} and f_{MAJ}.

For the local collision defined in Table 4, we can now derive the number of conditions and the corresponding probabilities such that the local collision holds for the original SHA-1. We refer to conditions that contain only expanded message words as *easy* conditions since we can easily fulfill them. Conditions that include state variables are considered to be *hard* conditions. For the analysis we can assume without loss of generality that the sign of the disturbance is positive, *i.e.* $W'_i = +2^j$. If the disturbance is -2^j, we get the same results by just flipping all the other signs. The propagation of the disturbance and corrections is shown in the left part of Figure 3.

Disturbance in step i. In step i, where the disturbance is introduced, it is required that the disturbance propagates to state variable A_{i+1} without causing a carry in the difference, *i.e.* $A'_{i+1} = W'_i = +2^j$. This occurs with probability 1/2. If the disturbance is introduced at bit position $j = 31$, it propagates to A'_{i+1} with probability 1.

Correction in step $i+1$. As shown in Figure 3, the difference in state variable A is rotated to the left by 5 positions. Therefore, the correction is $W'_{i+1} = -2^{j+5}$. It follows from Table 2 that if the sign of the correction is the opposite of the sign of the disturbance, then the correction occurs with probability 1. We can ensure the negative sign of the correction with condition CW_{i+1}: $W_{i+1,j+5} \oplus W_{i,j} = 1$. This condition is in W only and we can easily fulfill it.

Correction in step $i+2$. In this step, we have to consider the modular addition and the function f. As described in Table 3, f_{XOR} flips the sign of the input difference with probability 1/2. Therefore, for $B'_{i+2} = +2^j$ the output difference of f_{XOR} can be either $+2^j$ or -2^j depending on C_{i+2} and D_{i+2}. Since we cannot easily influence the values of C_{i+2} and D_{i+2} the probability for the correction is 1/2.

For f_{MAJ} we get the same probability as for f_{XOR} by defining a condition in W only. For the input difference $B'_{i+2} = +2^j$ the possible output difference

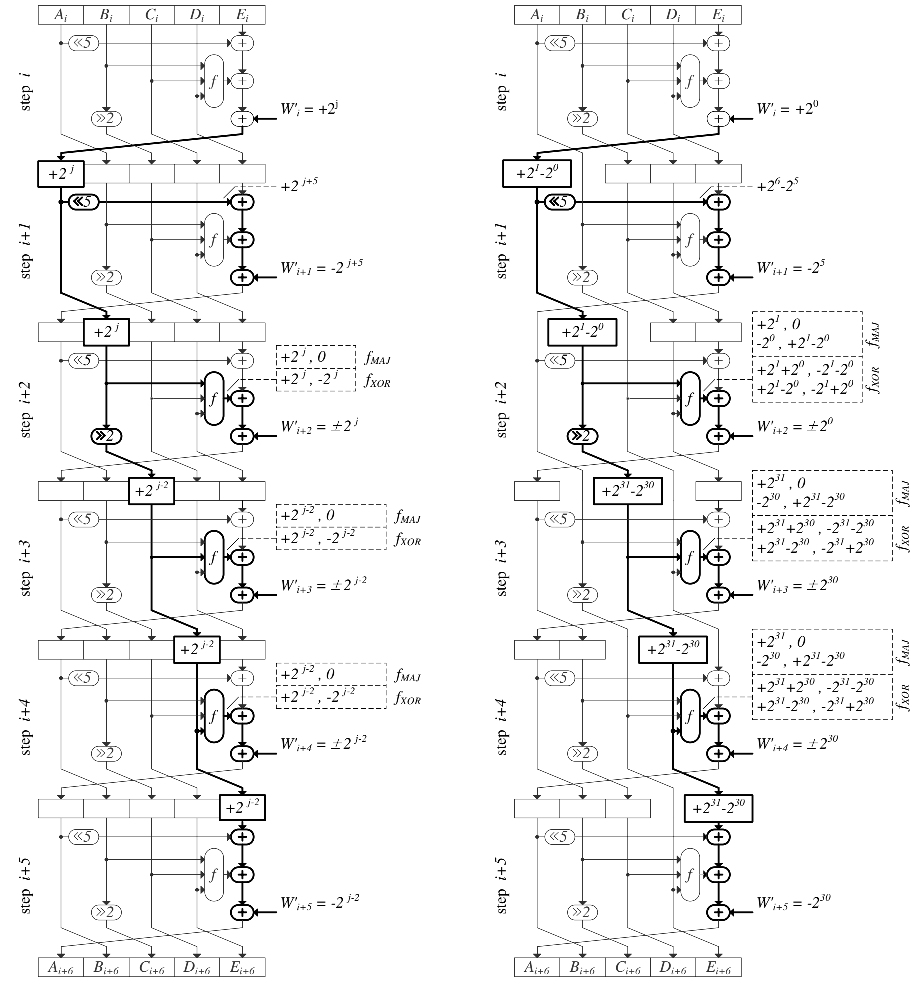

Fig. 3. On the left, a local collision with disturbance in bit position j. No carry occurs in step i. On the right a local collision with disturbance in bit position $j = 0$. In step i a carry occurs. The differences in the dashed rectangles are the possible output differences of f_{XOR} and f_{MAJ}

of f_{MAJ} is either $+2^j$ or 0. This results in a probability of 1/4. However, if the sign of the correction is negative, then the correction has a probability of 1/2. This can be ensured by fulfilling condition CW_{i+2}: $W_{i+2,j} \oplus W_{i,j} = 1$.

Correction in step $i+3$ and $i+4$. These steps are the same as step $i+2$ except that the difference $+2^j$ is rotated to the right by 2 positions, *i.e.* $+2^{j-2}$. For f_{XOR} we get a probability of 1/2 in each step. For f_{MAJ} we also get the probability 1/2 by fulfilling the following easy conditions in W only: CW_{i+3}: $W_{i+3,j-2} \oplus W_{i,j} = 1$, and CW_{i+4}: $W_{i+4,j-2} \oplus W_{i,j} = 1$.

Correction in step $i+5$. If all corrections have taken place in the previous steps the signed bit difference is in state variable E. As it can be seen in Figure 3, E'_{i+5} is the same difference as $A'_{i+1} = +2^j$ rotated by 2 to the right, *i.e.* $E'_{i+5} = +2^{j-2}$. We only have to consider the modular addition. As in step $i+1$, we can fulfill condition CW_{i+5}: $W_{i+5,j-2} \oplus W_{i,j} = 1$ such that the correction has negative sign. Hence, the correction in step $i+5$ has probability 1.

Local collision with best probability. With the above described probabilities for each step of the local collision we can define a local collision that has the best probability for f_{XOR}. Assume the disturbance is introduced in bit position $j = 1$. In step i we have a probability of 1/2. Since we can easily fulfill condition CW_{i+1} we have a probability of 1 in step $i+1$. In step $i+2$ the probability is 1/2. Now, for steps $i+3$ to $i+5$ the disturbance is rotated to bit position $j = 31$. Since a carry in the difference can be ignored (addition mod 2^{32}), we get a total probability of 2^{-2} for a local collision with disturbance in bit position $j = 1$.

Summary of probabilities of local collisions. Table 5 summarizes the probabilities for all possible local collisions with a single-bit disturbance and lists the easy conditions in W that have to be fulfilled. For the discussion so far we only considered probabilities and easy conditions. However, the probabilities for the modular addition and the functions f_{MAJ} and f_{XOR} can also be described in terms of so-called *hard* conditions. Each single condition is fulfilled with probability 1/2. Consider for instance f_{MAJ}. The input difference $B'_i = +2^j$ leads to the output difference $+2^j(C_i \oplus D_i)$ (see Table 3). In order to ensure that the difference propagates, we require that $C_i \oplus D_i = 1$. Since we cannot easily influence the values of C_i and D_i, the condition is fulfilled with probability 1/2. The same can be done for the other cases. For a local collision with disturbance in bit position $j = 1$, we have a probability of 2^{-4}. In other words there are 4 *hard* conditions that we cannot easily fulfill.

With the probabilities listed in Table 5 the complexity of the attack on SHA-1 can be determined. For the description we follow the work of Wang *et al.* [19]. For the disturbance vector [19, Table 5] we compute the product of all probabilities for each disturbance bit to determine the overall probability and hence the attack complexity.

Table 5. Probabilities for local collisions in SHA-1

	probability		easy conditions on W	
disturbance	f_{XOR}	f_{MAJ}	f_{XOR}	f_{MAJ}
$j = 1$	2^{-2}	2^{-4}	CW_{i+1}	CW_{i+1}, CW_{i+2}
$j = 26$	2^{-4}	2^{-4}	CW_{i+5}	$CW_{i+2}, CW_{i+3}, CW_{i+4}, CW_{i+5}$
$j = 31$	2^{-3}	2^{-3}	CW_{i+1}, CW_{i+5}	$CW_{i+1}, CW_{i+3}, CW_{i+4}, CW_{i+5}$
$j = 0, 2, \ldots, 25$ $j = 27, \ldots, 30$	2^{-4}	2^{-4}	CW_{i+1}, CW_{i+5}	$CW_{i+1}, CW_{i+2}, CW_{i+3}, CW_{i+4}, CW_{i+5}$

3.3 Accurate Probability Computation

In Section 3.2, we determined the probabilities of local collisions with disturbances introduced at different bit positions. For the analysis we did not allow carries in step i where the disturbance is introduced. This restriction can actually be relaxed. In the following we will analyze the impact of carries in step i on the probability of local collisions. We will show that the probabilities are actually higher for most bit positions of the disturbance.

Single bit disturbance. We start with a disturbance in bit position $j = 0$. As shown in Table 5 this results in a probability of 2^{-4}. Now consider that a carry occurs in the difference in step i, *i.e.* the disturbance $W'_i = +2^0$ propagates to $A'_{i+1} = +2^1 - 2^0$. This case is shown on the right hand side in Figure 3.

The carry in step i occurs with probability 1/4. The difference in bit position $j = 1$ can be seen as a new disturbance that leads to a second local collision with a certain probability. To cancel out the difference $A'_{i+1} = +2^1$ we require that the corrections in the consecutive steps also produce a carry in the difference. As described in Section 3.2, we fulfill condition CW_{i+1} to ensure that $W'_{i+1} = -2^5$. Therefore, the differences cancel out with probability 1 since $(+2^6 - 2^5) + (-2^5) = 0$ (as shown in Table 2, $-2^5 + (-2^5) = -2^6$ and hence $2^6 - 2^6 = 0$). For steps $i+2$ to $i+4$ we first consider f_{XOR}. In step $i+2$ we have a probability of 1/4 because f_{XOR} flips the sign of a bit difference with probability 1/2. Since we have two bit differences this results in a probability of 1/4. The same holds for steps $i+3$ and $i+4$. However, since the disturbance is introduced in bit position $j = 0$, the second difference caused by the carry is rotated to bit position $j = 31$ in step $i+2$. We can ignore carries in this bit position and hence the sign in bit position $j = 31$ has no impact. Therefore, we get a probability of 1/2 for each step. We can do the same analysis for f_{MAJ}. As already mentioned, f_{MAJ} preserves the sign of the input difference but the difference propagates only with probability 1/2. Therefore, we cannot exploit bit position $j = 31$—the probability for steps $i + 3$ and $i + 4$ is 1/4 each. For step $i + 2$ the probability is 1/4 since CW_{i+2} is fulfilled. In step $i + 5$ we have a probability of 1 for f_{XOR} and f_{MAJ} based on the same reasoning as for step $i + 1$. With the results of this analysis we can update the probability of Section 3.2. The best probability for f_{XOR} and f_{MAJ}

with a disturbance in bit position $j = 0$ is:

$$p(f_{XOR}, j = 0) = 2^{-4} + 2^{-6} = 2^{-3.6781} , \quad (6)$$
$$p(f_{MAJ}, j = 0) = 2^{-4} + 2^{-8} = 2^{-3.9125} . \quad (7)$$

Uncorrectable carries. Let us now consider the case where two carries in step i occur, *i.e.* $W'_i = +2^0$ propagates to $A'_{i+1} = +2^2 - 2^1 - 2^0$. Two carries occur with probability 1/8. If we work with the difference in bit position $j = 2$, we encounter the following problem, which we refer to as *uncorrectable carries*. In step $i+2$ the difference is rotated by two positions to the right, *i.e.* $-2^{31}-2^{30}+2^0$. It is not possible to correct the difference $+2^0$ in step $i + 3$ anymore since the correction takes place in bit position $j = 30$. For f_{MAJ}, uncorrectable carries for this example take place only in step $i + 5$. This is due to the fact that the difference $+2^0$ is blocked by f_{MAJ} with probability 1/2 in steps $i + 2$ to $i + 4$. However, in step $i + 5$ we cannot correct the difference $+2^0$ since the correction takes place in $j = 30$. Therefore, the probabilities given in (6) and (7) are the best probabilities for both functions with a disturbance in $j = 0$.

If we perform the carry analysis for bit position $j = 1$, we also encounter uncorrectable carries as for the disturbance in $j = 0$. Namely, a carry in step i cannot be corrected anymore in step $i + 3$ (step $i + 5$ for f_{MAJ}, respectively) and therefore, a carry does not increase the probability for a local collision with disturbance in $j = 1$ for both f_{XOR} and f_{MAJ}. Uncorrectable carries can also occur due to the left rotation by 5 in step $i + 1$. A disturbance in $j = 26$ that leads to a carry in step i cannot be corrected anymore in step $i + 1$ since the correction W'_{i+1} takes place in bit position $j = 31$ but the carry is rotated to $j = 0$.

Carries that improve the probability of local collisions. After determining the probabilities for $j = 0$ and $j = 1$, we describe now the impact of carry effects for disturbances in bit position $j = 2, \ldots, 31$. Due to uncorrectable carries after bit position $j = 26$ we have to analyze the probability for $j = 2, \ldots, 26$ and $j = 27, \ldots, 31$ separately. We start the explanation for f_{XOR}. For $2 \leq j \leq 26$ we have the same probability in steps i, $i+2$, $i+3$, and $i+4$, namely the probability that no carry occurs and the probabilities for all possible carries. Note that the probability in steps $i+1$ and $i+5$ is 1 since we fulfill the easy conditions CW_{i+1} and CW_{i+5} (see Section 3.2). For $27 \leq j \leq 31$ we have the same except that the probability in step $i+2$ is increased by a factor of 2 if the carry in step i reaches bit position $j = 31$. For f_{MAJ} we also assume that the easy conditions in W are fulfilled. Then we get the same probabilities as for f_{XOR} with the difference that for $27 \leq j \leq 31$ we cannot exploit bit position $j = 31$. In (8) and (9) we give the formulae to compute the accurate probability for a local collision including all carry effects. Probability bounds for (8) and (9) are given in Appendix A. For a disturbance in bit position $j = 3$ the probability for both f_{XOR} and f_{MAJ} is

$2^{-3.9068}$ instead of 2^{-4} which is the probability derived by counting conditions.

$$p(f_{XOR},j)=\begin{cases}2^{-2} & \text{for } j=1\\ 2^{-4}+2^{-6} & \text{for } j=0\\ \sum_{k=1}^{27-j}2^{-4k} & \text{for } j=2,\ldots,26\\ 2\cdot 2^{-4\cdot(32-j)}+\sum_{k=1}^{31-j}2^{-4k} & \text{for } j=27,\ldots,31\end{cases} \tag{8}$$

$$p(f_{MAJ},j)=\begin{cases}2^{-4} & \text{for } j=1\\ 2^{-3} & \text{for } j=31\\ 2^{-4}+2^{-8} & \text{for } j=0\\ \sum_{k=1}^{27-j}2^{-4k} & \text{for } j=2,\ldots,26\\ \sum_{k=1}^{32-j}2^{-4k} & \text{for } j=27,\ldots,30\end{cases} \tag{9}$$

3.4 Disturbances in consecutive bit position.

If we have a look at the disturbance vector in [19, Table 5] or [13, Table 7] there occur disturbances in consecutive bit positions, *i.e.* $W_i' = +2^{j+1}+2^j$ for f_{XOR}. For the explanation we take the concrete case with disturbance $W_i' = -2^1+2^0$, and the five corrections $W_{i+1}' = +2^6-2^5$, $W_{i+2}' = +2^1-2^0$, $W_{i+3}' = +2^{31}+2^{30}$, $W_{i+4}' = +2^{31}+2^{30}$, and $W_{i+5}' = +2^{31}-2^{30}$. In a straightforward way we can just treat them as separate disturbances and compute the probability based on (8). This results in a probability of

$$p(f_{XOR},-2^1+2^0)=\underbrace{2^{-2}}_{j=1}\cdot(\underbrace{2^{-4}+2^{-6}}_{j=0})=2^{-5.678}\ . \tag{10}$$

However, by performing a detailed analysis we show that the probability for this case can be improved to $p(f_{XOR},-2^1+2^0) = 2^{-3.678}$ by defining two additional conditions in W only, referred to as CW_i and $CW1_{i+2}$. We assume that the easy conditions described in Section 3.2 are fulfilled. If no carry occurs in step i, both disturbances are corrected with probability 2^{-6}. This follows from Section 3.2. Now consider the case that a carry occurs in step i. Assume that in step i the disturbances have opposite signs, *e.g.* $W_i' = -2^1+2^0$. This can be ensured by fulfilling the new condition CW_i: $W_{i,1}\oplus W_{i,0}=1$. If a carry occurs in bit position $j=0$ the difference that propagates to A_{i+1}' is -2^0 since the positive sign of the carry (see Table 2) cancels the negative difference in $j=1$. This occurs with probability 1/2. In step $i+1$ the probability is 1 since CW_{i+1} is fulfilled. In step $i+2$ we can increase the probability to 1/2 if the additional condition $CW1_{i+2}$: $W_{i+2,1}\oplus W_{i+2,0}=1$ is fulfilled. This is based on the same reasoning as for step i. For the remaining steps $i+3$ to $i+4$ we get a probability of 1/2 for each step. Again, in step $i+5$ we have a probability of 1. Hence we have a total probability of 2^{-4} for the case that a carry occurs in step i. Therefore, the total probability for the disturbance $+2^1-2^0$ or -2^1+2^0 is

$$p(f_{XOR},-2^1+2^0)=\underbrace{2^{-4}}_{\text{carry in j=0}}+\underbrace{2^{-6}}_{\text{no carry in step } i}=2^{-3.6781}\ . \tag{11}$$

Table 6. Update on complexity for collision attack on SHA-1

[19, Table 9]				our work
disturbance bit position	disturbance index	number of conditions	estimated probability	accurate probability
$j = 1$	$23, 24, 27, 28, 32, 35, 36$	$2 \cdot 7 = 14$	2^{-14}	2^{-14}
$j = 0$	$25, 29, 33$	$4 \cdot 3 = 12$	2^{-12}	$2^{-11.0343}$
$j = 1$	$39, 43, 45, 47, 49$	$4 \cdot 5 = 20$	2^{-20}	2^{-20}
$j = \{2, 3, 4, 5, 7\}$	$65, 68, 71, 73, 74$	$4 \cdot 5 = 20$	2^{-20}	$2^{-5 \cdot 3.9068} = 2^{-19.534}$
		total	2^{-66}	$2^{-64.5683}$

Wang *et al.* use a probability of 2^{-4} for their estimation. For disturbances in other consecutive bit positions the same analysis can be performed. For f_{XOR} the analysis is given in Appendix B.

3.5 Update of Attack Complexity by Wang *et al.*

With the above analysis we covered all cases of disturbances that occur in the disturbance vector of [19]. Since they count conditions in the last 60 steps of SHA-1 the overall probability can be updated based on (8) and (9). Table 6 lists the comparison with [19, Table 9].

As it can be seen in Table 6 the probability is by a factor of approx. 2.7 higher than estimated in [19]. Note that we did not count the disturbances in step $i = 21$ and step $i = 77$ since some of the conditions are fulfilled due to message modification or truncation. This means that the path of the disturbance is fixed and we cannot exploit any carry effects.

In order to determine the overall probability, we assume that the probabilities of local collisions are independent. To confirm this assumption, we have performed several computer measurements for a few overlapping local collisions. The measurement results match the computed probabilities.

3.6 Importance of Carry Effects

In the case of SHA-1, the improvement of the attack complexity is rather small. This is due to the fact that the disturbance vector is very sparse and the disturbances are introduced in bit positions where we cannot exploit any carry effects due to uncorrectable carries, *e.g.* bit position $j = 1$.

Consider for instance the hash function SHA1-IME [8]. Jutla and Patthak claim to improve the collision resistance of SHA-1 by modifying the existing message expansion with the goal to increase the minimum Hamming weight. By using a computer aided proof they show that the minimum weight in the last 60 steps of the message expansion of SHA1-IME is at least 75. It is clear that the overall complexity increases with a higher weight in the disturbance vector. However, due to the higher weight also the impact of carry effects as shown in this section increases. Therefore, our way of looking at probabilities instead of conditions gives a more accurate complexity estimation.

4 Conclusion and Further Work

In this article we analyzed local collisions and corresponding probabilities in detail. We showed that it is more accurate to consider probabilities instead of conditions for the estimation of the overall attack complexity for collision attacks on SHA-1. This is due to the fact that carry effects increase the probability. Based on the accurate probability computation we updated the complexity of the collision attack on SHA-1 presented by Wang *et al.* Currently we are investigating the impact of our approach on SHA1-IME and local collisions in SHA-256.

Acknowledgements

We would like to thank Christophe De Cannière for fruitful discussions and comments on this article.

References

1. Eli Biham and Rafi Chen. Near-Collisions of SHA-0. In Matthew K. Franklin, editor, *CRYPTO 2004, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, volume 3152 of *LNCS*, pages 290–305. Springer, 2004.
2. Eli Biham, Rafi Chen, Antoine Joux, Patrick Carribault, Christophe Lemuet, and William Jalby. Collisions of SHA-0 and Reduced SHA-1. In Ronald Cramer, editor, *EUROCRYPT 2005, Aarhus, Denmark, May 22-26, 2005. Proceedings*, volume 3494 of *LNCS*, pages 36–57. Springer, 2005.
3. Florent Chabaud and Antoine Joux. Differential Collisions in SHA-0. In Hugo Krawczyk, editor, *CRYPTO '98, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*, volume 1462, pages 56–71. Springer, 1998.
4. Magnus Daum. *Cryptanalysis of Hash Functions of the MD4-Family*. PhD thesis, Ruhr Universität Bochum, 2005. Available at `http://http://www.cits.rub.de/imperia/md/content/magnus/dissmd4.pdf`.
5. Hans Dobbertin. Cryptanalysis of MD4. In Bart Preneel, editor, *Fast Software Encryption, Cambridge, UK, February 21-23, 1996, Proceedings*, volume 1039 of *LNCS*, pages 53–69. Springer, 1996.
6. Hans Dobbertin. Cryptanalysis Of MD4. *Journal of Cryptology*, 11(4):253–271, 1998.
7. Charanjit S. Jutla and Anindya C. Patthak. A Matching Lower Bound on the Minimum Weight of SHA-1 Expansion Code. Cryptology ePrint Archive, Report 2005/266, 2005. `http://eprint.iacr.org/`.
8. Charanjit S. Jutla and Anindya C. Patthak. A Simple and Provably Good Code for SHA Message Expansion. Cryptology ePrint Archive, Report 2005/247, 2005. `http://eprint.iacr.org/`.
9. Vlastimil Klima. Finding MD5 Collisions on a Notebook PC Using Multi-message Modifications, 2005. Preprint, available at `http://eprint.iacr.org/2005/102`.
10. Arjen Lenstra, Xiaoyun Wang, and Benne de Weger. Colliding X.509 Certificates, 2005. Preprint, available online at `http://eprint.iacr.org/2005/067`.
11. Krystian Matusiewicz and Josef Pieprzyk. Finding good differential patterns for attacks on SHA-1. Cryptology ePrint Archive, Report 2004/364, 2004. `http://eprint.iacr.org/`.

12. National Institute of Standards and Technology (NIST). FIPS-180-2: Secure Hash Standard, August 2002. Available online at `http://www.itl.nist.gov/fipspubs/`.
13. Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Exploiting Coding Theory for Collision Attacks on SHA-1. In Nigel P. Smart, editor, *Cryptography and Coding, Cirencester, UK, December 19-21, 2005, Proceedings*, volume 3796 of *LNCS*, pages 78–95. Springer, 2005.
14. Bart Preneel. *Analysis and Design of Cryptographic Hash Functions*. PhD thesis, Katholieke Universiteit Leuven, 1993.
15. Vincent Rijmen and Elisabeth Oswald. Update on SHA-1. In Alfred Menezes, editor, *CT-RSA 2005, San Francisco, CA, USA, February 14-18, 2005, Proceedings*, volume 3376 of *LNCS*, pages 58–71. Springer, 2005.
16. Xiaoyun Wang, Dengguo Feng, Xuejia Lai, and Xiuyuan Yu. Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD, August 2004. Preprint, available at `http://eprint.iacr.org/2004/199`.
17. Xiaoyun Wang, Xuejia Lai, Dengguo Feng, Hui Chen, and Xiuyuan Yu. Cryptanalysis of the Hash Functions MD4 and RIPEMD. In Ronald Cramer, editor, *EUROCRYPT 2005, Aarhus, Denmark, May 22-26, 2005. Proceedings*, volume 3494 of *LNCS*, pages 1–18. Springer, 2005.
18. Xiaoyun Wang, Andrew Yao, and Frances Yao. New Collision Search for SHA-1, August 2005. Presented at rump session of CRYPTO 2005.
19. Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding Collisions in the Full SHA-1. In Victor Shoup, editor, *CRYPTO 2005, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, volume 3621 of *LNCS*, pages 17–36. Springer, 2005.
20. Xiaoyun Wang and Hongbo Yu. How to Break MD5 and Other Hash Functions. In Ronald Cramer, editor, *EUROCRYPT 2005, Aarhus, Denmark, May 22-26, 2005. Proceedings*, volume 3494 of *LNCS*, pages 19–35. Springer, 2005.
21. Xiaoyun Wang, Hongbo Yu, and Yiqun Lisa Yin. Efficient Collision Search Attacks on SHA-0. In Victor Shoup, editor, *CRYPTO 2005, Santa Barbara, California, USA, August 14-18, 2005, Proceedings*, volume 3621 of *LNCS*, pages 1–16. Springer, 2005.

A Probability Bounds for Single-Bit Disturbances

Based on formulae (8) and (9) in Section 3.3, the probability of f_{XOR} and f_{MAJ} can be bounded as follows. We know that

$$\sum_{k=1}^{27-j} 2^{-4k} = 2^{-4}\frac{1-2^{-4(28-j)}}{1-2^{-4}} \leq \frac{2^{-4}}{1-2^{-4}} = \frac{1}{2^4-1}\ ,$$

$$\sum_{k=1}^{32-j} 2^{-4k} = 2^{-4}\frac{1-2^{-4(33-j)}}{1-2^{-4}} \leq \frac{2^{-4}}{1-2^{-4}} = \frac{1}{2^4-1}\ ,\ \text{and}$$

$$2\cdot 2^{-4(32-j)} + \sum_{k=1}^{31-j} 2^{-4k} =$$

$$2^{-4(32-j)+1} + 2^{-4}\frac{1-2^{-4(32-j)}}{1-2^{-4}} \leq 2^{-3} + \frac{2^{-4}}{1-2^{-4}} = \frac{1}{2^3} + \frac{1}{2^4-1}\ .$$

Therefore, we get the following bounds on the probability for f_{XOR} and f_{MAJ}:

$$\frac{1}{2^4} \le \quad p(f_{XOR}, j) \quad \le \frac{1}{2^4 - 1} \text{ for } j = 2, \ldots, 26 \ , \tag{12}$$

$$\frac{1}{2^4} \le \quad p(f_{XOR}, j) \quad \le \frac{1}{2^3} + \frac{1}{2^4 - 1} \text{ for } j = 27, \ldots, 31 \ , \tag{13}$$

$$\frac{1}{2^4} \le \quad p(f_{MAJ}, j) \quad \le \frac{1}{2^4 - 1} \text{ for } j = 2, \ldots, 26 \text{ and } j = 27, \ldots, 30 \ , \tag{14}$$

where the lower bound for the probability 2^{-4} is derived by counting conditions. For instance, if we compute the probability for a disturbance in bit position $j = 3$ we get for both f_{XOR} and f_{MAJ} a probability of $2^{-3.9068}$ instead of 2^{-4}.

B Probabilities for Disturbances in Consecutive Bit Position

Here we give the probabilities for disturbances in consecutive bit positions for f_{XOR}. This is the generalization of the case presented in Section 3.4. Again, we have to consider uncorrectable carries. Uncorrectable carries occur if the disturbances are in bit position $j = 2, 1$ and $j = 27, 26$. In these cases, we get the probability of both disturbances without carry. If $j = 2, 1$, we obtain a probability of $2^{-4}2^{-2} = 2^{-6}$ and $j = 27, 26$ results in $2^{-4}2^{-4} = 2^{-8}$. Let us now consider disturbances in consecutive bit positions from $j = 2, \ldots 25$, *i.e.* the tuples $j = (3, 2), (4, 3), \ldots, (26, 25)$, and from $j = 27, \ldots, 30$, *i.e.* the tuples $j = (28, 27), (29, 28), (30, 29), (31, 30)$. The formulae for all cases are given in (15), where j refers to the right entry of the tuple.

$$p(f_{XOR}, (j+1, j)) = \begin{cases} 2^{-4} + 2^{-6} & \text{for } j = 0 \\ 2^{-4} + 2^{-8} & \text{for } j = 1 \text{ and } j = 26 \\ \sum_{k=1}^{27-j} 2^{-4k} & \text{for } j = 2, \ldots, 25 \\ 2 \cdot 2^{-4(32-j)} + \sum_{k=1}^{31-j} 2^{-4k} & \text{for } j = 27, \ldots, 30 \end{cases} \tag{15}$$