# Enhanced Parallel Hash Function Algorithm Based on 3C Construction (EPHFA-3C) 

Roayat Ismail Abdelfatah ${ }^{\text {a }}$, Esraa Abdelkhalek B. ${ }^{\text {b* }}$, M. E. Nasr ${ }^{\text {c }}$<br>${ }^{a, b, c}$ Electronics and Electrical Communications Engineering Dept., Faculty of Engineering, Tanta University, Tanta, Egypt<br>${ }^{a}$ Email: roayatismail2016@gmail.com<br>${ }^{b}$ Email: israa.baqa@f-eng.tanta.edu.eg<br>${ }^{c}$ Email: menasr2001@yahoo.com


#### Abstract

The hash function is a function that can convert data from variable size to fixed-size data that can be used in security of communication like, authentication, digital signature and integration. In this paper, a parallel, secure and fast hash function algorithm that is based on 3C construction is proposed. It is an enhancement for the MD construction. This enhancement makes the construction more resistant to the extension and multi-blocks attacks. The parallel structure of the algorithm improves the speed of hashing and reduces the number of operations. The simulation analysis such as hashes distribution, confusion and diffusion properties, and collision resistance are executed. Based on the results, our proposed hash algorithm is efficient, simple, and has strong security compared with some recent hash algorithms.


Keywords: Hash function; parallel hashing; 3C construction; collision resistance.

## 1. Introduction

As the significant growth of computer and Internet technology, multimedia communication plays an important role in many fields in our social community. Multimedia data security is becoming extremely significant in wired and wireless communications, such as file sharing, online payments, message authentication, and watermarking [1-5]. Hash algorithm is already proved to solve these problems accurately and efficiently. The hash function can be divided into two classes: 1) Unkeyed hash function: hash determines a single input parameter, message; and 2) Keyed hash function: has two different inputs, a message and a secret key [6-8].

[^0]To be an efficient cryptographic algorithm, the hash function needs to achieve following properties: 1) Preimage resistance (one-way): this means that a hash function would be difficult to reverse computationally. In other words, for any hash function H , generates a hash value D , it must be a hard process to find an input value X that hashes to D . This feature protects against the attacker who has a hash value only and trying to get the input. 2)Second pre-image resistance: this means; given an input and its corresponding hash, it would be difficult to get -another input with the same hash. In other words, for an input X and a hash function H that generates hash value $\mathrm{H}(\mathrm{X})$, it should be hard to find any different input value $Y$ such that $H(Y)=H(X)$. This feature of hash function protects against the attacker that has an input and its hash, and trying to substitute another value as valid value instead of the original input value. 3) Collision resistance: this means it should be difficult to find two distinct inputs of any length, which have the same hash. In other words, it is difficult to find two distinct inputs X and Y that achieve $\mathrm{H}(\mathrm{X})=\mathrm{H}(\mathrm{Y})$. This feature of collision free ensures that these collisions would be difficult to detect, and makes the hash very hard for the attacker to get two different input values with the same hash. In addition, if a hash function supposes the collision-resistant property, then it supposes second pre-image resistant [9]. The conventional hash functions such as MD5 [10] and SHA-1 [11] are based on logical operations, multi-round operations, and digital algebraic operations that significantly affect the security, as attacks on these algorithms have been found in. Multi-block collision attacks (MBCA) were discovered on the Merkle-Damgard (MD)-structure that traditional hash functions MD5, SHA-0 and SHA-1 based on it [1214]. SHA-2 hash algorithms family that announced in 2002 by NIST replaces the SHA-1 [10]. SHA-2 variants were analyzed and found to contain certain inefficiencies versus the attacks investigated in[16,17]. One of the reasons for this is that the popular structure of Merkle-Damgard was the basic framework behind the creation of these hash algorithms. The (MD) structure as shown in Fig. 1 takes the input message then divide it to $\mathrm{N}-1$ fixed-sized block of $m$ bits each, and padding the last block to $m$ bits. In addition, the last block contains the length of the hash function input, this makes the task of the attacker more difficult. In this case, the attacker should find two different messages of equal length that have the same hash or find two messages of different lengths, which, when their length added to the message, have the same hash value. The hash algorithm uses a repeated compression function f , which deals with two inputs; input from the previous stage of n -bit called chaining variables, and m-bit block, and generates output of $n$-bit. The chaining variable has an initial value at the beginning of hashing which considered as portion of the algorithm and the hash value is the final value of the variable chaining $[18,19]$ so the hash function can be defined as:
$C V_{0}=I V=$ initial value of $n-$ bit
$C V_{i}=f\left(C V_{i-1}, M_{i-1}\right) \quad 1 \leq i \leq N$
$\mathrm{H}(M)=C V_{N}$

Where a message m involving the blocks $M_{0}, M_{1}, \ldots, M_{N-1}$ is the input of the hash function. Note that there must be collisions for any hash function, because size of message at least similar to the block length $m$ is converted into a hash value of length $n$, where $m>n[19]$.


Figure 1: Merkle-Damgard iterated construction [18]

Hash function cryptographic analysis shows that the M-D hash structure is not resistant to fix point attack, multi-blocks attack, and extend attack; furthermore, there are some small impairment in compression function may cause failure in the hash algorithm. Hence, some enhanced structures such the generic 3C construction is very significant [18, 19]. The generic 3C hash function structure as appeared in Fig 2 formed of two compression functions: the function $f$, which repeated in the series and the function $f^{\prime}$ that iterated in the accumulation series. The function $f^{\prime}$ is a common function that can be similar to the compression function $f$ applied in the cascade chain.


Figure 2: The generic 3C hash function construction [19]

Firstly, based on the M-D construction the input message is handled in repeated way in the compression function $f$. Then padding the output of this function using the usual padding method Z-PAD (attaching a bit 1 and some 0 's then add the encoded binary of z-length representation) to make the length of the last block equal to the block length of $m$ bits. Finally, the cumulative of the function $f^{\prime}$ output is also padded and entered to the external application $f$ as input. Note that, the padding function is applied twice for the 3 C structure: firstly, in the cascaded structure and secondly, on Z for the block of accumulative chain. This padding is indicated as ZPAD operation in Fig 2. The structure is called 3C construction because at least three compression function applications are needed to handle the message when it has only one block. The single block is processed by first application the, the second deal with the padded block and the last compression function application deal with
the accumulative chain function block [19].

The main contribution in this paper is to enhance 3 C construction from two sides:

- Information of message length is attached to the padded message so the proposed scheme is resistant to length-extension attack and meet-in-the-middle attack.
- The standard MD sequential iteration structure is changed to parallel iteration structure. In this case, the enhanced hash algorithm will have more advantage in term of speed when dealing with large files.
- New design for step function that is simple and easy to computation so the proposed algorithm has great advantage in speed.

After this introduction, the remaining of the paper is arranged as follow: Sect. II illustrates our proposed hash algorithm, Sect. III presents implementation analysis, Sect. IV covers the comparative analysis, and finally Sect. V includes the conclusion.

## 2. Proposed algorithm

In this section, we briefly illustrate the five phases of the proposed hash algorithm:

### 2.1. Message padding

Assuming that the message size is $l_{1}$, and the message block size is $m=512$ bits. First, padding the input message M as follow: a bit of " 1 " is added to the end of the message, then, attaching some of " 0 " after the " 1 " bit that makes the padding message size satisfies 448 modulo 512 . Finally, $l_{1}$ is converted to binary representation and inserted at the ending of the message. The message block after padding is shown in Fig. 3.


Figure 3: Message padding

### 2.2. Message blocking

After padding the message, the input message M is divided into N blocks each of 512 -bit with $M=$ $M_{0}\left\|M_{1}\right\| M_{2}\|\ldots.\| M_{N-1}$. During message block partition processing, every message block is divided into sixteen 32-bit messages with $M_{i}=M_{i}^{0}\left\|M_{i}^{1}\right\| M_{i}^{2}\|\ldots\| M_{i}^{15}$.

### 2.3. Initialization of chaining variables

We define eight chaining variables with $S, T, U, V, W, X, Y$, and $Z$. The length of each chaining variable is 32 bits. The initial value of the eight chaining variables is listed as.
$S_{0}=6 \mathrm{a} 09 \mathrm{e} 667$
$T_{0}=$ bb67ae 85
$U_{0}=3 \mathrm{c} 6 \mathrm{ef} 372$
$V_{0}=\mathrm{a} 54 \mathrm{ff} 53 \mathrm{a}$
$W_{0}=510 \mathrm{e} 527 \mathrm{f}$
$X_{0}=9 \mathrm{~b} 05688 \mathrm{c}$
$Y_{0}=1 \mathrm{f} 83 \mathrm{~d} 9 \mathrm{ab}$
$Z_{0}=5 b e 0 c d 19$

These are defined by having the fractional portion of square roots for the first eight primes in hexagonal representation.

### 2.4. Parallel iterative structure design

The basic structure of hash algorithm is the iteration structure. As illustrated in Fig. 4, the parallel iteration structure is consisted of two procedures: message preprocessing (illustrated in Sect. II-a), compression function (will be illustrated in Sec. II-e).


Figure 4: Parallel iterative structure

### 2.5. Compression Function

Compression function of our proposed algorithm is consists of three-compression functions: two of the parallel compression functions $f$, which iterated in the cascade series and the function $f^{\prime}$ that is iterated in the accumulative series. For the two parallel compression functions, the input of the first function is the first message block and the next message block is the input of second compression function and hence. The function $f$ has two parallel branch function; Branch1 and Branch2 as shown in Fig. 5 therefore the attacker who attemps to fracture the function must target the two branches simultaneously. The two branches message words ordering are:

Branch1: $M_{i}=M_{i}^{0}\left\|M_{i}^{1}\right\| M_{i}^{2}\|\ldots\| M_{i}^{15}$

Branch2: $M_{i}=M_{i}^{15}\left\|M_{i}^{14}\right\| M_{i}^{13}\|\ldots\| M_{i}^{0}$, where the order is reversed.


Figure 5: Compression function (f)

### 2.5.1. Constants

For compression function, there are sixteen constants will be defined to use. For Branch1 the constants are ordered as following:
$\beta_{0}=428 \mathrm{a} 2 \mathrm{f} 98$

$$
\beta_{2}=\mathrm{b} 5 \mathrm{c} 0 \mathrm{fbcf}
$$

$\beta_{3}=\mathrm{e} 9 \mathrm{~b} 5 \mathrm{dba} 5$

$$
\beta_{1}=71374491
$$

$\beta_{4}=3956 c 25 b$
$\beta_{5}=59 \mathrm{f} 111 \mathrm{f} 1$
$\beta_{6}=923 f 82 \mathrm{a} 4$
$\beta_{7}=\mathrm{ab} 1 \mathrm{c} 5 \mathrm{ed} 5$
$\beta_{8}=\mathrm{d} 807 \mathrm{aa} 98$
$\beta_{9}=12835 \mathrm{~b} 01$
$\beta_{10}=243185$ be
$\beta_{11}=550 \mathrm{c} 7 \mathrm{dc} 3$
$\beta_{12}=72 \mathrm{be} 5 \mathrm{~d} 74$
$\beta_{13}=80 \mathrm{deb} 1 \mathrm{fe}$
$\beta_{14}=9 \mathrm{bdc} 08 \mathrm{a} 7$
$\beta_{15}=\mathrm{c} 19 \mathrm{bf} 174$

The order of constants are reversed for Branch2.

The aim of using these constants is to upset the attacker that attempt to get the best differential characteristics with high relative possibility. Therefore, we choose the constants which form the first thirty-two bits for the fractional portions of the cube roots for the first sixteen four prime numbers.

### 2.5.2. Step function

The compression function f is iterated four times. For $k^{t h}$ step, the input registers is divided into eight 32 -bit words : $S_{k}, T_{k}, U_{k}, V_{k}, W_{k}, X_{k}, Y_{k}$, and $Z_{k}$. The inputs of $k+1$ step calculated as:
$S_{k+1}=\left[\left[\left(Y_{k} \oplus Z_{k}\right) \oplus M_{2 k+9}\right]^{\lll r_{7}}\right] \oplus Z_{k}^{\lll r_{8}}$
$T_{k+1}=\left[\left(S_{k} \oplus T_{k}\right) \oplus M_{2 k}\right]+\beta_{4 k}$
$U_{k+1}=\left[\left[\left(S_{k} \oplus T_{k}\right) \oplus M_{2 k}\right]^{\lll r_{1}}\right] \oplus T_{4 k+1}^{\ll r_{2}}$
$V_{k+1}=\left[\left(U_{k} \oplus V_{k}\right) \oplus M_{2 k+1}\right]+\beta_{4 k+1}$
$W_{k+1}=\left[\left[\left(U_{k} \oplus V_{k}\right) \oplus M_{2 k+1}\right]^{\lll r_{3}}\right] \oplus V_{k}^{\lll r_{4}}$
$X_{k+1}=\left[\left(W_{k} \oplus X_{k}\right) \oplus M_{2 k+8}\right]+\beta_{4 k+2}$
$Y_{k+1}=\left[\left[\left(W_{k} \oplus X_{k}\right) \oplus M_{2 k+8}\right]^{\lll r_{5}}\right] \oplus X_{k}^{\lll r_{6}}$
$Z_{k+1}=\left[\left(Y_{k} \oplus Z_{k}\right) \oplus M_{2 k+9}\right]+\beta_{4 k+3}$


Figure 6: Step function structure

In Fig. $6,(\lll x)$ donates $x$ bits left shift rotation, $\oplus$ is XOR logic function, and $(+)$ is addition mod $2^{32}$.

### 2.5.3. Shift rotation variables

For $k=0$, the initial values are defined as:

| $r_{1}=5$ | $r_{2}=9$ | $r_{3}=3$ | $r_{4}=11$ |
| :--- | :--- | :--- | :--- |
| $r_{5}=8$ | $r_{6}=13$ | $r_{7}=7$ | $r_{8}=10$. |

For $k=1,2,3$ the shift rotations values calculated as:
$r_{1}=S_{k}+T_{k}+U_{k}$
$r_{2}=T_{k}+U_{k}+V_{k}$
$r_{3}=U_{k}+V_{k}+W_{k}$
$r_{4}=V_{k}+W_{k}+X_{k}$
$r_{5}=W_{k}+X_{k}+Y_{k}$
$r_{6}=X_{k}+Y_{k}+Z_{k}$
$r_{7}=Y_{k}+Z_{k}+S_{k}$
$r_{8}=Z_{k}+S_{k}+T_{k}$.

## 3. Experimental analysis and simulation results

In this section, We estimate our parallel hash function in forms of hash values uniform distribution, hash value sensitivity to delicate changes in the original message, properties of confusion and diffusion, and collision tests.

### 3.1. Hash sensitivity

We randomly choose a text and the simulation of sensitivity is done under nine conditions:

Condition1: The original message "A hash function is any function that can be used to map data of arbitrary size to fixed-size values. These values called digests. ".

Condition2: The first character changed to lowercase.

Condition3: Add a number ' 1 ' before the sentence.

Condition4: Change the word 'arbitrary' to 'variable'.

Condition5: Remove dash between 'fixed-size'.

Condition6: Change the dot after the word 'values' to comma.

Condition7: Change the first letter of the word 'digests' to capital letter.

Condition8: Change the last character after the word 'digests' from '.' to '!'.

Condition9: Add a whitespace before the last character ' $\because$ '.

The corresponding 256-bit hash values in hex format are the following:

Condition1: 4B1633D0FF7BB26695ADA088D65F5E9658A37883AF16A746B41E5FBED6377001

Condition2: EF5BD59DFA138F9A009D8A83A17F2F466D8BAA5EA0825E01CD60D828903D7AB2

Condition3: E43DC0208B12E4F7762E242DC92523A9F3A70934B3303FDEA685CD8C656A3A03

Condition4: 8E328BE1A2BD0D8EA43445260B1F7347C7F9406BBEA7FD80D88CD279AA8CEC74

The corresponding hash values graphic in binary format is illustrated in Fig. 7:


Figure 7: Hash values in binary format under nine condition

The hex codes of the hashes and their binary representations graphics show that even a single bit of modification in the original message results in disastrous variation in the hash code. Based on these graphics, our proposed hash satisfies the one-way cryptographic hash feature.

### 3.2. Confusion and diffusion statistical analysis

Confusion and diffusion are identified by Claude Shannon as the main features of cipher security that are considered as the main design requirements of any cryptographic hash algorithm [20]. The feature of confusion indicates to the relationship between the message and its hash value should be unpredictable and complex; whereas diffusion indicates to the hash value should be extremely based on the message. The following test is implemented in order to catch the qualitative features of the diffusion and confusion for our proposed hash: First, select a message randomly and its corresponding hash value is produced; then change a bit from the message randomly and produce the new hash value. Finally, compare the two hash values and count the number of various bits in the two hash values that placed at the same location. Generally, statistical analysis is based on the following equations:

Number of average bit change:
$B_{c}=\frac{1}{N} \sum_{i=1}^{N} B_{i}$

Standard deviation of bit change:
$\Delta B_{c}=\sqrt{\frac{1}{N-1} \sum_{i=1}^{N}\left(B_{i}-B_{c}\right)^{2}}$

Percentage of bit change:
$P=\frac{B_{c}}{N} \times 100 \%$

Standard deviation of $P$ :
$\Delta P=\sqrt{\frac{1}{N-1} \sum_{i=1}^{N}\left(\frac{B_{i}}{n}-P\right)^{2}} \times 100 \%$

The test is applied on a sample size of 1024 and 2048 bits and the results are tabulated in Table 1.

Table 1: Statistical results of changed bits for $\mathrm{n}=1024$ and 2048 bits

| n | $\mathrm{n}=1024$ | $\mathrm{n}=2048$ | Mean |
| :--- | :--- | :--- | :--- |
| $B_{\min }$ | 122 | 118 | 120 |
| $B_{\max }$ | 136 | 144 | 140 |
| $\overline{B_{c}}$ | 128.25 | 134.5 | 131.375 |
| $P(\%)$ | $50.0977 \%$ | $52.5391 \%$ | $51.3184 \%$ |
| $\Delta B_{c}$ | 5.06388 | 7.91021 | 6.847 |
| $\Delta P(\%)$ | $1.97808 \%$ | $3.08993 \%$ | $2.534 \%$ |

The test is performed on our proposed scheme N times, where $\mathrm{N}=256,512,1024$ and 2048. The test messages are 2048 bits in the length and the result is listed in Table 2 for 256-bit hashes.

Table 2: Statistical results of changed bits for $\mathrm{N}=256,512,1024$ and 2048 times

| N | $\mathrm{N}=256$ | $\mathrm{~N}=512$ | $\mathrm{~N}=1024$ | $\mathrm{~N}=2048$ | Mean |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $B_{\min }$ | 110 | 106 | 106 | 106 | 107 |
| $B_{\max }$ | 148 | 148 | 150 | 150 | 149 |
| $\overline{B_{c}}$ | 128.203 | 128.129 | 127.975 | 127.899 | 128.0515 |
| $P(\%)$ | $50.0793 \%$ | $50.0504 \%$ | $49.9901 \%$ | $49.9607 \%$ | $50.0201 \%$ |
| $\Delta B$ | 7.90469 | 7.8881 | 8.02607 | 7.9303 | 7.8235 |
| $\Delta P(\%)$ | $3.08777 \%$ | $3.08129 \%$ | $3.13518 \%$ | $3.09778 \%$ | $3.0407 \%$ |

As shown in Table 2, the mean changed bit number for the proposed hash algorithm is $\overline{B_{c}}=128.0515$ and the mean changed probability $P=50.02 \%$ these results of our scheme are around to the ideal values of 128 bits and $50 \%$, respectively. Also, the values of standard deviation $\Delta B$ and $\Delta P$ are very small that indicates to a high capability for confusion and diffusion. Fig. 8 illustrate the behavior of distribution of changed bit number as (a) the $B_{i}$ graphic indicates to its value is equally distributed, and as illustrated in Fig. 8(b), the normal distribution of $B_{i}$ centered at the ideal value of 128 . The proposed hash function results shows that it has close-ideal statistical characteristics in form of confusion and diffusion capability, where even one bit change from the plaintext will result in a totally distinct message digest.


Figure 8: Changed bit number spreading: a) plot of $\boldsymbol{B}_{\boldsymbol{i}}$ and b) histogram of $\boldsymbol{B}_{\boldsymbol{i}}$

### 3.3. Collision resistance analysis

When two distinct input messages are mapped to precisely same hash value a collision occurs, so the aim of the collision attack is to seek to find two distinct messages that result in collision. One of the most important features of efficient encryption algorithms is collision resistance. In order to prove the highly collision resistant of our proposed scheme and its produced hashes, a test is performed as follow: First, produce a message randomly along with its corresponding hash value and saved in ASCII format. Then, choose a bit from the original message and replace it so a new message is generated with a small different. Next, hash the new generated message and store the corresponding digest in ASCII format. Finally, compare the two message hashes and count the number of ASCII characters that located in the same places and have the same values (number of hits). In this paper, the test is performed $\mathrm{N}=2048$ times, and plotted the distribution number of hits in Fig. 9. As shown in Fig. 9 and Table 3, no hit happens in 1820 tests, one hit occurs in 214 tests, and 12 tests hit twice. As the maximum number of hits is only 2, the collision of the proposed algorithm is very small.

Table 3: Number of hits in Collision test for $\mathrm{N}=2048$

| Number of equal characters | 0 | 1 | 2 | 3 | 4 | 5 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| The proposed algorithm | 1820 | 214 | 12 | 0 | 0 | 0 |



Figure 9: Distribution of number of hits for $\mathrm{N}=2048$

In addition, the absolute difference between two distinct hashes is also defined as:
$d=\sum_{i=1}^{N}\left|t\left(e_{i}\right)-t\left(e_{i}^{\prime}\right)\right|$

Where $e_{i}$ indicate to the $i^{\text {th }}$ character of ASCII of the original hash value while $e_{i}^{\prime}$ donate the $i^{\text {th }}$ ASCII character of the modified hash value, and the function $t$ () maps the results into the corresponding decimal values. The collision test is performed $\mathrm{N}=2048$ times, and the results is listed in Table 4. The values of maximum, minimum, and mean of the absolute difference d of two different hash values are 4032, 1726, and 2729.9, respectively and the mean/character of absolute difference d of two hash values for our scheme is 85.3094 that is very near to the ideal theoretical mean/character value 85.3333 .

Table 4: Absolute differences d between two hash values

| Maximum | Minimum | Mean | Mean <br> Character |
| :--- | :--- | :--- | :--- | :--- |
| 4032 | 1726 | 2729.9 | 85.31 |

## 4. Comparative Analysis

### 4.1. Statistical Performance Comparison

Comparison between the proposed algorithm and some relevant and important hash algorithms is done based on
security evaluation. The statistical results of all chosen hash algorithms are reported in Table 5 a). As listed in Table 5 a), the values of $\bar{B}$ and P of the proposed scheme are near-ideal, and standard deviation of $\Delta B$ is smaller than traditional hash algorithms of SHA256, Keccak-256 and other recent hash function of PLHF [18]. The proposed algorithm also has smaller standard deviation of $\Delta P$ than traditional algorithms of SHA1, SHA2, Keccak-256, and other recent algorithm of PLHF [21]. The proposed algorithm also mapped to variable output length of 128 -bit and the statistical performance is compared with some recent hash algorithm such as Li-Ge [22], and Je and his colleagues [23]. The results are shown in Table 5 b) and it is very close to the ideal result.

Table 5: Comparison on statistical performance a) Hash-256 b) Hash-128
a) Hash-256 bit

| Algorithm | Statistical performance of the algorithms |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :---: |
|  | $\boldsymbol{B}_{\boldsymbol{c}}$ | $\overline{\boldsymbol{B}}$ | $\boldsymbol{P}(\%)$ | $\Delta \boldsymbol{B}$ | $\Delta \boldsymbol{P}(\%)$ |  |
| SHA1 | $49-114$ | 80.16 | 50.10 | 6.24 | 3.90 |  |
| SHA256 | $71-193$ | 128.61 | 50.24 | 10.83 | 4.23 |  |
| Keccak-256 | $80-207$ | 129.34 | 50.52 | 10.21 | 3.99 |  |
| PLHF [21] | $91-168$ | 127.91 | 49.96 | 8.93 | 3.49 |  |
| Proposed | $106-150$ | 128.05 | 50.02 | 7.82 | 3.04 |  |

b) Hash-128 bit

| Algorithm | Statistical performance of the algorithms |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :---: |
|  | $\boldsymbol{B}_{\boldsymbol{c}}$ | $\overline{\boldsymbol{B}}$ | $\boldsymbol{P}(\%)$ | $\Delta \boldsymbol{B}$ | $\Delta \boldsymbol{P}(\%)$ |  |
| MD5 | N/A | 64.03 | 50.02 | 5.66 | 4.42 |  |
| Li-Ge [22] | $42-85$ | 63.87 | 49.90 | 5.58 | 4.36 |  |
| Je and his | $45-86$ | 63.99 | 49.99 | 5.64 | 4.38 |  |
| colleagues [23] |  |  | 49.88 | 5.67 | 4.43 |  |

### 4.2. Collision Resistance Comparison

Collision resistance test of this algorithm and other hash functions is running 2000 times, and the results are listed in Table 6 as a distance between two hash values. Also, the distances between two different hash values are calculated as $D_{\text {hash }}=\sum_{i=1}^{80} \mid$ hex $\left(C_{1}\right)$-hex $\left(C_{2}\right) \mid$, where $C_{1}, C_{2}$ are the two-hexadecimal characters that located at the same location in two output hash values. The maximum and the minimum $D_{\text {hash }}$ can be calculated using the same random message, and the average $D_{\text {hash }}$ can be obtained by:
$D_{\text {average }}(a v g)=\frac{\sum_{i=1}^{2000} D_{\text {hash }_{i}}}{2000}$

Table 6: Distances between two hash values in different schemes

| Algorithm | Size of output | $\boldsymbol{N}_{\boldsymbol{s}}(\max )$ | $\boldsymbol{D}_{\text {hash }}(\max )$ | $\boldsymbol{D}_{\text {hash }}(\mathrm{min})$ | $\boldsymbol{D}_{\boldsymbol{a v g}}(\mathrm{avg})$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| SHA1 | 160 | 2 | 2556 | 857 | 85.4 |
| SHA256 | 256 | 2 | 4098 | 1365 | 85.375 |
| Keccak-256 | 256 | 2 | 1405 | 1368 | 85.406 |
| PLHF [21] | 256 | 1 | 4096 | 1364 | 85.344 |
| Proposed | 256 | 2 | 4032 | 1726 | 85.309 |

As shown in Table 6, the fluctuation of PLHF is small, which means it is more stable and has stronger resistance against random collision attack than other algorithms. Its security parameter, $D_{\text {average }}(a v g)$, is approximately equal to the optimal value 85.33 which is superior to other parallel schemes such SHA1, SHA256, Keccak-256 and PLHF.

### 4.3. Speed Analysis Comparison

Speed analysis is conducted by comparing the number of operations between this scheme and traditional hash function SHA-256 for one block of 512-bits. The comparison of total number of operations is listed in Table 7.

Table 7: Number of operation comparison

| Operation | SHA-256 | PLHF [21] | Proposed |
| :--- | :--- | :--- | :--- |
| Addition (+) | 600 | 560 | 80 |
| Bitwise Operation $(\oplus, \wedge, \vee)$ | 1024 | 0 | 243 |
| Multiplication | 0 | 320 | 0 |
| Shift operation $(\ll, \gg,\langle\ll, \ggg)$ | 672 | 280 | 160 |
| Total | 2296 | 1160 | 483 |

As illustrate in Table 7, the number of operation of SHA-256 [19] is approximately five times of the number of operation of the proposed hash also, the proposed hash has less than half number of operation of PLHF[21]. So, These result show how the proposed hash is much fast.

## 5. Conclusion

This paper propose an efficient 256-bit cryptographic hash function algorithm that based on enhanced generic 3 C hash function structure. The algorithm is achieved by adjusting the M-D iterative structure to be more robust against the extension attacks and differential multi-blocks attacks. Further, parallelization is implemented in this paper to reduce the number of operations and hence improve the speed of hashing algorithm. Based on
experiments and security analysis, the proposed hash function achieves the security requirements and has other advantages such as high resistance to collision attacks and great statistical diffusion and confusion performance compared with conventional schemes.

## References

[1] National Bureau of Standards (1999) Data encryption Li, Yantao, and Guangfu Ge. "Cryptographic and parallel hash function based on cross coupled map lattices suitable for multimedia communication security." Multimedia Tools and Applications 78.13 (2019), pp 17973-17994.
[2] Deng S., Xiao D., Li, Y., \& Peng, W. "A novel combined cryptographic and hash algorithm based on chaotic control character." Communications in Nonlinear Science and Numerical Simulation 14.11 (2009), pp 3889-3900.
[3] Deng, Shaojiang, et al. "Analysis and improvement of a hash-based image encryption algorithm." Communications in Nonlinear Science and Numerical Simulation 16.8 (2011), pp 32693278.
[4] Schneider, Marc, and Shih-Fu Chang. "A robust content based digital signature for image authentication." Proceedings of third IEEE International Conference on Image Processing. Vol. 3. IEEE (1996), pp 227-230.
[5] Mihçak, M. Kıvanç, Ramarathnam Venkatesan, and Tie Liu. "Watermarking via optimization algorithms for quantizing randomized semi-global image statistics." Multimedia Systems 11.2 (2005), pp 185-200.
[6] Li, Yantao. "Collision analysis and improvement of a hash function based on chaotic tent map." Optik 127.10 (2016), pp 4484-4489.
[7] Shi-Hong, Wang, and Shan Peng-Yang. "Security analysis of a one-way hash function based on spatiotemporal chaos." Chinese Physics B 20.9 (2011), pp 090504-090507.
[8] Wang, Shihong, Da Li, and Hu Zhou. "Collision analysis of a chaos-based hash function with both modification detection and localization capability." Communications in Nonlinear Science and Numerical Simulation 17.2 (2012), pp 780-784.
[9] Ahmad, Musheer, et al. "A simple secure hash function scheme using multiple chaotic maps." 3D Research 8.2 (2017), pp 13.
[10] Rivest, Ron. "The md5 message-digest algorithm (rfc 1321)." Internet Activities Board (1992).
[11] Standard, Secure Hash. "FIPS Pub 180-1." National Institute of Standards and Technology 17 (1995): 15.
[12] Mendel, Florian, Tomislav Nad, and Martin Schläffer. "Improving local collisions: new attacks on reduced SHA-256." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg (2013), pp 262-278.
[13] Stevens, Marc. "New collision attacks on SHA-1 based on optimal joint local-collision analysis." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg (2013), pp 245-261.
[14] Wang, Xiaoyun, Yiqun Lisa Yin, and Hongbo Yu. "Finding collisions in the full SHA-1." Annual international cryptology conference. Springer, Berlin, Heidelberg, (2005), pp 17-36.
[15] FIPS, NIST. 180-2, Secure Hash Standard, Federal Information Processing Standard (FIPS). publication 180-2. Technical report, Department of Commerce, 2002.
[16] Sanadhya, Somitra Kumar, and Palash Sarkar. "New collision attacks against up to 24 -step SHA2." International conference on cryptology in India. Springer, Berlin, Heidelberg (2008), pp 91-103.
[17] Khovratovich, Dmitry, Christian Rechberger, and Alexandra Savelieva. "Bicliques for preimages: attacks on Skein-512 and the SHA-2 family." International Workshop on Fast Software Encryption. Springer, Berlin, Heidelberg (2012), pp 244-263.
[18] Gauravaram, Praveen, Millan, W., Dawson, E., \& Viswanathan. "Constructing secure hash functions by enhancing Merkle-Damgård construction." Australasian Conference on Information Security and Privacy. Springer, Berlin, Heidelberg (2006), pp 407-420.
[19] Elkamchouchi, Hassan M., Mohamed E. Nasr, and Roayat Ismail Abdelfatah. "A new secure and fast hashing algorithm (SFHA-256)." 2008 National Radio Science Conference. IEEE (2008), pp 1-8.
[20] Selent, Douglas. "Advanced encryption standard." Rivier Academic Journal 6.2 (2010), pp 1-14.
[21] Yang, Yijun, Chen, F., Sun, Z., Wang, S., Li, J., Chen, J., \& Ming, Z. "Secure and efficient parallel hash function construction and its application on cloud audit." Soft Computing 23.18 (2019), pp 89078925.
[22] Li, Yantao, and Guangfu Ge. "Cryptographic and parallel hash function based on cross coupled map lattices suitable for multimedia communication security." Multimedia Tools and Applications 78.13 (2019), pp 17973-17994.
[23] Teh, Je Sen, Azman Samsudin, and Amir Akhavan. "Parallel chaotic hash function based on the shuffle-exchange network." Nonlinear Dynamics 81.3 (2015), pp 1067-1079.


[^0]:    * Corresponding author.

