Practical Collision and Preimage Attack on DCH-n

Mario Lamberger and Florian Mendel

Institute for Applied Information Processing and Communications (IAIK),
Graz University of Technology, Inffeldgasse 16a, A-8010 Graz, Austria

Abstract. In this paper, we show practical collision and preimage attacks on DCH-n. The attacks are based on the observation of Khovratovich and Nikolic that the chaining value is not used in the underlying block cipher. Based on this observation, we show a trivial collision resp. multi-collision attack on DCH-n and a preimage attack with a complexity of about 521 compression function evaluations.

1 Description of DCH-n

The hash function DCH-n is an iterated hash function based on the Merkle-Damgaard design principle. It processes message blocks of 512 bits (504 bits message input, 8 bits dithering input) and produces a hash value of $n = 224, 256, 384$ or 512 bits. In each iteration the compression function f is used to update the chaining value of 512 bits as follows:

$$H_{i+1} = f(H_i, M_i) = H_i \oplus M_i \oplus g(M_i) \ ,$$

where $g(M)$ is some non-linear transformation. For a detailed description of DCH-n we refer to [3].

2 Cryptanalysis

In this section, we will present our collision and preimage attack on DCH-n. The attack is an extension of the attack of Khovratovich and Nikolic [1] and is based on similar principles as the attacks on SMASH [2]. Let $\gamma_i(M_i) = g(M_i) \oplus M_i$. Then the above equation can be rewritten as:

$$H_{i+1} = H_0 \oplus \gamma_0(M_0) \oplus \gamma_1(M_1) \oplus \cdots \oplus \gamma_i(M_i) \tag{1}$$

Note that the γ_i are different since in DCH-n an 8-bit dithering is used in each message block to compute $M_i \oplus g(M_i)$. The only thing we need to know about dithering method is that the last 5 bits are a counter and that the sequence guiding the first 3 bits changes every time the counter resets.

2.1 Collision Attack

We now describe the collision attack. Assume we are given a message consisting of $2^8 + 1$ message blocks, $m = M_0 \| M_1 \| \ldots \| M_{2^8}$. Each $M_i = m_i \| M_i'$, where m_i is the dithering of the $i - th$ message block. Since only 2^8 different dithering blocks exist, there exist $0 \leq i, j \leq 2^8$ with $i \neq j$ such that $m_i = m_j$. But this also implies that $\gamma_i = \gamma_j$. Based on (1) we have with $k = 2^8$

$$H_{k+1} = H_0 \oplus \gamma_0(M_0) \oplus \gamma_1(M_1) \oplus \cdots \oplus \gamma_k(M_k).$$

So setting $M_i' = M_j' = a$ for the above $i \neq j$ implies that these blocks don't contribute to the value H_{k+1}. Hence, we can trivially construct collisions for DCH-n. Note that the messages in the colliding message pair consist of $2^8 + 1$ message blocks.

Every choice of $a \in \{0,1\}^{504}$ leads to a collision. Hence, we can trivially construct t-collisions (for $0 < t < 2^{504}$) for DCH-n. Note that all these attacks apply to DCH-n for all output sizes.

2.2 Preimage Attack

In a similar way as in the collision attack, we can also construct preimages for DCH-n. The attack is based on the observation that the outputs of DCH-n form a vector space of dimension n over $GF(2)$ (cf. also [2]). Hence, we only need to compute a basis of the output vector space to construct preimages for DCH-n. The only technicality we have to take care of is the dithering of the message blocks.

We assume $n = 512$ since the other output lengths of DCH-n result from truncations of the $n = 512$ version.

As in the collision case we start by finding different indices (i, j) for which the dithering m_i and m_j is the same. For the preimage case, we need to find $n = 512$ such pairs. We will construct a preimage of length $N+1$ message blocks.

Then, the attack can be summarized as follows:

1. Assume we want to construct a preimage for h consisting of $N + 1$ message blocks. Then, we have to find a message M such that:

$$h = H_0 \oplus \bigoplus_{i=0}^{N} \gamma_i(M_i) \ .$$

2. Choose the last message block M_N such that the padding is correct.
3. Once, we have fixed the last message block, we have to find the remaining message blocks M_i for $0 \leq i < N$ such that:

$$\bigoplus_{i=0}^{N-1} \gamma_i(M_i) = h \oplus H_0 \oplus \gamma_N(M_N) \ . \tag{2}$$

4. N will be chosen such that among the remaining N message blocks we have ℓ index pairs $(i_0, j_0), \ldots, (i_{\ell-1}, j_{\ell-1})$ satisfying $\gamma_{i_k} = \gamma_{j_k}$ (where every i_k, j_k is unique).
5. Compute ℓ vectors $a^k = \gamma_{i_k}(M_0^k) \oplus \gamma_{j_k}(M_1^k)$ with arbitrary values for M_0^k and M_1^k and save the triple (a^k, M_0^k, M_1^k) in a list L.
6. From the set of $\ell \geq n$ vectors a^k compute a basis of the output vector space of DCH-n. The probability for $\ell \geq n$ vectors to contain n vectors which are linearly independent is

$$\prod_{i=0}^{n-1} \frac{2^\ell - 2^i}{2^\ell} = \prod_{i=0}^{n-1} (1 - 2^{i-\ell}) \geq 2^{-\frac{2^n-1}{2^{\ell-1}}}.$$

   This means that we can basically construct such a basis with a complexity of $2 \cdot \ell$ compression function evaluations. This can be reduced to $\ell+1$ evaluations of the compression function by fixing the block M_0^k and letting only the block M_1^k vary when generating the basis of the output vector space.
   E. g. choosing $n = 512$ and $\ell = 520$ we already get a probability of 0.9961 for finding a basis and thus need 521 compression function evaluations. Note, that constructing the basis is a one time effort.
   Let $B = \{a^{k_0}, \ldots, a^{k_{n-1}}\}$ denote the basis for the output vector space and let $\mathcal{I} = \bigcup_{k=0}^{n-1} i_k \cup j_k$ be the union of all the indices contributing to the basis vectors. (For simplicity we assume that the first n pairs correspond to the basis vectors.)
7. We divide the indices $\mathcal{N} = \{0, \ldots, N-1\}$ into $\mathcal{I}$ and $\mathcal{N} \setminus \mathcal{I}$. For every index i in $\mathcal{N} \setminus \mathcal{I}$ we set $M_i' = 0 \ldots 0$. These are the indices not contributing to the basis. From (2) we thus get

$$\bigoplus_{\mathcal{I}} \gamma_i(M_i) = h \oplus H_0 \oplus \gamma_N(M_N) \bigoplus_{\mathcal{N}\setminus\mathcal{I}} \gamma_i(m_i \| 0 \ldots 0).$$

   Once a basis and the indices $\mathcal{I}$ are computed, the right side of the equation is completely known and thus we have

$$\bigoplus_{\mathcal{I}} \gamma_i(M_i) = c$$

8. An arbitrary c can be represented with respect to this basis $c = x_0 a^{k_0} + \cdots + x_{n-1} a^{k_{n-1}}$ by solving the linear system over $GF(2)$. Now we choose the blocks M_i for $i \in \mathcal{I}$ as follows:
   – If $x_k = 0$ for $0 \leq k < n$ set $M_{i_k} = \alpha$ and $M_{j_k} = \alpha$ for some arbitrary value of α (as in the collision attack). Since in such a case, γ_{i_k} and γ_{j_k} are equal, these two values cancel out and don't contribute to the result.
   – If $x_k = 1$ for $0 \leq k < n$ set $M_{i_k} = M_0^k$ and $M_{j_k} = M_1^k$ such that $\gamma_{i_k}(M_0^k) \oplus \gamma_{j_k}(M_1^k) = a^k$ for $0 \leq k < n$.
9. What remains is to say how large N has to be. We need to guarantee that among all indices from $0, \ldots, N-1$ we can find ℓ pairs as described above. If

we take a look at the 8-bit dithering strings m_i for $i = 0, \ldots, N-1$ we know, that the 3 non-counter bits can only have 8 different values $0, 1, \ldots 7$ (actually 6 for the concrete Hanoi sequence). Let $n_0, \ldots, n_7$ denote the frequencies of the value $0, \ldots, 7$ in the non-counter part. Assume $N = 32 \cdot \sum_{i=0}^{7} n_i$. Then, the number of valid pairs (i_k, j_k) is

$$32 \cdot \sum_{i=0}^{7} \left\lfloor \frac{n_i}{2} \right\rfloor = 32 \cdot \sum_{i=0}^{7} \left(\frac{n_i}{2} - \left\{ \frac{n_i}{2} \right\} \right) \geq \frac{N}{2} - 2^7.$$

Therefore, $N = 2 \cdot \ell + 2^8$ is a valid choice of N. For $\ell = 520$ as above we therefore get a preimage of length 1297 blocks.

Hence we can construct a preimage by solving a linear system of equations of dimension $n \times n$ over $GF(2)$. Constructing the basis has a complexity of $\ell + 1$ compression function evaluations and is a one time effort.

Furthermore, the preimage attack can be used to construct second preimages for DCH-n with the same complexity. Note that by using the above described method, preimages (or second preimages) always consist of $N + 1 = 2\ell + 2^8 + 1$ message blocks.

3 Conclusion

We showed, that it is trivial to construct collisions and (second) preimages for DCH-n. Furthermore, the presented attack applies to all similar constructions not introducing the chaining variable into the compression function.

Acknowledgements

The authors wish to thank David Wilson for useful comments and discussions.

References

1. Dmitry Khovratovich and Ivica Nikolic. Cryptanalysis of DCH-n, 2008. Available online: `http://lj.streamclub.ru/papers/hash/dch.pdf`.
2. Mario Lamberger, Norbert Pramstaller, Christian Rechberger, and Vincent Rijmen. Analysis of the Hash Function Design Strategy Called SMASH. *IEEE Transactions on Information Theory*, 54(8):3647–3655, 2008.
3. David A. Wilson. The DCH Hash Function. Submission to NIST, 2008. Available online: `http://web.mit.edu/dwilson/www/hash/dch/Supporting_Documentation/dch.pdf`.