The Second-Preimage Attack on MD4

Hongbo Yu, Gaoli Wang, Guoyan Zhang, and Xiaoyun Wang*

School of Mathematics and System Sciences,
Shandong University, Jinan 250100, China
yhb@mail.sdu.edu.cn, xywang@sdu.edu.cn

Abstract. In Eurocrypt'05, Wang et al. presented new techniques to find collisions of Hash function MD4. The techniques are not only efficient to search for collisions, but also applicable to explore the second-preimage of MD4. About the second-preimage attack, they showed that a random message was a weak message with probability 2^{-122} and it only needed a one-time MD4 computation to find the second-preimage corresponding to the weak message. A weak message means that there exits a more efficient attack than the brute force attack to find its second-preimage. In this paper, we find another new collision differential path which can be used to find the second-preimage for more weak messages. For any random message, it is a weak message with probability 2^{-56}, and it can be converted into a weak message by message modification techniques with about 2^{27} MD4 computations. Furthermore, the original message is close to the resulting message (weak message), i.e, the Hamming weight of the difference for two messages is about 44.

Keywords: Hash function, collision differential path, second-preimage, weak message.

1 Introduction

In 1990[1], Rivest introduced the hash function MD4 which is the first dedicated hash function. After MD4, many hash functions such as MD5[2], HAVAL[3], RIPEMD [4], SHA-0[5], SHA-1[6], SHA-256[7] were designed subsequently.

For a hash function h with inputs x, x' and outputs y, y', three potential security properties should be satisfied:

1. **Preimage resistance:** for any pre-specified output y, it is computationally infeasible to find an input x such that $h(x) = y$.
2. **Second-preimage resistance:** for any input x, it is computationally infeasible to find another input x' such that $h(x) = h(x')$
3. **Collision resistance:** it is computationally infeasible to find any two distinct inputs x, x' with the same output, i.e., $h(x) = h(x')$.

* Supported by the National Natural Science Foundation of China(NSFC Grant No.90304009) and 973 Project(No.2004CB318000).

Y.G. Desmedt et al. (Eds.): CANS 2005, LNCS 3810, pp. 1–12, 2005.

The original design purpose of MD4 is that there is no better collision attack than the birthday attack which should take about 2^{64} MD4 computations to find a collision, and no better attack than brute force attack which should take 2^{128} MD4 computations to find a preimage corresponding to a pre-specified hash-value or the second preimage corresponding to a given message. The existing attack reveals that MD4 fails to reach the designer's goals both on collision resistance and second-preimage resistance. In 1996, Dobbertin presented a successful attack on MD4 which find a collision with probability 2^{-22}[8]. In 1998, H.Doberrtin[9] showed that the first two (out of the total three) rounds of MD4 are not one-way. This means it is possible to find the preimage and the second-preimage for the first two rounds of MD4. Wang et al. [11] described a new kind of collision attack on the hash function MD4 and RIPEMD which is also applied to break MD5[10], HAVAL-128[14], SHA-0[12] and SHA-1[13]. Simultaneously, the collision attack on MD4 [11] can be used to explore the second-preimage attack on MD4, and the main results are as follows:

1. A random message is a weak message with probability 2^{-122}. For a weak message, it only needs a one-time MD4 computation to find a second-preimage of the resulting hash value.
2. Any message M can be modified with the basic message modification techniques. The resulting message M' is a weak message with probability 2^{-23}. M and M' are close and the Hamming weight of the difference for two messages is 50 on average.
3. Under the advanced message modification, any message M can be modified into M' which is a weak message with probability 2^{-2} to 2^{-6}. However, the Hamming weight of their difference increases quickly up to 110.

In this paper, we give a further research on the second-preimage attack on MD4. Our results are as follows:

1. We find a new differential path which is efficient to find more weak messages. Utilizing this path, any message M is a weak message with probability 2^{-56}. For a weak message, it only needs a one-time MD4 computation to find a second-preimage.
2. For any message, we apply message modification techniques to convert it into a weak message with 2^{27} MD4 computations, the Hamming weight for their difference is about 44.

The paper is organized as follows: In section 2, we describe MD4 details. In section 3, we give some basic properties of nonlinear round functions for MD4 and some notations. Our main results are introduced in section 4. We summarize the paper in section 5.

2 Description of MD4

The message digest algorithm MD4 takes a message of length less than 2^{64} bits and produces a 128-bit hash value. The input message is padded and then processed in 512-bit blocks by Damgard/Merkle iterative structure. Each iteration