Cryptanalysis of MD5 Compress

Hans Dobbertin

German Information Security Agency
e-mail: dobbertin@skom.rhein.de

May 2, 1996

In 1991 the hash function MD5 was introduced by Ron Rivest as a strengthened version of MD4. Beside some other modifications, the number of rounds is extended from three to four.

In this short note we report about an attack on the compress function of MD5, which is based on similar methods as previous attacks on RIPEMD, MD4 and the 256-bit extension of MD4 (see [4], [5]). Below we give a **collision**[1] **of the compress function of MD5**.

Recall that in 1993 Bert den Boer and Antoon Bosselaers [3] showed how **pseudo-collisions** (in our terminology) of MD5 compress can be found. Matt Robshaw ([8], Remark 1 in Section 4.1) commented on this attack as follows:

> "As it stands, the pseudo-collision arises from initializing the four-word buffer at the start of MD5 to two different values. These values differ only in the MSB of each of the four words. The same message is used for both sets of buffer values and the same message digest is obtained.
> A far more serious flaw would be if it were possible to choose one initial starting value for the buffer, not necessarily the one given in the algorithm, and then choose two different messages, perhaps differing in only a few bits of one word, so that the same message digest is obtained."

The latter describes precisely what is done now. We think that this might be reason enough to substitute MD5 in future applications.

Alternatives for MD5 are SHA-1 [1] and on the other hand RIPEMD-160 [6], which has been designed as a strengthened version of RIPEMD [2] by Antoon Bosselaers, Bart Preneel and the author taking account of the recent analysis of MD4-like hash functions.

---

[1] Using the term "collision of a compress function" we assume that the initial value is the same for both inputs, i.e. an initial value IV and two different inputs X and $\tilde{X}$ are given such that

$$\mathsf{compress}(IV; X) = \mathsf{compress}(IV; \tilde{X}).$$

On the other hand we use the term "pseudo-collision" if two different initial values $IV, \tilde{IV}$ and (possibly identical) inputs $X, \tilde{X}$ are given such that

$$\mathsf{compress}(IV; X) = \mathsf{compress}(\tilde{IV}; \tilde{X}).$$

Pseudo-collisions are of much less practical importance than collisions.

Collision for the compress function of MD5. Use the initial value

$$IV = \texttt{0x12AC2375 0x3B341042 0x5F62B97C 0x4BA763ED},$$

and define the first input $X = (X_i)_{i<16}$ by setting:

$X_0 =$ `0xAA1DDA5E`	$X_4 =$ `0x1006363E`	$X_8 =$ `0x98A1FB19`	$X_{12} =$ `0x1326ED65`
$X_1 =$ `0xD97ABFF5`	$X_5 =$ `0x7218209D`	$X_9 =$ `0x1FAE44B0`	$X_{13} =$ `0xD93E0972`
$X_2 =$ `0x55F0E1C1`	$X_6 =$ `0xE01C135D`	$X_{10} =$ `0x236BB992`	$X_{14} =$ `0xD458C868`
$X_3 =$ `0x32774244`	$X_7 =$ `0x9DA64D0E`	$X_{11} =$ `0x6B7A669B`	$X_{15} =$ `0x6B72746A`

The second input $\tilde{X} = (\tilde{X}_i)_{i<16}$ is defined by setting $\tilde{X}_i = X_i$ $(i < 16, i \neq 14)$ and $\tilde{X}_{14} = X_{14} + 2^9$. Then we have a collision, i.e.

$$\textsf{MD5-compress}(IV; X) = \textsf{MD5-compress}(IV; \tilde{X}),$$

and this common compress value is

$$\texttt{0xBF90E670 0x752AF92B 0x9CE4E3E1 0xB12CF8DE}.$$

The computation of such a collision takes about 10 hours on a Pentium PC.

References

1. FIPS 180-1, *Secure hash standard,* NIST, US Department of Commerce, Washington D.C., April 1995.
2. RIPE Consortium, *Ripe Integrity Primitives – Final report of RACE Integrity Primitives Evaluation (R1040)*, LNCS 1007, Springer-Verlag, 1995.
3. B. den Boer, A. Bosselaers, *Collisions for the compression function of MD5*, Advances in Cryptology, Proc. Eurocrypt'93, LNCS 765, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 293–304.
4. H. Dobbertin, *RIPEMD with two-round compress function is not collision-free*, Journal of Cryptology, to appear.
5. H. Dobbertin, *Cryptanalysis of MD4*, Fast Software Encryption, LNCS 1039, D. Gollmann, Ed., Springer-Verlag, 1996, pp. 53–69.
6. H. Dobbertin, A. Bosselaers, and B. Preneel: *RIPEMD-160: A strengthened version of RIPEMD*, Fast Software Encryption, Cambridge Workshop, LNCS 1039, D. Gollmann, Ed., Springer-Verlag, 1996, pp. 71-82.[2]
7. R. Rivest, *The MD5 message digest algorithm*, RFC 1321, April 1992.
8. M. Robshaw, *On pseudo-collisions in MD5*, Technical Report TR-102, version 1.1, RSA Laboratories, July 1994.

[2] This first publication of [6] still contains some bugs. The corrected version as well as a C reference implementation of RIPEMD-160 can be obtained under `ftp.esat.kuleuven.ac.be` in the directory `/pub/COSIC/bosselae/ripemd/`.