Finding Collisions in Block Cipher-based Iterative Hash Function Schemes Using Iterative Differential

Bety Hayat Susanti*, *Member, IAENG,* Mohammad Heading Nor Ilahi, Amiruddin Amiruddin, and Sa'aadah Sajjana Carita

***Abstract*—Hash function has a fundamental role in modern cryptography as a tool to ensure integrity services in the exchange of digital information. The hash function allows one to easily verify whether or not an input data is mapped to a given or stored hash value. One type of hash function is one that uses only messages as input values called Modification Detection Codes (MDCs). Good MDCs must meet the preimage resistance, second-preimage resistance, and collision resistance properties. One type of MDCs hash function is the Preneel-Govaerts-Vandewalle (PGV) scheme, which is one of the most common iterative MDCs utilizing block cipher as its compression function. PGV has 64 schemes for building hash functions that have the property of collision resistance, which is the difficulty of finding two different inputs that have the same hash value. Of the 64 schemes, it is claimed that there are 12 secure schemes, even though there are no formal proofs of the claim. In this study, we showed that iterative differential characteristics can be utilized for finding collision on the 12 claimed-to-be-secure schemes of PGV hash function.**

***Index Terms*—block cipher, hash function, iterative differential, MDC, PGV.**

I. Introduction

THE hash function is a function h having at least two properties i.e., compression function and ease to compute. The hash function h maps any arbitrary length of string to a fixed length string. The hash function has an important role in the world of modern information technology. The hash functions are often used in hash table, a data structure that is commonly used on computer devices for fast data retrieval. In addition, hash functions are also very useful in the field of cryptography. The hash function allows an entity to easily verify whether or not an input data is mapped to a given or stored hash value. However, if the input data are unknown, it will be very difficult to reconstruct the input data or look for alternative input data that has the same hash value. These characteristics can be used to ensure the integrity of the data that has been sent. Some of the uses of hash functions are modification detection, message authentication, digital signatures, universal functions, entropy extraction and key derivation, password hashing, data identification, key updates, proof-of-work systems, and timestamping [1].

The hash function is divided into two types based on key usage [9], [7], i.e., Modification Detection Codes (MDCs), or unkeyed hash functions, which require only messages as input values to generate hash values and Message Authentication Codes (MACs), or keyed hash functions, which require messages and keys as input to produce a hash value. In [9], Menezes *et al.* categorized iterative hash functions based on the property of the operation which consists of internal functions. The three most common categories of iterative hash functions are block cipher-based hash functions, dedicated hash functions (specifically designed for hashing), and modular arithmetic hash functions. Focusing on hash functions based on keyless block ciphers (MDCs), Preneel, Govaerts, and Vandewalle (PGV) introduced 64 ways to construct single-block-length hash functions from a block cipher. Of the 64 schemes proposed, 12 schemes were claimed to be secure (fulfilling the collision-resistance property), but these claims have not been formally proven [10]. For the ease of further writing, these 12 claimed-to-be-secure schemes of PGV are called 12 secure schemes of PGV.

Ideally, a hash function must be resistant to all cryptanalytic attacks. In theoretical cryptography, the security level of a hash function is determined using preimage resistance, second preimage resistance, and collision resistance properties [1]. Collision resistance is a condition where it is very difficult to find two different input messages m and m' with the same hash value, $h(m) = h(m')$ [1]. Menezes *et al.* [9] has stated that the hash function is a many-to-one function, where the existence of collisions is unavoidable. A collision attack against a hash function is an attempt to find two different message inputs that produce the same hash value [12]. We took 12 secure schemes of the PGV hash function to be analyzed in this study considering that the PGV hash function scheme is the most common iterative hash function scheme. We tested whether the 12 PGV schemes really fulfilled the collision resistance property by carrying out collision attacks utilizing the iterative differential characteristic information of the block cipher which is used as its compression function. A block cipher is a symmetric cryptosystem in which a block of plaintext is treated as a whole and used to produce a ciphertext block of equal length

Manuscript received April 20^{th}, 2020; revised May 18^{th}, 2021.

*Corresponding author

Bety Hayat Susanti is an Assistant Professor of Cryptographic Engineering Department, Politeknik Siber dan Sandi Negara, Jl. Raya Haji Usa, Putat Nutug, Ciseeng, Putat Nutug, Bogor, Jawa Barat 16120, Indonesia (e-mail: bety.hayat@poltekssn.ac.id, bety.hayat@bssn.go.id).

Mohammad Heading Nor Ilahi is a Research Assistant at Badan Siber dan Sandi Negara, Jalan Raya Muchtar, Bojongsari, Depok 16518, Indonesia (e-mail: mohammad.heading@bssn.go.id).

Amiruddin Amiruddin is an Associate Professor of Cyber Security Department, Politeknik Siber dan Sandi Negara, Jl. Raya Haji Usa, Putat Nutug, Ciseeng, Putat Nutug, Bogor, Jawa Barat 16120, Indonesia (e-mail: amir@poltekssn.ac.id).

Sa'aadah Sajjana Carita is a Lecturer of Cryptographic Engineering Department, Politeknik Siber dan Sandi Negara, Jl. Raya Haji Usa, Putat Nutug, Ciseeng, Putat Nutug, Bogor, Jawa Barat 16120, Indonesia (e-mail: ss.carita@poltekssn.ac.id).

[12]. Advanced Encryption Standard (AES) is a block cipher algorithm that is intended to replace DES as the approved standard for a wide range of applications both in software [3] and hardware implementation [5], [14].

In the differential cryptanalysis of the block cipher, iterative differential is often found, that is, the condition with the value of the input difference repeating in a certain round, so that the output difference value in the round is equal to the value of the input difference [8]. In this research, we used theoretical and empirical methods. The empirical method was done by applying the PRESENT block cipher as a compression function. PRESENT was designed by Bogdanov *et al.* (2007) which later became a lightweight standard algorithm established by the International Standard Organization (ISO) based on ISO / IEC 29192 in 2012 [4]. PRESENT is an algorithm that has an SPN structure with a 64-bit block size. Differential cryptanalysis on PRESENT-80 was introduced by Wang [13], who found 4 input differences that formed 4-round iterative differential on PRESENT-80 with probability of 2^{-18}. Information about 4-round iterative characteristics on PRESENT-80 can be used to find collisions of hash functions that use a 4-round PRESENT-80 PGV scheme.

In this research, we focused only on the unkeyed hash function, MDCs. We generalized to test the property of collision resistance of block cipher-based iterative hash function scheme by utilizing the iterative characteristics of the block cipher used. We tested it using theoretical and empirical methods. Theoretical method was carried out by mathematically proving 12 PGV hash function schemes to prove whether or not they still fulfil the collision resistance character if the block cipher used has an iterative differential. The empirical method was carried out by attempting collision attacks using a C programming language. We showed that the use of block ciphers that have iterative differentials such as PRESENT can provide clear indication for finding collisions. We used sample pairs of different messages m and m' as many as 2^{18}. $\Delta m = m \oplus m'$ is the input difference that match the iterative characteristics of PRESENT as described by Wang. In this study, we assumed that Δm is already known by the attackers.

II. Theoretical background

A. Modification Detection Codes (MDC)

Modification Detection Codes (MDCs) are also called manipulation detection codes. The purpose of MDCs is to provide representative images or hash values of a message. Secure MDCs should meet the following characteristics [1]:

1) Preimage resistance (one-wayness) – given an output z, it is difficult to find input message x so that $z = h(x)$.
2) Second preimage resistance (weak collision resistance) – given an input message x_1, it is difficult to find another input message x_2 where $x_1 \neq x_2$ having the same hash value, $z_1 = h(x_1) = h(x_2) = z_2$.
3) Collision resistance (strong collision resistance) – it is difficult to find two different inputs $x_1 \neq x_2$ with $h(x_1) = h(x_2)$.

B. PRESENT

PRESENT, designed by Bogdanov *et al.* in 2007 [4], is an ultra-lightweight block cipher with 64-bit block size and having iteration of 31-round. PRESENT can support the use of two key lengths, 80-bit and 120-bit. PRESENT only uses one 4-bit S-Box that is applied 16 times in each round. PRESENT has three components, i.e., AddroundKey, substitution, and permutation. For more detailed discussion about PRESENT, refer to [8].

C. PGV Hash Function Schemes

In 1993, Preneel, Govaerts, and Vandewalle (PGV) proposed synthetic approach to design a single block length hash function based on block cipher [10]. They found how to establish hash function $H : \{0,1\}^* \rightarrow \{0,1\}^n$ using a compression function $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$, which was derived from block cipher $E : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$. They proposed 64 basic ways to establish hash function based on a block cipher. They claimed that 12 of 64 schemes were secure schemes. Detailed description of the PGV hash function can be found in [10]. The function expression of the 12 secure schemes of PGV is listed in Table I.

TABLE I
FUNCTION EXPRESSIONS OF 12 SECURE PGV SCHEMES

No	Function expressions
1	$E(H_{i-1}, X_i) \oplus X_i$
2	$E(H_{i-1}, X_i \oplus H_{i-1}) \oplus X_i \oplus H_{i-1}$
3	$E(H_{i-1}, X_i) \oplus X_i \oplus H_{i-1}$
4	$E(H_{i-1}, X_i \oplus H_{i-1}) \oplus X_i$
5	$E(X_i, H_{i-1}) \oplus H_{i-1}$
6	$E(X_i, X_i \oplus H_{i-1}) \oplus X_i \oplus H_{i-1}$
7	$E(X_i, H_{i-1}) \oplus X_i \oplus H_{i-1}$
8	$E(X_i, X_i \oplus H_{i-1}) \oplus H_{i-1}$
9	$E(X_i \oplus H_{i-1}, X_i) \oplus X_i$
10	$E(X_i \oplus H_{i-1}, H_{i-1}) \oplus H_{i-1}$
11	$E(X_i \oplus H_{i-1}, X_i) \oplus H_{i-1}$
12	$E(X_i \oplus H_{i-1}, H_{i-1}) \oplus X_i$

D. Iterative Differential of PRESENT-80

In the differential cryptanalysis of the block cipher, a characteristic is called iterative if the value of input difference is equal to the value of output difference. In 1991, Biham and Shamir gave a formal definition of iterative characteristics to DES, a block cipher with a Feistel structure [2]. In 1994, Knudsen provided a general formal definition of iterative characteristics for block ciphers. A clearer discussion of the iterative characteristics of the block cipher, can refer to [8]. Information about iterative characteristics contains input difference $\Delta m = m \oplus m'$, output difference $\Delta E_k(m) = E_k(m) \oplus E_k(m')$, probability p where m and m' are different messages, and E_k is a block cipher. This information can be used to find collisions on block cipher-based PGV schemes that have iterative differential. A collision attack on hash function H of 4-round PRESENT-80-based PGV schemes requires input difference that satisfies $\Delta H = H(m) \oplus H(m') = 0$. The collision attack in this study utilized iterative characteristics found by Wang for iterative differential 4-round PRESENT so that we can find $H(m) = H(m')$ with probability 2^{-18}.

III. Research method

This experiment was conducted by implementing 4-round PRESENT-80 block cipher into 12 secure schemes of PGV

using C programming language and Dev C++ Compiler. Then, we conducted collision-resistance tests on the 12 secure schemes of PGV hash function using chosen messages m and m' that satisfied Wangs iterative characteristics [8]. The number of chosen message pairs needed in this attack was 2^{18}. The simulation attack was conducted through the following steps:

1) Create simulations of 12 secure schemes of PGV hash function based on 4-round PRESENT-80.
2) Generate 2^{18} chosen message pairs m and m' with hash difference value $\Delta m = m \oplus m'$ as Wang explained, i. e., $0000\ 0000\ 0000\ 4004_{16}$, $0000\ 0101\ 0000\ 0000_{16}$, $0000\ 0009\ 0000\ 0009_{16}$, and $0500\ 0000\ 0000\ 0500_{16}$. The method of generating samples in this study was the same as the method by Ilahi *et al.* [6] but the generated samples was different. Using our own samples, we can find collisions in full for the four input differences, whereas using Ilahi *et al.*'s samples, we cannot find collisions for all four input differences in the same way.
3) Calculate the hash value of those 2^{18} chosen message pairs with 12 secure PGV schemes.
4) Do XOR operation for all the hash value pairs.
5) Check if the XOR result is $0000\ 0000\ 0000\ 0000_{16}$. This means that we have found a collision from that message pair.

IV. Results and discussion

In this section, we describe two methods used for conducting collision attacks.

A. Theoretical method

This subsection describes the mathematical proof of collision attacks on 12 variants of PGV hash function schemes based on any block cipher algorithm having iterative characteristics. The following is a formal definition of iterative characteristics for block ciphers introduced by Knudsen [8].

Definition IV.1. *[8] For an iterative block cipher, an s-round iterative characteristics is an s-tuple $(\Delta C_i, \ldots, \Delta C_{i+s})$ with $\Delta C_i = \Delta C_{i+s}$.*

In Definition IV.1, ΔC is the difference value and i is the round state. Based on the definition, Knudsen explained that if a block cipher has an s-round iterative characteristic, then the difference value of the output will repeat with period s. In other words, the input difference (difference value before entering the 1^{st} round) will be equal to the output difference in the round s. Biham and Shamir have provided a definition of iterative characteristics, but is limited to block ciphers with Feistel structures [2]. Not only for Feistel structure, Knudsen also provided a more general definition of iterative differential or iterative characteristics (see Definition IV.1).

In fact, an iterative characteristic is found using several ways. Wang found four iterative characteristics for PRESENT-80 (SPN) but without giving an explanation of how he found it. In general, the search for iterative characteristics in a block cipher is done by trying all possible input differences and seeing whether the input difference value recurs in a certain round. However, Setianingsih explained that to find the iterative characteristics of 1- to 5-round PRESENT, she used inputs that caused one active s-box and two active s-boxes. Then, look for the input difference in the active s-box where those that can cause repetition of the difference value according to the Differential Distribution Table (DDT) is called iterative characteristics [11]. In general, if a block cipher algorithm has an iterative differential, a differential cryptanalysis can be performed against the block cipher algorithm, similar to the differential cryptanalysis conducted by Wang against PRESENT. Collision on an iterative hash function based on a block cipher with a certain structure can be searched by utilizing the iterative differential of the block cipher. Theorem IV.2 explains the relationship between the iterative characteristics of the block cipher and the search for collisions in a block cipher-based hash function scheme.

Theorem IV.2. *Let E_k be a block cipher algorithm, k be an encryption key, and x be an input of block cipher algorithm derived from m or $m\oplus$ initial value (IV) where m is a message. Let $feedforward$ be a value of one of the m or $m \oplus IV$. Given a block cipher-based hash function $f(x) = E_k(x) \oplus feedforward$. If there is an s-round iterative differential in E_k with probability p and input difference $\Delta x = x \oplus x'$ is an iterative characteristic of E_k where $x \neq x'$, then $f(x) = f(x')$.*

Proof: It will be proven that $f(x) = f(x')$ with probability p if there is an s-round iterative differential in E_k. Based on Definition IV.1, an s-round iterative differentials in E_k have a condition of $(E_k(x) \oplus E_k(x')) = (x \oplus x')$ with probability p. The condition of $f(x) = f(x')$ is equivalent to $f(x) \oplus f(x') = 0$. So, it will be proven that $f(x) \oplus f(x') = 0$ with probability p. Block cipher and feedforward of $f(x)$ are mutually independent, where the variation of feedforward will not affect the block cipher operations. Thus, the probability of the collision occurrence equals to that of the iterative differential occurrence, that is, p.
Based on Definition IV.1, we have the following result. $f(x) \oplus f(x') = E_k(x) \oplus feedforward \oplus E_k(x') \oplus feedforward = (E_k(x) \oplus E_k(x')) \oplus (feedforward \oplus feedforward) = (E_k(x) \oplus E_k(x'))$.
We consider two feedforward cases to prove the occurrence of collision, as the following:

1) $feedforward = m$
   $f(x) \oplus f(x') = ((E_k(x) \oplus E_k(x')) \oplus (m \oplus m')$ based on Definition IV.1, we get $f(x) \oplus f(x') = 0$ with probability p.
2) $feedforward = m \oplus IV$
   $f(x) \oplus f(x') = ((E_k(x) \oplus E_k(x')) \oplus ((m \oplus IV) \oplus (m' \oplus IV))$ based on Definition IV.1, we get $f(x) \oplus f(x') = 0$ with probability p.

Thus, there is a collision, $f(x) = f(x')$ with probability p. ■

B. Classification of 12 secure schemes of PGV hash function

Based on Theorem IV.2, if the block cipher used as a compression function in 12 PGV schemes has iterative differential with probability p, it is expected to find collisions for message pairs that have a difference value according to the iterative characteristics of the block cipher with probability

p or with the number of chosen message pairs as many as $\frac{1}{p}$. Actually, there are three feedforward possibilities, i.e., m, IV, or $m \oplus IV$. According to the hypothesis stated in Theorem IV.2, collisions will not be found if the PGV hash function scheme uses feedforward IV because the absence of the element m will cause the input difference not suitable for forming iterative differential.

We formed 4 groups in classifying the 12 secure schemes of PGV hash function i.e., General-Scheme1, General-Scheme2, General-Scheme3, and General-Scheme4. The classification is based on structural analysis in each scheme. If we look carefully, there is a structural pattern in the 12 PGV schemes that can distinguish the schemes based on the plaintext input and encryption keys for E_k as well as the $feedforward$ form that builds the scheme.

The following are the detailed description of the 4 groups of the classification of 12 secure schemes of PGV hash functions based on plaintext input and encryption keys for E_k. Please note that in each of the schemes, H_i is the i^{th} PGV hash function, m and m' are different pair of messages, message is the input for the hash function, and the plaintext is the input for the block cipher (the value before entering the compression function, E).

a. **General-Scheme1** is given in Figure 1 with m as a plaintext input and IV as the fixed encryption key for E_k.

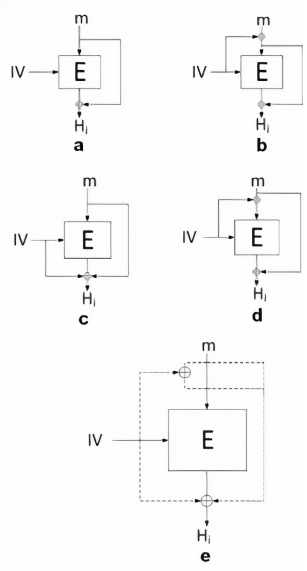

Fig. 1. a. PGV-1 scheme, b. PGV-2 scheme, c. PGV-3 scheme, d. PGV-4 scheme, and e. General-Scheme1.

$$\mathrm{H}_1(m) = E_{IV}(m) \oplus m \quad (1)$$

The PGV-1 scheme uses m as the plaintext and IV as an encryption key of block cipher E (see Figure 1.a.). In the PGV-1 scheme, m is encrypted with E_{IV} so that it becomes $E_{IV}(m)$. After that, $E_{IV}(m)$ is XORed with m. Function expressions for the PGV-1 scheme can be seen in Equation (1). Furthermore, it will be proven that there is a collision in the PGV-1 hash function scheme based on the s-round block cipher E_k, if E_k used has an s-round iterative differential.

Remember that the requirement to carry out collision attacks using an iterative differential approach is that there is the difference in input m corresponds to the iterative characteristics of E_k. Collision is found when with two different messages m and m', the hash value is equal to $H_1(m) = H_1(m')$. In other words, the difference in the hash value of both messages is $\Delta H_1 = H_1(m) \oplus H_1(m') = 0$. In the first iteration, all PGV schemes require initial value IV. In this scheme, IV is used as the encryption key for the block cipher E, so that

$$\begin{aligned}\Delta H_1 &= H_1(m) \oplus H_1(m') \\ &= E_{IV}(m) \oplus m \oplus E_{IV}(m') \oplus m' \\ &= (E_{IV}(m) \oplus E_{IV}(m')) \oplus (m \oplus m').\end{aligned}$$

If the input difference $\Delta m = m \oplus m'$ satisfies the criteria according to the iterative characteristics of the block cipher E_{IV}, then $E_{IV}(m) \oplus E_{IV}(m') = (m \oplus m')$ with probability p, we get $\Delta H_1 = 0$. Thus, it can be concluded that using an iterative differential approach, collision can be found in the PGV-1 hash function scheme with the number of chosen message pairs as many as $\frac{1}{p}$.

The structure of the PGV-2 scheme is shown in Figure 1.b., it appears that the feedforward used by this scheme is different from the PGV-1 scheme.

$$\mathrm{H}_2(m) = E_{IV}(m \oplus IV) \oplus m \oplus IV \quad (2)$$

IV is used as an encryption key and $(m \oplus IV)$ is used as a plaintext input for $E_{k=IV}$ to produce a ciphertext $E_{IV}(m \oplus IV)$. Equation (2) is the function expression for the PGV-2 scheme. The proof for the PGV-2 scheme is similar to the proof for the PGV-1 scheme. Collision is found when different messages m and m' have the same hash value $H_2(m) = H_2(m')$. In other words, the difference between the two hash values is $\Delta H_2 = H_2(m) \oplus H_2(m') = 0$. Thus,

$$\begin{aligned}\Delta H_2 &= H_2(m) \oplus H_2(m') \\ &= E_{IV}(m \oplus IV) \oplus m \oplus IV \oplus E_{IV}(m' \oplus IV) \oplus m' \oplus IV \\ &= (E_{IV}(m \oplus IV) \oplus E_{IV}(m' \oplus IV)) \oplus (m \oplus IV) \oplus (m' \oplus IV).\end{aligned}$$

Although the message input m is XORed with IV, if the input difference $\Delta m = m \oplus m'$ satisfies the criteria according to iterative characteristics with probability p, then there is still $E_{IV}(m \oplus IV) \oplus E_{IV}(m' \oplus IV) = (m \oplus IV) \oplus (m' \oplus IV)$ with probability p. This is because the difference value will not change even though each m and m' is XORed with IV, so $\Delta H_2 = 0$. Thus, it can be concluded that using an iterative differential approach, a collision can be found in the PGV-2 hash function scheme with the

number of chosen message pairs as many as $\frac{1}{p}$.

Figure 1.c. shows the structure of the PGV-3 scheme. It can be seen that IV is used as the encryption key and m is used as a plaintext input for $E_{k=IV}$, to produce a ciphertext $E_{IV}(m)$. Thereafter, to produce the hash value, $E_{IV}(m)$ is XORed with m and IV as expressed in Equation (3). It will be proven that there is a collision in the PGV-3 block cipher-based hash function scheme, s-round E_{IV}, if E_{IV} which is used has an s-round iterative differential.

$$\mathrm{H}_3(m) = E_{IV}(m) \oplus m \oplus IV \quad (3)$$

Similar to the proof of PGV-1 and PGV-2, collision is found when different messages m and m' have the same hash value $H_3(m) = H_3(m')$. In other words, the difference between the hash values is $\Delta H_3 = H_3(m) \oplus H_3(m') = 0$.
The PGV-3 scheme has a similar structure to PGV-1 but the difference is that there is an additional XOR with IV after the encryption process using E_{IV}. In the PGV-3 scheme, IV is used as a key, so that

$$\begin{aligned}\Delta H_3 &= H_3(m) \oplus H_3(m') \\ &= E_{IV}(m) \oplus m \oplus IV \oplus E_{IV}(m') \oplus m' \oplus IV \\ &= (E_{IV}(m) \oplus E_{IV}(m')) \oplus (m \oplus m') \oplus (IV \oplus IV).\end{aligned}$$

If the input difference $\Delta m = m \oplus m'$ satisfies the iterative characteristics of E_k, then $E_{IV}(m) \oplus E_{IV}(m') = (m \oplus m')$ with probability p. Consequently, with the probability p, there exists $\Delta H_3 = 0 \oplus (IV \oplus IV) = 0$.
This showed that using an iterative differential approach, a collision can be found in the PGV-3 hash function scheme with the number of chosen message pairs as many as $\frac{1}{p}$.

Figure 1.d. shows that the PGV-4 scheme uses $(m \oplus IV)$ as a plaintext input and IV is used as the encryption key of E. In the PGV-4 scheme, $(m \oplus IV)$ is encrypted with $E_{k=IV}$ so it becomes $E_{IV}(m \oplus IV)$. Next, to produce the hash value, $E_{IV}(m \oplus IV)$ is XORed with m. The function of the PGV-4 is given in Equation (4). The structure of the PGV-4 scheme is similar to the PGV-2 scheme, but the difference is that to produce the hash value of the PGV-4 scheme, we did not XOR $E_{IV}(m \oplus IV)$ with $m \oplus IV$ instead we use m.

$$\mathrm{H}_4(m) = E_{IV}(m \oplus IV) \oplus m \quad (4)$$

Proving a collision in the PGV-4 hash function scheme based on the s-round block cipher $E_{k=IV}$ is similar to that of the PGV-2 scheme. If E_{IV} used has an s-round iterative differential, it is expected that the PGV-4 s-round-based scheme has collisions. Input difference Δm must be known so that collision attacks using an iterative differential approach can be carried out. Collision is found when using different messages m and m', we obtain the same hash value $H_4(m) = H_4(m')$. In other words, the difference in both hash values is $\Delta H_4 = H_4(m) \oplus H_4(m') = 0$.

$$\begin{aligned}\Delta H_4 &= H_4(m) \oplus H_4(m') \\ &= E_{IV}(m \oplus IV) \oplus m \oplus E_{IV}(m' \oplus IV) \oplus m' \\ &= (E_{IV}(m \oplus IV) \oplus E_{IV}(m' \oplus IV)) \oplus (m \oplus m').\end{aligned}$$

If the input difference $\Delta m = m \oplus m'$ satisfies the iterative characteristic criteria of $E_{k=IV}$ with probability p, then there is $E_{IV}(m \oplus IV) \oplus E_{IV}(m' \oplus IV) = (m \oplus m')$ with probability p. This is because the difference value will not change even though each m and m' is XORed with IV, so that $\Delta H_4 = 0$. Thus, it can be concluded that using an iterative differential approach, a collision can be found in the PGV-4 hash function scheme with the number of chosen message pairs as many as $\frac{1}{p}$.

b. **General-Scheme2** is given in Figure 2 where IV as a plaintext input and m as the fixed encryption key for E_k.
In Figure 2.e., General-Scheme2 is given based on the plaintext input and encryption key for E_k. In this general scheme, no chosen message pairs are collided because although it looks similar to General-Scheme1, General-Scheme2 uses IV as a plaintext input for E_k. In addition, m and m' which are two different messages are used as keys for the block cipher E_k, giving rise to an unbalanced comparison. The PGV scheme that has a common form like this is the PGV-5, 6, 7, and 8 schemes.
This scheme uses IV as an input for E_k encryption, so that it can be ascertained to have an input difference $\Delta IV = IV \oplus IV' = 0$. In other words, the two plaintext inputs for E_k have the same value, IV, so that the PGV hash function having this general structure will not find collisions using collision attacks through an iterative differential approach. Even though E_k used has an iterative differential, the use of the input difference $\Delta m = m \oplus m'$ that matches the iterative characteristics of E_k is not used properly. Actually, m and m' are used as an encryption key for E_k.
The structure of the PGV-5 scheme is shown in Figure 2.a.

$$\mathrm{H}_5(m) = E_m(IV) \oplus IV \quad (5)$$

The PGV-5 scheme is similar to the PGV-1 scheme, but the inputs used in the two schemes are different. PGV-5 scheme uses IV as a plaintext input and m as an encryption key of E. After encrypted using the m and key, then the ciphertext is formed $E_m(IV)$. To generate a hash value,

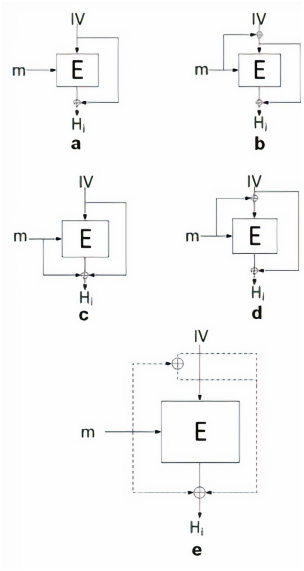

Fig. 2. a. PGV-5 scheme, b. PGV-6 scheme, c. PGV-7 scheme, d. PGV-8 scheme, and e. General-Scheme2.

$E_m(IV)$ was XORed with IV. Equation (5) shows the function expression of the PGV-5 scheme. Next, it will be proven that there is no collision in the PGV-5 hash function scheme based on the s-round block cipher E_k if the E_k used has an s-round iterative differential. Collisions are found if with different messages m and m', we get the same hash value, $H_5(m) = H_5(m')$. In other words, the difference in the hash value of both messages is $\Delta H_5 = H_5(m) \oplus H_5(m') = 0$.

$$\begin{aligned}\Delta H_5 &= H_5(m) \oplus H_5(m') \\ &= E_m(IV) \oplus IV \oplus E_{m'}(IV) \oplus IV \\ &= (E_m(IV) \oplus E_{m'}(IV)) \oplus (IV \oplus IV).\end{aligned}$$

The use of the input difference that has an iterative characteristic $\Delta m = m \oplus m'$ has no effect because the chosen message pair for an s-round block cipher E_k is equal to IV, so it has an input difference value $\Delta IV = IV \oplus IV = 0$. The message input is the same, namely IV, but the encryption key pair used in E_k is different,i.e., m and m'. The use of a different encryption key for the same message clearly results in a different hash value. This results in $(E_m(IV) \oplus E_{m'}(IV)) \neq (IV \oplus IV)$, so that $\Delta H_5 \neq 0$. Thus, it can be concluded that with the number of chosen message pairs sample of $\frac{1}{p}$ using an s-round iterative differential of block cipher E_k, no collisions found in the PGV-5 hash function scheme based on the s-round block cipher E_k.

The structure of the PGV-6 scheme is shown in Figure 2.b. It will be proven that there is no collision in the PGV-6 hash function scheme based on the s-round block cipher $E_{k=m}$ if E_k has an iterative differential s-round. The PGV-6 scheme has a similar structure to PGV-2 scheme (see Figure 1.b). Similar to the case of the PGV-1 and PGV-5 schemes, the difference between the PGV-2 and PGV-6 schemes is that in the PGV-2 scheme, m is used as message input and IV as the encryption key of $E_{k=IV}$, the PGV-6 scheme uses IV as message input while m is used as the encryption key. The function expression of the PGV-6 scheme is shown in Equation (6).

$$\mathrm{H}_6(m) = E_m(m \oplus IV) \oplus m \oplus IV \quad (6)$$

The requirement to carry out a collision attack using an iterative differential approach is to know the difference input, Δm according to the iterative characteristics of E_m. Collisions are found if with different messages m and m', we obtain the same hash value, $H_6(m) = H_6(m')$. In other words, the difference between the two hash values is $\Delta H_6 = H_6(m) \oplus H_6(m') = 0$.

$$\begin{aligned}\Delta H_6 &= H_6(m) \oplus H_6(m') \\ &= E_m(m \oplus IV) \oplus m \oplus IV \oplus E_{m'}(m' \oplus IV) \oplus m' \oplus IV \\ &= (E_m(m \oplus IV) \oplus E_{m'}(m' \oplus IV)) \oplus (m \oplus IV) \oplus (m' \oplus IV).\end{aligned}$$

$\Delta m = m \oplus m'$ is the input difference which corresponds to the iterative characteristics of $E_{k=m}$ with probability p. The message input in the PGV-6 scheme is $(m \oplus IV)$ and $(m' \oplus IV)$ so that both have a difference value according to the iterative characteristics of E_m, namely

$$\begin{aligned}\Delta m &= (m \oplus IV) \oplus (m' \oplus IV) \\ &= (m \oplus m') \oplus (IV \oplus IV) \\ &= (m \oplus m') \oplus 0 \\ &= m \oplus m'.\end{aligned}$$

Even though, it has a difference value according to the iterative characteristics of E_m, the encryption keys used are different, namely m and m', so that the parameters to form an iterative differential in the s-round E_m is not fulfilled. This results in $(E_m(m \oplus IV) \oplus E_{m'}(m' \oplus IV)) \neq (m \oplus IV) \oplus (m' \oplus IV)$, so that $\Delta H_6 \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher $E_{k=m}$, collisions are not found in the PGV-6 based s-round hash function scheme with the number of chosen message pairs as many as $\frac{1}{p}$.

The structure of the PGV-7 scheme is shown in Figure 2.c., it can be seen that the PGV-7 scheme looks similar to the PGV-3 scheme but the use of IV and m as the input between the two is reversed. This case is the same as the difference between the PGV-1 with PGV-5 schemes and the PGV-2 with PGV-6 schemes. In the PGV-7 scheme, to produce a hash value, $E_m(IV)$ is XORed with $(m \oplus IV)$. The function expression of the PGV-7 scheme is shown in Equation (7).

$$\mathrm{H}_7(m) = E_m(IV) \oplus m \oplus IV \quad (7)$$

It will be proven that there is no collision in the PGV-7 hash function scheme based on the s-round block cipher $E_{k=m}$ if E_k has an iterative differential s-round. The input difference, Δm which corresponds to the iterative characteristics of s-round block cipher E_m must be known so that collision attacks using an iterative differential approach can be carried out. Collisions are found when with different messages m and m', the same hash value is obtained $H_7(m) = H_7(m')$. In other words, the difference between the two hash values is $\Delta H_7 = H_7(m) \oplus H_7(m') = 0$. The PGV-7 scheme uses IV as message input and m is used as the encryption key, so that

$\Delta H_7 = H_7(m) \oplus H_7(m')$
$= E_m(IV) \oplus m \oplus IV \oplus E_{m'}(IV) \oplus m' \oplus IV$
$= (E_m(IV) \oplus E_{m'}(IV)) \oplus (m \oplus m') \oplus (IV \oplus IV)$.

The use of input difference, according to the iterative characteristics, $\Delta m = m \oplus m'$ has no effect because the chosen message pair used is the same, namely IV so that the difference value does not match Δm. The chosen message pairs that have a difference value according to the iterative characteristics, i.e., m and m' which should be used as plaintext input are used as encryption keys so that an iterative differential is not formed. Also, using different encryption keys for the same message will obviously result in different hash values, this results in $(E_m(IV) \oplus E_{m'}(IV)) \neq (IV \oplus IV)$, so that $\Delta H_7 \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher $E_{k=m}$, collisions are not found in the PGV-7 hash function scheme based s-round E_m with the number of chosen message pairs as many as $\frac{1}{p}$.

The structure of the PGV-8 scheme is shown in Figure 2.c. It can be seen in Figure 2.c, after $(m \oplus IV)$ is encrypted using $E_{k=m}$, it will produce ciphertext $E_m(m \oplus IV)$. To generate hash values using the PGV-8 scheme, $E_m(m \oplus IV)$ is XORed with IV. The function expression for the PGV-8 scheme can be seen in Equation (8). Next, it will be proven that there is no collision in the PGV-8 hash function scheme based on the s-round block cipher $E_{k=m}$ if E_m used has an iterative differential s-round.

$$\mathrm{H}_8(m) = E_m(m \oplus IV) \oplus IV \quad (8)$$

The PGV-8 scheme has a schematic structure similar to that of the PGV-4. Similar to the case of the PGV-2 and PGV-6 schemes, the difference between the PGV-4 and PGV-8 schemes is that in the PGV-4 scheme m is used as message input and IV as the encryption key, while in the PGV-8 scheme IV is used as message input while m is used as an encryption key. Actually, the PGV-8 scheme has similarities with the PGV-2, PGV-4, and PGV-6 schemes, namely the message input for encryption is $m \oplus IV$ so that there is a similarity in the method of proof. The requirement to carry out a collision attack using an iterative differential approach is that the input difference input is known according to the iterative characteristics of $E_{k=m}$. Collisions are found when from different messages m and m', the same hash value is obtained $H_8(m) = H_8(m')$, in other words, the difference between the two hash values is $\Delta H_8 = H_8(m) \oplus H_8(m') = 0$.

$\Delta H_8 = H_8(m) \oplus H_8(m')$
$= E_m(m \oplus IV) \oplus IV \oplus E_{m'}(m' \oplus IV) \oplus IV$
$= (E_m(m \oplus IV) \oplus E_{m'}(m' \oplus IV)) \oplus (IV \oplus IV)$

Similar to the PGV-6 proof, $\Delta m = m \oplus m'$ is the input difference which corresponds to the iterative characteristics of E_m with probability p. The message input in the PGV-8 scheme is $(m \oplus IV)$ and $(m' \oplus IV)$ so that both have a difference value in accordance with the iterative characteristics of $E_{k=m}$, namely $\Delta m = (m \oplus IV) \oplus (m' \oplus IV) = (m \oplus m') \oplus (IV \oplus IV) = (m \oplus m') \oplus 0 = m \oplus m'$. Even though it has a difference value according to the iterative characteristics of $E_{k=m}$, the use of different encryption keys m and m' causes the parameters to form an iterative differential of s-round block cipher E_m are not fulfilled. The encryption key used must be the same because if the keys are different, it will create an unbalanced ratio between $H_8(m)$ and $H_8(m')$, this results in $(E_m(m \oplus IV) \oplus E_{m'}(m' \oplus IV)) \neq (IV \oplus IV)$, so $\Delta H_8 \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher $E_{k=m}$, collisions are not found in the PGV-8 hash function scheme with the number of chosen message pairs as many as $\frac{1}{p}$.

c. **General-Scheme3** is given in Figure 3 where m as a plaintext input and $m \oplus IV$ as the fixed encryption key for E_k.

In General-Scheme3, there cannot be found pairs of collided chosen messages. The PGV schemes that have a common form like this are the PGV-9 and PGV-11. Unlike General-Scheme1 and

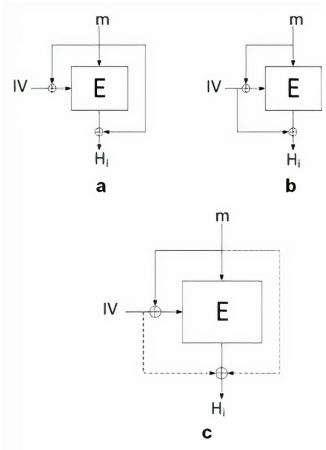

Fig. 3. a. PGV-9 scheme, b. PGV-11 scheme, and c. General-Scheme3.

General-Scheme2, General-Scheme3 has feedforward which causes the encryption key of E_k not to be fixed. The input for encryption key of E_k comes from the value of XORing the IV and the message, so the encryption key of E_k will always change depending on the message, $m \oplus IV$. This scheme uses m as a plaintext input for E_k encryption, so that it can be ascertained that iterative differential can be used properly. However, because the key of E_k is not fixed, different messages m and m', of course, will be processed with a different encryption key of E_k. Thus, $H(m) \neq H(m')$.
Figure 3.a. shows the scheme of the PGV-9 scheme. It can be seen that in general, the structure of the PGV-9 scheme differs from the PGV-1 to PGV-8. There is an XOR function before the encryption key is used, i.e., IV is XORed with m so that the encryption key is $(m \oplus IV)$. In this scheme, m is used as a plaintext input for $E_{(m \oplus IV)}$ to produce a ciphertext $E_{m \oplus IV}(m)$. Equation (9) shows the function expression of the PGV-9 scheme.

$$\mathrm{H}_9(m) = E_{m \oplus IV}(m) \oplus m \quad (9)$$

Next, it will be proven that there is no collision in the PGV-9 hash function scheme if E_k used has an s-round iterative differential. In order for the attack to succeed, the input difference Δm corresponding to the iterative characteristics of E_k must be known. Collisions are found if with different messages m and m', the same hash value is obtained $H_9(m) = H_9(m')$. In other words, the difference in the hash value of both messages is $\Delta H_9 = H_9(m) \oplus H_9(m') = 0$. In addition to the existence of feedforward to form the key for encryption, the PGV-9 scheme also has other similarities to the PGV-1 scheme, that is after m is encrypted into ciphertext, to produce a hash value, the ciphertext is XORed again with m.

$$\begin{aligned}\Delta H_9 &= H_9(m) \oplus H_9(m') \\ &= E_{m \oplus IV}(m) \oplus m \oplus E_{m' \oplus IV}(m') \oplus m' \\ &= (E_{m \oplus IV}(m) \oplus E_{m' \oplus IV}(m')) \oplus (m \oplus m').\end{aligned}$$

Although the message input for s-round E_k, i.e., m and m' has a different value according to the iterative characteristics of E_k, the two encryption keys used $(m \oplus IV)$ and $(m' \oplus IV)$ are different, so the parameters for forming an iterative differential at s-round E_k is not fulfilled. Remember that the encryption key used must be the same, if the key is different then there will be an unbalanced comparison between $H_9(m)$ and $H_9(m')$, and this results in $(E_{(m \oplus IV)}(m) \oplus E_{(m' \oplus IV)}(m')) \neq (m \oplus m')$, so that $\Delta H_9 \neq 0$. Thus, it can be concluded that using an iterative differential approach in the s-round block cipher E_k, collision is not found in the PGV-9 hash function scheme based on the s-round block cipher E_k with the number of chosen message pairs as many as $\frac{1}{p}$.

The PGV-11 scheme is slightly different from the previous PGV scheme (PGV-1 to PGV-10). In the PGV-1 to PGV-10 schemes, to generate a hash value, there is a plaintext input or a message that will be XORed with ciphertext while in the PGV-11 scheme there is no such thing. The structure of the PGV-11 scheme is shown in Figure 3.b.

$$\mathrm{H}_{11}(m) = E_{m \oplus IV}(m) \oplus IV \quad (10)$$

To generate a hash value in the PGV-11 scheme, IV is XORed with $E_{m \oplus IV}(m)$, the function expression is shown in Equation 11. Next, it will be proven that there is no collision in the block cipher-based PGV-11 hash function scheme s-round $E_{k=m \oplus IV}$ if $E_{m \oplus IV}$ is used which has an s-round iterative differential. For the attack to be successful, the input difference Δm according to the iterative characteristics of $E_{k=m \oplus IV}$ must be known. If there is $H_{11}(m) = H_{11}(m')$ then collision occurs for the message pairs m and m', in other words, the difference between the hash values of both messages is $\Delta H_{11} = H_{11}(m) \oplus H_{11}(m') = 0$. We used $(m \oplus IV)$ as the encryption key, so that

$$\begin{aligned}\Delta H_{11} &= H_{11}(m) \oplus H_{11}(m') \\ &= E_{m \oplus IV}(m) \oplus IV \oplus E_{m' \oplus IV}(m') \oplus IV \\ &= (E_{m \oplus IV}(m) \oplus E_{m' \oplus IV}(m')) \oplus (IV \oplus IV)\end{aligned}$$

Even though the message input for s-round $E_{m \oplus IV}$, namely m and m' has a difference value according to the iterative characteristics of E_k, the two encryption keys used $(m \oplus IV)$ and $(m' \oplus IV)$ are different so that the parameters to form an iterative differential in the s-round $E_{m \oplus IV}$ are not fulfilled. The encryption key used must be the same, because if the key is different, it will create an unbalanced ratio between $H_{11}(m)$ and $H_{11}(m')$, this results in $(E_{m \oplus IV}(m) \oplus E_{m' \oplus IV}(m')) \neq (IV \oplus IV)$, so that $\Delta H_{11} \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher $E_{k=m \oplus IV}$, collisions are not found in the PGV-11 hash function scheme based on s-round $E_{m \oplus IV}$ with the number of chosen message pairs as many as $\frac{1}{p}$.

d. **General-Scheme4** is given in Figure 4 where IV as a plaintext input and $m \oplus IV$ as the fixed encryption key for E_k.
The PGV-10 scheme has a similar structure with the PGV-9 scheme (see Figure 4.a.), only the use of inputs is different between the two. Equation (7) is a function expression to produce a hash value using the PGV-10 scheme.

$$\mathrm{H}_{10}(m) = E_{m \oplus IV}(IV) \oplus IV \quad (11)$$

Next, it will be proven that there is no collision in the PGV-10 hash function scheme based on s-round block cipher if the E_k used has an s-round iterative differential. Attack on collisions using an iterative differential approach requires the input difference Δm according to the iterative characteristics of E_k. By definition, the collision is found if with different messages m and m', the same hash value is obtained $H_{10}(m) = H_{10}(m')$, or in other words, the difference in both hash values is $\Delta H_{10} = H_{10}(m) \oplus H_{10}(m') = 0$ as in the following equations.

$$\begin{aligned}\Delta H_{10} &= H_{10}(m) \oplus H_{10}(m') \\ &= E_{m\oplus IV}(IV) \oplus IV \oplus E_{m'\oplus IV}(IV) \oplus IV \\ &= (E_{m\oplus IV}(IV) \oplus E_{m'\oplus IV}(IV)) \oplus (IV \oplus IV).\end{aligned}$$

Actually, the PGV-10 scheme has many similarities

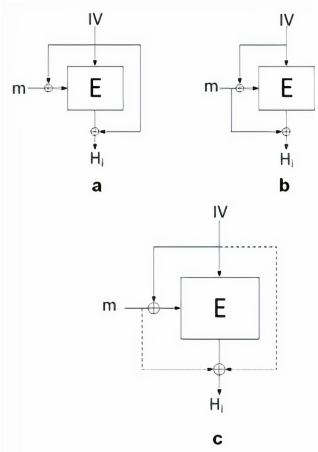

Fig. 4. a. PGV-10 scheme, b. PGV-12 scheme, and c. General-Scheme4.

with the PGV-5 scheme as well as the PGV-1 and PGV-9 schemes. The PGV-10 scheme has an addition from PGV-5, namely that the PGV-5 scheme uses IV as an encryption key for E_k while the PGV-9 scheme uses $(m \oplus IV)$ as an encryption key for E_k. The use of input difference which has an iterative characteristic $\Delta m = m \oplus m'$ has no effect because the message input for s-round block cipher E_k is the same, namely IV, so that it has an input difference $\Delta IV = IV \oplus IV = 0$. The message input is the same, namely IV, but the encryption key pair used in E_k is different, $(m \oplus IV)$ and $(m' \oplus IV)$. The use of a different encryption key for the same message will result in a different hash value, and this results in $(E_{m\oplus IV}(IV) \oplus E_{m'\oplus IV}(IV)) \neq (IV \oplus IV)$, so $\Delta H_{10} \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher E_k, collisions are not found in the PGV-10 hash function based on the s-round block cipher E_k with the number of chosen message pairs as many as $\frac{1}{p}$.

The schematic structure of PGV-12 is shown in Figure 4.b. and Equation (12) shows a function expression for generating hash values using the PGV-12 scheme. It will be proven that there is no collision in the PGV-12 hash function scheme based on the s-round block cipher $E_{k=m\oplus IV}$ if $E_{m\oplus IV}$ is used which has an s-round iterative differential.

$$\mathrm{H}_{12}(m) = E_{m\oplus IV}(IV) \oplus m \quad (12)$$

Collision attack using an iterative differential approach requires a difference input Δm according to the iterative characteristics of $E_{m\oplus IV}$. Collisions are found when with different messages m and m', the same hash value is obtained $H_{12}(m) = H_{12}(m')$, in other words the difference between the two hash values is $\Delta H_{12} = H_{12}(m) \oplus H_{12}(m') = 0$. Similar to the PGV-11 scheme, the PGV-12 scheme uses $(m \oplus IV)$ as the encryption key, so that it is obtained

$$\begin{aligned}\Delta H_{12} &= H_{12}(m) \oplus H_{12}(m') \\ &= E_{m\oplus IV}(IV) \oplus m \oplus E_{m'\oplus IV}(IV) \oplus m' \\ &= (E_{m\oplus IV}(IV) \oplus E_{m'\oplus IV}(IV)) \oplus (m \oplus m').\end{aligned}$$

The use of input difference which has iterative characteristics $\Delta m = m \oplus m'$ has no effect because the message input for the s-round block cipher $E_{m\oplus IV}$ is the same, namely IV so that it has a difference input $\Delta IV = IV \oplus IV = 0$. Both input messages are the same, namely IV but the encryption key pair used is different, $(m \oplus IV)$ and $(m' \oplus IV)$. Using different encryption keys for the same message will obviously result in different hash values, this results in $(E_{m\oplus IV}(IV) \oplus E_{m'\oplus IV}(IV)) \neq (m \oplus m)$, so that $\Delta H_{12} \neq 0$. Thus, it can be concluded that using an iterative differential approach on the s-round block cipher $E_{k=m\oplus IV}$, collisions are not found in the PGV-12 based on the s-round block cipher $E_{m\oplus IV}$ with the number of chosen message pairs as many as $\frac{1}{p}$.

C. Empirical method

Wang found four input differences to form 4-round iterative differential in PRESENT [13]. The four values of the input difference are $0000\ 0000\ 0000\ 4004_{16}$, $0000\ 0101\ 0000\ 0000_{16}$, $0000\ 0009\ 0000\ 0009_{16}$, and $0500\ 0000\ 0000\ 0500_{16}$. Wang explained that iterative differential can be formed with probability 2^{-18}. However, Wang did not explain how to get probabilities of 2^{-18}. The following is the probability search process.

Figure 5 shows the 4-round iterative differential path on PRESENT for the input difference $0000\ 0000\ 0000\ 4004_{16}$. The active bit position is indicated by a thick line. Based on the iterative differential path in Figure 5, it can be seen that the input difference is $0000\ 0000\ 0000\ 4004_{16}$ repeated in the 4^{th} round. There are three components that make up a round of PRESENT algorithm, i.e., XOR subkeys, substitution boxes (s-boxes), and permutations. Because the XOR subkey does not affect the formation of iterative differential, this component can be ignored.

There are only two influential components, i.e. substitution and permutation. The difference value can change after going through these two components. After passing

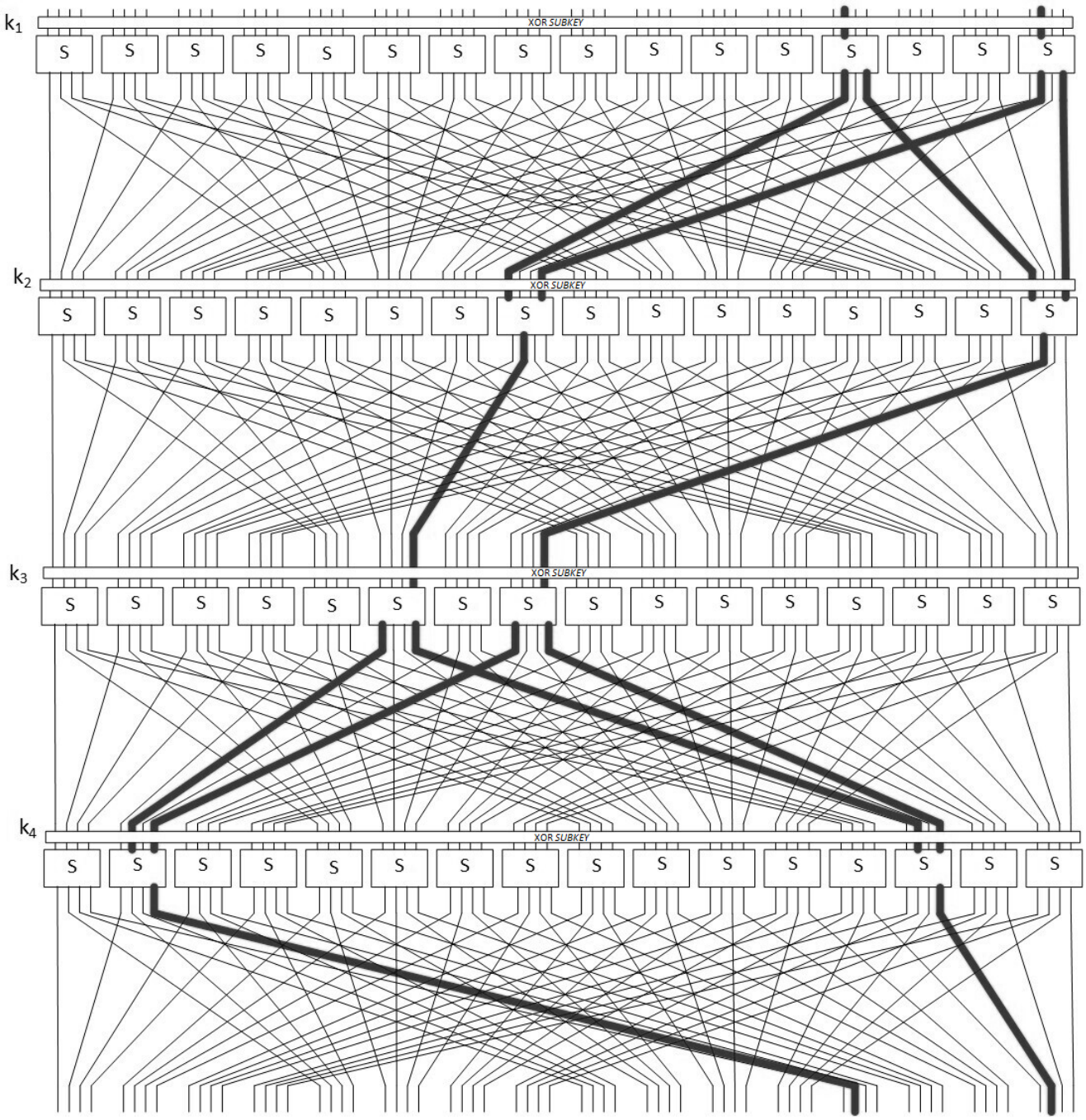

Fig. 5. Flow of iterative differential of 4-round PRESENT for input difference 0000 0000 0000 4004_{16}.

the s-box in the 1^{st} round, based on the Differential Distribution Table (DDT) in Table II the input difference is 0000 0000 0000 4004_{16} changing to 0000 0000 0000 5005_{16} with a probability of $\frac{4}{16} \times \frac{4}{16} = \frac{1}{4} \times \frac{1}{4} = \frac{1}{2^4}$. Furthermore, after passing permutation, the difference value changes to 0000 0009 0000 0009_{16} with probability 1. After passing the 1^{st} round, the input difference is 0000 0000 0000 4004_{16} will change to 0000 0009 0000 0009_{16} with a probability of $1 \times \frac{1}{2^4} = \frac{1}{2^4}$. After passing the 2^{nd} s-box round, based on DDT in Table II, the difference value changes to 0000 0004 0000 0004_{16} with a probability of $\frac{4}{16} \times \frac{4}{16} = \frac{1}{4} \times \frac{1}{4} = \frac{1}{2^4}$. Then, after passing permutation, the difference value changes to 0000 0101 0000 0000_{16} with probability 1, so that after passing the 1^{st} round and the 2^{nd} round, the input difference 0000 0000 0000 4004_{16} will change to 0000 0101 0000 0000_{16} with a total probability of $\frac{1}{2^4} \times \frac{1}{2^4} \times 1 = \frac{1}{2^8}$. After passing the 3^{rd} s-box round, based on DDT in Table II the difference value 0000 0101 0000 0000_{16} changes to 0000 0909 0000 0000_{16} with a probability of $\frac{4}{16} \times \frac{4}{16} = \frac{1}{4} \times \frac{1}{4} = \frac{1}{2^4}$. Furthermore, after passing permutation the value of the difference changes to 0500 0000 0000 0500_{16} with probability 1 so that after passing the 1^{st}, 2^{nd}, and 3^{rd} round the input difference value is 0000 0000 0000 4004_{16} changing to 0500 0000 0000 0500_{16} with probability $\frac{1}{2^8} \times \frac{1}{2^4} \times 1 = \frac{1}{2^{12}}$.

After passing the s-box in the 4^{th} round, the difference value changes to 0100 0000 0000 0100_{16} with a probability of $\frac{2}{16} \times \frac{2}{16} = \frac{1}{8} \times \frac{1}{8} = \frac{1}{2^6}$. Furthermore, after passing permuta-

TABLE II
DIFFERENTIAL DISTRIBUTION TABLE (DDT) OF S-BOX PRESENT [13]

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	4	0	0	0	4	0	4	0	0	0	4	0	0
2	0	0	0	2	0	4	2	0	0	0	2	0	2	2	2	0
3	0	2	0	2	2	0	4	2	0	0	2	2	0	0	0	0
4	0	0	0	0	0	4	2	2	0	2	2	0	2	0	2	0
5	0	2	0	0	2	0	0	0	0	2	2	2	4	2	0	0
6	0	0	2	0	0	0	2	0	2	0	0	4	2	0	0	4
7	0	4	2	0	0	0	2	0	2	0	0	0	2	0	0	4
8	0	0	0	2	0	0	0	2	0	2	0	4	0	2	0	4
9	0	0	2	0	4	0	2	0	2	0	0	0	2	0	4	0
A	0	0	2	2	0	4	0	0	2	0	2	0	0	2	2	0
B	0	2	0	0	2	0	0	0	4	2	2	2	0	2	0	0
C	0	0	2	0	0	4	0	2	2	2	2	0	0	0	2	0
D	0	2	4	2	2	0	0	2	0	0	2	2	0	0	0	0
E	0	0	2	2	0	0	2	2	2	2	0	0	2	2	0	0
F	0	4	0	0	4	0	0	0	0	0	0	0	0	0	4	4

tion, the difference value changes to $0000\ 0000\ 0000\ 4004_{16}$ with probability 1. Remember that the value of the difference after passing the 4^{th} round is the same as the input difference value. In the 4^{th} round, there is a repetition of the value of difference, the 4^{th} difference round value is equal to the value of the input difference so it forms an iterative differential 4-round with a total probability is $\frac{1}{2^{12}} \times \frac{1}{2^6} \times 1 = \frac{1}{2^{18}}$.

In summary, looking for collisions requires four steps. First, generating chosen message pairs whose difference value is the same as the four difference values explained by Wang. Second, calculate the hash value of the two chosen messages. Third, XOR the second hash value. Fourth, check the results of XOR, if the XOR result is $0000\ 0000\ 0000\ 0000_{16}$ then a collision occurs for the chosen message pair.

From the collision attack simulations, collisions were found only in schemes of PGV 1, PGV 2, PGV 3, and PGV 4. The results of these collisions can be seen in Table III. The column "Differences (Hex)" shows 4 input differences (in hexadecimal) for 4-round iterative differential PRESENT-80 as explained by Wang. Columns "m (Hex)" and "m' (Hex)" respectively show the first message and the second message. The column "$H(m)$ (Hex)" and column "$H(m')$ (Hex)" respectively show the hash value for m and m'. The underlined digits show the position of active difference. Using the sample that we generated (our sample), we found collisions for the four input differences, whereas using the sample generated by Ilahi *et al.* (Ilahi's sample), there could be no collisions for the input difference $0000\ 0101\ 0000\ 0000_{16}$ (in PGV-1 and PGV-3) and $0500\ 0000\ 0000\ 0500_{16}$ (in PGV-1, PGV-2, PGV-3, and PGV-4). The total number of collided messages pairs between what we found and those found by Ilahi *et al.* are shown in Table IV. The column "sample" shows sample ownership with two subjects compared. The column "Differences (Hex)" shows the 4 input differences for 4-round iterative differential PRESENT-80. The column "Number of collisions" shows the number of collisions found using this attack.

It can be seen in Table IV that using samples from Ilahi *et al.* [6], the total number of collisions found was 20 (twenty) pairs of messages. However, the distribution of collisions was found to be uneven. Using samples of Ilahi *et al.*'s [6], no

TABLE IV
COMPARISON OF THE NUMBER OF COLLISIONS FOUND USING OUR SAMPLE AND ILAHI *et al.*'S SAMPLE

Sample used	Differences(Hex)	♯ of collisions
Our sample	0000 0000 0000 4004	2
	0000 0101 0000 0000	2
	0000 0009 0000 0009	6
	0500 0000 0000 0500	6
Ilahi *et al.*'s sample [6]	0000 0000 0000 4004	10
	0000 0101 0000 0000	2
	0000 0009 0000 0009	8
	0500 0000 0000 0500	-

collisions can be found in schemes of PGV-1, 2, 3, and 4 for input differences of $0500\ 0000\ 0000\ 0500_{16}$. Using our sample, we found collisions for the four input differences in the PGV-1 and PGV-3 schemes. Overall, we found 2 (two) pairs of chosen messages in the PGV-1 and PGV-2 and 6 (six) pairs of chosen message collisions in the PGV-3 and PGV-4 schemes.

V. CONCLUSION

In this paper, we showed a different approach to finding collisions in 12 secure schemes of PGV hash functions. The use of block ciphers that have iterative differentials such as PRESENT can lead us to find collisions. The results of the attacks showed that out of 12 PRESENT-80 4-round PGV hash function schemes, there were 4 (four) schemes having collided message pairs so that the four schemes did not meet the collision resistance property. This means that the selection of a good block cipher to build a secure scheme of PGV hash function is very important. The use of block ciphers that have an iterative differential is vulnerable to collision attacks. Therefore, a block cipher without iterative differential is a good indication of good block ciphers.

ACKNOWLEDGMENT

This research was supported by Politeknik Siber dan Sandi Negara, Bogor, Indonesia.

REFERENCES

[1] J. Aumasson, *et al.*, "The hash function BLAKE," Springer, 2014.

TABLE III
RESULT OF COLLISIONS FOUND AFTER ATTACKS ON 12 SECURE SCHEMES OF PGV

Scheme	Differences (Hex)	$m(Hex)$	$m'(Hex)$	$H(m) = H(m')$ (Hex)
PGV 1	0000 0000 0000 4004	4859 0609 $4f01$ $\underline{1af5}$	4859 0609 $4f01$ $\underline{5af1}$	$ba1d$ $9ed0$ $130c$ $fe7c$
	0000 0101 0000 0000	$23f6$ $\underline{3495}$ 1439 $19b8$	$23f6$ $\underline{3594}$ 1439 $19b8$	$97fa$ $65a0$ $860b$ $dfde$
	0000 0009 0000 0009	1274 $\underline{7804}$ 3571 $2ac\underline{2}$	1274 $\underline{780d}$ 3571 $2ac\underline{b}$	$7e42$ 6903 $0ec4$ $09a5$
	0500 0000 0000 0500	$\underline{1607}$ $6f3d$ $50ac$ $\underline{421a}$	$\underline{1307}$ $6f3d$ $50ac$ $\underline{471a}$	$a76d$ $9e95$ 7939 $e56f$
		$\underline{2899}$ $1ff6$ $20ae$ $\underline{590a}$	$\underline{2d99}$ $1ff6$ $20ae$ $\underline{5c0a}$	$d551$ $4f77$ 1601 $534e$
		$\underline{136f}$ $352e$ $47d9$ $\underline{7712}$	$\underline{166f}$ $352e$ $47d9$ $\underline{7212}$	$de77$ $54c6$ $e400$ $be83$
PGV 2	0000 0000 0000 4004	–	–	–
	0000 0101 0000 0000	–	–	–
	0000 0009 0000 0009	7144 $63a\underline{a}$ $51b5$ $51c\underline{c}$	7144 $63a\underline{3}$ $51b5$ $51c\underline{5}$	2936 $c662$ $9e23$ $fa60$
		$719b$ $79f\underline{9}$ $4d0d$ 1839	$719b$ $79f\underline{0}$ $4d0d$ 1830	$0c1a$ $b403$ $0fd5$ $10e7$
	0500 0000 0000 0500	$6\underline{3}45$ $1c8c$ $179c$ $5\underline{6}b1$	$6\underline{6}45$ $1c8c$ $179c$ $5\underline{3}b1$	$94bd$ $1e37$ $57cf$ $440d$
PGV 3	0000 0000 0000 4004	4859 0609 $4f01$ $\underline{1af5}$	4859 0609 $4f01$ $\underline{5af1}$	$0b1c$ $81c0$ $025d$ $ef4f$
	0000 0101 0000 0000	$23f6$ $\underline{3495}$ 1439 $19b8$	$23f6$ $\underline{3594}$ 1439 $19b8$	$26fb$ $7ab0$ $975a$ $ceed$
	0000 0009 0000 0009	1274 $\underline{7804}$ 3571 $2ac\underline{2}$	1274 $\underline{780d}$ 3571 $2ac\underline{b}$	$cf43$ 7613 $1f95$ 1896
	0500 0000 0000 0500	$\underline{1607}$ $6f3d$ $50ac$ $\underline{421a}$	$\underline{1307}$ $6f3d$ $50ac$ $\underline{471a}$	$166c$ 8185 6868 $f45c$
		$\underline{2899}$ $1ff6$ $20ae$ $\underline{590a}$	$\underline{2d99}$ $1ff6$ $20ae$ $\underline{5c0a}$	6450 5067 0750 $427d$
		$\underline{136f}$ $352e$ $47d9$ $\underline{7712}$	$\underline{166f}$ $352e$ $47d9$ $\underline{7212}$	$6f76$ $4bd6$ $f551$ $afb0$
PGV 4	0000 0000 0000 4004	–	–	–
	0000 0101 0000 0000	–	–	–
	0000 0009 0000 0009	7144 $63a\underline{a}$ $51b5$ $51c\underline{c}$	7144 $63a\underline{3}$ $51b5$ $51c\underline{5}$	3926 $d773$ $9f22$ $eb71$
		$719b$ $79f\underline{9}$ $4d0d$ 1839	$719b$ $79f\underline{0}$ $4d0d$ 1830	$1c0a$ $a512$ $0ed4$ $01f6$
	0500 0000 0000 0500	$6\underline{3}45$ $1c8c$ $179c$ $5\underline{6}b1$	$6\underline{6}45$ $1c8c$ $179c$ $5\underline{3}b1$	$84ad$ $0f26$ $56ce$ $551c$

[2] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *Journal of Cryptology* vol. 4, pp. 3–72, 1991.

[3] B. Bilgin, *et al.,* "Trade-Offs for Threshold Implementations Illustrated on AES," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* vol. 34, no. 7, pp. 1188–1200, 2015.

[4] A. Bogdanov, *et al.,* "PRESENT: an ultra-lightweight block cipher," *CHES 2007*, pp. 450–466, 2007.

[5] U. Farooq and M. F. Aslam, "Comparative analysis of different AES implementation techniques for efficient resource usage and better performance of an FPGA", *Journal of King Saud University - Computer and Information Sciences* vol. 29, Issue 3, pp. 295–302, 2017.

[6] M. H. N. Ilahi, *et al.,* "Collision attack on 4 secure PGV hash function schemes based on 4-round PRESENT-80 with iterative differential approach," *16th International Conference on Quality in Research (QiR): International Symposium on Electrical and Computer Engineering*, IEEE, 2019.

[7] N. Kishore and B. Kapoor, "Attacks on and Advances in Secure Hash Algorithms," *IAENG International Journal of Computer Science* vol. 43, no.3, pp. 326–335, 2016.

[8] L. R. Knudsen, "Block ciphers - analysis, design, and applications," *Aarhus University*, 1994.

[9] A. J. Menezes, *et al., Handbook of applied cryptography*, Boca Raton, 1997.

[10] B. Preneel, *et al.,* "Hash functions based on block ciphers: a synthetic approach," *CRYPTO'93*, vol. 13, pp. 368–378, 1993.

[11] A. Y. Setianingsih, "Differential cryptanalysis of DES-like cryptosystems," *unpublished thesis*, Sekolah Tinggi Sandi Negara, 2016.

[12] W. Stallings, *Cryptography and network security : principle and practice*, 7th edn, Pearson Education, 2017.

[13] M. Wang, "Differential cryptanalysis of PRESENT," Jinan, 2007.

[14] X. Zhang, *et al.,* "Hardware Implementation of Compact AES S-box," *IAENG International Journal of Computer Science* vol. 42, no.2, pp.125–131, 2015.

Bety Hayat Susanti (M'2020) is an Assistant Professor at Politeknik Siber dan Sandi Negara. She received PhD Degree in Mathematics from Institut Teknologi Bandung in 2019. She obtained Bachelor Degree in Mathematics from Universitas Indonesia in 2000 and Master Degree in Economics from Universitas Indonesia in 2005. Her research interest is cryptography, cryptanalysis, discrete mathematics, and graph theory. She is a member of the Indonesian Combinatorial Society (InaCombS) and International Association of Engineers (IAENG).

Mohammad Heading Nor Ilahi is a Research Assistant at Badan Siber dan Sandi Negara with Bachelor Degree in Cryptographic Engineering from Sekolah Tinggi Sandi Negara in 2019. His researches are in fields of cryptography and cryptanalysis.

Amiruddin Amiruddin is an Associate Professor at Politeknik Siber dan Sandi Negara with PhD Degree in Electronics Engineering from Universitas Indonesia (2018). He obtained Bachelor Degree in Informatics from Universitas Budi Luhur in 2003 and Master Degree in Information Technology from Universitas Indonesia in 2007. His research interest is cybersecurity, network security, information technology, cryptography, data mining, and internet of things. He is a member of Asosiasi Pendidikan Tinggi Informatika dan Komputer (APTIKOM).

Sa'aadah Sajjana Carita is a Lecturer at Politeknik Siber dan Sandi Negara with Master Degree in Mathematics from University of Aix-Marseille (2016) and Institut Teknologi Bandung (2017). She obtained Bachelor Degree in Mathematics from Institut Teknologi Bandung in 2013. Her research interest is elliptic curve cryptography and post quantum cryptography. She is a member of the International Association for Cryptologic Research (IACR).