JOURNAL OF COMPUTING TECHNOLOGIES

ISSN 2278 – 3814

Available online at www.jctjournals.com

Volume 1 Issue 2, June 2012

Advanced MD5 Encryption with Extra Compress Function

Mr. Deepak Singh Rajput[1], Mr. Shitanshu Jain[2]
[1&2]Computer Science & Engineering,
[1&2]RGPV Bhopal (M.P.), India
[1&2]Gyan Ganga College of Technolgy,Jabalpaur(M.P.),India
[1]*deepakrajput16@gmail.com*
[2]*Shitanshu_jn@yahoo.com*

Abstract— **recent breakthroughs in cryptanalysis of standard hash functions like SHA-1 and MD5 raise the need for alternatives. In the past few years, there have been significant research advances in the analysis of hash functions and it was shown that none of the hash algorithm is secure enough for critical purposes whether it is MD5 or SHA-1. Nowadays scientists have found weaknesses in a number of hash functions, including MD5, SHA and RIPEMD so the purpose of this paper is combination of some function to reinforce these functions and also increasing hash code length up to 160 that makes stronger algorithm against collision attests.**

Keywords—***MD5 Algorithm; Compress Function, SHA Algorithm, Brute force attack.***

1. INTRODUCTION

MD5 was the last in a succession of cryptographic hash functions designed by Ron Rivest[1] in the early 1990s. It is a widely-used well-known 128-bit iterated hash function, used in various applications including SSL/TLS, IPSec, and many other cryptographic protocols. It is also commonly-used in implementations of time stamping mechanisms, commitment schemes, and integrity-checking applications for online software, distributed systems, and random-number generation. It is even used by the Nevada State. Gaming Authority to ensure slot-machine ROMs have not been tampered with.

Hash functions are one-way functions with as input a string of arbitrary length (the message) and as output a fixed length strings (the hash value). The hash value is a kind of signature for that message. One-way functions work in one direction, meaning that it is easy to compute the hash value from a given message and hard to compute a message that hashes to a given hash value.A hash collision is a pair of different messages m1 6= m2 having the same hash value H (m1) = H (m2). Therefore second pre-image resistance and collision resistance are also known as weak and strong collision resistance, respectively. Since the domain of a hash function is much larger (can even be infinite) than its range, it follows from the pigeonhole principle that many collisions must exist. A brute force attack can find a pre-image or second pre-image for a general hash function with n-bit hashes in approximately 2n hash operations. Because of the birthday paradox a brute force approach to generate collisions will succeed in approximately 2(n/2) hash operations. Any attack that requires less hash operations than the brute force attack is formally considered a break of a cryptographically hash function [2].
MD stands for „Message Digest" and describes a mathematical function that can take place on a variable length string. The number 5 simply depicts that MD5 was the successor to MD4. MD5 is essentially a checksum that is used to validate the authenticity of a file or a string and this is one of its

most common uses. Let‟s take a look at a working example. Let‟s say you have released some software or a program that you want people to freely distribute, this is all good and well but what if someone was to tamper with your application with malicious intent? For example what if they added malware onto your program, how would people know? Well if you had taken an MD5 checksum of your original program and made this information public, then when people downloaded your software could then check their downloaded file and check that the MD5 checksum matches yours. If it does then great! If not then it means your program has been tampered with. MD5, with the full name of the Message-digest Algorithm 5, is the fifth generation on behalf of the message digest algorithm. In August 1992, Ronald L.Rivest[1] submitted a document to the IETF (The Internet Engineering Taskforce) entitled "The MD5Message-Digest Algorithm", which describes the theory of this algorithm. For the publicity and security of algorithm, it has been widely used to verify data integrity in a variety of program languages since the 1990s.MD5 was developed from MD, MD2, MD3 and MD4. It can compress any length of data into an information digest of 128 bits while this segment message digest often claims to be a digital fingerprint of the data. This algorithm makes use of a series of non-linear algorithm to do the circular operation, so that crackers cannot restore the original data. In cryptography, it is said that such algorithm as an irreversible algorithm, can effectively prevent data leakage caused by inverse operation. Both the theory and practice have good security, because the use of MD5 algorithm does not require the payment of any royalties, time, and cost less which make it be widely used in the general non-top-secret applications. But even the top-secret area, MD5 may well be an excellent Intermediate technology [1].

SHA (Secure Hash Algorithm) is a 160-bit hash function published in 1993 as the Secure Hash Standard by NIST (The National Institute of Standards and Technology) [3]. It is based on MD5 and is mainly used in digital signature schemes. It hashes onto 160 bits and uses the Merkle- Damgard construction from a $160 \times 512 \rightarrow 160$ compression function. As for MD5, the compression functions of SHA and SHA-1 are made from an "encryption function" by the Davies-Meyer scheme.

A 160-bit hash function has a security level on the order of 80 bits, so SHA-1 is designed to match the security level that uses an 80-bit secret key [1]. SHA-0 analyzed by Chabaud and Joux using differential methods (local collisions and disturbance vectors) and they found a collision attack on SHA-0 of complexity 261 [4]. Biham and Chen found near collisions on SHA-0 in complexity 240 [5]. The work of Biham, Joux, and Chen included the first real collision of SHA-0 [5] therefore; the migration to more secure hash functions should be accelerated.

In 2001, NIST developed three new hash functions SHA-256, 384, and 512 whose hash value sizes are 256, 384, and 512 bits, respectively. These hash functions are standardized with SHA-1 as SHS (Secure Hash Standard) [6, 7], and a 224-bit hash function, SHA-224, based on SHA-256, was added to SHS in 2004 but moving to other members of the SHA family may not be a good long term solution [8].

2. MODIFIED MD5 ALGORITHM

MD5algorithm is co-invented by Rivets in MIT Computer Science Laboratory and RSA Data Security Company. MD5 is a non-reversible encryption algorithm [3]. It is widely applied in many aspects, including digital signature, encryption of information in a database and encryption of communication information. It makes large amounts of information to be compressed into a confidential format before signing the private key by digital signature soft (that is, any length byte string is transformed into a certain length of big integer).A brief description of new modified MD5 algorithm as follows: MD5algorithm divides plaintext input into blocks each which has512-bit, and each block is again divided into sixteen 32-bitmessage words, after a series of processing, the outputs of the algorithm consist of eight 32-bit message words. After these eight 32-bit message words are cascaded, the algorithm generates a 128-bit hash value which is the required cipher text. Specific steps are as follows [11, 12, 13].

2.1. PADDING-BIT

Without loss of generality, supposes that the original data at the source has k bits (mk-1, mk-2… m0), where mi_ {0, 1}. For MD5 algorithm, its k bits data must be processed in 512-bit message block, so if the length of source is less than that length, padding is always added until its length in bits is congruent to 448 modulo 512 (length ≡ 448 mod 512). The padding consists of a single 1-bit followed by the necessary number of 0-bits.

2.2. PADDING LENGTH OF DATA

A 64-bit representation of the length on bits of the original message is appended to the result of above step. It is present by two 32-bit digits. At this time, the length of message is filled to a multiple of 512.

2.3. INITIALIZE MD5 STANDARD PARAMETERS

Eight 32-bit integers A, B, C, D, E are called chaining variables, used to calculate the message digest, are initialized by hexadecimal number

A=0x01234567

B=0x89abcdef
C=0xfedcba98
D=0x76543210
E=0x12ac2375

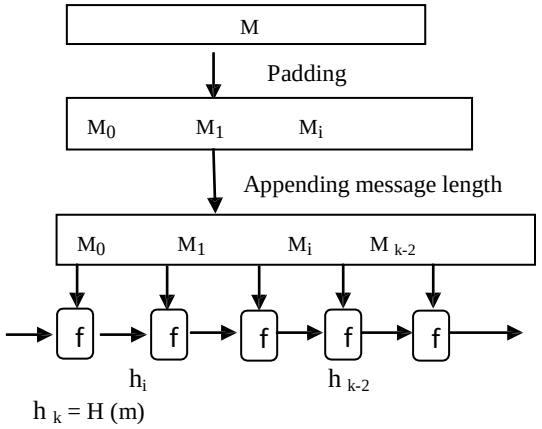

Figure 1: Working principle of an iterated hash function

2.4. BIT OPERATION FUNCTIONS

We define eight bit operation functions J, K, L, M, N, O, P and Q respectively in two group first J, K, L, M and other one N, O, P, Q, in which x, y, z are three 32-bit integers. The operation is as follows:

Group 1:

$J(x,y,z) = (x \wedge y) \vee ((\neg x) \wedge z)$ ……..…1
$K(x,y,z) = (x \wedge z) \vee (y \wedge (\neg z))$……...2
$L(x,y,z) = x \oplus y \oplus z$……...3
$M(x, y, z) = y \oplus (x \vee (\neg z))$……..……4

Group 2:

$N(x,y,z) = (x \wedge y) \vee ((\neg x) \wedge z)$ ……..…5
$O(x,y,z) = x \oplus y \oplus z$……..……..6
$P(x,y,z) = (x \wedge y) \oplus (y \wedge z) \vee (z \wedge x)$...…...7
$Q(x, y, z) = x \oplus y \oplus z$ ……..….
……..8

In eight functions, if the corresponding bits of x, y and z are independent and uniform, then each bit of the results should be independent and uniform as well. For

$$x = \sum_{i=1}^{32} x_i 2^{i-1} \in \frac{Z}{[(2]^{32})}, x^i \in \{0,1\}$$

We call x^i the i^{th} bit of x.

2.5. MAIN TRANSFORMATION PROCESS

The number of main looping this algorithm is the number of 512-bit information groups. The main loop have five rounds, each round carries out 80 operations. All operation run in a two group that assigned to another five chaining values: a0=A, b0=B, c0=C, d0=D, e0=E. One of the chaining values is updated in each step and computation is continued in sequence. Here we have defined five rounds composite functions of main loop nFF, FF, nGG, GG, nHH, HH, nII and II respectively, which change from nF, F, nG, G, nH, H, nI and I. the operation is as follows:

nFF →a = b + ((a + N (b c d) + Mi +ti)) ……..…9
FF →a = b + ((a + J (b c d) + Mi +ti)<< s)……..……10

nGG→a = b + ((a + O (b c d) + Mi +ti))……...……..11
GG→a = b + ((a + K (b c d) + Mi +ti)<<s)…….……12

nHH→a = b + ((a + P (b c d) + Mi +ti)s)……..……..13
HH →a = b + ((a + L (b c d) + Mi +ti)<<s)……....14

nII →a = b + ((a + Q (b c d) + Mi +ti))……..15
II →a = b + ((a + M (b c d) + Mi +ti) <<s)…….16

Where, + is addition modulo 2^{32}, Mi (0_i_15) is a 32-bit message word and the 512-bit message block is divided into 16 32-bit message words. x _ s is the left shift rotation of x by s bits. The ti and s are step-dependent constants, ti has the following options: in i-th step, ti is the integer part of 4294967296×abs (sin (i)), $4294967296=2^{32}$.

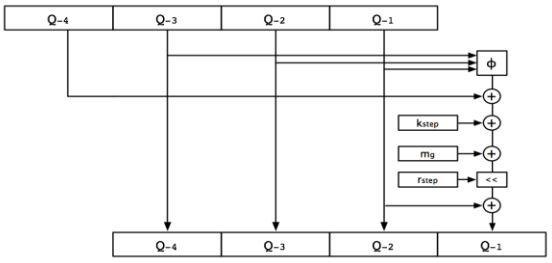

Figure 2: Standard step of compression function in MD5

mentioned in above. Each group has four compression functions, so the total operations are in 160 steps. The above five chaining variables are

After all of these steps, A, B, C, D, E, add a, b, c, d, e Respectively, then the algorithm is continued to run the next 512-bit message block, the final output is A, B, C, D, E of cascading. Application of MD5 algorithm is to generate a message digest of information in order to prevent tampering. The results of some strings are showing in table 2. We view the entire file as a large text message, and result in a unique message digest by the irreversible string transform method. In the future, if the contents of file are changed, we only recalculate message digest of this file, and will find the difference from the original message digest. There by, we can sure the checked file is incorrect. [14]

2.6. OUTPUTS

The *L* 512-bit blocks have been processed, the output from L^th age is the 160-bit message digest. The results of some strings are showing in table 2.

3. STRENGTH OF MD5-160

Cryptographic [9] hash functions are usually designed to be collision resistant. But many hash functions that were once thought to be collision resistant were later broken. MD5 and SHA-1 in particular both have published techniques more efficient than brute force for finding collisions. However, some compression functions have a proof that finding collisions is at least as difficult as some hard mathematical problem (such as integer factorization or discrete logarithm). Those functions are called provably secure. Every hash function with more inputs than outputs will necessarily have collisions. This hash function MD5-160 that produces 160 bits of output from an arbitrarily large input. Since it must generate one of 2^{160} outputs for each member of a much larger set of inputs, the pigeonhole principle[18] guarantees that some inputs will hash to the same output. Collision resistance doesn't mean that no collisions exist; simply that they are hard to find. The birthday "paradox" places an upper bound on collision resistance: if a hash function produces N bits of output, an attacker who computes "only" 2N/2 hash operations on random input is likely to find two matching outputs. If there is an easier method than this brute force attack, it is typically considered a flaw in the hash function.

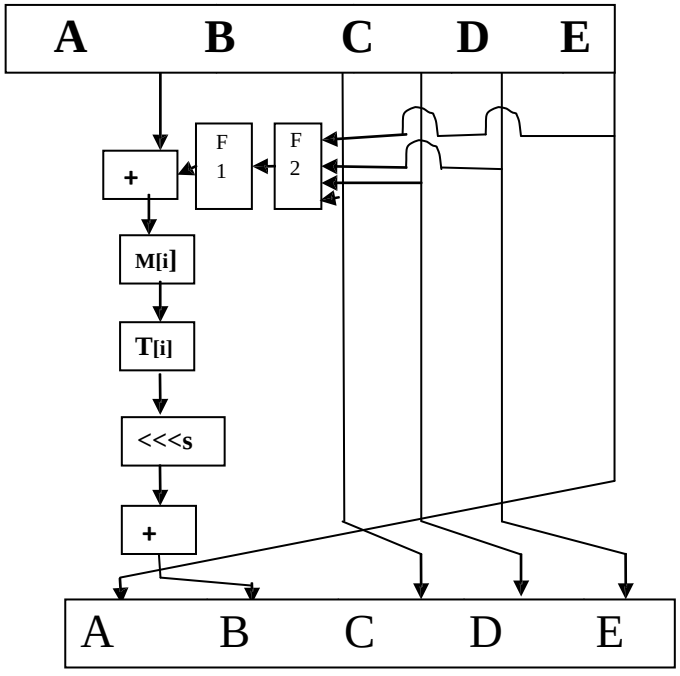

Figure 3: Enhanced MD5- 160 output

4. OTHER ATTECK

4.1. BRUTE FORCE ATTACK

In cryptography, a brute force attack or exhaustive key search is a strategy that can in theory be used against any encrypted data[15] by an attacker who is unable to take advantage of any weakness in an encryption system that would otherwise make his/her task easier. It involves systematically checking all possible keys until the correct key is found. In the worst case, this would involve traversing the entire search space.

4.2. RAINBOW TABLE

Tables are usually used in recovering the plaintext password, up to a certain length consisting of a limited set of characters. It is a form of time-memory tradeoff, using less CPU at the cost of more storage. Proper key derivation functions employ salt to make this attack infeasible. Rainbow tables are a refinement of an earlier, simpler algorithm by Martin Hellman[16] that used the inversion of hashes by looking up recomputed hash chains.

Symmetric key length vs. brute force combinations

Key size in bits[2]	Permutations	Brute force time for a device checking 256 permutations per second
8	28	0 milliseconds
40	240	0.015 milliseconds
56	256	1 second
64	264	4 minutes 16 seconds
128	2128	149,745,258,842,898 years

Table 4.1 Combinations Time

4.3. BIRTHDAY ATTACK

A Birth day attack is a name used to refer to a class of brute-force attacks. It gets its name from the surprising result that the probability that two or more people in a group of 23 share the same birthday is greater than 1/2; such a result is called a birthday paradox. If some function, when supplied with a random input, returns one of k equally-likely values, then by repeatedly evaluating the function for different inputs, we expect to obtain the same output after about 1.2k1/2. For the above birthday paradox, replace k with 365. Birthday attacks are often used to find collisions of hash functions.

5. RESULT

We have presented a new hash function based on Double-Davies-Meyer that satisfied Merkle-

Damgard condition. Security of this new algorithm is higher than SHA-1 and MD5 because even if local collision happened in the middle of SHA second algorithm will fade it with an acceptable diffusion so at the start of next round there are no equal states with previous round. It means that Hi – Hi-1 ≠ 0 We chose some messages that has already shown as collision in MD5 and SHA-1 and changed them by XOR and Addition but we did not find any collision. Sophisticated message modification techniques were applied to achieve the necessary conditions on the chaining variables and the message bits. The differential path for the improved algorithm is different from the previous differential path so it is resistant against local collision and differential attack.

In this scheme brute force attack takes 256 Permutations to break algorithm which takes more time than 160 bit hash function as show in the Table1. Even the last scheme is 160 bits and need 280 bit for birthday paradox but it is strong enough to the first and second pre-image attack. We can extend the length of hash to 512 or 640 to be more resistible against birthday attack which comparing with its ancestors, it is more powerful. Even this scheme is 160 bits and need 280 bit for birthday paradox but it is strong enough to the first and second pre image attack. We can extend the length of hash to 256 or 512 to be more resistible against birthday attack which comparing with its ancestors, it is more powerful. In cryptography, a brute force attack or exhaustive key search is a strategy that can in theory be used against any encrypted data[15] by an attacker who is unable to take advantage of any weakness in an encryption system that would otherwise make his/her task easier. It involves systematically checking all possible keys until the correct key is found. In the worst case, this would involve traversing the entire search space.

Message	**MD5 (128bits)**	**SHA-1 (160bits)**	**Enhanced MD5(160bits)**
""	D41D8CD9 8F00B204 E9800998 ECF8427E	DA39A3EE 5E6B4B0D 3255BFEF 95601890 AFD80709	25A12C0F 20F9CA9E A78B1A92 FCE44ED2 88C903D0
"a"	0CC175B9 C0F1B6A8 31C399E2 69772661	86F7E437 FAA5A7FC E15D1DDC B9EAEAEA 377667B8	FE0173D2 6A66A447 BFBB6511 4ADEA4ED 408C1925
"ABCDEF GHIJKLM NOPQRST UVWXYZ abcdefg hijklmn opqrstu vwxyz 0123456 789"	D174AB98 D277D9F5 A5611C2C 9F419D9F	761C457B F73B14D2 7E9E9265 C46F4B4D DA11F940	7BFDB0BB AC222369 0B6659A2 61AF9C51 A5B500AD

Table 5.1 Result Table

6. CONCLUSION

In this paper, we have proposed a new message digest algorithm based on previous algorithms that can be used in any message integrity or signing applications. The simplified MD5 message-digest algorithm is simple to implement, and provides a "fingerprint" or message digest of a message of random length. It is conclusion that the complexity of coming up with two messages having the same message digest is on the order of 2^{64} operations, and that the complexity of coming up with any message having a given message digest is on the order of 2^{160} operations. The Message Digest 128 bit algorithm is slightly cheaper to compute, however Message Digest Algorithm 128 and SHA-1 are currently very vulnerable to collision attacks. Message digest compresses any stream of bytes into a 160 bit value with extra compression function. This compression only goes one way. If you give the hash of a random stream of bytes to someone, there is no theoretical way for them to go back to unique stream of bytes.

.

7. REFERENCES

[1] R. Rivest. The MD5 Message-Digest Algorithm [rfc1321].

[2]. S.Vaudenay "A Classical Introduction to Cryptography Applications for Communications Security" Springer, 2006, P 74.

[3]. NIST, "Secure Hash Standard," FIPS PUB 180, May. 1993.

[4]. F. Chabaud, A. Joux. "Differential Collisions in SHA-0". In Advances in Cryptology CRYPTO˝98, Santa Barbara, CA, Lecture Notes in Computer Science 1462. Springer-Verlag, NY, pp. 56–71, 1998.

[5]. E. Biham, R. Chen, A. Joux, P. Carribault, W. Jalby and C. Lemuet. "Collisions in SHA-0 and Reduced SHA-1- In Advances in Cryptology" – Eurocrypt˝05, Springer-Verlag, 2005.

[6]. NIST FIPS PUB 180-1. Oct.2001.

[7]. NIST, "Secure Hash Standard (SHS)", FIPS PUB 180-2, 2002.

[8]. S. Chang, M. Dworkin, Workshop Report, The First Cryptographic Hash Workshop, Report prepared, NIST 2005.

[9] http://www.springer.com/978-0-387-25464-7."A Classical introduction to cryptography"

[10]. F. Chabaud, A. Joux. "Differential Collisions in SHA-0". In Advances in Cryptology CRYPTO˝98,

Santa Barbara, CA, Lecture Notes in Computer Science 1462. Springer-Verlag, NY, pp. 56–71, 1998.
[11] Rivest R L. The MD5 message digest algorithm [EB/OL].
[12]Xiaoyun Wang, Dengguo, k., m., m, HAVAL-128 and RIPEMD], Cryptology ePrint Archive Report 2004/199, 16 August 2004,
[13] J. Black, M. Cochran, T. Highland: A Study of the MD5 Attacks: Insights and Improvements, March 3, 2006
[14]Tao Xie and DengguoFeng (30 May 2009). How to Find Weak Input Differences for MD5 Collision Attacks.
[15]ChristofPaar, Jan Pelzl, Bart Preneel (2010). Understanding Cryptography: A Textbook for Students and ractitioners.Springer.p. 7. ISBN 3642041000.
[16]M.E. Hellman, H.R. Amirazizi, "A Cryptanalytic Time - Memory Trade-Off," IEEE Transactions on Information Theory, vol. 34-3, pp. 505-512, 1988
[17]. X. Wang, X. D. Feng, X. Lai and H. Yu., "Collisions for Hash Functions MD4, MD5, HAVAL-128 and RIPEMD," Cryptology ePrint Archive: Report 2004/199, Aug. 2004 http://eprint.iacr.org/2004/199/
[18]. pigeonhole principle "http://www.math.ust.hk/~mabfchen/Math391I/Pigeonhole.pdf"