A Lightweight Hash Function Resisting Birthday Attack and Meet-in-the-middle Attack

Shenghui Su¹, Tao Xie², and Shuwang Lü³

¹College of Computers, Beijing University of Technology, Beijing 100124
²School of Computers, National University of Defense Technology, Changsha 410073
³Graduate School, Chinese Academy of Sciences, Beijing 100039

Abstract: To be paired with a lightweight digital signing scheme of which the modulus length is between 80 and 160 bits, a new non-Merkle-Damgård structure (non-MDS) hash function is proposed by the authors based on a multivariate permutation problem (MPP) and an anomalous subset product problem (ASPP) to which no subexponential time solutions are found so far. It includes an initialization algorithm and a compression algorithm, and converts a short message of \( n \) bits treated as only a block into a digest of \( m \) bits, where \( 80 \leq m \leq 232 \) and \( 80 \leq m \leq n \leq 4096 \). Analysis shows that the new hash is one-way, weakly collision-free, and strongly collision-free along with a proof, and its security against existing attacks such as birthday attack and meet-in-the-middle attack gets the \( O(2^m) \) magnitude. Running time of its compression algorithm is analyzed to be \( O(nm^2) \) bit operations. A comparison with the Chaum-Heijst-Pfitzmann hash based on a discrete logarithm problem is made. Especially, the new hash with short input and small computation may be used to reform a classical hash with an \( m \)-bit output and an \( O(2^{m/2}) \) magnitude security into a compact hash with an \( m/2 \)-bit output and the same security. Thus, it opens a door to convenience for utilization of lightweight digital signing schemes.

Keywords: Hash function; Compression algorithm; Merkle-Damgård structure; Provable security; Birthday attack; Meet-in-the-middle attack

1 Introduction

In recent years, the ECC-160 digital signing scheme, an analogue of the ElGamal digital signing scheme based on a discrete logarithm problem (DLP) in an elliptic curve group over a finite field [1][2], and some lightweight digital signing schemes — the optimized version of the REESSE1+ digital signing scheme [3] for example have been utilized for RF ID (Radio Frequency Identity) tags or non-RF ID tags [4][5][6].

While a RF ID tag contains an IC chip which is used to store signatures and other data, an non-RF ID tag contains no IC chip because a short signature from a lightweight or ultra-lightweight signing scheme may be symbolized in short length, and printed directly on a papery tag or label. Now, such tags are applied to the identification, authentication, or anti-forgery of financial-notes, certificates, diplomas, and commodities, particularly including food and drug.

It is well understood that we first need to extract the digest of a message by employing a hash function before signing the message. Traditionally, a hash function consists of a compression function and the Merkle-Damgård structure (MDS) [7][8]. Let \( \hat{h} \) be a hash function, and generally, it has the following four properties [9][10]:

1. given a message \( \mu \), it is very easy to calculate the message digest \( \sigma = \hat{h}(\mu) \), where \( \sigma \) is also called a hash output;
2. given a digest \( \sigma \), it is very hard to calculate the message \( \mu \) according to \( \sigma = \hat{h}(\mu) \), namely \( \hat{h} \) is one-way;
3. given any arbitrary message \( \mu \), it is computationally infeasible to find another message \( \mu' \) such that \( \hat{h}(\mu) = \hat{h}(\mu') \), namely \( \hat{h} \) is weakly collision-free;
4. it is computationally infeasible to find two arbitrary messages \( \mu \neq \mu' \) such that \( \hat{h}(\mu) = \hat{h}(\mu') \), namely \( \hat{h} \) is strongly collision-free.

The word “infeasible” means that some problem cannot be solved at least in polynomial time. Sometimes, 4 is optional with some users of a hash function because 1, 2, and 3 are enough for most of applications of the users.

At present, SHA-1, SHA-256, and SHA-384 announced by NIST are among the hash functions that are believed to be secure [9][11] though they each cannot resist birthday attack of which the time

* This work is supported by MOST with Project 2007CB311100 and 2009AA01Z441. Corresponding email: reesse@126.com.
complexity is $O(2^{m/2})$ that means that the security of each them is nearly the $O(2^{m/2})$ magnitude, where $m$ is the bit-length of a message digest namely a hash output. It is well known that the output bit-lengths of these three functions are 160, 256, and 384 respectively.

When any of the three is practically paired with a lightweight signing scheme of which the modulus bit-length is between 80 and 160, its output must be adjusted to the range of the modulus bit-length of the signing scheme with its security unchanged or corresponding to the signing scheme.

The modulus bit-length of the optimized REESSE1+ signing scheme based on a transcendental logarithm problem and a polynomial root finding problem is 80 [3], and its security is the $2^{80}$ magnitude. When SHA-1 is paired with this signing scheme, the output of SHA-1 must be adjusted to 80 bits with its security unchanged. Again when SHA-256 is paired with ECC-160, the output of SHA-256 must be adjusted to 160 bits with its security being at least the $2^{80}$ magnitude.

Therefore, it is a problem in practice how to adjust a message digest from a classical hash function to the range of the modulus bit-length of a host signing scheme and to keep the security of the message digest being unchanged or corresponding to the host signing scheme.

To settle this problem, the authors design a new non-MDS hash function called JUNA which includes two algorithms: an initialization algorithm and a compression algorithm, converts a short message or a message digest of $n$ bits into an output string of $m$ bits, where $80 \leq m \leq 232$ and $80 \leq m \leq n \leq 4096$, and moreover ensures that the security of the output string against collision attacks gets the $O(2^n)$ magnitude.

This paper has two dominant novelties: ① designing the initialization algorithm based on a multivariate permutation problem which only has an exponential time solution currently, and makes the new hash function be able to resist a birthday attack; ② designing the compression algorithm based on an anomalous subset product problem which also only has an exponential time solution currently, and makes the new hash function be able to resist other classical attacks, especially including a meet-in-the-middle attack. The significance of the paper lies in the thing that a new non-MDS hash function with an $m$-bit output and the $O(2^n)$ magnitude security is first proposed by the authors while a classical iterative hash function is with an $m$-bit output and only the $O(2^{m/2})$ magnitude security.

In Section 2 of the paper, several relevant definitions are given. In Section 3, the two algorithms of the new hash function are described. In Section 4, the security of the new hash function is analyzed. In Section 5, the running time of the compression algorithm of the new hash is dissected, a comparison with another non-MDS hash, the Chaum-Heijst-Pfitzmann hash based on a discrete logarithm problem, is made, and the reformation of a classical hash function is illustrated.

Throughout the paper, unless otherwise specified, an even number $n \geq 80$ is the bit-length of a short message (a message digest) or the item-length of a sequence, the sign % denotes “modulo”, $|\cdot|$ means the size of a set, and gcd($x$, $y$) represents the greatest common divisor of two integers $x$ and $y$. Without ambiguity, “$\% M$” is usually omitted in expressions.

## 2 Several Definitions

Before the two algorithms of the new non-MDS hash function are described, three relevant definitions are presented.

### 2.1 A Coprime Sequence

**Definition 1:** If $A_1, \ldots, A_n$ are $n$ pairwise distinct positive integers such that $\forall A_i$ and $A_j$ ($i \neq j$), either \(\gcd(A_i, A_j) = 1\) or $\gcd(A_i, A_j) = F \neq 1$ with $(A_i / F) \not\in A_i$ and $(A_j / F) \not\in A_j \forall k (k \neq i, j) \in [1, n]$, these integers are called a coprime sequence, denoted by $\{A_1, \ldots, A_n\}$, and shortly $\{A_i\}$.

Notice that the elements of a coprime sequence are not necessarily pairwise coprime, but a sequence of the elements are pairwise coprime is a coprime sequence.

For example, $\{21, 15, 29, 23, 11, 17, 19, 13\}$ and $\{23, 7, 11, 3, 19, 13, 5, 17\}$ are two coprime sequences separately.

**Property 1:** Let $\{A_1, \ldots, A_n\}$ be a coprime sequence. If randomly select $k \in [1, n]$ elements $A_{i_1}, \ldots, A_{i_k}$ from the sequence, then the mapping from a subset $\{A_{i_1}, \ldots, A_{i_k}\}$ to a subset product $G = \prod_{i=1}^{n} A_i^{b_i}$ is one-to-one, namely the mapping from $b_1 \ldots b_n$ to $G = \prod_{i=1}^{n} A_i^{b_i}$ is one-to-one, where $b_1 \ldots b_n$ is a bit string. Refer to [3] for its proof.
2.2 A Bit Shadow and a Bit Long-Shadow

**Definition 2:** Let $b_1 \ldots b_n \neq 0$ be a bit string. Then $b_i$ with $i \in [1, n]$ is called a bit shadow if it comes from such a rule: \( \circ \ b_i = 0 \) if $b_i = 0$; \( \circ \ b_i = 1 \) + the number of successive 0-bits before $b_i$ if $b_i = 1$; or \( \circ \ b_i = 1 \) + the number of successive 0-bits after the rightmost 1-bit if $b_i$ is the leftmost 1-bit.

Notice that (3) of this definition is slightly different from that in [3].

**Fact 1:** Let $b_1 \ldots b_n$ be the bit shadow string of $b_1 \ldots b_n \neq 0$. Then there is $\sum_{i=1}^{n} b_i = n$.

**Proof:**

According to Definition 2, every bit of $b_1 \ldots b_n$ is considered into $\sum_{i=1}^{n} b_i$, where $b_i$, ..., $b_n$ are 1-bit shadows in the string $b_1 \ldots b_n$, and there is $\sum_{i=1}^{n} b_i = n$.

On the other hand, there is $\sum_{i=1}^{n} b_i = 0$, where $b_1$, ..., $b_n$ are 0-bit shadows.

In total, there is $\sum_{i=1}^{n} b_i = n$.

**Property 2:** Let $\{A_1, \ldots, A_n\}$ be a coprime sequence, and $b_1 \ldots b_n$ be the bit shadow string of $b_1 \ldots b_n \neq 0$. Then the mapping from $b_1 \ldots b_n$ to $G = \prod_{i=1}^{n} A_i$ is one-to-one.

**Proof:**

Step 1.

Let $b_1 \ldots b_n$ and $b' \ldots b'_n$ be two different nonzero bit strings, and $b_1 \ldots b_n$ and $b'_1 \ldots b'_n$ be the two corresponding bit shadow strings.

If $b_1 \ldots b_n = b'_1 \ldots b'_n$, then by Definition 2, there is $b_1 \ldots b_n = b'_1 \ldots b'_n$.

In addition, for any arbitrary bit shadow string $b_1 \ldots b_n$, there always exists a preimage $b_1 \ldots b_n$. Thus, the mapping from $b_1 \ldots b_n$ to $b_1 \ldots b_n$ is one-to-one.

Step 2.

Obviously the mapping from $b_1 \ldots b_n$ to $\prod_{i=1}^{n} A_i$ is surjective.

Again presuppose that $\prod_{i=1}^{n} A_i = \prod_{i=1}^{n} A_i$ for $b_1 \ldots b_n \neq b'_1 \ldots b'_n$.

Since $\{A_1, \ldots, A_n\}$ is a coprime sequence, and $A_i$ either equals 1 with $b_i = 0$ or contains the same prime factors as those of $A_i$ with $b_i = 0$, we can obtain $b_1 \ldots b_n = b'_1 \ldots b'_n$ from $\prod_{i=1}^{n} A_i = \prod_{i=1}^{n} A_i$, which is in direct contradiction to $b_1 \ldots b_n \neq b'_1 \ldots b'_n$.

Therefore, the mapping from $b_1 \ldots b_n$ to $\prod_{i=1}^{n} A_i$ is injective [12].

In summary, the mapping from $b_1 \ldots b_n$ to $\prod_{i=1}^{n} A_i$ is one-to-one, and further the mapping from $b_1 \ldots b_n$ to $\prod_{i=1}^{n} A_i$ is also one-to-one.

**Definition 3:** Let $b_1 \ldots b_n$ be the bit shadow string of $b_1 \ldots b_n \neq 0$. Then $b_i = b_i 2^x$ with $i \in [1, n]$ is called a bit long-shadow, where $x = \sum_{i=1}^{n} b_i 2^x$ and $x = \sum_{i=1}^{n} b_i = n$.

According to Definition 3, it is not difficult to understand that for every $b_i$, there is $0 \leq b_i \leq n$ when $b_1 \ldots b_n \neq 0$.

For example, let $b_1 \ldots b_n = 01010110$, then $b_1 \ldots b_n = 06020410$.

**Fact 2:** Let $b_1 \ldots b_n$ be the bit long-shadow string of $b_1 \ldots b_n \neq 0$. Then there is $n \leq \sum_{i=1}^{n} b_i \leq 2n$.

**Proof:**

By Definition 3 and Fact 1, we have

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} b_i 2^x$$

If every $b_i = 1$, namely every $x_i = 1$, then

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} b_i 2^x = 2 \sum_{i=1}^{n} b_i = 2n.$$

Again, by Definition 3, not all the bits of $b_1 \ldots b_n$ are zero.

If there exists only one nonzero bit in $b_1 \ldots b_n$, namely every $x_i = 1$ with $x \in [1, n]$ for example, then

$$\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} b_i 2^x = b_i 2^x = b_i = n,$$

where $x = b_i 2^x$ is due to $b_i$ being the unique nonzero bit.

Thus, it holds that $n \leq \sum_{i=1}^{n} b_i \leq 2n$.

**Property 3:** Let $b_1 \ldots b_n$ be the bit long-shadow string of $b_1 \ldots b_n \neq 0$. Then the mapping from $b_1 \ldots b_n$ to $b_1 \ldots b_n$ is one-to-one.

**Proof:**

On one hand, assume that a bit string $b_1 \ldots b_n \neq 0$ is known.
It is understood from Definition 3 that $m = 2^i$ for each $i$, where $m = b_1^i \cdot \ldots \cdot b_n^i$.

According to $m = b_1^i \cdot \ldots \cdot b_n^i$ and $m = 0$ with $b_i = 0$, where $m = b_1^i \cdot \ldots \cdot b_n^i$, we can determine $b_i$ for $i = 1, \ldots, n$ as follows.

1. Case of $b_i = 0$
   - $b_i = 0$, then $b_i = 0$, and set $b_i = 0$.
2. Case of $b_i \neq 0$
   - $b_i \neq 0$, and set $b_i = 1$.

In this way, the value of every $b_i$ can be determined uniquely.

In summary, the mapping from $b_1, \ldots, b_n$ to $b_1, \ldots, b_n$ is one-to-one.

## 2.3 A Lever Function

The designing of the initialization algorithm of the new hash function is based on the hard problem $C_i = (A_i \cdot \mathcal{W}^{(\ell_i)}(x) \pmod{M})$ for $i = 1, \ldots, n$ which is first used for the REESSE1 asymmetric cryptosystem, where the exponent $\ell(i)$ is called a lever function [3].

In the paper, we still borrow the concept of the lever function but a public key is regarded as an "a public terminal", and is small at "a private terminal".

Definition 4: The secret parameter $\ell(i)$ in the transform of a non-MDS hash function is called a lever function, if it has the following features:

1. $\ell(.)$ is an injection from the domain $\{1, \ldots, n\}$ to the codomain $\mathcal{W} \subset \{\mathcal{M}\}$, where $\mathcal{M}$ is large;
2. The mapping between $i$ and $\ell(i)$ is established randomly without an analytical expression;
3. An attacker has to be faced with all the permutations of elements in $\mathcal{W}$ when inferring a related private parameter from an initial value;
4. The owner of the private parameter only need to consider the polynomial arithmetic of elements in $\mathcal{W}$ when doing a certain computation.

Feature 1 and 3 make it clear that if $n$ is large enough, it is infeasible for the attacker to search all the permutations of elements in $\mathcal{W}$ exhaustively while the computation by the owner of the private parameter is feasible in polynomial time in $n$. Thus, the amount of calculation on $\ell(.)$ is large at "a public terminal", and is small at "a private terminal".

Notice that the number of all the elements of $\mathcal{W}$ namely the size of $\mathcal{W}$ is not less than $n$.

Property 4 (Indeterminacy of $\ell(.)$): Let $d = 1$ and $C_i = (A_i \cdot \mathcal{W}^{(\ell_i)}(x) \pmod{M})$ with $\ell(i) \in \mathcal{W} = \{1, \ldots, n\}$ and $\mathcal{A}_i \in \mathcal{A} = \{2, \ldots, \mathcal{P}\}$, where $\mathcal{P} \geq 1201$ for $i = 1, \ldots, n$. Then $\forall \mathcal{W} \not{\equiv} \mathcal{M} \in \{1, \mathcal{M}\}$, and $\forall x, y, z (x \neq y \neq z) \in \{1, \mathcal{W}\}$, there is $\ell(x) + \ell(y) = \ell(z)$, where $\mathcal{W}$ is large; $\ell(x) + \ell(y) = \ell(z)$, where $\mathcal{W}$ is large.

Proof:

1. It is easy to understand that $\mathcal{W}^{(\ell(x))} \equiv \mathcal{W}^{(\ell(y))} \equiv \mathcal{W}^{(\ell(z))} \equiv \mathcal{W}^{(\ell(\ell(z)))} \pmod{\mathcal{M}}$.
2. However, it should be noted that when $\mathcal{W} \equiv \mathcal{M}$, there is $\ell(x) + \ell(y) + \ell(z) \equiv \ell(z) + \ell(y) + \ell(x) \pmod{\mathcal{M}}$.

In light of group theories, $\ell(x) \in \{2, \ldots, \mathcal{P}\}$, the congruence $C_i = A_i \cdot \mathcal{W}^{(\ell(x))} \pmod{M}$ has a solution. Then, $\ell(x)$ may be taken through $\mathcal{O}_k$.

Further, from the congruences $C_i = A_i \cdot \mathcal{W}^{(\ell(x))} \pmod{M}$ and $C_j = A_j \cdot \mathcal{W}^{(\ell(y))} \pmod{M}$, we can obtain many
Because of {+/, …, +/} 1, …, Al, a short message or a message digest from a classical hash function, and there are 80 logarithm problem, and proved to be strongly collision-free [16].

3 Design of the New Non-MDS Hash Function

The Chaum-Heijst-Pfitzmann hash function, a non-MDS one, is appreciable. It is based on a discrete logarithm problem, and proved to be strongly collision-free [16].

The new non-MDS hash function is composed of two algorithms which contain two main parameters m and n, where m denotes the bit-length of a modulus used in the new hash, n denotes the bit-length of a short message or a message digest from a classical hash function, and there are 80 ≤ m ≤ 232 with 80 ≤ m ≤ n ≤ 4096.

Additionally, A and Ω are two integral sets, and their lengths should be selected in conformity to the values of m and n such that 2n|Ω|λ |Ω| ≥ 2n with 210 ≤ |Ω| ≤ 232 and n ≤ a ≤ 232 (see Section 4.2.1), where n = |Ω|, and 210 ≤ |Ω| ≤ 232 means 10 ≤ lg |Ω| ≤ 32.

For example, as m = 80 ≤ n, there should be |Ω| = 210 and |Ω| = n; as m = 96 ≤ n, should |Ω| = 212 and |Ω| = n; as m = 112 ≤ n, should |Ω| = 214 and |Ω| = n; as m = 128 ≤ n, should |Ω| = 216 and |Ω| = 212; and m = 232 ≤ n, should |Ω| = 232 and |Ω| = 232.

Notice that in the arithmetic modulo M, –x represents M – x.

3.1 Initialization Algorithm

This algorithm is employed by an authoritative third party or the owner of a key pair, and only needs to be executed one time.

INPUT: the bit-length m of a modulus with 80 ≤ m ≤ 232;
the item-length n of a sequence with 80 ≤ m ≤ n ≤ 4096;
the maximal prime P with 10 ≤ lg |P| ≤ 32;
the size a of the set with 2n|Ω|λ |Ω| ≥ 2n and n ≤ a ≤ 232.

S1: Produce A ← {2, 3, …, Π}.

S2: Find a prime M with |lg M| = m such that M / 2 is a prime, or the least prime factor of M / 2 > 4n(2a + 3).

S3: Pick W ∈ (1, M) making |W| ≥ 2n|lg |P| |.| Pick 3 ∈ (1, M) making gcd(δ, M) = 1.

S4: Randomly yield Ω ← {+/– 5, +/- 7, …, +/- (2a + 3)}.

Randomly select a distinct Ω(i) ∈ Ω for i = 1, …, n.

S5: Compute Ci ← (A,W(i)|M % M for i = 1, …, n.

OUTPUT: an initial value (C(i), M) which is public to the people.

A private parameter (A, W, δ) may be discarded, but must not be divulged.

Assume that there is Ci = Ci with i ≠ j. Then (A, W(i)|δ = (A, W(j)|δ ( % M), and W(i) = W(j) ( % M).

Because of M / 2 = a prime or the least prime factor of M / 2 > 4n(2a + 3), the probability that the case W(i) = W(j) = A, A–1 ( % M), namely Ci = Cj occurs is 1/2n.

At S3, to seek W, let W = g |Ω| /F ( % M), where g is a generator of (ZM, ·) obtained through Algorithm 4.80 in Section 4.6 of [9], and F < 2|lg |P| is a factor of M.

At S4, Ω = {+/– 5, +/- 7, …, +/- (2a + 3)} indicates that Ω is one of 2d potential sets, indeterminate, and unknown to the public, where “+/–” means the selection of the “+” or “–” sign.

Definition 5: Given (A(i), M), seeking the original (A(i), W(i), δ) from Ci = (A(W(i)|δ ( % M) with Ai ∈ {2, 3, …, Π} 10 ≤ |lg |P| ≤ 32) and (i(i) ∈ {+/– 5, +/- 7, …, +/- (2a + 3) | n ≤ a ≤ 232} for i = 1, …, n is referred to as a multivariate permutation problem, shortly MPP [3].

Property 5: The MPP Ci = (A(W(i)|δ ( % M) with Ai ∈ {2, 3, …, Π} 10 ≤ |lg |P| ≤ 32) and (i(i) ∈ {+/– 5, +/- 7, …, +/- (2a + 3) | n ≤ a ≤ 232} for i = 1, …, n is computationally at least equivalent to the
DLP in the same prime field.

See Section 4.1 for its proof.

3.2 Compression Algorithm

This algorithm is employed by a person who wants to obtain a short message digest.
INPUT: an initial value (\(\{C_1, \ldots, C_n\}, M\)), where \(\lceil \log M \rceil = m\) with \(80 \leq m \leq n \leq 4096\); a short message (or a message digest from a classical hash function) \(b_1 \ldots b_n \neq 0\).

S1: Set \(k \leftarrow 0, i \leftarrow 1\).
S2: If \(b_i = 0\) then
   S2.1: let \(k \leftarrow k + 1, b_i \leftarrow 0\)
   else
   S2.2: if \(i = k + 1\) then let \(\hat{s} \leftarrow i\);
   S2.3: let \(b_i \leftarrow k + 1, k \leftarrow 0\).
S3: Let \(i \leftarrow i + 1\).
If \(i \leq n\) then go to S2.
S4: Compute \(b_i \leftarrow b_i + k\).
S5: Compute \(g \leftarrow \prod_{i=1}^{n} C_i^{b_i} \mod M\),
where \(b_i = b_i 2^\hat{s}\) with \(\hat{a}_i = b_i (2^ {-1})^{(\hat{i} (\hat{s} + 1))} \mod (n/2)\).

OUTPUT: a digest \(g = \prod_{i=1}^{n} C_i^{b_i} \mod M\) of which the bit-length is \(m\).
It is easily known from Definition 3 that the max of \(\{b_1, \ldots, b_n\}\) is less than or equal to \(n\) when \(b_1 \ldots b_n \neq 0\).

Definition 6: Given \((g, M)\), seeking the original \(b_1 \ldots b_n\) from \(g = \prod_{i=1}^{n} C_i^{b_i} \mod M\), where \(b_i = b_i 2^\hat{s}\) with \(\hat{a}_i = b_i (2^ {-1})^{(\hat{i} (\hat{s} + 1))} \mod (n/2)\) and \(b_i\) being a bit shadow is referred to as an anomalous subset product problem, shortly ASPP [3].

Property 6: The ASPP \(g = \prod_{i=1}^{n} C_i^{b_i} \mod M\), where \(b_i = b_i 2^\hat{s}\) with \(\hat{a}_i = b_i (2^ {-1})^{(\hat{i} (\hat{s} + 1))} \mod (n/2)\) and \(b_i\) being a bit shadow is computationally at least equivalent to the DLP in the same prime field.

See Section 4.3 for its proof.

4 Security Analysis of the New Non-MDS Hash Function

Because a hash function must be one-way, weakly collision-free, and sometimes required to be strongly collision-free, the new non-MDS hash function should also be at least one-way and weakly collision-free.

It is should be noted that \(\lceil \log M \rceil = m\), but not \(n\), is the security dominant parameter of the new non-MDS hash function.

Definition 7: Let \(A\) and \(B\) be two computational problems. \(A\) is said to reduce to \(B\) in polynomial time, written as \(A \leq_T^r B\), if there is an algorithm for solving \(A\) which calls, as a subroutine, a hypothetical algorithm for solving \(B\), and runs in polynomial time, excluding the time of the algorithm for solving \(B\) [9][17].

The hypothetical algorithm for solving \(B\) is called an oracle. It is easy to understand that no matter what the time complexity of the oracle is, it does not influence the result of the comparison.

\(A \leq_T^r B\) means that the difficulty of \(A\) is not greater than that of \(B\), namely the time complexity of the fastest algorithm for solving \(A\) is not greater than that of the fastest algorithm for solving \(B\) when all polynomial times are treated as the identical magnitude. Concretely speaking, if \(A\) cannot be solved in polynomial or subexponential time, correspondingly \(B\) cannot also be solved in polynomial or subexponential time; and if \(B\) can be solved in polynomial or subexponential time, correspondingly \(A\) can also be solved in polynomial or subexponential time.

Definition 8: Let \(A\) and \(B\) be two computational problems. If \(A \leq_T^r B\) and \(B \leq_T^r A\), then \(A\) and \(B\) are said to be computationally equivalent, written as \(A =_T^r B\) [9][17].

\(A =_T^r B\) means that either if \(A\) is a intractability with a certain complexity on a condition that its dominant variable approaches a large number, \(B\) is also a intractability with the same complexity on the identical condition; or both \(A\) and \(B\) can be solved in linear or polynomial time.

Obviously, Definition 7 and 8 gives a partial order relation among the complexities or difficulties of computational problems [18], and suggest a reductive proof method called polynomial time Turing
In addition, for convenience sake, let $\hat{H}(y = f(x))$ represent the complexity or difficulty of the problem of solving $y = f(x)$ for $x$.

### 4.1 Proof of Property 5

In Section 3.1, the MPP is defined as $C_i = (A_i W^{(i)})^\delta (\% M)$ with $A_i \in A = \{2, 3, \ldots, P\}$ and $\ell(i) \in \Omega = \{+/−5, +/−7, \ldots, +/−(2n + 3)\}$ for $i = 1, \ldots, n$. What follows is the proof of Property 5, a property of the MPP.

**Proof:**

Firstly, systematically consider $C_i = (A_i W^{(i)})^\delta (\% M)$ for $i = 1, \ldots, n$.

Assume that each $g_i = A_i W^{(i)} (\% M)$ with $\ell(i) \in \{+/−5, +/−7, \ldots, +/−(2n + 3)\}$ for $i = 1, \ldots, n$. What follows is the proof of Property 5, a property of the MPP.

Now let $\beta = g_i^\delta (\% M)$ and $z_i = \delta x_i$. Then

$C_i = \beta_i \delta_i (\% M) = \beta_i^\delta (\% M)$ for $i = 1, \ldots, n$.

The above expression corresponds to the fact that in the ElGamal cryptosystem where many users share the modulus and a key generator, User 1 acquires a private key $z_1$ and a public key $C_1$, ..., and User $n$ acquires a private key $z_n$ and a public key $C_n$. It is well known that in this case, the attack of an adversary is still faced with the DLP, namely seeking $z_i$ from the simultaneous equation $C_i = g_i^{\delta_i} (\% M)$ for $i = 1, \ldots, n$ is computationally equivalent to the DLP.

Thus, when every $g_i$ is weakened to a constant, seeking $\delta$ from $C_i = g_i^\delta (\% M)$ for $i = 1, \ldots, n$ is computationally equivalent to the DLP, which indicates that when every $g_i$ is not a constant, seeking $g_i$ and $\delta$ from $C_i = g_i^{\delta_i} (\% M)$ for $i = 1, \ldots, n$ is computationally at least equivalent to the DLP.

Secondly, singularly consider a certain $C_i$ for which the subscript $i$ is designated.

Assume that $\tilde{\hat{O}}_{\hat{a}}(C_i, M, g)$ is an oracle on solving $C_i = g_i^\delta (\% M)$ for $g_i$ and $\delta$, where $i$ is in $\{1, \ldots, n\}$, and $g$ is a constraint on $g_i$ such that the original $g_i$ and $\delta$ can be found.

Let $y = g_i^\delta (\% M)$ be the DLP. Then, by calling $\tilde{\hat{O}}_a(y, M, g)$, $x$ can be obtained.

According to Definition 7, there is

$\hat{H}(y = g_i^\delta (\% M)) \leq \hat{H}(\tilde{\hat{O}}_a(y, M, g))$,

which indicates that when only a certain $g_i$ is known, seeking $g_i$ and $\delta$ from $C_i = g_i^\delta (\% M)$ is computationally at least equivalent to the DLP.

Integrally, seeking the original $(\beta_1, \ell(i))$, $W$, and $\delta$ from $C_i = (A_i W^{(i)})^\delta (\% M)$ for $i = 1, \ldots, n$ is computationally at least equivalent to the DLP in the same prime field.

### 4.2 Security of the Initialization Algorithm

Clearly, the security of the initialization algorithm depends on the security of the MPP $C_i = (A_i W^{(i)})^\delta (\% M)$ with $A_i \in A = \{2, 3, \ldots, P\}$ and $\ell(i) \in \Omega = \{+/−5, +/−7, \ldots, +/−(2n + 3)\}$ for $i = 1, \ldots, n$.

In [3], we analyze the security of the MPP $C_i = (A_i W^{(i)})^\delta (\% M)$ with $A_i \in \{2, 3, \ldots, P\}$ and $\ell(i) \in \Omega = \{5, 7, \ldots, (2n + 3)\}$ for $i = 1, \ldots, n$ from the three aspects, discover no subexponential time solution to it, and contrarily, find some evidence which inclines people to believe that the MPP is computationally harder than the DLP.

Considering that the set $\Omega$ is different from the old in [3], and the range of $P$ is larger than the old in [3], we will analyze the security of the MPP with different restrictions additionally.

#### 4.2.1 Ineffectualness of Presupposing $\ell(x_1) + \ell(x_2) = \ell(y_1) + \ell(y_2)$

Because of $\Omega = \{+/−5, +/−7, \ldots, +/−(2n + 3)\}$, when the absolute values $|\ell(x_1)|, |\ell(x_2)|, |\ell(y_1)|, |\ell(y_2)|$ are determined, the value $\ell(x_1) + \ell(x_2) = (\ell(y_1) + \ell(y_2))$ has $2^4 = 16$ possible cases, which enhances the indeterminacy of the lever function, and increases the complexity of an attack task for cracking the MPP to some extent.
Adversaries may try to eliminate $W$ through judging $\ell(x_1) + \ell(x_2) = \ell(y_1) + \ell(y_2)$.

$\forall x_1, x_2, y_1, y_2 \in [1, n]$, presuppose that $\ell(x_1) + \ell(x_2) = \ell(y_1) + \ell(y_2)$ holds.

Let $G_2 = C_{i_2}C_{i_3}(C_{i_2}C_{i_3})^{-1} \pmod{M}$, namely

$$G_2 = (A_{i_2}A_{i_3}(A_{i_2}A_{i_3})^{-1})^\delta \pmod{M}.$$ 

If the adversaries divine the values of $A_{i_1}, A_{i_2}, A_{i_3}, A_{i_4}$, and compute $u, v_{x_1}, v_{x_2}, v_{y_1}, v_{y_2}$ in at least $L_M[1/3, 1.923]$ time such that

$$G_2 = g^x, A_{i_1} = g^{v_{x_1}}, A_{i_2} = g^{v_{x_2}}, A_{i_3} = g^{v_{y_1}}, A_{i_4} = g^{v_{y_2}} \pmod{M},$$

where $g$ is a generator of $(\mathbb{Z}_M^*, \cdot)$, then

$$u = (v_{x_1} + v_{x_2} - v_{y_1} - v_{y_2})^\delta \pmod{M}.$$ 

If gcd$(v_{x_1} + v_{x_2} - v_{y_1} - v_{y_2}, M) \mid u$, the congruence in $\delta$ has solutions. Because each of $A_{i_1}, A_{i_2}, A_{i_3}, A_{i_4}$ may traverse the interval $A$, and the subscripts $x_1, x_2, y_1, y_2$ are unfixed, the number of potential values of $\delta$ is about $n^4A^4$. Notice that the number of non-repeated values of $\delta$ will be less than $2^n$.

In succession, need to seek $W$. Now, the most effectual approach to seeking $W$ is that for every $i$, the adversaries fix a value of $\delta_i$, divine $A_i$ and $\ell(W)$, and find the set $V_i$ according to $C_i = (A_iW^{(0)})^\delta \pmod{M}$, where $V_i$ is the set of possible values of $W$ meeting $C_i = (A_iW^{(0)})^\delta \pmod{M}$ for $i = 1, \ldots, n$. If there exist $W_i \in V_i, \ldots, W_n \in V_n$ being pairwise equal, the divination of $\delta, \{A_i\}$, and $\{\ell(i)\}$ is right; else fix another value of $\delta$, repeat the above process.

Notice that due to $\delta / 2 = \ell(W) / 2 > 4n(2^\delta + 3)$, $W^{(0)} = C_i^{1/4}A_i^{-1} \pmod{M}$ can be solved in polynomial time, and besides letting $W = g^\delta \pmod{M}$ is unnecessary.

It is not difficulty to understand that the size of every $V_i$ is about $(2^i)(2_i^i)A_i^i$.

In summary, the time complexity of the above attack task is

$$F = (n + 1A_i)\log_2(1 / 3, 1.923) + (n^4A^4) + (n^4A^4)(2^i2^iA_i^n) \approx 2n^4A^4\log_2(3, 1.923).$$

Concretely speaking,

- For $m = n = 80$, $|A_i| = 2^{10} \& |2^i| = 80, F > 2(2^{10})^5(2^{10})^5 = 2^{88} > 2^n$.
- For $m = n = 96$, $|A_i| = 2^{12} \& |2^i| = 96, F > 2(2^{12})^5(2^{12})^5 = 2^{100} > 2^n$.
- For $m = n = 112$, $|A_i| = 2^{14} \& |2^i| = 112, F > 2(2^{14})^5(2^{14})^5 = 2^{112} = 2^n$.
- For $m = n = 128$, $|A_i| = 2^{16} \& |2^i| = 128, F > 2(2^{16})^5(2^{16})^5 = 2^{128} = 2^n$.
- For $m = n = 232$, $|A_i| = 2^{32} \& |2^i| = 232, F > 2(2^{32})^5(2^{32})^5 = 2^{232} > 2^n$.

Thus, the time complexity of the attack by presupposing $\ell(x_1) + \ell(x_2) = \ell(y_1) + \ell(y_2)$ is not less than $O(2^n)$ when $|A_i|$ and $|2^i|$ are chosen suitably.

### 4.2.2 Ineffectualness of Guessing $\|W\|$ 

Owing to $80 \leq \lceil \log_2 M \rceil \leq 232$, $\mathcal{M}$ can be factorized in tolerable subexponential time, and further a value of $\|W\|$ can be guessed.

Adversaries may try to eliminate $W$ through $W^{\|W\|} = 1 \pmod{M}$.

Raising either side of every equation $C_i = (A_iW^{(0)})^\delta \pmod{M}$ to the $\|W\|$-th power yields

$$C_i^{\|W\|} = (A_iW^{(0)})^{\delta \|W\|} \pmod{M}.$$ 

Suppose that the value of every $A_i \in A = \{2, 3, \ldots, P | 10 \leq \lceil \log_2 P \rceil \leq 32\}$ is guessed, or the possible values of every $A_i$ are traversed.

Let $C_i = g^x \pmod{M}$, and $A_i = g^y \pmod{M}$, where $g$ is a generator of $(\mathbb{Z}_M^*, \cdot)$. Then

$$u_i\|W\| = v_i\|W\|\delta \pmod{M} \quad (i = 1, \ldots, n).$$

Notice that $u_i \neq v_i \delta \pmod{M}$, and $\{v_1, \ldots, v_n\}$ is not a super increasing sequence.

The above congruence is seemingly the MH transform [20]. Actually, $\{v_i\|W\|, \ldots, v_n\|W\|\}$ is not a super increasing sequence, and moreover there is not necessarily $\lceil \log(u_i\|W\|) \rceil = \lceil \log \mathcal{M} \rceil$.

Because $v_i\|W\| \in [1, \mathcal{M}]$ is stochastic, the inverse $\delta^{-1} \pmod{M}$ need not be close to the minimum $\mathcal{M} / (u_i\|W\|)$, $2\mathcal{M} / (u_i\|W\|)$, $\ldots$, or $(u_i\|W\| - 1)\mathcal{M} / (u_i\|W\|)$. Namely $\delta^{-1}$ may lie at any integral position of the interval $[k\mathcal{M} / (u_i\|W\|), (k + 1)\mathcal{M} / (u_i\|W\|)]$, where $k = 0, 1, \ldots, u_i\|W\| - 1$, which illustrates that the accumulation points of minima do not exist. Further observing, in this case, when $i$ traverses the interval $[2, n]$, the number of intersections of the intervals containing $\delta^{-1}$ is likely the max of $\{u_i\|W\|, \ldots, u_n\|W\|\}$ which is promisingly close to $\mathcal{M}$. Therefore, the Shamir attack by the accumulation point of minima is fully ineffectual [21].

Even if find out $\delta^{-1}$ through the Shamir attack method, because each of $\{v_i, \ldots, v_n\}$ has $\|W\|$ solutions, the number of potential sequences $\{g^{u_1}, \ldots, g^{u_n}\}$ is up to $\|W\|^n$. Because of needing to verify whether
\{g^n, \ldots, g^n\} is a coprime sequence for each different sequence \{v_1, \ldots, v_n\}, the number of possible coprime sequences is in proportion to \(|W|^n\). Hence, the initial \{A_1, \ldots, A_n\} cannot be determined in subexponential time. Further, the value of \(W\) cannot be computed, and the values of \(|W|\) and \(\delta^{-1}\) cannot be verified, which indicates that the MPP can also be resistant to the Shamir attack by the accumulation point of minima.

Additionally, the adversaries may divine the value of \(A_i\) in about \(O(|A|!\) time with \(i \in [1, n]\), and compute \(\delta\) by \(v_i|W| = u_i|W|/\delta\) \((mod\ M)\). However, because of \(|W| \mid A_i\), the equation will have \(|W|\) solutions. Therefore, the time complexity of finding the original \(\delta\) is at least

\[ T = (n + |A|)\nu M [1 / 3, 1.923] + \frac{|A||W|}{\delta} \]

\[ \geq (n + |A|)\nu M [1 / 3, 1.923] + 2^{\frac{|W|}{2}} > 2^m. \]

It is also not less than \(O(2^m)\).

### 4.3 Proof of Property 6

In Section 3.2, the ASPP is defined as \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\), where \(b_i = b_i2^\phi\) with \(\phi = b_{i-1}|2^{v_i}|/|v_i/2\) and \(b_i\) being a bit shadow. What follows is the proof of Property 6, a property of the ASPP.

**Proof:**

Assume that \(\tilde{O}(\mathcal{G}, C_1, \ldots, C_m, M)\) is an oracle on solving \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) for \(b_1\ldots b_n\), where \(b_1\ldots b_n\) is the bit-long-shadow string of \(b_1\ldots b_n\).

Particularly, when \(C_1 = \ldots = C_n = C\), define

\[ \mathcal{G} = \prod_{i=1}^n C_i^{b_i} = \prod_{i=1}^n (C^{a_i})^{b_i} (%M) \]

with \(0 \leq b_i \leq n\), and define the corresponding oracle as \(\tilde{O}(\mathcal{G}, C^{a_1}, \ldots, C^{a_n}, M)\).

Let \(\mathcal{G}_i = \prod_{i=1}^n C_i^{b_i} (%M)\) be of the subset product problem (SPP) \([3][22][23]\).

Since there is \(0 \leq b_i \leq b_n\) and the mapping from \(b_1\ldots b_i\) to \(b_1\ldots b_n\) is one-to-one, by calling \(\tilde{O}(\mathcal{G}_i, C_1, \ldots, C_m, M)\), we can find \(b_1\ldots b_n\).

By Definition 7, there is

\[ \tilde{H}(\mathcal{G}_i) = \prod_{i=1}^n C_i^{b_i} (%M) \leq \tilde{H}(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)). \]

By Property 5 in \([3]\), there is

\[ \tilde{H}(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)) \leq \tilde{H}(\mathcal{G}_i) = \prod_{i=1}^n C_i^{b_i} (%M)). \]

Further, by transitivity, there is

\[ \tilde{H}(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)) \leq \tilde{H}(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)). \]

Therefore, solving \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) for \(b_1\ldots b_n\) is at least equivalent to the DLP in the same prime field in computational complexity. \(\square\)

### 4.4 Security of the Compression Algorithm

The compression algorithm of which the input message is treated as only a block is the main body of the new non-MDS hash function, and thus, through it the four natural properties of the new hash function are embodied dominantly.

Clearly, the security of the compression algorithm depends on the security of the ASPP \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\), where \(b_i = b_i2^\phi\) with \(\phi = b_{i-1}|2^{v_i}|/|v_i/2\) and \(b_i\) being a bit shadow.

In \([3]\), we analyze the security of the ASPP \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) from the three aspects, discover no subexponential time solution to it, and contrarily, find some evidence which inclines people to believe that \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) is computationally harder than the DLP. Due to \(b_i = b_i2^\phi \geq b_n\), the security conclusion about \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) is also suitable for \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) which is just another form of the ASPP. Hence \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) has no subexponential time solution at present.

In what follows, we will analyze whether the compression formula \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\) satisfies the four natural properties of a hash function, and especially resists the three classical attacks or not.

In terms of Section 3.2, given the initial value \((\{C_i\}, M)\) and a short message \(b_1\ldots b_n\), it is transparently easy to calculate the digest \(\mathcal{G} = \prod_{i=1}^n C_i^{b_i} (%M)\).
4.4.1 Compression Algorithm Is Computationally One-way

Let \( C_l = g^{x_l} \mod M \), \( \ldots \), \( C_n = g^{x_n} \mod M \), where \( g \) is a generator of the group \( \mathbb{Z}_M^\times \), and easily found when \( \lfloor \log M \rfloor < 1024 \).

Then, solving \( g = \prod_{i=1}^{n} C_i b_i \mod M \) for \( b_1 \ldots b_n \), namely \( b_1 \ldots b_n \), is equivalent to solving

\[
\prod_{i=1}^{n} b_i u_i + \prod_{i=1}^{n} b_i u_i = v \mod M,
\]

which is called an anomalous subset sum problem, shortly ASSP [3], and computationally at least equivalent to a subset sum problem (SSP) due to \( \prod_{i=1}^{n} b_i 2^{x_i} \geq b_i \geq b_i \in [0, 1] \).

The SSP has been proved to be NP-complete in its feasibility recognition form, and its computational version, especially the high-density or big-length version, is NP-hard [9][24]. Hence, solving ASSP is at least NP-hard.

Moreover in the non-MDS hash function, there is \( n \geq m = \lfloor \log M \rfloor \) and \( n \geq b_i \geq b_i \in [0, 1] \). The knapsack density relevant to the ASSP \( \prod_{i=1}^{n} b_i u_i = v \mod M \) roughly equals

\[
D = \sum_{i=1}^{n} \lfloor \log n \rfloor / \lfloor \log M \rfloor = n \log n / m > \log n > 1,
\]

which means that there exists many solutions to \( \prod_{i=1}^{n} b_i u_i = v \mod M \), namely the original solution cannot be determined, or will not occur in a reduced lattice base defined by LLL [25]. Notice that only such a \( \langle b_1, \ldots, b_n \rangle \) from which a right bit string can be deduced will be a reasonable solution vector. Experiments show that when \( D > 1 \), the probability that the original solution or a reasonable solution is found through LLL lattice base reduction is almost zero [26].

Hence, LLL lattice base reduction attack on ASSP [25][27] is utterly ineffectual, which illustrates that even although a DLP with the modulus bit-length less than 1024 can be solved, the original or a reasonable \( b_1 \ldots b_n \) cannot be found yet in DLP subexponential time, namely \( g = \prod_{i=1}^{n} C_i b_i \mod M \) is computationally one-way.

4.4.2 Compression Algorithm Is Weakly Collision-free

Assume that \( b_1 \ldots b_n \neq 0 \) is a short message or a message digest from a classical hash function. By Definition 3, we easily understand that \( b_i = b_i 2^{x_i} \leq n \forall i \in [1, n] \).

Given a short message \( b_1 \ldots b_n \neq 0 \), and let \( b_1 \ldots b_n \neq 0 \) be another short message to need to be found.

Let \( b_1 \ldots b_n \) be the bit long-shadow string of \( b_1 \ldots b_n \), and \( b_1 \ldots b_n \) be the bit long-shadow string of \( b_1 \ldots b_n \).

Let \( \hat{h} \) be the compression algorithm of the new non-MDS hash function described in Section 3.2.

Hence, we have

\[
g = \hat{h}(b_1 \ldots b_n) = \prod_{i=1}^{n} C_i b_i \mod M,
\]

and

\[
g' = \hat{h}(b_1' \ldots b_n') = \prod_{i=1}^{n} C_i b_i' \mod M,
\]

where \( b_i = b_i 2^{x_i} \) with \( x_i = b_i - (\lfloor \log n \rfloor - \| b_i \|_2) \) and \( b_i' = b_i' 2^{x_i'} \) with \( x_i' = b_i' - (\lfloor \log n \rfloor - \| b_i' \|_2) \).

If \( g = g' \), there is \( \prod_{i=1}^{n} C_i b_i = \prod_{i=1}^{n} C_i b_i' \mod M \).

Observe an extreme case. Assume that \( C_1 = \ldots = C_n = C \).

Owing to the max of \( 0 \leq b_i \leq n \), we define logically

\[
\prod_{i=1}^{n} C_i b_i = \prod_{i=1}^{n} C_i (a + 1)^{x_i - b_i} \mod M.
\]

Under the circumstances, if \( g = g' \), then there is

\[
\prod_{i=1}^{n} C_i (a + 1)^{x_i' - b_i'} = \prod_{i=1}^{n} C_i (a + 1)^{x_i - b_i} \mod M,
\]

namely

\[
C^{\sum_{i=1}^{n} (a + 1)^{x_i' - b_i'}} = C^{\sum_{i=1}^{n} (a + 1)^{x_i - b_i}} \mod M.
\]

Let \( z = \sum_{i=1}^{n} b_i (n + 1)^{x_i - b_i} \mod M \), and \( z' = \sum_{i=1}^{n} b_i' (n + 1)^{x_i' - b_i'} \mod M \).

Correspondingly,

\[
C^z = C^{z'} \mod M.
\]

We need to solve the above equation for \( z' \).

If the order \( |C| \) is known, let \( z' = z + k|C| \), where \( k \geq 1 \) is an integer. Once a fit \( k \) is found, there will be \( C^z = C^{z'} \mod M \), and a bit string can be inferred from \( b_1 \ldots b_n \). However, seeking \( |C| \) is of the integer factorization problem (IFP) at present because the prime factors of \( M \) must be known.

In practice, \( C_1, \ldots, C_n \) that are produced through the algorithm in Section 3.1 are pairwise unequal, which implies that for any given short message \( b_1 \ldots b_n \), seeking another short message \( b_1' \ldots b_n' \) such that
The Chaum-Heijst-Pfitzmann hash function, it is resistant to single-block differential attack \cite{28}.

Therefore, we say that the new non-MDS hash function is weakly collision-free.

Again because the new hash function is non-MDS, and based on the intractabilities, like the Chaum-Heijst-Pfitzmann hash function, it is resistant to single-block differential attack \cite{28}.

### 4.4.3 Compression Algorithm Is Resistant to Birthday Attack

First, observe an example of whether any two students in a class have the same birthday.

Suppose that the class has 23 students. If a teacher specifies a day (say February 12), then the chance that at least one student has the same birthday as any other student is around \((1 – (365 / 365)^{23})\) \approx 50.73 \%, which prompts birthday attack on hash functions.

Birthday attack is widely exploited for finding any two messages \(\mathfrak{q}\) and \(\mathfrak{q}'\) such that \(\mathfrak{h}(\mathfrak{q}) = \mathfrak{h}(\mathfrak{q}')\), namely \(\mathfrak{q}, \mathfrak{q}'\) is a collision, where \(\mathfrak{h}\) is a hash function \cite{29}. If the bit-length of a message digest is \(m\), an adversary can find a collision \(\mathfrak{q}, \mathfrak{q}'\) such that \(\mathfrak{h}(\mathfrak{q}) = \mathfrak{h}(\mathfrak{q}')\) with probability 50\% in roughly \(1.1774 \times 2^{m/2}\) time, namely with input of \(1.1774 \times 2^{m/2}\) random messages \cite{30}.

However, to the new non-MDS hash, a collision is transformed into a mapping.

**Theorem 1**: The new non-MDS hash function is resistant to birthday attack on the assumption that the MPP and ASPP have only exponential time solutions.

**Proof:**

Let \(b_1\ldots b_n\) and \(b'_1\ldots b'_n\) be two arbitrary different short messages, and \(\mathfrak{q}_1\ldots \mathfrak{q}_n\) and \(\mathfrak{q}'_1\ldots \mathfrak{q}'_n\) be two related bit-long-shadow strings.

Suppose that \(\mathfrak{g}' = \mathfrak{g}\), namely \(\prod_{i=1}^{n} C_{\mathfrak{g}'}^b = \prod_{i=1}^{n} C_{\mathfrak{g}}^b\) \((\% M)\).

Therefore, we say that the new non-MDS hash function is weakly collision-free.

Further, \(W^k \prod_{i=1}^{n} (A, W^{r(i)^{\mathfrak{g}}}b_i) = \prod_{i=1}^{n} (A, W^{r(i)^{\mathfrak{g}'}}b_i)\) \((\% M)\).

Without loss of generality, let \(\bar{k} \geq \bar{g}'\). Because \((\mathfrak{M}, \cdot)\) is an Abelian group, we have \(W^{k-g'} = \prod_{i=1}^{n} A^{\delta_i} \left(\prod_{i=1}^{n} A^{\delta_i}\right)^{-1} \((\% M)\).

Due to \(\mathfrak{M}/2 = \text{a prime or the least prime factor of } \mathfrak{M}/2 > 4n(2\bar{a} + 3)\), there is \(W^{k-g'} = \prod_{i=1}^{n} A^{\delta_i} \left(\prod_{i=1}^{n} A^{\delta_i}\right)^{-1} \((\% M)\).

\(k\in\{0, 46\}\) is a integer, \((k-g')/2^k\) is a prime, and \(W\in(1, M)\) is a component of a private key is determinate, which manifests that if \(\mathfrak{g}_1\ldots \mathfrak{g}_n\) and \(\mathfrak{g}'_1\ldots \mathfrak{g}'_n\) satisfy (1), there will be \(\mathfrak{g}' = \mathfrak{g}\).

Therefore, only if \(\Psi(b_1\ldots b_n, b'_1\ldots b'_n) = W^{2^k}\) with \(k\in\{0, 46\}\), can there exists \(\mathfrak{g}' = \mathfrak{g}\). Obviously, \(\forall (b_1\ldots b_n, b'_1\ldots b'_n) \in \{0, 1\}^n\times\{0, 1\}^n\), the probability that \(\Psi(b_1\ldots b_n, b'_1\ldots b'_n) = W^{2^k}\) is nearly \(1/2^n\).

Further, let \(\theta\) be the number of needed inputs \((b_1\ldots b_n, b'_1\ldots b'_n)\)'s to find at least a \((b_1\ldots b_n, b'_1\ldots b'_n)\)'s such that \(\Psi(b_1\ldots b_n, b'_1\ldots b'_n) = W^{2^k}\) with probability 50\%, which is equivalent to finding any two messages \(b_1\ldots b_n\) and \(b'_1\ldots b'_n\) such that \(\mathfrak{h}(b_1\ldots b_n) = \mathfrak{h}(b'_1\ldots b'_n)\) with probability 50\%. Then \(\theta\) satisfies \(1-(2^{2n-k})/\left(2^{m}\right)^2\) \approx 50\%. Through computation, find that \(\theta\) is nearly \(2^{m-1}\) with \(k\in\{0, 46\}\).

The \(2^{m-1}\) is far larger than the threshold \(1.1774\times 2^{m/2}\) for the effective birthday attack. The reason is...
that a hidden restriction is imposed on the input \((b_1, \ldots, b_n, b'_1, \ldots, b'_n)\), which is easily understood as the number of students of the class needs to be increased for finding with probability 50% any two students who have both the same birthday and the same gender.

Additionally, because a private key \(\{\mathcal{A}_i\}, \{\mathcal{E}(i)\}, W, d\) is unknown for the adversary, and the MPP is intractable, it is also infeasible that the adversary finds specific \(b_1, \ldots, b_n\) and \(b'_1, \ldots, b'_n\) such that (1) holds by utilizing the private key.

Therefore, the new non-MDS hash can be resistant to the birthday attack, and at present, its security is nearly the \(O(2^n)\) magnitude, but not \(O(2^{n/2})\).

### 4.4.4 Compression Algorithm Is Resistant to Meet-in-the-middle Attack

Meet-in-the-middle dichotomy used for attack on an intended expansion of a block cipher was first developed by Diffie and Hellman in 1977 [31]. Section 3.10 of [9] brings forth a meet-in-the-middle attack algorithm for solving a subset sum problem.

**INPUT:** a set of positive integers \(\{c_1, c_2, \ldots, c_n\}\) and a positive integer \(s\).

1. Set \(t ← \lceil n / 2 \rceil\).
2. Construct a table with entries \((\sum_{i=1}^{t} c_i b_i, (b_1, b_2, \ldots, b_t))\) for \((b_1, b_2, \ldots, b_t) ∈ (\mathbb{Z}_2)^t\).
3. Sort this table by the first component.
4. For each \((b_{t+1}, b_{t+2}, \ldots, b_n) ∈ (\mathbb{Z}_2)^{n-t}\), do the following:
   - Compute \(r ← s - \sum_{i=1}^{t} c_i b_i\) and check, using a binary search, whether \(r\) is the first component of some entry in the table;
   - If \(r = \sum_{i=1}^{t} c_i b_i\), then return (a solution is \((b_1, b_2, \ldots, b_n)\)).

**OUTPUT:** \(b_i ∈ \{0, 1\}, 1 ≤ i ≤ n\), such that \(\sum_{i=1}^{n} c_i b_i = s\), provided such \(b_i\) exist.

It is not difficult to understand that the time complexity of the above algorithm is \(O(n2^{n/2})\).

### 4.4.5 Compression Algorithm Is Resistant to Multi-block Differential Attack

The [32] and [33] show that multi-block near differential attack is effective on the iterative hash functions MD5, SHA-0, SHA-1, and SHA-256 which have multiple block-inputs and the Merkle-Damgård-Iteration structure [7][8].

It is well known that MD5, SHA-0, or SHA-1 will execute a number of rounds of inner iteration for each input block, and each round of the inner iteration consists of linear arithmetics and logic operators such as addition, shift, exclusive or etc.

The input of the new non-MDS hash function is a short message which may be treated as only one block. Its inner iteration consists of at most 2n modular multiplications which is nonlinear and intricate,
which indicates that the differential analysis of \( g = \prod_{i=1}^{n} C_i^{b_i} \) (% \( M \)) loses a basis.

Furthermore, in the new non-MDS hash, the inner nonlinear iteration leads to the fierce snowslide effect and strong noninvertibility (see Section 4.4.1), and makes it impossible to derive a set of sufficient conditions which ensure that the collision differential characteristics hold for two short messages which are expected to produce a collision.

Therefore, the new non-MDS hash is substantially distinct from the classical hashes MD5, SHA-0, SHA-1 etc, and the multi-block near differential attack suitable for the classical hashes will be utterly ineffective on the new non-MDS hash function.

### 4.4.6 Compression Algorithm Is Strongly Collision-free

Firstly, it is known from Section 4.4.2 that the new non-MDS hash function \( lh \) is weakly collision-free.

Secondly, for any arbitrary short message \( h_1 \ldots h_n \), if want to find another short message \( h_1' \ldots h_n' \) such that \( lh(h_1 \ldots h_n) = lh(h_1' \ldots h_n') \), adversaries must take \( g_1' \ldots g_n' \) from

\[
\prod_{i=1}^{n} C_i^{h_i} = \prod_{i=1}^{n} C_i^{h_i'} \quad (% \( M \)),
\]

and further acquire the bit string \( b_1' \ldots b_n' \). It is known from Section 4.4.2 that such a collision problem is computationally harder than IFP at present.

Thirdly, the new non-MDS hash is resistant to classical or efficient attacks in common use — the birthday attack, meet-in-the-middle attack, and multi-block differential attack for example.

Lastly, any subexponential time algorithm for solving the ASPP \( g = \prod_{i=1}^{n} C_i^{h_i} \) (% \( M \)) is not found yet [34], and the most efficient method of solving \( g = \prod_{i=1}^{n} C_i^{h_i} \) (% \( M \)) is brute force attack so far. The analysis manifests that the security of the new non-MDS hash gets the \( O(2^n) \) magnitude at present.

In sum, the new hash function is strongly collision-free. Further, we give a related theorem.

**Theorem 2:** If any arbitrary collision of the new non-MDS hash function can be found in subexponential time, the ASPP \( \prod_{i=1}^{n} C_i^{h_i} = 1 \) (% \( M \)) can be solved efficiently, where \( h_i \in [−n, n] \) is the difference of two bit long-shadows at the same position.

**Proof:**

According to Definition 3, it is easy to understand that for every \( b_i \), there is \( 0 \leq b_i \leq n \).

Let \( b_1 \ldots b_n \neq b_1' \ldots b_n' \neq 0 \) be two arbitrary bit strings, \( b_1 \ldots b_n \) and \( b_1' \ldots b_n' \) be respectively two corresponding bit long-shadow strings.

Again let \( \bar{h}_i = b_i - b_i' \), and then there is \( \bar{h}_i \in [−n, n] \).

Since the interval \([−n, n]\) is wider than \([0, n]\), similar to \( g = \prod_{i=1}^{n} C_i^{h_i} \) (% \( M \)), the ASPP \( \prod_{i=1}^{n} C_i^{\bar{h}_i} = 1 \) (% \( M \)) with \( \bar{h}_i \in [−n, n] \) has no subexponential time solution [34], and is only faced with brute force attack.

Assume that \( \prod_{i=1}^{n} C_i^{b_i} = \prod_{i=1}^{n} C_i^{\bar{b}_i} \) (% \( M \)) is a found collision between two arbitrary bit strings \( b_1 \ldots b_n \) and \( b_1' \ldots b_n' \) in subexponential time.

From \( \prod_{i=1}^{n} C_i^{b_i} = \prod_{i=1}^{n} C_i^{\bar{b}_i} \) (% \( M \)), we have

\[
\prod_{i=1}^{n} C_i^{b_i - \bar{b}_i} = 1 \quad (% \( M \)).
\]

Let \( \bar{h}_i = b_i - b_i' \in [−n, n] \), and then

\[
\prod_{i=1}^{n} C_i^{\bar{h}_i} = 1 \quad (% \( M \)),
\]

which means that the ASPP \( \prod_{i=1}^{n} C_i^{\bar{h}_i} = 1 \) (% \( M \)) can be solved efficiently in subexponential time. It is in direct contradiction to the fact.

Therefore, the new non-MDS hash function is strongly collision-free. 

\[\square\]

### 5 Applicability of the New Non-MDS Hash Function

The new non-MDS hash function may be applied in practice, which can be seen from three aspects.

#### 5.1 Running Time of the Compression Algorithm

Suppose that running time is measured in the number of bit operations. Then it is easy to understand that the running time of a modular multiplication is \( O(2 \log^2 M) \) bit operations.

The initialization algorithm in Section 3.1 is one-shot, and not real-time, and thus it is unnecessary to care about its running time.

---

13
In what follows, we consider the running time of the compression algorithm in Section 3.2.

Because of \( n \leq \sum_{i=1}^{n} p_i \leq 2n \) for a nonzero bit string \( b_1, \ldots, b_n \), the compression algorithm takes at most \( 2n \) modular multiplications, namely the running time of the compression algorithm is \( O((2n)2\log^2 M) = O(4n^2) \) bit operations which is relatively small.

5.2 Comparison with the Chaum-Heijst-Pfitzmann Hash

The Chaum-Heijst-Pfitzmann hash function is provably secure, and defined as follows [16]:

\[
\hat{h}(w_1, w_2) = \alpha^{w_1} \beta^{w_2} \mod p \quad \left( \{0, \ldots, q-1\}^2 \rightarrow \mathbb{Z}_p \setminus \{0\} \right),
\]

where \( w_1 \) and \( w_2 \) are the two complementary parts of a short message, \( p \) and \( q = (p-1)/2 \) are two big primes, and \( \alpha \) and \( \beta \) are two generators of the group \( (\mathbb{Z}_p^*, \cdot) \). Hence, the Chaum-Heijst-Pfitzmann function based on the difficulty of the DLP \( \beta = \alpha^x \mod p \) compresses a short message of \( 2(\lceil \log p \rceil - 1) \) bits into a digest of \( \lceil \log p \rceil \) bits.

Let \( \lceil \log p \rceil = 1024 \), and then the time complexity of computing \( \log_{\alpha} \beta \mod p \) is \( 2^{80} \) according to the subexponential time \( L_p[1/3, 1.923] \) [9], which means that the security of the Chaum-Heijst-Pfitzmann hash is the \( 2^{80} \) magnitude when \( \lceil \log M \rceil = 80 \). Besides, let the bit-length \( n = 2046 \) of a short message \( (w_1, w_2) = (b_1, \ldots, b_{1023}, b_{1024}, \ldots, b_{2046}) = b_1 \ldots b_n \neq 0 \).

Under the same security, may draw a comparison between the new non-MDS hash and the Chaum-Heijst-Pfitzmann hash.

- **Running time of the compression algorithm**
  - The Chaum-Heijst-Pfitzmann hash: \( 2(\lceil \log p \rceil^1) = 2(4(1024)^1) = 858934592 \) bit operations.
  - The new non-MDS hash: \( 4mn^2 = 4(2048)80^2 = 52428800 \) bit operations.

- **Compression rate**
  - The Chaum-Heijst-Pfitzmann hash: \( 1024 / 2046 \approx 50.05\% \).
  - The new non-MDS hash: \( 80 / 2046 \approx 3.91\% \).

- **Resisting birthday attack**
  - The number of inputs \( (w_1, w_2) \)'s needed by birthday attack on \( \hat{h}(w_1, w_2) = \alpha^{w_1} \beta^{w_2} \) \( (\% \) is about \( 2^{(\lceil \log p \rceil - 1) / 2} = 2^{512} \), larger than \( 2^{80} \) which is the security magnitude of the DLP \( \beta = \alpha^x \mod p \), and thus the Chaum-Heijst-Pfitzmann hash function cannot resist the birthday attack.
  - The number of inputs \( b_1 \ldots b_n \)'s needed by birthday attack on \( \hat{h}(b_1, \ldots, b_n) = \prod_{i=1}^{n} C_i^b \mod M \) is about \( 2^{(\lceil \log M \rceil - 1) / 2} = 2^{80} \), smaller than \( 2^{80} \) which is the security magnitude of the ASPP \( g = \prod_{i=1}^{n} C_i^b \mod M \), and thus the new non-MDS hash function can resist the birthday attack.

- **Provable security**
  - On the assumption that the DLP has a subexponential time solution, the Chaum-Heijst-Pfitzmann hash function is proved to be strongly collision-free in subexponential time.
  - Likewise, on the assumption that the ASPP has an exponential time solution, the new non-MDS hash function is also proved to be strongly collision-free in exponential time.

In summary, the new non-MDS hash has some advantages over the Chaum-Heijst-Pfitzmann one, and relatively the former may be regarded as lightweight.

5.3 Reformation of a Classical Hash Function

Because the new non-MDS hash function is resistant to birthday attack and meet-in-the-middle attack, a classical hash function of which the output is \( m \) bits, and the security is intended to be the \( 2^{m/2} \) magnitude may be reformed into a compact hash function of which the output is \( m/2 \) bits, and the security is still equivalent to the \( 2^{m/2} \) magnitude [35].

For example, let \( b_1 \ldots b_{128} \) be the output of MD5, \( y_1 \ldots y_{128} \) be its bit long-shadow string, and \( \lceil \log M \rceil = 64 \). Then, regard \( g = \prod_{i=1}^{128} C_i^b \mod M \) as the 64-bit output of the reformed MD5 with the equivalent security, where \( C_i = (A_iW^i)^{y_i} \mod M \) which is produced by the algorithm in Section 3.1.

Again for example, let \( b_1 \ldots b_{160} \) be the output of SHA-1, \( y_1 \ldots y_{160} \) be its bit long-shadow string, and \( \lceil \log M \rceil = 80 \). Then, regard \( g = \prod_{i=1}^{160} C_i^b \mod M \) as the 80-bit output of the reformed SHA-1 with the equivalent security.
The above two examples indicate that we may exchange time for space when the related security remains unchanged.

6 Conclusion

In the paper, the authors propose a new non-MDS hash function which contains the initialization algorithm and the compression algorithm, and converts a short message or a message digest of \( n \) bits into a string of \( m \) bits, where \( 80 \leq m \leq 232 \) and \( 80 \leq m \leq n \leq 4096 \).

The authors prove that both the MPP and the ASPP are computationally at least equivalent to the DLP in the same prime field, and analyze the security of the new non-MDS hash function. The analysis shows that the new non-MDS hash is computationally one-way, weakly collision-free, and strongly collision-free. Moreover, at present, any subexponential time algorithm for attacking the new non-MDS hash is not found, and its security gets be the \( O(2^m) \) magnitude.

Especially, the analysis illustrates that the new non-MDS hash function is resistant to birthday attack and meet-in-the-middle attack. By utilizing this characteristic, one can reform a classical hash function with an \( m \)-bit output and an \( O(2^{m/2}) \) magnitude security into a compact hash function with an \( m / 2 \) bit output and the equivalent security.

Simultaneously, the authors dissect the running time of compression algorithm of the new non-MDS hash function, and it is \( O(nm^2) \) bit operations.

The new non-MDS hash function opens a door to convenience for the utilization of a lightweight digital signing scheme of which the modulus length is not greater than 160 bits.

Acknowledgment

The authors would like to thank the Academicians Jiren Cai, Zhongyi Zhou, Jianhua Zheng, Changxiang Shen, Zhengyao Wei, Binxing Fang, Guangnan Ni, Andrew C. Yao, and Xicheng Lu for their important guidance, advice, and suggestions.

The authors also would like to thank the Professors Dingyi Pei, Jie Wang, Ronald L. Rivest, Moti Yung, Adi Shamir, Dingzhu Du, Mulan Liu, Huaguang Zhang, Dengguo Feng, Yixian Yang, Maozhi Xu, Hanliang Xu, Yuejia Lai, Yongfei Han, Yuyu Hu, Dongdai Lin, Chunqun Wu, Rongguan Feng, Ping Liu, Jianfeng Ma, Lusheng Chen, Wenbao Han, Bogang Lin, Lequan Min, Qibin Zhai, Hong Zhu, Renji Tao, Zhiying Wang, Quanyuan Wu, and Zhichang Qi for their important counsel, suggestions, and corrections.

References


