Architecture based on MD5 and MD5-512 Bit Applications

Vandana Pandey
Research Scholar (Computer Science Engineering, B.T. Kumaon Institute of Technology, Dwarahat, Uttrakhand, India)

V.K Mishra
Assit. Professor (Computer Science Engineering, B.T. Kumaon Institute of Technology, Dwarahat, Uttrakhand, India)

ABSTRACT

Researchers have found many flaws in hash function such as MD5 algorithm. It has been one of the most widely used hash algorithm and it has been indicated that there is a security threat in the algorithm. Furthermore, a MD5 hash output may not offer sufficient protection in the near future. So in this paper, to provide higher security protection for MD5 and MD5-512 bit algorithm unified architecture is developed to make stronger algorithm. These two algorithm structures are quite similar but different in speed and security level. It follows that they can be merged together to give new design that can perform two hash function. Thus a incorporated design allows applications to switch from one algorithm to another based on different necessities. MD5-512 design accepts the same input format as that of MD5.

Keywords

MD5, Hash Function, MD5-512.

1. INTRODUCTION

A hash function is an algorithm that takes an random block of data and returns a fixed-size bit string, the (cryptographic) hash value, such that any (accidental or intentional) change to the data will (with very high probability) change the hash value. The data to be encoded are often called the "message," and the hash value is sometimes called the message digest or fingerprint Fig1. A hash function produce the fingerprint or hash code of a message that is used for verification purpose such as for information security applications, in digital signatures, message authentication codes (MACs), and other forms of authentication. In cryptography, a collision attack on a cryptographic hash tries to find two arbitrary inputs that will produce the same hash value known as a hash collisions. If x_1 and x_2 are two different messages, it is possible that $h(x_1) = h(x_2)$. This is called a collision. Collision resolution is the most important issue in hash implementations. To avoid the collision, the MD5 hashing algorithm was developed so that the collision probability becomes smaller.

A hash function is called secure if it fulfilled the following conditions are:

- It is impossible to find a message that corresponds to a given hash code. This is sometimes referred to as the one-way property of a hash function
- It is impossible to find two different messages that generate the same hash code value. This is also referred to as the strong collision resistance property of a hash function.
- It is impossible to alter a message without changing the hash.

The concept of MD5 is similar to the hash function. The MD5 algorithm takes an input of variable length and produce a message digest that is 128 bits long known as output. The message digest called the “hash” or “fingerprint” of the input. A 128-bit output found security hazard in the algorithm. Moreover, a 128-bit hash output may not offer a sufficient amount of security protection in the near future. To provide higher security protection, extend the hash output that accepts the same input format as that of MD5, and produces a 512-bit [2]. Since MD5 is still the most broadly used hash algorithm, improved the current implementation in the future to MD5-512 bit is much easier with a integrated structural design.

Motivated by the above examinations, in this paper we develop a unified architecture for MD5 and MD5-512. Comparison with other unified architectures indicates that the proposed architecture is area- efficient.

The content of Message-Digest is strongly related to the original data. Once the original data were altered, the content of Message-Digest integrity different to the old [5]. The output from the original data is unalterable, which means it can not be reverted to original data or can not be accomplished through reverse-computing. So it usually can not only be used for data reliability validation but also for content encryption [3]. Therefore the algorithm based on this principle can provide more robust protection about data. MD5 is used in many situations where a long message requires to be processed and compared rapidly i.e. mostly used in

creation and verification of digital signatures. MD5 was developed from MD, MD2, MD3 and MD4 [4]. It is a improved version in all MD's algorithms. The input message is divided into chunks of 512-bit blocks. And the process of MD5 contains the following steps: Preparing the input (Padding bits, appending data length), initializing MD5 buffer, processing message in 16-word block and output [1, 8].

Step1. Padding bits:

Pad the message with: first the '1'-bit, next as many '0' bits until the resulting bit length equals 448 mod 512, and finally the bit length of the original message as a 64-bit integer. The total bit length of the padded message is 512N for a positive integer N.

The padded message is partitioned into N successive 512-bit blocks M_1, M_2..... M_N.

Step2. The Initializing MD5 buffer:

A buffer (4 registers) is used to hold the intermediate and final result of the hash function that are each 32 bits long known as chaining variable. The buffers (A, B, C, and D) are defined as:

A= 08 ab 32 ef

B= 98 ba dc f4

C= fe dc ba 98

D= 76 54 32 10

Step3. Processing the blocks:

The contents of the four buffers (A, B, C and D) are now mixed with the words of the input, using the four auxiliary functions (F, G, H and I). There are four rounds, each involves 16 basic operations. The auxiliary function F is applied to the four buffers (A, B, C and D), using message word M_i and constant K_i. The item "<<<s" denotes a binary left shift by *s* bits. The four auxiliary functions that each take as input three 32-bit words and produce as output one 32-bit word. They apply the logical operators and, or, not and xor to the input bits.

F(X, Y, Z) = XY v not(X) Z

G(X, Y, Z) = XZ v Y not (Z)

H(X, Y, Z) = X xor Y xor Z

I(X, Y, Z) = Y xor (X v not (Z))

The bits of X, Y, and Z are autonomous and balanced the each bit of F(X, Y, Z) will be autonomous and balanced. The functions G, H and I are similar to the function F, in that they act in "bitwise parallel" to produce their output from the bits of X, Y, and Z, in such a manner that if the corresponding bits of X, Y, and Z are autonomous and balanced, then each bit of G(X,Y,Z), H(X,Y,Z), and I(X,Y,Z) will be autonomous and balanced.

Step5. Output:

After all rounds have been performed, the buffers A, B, C and D contain the MD5 digest of the original input.

2. CHARACTERSTICS OF MESSAGE DIGEST

- For any messages, the message digest always be the same and it should be very easy to calculate.
- It should be very difficult to find the original message by a given message digest because of one way encryption [5].
- MD5 encryption algorithm provide high security even if the encrypted data were leaked, data itself would not leak the true meaning [5].
- MD5 is one-way irreversible encryption mean once a data is encrypted, the value is generated and cannot be decrypted and reversed to the initial data [10].
- Comparing to other digest algorithms, MD5 is simple to implement, and provides a "fingerprint" or message digest of a message of arbitrary length.
- It performs very fast on 32-bit machine.
- For any two messages if we calculate their message digest then these two messages digest must be different.

3. REVIEW OF HASH ALGORITHM

Hash algorithms are important components in many cryptographic applications and security protocol suites. The MD5 Message-Digest Algorithm is a widely used cryptographic hash function that generates a 128-bit hash value., defined in RFC 1321[7], MD5 has been exploit in a wide variety of security applications, and is also commonly used to check data integrity. MD5 was designed by Ron Rivest in 1991[1] to substitute an earlier hash function, MD4 [9]. An MD5 hash is typically expressed as a hexadecimal number, 32 digits long. It is one in a sequence of message digest algorithms designed by Professor Ronald Rivest of MIT. When systematic work indicated that MD5's predecessor MD4 was likely to be insecure, It was designed to be a secure replacement. (Weaknesses were indeed later found in MD4 by Hans Dobbertin.) So the MD5 was developed. First the padding scheme is used in hash algorithm. Padding can be applied to input messages of any size and divides the input in blocks of 512 bits each. 64 bits are inserted at the end of the last block. Each block is divided into 16 words of 32 bits each. The algorithm uses a buffer that is made up of four words that are each 32 bits long. These words are called A, B, C and D known as chaining variable. Then hash algorithm process each data block iteratively. The input of the first data block is initial value; the input of the next data block is output of the previous data block. After all rounds have been performed, the output of the last data block is the MD5 hash value of the whole data [1].

4. MD5 (MESSAGE DIGEST ALGORITHM)

MD5 is a widely used cryptographic function with a 128- bit hash value. MD5 has been employed in a wide variety of security applications, and is also commonly used to check the integrity of files. An MD5 hash is typically expressed as a 32-digit hexadecimal number. MD5 algorithm is one of the most common Hash function [1, 6]. Its basic principle is to process the input information divided groups by 512 bits, and each group divided into 16 sub-groups with 32 bits. After a series of processing, the algorithm output composed by 4 groups with 32 bits, and cascade this 4 groups will generate a hash value with 128 bits. In the process of MD5 algorithm, fill information first to make its length 64 less than the multiple numbers of 512. The filling method is to attach a 1 and millions of 0, and add an information length before filling indicated by binary system with 64 bits. These two steps are to make the information length be the integer multiple of 512, and ensure the difference after different information filling Fig2.

PROPOSED IDEA

The basic principle of MD5-512[1, 7] is to process the input information divided groups by 512 bits, and each group divided into 4 sub-groups with 128 bits. After a series of processing, the algorithm output composed by 4 groups with 128 bits, and cascade after certain operations this 4 groups will generate a hash value with 512 bits. In the process of MD5 algorithm, fill information first to make its length128 less than the multiple numbers of 512. The filling method is to attach a 1 and millions of 0, and add an information length before filling indicated by binary system with 64 bits. These two steps are to make the information length be the integer multiple of 512, and ensure the difference after different information filling Fig3.

ADVANTAGE

By combining together to give new design that can perform two hash functions. This approach has the following advantages.

- Firstly, the integrated design is a source- efficient implementation when different hash algorithms are required in applications. Applications can switch to either algorithm based on different necessity.
- Second, since MD5 is still the most extensively-used hash algorithm, raised the current implementation in the future to MD5-512 is much easier with a integrated structural design.

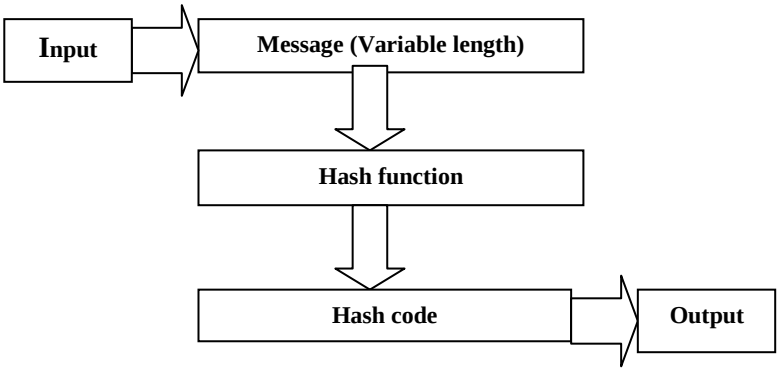

Figure1. Message Digest concept

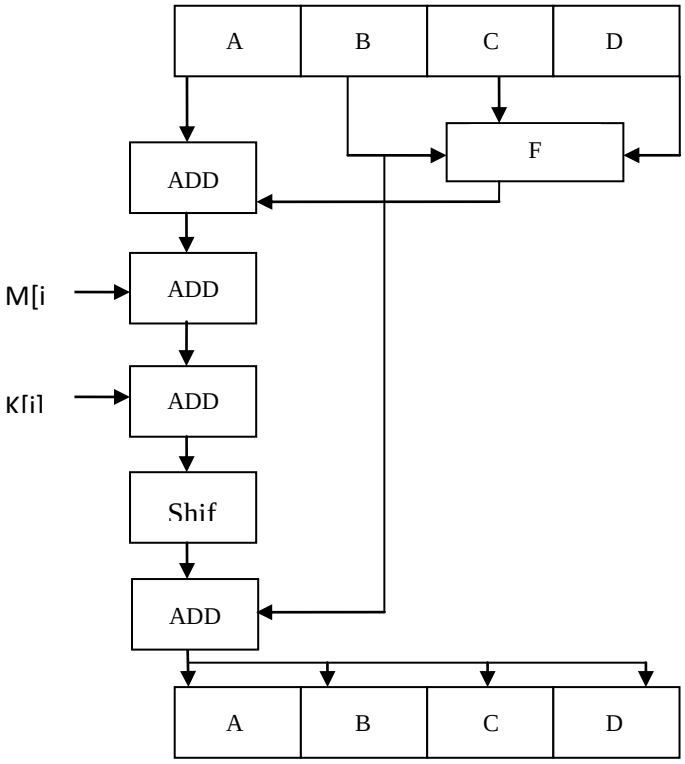

Figure2. MD5 operation (128 bit)

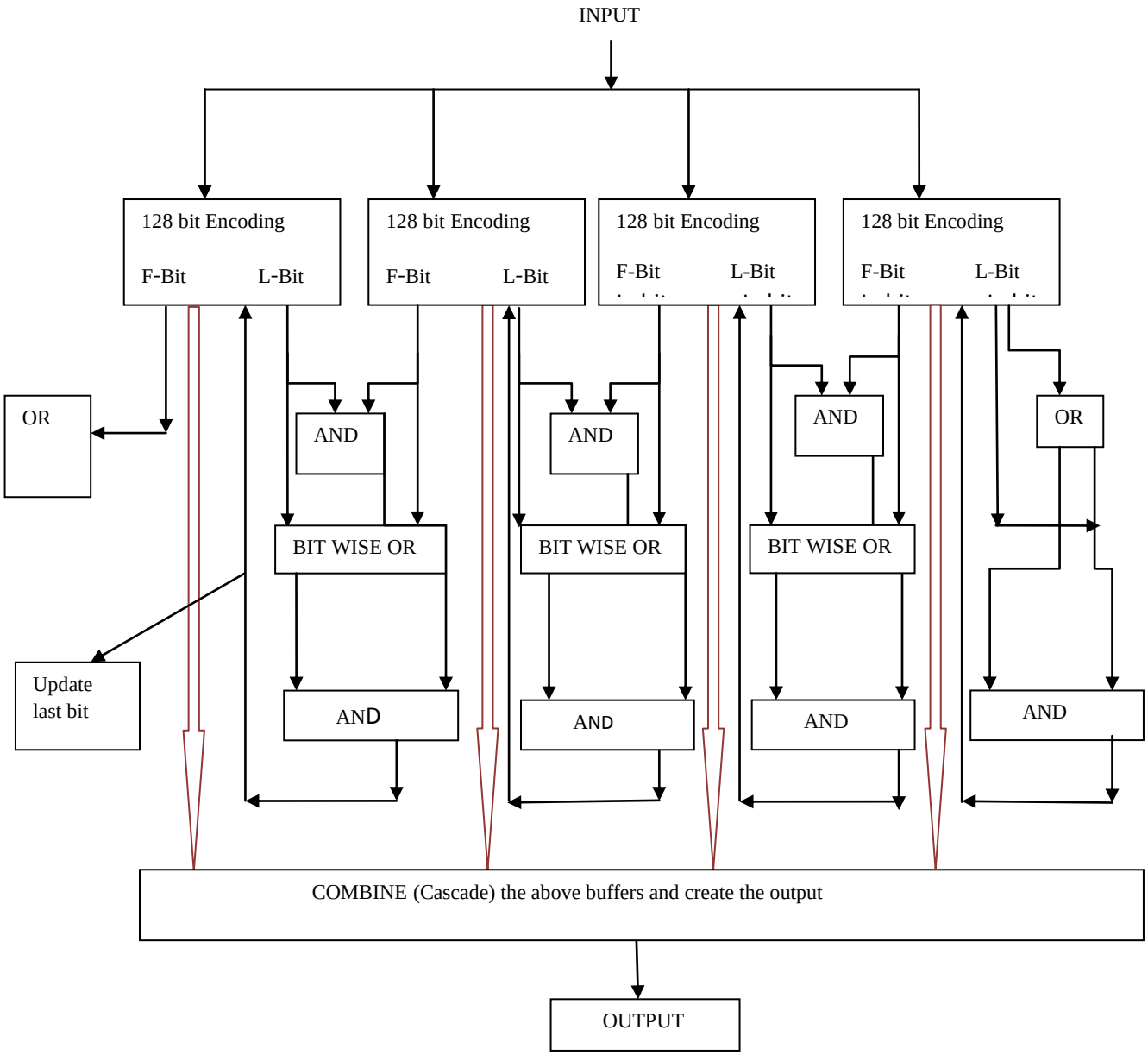

Figure3. MD5-512 bit operation

5. CONCLUSION

In this paper, a unified architecture for MD5 and MD5-512 bit hash algorithm is developed. These two algorithms are different in speed and security level. Here, MD5-512 provides higher security protection and accepts the same input format as that of MD5. It has been shown that the MD5-512 bit design is resource-efficient. Comparison with other unified architectures indicates that the proposed architecture is area-efficient. Here we have used MD5-128 bit algorithm as a basic element to design MD5-512 bit architecture.

6. REFERENCES

[1]. R.Rivest, "The MD5 message-digest algorithm," IETF RFC 1321, 1992.

[2] H. Dobbertin , A. Bosselaerand B. Preneel.1996. "RIPEMD-160: Astrengthened Version of RIPEMD, Fast Software Encryption," LNCS 1039,pp. 71-92, Springer-Verlag.

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

[4]. Zhao Yong-Xia, Zhen Ge. 2010. "MD5 Research,"IEEE, pp. 271-273.

[5]. Hancheng LIAO. 2008. "Image Retrieval Based on MD5". IEEE, pp. 987-991.

[6]. W.STALLINGS, Cryptography and Network Security, 2nd ed..New York: Prentice-Hall, 1997.

[7].Chiu-Wah Ng, Tung-Sang Ng and Kun-Wah Yip.2004. "A UNIFIED ARCHITECTURE OF MD5 AND RIPEMD-160 HASH ALGORITHMS". IEEE, pp.889-892.

[8].Wikipedia,"MD5",[EB/OL].http://en.wikipedia.org/wiki/MD5.

[9].R.L. Rivest. 1991 The MD4 message digest algorithm, Advances in Cryptology, Crypto'90, Springer-Verlag, pp. 303-311

[10]. Russell Impagliazzo, Leonid A. Levin, and Michael Luby.1989. Pseudo-random Generation from one-way functions (Extended Abstracts), STOC, ACM, pp. 12–24.