The Neighborhood of a MD4 Collision

Alexander Scheel
alexander.m.scheel@gmail.com
Iowa State University

Eric Rozier
erozier@iastate.edu
Iowa State University

Abstract

In 2005, Wang et al. published the details of a collision attack against the MD4 hash function. We extend their work by showing that not only is the attack state independent and differential independent, but also that the neighborhood of the attack is populated with other classes of collisions. We describe an efficient encoding of this problem into SAT, and present example classes in the neighborhood of Wang's.

1 Introduction

The MD4 algorithm was developed by Ron Rivest around 1990 and is standardized in RFC 1320 [4]. While largely known to be insecure, and thus having fallen out of use, its simple construction leads to fast implementations and easy encoding as a SAT problem. An attack against a chosen-state MD4 was published by Dobbertin in 1998 [1]. In 2004, Wang et al. demonstrated a collision attack against the full MD4 [6], before publishing the details in 2005 [7]. With reduced time complexity compared to Dobbertin, this class of collision went on to inspire other attacks such as Sasaki et al., which pushed the complexity of collisions to less than two MD4 evaluations [5].

Outside of work published in 2006 by Mironov and Zhang [3], the authors have found little evidence of the application of SAT solvers to hash function collisions and MD4 specifically. The Mironov and Zhang work was focused on the feasibility of such an encoding and the time for finding new examples of collisions in the Wang et al. class.

We present the following four improvements in collisions.

1. Wang et al.'s attack is differential independent.
2. Wang et al.'s attack is state independent.
3. Wang et al.'s attack is highly malleable.
4. The neighborhood of Wang et al.'s attack contains other collision classes.

2 Terminology

We first begin by giving a brief set of terminology:

A **differential** is the signed XOR difference between two input blocks to a hash function.

The **differential path** is the set of state transition constraints which produce a full or partial collision.

The **constraints** of a collision are the necessary and/or sufficient boolean formulas for generating a collision, typically as a function of the state blocks.

A **state transition constraint** is a constraint on the intermediate state variables, i.e., round variables. A constraint is **weak** if it is a any difference, and a constraint is **strong** if it is a signed difference. A equality constraint can either be weak or strong. This is roughly equivalent to a differential path.

Input block constraints, **input state constraints**, and **output state constraints** are all analogously defined as state transition constraints.

A **property** of a collision is any constraint placed upon the input block which exists independent from the constraints of a collision.

An attack is **constraint-based** if there exists a SAT encoding of the attack which can generate multiple collisions.

A collision is **differential-independent** if there are multiple input block differentials which satisfy the state transition constraints.

A collision is **malleable** if collisions can be found with different properties such as ASCII or JSON input blocks, or with fixed internal text.

A collision is **state-independent** if there exist alternative starting states which produce collisions. A state-independent collision is termed **strong** if one block can collide under two or more starting states.

A **collision class** is defined by the differential path and optionally differential. The **neighborhood of a collision class** is a set of alternative differential paths with low distance between differences. For example, consider the differential path with $\Delta(I_{0...48}) = 0$; then the differential path with $\Delta(I_0) = 1$ has low distance to $\Delta(I_0) = 0$, and hence is said to be in the neighborhood of $\Delta(I_{0...48})$.

3 Notation

As we are not concerned with the theoretical underpinnings of the SAT models, nor will we use the model to provide general proofs, we use the following notations. This notation is heavily influenced by our use of the $bc2cnf$ utility by Junttila [2].

A constraint, $c = (n, e)$ is a mapping between a name, n, and an equation, e, of boolean clauses. These equations can be simple $(a \wedge b)$ or complex $(a \oplus (c \wedge d) \oplus (a \vee b))$. A model, $S = (C, A)$, is an ordered pair of a set, C, of constraints and a set, A, a set of names which must evaluate to true. Note that input variables are implicit and thus left unspecified.

We construct the base model for MD4 as follows. Let $s_0..s_{127}$ be the input state variables. Let $b_0..b_{511}$ be the input block. We define $i_0..i_{1535}$ as the bits of the intermediate state variables generated via MD4, per RFC 1320 [4], and $o_0..o_{128}$ as the resulting output state. With the s's and b's left unspecified, the model for MD4 is thus $C_{MD4} = \{(i_0, e(i_0)), ..., (o_{128}, e(o_{128}))\}$. As we are interested in collisions, we use C_{MD4,h_1} to denote the construction of MD4 with variables i, o, s, and b prefixed with h_1.

We then define I_{md4} to be the set of initial state values, with I_{md4,h_1} to be those state values with names prefixed by h_1. We also define $C_{collision,h_1,h_2}$ to be the two-way collision between h_1 and h_2; that is, all output bits are equal.

4 Differential Independent

We note that Wang et al.'s attack on MD4 is differential independent. Let C_{wangs,h_1,h_2} be the set of the state transition constraints on page 15 and sufficient constraints on page 16 [7]. Let $\bar{\Delta}_{wangs}$ be the negation of the specified differential on page 7 [7]. Then we claim that

$$S = (\{C_{MD4,h_1} \cup C_{MD4,h_2} \cup C_{wangs,h_1,h_2} \cup \bar{\Delta}_{wangs} \cup C_{collision,h_1,h_2} \cup I_{md4,h_1} \cup I_{md4,h_2}\}, \{C_{wangs}, \bar{\Delta}_{wangs}, C_{collisionh_1,h_2}\})$$

is a model which encodes a collision between h_1 and h_2 satisfying the conditions of Wang's attack but with a different differential.

Claim: the model S above is satisfiable. We give the following examples as proof.

By negating each found differential and re-running we can find all possible differentials. This is a total of 238 differentials. We note the signed block differences below.

5 State Independent

We note that Wang et al.'s is strongly state independent. We construct a model S with four instances of MD4, h_1, h_2, h_3, h_4, which share an input block and have a nonzero input state delta.

We thus claim that

$$S = (\{C_{MD4,h_1}, C_{MD4,h_2}, C_{MD4,h_3}, C_{MD4,h_4}, C_{wangs,h_1,h_2}, C_{wangs,h_3,h_4}, \Delta_{wangs,h_1},$$

encodes the above properties and is satisifable. We reproduce several examples below.

6 Highly Malleable

In the above style above, we note that Wang et al.'s is malleable, and highly malleable under alternative differentials. We produce the following examples; though we note that it is not possible with SAT prove that the attack is fully malleable and can produce collisions with any property.

7 Neighborhood

References

[1] DOBBERTIN, H. Cryptanalysis of md4. *Journal of Cryptology 11*, 4 (Sep 1998), 253–271. https://link.springer.com/content/pdf/10.1007/s001459900047.pdf.

[2] JUNTTILA, T. Tools for constrained boolean circuits. https://users.ics.aalto.fi/tjunttil/circuits/.

[3] MIRONOV, I., AND ZHANG, L. *Applications of SAT Solvers to Cryptanalysis of Hash Functions*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 102–115. https://eprint.iacr.org/2006/254.pdf.

[4] RIVEST, R., AND RSA DATA SECURITY, I. The MD4 Message-Digest Algorithm. RFC 1320, IETF, April 1992. https://tools.ietf.org/html/rfc1320.

[5] SASAKI, Y., WANG, L., OHTA, K., AND KUNIHIRO, N. *New Message Difference for MD4*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 329–348. https://link.springer.com/content/pdf/10.1007/978-3-540-74619-5_21.pdf.

[6] WANG, X., FENG, D., LAI, X., AND YU, H. Collisions for hash functions md4, md5, haval-128 and ripemd. Cryptology ePrint Archive, Report 2004/199, 2004. http://eprint.iacr.org/2004/199.

[7] WANG, X., LAI, X., FENG, D., CHEN, H., AND YU, X. *Cryptanalysis of the Hash Functions MD4 and RIPEMD*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 1–18. http://www.infosec.sdu.edu.cn/uploadfile/papers/Cryptanalysis%20of%20the%20Hash%20Functions%20MD4%20and%20RIPEMD.pdf.