ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

An Enhanced Message Digest Hash Algorithm for Information Security

[1]Akanksha Rawat, [2]Deepak Agrawal

[1]Research Scholar, [2]Professor, TITECH Jabalpur, India

Abstract: **Information is an important commodity in the world of Electronic communication. To achieve a secure communication between communicating parties, the protection of authenticity and integrity of information is necessary. Cryptographic hash functions play a central role in cryptology. A cryptographic hash function takes an input of arbitrary large size and returns a small fixed size hash value. It satisfies three major cryptographic properties: preimage resistance, second preimage resistance and collision resistance. Due to its cryptographic properties hash function has become an important cryptographic tool which is used to protect information authenticity and integrity. This thesis presents a review of cryptographic hash functions. The thesis includes various applications of hash functions. It gives special emphasis on dedicated hash functions MD5. 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 thesis is combination of some function to reinforce these functions and also increasing hash code of message digest of length up to 160 that makes stronger algorithm against collision and brute force attacks.**

Keywords: **Hash Algorithm, Information Security.**

1. INTRODUCTION

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.

The MD5 message digest algorithm was developed by Ron Rivest at MIT. Until the last few years when both burst force and cryptanalytic concerns have arose , MD5 was most widely used secure hash algorithm. It is a widely-used 128-bit hash function, used in various applications Including SSL/TLS, IPSec, and many other cryptographic protocols. The MD5 algorithm breaks a sleeve into 512 bit input blocks. Every block is run from side to side a series of functions to produce a exceptional bit hash value for the sleeve [1].

2. HASH FUNCTION

A hash function H is a transformation that takes a variable-size input m and proceeds a fixed-size string, which is called the hash value h .Hash functions with just this property have a variety of general computational uses, but when working in

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

cryptography the hash functions are regular chosen to have some supplementary properties. This is a contract in lots of programming languages that allocate the user to dominate equality and hash functions for an object, that if two objects are the same their hash codes must be the same. Hash functions compress a n (arbitrarily) large number of bits into a small number of bits.

The hash function properties are:-

- Output does not reveal information on input.
- Hard to find collisions (different messages with same hash).
- One way cannot be reversed.

3. MD5 ALGORITHM

MD5 stands for Message-Digest Algorithm 5. MD5 algorithm 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 MD5 algorithm as follows: MD5 algorithm divides plaintext input into blocks each which has 512- bit, and each block is again divided into sixteen 32-bit message words, after a series of processing, the outputs of the algorithm consist of four 32-bit message words. After these four 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 [3, 4, 5].

(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 448modulo 512 (length≡448 mod 512). The padding consists of a single 1-bit followed by the necessary number of 0-bits.

(2) Padding the 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-bitdigits. At this time, the length of message is filled to amultiple of 512.

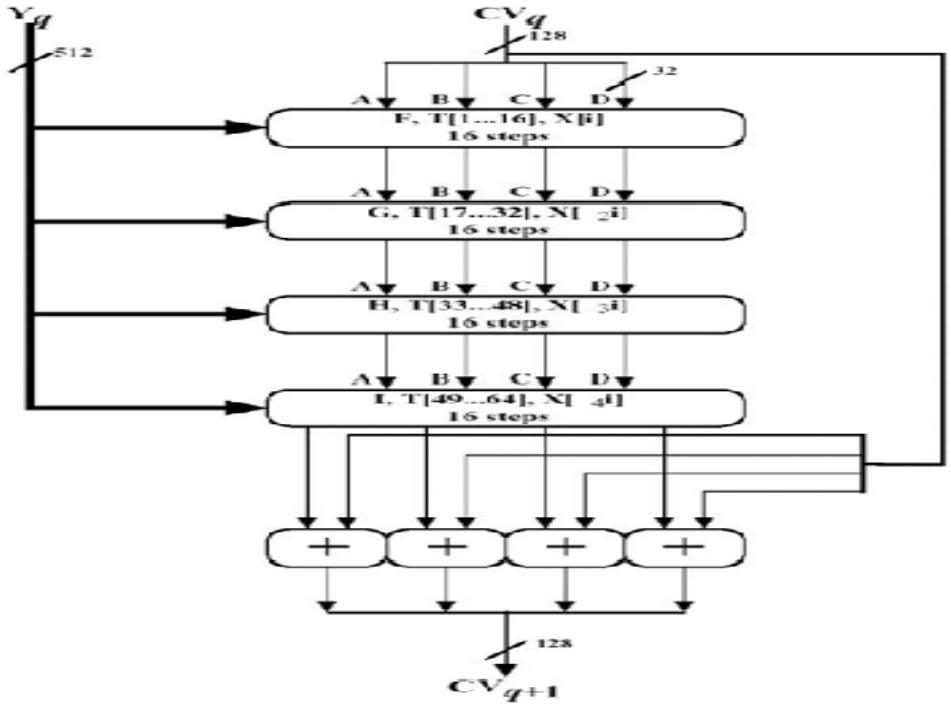

Figure 1: Working principle of an iterated hash function

(3) Initialize MD5 Parameters: four 32-bit integers A, B, C, D are called chaining variables, used to calculate themes age digest, are initialized by hexadecimal number

A=0x01234567

B=0x89abcdef

C=0xfedcba98

D=0x76543210

(4) Bit operation functions: We define four bit operation functions, F, G, H and I respectively, in which x, y, z are three 32 -bit integers. The operation is as follows:

$F(x, y, z) = (x \wedge y) \vee ((\neg x) \wedge z)$ ……..…….1

$G(x, y, z) = (x \wedge z) \vee (y \wedge (\neg z))$……..….2

$H(x, y, z) = x \oplus y \oplus z$..…..……..…..….3

$I(x, y, z) = y \oplus (x \vee (\neg z))$…. ……..……...4

(5) Main transformation process: The number of main loop in this algorithm is the number of 512-bit information groups. The main loop have four rounds, each round carries out 16 operations, so the total of operations are64 steps. The above four chaining variables are assigned to another four chaining values: a0=A, b0=B, c0=C, d0=D. One of the chaining values is updated in each step and computation is continued in sequence. Here we have defined four rounds composite functions of main loop FF, GG, HH, and II respectively, which change from F, G, H and I. the operation is as follows:

FF →a = b + ((a + F (b c d) Mi+ ti)<< s)……..…5

GG →a = b + ((a + H (b c d))Mi+ ti<<s)…….....6

HH →a = b + ((a + H (b c d))Mi+ ti<<s)…..…...7

II →a = b + ((a + I (b c d)) Mi+ ti<<s)……..…8

Where, + is addition module232, 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 = 232.

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

4. SECURITY ATTACKS

Security attack can be classified in two categories Active attack and Passive attack.

4.1 Active attack:

An intrusion into a computer network which attempts to delete or modify the data stored on the computers which form part of the network. This is one of the most important forms of strike since many companies operations deprecatory depend on data. This type of attack requires the assailant to be able to transmit data to one or both of the parties, or block

the data stream in one or both directions. It is also feasible that the assailant is detected between the communicating parties. In this condition the assailant can terminate all or parts of the data sent by the communicating parties. This assailant can e.g. try to take the place of the client (or server) when the authentication method has been executed. Without probity checks of the received data, the server will not find out that the origin of the data is not the validated person. A wise programmer can, with not too much attempt, execute a system like this on a computer acting as a gateway (bridge) between two subnets.

4.2 Passive Attack:

An intrusion into a computer network which reads the data passing along some of the transmission lines without modifying it. A passive attack is an attack where an unauthorized assailant monitors or listens in on the communication between two parties. A passive attack on a cryptosystem is one in which the cryptanalyst cannot interact with any of the parties intricate, attempting to break the system solely based upon observed data. This can also include known plain text attacks where [2] both the plain text and its comparable cheaper text are available.

A pounce is the procedure of trying to find out a way around security controls on a computer system. It can be active or passive. An active attack is an attack in which the assailant manipulates data and adds unauthorized data. In a passive attack, the assailant only monitors and/or records data.

4.3 Password Guessing Attack:

A different type of network attack is Password Guessing attack. Here a permissible users access rights to a computer and network capitals are compromised by recognizing the user id/password combination of the permissible user. This attack takes place when an unauthorized user frequently tries to log on to a computer or network by approximating usernames and passwords. Many password-guessing programs that used to guess passwords are available on the Internet. Following are the types of password guessing attacks.

4.4 Brute Force Attack:

In cryptography, a brute-force attack, or comprehensive key search, is a master plan that can, in theory, be employed against any converted data. Such an attack might be deployed when it is not feasible to take favor of other weaknesses in an encryption system that would convert it in easier task. The key length used in the encryption regulates the practical employability of executing a brute-force attack, with longer keys exponentially more tough to crack than shorter ones. Brute-force attacks can be made less successfully by confusing the data to be encoded, something that makes it tougher for an attacker to identify when he/she has break the code.

4.5 Dictionary Attack:

A dictionary attack uses a specified technique of continuously trying all the words in a comprehensive list called a dictionary. In difference with a brute force attack, where a large portion key space is searched specifically, a dictionary attack strives only those probability which are most likely to triumph, typically derived from a list of words for example a dictionary.

4.6 Denial-of-Service Attack:

A Denial-of-Service (DoS) attack sources a negative effect on the staging of a computer or network. This attack is depicting to bring loss of network connectivity and services by absorbing the bandwidth of the user's network. It can also define as network saturation attack or bandwidth consumption attack. Attacker makes Denial-of-Service attacks by introducing a large number of protocol packets to a network. DoS (Denial-of-Service) attacks are possibly the nastiest, and most difficult to find.

4.7 Replay Attack:

Replay attack is a type of attack in which assailants capture packets containing passwords or digital signatures whenever packets pass between two hosts on a network. These assailants then filter the data and extract the passwords, encryption keys, or digital signatures from the captured packets. In an attempt to obtain an authenticated connection, the assailants then resend this information to the system.

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

5. LITERATURE REVIEW/ WORK DONE

MD5 is one in a series of message digest algorithms designed by Professor Ronald Rives of MI (Rivest, 1994). When analytic work indicated that MD5's predecessor MD was likely to be insecure, MD5 was designed in 1991 to be a secure replacement. (Weaknesses were indeed later found in MD4 by Hans Dobbertin.)

In 1993, Den Boer and Bosselaers gave an early, although limited, result of finding a "pseudo-collision" of the MD5 compression function; that is, two different initialization vector which produce an identical digest. At Crypto '91 Ronald L. Rivest introduced the MD5 MessageDigest Algorithm as a strengthened version of MD4, differing from it on six points. Four changes are due to the two existing attacks on the two round versions of MD4. The other two changes should additionally streng then MD5. However both these changes cannot be described as well-considered. One of them results in an approximate relation between any four consecutive additive constants. The other allows creating collisions for the compression function of MD5.

Bert den Boer and Antoon Bosselaers implemented a C program algorithm that establishes a work load offinding about 216 collisions for the first two rounds of the MD5 compression function to find a collision for the entire four round functions. On a 33MHz 80386 based PC the mean run time of this program is about 4 minutes.[9]

In 1996, Dobbertin announced a collision of the compression function of MD5 (Dobbertin, 1996). While this was not an attack on the full MD5 hash function, it was close enough for cryptographers to recommend switching to a replacement, such as SHA- or RIPEMD-16Hans Dobbert in proposed an attack on the compress function of MD5, which is based on similar methods as previous attacks on RIPEMD, MD4and the 256-bit extension of MD4 (see [18], [19]). Below we give a collision1of the compress function of MD5. [10]

Since the first feasible collision differential was given for MD5 in 2004 by Wang et al, a lot of work has been concentrated on how to improve it, but the researches on how to select weak input deference for MD5 collision attack are only sporadically scattered in literature. Collision resistance of several hash functions was broken by Wang et al. The strategy of determining message differential is the most important part of collision attacks against hash functions. So far, there are only three other message differentials attack published, one of which is 6 bits difference and two are 1 bit difference. [4]Wang et al. revealed the first MD5 collision with 6 bits message differences in two-block message. Later, a great many of researches greatly improved the complexity of computations and the best result is 230 MD5 operations. However, no other message. Differentials are found to generate collision before 2008.

In 2004, more serious flaws were discovered, making further use of the algorithm for security purposes questionable; specifically, a group of researchers described how to create a pair of files that share the same MD5 checksum.[4][5] Further advances were made in breaking MD5 in 2005, 2006, and 2007.[6] In an attack on MD5 published in December 2008, a group of researchers used this technique to fake SSL certificate validity.[7][8]

The size of the hash—128 bits—is small enough to contemplate a birthday attack. MD5CR was a distributed project started in March 2004 with the aim of demonstrating that MD5 is practically insecure by finding a collision using a birthday attack. In 2009, the United States Cyber Command used an MD5 hash of their mission statement as a part of their official emblem.[15

On December 24, 2010, Tao Xie and Dengguo Feng announced the first published single-block MD5 collision (two 64-byte messages with the same MD5 hash).[16 Previous collision discoveries relied on multi-block attacks. For "security reasons", Xie and Feng did not disclose the new attack method. They have issued a challenge to the cryptographic community, offering a US$ 10,000 reward to the first finder of a different 64-byte collision before January 1, 2013.

In 2011 an informational RF was approved to update the security considerations in RFC 132 (MD5) and RFC 210 (HMAC-MD5). HMAC-MD5 describes a keyed-MD5 mechanism (called HMAC-MD5) for use as a message authentication code (or, integrity check value). It is mainly intended for integrity verification of information transmitted over open networks (e.g., Internet) between parties that share a common secret key. The proposed mechanism combines the (key-less) MD5 hash function [RFC-132] with a shared secret key. [17]

In, January 29, 2012, Cryptology Group, CWI, announced the collision attack on MD5.
Xie Tao, Fanbao Liu, and Dengguo Feng (2013). Found Fast Collision Attack on MD5.

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

6. PROPOSED FRAMEWORK

In this dissertation work, we have proposed an advanced message digest hash algorithm, which is better to dealt with pre-image attack in contrast to actual MD5 hash algorithm. Our advanced message digest hash algorithm enclosed following features-

1. We have used 4 rounds each of 20 operations (actual MD5 hash algorithm also uses 4 rounds but each of 16 operations.
2. Because we have used 4 rounds each of 20 operations, so that we uses 5 word buffers A, B, C, D and E each of 32 bits for computing message digest.
3. Fixed length 160-bit message digest is produced by the proposed algorithm as output, which is more robust and secure as compared to 128-bit message digest produced by the actual MD5 hash algorithm.

6.1 Proposed Algorithm:

The proposed algorithm is depicted by the following steps:

6.2.1 Padding Bits:

The message is "padded" (extended) so that its length (in bits) is congruent to 448, modulo 512. That is, the message is extended so that it is just 64 bits shy of being a multiple of 512 bits long. Padding is always performed, even if the length of the message is already congruent to 448, modulo 512.Padding is performed as follows: a single "1" bit is appended to the message, and then "0" bits are appended so that the length in bits of the padded message becomes congruent to 448, modulo 512. In all, at least one bit and at most 512 bits are appended [11].

A 64-bit representation of b (the length of the message before the padding bits were added) is appended to the result of the previous step. In the unlikely event that b is greater than 2^64, then only the low-order 64 bits of b are used [11].

6.2.3 Initialize MD Buffer:

A Five-word buffer (A, B, C, D, E) is used to compute the message digest. Here each of A, B, C, D, E is a 32-bit register. These registers are initialized to the following values in hexadecimal, low-order bytes first.

Word A: 01 23 45 67

Word B: 89 ab cd ef

Word C: fe dc ba 98

Word D: 76 54 32 10

Word E: c3 d2 d1 f0

This step uses a 80-element table T [1 ... 80] constructed from the sine function.

6.2.4 Process Message in 16-Word Blocks:

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

Group 1:

$F(x,y,z) = (x \wedge y) \vee ((\neg x) \wedge z)$ …….....1

$G(x,y,z) = (x \wedge z) \vee (y \wedge (\neg z))$……...2

$H(x,y,z) = x \oplus y \oplus z$..……...3

$I(x, y, z) = y \oplus (x \vee (\neg z))$……4

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

Group 2:

J(x,y,z) =(x∧ y)∨ ((¬x) ∧ z) ……...…...5

K(x,y,z) = x⊕y⊕z……..……6

L(x,y,z) =(x∧ y) ⊕(y∧ z) ∨ (z∧ x)...….....7

M(x, y, z) = x⊕y⊕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.

7. COMPARISON OF COMPRESSION FUNCTION

The modified message algorithm has eight compression functions in two groups. Four function in each group. It is compared with the existing message digest which generates 128 bits secure output and sha-1 which generates 160 bits secure output. The existing message digest md5 and sha1 perform four compression functions whereas the modified message algorithm has eight compression functions. The comparison of compression function of modified message digest, existing message digest and sha-1 is shown here:

8. ADVANCED MESSAGE DIGEST COMPRESSION FUNCTION

Group 1:

J(x,y,z) =(x∧y)∨ ((¬x) ∧ z) …….....1

K(x,y,z) = (x∧ z) ∨(y∧ (¬z))…….....2

L(x,y,z) = x ⊕ y ⊕ z..…….....3

M(x, y, z) = y ⊕ (x ∨ (¬z))……4

Group 2:

N(x,y,z) =(x∧y)∨ ((¬x) ∧ z) ……...….....5

O(x,y,z) = x ⊕ y⊕ z……..…….6

P(x,y,z) =(x∧y) ⊕ (y∧z) ∨ (z∧x)...….....7

Q(x, y, z) = x ⊕ y ⊕ z …….…. ……8

Message Digest 5 Compression Function

J(x,y,z) =(x∧y)∨ ((¬x) ∧ z) ……...…….1

K(x,y,z) = (x∧ z) ∨(y∧ (¬z))……...2

L(x,y,z) = x ⊕ y ⊕ z..……3

M(x, y, z) = y⊕(x ∨ (¬z))…….…4

SHA-1 Compression Function

N(x,y,z) =(x∧y)∨ ((¬x) ∧ z) ……...……1

O(x,y,z) = x ⊕ y ⊕ z……..……..2

P(x,y,z) =(x∧y) ⊕ (y∧z) ∨ (z∧x)...…….....3

Q(x, y, z) = y⊕(x ∨ (¬z))……...4

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 2, Issue 1, pp: (54-62), Month: January - March 2015, Available at: www.paperpublications.org

9. ALGORITHM COMPARISON [19, 20]

Table 1.1 Comparison of MD5, SHA-1 and Advanced Message Diegest

Function	MD5	SHA-1	Advanced Message Digest
Block length	512 bit	512 bit	512 bit
Algorithm length	128 bit	160 bit	160 bit
Rotation steps	64 steps	80 steps	160 steps
Initialization variables	4	5	5
Compression functions	4	4	8
Brute Force Attack Operations	2^{64}	2^{64}	2^{160}
Brute Force Attack Time($4.7*10^{9}$ Hash/Sec)	$53*10^{-11}$ yrs	38 yrs	2^{85} yrs
Security against cryptanalysis	Vulnerable to attacks	Vulnerable to attacks	Not to be vulnerable yet

10. CONCLUSION AND FUTURE WORK

At present work, it is proposed an enhanced message digest algorithm, which is better to dealt with pre-image attack in contrast to actual Message Digest hash algorithm.

Our enhanced Message Digest hash algorithm enclosed following features-

1. We have used 4 rounds each of 20 operations (actual MD5 hash algorithm also uses 4 rounds but each of 16 operations.
2. Because we have used 4 rounds each of 20 operations, so that we uses 5 word buffers A, B, C, D and E each of 32 bits for computing message digest.
3. Fixed length 160-bit message digest is produced by the proposed algorithm as output, which is more robust and secure as compared to 128-bit message digest produced by the actual MD5 hash algorithm.

We observed that the Message Digest algorithm with 128 bit message digest length is slightly cheaper to compute, but it is currently very vulnerable to collision attacks, as compared to Message Digest algorithm with 160 bit message digest length. Based on the above analyzed results we conclude that the enhanced Message Digest Algorithm is more strong, secure and better to dealt with pre-image attack in contrast to actual MD5 hash algorithm. In future we have to analyze the Message Digest algorithm with higher bit values then proposed algorithm and were compare the results. We can extend the length of hash to 256 or 512 bits to be more resistible against birthday attack.

REFERENCES

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

[2] Tao Xie and Dengguo Feng (30 May 2009). How to Find Weak Input Differences for MD5 Collision Attack.

[3] Rivest R L. The MD5 message digest algorithm [EB/OL].

[4] Xiaoyun Wang, Dengguo, k., m., m, HAVAL-128 and RIPEMD], Cryptology ePrint Archive Report 2004/199, 16 August 2004,

[5] J. Black, M. Cochran, T. Highland: A Study of the MD5 Attacks: Insights and Improvement, March 3, 2006

[6] Marc Stevens, Arjen Lenstra, Benne de Weger: Vulnerability of software integrity and code signing applications to chosen-prefix collisions for MD, Nov 30, 2007.

[7] Sotirov, Alexander; Marc Stevens, Jacob Appelbaum, Arjen Lenstra, David Molnar, Dag Arne Osvik, Benne de Weger (2008-12-30). "MD5 considered harmful today. Announce at the 25th Chaos Communication Congress.

[8] Stray, Jonathan (2008-12-30). "Web browser flaw could put e-commerce security at risk.

[9] Bert den Boer, Antoon Bosselaers. Collisions for the compression function on Md5.

[10] Hans Dobbertin. Cryptanalysis of MD5 Compress

[11] Philip Hawkes and Michael Paddon and Gregory G. Rose: Musings on the Wang et al. MD5 Collision, 13 Oct 2004.

[12] Arjen Lenstra, Xiaoyun Wang, and Benne de Weger: Colliding X.509 Certificate, Cryptology ePrint Archive Report 2005/067, 1 March 2005.

[13] Vlastimil Klima: Finding MD5 Collisions – a Toy for a Noteboo, Cryptology ePrint Archive Report 2005/075, 5 March 2005, revised 8 March 2005.

[14] Vlastimil Klima: Tunnels in Hash Functions: MD5 Collisions within a Minute, Cryptology ePrint Archive Report 2006/105, 18 March 2006, revised 17 April 2006.

[15] http://www.wired.com/dangerroom/2010/07/code-cracked-cyber-command-logos-mystery-solved Code Cracked! Cyber Command Logo Mystery Solved

[16] http://eprint.iacr.org/2010/64

[17] RFC 615, Updated Security Considerations for the MD5 Message-Digest and the HMAC-MD5 Algorithms

[18] H. Tiwari and K. Asawa, 2010, “A Secure Hash Function MD-192 with Modified Message Expansion”, IJCSIS, Vol. 7, No. 2, pp. 108-111.

[19] http://blog.codinghorror.com/speed-hashing/

[20] www.troyhunt.com201206our-password-hashing-has-no-clothes.html.