A review of Comparative Study of MD5 and SHA Security Algorithm

Surbhi Aggarwal
Student
YMCA University
Faridabad

Neha Goyal
Astt. Professor
ITM University
Gurgaon

Kirti Aggarwal
Astt. Professor
MRCE, Faridabad

ABSTRACT
Security algorithms or cryptography enables secure communication between two parties in the presence of a third-party or an eavesdropper. It assures the recipient of the message of the authenticity of the claimed source, protects the message against the unauthorized release of the message content by the adversaries, limits the access to authorized users, protects against sender/receiver denying sending/receiving a message.MD5 and Security Hash algorithms (SHA), cryptographic hash algorithms are one-way hashing functions which are easier to compute but are much harder to reverse and would take around millions of years to compute the authentic or veritable message content. This research paper aims to analyze and juxtapose the two hash algorithms, MD5 and SHA, using various key features and performance metrics. Their features have also been highlighted in order to provide the researchers a better comparison picture so that they can reach to the final upshot, which algorithm has superseded the other.

General Terms
Security, cryptography, hash algorithm

Keywords
MD5, SHA, hash

1. INTRODUCTION
In current scenario, where the number of internet users is widely increasing, internet has become the primary medium of communication. So, user's data attains the greatest priority in the field of data communication. To keep the network usage reliable, data integrity, data authentication, non-repudiation, data confidentiality is of utmost importance.

Cryptography is an important technique used for secure communication in the presence of third party or eavesdroppers. It provides all the paramount aspects of information security such as data confidentiality, data integrity, authentication and non-repudiation. [3]

Cryptography is defined as a process of making foremost piece of information indecipherable to attackers and available only to the intended recipients. So, now the data can be easily transferred securely without the threat of information being leaked or compromised. [10]

The popular methods of cryptography are:

1. Symmetric-key cryptosystem, in which the same public key is used by both the sender to send the message and the receiver to retrieve the message, that is, the same key is used both for encryption and decryption. The various symmetric-key cryptosystems are: DES (Modes: ECB, CBC, CFB, OFB, CM), 3DES, AES, IDEA, Blowfish, RC4, RC5, CAST, SAFER, Twofish. [11]

2. Asymmetric-key cryptosystem, in which two different keys are used for encryption and decryption. The sender uses the public key to encrypt the message and the receiver uses the private key to decrypt the message. The various asymmetric-key cryptosystems are: Diffie-Hellman, RSA, El Gamal , Elliptic Curve Cryptography (ECC).[11]

3. Hybrid cryptosystem: Combines strengths of both methods (symmetric and asymmetric-key cryptosystems). Asymmetric distributes symmetric key, also known as a session key. Symmetric provides bulk encryption. The example of a hybrid cryptosystem is SSL[11]

The shortcomings of a symmetric-key algorithm:

• Key-exchange becomes a problem.

• Trust problems among the intended parties occur.

• More damage is caused, when someone gets their hands on a symmetric key because they can decrypt everything encrypted with that key.

• As the number of participants using the secret key increases, the risk of damage and the consequences of this damage increases. [10]

The shortcomings of a asymmetric-key algorithm:

• Since asymmetric-keys must be many times longer than the secret-key in symmetric-key algorithm, asymmetric-keys are more computationally costly.

• They are susceptible to attacks in less than brute-force time.

• It is also vulnerable to man in the middle attack.

• Many public key systems use a third party to certify the reliability of public keys. [10]

The message-digest or one-way hashing functions were then proposed as an alternative to fulfill all the aspects of information security because of the following features:

•It is computationally easy to calculate the hash of any given message.

• With the same hash, there can never be two messages associated with it.

• Message cannot be changed without any changes in the hash value.

• It is infeasible to generate the message with the given hash value. [3]

The two cryptographic hashing algorithms widely known are:

- MD5
- Security Hash Algorithm (SHA).

1.1 MD5

It is one of message digest algorithm given by Professor Ronald Rivest in 1991 to be a secure replacement of its predecessor MD4 [4].

- Input: message of arbitrary length.
- Output: 128 bit hash code.

The input message is broken up into chunks of 512-bit blocks (sixteen 32-bit words).In case the message is not an integer multiple of 512-bit blocks, the message is padded so that its length is divisible by 512.

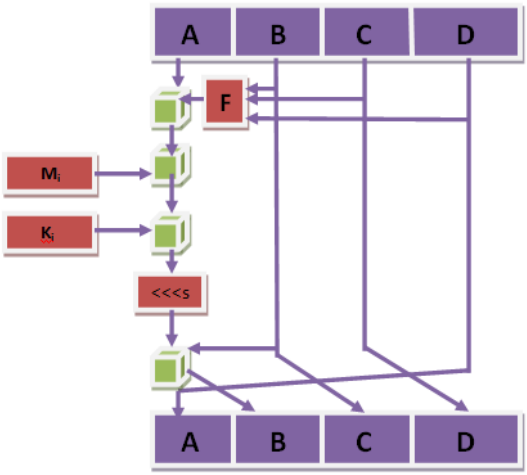

Fig.1 MD5 Algorithm

One MD5 operation: MD5 consists of 64 of these operations, grouped in four rounds of 16 operations. F is a nonlinear function; one function is used in each round. Mi denotes a 32-bit block of the message input, and Ki denotes a 32-bit constant, different for each operation. s denotes a left bit rotation by s. [3]

The main MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted A, B, C, and D. These are initialized to certain fixed constants.

A = 0x67452301

B = 0xEFCDAB89

C = 0x98BADCFE

D = 0x10325376

The main algorithm then uses each 512-bit message block in turn to modify the state. The processing of a message block consists of four similar stages, termed rounds; each round is composed of 16 similar operations based on a non-linear function F, modular addition and left rotation. The above figure illustrates one operation within a round. There are four possible functions F; a different one is used in each round:

F(B,C,D)=(B AND C) OR (NOT B AND D)

G(B,C,D)= (B AND D) OR (C AND NOT D)

H(B,C,D)= B XOR C XOR D

I(B,C,D)= C XOR (B OR NOT D)

The output is called a hash value, a fingerprint or a message digest. [1][3][4]

1.2 Secure Hash Algorithm (SHA)

• SHA-0: It was removed soon after publication because of "paramount flaw" and was replaced by a revised version SHA-1.

• SHA-1:It works similar to MD5 and produces a 160-bit message digest.. This was designed by the National Security Agency (NSA) to be part of the Digital Signature Algorithm. It was no longer used for most cryptographic uses after 2010 because of the cryptographic weaknesses discovered in the working.

• SHA-2: It were also formulated by the NSA.A family of two similar hash functions, with different block sizes, known as SHA-256 and SHA-512,they differ in the word size; SHA-256 uses 32-bit words where SHA-512 uses 64-bit words.

• SHA-3: It was proposed in 2012 after a public competition among non-NSA designers. It supports the same hash lengths as SHA-2, and its internal structure differs significantly from the rest of the SHA family.

1.2.1 SHA-1

- Input: message of arbitrary length.
- Output: 160 bit hash code.
- The input message is broken up into chunks of 512-bit blocks (sixteen 32-bit words).In case the message is not an integer multiple of 512-bit blocks, the message is padded so that its length is divisible by 512[5].
- The padding works as follows: Pad the message with a single 1 followed by 0's until the final block has 448 bits and append the size of the original message as an unsigned 64-bit integer.

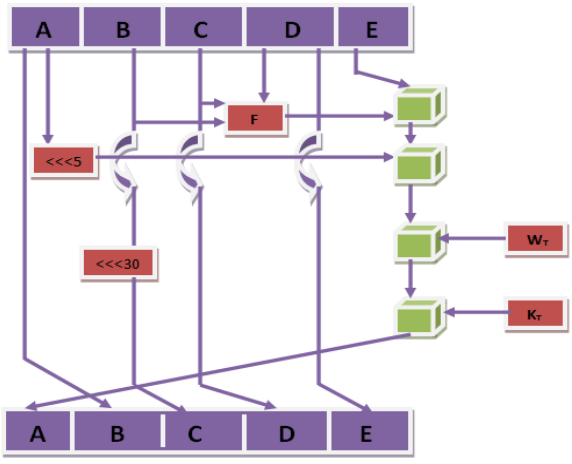

Fig.2 SHA-1 Algorithm

One iteration within the SHA-1 compression function: A, B, C, D and E are 32-bit words of the state;

F is a nonlinear function that varies; n denotes a left bit rotation by n places;

n varies for each operation; Wt is the expanded message word of round t; Kt is the round constant of round t; denotes addition modulo 232.

The SHA algorithm operates on a 160-bit state, divided into five 32-bit words, denoted h0, h1, h2, h3, and h4. These are initialized to certain fixed constants.

h0 =0x67452301

h1= 0Xefcdab89

h2=0x98BADCFE

h3=0x10325476

h4=0XC3D2E1F0 [3]

The main algorithm then uses each 512-bit message block in turn to modify the state. The processing of a message block consists of 80 similar operations based on a non-linear function F, modular addition and left rotation.

A=h0, B=h1, C=h2, D=h3, E=h4

From iteration 16 to 79

w[i]= (w[i-3] xor w[i-8] xor w[i-14] xor w[i-16]) leftrotate1

There are four possible functions F; a different one is used in each round:

F(B,C,D)=(B AND C) OR (NOT B AND D)

G(B,C,D)=B XOR C XOR D

H(B,C,D)=(B AND C) OR (B AND D) OR (C AND D) [5]

I(B,C,D)=B XOR C XOR D

SHA1 requires 80 processing constant words defined as:

K(t) = 0x5A827999 , (0 <= t <= 19)

K(t) = 0x6ED9EBA1, (20 <= t <= 39)

K(t) = 0x8F1BBCDC ,(40 <= t <= 59)

K(t) = 0xCA62C1D6, (60 <= t <= 79)

2. Why SHA Supersedes MD5?

MD5 generally processes smaller strings storing passwords, credit card numbers, or other sensitive data in databases system such as MySql.

Hash collisions: A hash collision occurs any time that two given inputs produce the same hash output.

MD5

- It is easy to produce collision in MD5.
- There are devastating collision attacks on MD5.
- These attacks mean that MD5 provides essentially no security against collisions.

SHA

- It is not easy to produce SHA-1 collisions.
- No collision for SHA-1 has been produced yet. SHA-256, which is much more "massive" (many more operations than SHA-1, yet with a similar structure), and is currently unbroken.
- MD5 cannot be implemented in existing technology at exceeding rates than the existing, i.e., at rates in excess of 256 Mbps in hardware, or 86 Mbps in software. While, SHA can be implemented in existing technology at exceeding rates than the existing.
- Speed: Assembly optimized versions of SHA-1 is consistently faster than MD5.
- MD5 is a proposed authentication option in IPv6, a protocol that should support existing networking technology, which is capable of 130 Mbps UDP.
- SHA1 appears to be much more secure. While there are some known attacks on SHA1, they are much less serious than the attacks on MD5.

These days, instead of using MD5 or SHA1, on which there are known attacks, there are probably even better modern hash functions, like SHA256. Those have no known attacks of any practical relevance.

Reverse hashing: In order to validate that data has not been altered, a new hash is generated with the received data and is matched with the original hash.It is not easily possible to generate a hash for an altered set of data that matches the hash of the original.

MD5:

- Unfortunately, MD5 is thoroughly compromised in this regard with there being multiple ways to relatively easily find alterations that can be put on the end or beginning of a payload to make it appear to be valid.

SHA:

- SHA-1 also has some minor compromises in this regard that were recently discovered, however they are less severe than the MD5 issues. Using something like SHA256 is even more secure as it does not currently have any known attacks against hash collision[7].

3. CONCLUSION AND FUTURE SCOPE

This paper proposes that the SHA algorithms should be given paramount importance in comparison to MD5 as SHA algorithms' performance is surpassing other cryptographic hash algorithm functions.

In the near future, new researches would be propounded proposing the same conclusion and more information would be amassed which could be used as a driving factor in the technological testing of the cryptographic hashing algorithms. This would result in the ultimate approved superiority of SHA algorithms above all cryptographic hash algorithms.

4. REFERENCES

[1] Andrew S. Tanenbaum, Pearson Publication, Fourth edition, Computer Networks.

[2] Professor Guevara Noubir, Fundamentals of Cryptography: Algorithms, and Security Services.

[3] Wikipedia, the free encyclopedia.

[4] Jerry li, MD5 Message Digest Algorithm. San Jose University, Computer Science department

[5] Ruth Betcher, Secure Hashing Algorithm.

[6] MD5 is faster than SHA-1.Journal Of Omnifarious-Myth.

[7] William Stalling7, Fourth Edition, Cryptography and Network Security (Various Hash Algorithms).

[8] http://stackoverflow.com/questions/2948156/algorithm-complexity-security-md5-or-sha1

[9] http://idrbtca.org.in/inf_crypto.htm

[10] Addam Schroll, ITNS and CERIAS CISSP Luncheon Series: Cryptography.

[11] http://science.opposingviews.com/advantages-disadvantages-symmetric-key-encryption-2609.html