Improved Collision and Preimage Resistance Bounds on PGV Schemes

Lei Duo[1] and Chao Li[2]

[1] Department of Science, National University of Defense Technology,
Changsha, China
Duoduolei@gmail.com
[2] Department of Science, National University of Defense Technology,
Changsha, China

Abstract. Preneel, Govaerts, and Vandewalle[14](PGV) considered 64 most basic ways to construct a hash function from a block cipher, and regarded 12 of those 64 schemes as secure. Black, Pogaway and Shrimpton[3](BRS) provided a formal and quantitative treatment of those 64 constructions and proved that, in black-box model, the 12 schemes ($group - 1$) that PGV singled out as secure really are secure. By stepping outside of the Merkle-Damgård[4] approach to analysis, an additional 8 ($group - 2$) of the 64 schemes are just as collision resistant as the first group of schemes. Tight upper and lower bounds on collision resistance of those 20 schemes were given. In this paper, those collision resistance and preimage resistance bounds are improved, which shows that, in black box model, collision bounds of those 20 schemes are same. In $Group - 1$ schemes, 8 out of 12 can find fixed point easily. Bounds on second preimage, multicollisions of Joux[6], fixed-point multicollisons[8] and combine of the two kinds multicollisions are also given. From those bound, $Group - 1$ schemes can also be deviled into two group.

Key Words: Hash Function, Block Cipher, M-D Construction

1 Introduction

Most of hash functions iterated a compression function by Merkle-Damgård structure with constant IV[13]. Building hash function based on block cipher gone back to Rabin[15], wherein one makes the compression function out of a block cipher. This topic had been systematically analyzed in [3, 10, 11, 14].

Block cipher hash has been less widely used, for a variety of reasons. Black, Rogaway, and Shrimpton[3] given some fresh light on the block cipher based hash and taken a proof-centric look at the 64 block cipher based compression function iterated by Merkel-Damgård structure. First summary of those 64 schemes was presented by Preneel, Govaerts, and Vandewalle[14]. Recently, some new double length block cipher based hash functions have been recommend[7, 12].

PGV Paper PGV paper[14] considered turning a block cipher $E : \{0, 1\}^n \times$

$\{0,1\}^n \rightarrow \{0,1\}^n$ into a hash function $H : (\{0,1\}^n)^* \rightarrow \{0,1\}^n$ using a compression function $F : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ derived from E. PGV considered all 64 compression functions F of the form $F(h_{i-1}, m_i) = E_a(b) \oplus c$, where $a, b, c \in \{h_{i-1}, m_i, h_{i-1} \oplus m_i, v\}$, in which $v \in \{0,1\}^n$ is a constant. Of the 64 such schemes, the authors of [14] regarded 12 as secure. Another 13 schemes they classify as $backward - attackable$. The remaining 39 schemes are subject to damaging attacks identified by [14] and others.

BRS Paper BRS paper[3] taken a more proof-centric look at the schemes from PGV, proved additional 8 schemes were collision resistant, divided the 20 schemes into two group where the $group-1$ scheme, $\{H_1, \ldots, H_{12}\}$, was the 12 schemes picked by PGV and the $group-2$ scheme, $\{H_{13}, \ldots, H_{20}\}$, was the new founded 8 schemes. For the new founded schemes, the hash function H immune to collision attack within the Merkle-Damgård paradigm, the compression functions were not immune to collision attack, the proves of collision resistant of $group-2$ used the assumptions of E was a black box model and H with fix start model. They also gave both upper and lower bounds for each.

Our PGV Results We reanalyze those 64 schemes, improve the upper and lower bounds that were given in BRS paper, using method based on graph theory, by which, hashing procedure is considered as directed graph drawing procedure and attacking method is considered path building method on directed graph. Tabel1 is contrast of bounds between BRS and ours. Table2 considers the second preimage bounds with plain padding and MD-strengthening padding.

Table 1. Summary of results on collision and preimage resistance bounds including BRS and ours. The adversary asks at most q query. Message padding is plain padding.

Category		Collision Resistance				Preimage Resistance			
		UP Bound		Low Bound		UP Bound		Low Bound	
		BRS	Our	BRS	Our	BRS	Our	BRS	Our
Group-1: $H_{[1..12]}$ 12 schemes	$H_{[1..4]}$	$\frac{q(q+1)}{2^n}$	$\frac{q(q+1)}{2^n}$	$\frac{0.039(q-1)(q-3)}{2^n}$	$\frac{q(q+1)}{2^{n+1}}$	$\frac{q}{2^{n-1}}$	$\frac{q}{2^{n-1}}$	$\frac{0.4q}{2^{n-1}}$	$\frac{q}{2^n}$
	$H_{[5..12]}$			$\frac{0.3q(q-1)}{2^n}$				$\frac{0.6q}{2^{n-1}}$	
Group-2:$H_{[13..20]}$		$\frac{3q(q+1)}{2^n}$		$\frac{0.3q(q-1)}{2^n}$		$\frac{9(q+3)^2}{2^n}$	$\frac{q(q+1)}{2^n}$	$\frac{0.15q^2}{2^n}$	$\frac{q(q+1)}{2^{n+2}}$

Short DiCycle Multicollisions This attack is an attack similar to Multicollisions, if Multicollisions is regarded as attack using short undirected cycle to build collisions, then Short DiCycle Multicollisions is attack using short directed cycle to build collisions. This attack was first given by Kelsey and Schneier[8]. Summary of Short DiCycle multicollisions, Combine of Joux's and Kelsey's Multicollisions, that are contrast with Joux's Multicollisions are given in Tabel3.

Table 2. Summary of bounds on second preimage resistance. Adversary asks at most q query with plain padding(PlainPD) and MD-strengthening padding(MD-SP). t is the first message length.

Second Preimage Attack	UP		Low	
	PlainPD	MD-SP	PlainPD	MD-SP
$H_{[1..4]}$	$\frac{q(t+1)}{2^n}$	$\frac{q}{2^{n-1}}$	$\frac{q(t+1)}{2^{n+2}}$	$\frac{q}{2^n}$
$H_{[5..12]}$		$\frac{q(t-1)}{2^n}$		$\frac{q(t-1)}{2^{n+2}}$
$H_{[13..20]}$	$\frac{q(q+2t+3)}{2^n}$	$\frac{q(q+t)}{2^n}$	$\frac{q(q+4t+3)}{2^{n+2}}$	$\frac{q(q+2t)}{2^{n+2}}$

Table 3. Summary of Short DiCycle multicollisions, Combine Multicollisions, that are contrast with Joux's Multicollisions. Message padding is MD-strengthening padding with time consuming(Time), minimum message length(MML) and maximum collide numbers(MCN).

Multicollisons	Joux's Multicollisions			Short Dicycle Multicollisions			Combine Multicollisions		
	Time	MML	MCN	Time	MML	MCN	Time	MML	MCN
$H_{[1..4]}$	$K2^{n/2}$	K	2^K	-	-	-	-	-	-
$H_{[5..12]}$				$2K2^{n/2}$	$L+K$	$S(L,K)^a$	$3K2^{n/2}$	L+2K	$2^K S(L,K)$
$H_{[13..20]}$				$2K2^{n/2}$	$2L+K$	$S_L(K)$	$3K2^{n/2}$	2L+2K	$2^K S(L,K)$

[a] $S(L,K) = \sum_{i_L=0}^{K} \sum_{i_{L-1}=0}^{i_L} \cdots \sum_{i_2=0}^{i_3} (i_2+1)$

2 Notations and Definitions

Let message block: $\overline{m} \in \{0,1\}^n$, message: $m = \overline{m}_1 \| \ldots \| \overline{m}_i \in \cup_{\iota=1}^{t} \{0,1\}^{n \cdot \iota}$ and instance of message: $m_i \in \cup_{\iota=1}^{t} \{0,1\}^{n \cdot \iota}$. If $\overline{m}', \overline{m}'' \in \{0,1\}^n$, then $\overline{m}' \| \overline{m}'' \in \{0,1\}^{2n}$ and $|\overline{m}'| = |\overline{m}''| = n$. Let $\overline{m}_1^{(t)} = \overline{m}_1 \| \ldots \| \overline{m}_1$, where $|\overline{m}_1^{(t)}| = t \cdot n$. Let $\mathbf{0}$ be n bit $'0'$. Let a hash function algorithm $H : \mathcal{M} \rightarrow \mathcal{Y}$ with initial value IV, for any $m \in \mathcal{M}$ with $H(m, IV)$. Let Block cipher $E : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ using notation $E(x,k)$ or $E_k(x)$, key $k \in \{0,1\}^n$.

64 Schemes We consider the schemes $F(h_{i-1}, m_i) = E_a(b) \oplus c$, where $a, b, c \in \{h_{i-1}, m_i, h_{i-1} \oplus m_i, v\}$, in which block cipher $E : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$. H_ι has compression function F_ι, in which H_ι is numbered as BRS[3]. Not loosing generally, we assume $v = \mathbf{0}$.

Message Padding Take the L-bit input message m ($L < 2^{n/2}$) and append a $'1'$ followed by '0' bits such that z is the smallest positive integer satisfying ($L + 1 + z \equiv n/2 \bmod n$) and, finally, append the binary representation of the length of the original message $\mathcal{P}(m)$. We call this padding method as MD-strengthening $m\|1\|0^{(z)}\|\mathcal{P}(m)$. The padding $m\|1\|0^{(\overline{z})}$ is called plain padding, in which $\overline{z}$ is the smallest positive integer satisfying ($L + 1 + \overline{z} \equiv 0 \bmod n$).

Ideal Cipher Model[16] A block cipher with the block length n and the key length κ is called an (n, κ) block cipher. Let $E : \{0,1\}^n \times \{0,1\}^\kappa \to \{0,1\}^n$ be an (n, κ) block cipher. Then, $E(k, \cdot)$ is a permutation for every $k \in \{0,1\}^\kappa$, and it is easy to compute both $E(k, \cdot)$ and $E(k, \cdot)^{-1}$. Let $\mathcal{B}_{n,\kappa}$ be the set of all (n, κ) block ciphers. In the ideal cipher model, E is assumed to be randomly selected from $\mathcal{B}_{n,\kappa}$. The encryption E and the decryption E^{-1} are simulated by the following two oracles. The encryption oracle E first receives a pair of a key and a plaintext as a query. Then, it returns a randomly selected ciphertext. On the other hand, the decryption oracle E^{-1} first receives a pair of a key and a ciphertext as a query. Then, it returns a randomly selected plaintext.

Adversary We consider a computationally unbounded adversary with access to either a E and E^{-1}. The adversarys "running time" is determined by her number of E queries. Our adversaries are probabilistic algorithms, and we concentrate on the expected running time. We will describe the running time asymptotically.

Advantage on Collision We write $x \xleftarrow{\$} S$ for the experiment of choosing a random element from the finite set S and calling it x. An adversary is an algorithm with access to one or more oracles. We write these as superscripts. The adversary $\mathcal{A}$ with the oracle E, E^{-1} is a collision-finding algorithm of H. The advantage of $\mathcal{A}$ finding collisions in H is:
$Adv_H^{coll}(\mathcal{A}) = Pr[E \xleftarrow{\$} \mathcal{B}_{n,\kappa}; (m, m') \xleftarrow{\$} \mathcal{A}^{E,E^{-1}} : m \neq m' \wedge H(m) = H(m')]$.

Advantage on Preimage The advantage of $\mathcal{A}$ finding preimage in H is:
$Adv_H^{pre}(\mathcal{A}) = Pr[E \xleftarrow{\$} \mathcal{B}_{n,\kappa}; \delta \xleftarrow{\$} \{0,1\}^n; m \leftarrow \mathcal{A}^{E,E^{-1}(\delta)} : \delta = H(m)]$.

Advantage on Second Preimage The advantage of $\mathcal{A}$ finding second preimage in H is:
$Adv_H^{sPre}(\mathcal{A}) = Pr[E \xleftarrow{\$} \mathcal{B}_{n,\kappa}; m \xleftarrow{\$} \{0,1\}^n; m' \leftarrow \mathcal{A}^{E,E^{-1}}(m) : H(m) = H(m')]$.
For $q > 1$ let $Adv_H^{attack}(q) = \max_{\mathcal{A}}\{Adv_H^{attack}(\mathcal{A})\}$, where $\mathcal{A}$ makes at most q queries to E, E^{-1} in total, $attack \in \{coll, pre, sPre\}$.

Simulating a Ideal Cipher Oracle[3] An adversary $\mathcal{A}$ access to a simulate ideal cipher oracle for E and E^{-1}, which is defined as follows:

```
Algorithm $SimulateOracles(\mathcal{A}, n)$
Initially, $i \leftarrow 0$ and $E_k(x) = undefined$ for all $(x, k) \in \{0,1\}^n \times \{0,1\}^n$
Run $\mathcal{A}^{?,?}$, answering oracle queries as follows:
   When $\mathcal{A}$ asks a query $(x, k)$ to its left oracle:
     $i \leftarrow i + 1; k_i \leftarrow k; x_i \leftarrow x; y_i \xleftarrow{\$} \overline{Range(E_k)}; E_k(x) \leftarrow y_i$; return $y_i$ to $\mathcal{A}$
   When $\mathcal{A}$ asks a query $(k, y)$ to its right oracle:
     $i \leftarrow i + 1; k_i \leftarrow k; y_i \leftarrow y; x_i \xleftarrow{\$} \overline{Domain(E_k)}; E_k(x_i) \leftarrow y$; return $x_i$ to $\mathcal{A}$
When $\mathcal{A}$ halts, outputting a string out: return $((x_1, k_1, y_1), \ldots, (x_i, k_i, y_i), out)$
```

Fig. 1. $Domain(E_k)$ is the set of points x where $E_k(x)$ is no longer $undefined$ and $\overline{Domain(E_k)} = \{0,1\}^n - Domain(E_k)$. $Range(E_k)$ is the set of points where $E_k(x)$ is no longer $undefined$ and $\overline{Range(E_k)} = \{0,1\}^n - Range(E_k)$.

Merkle-Damgård Graph Let $H : (\{0,1\}^{\kappa})^{*} \rightarrow \{0,1\}^{n}$ be a Merkel Damagård construction hash function with compression function $F : \{0,1\}^{n} \times \{0,1\}^{\kappa} \rightarrow \{0,1\}^{n}$ and initial value IV. Let Merkle-Damgård Graph be a directed graph $\overrightarrow{G} = (V_G, \overrightarrow{E}_G)$. If $h' = F(h, e)$, then h and h' are in vertex set $V_G \subseteq \{0,1\}^{n}$ and (h, e, h') or $h \xrightarrow{e} h'$, an arc begin from h point at h', is in edge set $\overrightarrow{E}_G = \{(h, e, h')\} \subseteq \{0,1\}^{n} \times \{0,1\}^{\kappa} \times \{0,1\}^{n}$.

Graph Drawing Attack Let $\mathcal{A}^{?,?}$ be an adversary attacking H_{ι}. We analyze the behavior of $\mathcal{A}$ when its left oracle is instantiated by $E \xleftarrow{\$} \mathcal{B}_{n,n}$ and its right oracle is instantiated by E^{-1}. Assume that $\mathcal{A}$ asks its oracles at most q total queries. $\mathcal{A}$ runs the algorithm $SimulateOracle(\mathcal{A}, n)$ and draws a directed graph $\overrightarrow{G}_H$. When $\mathcal{A}$ asks an $E-query(x, k)$ and this returns a value y, or when $\mathcal{A}$ asks an $E^{-1} - query$ of (k, y) and this returns x. Then $\mathcal{A}$ adds vertexes $f_1(x, k, y)$, $f_3(x, k, y)$ and an arc $(f_1(x, k, y), f_2(x, k, y), f_3(x, k, y))$ to $\overrightarrow{G}_H$. Time consuming of graph drawing procedure is neglected. For $h_i = F_{\iota}(h_{i-1}, \overline{m}_i)$, the relations among $h_i, \overline{m}_i, h_{i-1}$ and f_1, f_2, f_3 are that: $h_{i-1} = f_1(x, k, y)$, $\overline{m}_i = f_2(x, k, y)$ and $h_i = f_3(x, k, y)$, or saying $f_3(x, k, y) = F_i(f_1(x, k, y), f_2(x, k, y))$. Relation among $f_1(x, k, y), f_2(x, k, y), f_3(x, k, y)$ of those 20 schemes are in Fig2. Combine of running $SimulateOracle(\mathcal{A}, n)$ and Graph drawing procedure is showing in Fig3, named $GraphDrawing(\mathcal{A}, n)$.

ι	$f_1 =$	$f_2 =$	$f_3 =$	$h_i =$	ι	$f_1 =$	$f_2 =$	$f_3 =$	$h_i =$
1	k	x	$y \oplus x$	$E_{h_{i-1}}(m_i) \oplus m_i$	13	$x \oplus k$	x	y	$E_{w_i}(m_i)$
2	k	$x \oplus k$	$y \oplus x$	$E_{h_{i-1}}(w_i) \oplus w_i$	14	$x \oplus k$	x	$y \oplus k$	$E_{w_i}(m_i) \oplus w_i$
3	k	x	$y \oplus x \oplus k$	$E_{h_{i-1}}(m_i) \oplus w_i$	15	x	k	y	$E_{m_i}(h_{i-1})$
4	k	$x \oplus k$	$y \oplus x \oplus k$	$E_{h_{i-1}}(w_i) \oplus m_i$	16	x	$x \oplus k$	y	$E_{w_i}(h_{i-1})$
5	x	k	$y \oplus x$	$E_{m_i}(h_{i-1}) \oplus h_{i-1}$	17	x	k	$y \oplus k$	$E_{m_i}(h_{i-1}) \oplus m_i$
6	$x \oplus k$	k	$y \oplus x$	$E_{m_i}(w_i) \oplus w_i$	18	x	$x \oplus k$	$y \oplus k$	$E_{w_i}(h_{i-1}) \oplus w_i$
7	x	k	$y \oplus x \oplus k$	$E_{m_i}(h_{i-1}) \oplus w_i$	19	$x \oplus k$	k	y	$E_{m_i}(w_i)$
8	$x \oplus k$	k	$y \oplus x \oplus k$	$E_{m_i}(w_i) \oplus h_{i-1}$	20	$x \oplus k$	k	$y \oplus k$	$E_{m_i}(w_i) \oplus m_i$
9	$x \oplus k$	x	$y \oplus x$	$E_{w_i}(m_i) \oplus m_i$	21	k	x	$y \oplus k$	$E_{h_{i-1}}(m_i) \oplus h_{i-1}$
10	x	$x \oplus k$	$y \oplus x$	$E_{w_i}(h_{i-1}) \oplus h_{i-1}$	22	k	$x \oplus k$	$y \oplus k$	$E_{h_{i-1}}(w_i) \oplus h_{i-1}$
11	$x \oplus k$	x	$y \oplus x \oplus k$	$E_{w_i}(m_i) \oplus h_{i-1}$	23	k	x	y	$E_{h_{i-1}}(m_i)$
12	x	$x \oplus k$	$y \oplus x \oplus k$	$E_{w_i}(h_{i-1}) \oplus m_i$	24	k	$x \oplus k$	y	$E_{h_{i-1}}(w_i)$

Fig. 2. Rules for the functions of building vertexes and arc, in which the adversary gets query (x, k, y) then computes the value $f_1 := f_1(x, k, y), f_2 := f_2(x, k, y)$ and $f_3 := f_3(x, k, y)$. $w_i := h_{i-1} \oplus m_i$. The first and sixth columns are the number of those $Group - 1[1..12]$ and $Group - 2[13..20]$ and additional 4 schemes numbered [21..24].

Let $\overrightarrow{G}_H^{q}$ be $\overrightarrow{G}_H$ after q-th query, $\overrightarrow{G}_H$ begin with $\overrightarrow{G}_H^{0}$. Let $\mathcal{H}$ be connected subgraph of $\overrightarrow{G}_H$, $\overrightarrow{C}$ be directed cycle or loop in $\overrightarrow{G}_H$, C be undirected cycle or loop in G_H(on assuming the arc is undirected), and $\overrightarrow{P}$ be directed directed Path in $\overrightarrow{G}_H$.

In different attack, $\mathcal{A}$ uses following methods: $\mathcal{A}$ selects different G_H^0; $\mathcal{A}$ defines different event as success event; In i-th query, $\mathcal{A}$ adds restriction on selection of (x_i, k_i) for $E - query$, or on selection of $(y_i, k_i)$for $E^{-1} - query$.

```
$GraphDrawing(\mathcal{A}, n)$
    Initially, $i \leftarrow 0$ and $E_k(x) = undefined$ for all $(x, k) \in \{0,1\}^n \times \{0,1\}^n$,
        $G_H = G_H^0$.
    Run $\mathcal{A}^{?,?}$, answering oracle queries as follows:
        When $\mathcal{A}$ asks a query $(x \xleftarrow{\$} \{0,1\}^n,\ k \xleftarrow{\$} \{0,1\}^n)$ to its left oracle:
            $i \leftarrow i+1; k_i \leftarrow k; x_i \leftarrow x; y_i \xleftarrow{\$} \overline{Range(E_k)}; E_k(x) \leftarrow y_i$; return $y_i$ to $\mathcal{A}$;
        When $\mathcal{A}$ asks a query $(y \xleftarrow{\$} \{0,1\}^n,\ k \xleftarrow{\$} \{0,1\}^n)$ to its right oracle:
            $i \leftarrow i+1; k_i \leftarrow k; y_i \leftarrow y; x_i \xleftarrow{\$} \overline{Domain(E_k)}; E_k(x_i) \leftarrow y$; return $x_i$ to $\mathcal{A}$;
        $V_{G_H^i} \leftarrow V_{G_H^{i-1}} \cup \{f_1(x_i, k_i, y_i), f_3(x_i, k_i, y_i)\}$;
        $E_{G_H^i} \leftarrow E_{G_H^{i-1}} \cup \{(f_1(x_i, k_i, y_i), f_2(x_i, k_i, y_i), f_3(x_i, k_i, y_i))\}$;
    When $\mathcal{A}$ halts, outputting a string and Graph $G_H^i$:
        return $((x_1, k_1, y_1), \ldots, (x_i, k_i, y_i), G_H^i)$.
```

Fig. 3. Adversary $\mathcal{A}$ executes its (simulated) oracle to form a directed graph G_H to build attack on H_ι, where $G_H : V_{G_H} \subseteq \{0,1\}^n; E_{G_H} \subseteq \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n$.

Conventions We assume the adversary does not ask any oracle query in which the response is already known; namely, if $\mathcal{A}$ asks a query $E_k(x)$ and this return y, then $\mathcal{A}$ does not ask a subsequent query of $E_k(x)$ or $E_k^{-1}(y)$; and if $\mathcal{A}$ asks $E_k^{-1}(y)$ and this return x, then $\mathcal{A}$ does not ask a subsequent query of $E_k^{-1}(y)$ or $E_k(x)$. We also assume a successful adversary always outputs one or more messages m_i, which either collide or (2nd)preimages. Before finishing, the adversary asks all the oracles calls to compute all hash values $H(m_i, IV)$. In $E - query(x, k)$, $f_1(x, k, y)$ and $f_2(x, k, y)$ are not influenced the return value y, so before asking $E - query$, we use notation ? to represent the unknown y; namely, when x and k is known, $f_1(x, k, y)$ is known, then, before getting $E - query(x, k)$, we use $f_1(x, k, ?)$ to represent this value. And before getting $E^{-1} - query(k, y)$, we use notation ? to represent some unknown x.

3 Collision Resistance of PGV Schemes

BRS paper analyzed the group 1 schemes using the Merkle-Damgård paradigm, for their compression functions are collision resistant. Group 2 schemes were analyzed by graph theory. We use Merkel-Damagård drawing method to analyze those schemes, by which preimage, second preimage or collision finding attack is converted to special path finding algorithm. Theorem1 and Theorem2 based on the fact that, adversary $\mathcal{A}$ runs algorithm $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}^0_{H_\iota} =$

$\{IV\}$, if $\mathcal{A}$ gets connect subgraph $\mathcal{H} \subseteq G_{H_\iota}$ with a cycle or loop C in it and vertex IV in it, then he formes a collision attack on H_ι.

Collision on H is that two directed paths $\overrightarrow{P} = h_0 \overset{\overline{m}_1}{\rightarrow} \ldots \overset{\overline{m}_l}{\rightarrow} h_l$ and $\overrightarrow{P}' = h'_0 \overset{\overline{m}'_1}{\rightarrow} \ldots \overset{\overline{m}'_{l'}}{\rightarrow} h'{}_{l'}$ have same start $h_0 = h'_0$ and same end $h_l = h'_{l'}$, or $P \cup P'$ builds a connect graph, a cycle or loop and IV on it. However this way of collision does not consider the message padding. If H uses MD strengthening padding, then the length of P and P' should be equal or the message lengths are included in $\overline{m}_l$ and $\overline{m}_{l'}$. The proofs of Theorem 2 is on condition of plain padding, that is also holden in MD strengthening padding, by restricting $f_1(x_i, k_i, ?) = IV$ or including message length in $f_2(x_i, k_i, ?)$.

Theorem 1. *Fix $n \geq 1$, message padding is plain padding,*

Group 1 Scheme: $Adv^{coll}_{H_\iota}(\mathcal{A}) \leq q(q+1)/2^n$ *for any* $q \geq 1$ *and* $\iota \in [1..12]$.
Group 2 Scheme: $Adv^{coll}_{H_\iota}(\mathcal{A}) \leq q(q+1)/2^n$ *for any* $q \geq 1$ *and* $\iota \in [13..20]$.
Group 3 Scheme: $Adv^{coll}_{H_\iota}(\mathcal{A}) = 1$ *for any* $q \geq 2$ *and* $\iota \in [21..64]$.

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}^0_{H_\iota} = \{IV\}$. Let $\mathcal{E}$ be the event that, as a result of the adversary's queries, there be a connected subgraph $\mathcal{H} \subseteq G_{H\iota}$, which has a Cycle or loop C and vertex IV. Let assume $\{IV\}$ be a connected graph. Let $\mathcal{E}_i$ be the event that $\mathcal{E}$ occurs by the i-th query. Define $\mathcal{E}_0$ be the null event. Then $\Pr[\mathcal{E}] = \sum_{i=1}^{q} \Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0]$. We have $Adv^{coll}_{H_\iota}(\mathcal{A}) \leq \Pr[\mathcal{E}]$.

Claim $Adv^{coll}_{H_\iota}(\mathcal{A}) \leq \Pr[\mathcal{E}]$.
Meaning collision on H_ι at least is a connected subgraph $\mathcal{H} \subseteq G_H$, $\mathcal{H}$ has a cycle or loop C and vertex IV. $\mathcal{A}$ outputs colliding message $m = \overline{m}_1 \| \ldots \| \overline{m}_l$ and $m' = \overline{m}'_1 \| \ldots \| \overline{m}'_{l'}$; that is $H_i(m, IV) = H_i(m', IV)$. In path $\overrightarrow{P} = h_0 \overset{\overline{m}_1}{\rightarrow} h_1 \overset{\overline{m}_2}{\rightarrow} \ldots \overset{\overline{m}_l}{\rightarrow} h_l$ and $\overrightarrow{P}' = h'_0 \overset{\overline{m}'_1}{\rightarrow} h'_1 \overset{\overline{m}'_2}{\rightarrow} \ldots \overset{\overline{m}'_{l'}}{\rightarrow} h'_{l'}$, we have $h_0 = h'_0$, $h_l = h'_{l'}$ and $P \neq P'$. Then there exists at least one cycle or loop in $P \cup P'$ and $IV \in P \cup P'$. $P \cup P'$ is a connect subgraph.

Claim Let $\mathcal{H}_\alpha$ be connect subgraph in G_H and in each $\mathcal{H}_\alpha$, a cycle or a loop $C \subseteq \mathcal{H}_\alpha$ or $IV \in \mathcal{H}_\alpha$. Let $\mathcal{H}^q_\alpha \subseteq G^q_H$, $\cup\mathcal{H}^q_\alpha$ be union of all such connected subgraphs in G^q_H. Then $|V_{\cup\mathcal{H}^q_\alpha}| \leq q+1$.
If $\mathcal{H}_\alpha$ is a connect subgraph, then $|V_{\mathcal{H}_\alpha}| \leq |E_{\mathcal{H}_\alpha}| + 1$. If a cycle or loop $C \subseteq \mathcal{H}_\alpha$, then $|V_{\mathcal{H}\alpha}| \leq |E_{\mathcal{H}\alpha}|$. Since $|V_{\cup\mathcal{H}^q_\alpha}| = \sum |V_{\mathcal{H}^q_\alpha}| \leq \sum |E_{\mathcal{H}^q_\alpha}| + 1 \leq E_{G^q_H} + 1 = q+1$, we have $|V_{\cup\mathcal{H}^q_\alpha}| \leq q+1$.

Claim Let $Ver_{\cup\mathcal{H}^{i-1}_\alpha}(f_\iota(x,k,y))_i$ be event that $f_\iota(x_i, k_i, y_i) \in \cup\mathcal{H}^{i-1}_\alpha$. Then $Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \Pr[Ver_{\cup\mathcal{H}^{i-1}_\alpha}(f_1)_i \wedge Ver_{\cup\mathcal{H}^{i-1}_\alpha}(f_3)_i]$.
Let $\mathcal{E}$ occurs in i-th query. Then there exists a connected subgraph $\mathcal{H}$ with $IV \in \mathcal{H}$ and a cycle or loop C in $\mathcal{H}$ and $\mathcal{H} \subseteq G^i_H$. We will give proof of $\mathcal{H} - (f_1, f_2, f_3) \subseteq \cup\mathcal{H}^{i-1}_\alpha$. If that is true, then $f_1, f_3 \in \mathcal{H}^{i-1}_\alpha$. Firstly, if (f_1, f_2, f_3) is in cycle C, then $\mathcal{H} - (f_1, f_2, f_3)$ is connected graph and $IV \in \mathcal{H}$. Secondly, if (f_1, f_2, f_3) is not in cycle C, then at most two connected graph in

$\mathcal{H} - (f_1, f_2, f_3)$ denoted $\mathcal{H}_1$ and $\mathcal{H}_2$. Since $IV \in \mathcal{H}$, we have $IV \in \mathcal{H}_1 \cup \mathcal{H}_2$. Let assume $IV \in \mathcal{H}_1$. If there is a cycle in $\mathcal{H}_1$, then that is conflict with collision occur in i-th query. Since $(f_1, f_2, f_3) \notin C$, then there is a circle in $\mathcal{H}_2$. We have $\mathcal{H}_1 \cup \mathcal{H}_2 \subseteq \cup\mathcal{H}_\alpha^{i-1}$, that implies $f_1, f_3 \in \mathcal{H}_\alpha^{i-1}$.

Claim $\Pr[\mathcal{E}] \leq \frac{q(q+1)}{2^n}$, $\iota \in [1..12]$.

Given $\overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0$, the event $\mathcal{E}_i$ occurs at least in case that, the return vertex of i-th query has been exist in vertexes set $\cup\mathcal{H}_\alpha^{i-1}$.

$\Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i \wedge Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i] \leq \Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i | Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i]$.

If $\mathcal{E}_i$ occurs via an $E - query(x_i, k_i)$, then y_i is a random value from a set of size at least $2^n - (i-1)$. Then $f_3(x_i, k_i, y_i)$ is a random value from a set of size at least $2^n - (i-1)$. We also have $|V_{\cup\mathcal{H}_\alpha^{i-1}}| \leq i$. So,

$\Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \frac{i}{2^n - (i-1)}$.

Alternatively, let $\mathcal{E}_i$ occur via an $E^{-1} - query(y_i, k_i)$. For schemes $\iota \in [1..4]$, $\mathcal{A}$ can select $f_1(x_i, k_i, y_i) \in \cup\mathcal{H}_\alpha^{i-1}$ directly, that success probability is same as asking E-query. For schemes $\iota \in [5..12]$, $f_1(x_i, k_i, y_i)$ is a random value from a set of size at least $2^n - (i-1)$. Then

$$\begin{aligned} &\Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i \wedge Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i] \\ &= \Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i | Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i] \Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i] \end{aligned}$$

We have $\Pr[\mathcal{E}] \leq \sum_{i=1}^{q} \Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \sum_{i=1}^{q} \frac{i}{2^n - (i-1)} \leq \frac{q(q+1)}{2^n}$.

Claim $\Pr[\mathcal{E}] \leq \frac{q(q+1)}{2^n}$, $\iota \in [13..20]$.

If $\mathcal{E}_i$ occurs via an $E - query(x_i, k_i)$, then that probability is same as gruop1. Alternatively, if $\mathcal{E}_i$ occurs via an $E^{-1} - query(y_i, k_i)$, then $f_1(x_i, k_i, y_i)$ is a random value from a set of size at least $2^n - (i-1)$. So,

$\Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i \wedge Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i] \leq \Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i | Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i]$

We can select vertex $f_3(x_i, k_i, y_i)$ in $\cup\mathcal{H}_\alpha^{i-1}$. We have,

$\Pr[\mathcal{E}] \leq \sum_{i=0}^{q} \Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \sum_{i=1}^{q} \frac{i}{2^n - (i-1)} \leq \frac{q(q+1)}{2^n}$.

Claim $\Pr[\mathcal{E}] = 1$, $\iota \in [21..64]$.

$$\Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \Pr[Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_1)_i \wedge Ver_{\cup\mathcal{H}_\alpha^{i-1}}(f_3)_i]$$

In 2th query, we can directly select $f_1(x_i, k_i, y_i), f_3(x_i, k_i, y_i) \in V_{\mathcal{H}_\alpha^1}$. So we have $\Pr[\mathcal{E}] = 1$, $q \geq 2$. □

Theorem 2. *Fix $n \geq 1$, message padding is plain padding,*

Group 1 Scheme $Adv_{H_\iota}^{coll}(\mathcal{A}) \geq q(q+1)/2^{n+1}$ *for any* $q \geq 1$ *and* $\iota \in [1..12]$.
Group 2 Scheme $Adv_{H_\iota}^{coll}(\mathcal{A}) \geq q(q+1)/2^{n+1}$ *for any* $q \geq 1$ *and* $\iota \in [13..20]$.

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}_{H_\iota}^0 = \{IV\}$. Let $\mathcal{A}$ only ask $E - query(x, k)$. In each query, $\mathcal{A}$ selects x and k to satisfy $f_1(x, k, ?) \in V_{G_{H_\iota}}$. Then G_{H_ι} is connected graph and $IV \in G_{H_\iota}$, that is $G_{H_\iota}^i = \cup\mathcal{H}_\alpha^i$. Let $\mathcal{C}$ be the event of $IV \in G_{H_\iota}$ and there exists a cycle or loop in C. Let $\mathcal{C}_i$ be the event that $\mathcal{C}$ occurs by the i-th query. Define $\mathcal{C}_0$ be the null event. Then $\Pr[\mathcal{C}] = \sum_{i=1}^{q} \Pr[\mathcal{C}_i | \overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0]$. We have $Adv_{H_\iota}^{coll}(\mathcal{A}) \geq \Pr[\mathcal{C}]$.

Claim $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq \Pr[\mathcal{C}]$.

If adversary $\mathcal{A}$ build a cycle or loop C in i-th query. Then there are two directed path P and P' in $G^i_{H_\iota}$ with $P = h_0 \to h_1 \to \ldots \to h_l$, $P' = h_0 \to h'_1 \to \ldots \to h'_{l'}$, in which $h_l = h'_{l'} = f_3(x_i, k_i, y_i)$ and $P \neq P'$.

Claim $Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0] \geq \Pr[Ver_{G^{i-1}_{H_\iota}}(f_3)_i]$.

If $f_3(x_i, k_i, y_i) \in G^{i-1}_{H_\iota}$, then $E(G^i_{H_\iota}) = V(G^i_{H_\iota})$. There must exist a cycle or loop in G_{H_ι}.

Claim $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq q(q+1)/2^{n+1}$ for any $q \geq 1$ and $\iota \in [1..20]$.

Given $\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0$, the event $\mathcal{C}_i$ occurs in case that, the return vertex of i-th query has been exist in vertexes set $G^{i-1}_{H_\iota}$. If $\mathcal{E}_i$ occurs via an $E{-}query(x_i, k_i)$, then y_i is a random value from a set of size at most 2^n. Then $f_3(x_i, k_i, y_i)$ is a random value from a set of size at most 2^n. So,

$\Pr[\mathcal{C}] \geq \sum_{i=1}^{q} \Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0] \geq \sum_{i=1}^{q} \frac{i}{2^n} = \frac{q(q+1)}{2^{n+1}}$. □

Theorem 3. *Fix $n \geq 1$, message padding is MD strengthening padding,*

$$q(q-1)/2^n \geq Adv^{coll}_{H_\iota}(\mathcal{A}) \geq q(q-1)/2^{n+1}$$

for any $q \geq 1$ and $\iota \in [1..20]$.

4 Preimage Resistance of PGV Schemes

The proofs of Theorem 4 and Theorem 5 follow the fact that, a preimage is a directed path from IV to δ in graph $\overrightarrow{G}_{H_\iota}$. If adversary $\mathcal{A}$ finds a directed path from IV to δ, then he builds a preimage. The bounds on preimage attack not consider the MD-strengthening padding, if the padding is considered, the bounds are same. Because message length can be added in $f_2(x_i, k_i, y_i)$, before asking i-th query.

Theorem 4. *Fix $n \geq 1$, given δ, message padding is plain padding,*

Group 1 Scheme $Adv^{pre}_{H_\iota}(\mathcal{A}) \leq q/2^{n-1}$ *for any $q \geq 1$ and $\iota \in [1..12]$.*
Group 2 Scheme $Adv^{pre}_{H_\iota}(\mathcal{A}) \leq q(q+1)/2^n$ *for any $q \geq 1$ and $\iota \in [13..20]$.*
Others Scheme $Adv^{pre}_{H_\iota}(\mathcal{A}) = 1$ *for any $q \geq 1$ and $\iota \in [21..64]$.*

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}^0_{H_\iota} = \{IV, \delta\}$. Preimage finding is not finding a cycle or loop, it is finding a path. Let assume $\{IV\}$ and $\{\delta\}$ be connected subgraphs. Let $\mathcal{E}$ be the event that, as a result of the adversary's queries, there are formed a directed path $\overrightarrow{P}$ in $\overrightarrow{G}_{H_\iota}$, $IV \in \overrightarrow{P}$ and $\delta \in \overrightarrow{P}$. Let $\mathcal{E}_i$ be the event that $\mathcal{E}$ occurs by the i-th query. Define $\mathcal{E}_0$ be the null event. Then $\Pr[\mathcal{E}] = \sum_{i=1}^{q} \Pr[\mathcal{E}_i|\overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0]$. We have $Adv^{pre}_{H_\iota}(\mathcal{A}) \leq \Pr[\mathcal{E}]$.

Claim $Adv^{pre}_{H_\iota}(\mathcal{A}) \leq \Pr[\mathcal{E}]$. Implying preimage on H at least is a path $P \subseteq G_H$, in which $IV \in P$ and $\delta \in P$.

Claim Let $Arc_\delta(f_3(x,k,y))$ be event that $f_3(x,k,y)=\delta$.
Then $Pr[\mathcal{E}] \le \Pr[Arc_\delta(f_3(x,k,y))]$. That implies at least a directed arc point at δ.

Claim Let $\mathcal{H}_a$ be connect subgraph in G_H with $a \in \mathcal{H}_a$, in which $a \in \{IV, \delta\}$.
Let $\mathcal{H}_a^q$ be the connect graph after q-th query. Then $|V_{\mathcal{H}_{IV}^q \cup \mathcal{H}_\delta^q}| \le q+2$.

Claim $Pr[\mathcal{E}_i|\overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \le \Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i]$.

Claim $\Pr[\mathcal{E}] \le \frac{q}{2^{n-1}}$, $\iota \in [1..12]$.

Claim $\Pr[\mathcal{E}] \le \frac{q\cdot(q+1)}{2^n}$, $\iota \in [13..20]$.

Claim $\Pr[\mathcal{E}] = 1$, $\iota \in [21..64]$.
For given IV and δ, we have $Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] = 1$. □

Detail is in Appendix A.

Theorem 5. *Fix $n \ge 1$, given δ, message padding is plain padding,*

Group 1 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \ge q/2^n$ *for any* $q \ge 1$ *and* $\iota \in [1..12]$.
Group 2 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \ge q(q+1)/2^{n+2}$ *for any* $q \ge 1$ *and* $\iota \in [13..20]$.

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}_{H_\iota}^0 = \{IV, \delta\}$.

Claim $Adv_{H_\iota}^{coll}(\mathcal{A}) \ge q/2^n$ for any $q \ge 1$ and $\iota \in [1..12]$.
$\mathcal{A}$ only asks $E - query(x,k)$ with x_i and k_i satisfing $f_1(x_i, k_i, ?) \in \mathcal{H}_{IV}$. Let $\mathcal{C}$ be the event of $\delta \in \mathcal{H}_{IV}$ and at least a edge in $\mathcal{H}_{IV}$.
$Adv_{H_\iota}^{coll}(\mathcal{A}) \ge \Pr[\mathcal{C}] \ge \sum_{i=1}^q \Pr[Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] \ge \sum_{i=1}^q \frac{1}{2^n} = \frac{q}{2^n}$.

Claim $Adv_{H_\iota}^{coll}(\mathcal{A}) \ge q(q+1)/2^{n+2}$ for any $q \ge 1$ and $\iota \in [13..20]$.
$\mathcal{A}$ asks $E - query(x,k)$ and $E^{-1} - query(y,k)$, alternately. In odd-th query, $\mathcal{A}$ selects x and k satisfying $f_1(x,k,?) \in \mathcal{H}_{IV}$, in even-th query, $\mathcal{A}$ selects y and k satisfing $f_3(?,k,y) \in \mathcal{H}_\delta$. Let $\mathcal{C}$ be the event of $\delta \in \mathcal{H}_{IV}$ and at least a edge in $\mathcal{H}_{IV}$. If $\mathcal{C}$ occurs in $E - query$, then $Adv_{H_\iota}^{coll}(\mathcal{A}) \ge \Pr[\mathcal{C}] \ge \Pr[Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] \ge \sum_{i=1}^q \frac{\lfloor i/2 \rfloor + 1}{2^n}$. If $\mathcal{C}$ occurs in $E^{-1} - query$, then $Adv_{H_\iota}^{coll}(\mathcal{A}) \ge \Pr[\mathcal{C}] \ge \Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i] \ge \sum_{i=1}^q \frac{\lfloor i/2 \rfloor + 1}{2^n}$. □

Detail is in Appendix A.

5 Second Preimage Resistance of PGV Schemes

Second preimage bounds on H_ι with plain-padding and MD-strengthening are different, for the success events of those two attacks are different. Let the given message build a path $\overrightarrow{P}$. Second preimage attack with plain padding is also a direct path $\overrightarrow{P}'$ finding attack, for which the target is all vertexes in path $\overrightarrow{P}$.

Theorem 6. *Fix $n > 1$, $q \ge 1$, let $H(m, IV) = \delta$, $m = \overline{m}_1 \| \ldots \| \overline{m}_t$, message padding is plain padding,*

Group 1 Scheme: $\frac{(t+1)q}{2^n} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{(t+1)q}{2^{n-1}}$ *for* $\iota \in [1..12]$.
Group 2 Scheme: $\frac{q(4t+q+3)}{2^{n+2}} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{q(2t+q+3)}{2^n}$ *for* $\iota \in [13..20]$.

Proof. This proof is followed the proofs of Theorem 4 and Theorem 5. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. Let $m = \overline{m}_1 \| \ldots \| \overline{m}_t$, $\sigma_\iota := H(m_\iota, IV), m_\iota = \overline{m}_1 \ldots, \overline{m}_\iota, \iota \leq t$. Let $\sigma_0 := IV$. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}^0_{H_\iota} := \{\sigma_\iota | 0 \leq \iota \leq t\}$. $\mathcal{A}$ follows graph drawing method of Theorem 4 and Theorem 5.

UP Bound of Group 1 In q-query $\Pr[Arc_{\sigma_\iota}(f_3)] \leq \frac{q}{2^{n-1}}$, we have $Adv^{sPre}_{H_\iota}(q) \leq \sum_{\iota=0}^{t} \Pr[Arc_{\sigma_\iota}(f_3)] \leq \frac{q(t+1)}{2^{n-1}}$. In i-th query $\Pr[Ver_{\mathcal{H}^{i-1}_{\sigma_t}}(f_3)_i] \geq \frac{t+1}{2^n}$. We have, $Adv^{sPre}_{H_\iota}(q) \geq \sum_{i=1}^{q} \Pr[Ver_{\mathcal{H}^{i-1}_{\sigma_t}}(f_3)_i] \geq \frac{q(t+1)}{2^n}$.

UP Bound of Group 2 Since $\Pr[Ver_{\mathcal{H}^{i-1}_{\sigma_t}}(f_3)_i | Ver_{\mathcal{H}^{i-1}_{IV}}(f_1)_i] \leq \frac{i+t+1}{2^n-(i-1)}$. $Adv^{sPre}_{H_\iota}(q) \leq \sum_{i=0}^{q} \frac{i+t+1}{2^n-(i-1)} \leq \frac{q(q+2t+3)}{2^n}$. In i-th query, $\Pr[Ver_{\mathcal{H}^{i-1}_{\sigma_t}}(f_3)_i] \geq \frac{\lfloor \frac{i}{2} \rfloor + t+1}{2^n}$. We have, $Adv^{sPre}_{H_\iota}(q) \geq \sum_{i=1}^{q} \Pr[Ver_{\mathcal{H}^{i-1}_{\sigma_\iota}}(f_3)_i] \geq \frac{q(4t+q+3)}{2^{n+2}}$. □

Theorem 7 is based on the facts that, if the message length is considered, in each query, the target image set is not set $\{\sigma_\iota, 0 \leq \iota \leq t\}$, is a vertex in it. Then the bound will become same as preimage attack. However, in schemes [5..20], $\mathcal{A}$ can build a short direct cycle or loop $\overrightarrow{C}$, in this way, the length of second message can be controlled by $\mathcal{A}$. Let $L = |C|$. The target image set becomes $\{\sigma_{L\iota} | 2 \leq \iota \leq \lfloor t/L \rfloor\}$. C can found by precomputation with complexity of $\mathcal{O}(2^{n/2})$, which can be used in any second preimage attack, that was not included in complexity bounds of finding second preimage.

Theorem 7. *Fix* $n > 1$, $q \geq 1$, *let* $H(m, IV) = \delta$, $m = \overline{m}_1 \| \ldots \| \overline{m}_t$, *message padding is MD strengthening padding,*

Schemes [1..4] $\frac{q}{2^n} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{q}{2^{n-1}}$.
Schemes [5..12] $\frac{(t-1)q}{2^n} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{(t-1)q}{2^{n-1}}$.
Schemes [13..20] $\frac{q(q+1)}{2^{n+2}} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{q(q+1)}{2^n}$.

Proof. This proof is followed the proofs of Theorem 6 and Theorem 7. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. Let $\sigma_\iota := H(m_\iota, IV), m_\iota = \overline{m}_1 \ldots, \overline{m}_\iota, \iota \leq t$.

Schemes [1..4] For MD-strengthening, when message length is included in $f_2(x_i, k_i, y_i)$, the success event is $Arc\sigma_\iota$, not $Ver_{\mathcal{H}_{\sigma_\iota}}$.

Schemes [5..12] Before attacking H_ι, $\iota \in [5..12]$, $\mathcal{A}$ find message blocks $\overline{m}_1$ and $\overline{m}_2$ with $H_\iota(\overline{m}_1 \| \overline{m}_2^{(i)}, IV) = H_\iota(\overline{m}_1 \| \overline{m}_2^{(j)}, IV), \forall i, j > 0$, detail is given in next section, called expandable fixed point. Let $IV' = H_\iota(\overline{m}_1, IV)$. Let $G^0_H := \{\sigma_\iota, \iota \geq 2\}$. We have $\frac{(t-1)q}{2^n} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{(t-1)q}{2^{n-1}}$.

Schemes [13..20] Before attacking H_ι, $\iota \in [13..20]$, $\mathcal{A}$ find message blocks $\overline{m}_1$ and $\overline{m}_2$ with $H_\iota((\overline{m}_1 \| \overline{m}_2)^{(i)}, IV) = H_\iota((\overline{m}_1 \| \overline{m}_2)^{(j)}, IV)$, $\forall i, j > 0$, detail is given in next section. Let $G^0_H := \{\sigma_{2\iota} | 2\iota \leq t\}$. We have $\frac{(q+2t)q}{2^{n+2}} \leq Adv^{sPre}_{H_\iota}(\mathcal{A}) \leq \frac{(q+t)q}{2^n}$. □

6 Multicollisions on PGV Schemes

Multicollisions Multicollisions attack is first given by Joux[6], which is a way to produce a large number of messages that collide for an iterated hash function, with only a little more work than is needed to find a single pair of messages that collide. More precisely, using t collision $H(B_1\|\ldots\|B_l, IV) = H(B'_1\|\ldots\|B'_l, IV)$, where $B_i \neq B'_i$ and $l = 1, \ldots, t$, we can build 2^t-collisions in H. The time consuming is $t \cdot \mathcal{O}(2^{n/2})$. The collide block finding procedure is illustrated as Algorithm $CollisionBlock(\mathcal{A}, h, t)$:

```
$CollisionBlock(\mathcal{A}, h, t)$
    **For** $i := 1$ **to** $q$
        $\mathcal{A}$ selects $x_i, k_i$ with $f_1(x_i, k_i, ?) = h$, then asks $E - query(x_i, k_i)$.
    $\mathcal{A}$ success with $f_3(x_i, k_i, y_i) = f_3(x_j, k_j, y_j)$,
    return $\overline{m}_t \leftarrow f_2(x_i, k_i, y_i); \overline{m}'_t \leftarrow f_2(x_j, k_j, y_j); h_t \leftarrow f_3(x_i, k_i, y_i)$;
```

Fixed-Point Expandable Message A fixed point is a pair $(h_{i-1}, \overline{m}_i)$ such that $h_{i-1} = F(h_{i-1}, \overline{m}_i)$. Compression functions based on Davies-Meyer construction, such as the SHA family, MD4, MD5 and Tiger, have easily found fixed points. Kelesy and Schneier[8] gave a second preimage attack based on Fixed-Point expandable message. We call it expandable short dicycle in this paper, for, in schemes [13..20], fixed-point expandable message is not easy to be found, but a similar expandable short directed cycle as expandable fixed point can be found and attacks based on them.

Expandable Short DiCycle Expandable Short dicycle requires a short directed cycle or loop being build with desired prefix. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs algorithm $GraphDrawing(\mathcal{A}, n)$. The expandable short dicycle found algorithm is as follows:

```
$ExpandableDiCycle(\mathcal{A}, h, t, \iota)$
    **For** $i := 1$ **to** $2^{n/2}$
        $\mathcal{A}$ selects $(x_i, k_i)$ with $f_1(x_i, k_i, ?) = h$ asks $E - query(x_i, k_i)$.
    **For** $i := 1$ **to** $2^{n/2}$
        If $\iota \in \{5, 8, 10, 11\}$ Then
            $\mathcal{A}$ selects $(y'_i, k'_i)$ with $y'_i = \mathbf{0}$ asks $E^{-1} - query(y'_i, k'_i)$.
        If $\iota \in \{6, 7, 9, 12\}$ Then
            $\mathcal{A}$ selects $(y'_i, k'_i)$ with $y'_i = k'_i$ asks $E^{-1} - query(y'_i, k'_i)$.
        If $\iota \in [13..20]$ Then
            $\mathcal{A}$ selects $(y'_i, k'_i)$ with $f_3(?, k'_i, y'_i) = h$ asks $E^{-1} - query(y'_i, k'_i)$.
    $\mathcal{A}$ success with $f_3(x_i, k_i, y_i) = f_1(x'_j, k'_j, y'_j)$,
    return $\overline{m}_t \leftarrow f_2(x_i, k_i, y_i); \overline{m}_{tt} \leftarrow f_2(x'_j, k'_j, y'_j); h_t \leftarrow f_3(x'_i, k'_i, y'_i)$;
```

Short DiCycle Multicollisions Short DiCycle Multicollisions attack is first given in[8], which is a way to produce a large number of messages that collide for an iterated hash function, with only a little more work than is needed to find a single pair of expandable short dicyle. More precisely, using t short dicycle, we can build multicollisions in H_ι. The time consuming is $2t \cdot \mathcal{O}(2^{n/2})$.

$$H_\iota(\overline{m}_1 \| \overline{m}_{11}^{(k_1)} \| \ldots \| \overline{m}_t \| \overline{m}_{tt}^{(k_t)}) = H_\iota(\overline{m}_1 \| \overline{m}_{11}^{(k_1')} \| \ldots \| \overline{m}_t \| \overline{m}_{tt}^{(k_t')}), \iota \in [5..12]$$

With $\sum_{i=1}^{t} k_i = \sum_{i=1}^{t} k_i'$.

Let $l := \sum_{i=1}^{t} k_i$. Then adversary takes $2t2^{n/2}$ times finding $S_l(t)$, $l + t - length$ multicollisions, where $S_l(t) = \sum_{i_l=0}^{t} S_{l-1}(i_l)$.

$$S_l(t) = \sum_{i_l=0}^{t} S_{l-1}(i_l) = \sum_{i_l=0}^{t} \sum_{i_{l-1}=0}^{i_l} S_{l-2}(i_{l-1}) = \sum_{i_l=0}^{t} \sum_{i_{l-1}=0}^{i_l} \ldots \sum_{i_2=0}^{i_3} S_2(i_2).$$

Where $S_2(i) = i + 1$. For schemes [13..20], the minimum message length is $2l + t$ with $2t2^{n/2}$ complexity and $S_l(t)$ collisions.

Combine Multicollisions Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs algorithm $GraphDrawing(\mathcal{A}, n)$ building multicollision, where original multicollisions and short Dicycle multicollisions are combined.

$$\begin{aligned} &H_\iota(\overline{m}_1 \| \overline{m}_{11}^{(k_1)} \| \overline{m}_2 \| \overline{m}_3 \| \overline{m}_{33}^{(k_2)} \ldots \| \overline{m}_{2t-1} \| \overline{m}_{2t-12t-1}^{(k_t)} \| \overline{m}_{2t}) \\ = &H_\iota(\overline{m}_1 \| \overline{m}_{11}^{(k_1')} \| \overline{m}_2' \| \overline{m}_3 \| \overline{m}_{33}^{(k_2')} \| \ldots \| \overline{m}_{2t-1} \| \overline{m}_{2t-12t-1}^{(k_t')} \| \overline{m}_{2t}'), \iota \in [5..12] \end{aligned}$$

With $\sum_{i=1}^{t} k_i = \sum_{i=1}^{t} k_i'$. Let $l := \sum_{i=1}^{t} k_i$.Then adversary takes $\mathcal{O}(3t2^{n/2})$ times $E - query$ to get $\mathcal{O}(2^t(t+1)^{l-1})$, $2t + l - length$ multicollisions, where $l \geq t \geq 2$.

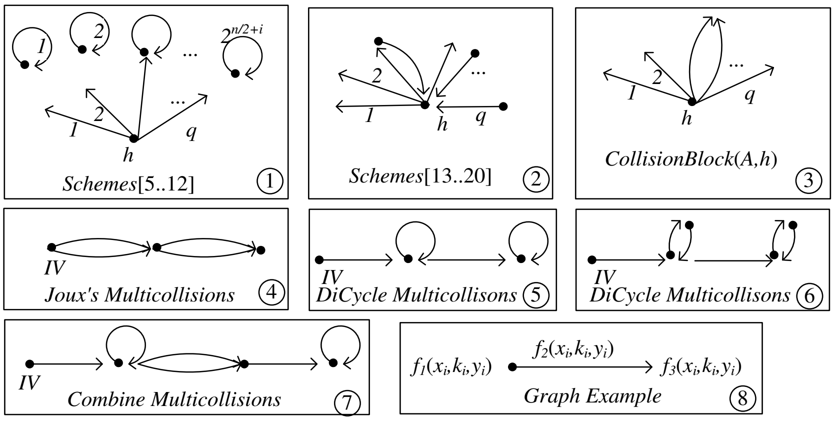

Fig. 4. Directed Cycle finding algorithms of schemes [5..12] and [13..20] are illustrated in subgraph 1 and 2. Undirected Cycle(Multicollision block) finding algorithm is given in subgraph 3. Joux's Multicollisions, Kelsey's Multicollisions on schemes [5..12] and [13..20] and combine of those two Multicollisions are presented in subgraph 4, 5, 6, 7, respectively.

7 Conclusions

In this paper, we give the bounds on PGV schemes against preimage, second preimage, collision and multicollisions, and that are improved by graph drawing method and short cycle build method. We omit the bounds of some new attacks including second preimage attack based on other expandable message[8] and preimage attack based on herding attack[9], for those bounds can be precise by similar way as second preimage attack. From the bounds, schemes [1..4] seems better than schemes [5..20], but more analysis is required.

References

1. E.Biham and R.Chen. Near-Collisions of SHA-0 and SHA-1. In Selected Areas in Cryptography-SAC 2004.
2. E.Biham and R.Chen. Near-Collisions of SHA-0,In Advances in Cryptology CRYPTO'2004, LNCS 3152,pp290-305,2004.
3. J.Black, P.Rogaway, and T.Shrimpton, "Black-box analysis of the block-cipher-based hashfunction constructions from PGV". In Advances in Cryptology - CRYPTO'02, volume 2442 of Lecture Notes in Computer Science. Springer-Verlag, 2002.pp.320-335.
4. I.Damgård. A design principle for hash functions. In G. Brassard, editor, Advances in Cryptology-CRYPTO' 89, volume 435 of Lecture Notes in Computer Science. Springer-Verlag, 1990.
5. J.Daemen and V.Rijmen: The Design of Rijndael: AES The Advanced Encryption Standard. Springer, 2002.
6. A.Joux. Multicollisions in iterated hash functions, application to cascaded constructions. Crypto 04, LNCS 3152, 306C316.
7. S. Hirose. Some plausible constructions of double-block-length hash functions. In Preproceedings of the 13th Fast Software Encryption Workshop (FSE 2006), pp. 231-246, 2006.
8. J. Kelsey and B. Schneier. Second preimages on n-bit hash functions for much less than 2n work. In R. Cramer, editor, EUROCRYPT 2005, LNCS 3494, pp.474-490, 2005.
9. John Kelsey, Tadayoshi Kohno. Herding Hash Functions and the Nostradamus Attack. In S. Vaudenay, editor, EUROCRYPT 2006, LNCS 4004, pp.183-200, 2006.
10. X.Lai and J.L.Massey: Hash functions based on block ciphers. In Advances in Cryptology Eurocrypt'92, Lecture Notes in Computer Science, Vol. 658. Springer-Verlag, Berlin Hei-delberg New York (1993) 55-70.
11. C. H. Meyer and S. M. Matyas. Cryptography: a New Dimension in Data Security. Wiley & Sons, 1982.
12. M. Nandi, W. Lee, K. Sakurai, and S. Lee. Security analysis of a 2/3-rate double length compression function in the black-box model. In Proceedings of the 12th Fast Software Encryption (FSE 2005), LNCS 35571, pp. 243-254, 2005.
13. B.Preneel, V. Rijmen, A.Bosselaers: Recent Developments in the Design of Conventional Cryptographic Algorithms. In State of the Art and Evolution of Computer Security and Industrial Cryptography. Lecture Notes in Computer Science, Vol 1528. Springer-Verlag, Berlin Heidelberg New York(1998) 106-131.

14. B. Preneel, R. Govaerts, and J. Vandewalle, " Hash functions based on block ciphers,", In Advances in Cryptology -CRYPTO'93, Lecture Notes in Computer Science,pages 368-378. Springer-Verlag, 1994.
15. M. O. Rabin. Digitalized Signatures. In R. A. Demillo, D. P. Dopkin, A. K. Jones, and R. J. Lipton, editors, Foundations of Secure Computation, pages 155-166, New York, 1978. Academic Press.
16. C.E. Shannon. "Communication theory of secrecy systems,", Bell System Technical Journal, 28:656C715, 1949.
17. X.Wang, H.Yu, How to Break MD5 and Other Hash Functions, EUROCRYPT'2005, Springer-Verlag, LNCS 3494, pp19-35, 2005.
18. X. Wang, X. Lai, D.Feng and H.Yu., Cryptanalysis of the Hash Functions MD4 and RIPEMD, EUROCRYPT 2005, Springer-Verlag,LNCS 3494, pp1-18, 2005.

A Bounds on Preimage

The following theorem is proof of Theorem4.

Theorem 8. *Fix $n \geq 1$, given δ, message padding is plain padding,*

Group 1 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \leq q/2^{n-1}$ *for any* $q \geq 1$ *and* $\iota \in [1..12]$.
Group 2 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \leq q(q+1)/2^n$ *for any* $q \geq 1$ *and* $\iota \in [13..20]$.
Others Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) = 1$ *for any* $q \geq 1$ *and* $\iota \in [21..64]$.

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι. $\mathcal{A}$ runs $GraphDrawing(\mathcal{A}, n)$ with $\overrightarrow{G}_{H_\iota}^0 = \{IV, \delta\}$. Preimage finding is not finding a cycle or loop, it is finding a path. Let assume $\{IV\}$ and $\{\delta\}$ be connected subgraphs. Let $\mathcal{E}$ be the event that, as a result of the adversary's queries, there are formed a directed path $\overrightarrow{P}$ in $\overrightarrow{G}_{H_\iota}$, $IV \in \overrightarrow{P}$ and $\delta \in \overrightarrow{P}$. Let $\mathcal{E}_i$ be the event that $\mathcal{E}$ occurs by the i-th query. Define $\mathcal{E}_0$ be the null event. Then $\Pr[\mathcal{E}] = \sum_{i=1}^{q} \Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0]$. We have $Adv_{H_\iota}^{pre}(\mathcal{A}) \leq \Pr[\mathcal{E}]$.

Claim $Adv_{H_\iota}^{pre}(\mathcal{A}) \leq \Pr[\mathcal{E}]$.
Implying preimage on H at least is a path $P \subseteq G_H$, in which $IV \in P$ and $\delta \in P$. Adversary $\mathcal{A}$ find message $m = \overline{m}_1 \| \ldots \| \overline{m}_l$ with $H(m, IV) = \delta$. Then we build path $P = h_0 \overset{\overline{m}_1}{\rightarrow} h_1 \overset{\overline{m}_2}{\rightarrow} \ldots \overset{\overline{m}_l}{\rightarrow} \delta$ we have $h_0 = IV$, $h_l = \delta$. Then there exists at least one path P, in which $IV \in P$ and $\delta \in P$.

Claim Let $Arc_\delta(f_3(x, k, y))$ be event that $f_3(x, k, y) = \delta$.
Then $Pr[\mathcal{E}] \leq \Pr[Arc_\delta(f_3(x, k, y))]$. That implies at least a directed arc point at δ.

Claim Let $\mathcal{H}_a$ be connect subgraph in G_H with $a \in \mathcal{H}_a$, in which $a \in \{IV, \delta\}$.
Let $\mathcal{H}_a^q$ be the connect graph after q-th query. Then $|V_{\mathcal{H}_{IV}^q \cup \mathcal{H}_\delta^q}| \leq q + 2$.

Claim $Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i]$.
If $\mathcal{E}$ occurs in i-th query, then there exists a Path P with $IV, \delta \in P$ and $P \subseteq G_H^i$. We will give proof of $P - (f_1, f_2, f_3) \subseteq \mathcal{H}_{IV}^{i-1} \cup \mathcal{H}_\delta^{i-1}$. If that is true, then $f_1 \in \mathcal{H}_{IV}^{i-1}$ and $f_3 \in \mathcal{H}_\delta^{i-1}$. Since (f_1, f_2, f_3) is in path P, two connected graph in $P - (f_1, f_2, f_3)$ denoted $\mathcal{H}_1$ and $\mathcal{H}_2$. Since $IV \in P$, we have $IV \in \mathcal{H}_1 \cup \mathcal{H}_2$. Let $IV \in \mathcal{H}_1$. If δ in $\mathcal{H}_1$, then that is conflict with path occur in i-th query. We have $f_1 \in \mathcal{H}_{IV}^{i-1}$ and $f_3 \in \mathcal{H}_\delta^{i-1}$.

Claim $\Pr[\mathcal{E}] \leq \frac{q}{2^{n-1}}$, $\iota \in [1..12]$.

If $Arc_\delta(f_3(x,k,y))$ occurs via an $E - query(x,k)$, then y is a random value from a set of size at least $2^n - (i-1)$. Then $f_3(x_i,k_i,y_i)$ is a random value from a set of size at least $2^n - (i-1)$. Let the $\mathcal{A}$ try q time, then

$$\Pr[Arc_\delta(f_3)] \leq \frac{q}{2^n - (i-1)}.$$

Alternatively, if $\mathcal{E}_i$ occurs via an $E^{-1} - query(y,k)$, then $f_3(x_i,k_i,y_i)$ is still a random value from a set of size at least $2^n - (i-1)$. Then

$$\Pr[\mathcal{E}] \leq \Pr[Arc_\delta(f_3)] \leq \frac{q}{2^n - (i-1)} \leq \frac{q}{2^{n-1}}.$$

Claim $\Pr[\mathcal{E}] \leq \frac{q\cdot(q+1)}{2^n}$, $\iota \in [13..20]$.

Given $\overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0$, the event $\mathcal{E}_i$ occurs in case that, the return vertex of i-th query has been exist in vertexes set $\mathcal{H}_{IV}^{i-1}$ and $\mathcal{H}_\delta^{i-1}$.

$$\Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] \leq \Pr[Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i | Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i].$$

If $\mathcal{E}_i$ occurs via an $E - query(x_i,k_i)$, then y_i is a random value from a set of size at least $2^n - (i-1)$. Then $f_3(x_i,k_i,y_i)$ is a random value from a set of size at least $2^n - (i-1)$. We also have $|V_{\mathcal{H}_\delta^{i-1}}| \leq i$. So,

$$\Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \frac{i}{2^n - (i-1)}.$$

Alternatively, if $\mathcal{E}_i$ occurs via an $E^{-1} - query(y_i,k_i)$, then $f_1(x_i,k_i,y_i)$ is a random value from a set of size at least $2^n - (i-1)$. Then

$$\Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1) \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] = \Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i | Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i]$$

$$\Pr[\mathcal{E}_i | \overline{\mathcal{E}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \leq \frac{i}{2^n - (i-1)}.$$

We have $\Pr[\mathcal{E}] \leq \sum_{i=1}^{q} \frac{i}{2^n-(i-1)} \leq \frac{q(q+1)}{2^n}$.

Claim $\Pr[\mathcal{E}] = 1$, $\iota \in [21..64]$.

For given IV and δ, we have $Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_\delta^{i-1}}(f_3)_i] = 1$. □

The following theorem is proof of Theorem5.

Theorem 9. *Fix $n \geq 1$, given δ, message padding is plain padding,*

Group 1 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \geq q/2^n$ *for any* $q \geq 1$ *and* $\iota \in [1..12]$.
Group 2 Scheme $Adv_{H_\iota}^{pre}(\mathcal{A}) \geq q(q+1)/2^{n+2}$ *for any* $q \geq 1$ *and* $\iota \in [13..20]$.

Proof. Let $\mathcal{A}^{?,?}$ be an adversary attacking H_ι.

Claim $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq q/2^n$ for any $q \geq 1$ and $\iota \in [1..12]$.

Followed the attack on Theorem 4, let adversary $\mathcal{A}$ only ask $E-query(x,k)$. In each query, select x_i and k_i to satisfy $f_1(x_i,k_i,?) \in \mathcal{H}_{IV}$. Let $\mathcal{C}$ be the event of $\delta \in \mathcal{H}_{IV}$ and at least a edge in $\mathcal{H}_{IV}$. Let $\mathcal{C}_i$ be the event that $\mathcal{C}$ occurs by the i-th query. Define $\mathcal{C}_0$ be the null event. Then $\Pr[\mathcal{C}] = \sum_{i=1}^{q} \Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0]$. We have $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq \Pr[\mathcal{C}]$. If adversary $\mathcal{A}$ builds a connected Graph $\mathcal{H}_{IV}$ with $\delta \in \mathcal{H}_{IV}$, then there is a path P from IV to δ. If $\mathcal{C}_i$ occurs via an $E-query(x_i,k_i)$, then y_i is a random value from a set of size at most 2^n. Then $f_3(x_i,k_i,y_i)$ is a random value from a set of size at most 2^n. So,

$$\Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_{\delta}^{i-1}}(f_3)_i] \geq \Pr[Ver_{\mathcal{H}_{\delta}^{i-1}}(f_3)_i] \geq \frac{1}{2^n}$$

$$\Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \geq \frac{1}{2^n}.$$

We have $\Pr[\mathcal{E}] \geq \sum_{i=1}^{q} \frac{1}{2^n} = \frac{q}{2^n}$.

Claim $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq q(q+1)/2^{n+2}$ for any $q \geq 1$ and $\iota \in [13..20]$.

In the proof of Theorem3, let adversary $\mathcal{A}$ ask $E - query(x,k)$ and $E^{-1} - query(y,k)$, alternately. In odd-th query, select x and k to satisfy $f_1(x,k,?) \in \mathcal{H}_{IV}$, in even-th query, select y and k to satisfy $f_3(?,k,y) \in \mathcal{H}_\delta$. Let $\mathcal{C}$ be the event of $\delta \in \mathcal{H}_{IV}$ and at least a edge in $\mathcal{H}_{IV}$. Let $\mathcal{C}_i$ be the event that $\mathcal{C}$ occurs by the i-th query. Define $\mathcal{C}_0$ be the null event. Then $\Pr[\mathcal{C}] = \sum_{i=1}^{q} \Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0]$. We have $Adv^{coll}_{H_\iota}(\mathcal{A}) \geq \Pr[\mathcal{C}]$. If adversary $\mathcal{A}$ build a connected Graph $\mathcal{H}_{IV}$ with $\delta \in \mathcal{H}_{IV}$, then there is a path P from IV to δ. Given $\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{C}}_0$, the event $\mathcal{C}_i$ occurs in case that, the return vertex of i-th query $f_3(x_i,k_i,y_i)$ has been exist in vertexes set $V_{\mathcal{H}_{IV}^{i-1}}$. If $\mathcal{C}_i$ occurs via an $E - query(x_i,k_i)$, then y_i is a random value from a set of size at most 2^n. Then $f_3(x_i,k_i,y_i)$ is a random value from a set of size at most 2^n. So,

$$\Pr[Ver_{\mathcal{H}_{IV}^{i-1}}(f_1)_i \wedge Ver_{\mathcal{H}_{\delta}^{i-1}}(f_3)_i] \geq \Pr[Ver_{\mathcal{H}_{\delta}^{i-1}}(f_3)_i]$$

$$\Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \geq \frac{\lfloor i/2 \rfloor + 1}{2^n}.$$

If $\mathcal{C}_i$ occurs via an $E^{-1} - query(y_i,k_i)$, then x_i is a random value from a set of size at most 2^n. Then

$$\Pr[\mathcal{C}_i|\overline{\mathcal{C}}_{i-1} \wedge \ldots \wedge \overline{\mathcal{E}}_0] \geq \frac{\lfloor i/2 \rfloor + 1}{2^n}.$$

We have $\Pr[\mathcal{E}] \geq \sum_{i=1}^{q} \frac{i}{2^{n+1}} = \frac{q(q+1)}{2^{n+2}}$. □