Improved indifferentiability security analysis of chopMD Hash Function

Donghoon Chang[1]* and Mridul Nandi[2]

[1] Center for Information Security Technologies (CIST)
Korea University, Seoul, Korea
dhchang@cist.korea.ac.kr
[2] CINVESTAV-IPN, Mexico City
mridul.nandi@gmail.com

Abstract. The classical design principle Merkle-Damgård [13, 6] is scrutinized by many ways such as Joux's multicollision attack, Kelsey-Schneier second preimage attack etc. In TCC'04, Maurer *et al.* introduced a strong security notion called as "indifferentiability" for a hash function based on a compression function. The classical design principle is also insecure against this strong security notion whereas chopMD hash is secure with the security bound roughly $\sigma^2/2^s$ where s is the number of chopped bits and σ is the total number of message blocks queried by a distinguisher. In case of $n = 2s$ where n is the output size of a compression function, the value σ to get a significant bound is $2^{s/2}$ which is the birthday complexity, where the hash output size is s-bit. In this paper, we present an improved security bound for chopMD. The improved bound shown in this paper is $(3(n-s)+1)q/2^s + q/2^{n-s-1} + \sigma^2/2^{n+1}$ where q is the total number of queries. In case of $n = 2s$, chopMD is indifferentiably-secure if $q = \mathrm{O}(2^s/(3s+1))$ and $\sigma = \mathrm{O}(2^{n/2})$ which are beyond the birthday complexity. We also present a design principle for an n-bit hash function based on a compression function $f : \{0,1\}^{2n+b} \to \{0,1\}^n$ and show that the indifferentiability security bound for this hash function is roughly $(3n+1)\sigma/2^n$. So, the new design of hash function is second-preimage and r-multicollision secure as long as the query complexity (the number of message blocks queried) of an attacker is less than $2^n/(3n+1)$ or $2^{n(r-1)/r}$ respectively.

1 Introduction

In TCC 2004, Maurer *et al.* [11] introduced the notion of indifferentiability which is more stronger notion than classical indistinguishability security notion. They have shown that if a cryptosystem $\mathcal{P}(\mathcal{G})$ based on a random oracle $\mathcal{G}$ is secure then the security of $\mathcal{P}(H^{\mathcal{F}})$ based on Merkle-Damgård (MD) [13, 6] hash function

* The first author was supported by the MIC(Ministry of Information and Communication), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute of Information Technology Advancement) (IITA-2008-(C1090-0801-0025))

H with a random oracle [1, 15] as an underlying compression function, is secure provided the hash function is indifferentiable. Informally, $H^{\mathcal{F}}$ is indifferentiable from random oracle if there is no efficient attacker (or distinguisher) which can distinguish $\mathcal{F}$ and the hash function based on it from a random oracle R and an efficient simulator of $\mathcal{F}$. Here R is a random oracle with (finite) domain and range same as that of H. In case of Indistinguishability, the distinguisher only needs to tell apart H from $\mathcal{G}$ without any help of oracle $\mathcal{F}$. Thus, the notion of indifferentiability is stronger and it is important when we consider attacks on a cryptosystem based on some ideal primitive where the attacker has some access on the computation of the primitive. In the case of hash function $H^{\mathcal{F}}$, the attacker can also compute $\mathcal{F}$ as it is a random oracle which can be computed publicly. On the other hand, if the attacker does not have that access (to the random oracle) then merely indistinguishability will suffice to preserve the security of the cryptosystem.

In Crypto 2005, Coron *et al.* [5] proved that the classical MD iteration is not indifferentiable with random oracle when the underlying compression function is random oracle. They have also stated indifferentiability for chopMD, prefix-free MD (or pfMD), NMAC construction, HMAC construction, and provided a bound for these as $\mathrm{O}(\sigma^2/2^n)$ where σ is the total number of message blocks queried by a distinguisher, the hash output size is n, and the number of chopped bits is n. Thus, according to their claim, chopMD is secure in this strong notion as long as the total number of message blocks queried is $\sigma = \mathrm{O}(2^{n/2})$. In Asiacrypt 2006, Chang *et al.* [4] also have provided a concrete security analysis of the many indifferentiable hash constructions. They have provided a security analysis for double length hash function based on prefix free padding. In Asiacrypt 2006, Bellare and Ristenpart [2] proposed an indifferentiably-secure domain extension called by EMD which also preserves pseudorandomness and collision resistance. Very recently in Asiacrypt 2007, Hirose *et al.* [7] introduced an indifferentiably-secure domain extension called by MDP which also preserves pseudorandomness, collision resistance and unforgeability. All of these constructions have bounds of the form of birthday collision probability. Recently in Crypto 2007, Maurer and Tessaro [12] firstly presented a construction which has security beyond the birthday barrier. Table 1 summarizes the security bound of above constructions.

Our Results. In this paper, we prove a better bound of chopMD which is beyond the birthday bound. We prove that chopMD is secure if $\sigma = \mathrm{O}(2^{n/2})$ and the number of queries $Q = \mathrm{O}(\mathbf{min}(2^{n-s-1}, \frac{2^s}{3(n-s)+1}))$, where n is the output length of the compression function and s is the chopped bit length and σ is the total number of message blocks queried. When $s = n/2$ our bound shows that chopMD is secure as long as the number of queries is less that $2^s/(3s+1)$ which is better than the original proposal (where security is guaranteed only when the number of queries is less than $2^{s/2}$). As a result we propose a wide pipe version of MD-hash function which is second-preimage and r-multicollision secure as long as the query complexity (the number of message blocks queried) of an attacker is less than $2^n/(3n+1)$ or $2^{n(r-1)/r}$ respectively. This hash function is more efficient to the Lucks' [10] wide pipe hash design as our hash function does not need the post-processor.

Domain Extensions	The value σ to get a significant bound
chopMD [5] prefix-free MD [4, 5] NMAC construction [5] HMAC construction [5] EMD [2] MDP [7]	$2^{s/2}$: the Birthday Bound
prefix-free chopMD [12]	2^s : Beyond the Birthday Bound
chopMD [This paper]	$2^s/(3s+1)$: Beyond the Birthday Bound

Table 1. Comparison of Indifferentiable Security when *the hash output size* is s and the chopped bit size is s and σ is the total number of message blocks queried by a distinguisher. Note that q is less than σ.

Organization. In section 2, we first state some important definitions and results related to our paper. We state an important result known as strong interpolation theorem in this section. Then in Section 3, we provide a concrete and improved security analysis for chopMD. As an application of the improved security analysis of chopMD, we propose a secure chopDBL hash design in section 4. Finally, we conclude.

2 Some Notations and Results

Counting. Let $\mathcal{F} := \text{Func}(n+b, n)$, the set of all functions $f : \{0,1\}^n \times \{0,1\}^b \rightarrow \{0,1\}^n$. It is easy to see that $|\mathcal{F}| = 2^{n2^{n+b}}$. Now, for any distinct a_i's, the number of functions f such that $f(a_1) = z_1, \cdots, f(a_q) = z_q$ is exactly $2^{n(2^{n+b}-q)}$ because, the outputs of q elements are fixed and the rest $(2^{n+b} - q)$ many outputs can be chosen in $(2^n)^{(2^{n+b}-q)}$ many ways. Thus, $\Pr_{\mathtt{u}}[\mathtt{u}(a_1) = z_1, \cdots, \mathtt{u}(a_q) = z_q] = \frac{1}{2^{nq}}$ where $\mathtt{u}$ is the uniform random function on $\mathcal{F}$ (an uniform random variable taking values on $\mathcal{F}$).

Inequalities. $\mathbf{P}(m, r) = m(m-1)\cdots(m-r+1)$ where $0 \leq r \leq m$. By our convention, $\mathbf{P}(m, 0) = 1$. We state two inequalities which will be used in this paper.

[ineq-1] For any $0 \leq a_i \leq 1$, $\prod_{i=1}^{k}(1 - a_i) \geq 1 - \sum_{i=1}^{k} a_i$. One can prove it by induction on k.

[ineq-2] $\mathbf{P}(m - x, r) \geq m^r \times (1 - \frac{(x+r)^2}{2m})$ where $m \geq x + r$. This is followed from ineq-1, by choosing $a_i = \frac{x+i}{m}, 0 \leq i \leq r - 1$.

MD-hash. We fix an initial value $\mathrm{IV} \in \{0,1\}^n$ throughout the paper. Given a function $f : \{0,1\}^{n+b} \to \{0,1\}^n$ we define

$$\mathrm{MD}^f(m_1, \cdots, m_\ell) = f(f(\cdots f(f(\mathrm{IV}, m_1), m_2), \cdots), m_\ell)$$

where $m_i \in \{0,1\}^b$. MD^f is popularly known as Merkle-Damgård hash function with underlying compression function f. We define $\mathrm{MD}^f(\lambda) = \mathrm{IV}$ where λ is the empty string. Given $p = (m_1, \cdots, m_\ell) \in (\{0,1\}^b)^\ell$ with $\ell \geq 1$ we define

- $\mathrm{last}(p) = m_\ell$.
- If $\ell \geq 2$ we write $\mathrm{cut}(p) = (m_1, \cdots, m_{\ell-1})$, otherwise $\mathrm{cut}(m_1) = \lambda$.
- Note that $p = (\mathrm{cut}(p), \mathrm{last}(p))$ and $\mathrm{MD}^f(p) = f(\mathrm{MD}^f(\mathrm{cut}(p)), \mathrm{last}(p))$.

chopMD. For $0 \leq s \leq n$ we define $\mathrm{chop}_s(x) = x_R$ where $x = x_L \parallel x_R$ and $|x_L| = s$. In this paper, we fix $0 < s < n$ and define $\mathrm{chopMD}^f(M) = \mathrm{chop}_s(\mathrm{MD}^f(M))$.

Padding. Note that both MD and chopMD have domain $(\{0,1\}^b)^+$. We write $||M|| = k$ if $M \in (\{0,1\}^b)^k$ and k is called as the number of blocks of M. We say M' is a prefix of M if $M', M \in (\{0,1\}^b)^+$ and $M = M' \parallel x$ for some $x \in (\{0,1\}^b)^*$. We say any injective function $\mathsf{pad} : \{0,1\}^* \to (\{0,1\}^b)^+$ as a padding rule. A padding rule pad is called a *prefix free* if $M_1 \neq M_2 \Rightarrow \mathsf{pad}(M_1)$ is not a prefix of $\mathsf{pad}(M_2)$. For any such prefix-free padding rule pad, pfMD is defined as follows. $\mathrm{pfMD}^f_{\mathsf{pad}}(M) = \mathrm{MD}^f(\mathsf{pad}(M))$. We also write $\mathrm{choppfMD}^f_{\mathsf{pad}}(M) = \mathrm{chop}_s(\mathrm{MD}^f(\mathsf{pad}(M)))$.

View. In this paper we consider a distinguisher $\mathcal{A}$ which has access of two oracles $\mathcal{O}_1$ and $\mathcal{O}_2$. We assume that $\mathcal{A}$ is deterministic and computationally unbounded[1]. We assume that all queries are distinct and it makes at most Q_i queries to the oracle $\mathcal{O}_i$. Suppose $\mathcal{A}$ makes M_i as $\mathcal{O}_1$-query and obtains responses h_i, $1 \leq i \leq q_1$. Similarly, the tuple of all query-responses of $\mathcal{O}_2$ is $((x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$. The combined tuple $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$ is called as the *view* of $\mathcal{A}$. We also denote $v_{\mathcal{O}_1, \mathcal{O}_2}$ to specify that the view is obtained after interacting with $\mathcal{O}_1$ and $\mathcal{O}_2$. We also denote i-th query-response pair by (X_i, Y_i), where $X_i = (x_j, m_j)$ or $X_i = M_j$ for a j. So we can define the first i query-response pairs of the tuple v by $v_i = ((X_1, Y_1), \cdots, (X_i, Y_i))$.

Advantage. Let $\mathtt{F}_i, \mathtt{G}_i$ be probabilistic oracle algorithms. We define advantage of the distinguisher $\mathcal{A}$ at distinguishing $(\mathtt{F}_1, \mathtt{F}_2)$ from $(\mathtt{G}_1, \mathtt{G}_2)$ as

$$\mathbf{Adv}_{\mathcal{A}}((\mathtt{F}_1, \mathtt{F}_2), (\mathtt{G}_1, \mathtt{G}_2)) = |\Pr[\mathcal{A}^{\mathtt{F}_1, \mathtt{F}_2} = 1] - \Pr[\mathcal{A}^{\mathtt{G}_1, \mathtt{G}_2} = 1]|.$$

Theorem 1. (Strong Interpolation Theorem) *If there is a set of good views $\mathcal{V}_{\text{good}}$ such that*

[1] computationally unbounded deterministic algorithms are as powerful as randomized algorithms.

1. for all $v \in \mathcal{V}_{\text{good}}$, $\Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \geq (1-\varepsilon) \times \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v]$ *and*
2. $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} \in \mathcal{V}_{\text{good}}] \geq 1 - \varepsilon'$

then for any $\mathcal{A}$ *we have* $\mathbf{Adv}_{\mathcal{A}}((\mathtt{F}_1, \mathtt{F}_2), (\mathtt{G}_1, \mathtt{G}_2)) \leq \varepsilon + \varepsilon'$.

Proof. Intuitively, a view of $\mathcal{A}^{\mathtt{G}_1,\mathtt{G}_2}$ is good with probability at least $1-\varepsilon'$. Moreover, $\mathcal{A}$ obtains a good view v with almost same probability for both pairs of oracles up to a factor of $(1-\varepsilon)$. Then intuitively the distinguishing advantage of $\mathcal{A}$ should be bounded by $\varepsilon+\varepsilon'$. More precisely, we prove it as in below where $\mathcal{V}^1$ denotes the set of all views v such that $\mathcal{A}$ returns 1 after obtaining the view[2]. $\mathcal{V}^0$ denotes the set of all views v such that $\mathcal{A}$ doesn't returns 1 after obtaining the view. And let $\alpha(v) = \Pr[\mathcal{A}(v_{i-1}) = X_i \text{ for all } 0 \leq i \leq q_1 + q_2]$. Our proof is directly from the idea explained in [3].

$$\begin{aligned}
&\Pr[\mathcal{A}^{\mathtt{G}_1,\mathtt{G}_2} = 1] - \Pr[\mathcal{A}^{\mathtt{F}_1,\mathtt{F}_2} = 1] \\
&\quad = \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] + \sum_{v \in \mathcal{V}^1 \setminus \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\qquad - \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] - \sum_{v \in \mathcal{V}^1 \setminus \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \\
&\quad \leq \varepsilon' + \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] - \sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \\
&\quad \leq \varepsilon' + \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) (\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] - \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v]) \\
&\quad \leq \varepsilon' + \varepsilon \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\quad \leq \varepsilon' + \varepsilon \textstyle\sum_{v \in \mathcal{V}^1 \cap \mathcal{V}_{\text{good}}} \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\quad \leq \varepsilon' + \varepsilon.
\end{aligned}$$

$$\begin{aligned}
&\Pr[\mathcal{A}^{\mathtt{F}_1,\mathtt{F}_2} = 1] - \Pr[\mathcal{A}^{\mathtt{G}_1,\mathtt{G}_2} = 1] = \Pr[\mathcal{A}^{\mathtt{G}_1,\mathtt{G}_2} \neq 1] - \Pr[\mathcal{A}^{\mathtt{F}_1,\mathtt{F}_2} \neq 1] \\
&\quad = \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] + \sum_{v \in \mathcal{V}^0 \setminus \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\qquad - \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] - \sum_{v \in \mathcal{V}^0 \setminus \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \\
&\quad \leq \varepsilon' + \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] - \sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \\
&\quad \leq \varepsilon' + \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) (\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] - \Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v]) \\
&\quad \leq \varepsilon' + \varepsilon \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \alpha(v) \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\quad \leq \varepsilon' + \varepsilon \textstyle\sum_{v \in \mathcal{V}^0 \cap \mathcal{V}_{\text{good}}} \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v] \\
&\quad \leq \varepsilon' + \varepsilon.
\end{aligned}$$

■

Indifferentiability

We give a brief introduction of indifferentiability and state significance of it. The following definition is a slightly modified version of the original definition [11, 5], where the condition that the maximum number of message blocks queried by a distinguisher is σ is not described.

[2] since $\mathcal{A}$ is deterministic algorithm the output of $\mathcal{A}$ is completely determined by the view.

Definition 1. *[11] A Turing machine C with oracle access to an ideal primitive $\mathcal{F}$ is said to be $(t_{\mathcal{A}}, t_S, q, \sigma, \varepsilon)$-indifferentiable from an ideal primitive $\mathcal{G}$ if there exists a simulator S such that for any distinguisher $\mathcal{A}$ it holds that :*

$$\mathbf{Adv}_{\mathcal{A}}((C,\mathcal{F}),(\mathcal{G},S)) = |\Pr[\mathcal{A}^{C,\mathcal{F}} = 1] - \Pr[\mathcal{A}^{\mathcal{G},S} = 1]| < \varepsilon$$

The simulator S is an interactive algorithm which has oracle access to $\mathcal{G}$ and runs in time at most t_S. The distinguisher $\mathcal{A}$ runs in time at most $t_{\mathcal{A}}$ and makes at most q queries. The total message blocks queried by $\mathcal{A}$ is at most σ.

The following Theorem [11] due to Maurer *et al.* is related to this paper. We explain the theorem for random oracle model of hash functions. Suppose a hash function (in some design of iteration) H based on a random oracle (or an ideal cipher) $\mathcal{F}$ is indifferentiable from a random oracle $\mathcal{G}$. Then a cryptosystem $\mathcal{P}$ based on the random oracle $\mathcal{G}$ is at least as secure as the cryptosystem $\mathcal{P}$ based on the hash function H in the random oracle model (or an ideal cipher model) $\mathcal{F}$. Here, $\mathcal{F}$ is the underlying compression function of H (or block-cipher in case of block cipher based hash function). The original theorem as stated below is a more general statement.

Theorem 2. *[11] Let $\mathcal{P}$ be a cryptosystem with oracle access to an ideal primitive $\mathcal{G}$. Let H be an algorithm such that $H^{\mathcal{F}}$ is indifferentiable from $\mathcal{G}$. Then cryptosystem $\mathcal{P}$ is at least as secure in the $\mathcal{F}$ model with algorithm H as in the $\mathcal{G}$ model.*

In this paper we consider $\mathcal{G}$ and $\mathcal{F}$ as the arbitrary input length random oracle R and the fixed input length random oracle f, respectively. And C is the chopMD hash function. If chopMD^f is $(t_{\mathcal{A}}, t_S, q, \sigma, \varepsilon)$-indifferentiable from the random oracle R, we also say that *the indifferentiability insecurity bound* of chopMD^f is ε.

3 Improved indifferentiability analysis of chopMD

Coron *et al.* [5] stated MD hash function is not indifferentiability-secure whereas prefix free MD construction or chopMD construction in random oracle (or in ideal cipher model) is secure against indifferentiability attack. In [5], they had proved the following statement for a distinguisher which makes queries whose total number of message blocks is σ. And u is the random oracle from the set of all $n+b$ bits to the set of n bits.

1. The indifferentiability insecurity for $\text{pfMD}^{\mathsf{u}}_{\mathsf{pad}}$ is upper bounded by $\mathrm{O}(\sigma^2/2^n)$ where pad is any prefix-free padding.
2. The indifferentiability insecurity for $\text{chopMD}^{\mathsf{u}}_{pad}$ is upper bounded by $\mathrm{O}(\sigma^2/2^s)$.

Very recently, Maurer and Tessaro considered the combination of prefix free MD and chopMD [12], i.e., $\text{choppfMD}^{\mathsf{u}}_{\mathsf{pad}}$. They proved that the indifferentiability insecurity for this combination is bounded by $\mathrm{O}(\sigma^2/2^n)$. This is an improved bound compare to the bound for chopMD. Since choppfMD outputs $n-s$ bits, the security bound is beyond the birthday barrier. A prefix-padding may cost extra overhead in terms of efficiency and designs. In this section, we show that the the prefix-padding is not necessary to obtain the similar kind of bound. In other words, we provide an improved bound of chopMD and the improved bound stated in this paper is $(3(n-s)+1)q/2^s + q/2^{n-s-1} + \sigma^2/2^{n+1}$ where q denotes the the maximum number of queries for two oracles and σ is the total number of message blocks queried by a distinguisher. If we choose $s = n/2$ then to have a significant advantage, the total number of blocks of all queries should be at least $2^s/(3s+1)$ which is far beyond the birthday attack complexity.

The organization of section 3 is as follows. In subsection 3.1, we define a set of good views $\mathcal{V}^r_{\text{good}}$ and give a lower bound of $\Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v]$ for all $v \in \mathcal{V}^r_{\text{good}}$, where $\mathtt{F}_1$ is chopMD$^{\mathtt{u}}$ and $\mathtt{F}_2$ is $\mathtt{u}$. In subsection 3.2, we give an upper bound of $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v]$ for all $v \in \mathcal{V}^r_{\text{good}}$, where $\mathtt{G}_1$ is the random oracle R and $\mathtt{G}_2$ is the simulator S^R described in subsection 3.2. In subsection 3.3, we compute ε and ε' such that for all $v \in \mathcal{V}_{\text{good}}$, $\Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \geq (1-\varepsilon) \times \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v]$ and $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} \in \mathcal{V}_{\text{good}}] \geq 1 - \varepsilon'$. Finally, based on Theorem 1 (strong interpolation theorem), we conclude in Theorem 3 that the indifferentiability insecurity bound of chopMD$^{\mathtt{u}}$ is $\varepsilon^* = \varepsilon + \varepsilon'$.

3.1 Interpolation probability of chopMD and its underlying random oracle

We first provide a lower bound on the number of functions when some inputs-outputs of f and MD^f are known. More precisely, we want to compute the number of functions f such that

$$\mathrm{MD}^f(M_j) = h_j \text{ and } f(a_i) = z_i, 1 \leq j \leq q_1, 1 \leq i \leq q_2$$

where $a_i \in \{0,1\}^{n+b}$ are distinct, $M_j \in (\{0,1\}^b)^+$ are distinct. Intuitively, we say the above set of relations is irreducible (see definition 2 in below) if $\mathrm{MD}^{\mathcal{O}_2}(M_i)$ is not determined from $\mathcal{O}_2(x_1, m_1) = z_1, \cdots, \mathcal{O}_2(x_{q_2}, m_{q_2}) = z_{q_2}$ and $\mathrm{MD}^{\mathcal{O}_2}(M_j) = h_j$ for all $j \neq i$. Thus, q_1 many outputs of MD^f add q_1 more restrictions on the outputs of f besides q_2 many input-output relations of f. Hence the number of functions f should be close to $2^{n(2^{n+b}-q_1-q_2)}$. In lemma 1 we will show that the number of functions is at least $(1-\nu) \times 2^{n(2^{n+b}-q_1-q_2)}$ for some positive ν (stated in the lemma 1) close to zero. The above statement is also equivalent to

$$\Pr_{\mathtt{u}}[\mathrm{MD}^{\mathtt{u}}(M_j) = h_j, \mathtt{u}(a_i) = z_i \ \forall 1 \leq j \leq q_1, 1 \leq i \leq q_2] \geq \frac{1}{2^{n(q_1+q_2)}} \times (1-\nu).$$

Definition 2. *The set of relations*

$$\begin{aligned}&\mathrm{MD}^{\mathcal{O}_2}(M_1) = h_1, \cdots, \mathrm{MD}^{\mathcal{O}_2}(M_{q_1}) = h_{q_1},\\ &\mathcal{O}_2(x_1, m_1) = z_1, \cdots, \mathcal{O}_2(x_{q_2}, m_{q_2}) = z_{q_2} \quad \cdots\cdots \ \text{(rel-A)}\end{aligned}$$

is said to be irreducible *if* $M_1, \cdots, M_{q_1} \in (\{0,1\}^b)^+$ *are distinct,* $(x_1, m_1), \cdots, (x_{q_2}, m_{q_2}) \in \{0,1\}^{n+b}$ *are distinct,* $h_1, \cdots, h_{q_1} \in \{0,1\}^n$ *are distinct from* x_i*'s and* IV *and finally the value of* $\mathrm{MD}^{\mathcal{O}_2}(M_i)$ *is not determined from the relations* $\mathcal{O}_2(x_1, m_1) = z_1, \cdots, \mathcal{O}_2(x_{q_2}, m_{q_2}) = z_{q_2}$*. A tuple of elements* $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$ *is* irreducible *if the above* rel-A *is irreducible*[3]*.*

Remark 1. Intuitively, it says that there is no redundant relation in rel-A. All the inputs of $\mathcal{O}_1$ and $\mathcal{O}_2$ are distinct. $\mathcal{O}_1(M_i) = \mathrm{MD}^{\mathcal{O}_2}(M_i)$ is also not determined from the relations $\mathcal{O}_2(x_1, m_1) = z_1, \cdots, \mathcal{O}_2(x_{q_2}, m_{q_2}) = z_{q_2}$. Moreover, as h_i's are distinct from x_i's and IV, $\mathrm{MD}^{\mathcal{O}_2}(M_i)$ is also not determined from $\mathcal{O}_2(x_1, m_1) = z_1, \cdots, \mathcal{O}_2(x_{q_2}, m_{q_2}) = z_{q_2}$ and $\mathrm{MD}^{\mathcal{O}_2}(M_j) = h_j$ for all $j \neq i$.

Lemma 1. *Let a tuple* $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$ *be irreducible then the number of functions* f *such that*

1. $\mathrm{MD}^f(M_1) = h_1, \cdots, \mathrm{MD}^f(M_{q_1}) = h_{q_1}$ *and*
2. $f(x_1, m_1) = z_1, \cdots, f(x_{q_2}, m_{q_2}) = z_{q_2}$*.*

is at least $\frac{|\mathcal{F}|}{2^{n(q_1+q_2)}} \times (1 - \frac{\sigma^2}{2^{n+1}})$ *where* σ *is the total number of message blocks queried. In other words,*

$$\Pr_{\mathtt{u}}[\mathrm{MD}^{\mathtt{u}}(M_1) = h_1, \cdots, \mathrm{MD}^{\mathtt{u}}(M_{q_1}) = h_{q_1}, \mathtt{u}(x_1, m_1) = z_1, \cdots, \mathtt{u}(x_{q_2}, m_{q_2}) = z_{q_2}]$$

$$\geq \frac{1}{2^{n(q_1+q_2)}} \times (1 - \frac{\sigma^2}{2^{n+1}}).$$

Proof. See the appendix. ∎

Now we compute the joint probability for $\mathrm{chopMD}^{\mathtt{u}}$ and $\mathtt{u}$. The next lemma is analogue version of Lemma 1 for chopMD hash function instead of MD hash function. Here, we allow collisions among outputs of chopMD. Intuitively, if $\mathrm{chopMD}^{\mathtt{u}}$ behaves as an uniform random function then $\Pr_{\mathtt{u}}[\mathrm{chopMD}^{\mathtt{u}}(M_j) = y_j, \mathtt{u}(x_i, m_i) = z_i, 1 \leq j \leq q_1, 1 \leq i \leq q_2]$ ideally should be $\frac{1}{2^{nq_2+(n-s)q_1}}$. Since $\mathrm{chopMD}^{\mathtt{u}}(M_j) = y_j$ has some influence on the intermediate computations we would rather expect a probability close to the above probability. In lemma 2 we show that the for a given choices of inputs and outputs satisfying some conditions (stated in the lemma 2) the above probability is at least $\frac{1-\Delta}{2^{nq_2+(n-s)q_1}}$ for some positive Δ (defined in the lemma 2) which is close to zero for reasonable choices of parameters.

Lemma 2. *The number of functions* f *such that*

1. $\mathrm{chopMD}^f(M_1^1) = \cdots = \mathrm{chopMD}^f(M_{r_1}^1) = y^1, \cdots, \mathrm{chopMD}^f(M_1^t) = \cdots = \mathrm{chopMD}^f(M_{r_t}^t) = y^t$ *and*

[3] From the definition it is clear that irreducibility of the relation does not depend on the choice of the functions or oracles $\mathcal{O}_1$ and $\mathcal{O}_2$. This only depends on M_j's, h_j's, (x_i, m_i)'s and z_i's, $1 \leq j \leq q_1$ and $1 \leq i \leq q_2$.

2. $f(x_1, m_1) = z_1, \cdots, f(x_{q_2}, m_{q_2}) = z_{q_2}$.

is at least $|\mathcal{F}| \times \frac{1-\Delta}{2^{nq_2+(n-s)q_1}}$ *where*

$$\Delta = \frac{r(q_1+q_2)}{2^s} + \frac{\sigma^2}{2^{n+1}}, \quad r = \max_i r_i, \quad \sum_i r_i = q_1.$$

Here, σ *is the total number of message blocks queried.* M^i_j*'s are distinct elements from* $(\{0,1\}^b)^+$ *such that the value of* $\mathrm{MD}^f(M_i)$ *is not determined from the relations* $f(x_1, m_1) = z_1, \cdots, f(x_{q_2}, m_{q_2}) = z_{q_2}$*. The values of* (x_i, m_i)*'s are distinct elements from* $\{0,1\}^n \times \{0,1\}^b$*. In terms of probability, we have*

$$\Pr_{\mathtt{u}}[\mathrm{chopMD}^{\mathtt{u}}(M^i_j) = y^i, \mathtt{u}(x_1, m_1) = z_1, \cdots, \mathtt{u}(x_{q_2}, m_{q_2}) = z_{q_2}, \ \forall i, j] \geq \frac{1-\Delta}{2^{nq_2+(n-s)q_1}}.$$

Proof. See the appendix. ■

Definition 3. *A view* $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$ *is said to be* r*-good if* (x_i, m_i)*'s are distinct,* M_j*'s are distinct,* $\mathrm{MD}^{\mathcal{O}_2}(M_j)$ *is not determined from the relations* $\mathcal{O}_2(x_i, m_i) = z_i$ *and there is no* r*-multicollision in* $\mathrm{chop}(z_i)$*'s and* h_i*'s. The set of all* r*-good views is denoted by* $\mathcal{V}^r_{\mathrm{good}}$*.*

By using lemma 2 we have similar result for $\mathrm{chopMD}^{\mathtt{u}}$ and $\mathtt{u}$.

Proposition 1. *For any* r*-good view* $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$*, the probability that* v *is a view when* $\mathcal{A}$ *is interacting with* $\mathrm{chopMD}^{\mathtt{u}}$ *and* $\mathtt{u}$*, is at least* $\frac{1-\Delta}{2^{nq_2+(n-s)q_1}}$ *where* $\Delta = \frac{r(q_1+q_2)}{2^s} + \frac{\sigma^2}{2^{n+1}}$ *and* σ *is the total number of message blocks queried.*

3.2 Interpolation probability of A Simulator and Random Oracle

Now we define a simulator S which almost behaves as a random oracle. Moreover, for an $(n-s)$-bit outputting random oracle R, responses of MD^S will match with R. By the notation $x \in_R A$ we mean that x is chosen uniformly from A and it is independent with all previously defined random variables.

Definition of Simulator

Initialization :

1. A partial function $e_1 : \{0,1\}^{n+b} \to \{0,1\}^n$ initialized as empty,
2. a partial function $e_1^* = \mathrm{MD}^{e_1} : (\{0,1\}^b)^* \to \{0,1\}^n$ initialized as $e_1^*(\lambda) = \mathrm{IV}$.
3. a set $C = \{\mathrm{IV}\}$.

$\boxed{\text{On query } S^R(x, m) :}$

001 `if` $(e_1(x, m) = x')$
  `return` x';

002 `else if` $(\exists\, M', e_1^*(M') = x)$
  $y = R(M', m)$;
  `choose` $w \in_R \{0,1\}^s \setminus \{w' : w' \parallel y \in C \cup \{x\}\}$;
  `define` $e_1(x, m) = z := w \parallel y$;
  `define` $C = C \cup \{x, z\}$;
  `define` $e_1^*(M', m) = z$;
  `return` z;

003 `else`
  $y \in_R \{0,1\}^{n-s}$;
  `choose` $w \in_R \{0,1\}^s \setminus \{w' : w' \parallel y \in C \cup \{x\}\}$;
  `define` $e_1(x, m) = z := w \parallel y$;
  `define` $C = C \cup \{x, z\}$;
  `return` z;

In 002, we have $w \in_R \{0,1\}^s \setminus \{w' : w' \parallel y \in C \cup \{x\}\}$. This is not possible if and only if the above set becomes empty. Note that after i^{th} query the size of C is less than or equal to $(2i+1)$. Thus we assume that q_2, the maximum number of queries to the simulator (and hence for oracle $\mathcal{O}_2$) satisfies the condition $2q_2 + 2 < 2^s$ equivalently $q_2 \le 2^{s-1} - 2$.

Some important observations

Distinct Query. Suppose $\mathcal{A}^{\mathcal{O}_1, \mathcal{O}_2}$ is an oracle algorithm where $\mathcal{O}_1 = R$ and $\mathcal{O}_2 = S^R$. Note that S^R responses identically in identical queries and so does R. Same property is true for chopMD^f and f. Hence we assume that all queries to $\mathcal{O}_1$ and $\mathcal{O}_2$ are distinct.

chopMD$^{\mathbf{S}}$= R. All responses of S are distinct and distinct from IV and the first n-bits of all previous S-queries. Whenever $\text{MD}^S(M)$ is computable from the all previous query-responses, we have $\text{chopMD}^S(M) = R(M)$. Thus, $\text{chopMD}^{\mathcal{O}_2}(M) = \mathcal{O}_1(M)$ whenever $\text{chopMD}^{\mathcal{O}_2}(M)$ is computable from $\mathcal{O}_2$ query-responses only. Obviously this is true when $\mathcal{O}_1 = \text{chopMD}^f$ and $\mathcal{O}_2 = f$. Thus, we assume that $\mathcal{A}$ do not make any $\mathcal{O}_1$-query which is computable from the previous query-responses of $\mathcal{O}_2$. More particularly, we can remove all those $\mathcal{O}_1$-queries from the final view which are computable from the query-responses of $\mathcal{O}_2$.

Distribution. Because of the above two assumptions, the last $(n-s)$ bits of outputs of $S^R(\cdot)$ and outputs of $R(\cdot)$ are uniformly and independently distributed over the set $\{0,1\}^{n-s}$. By our assumption, whenever line002 is executed, $\mathcal{A}$ does not make (M', m)-query to R. Thus, *the output distribution of* $R(\cdot)$ *and* $S(\cdot)$ *are independent.*

Now, a typical view of $\mathcal{A}^{\mathcal{O}_1, \mathcal{O}_2}$ is a tuple

$$v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$$

where $\mathcal{O}_1(M_j) = h_j$ and $\mathcal{O}_2(x_i, m_i) = z_i$. Moreover, (x_i, m_i)'s are distinct, M_j's are distinct and $\mathrm{MD}^{\mathcal{O}_2}(M_j)$ is not determined from the relations $\mathcal{O}_2(x_i, m_i) = z_i$. Now we compute the joint interpolation probabilities for S and R. More precisely, $p := \Pr[R(M_j) = h_j \ \forall j \text{ and } S(x_i, m_i) = z_i \ \forall i]$. Since outputs of S and outputs of R are independently distributed, it is sufficient to compute the joint probabilities of S and R separately. Obviously, $\Pr[R(M_j) = h_j \ \forall j] = \frac{1}{2^{(n-s)q_1}}$. Now on i^{th} query of S, the response of (x_i, m_i) is z_i with probability at most $\frac{1}{2^{n-s}} \times \frac{1}{2^s - \ell_i}$ where

$$\begin{aligned}\ell_i = &|\{k : 1 \leq k \leq q_2, \mathrm{chop}(x_k) = \mathrm{chop}(z_i)\}| \\ &+ |\{k : 1 \leq k \leq q_2, \mathrm{chop}(z_k) = \mathrm{chop}(z_i)\}| + 1.\end{aligned}$$

ℓ_i is the upper bound of the size of the set $\{w' : w' \parallel y \in C \cup \{x\}\}$ appeared in the i^{th} query of S. Multiplying all these probabilities we obtain $\Pr[S(x_i, m_i) = z_i \ \forall i] \leq \frac{1}{2^{nq_2}} \times \frac{1}{1 - \sum_i \ell_i / 2^s}$.

It is easy to see that for any r-good view $\sum_i \ell_i \leq (2r+1)q_2$. Thus, we have proved the following result.

Proposition 2. *For any r-good view $v = ((M_1, h_1), \cdots, (M_{q_1}, h_{q_1}), (x_1, m_1, z_1), \cdots, (x_{q_2}, m_{q_2}, z_{q_2}))$, the probability that v is a view when $\mathcal{A}$ is interacting with the simulator S and a random oracle R, is at most $\frac{1}{2^{nq_2 + (n-s)q_1}} \times \frac{1}{1 - (2r+1)q_2/2^s}$.*

3.3 Indifferentiability Security Bound of chopMD

Now we compute ε and ε' such that for all $v \in \mathcal{V}_{\text{good}}$, $\Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v] \geq (1 - \varepsilon) \times \Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v]$ and $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} \in \mathcal{V}_{\text{good}}] \geq 1 - \varepsilon'$, where $\mathtt{F}_1$ is $\mathrm{chopMD}^{\mathtt{u}}$, $\mathtt{F}_2$ is $\mathtt{u}$, $\mathtt{G}_1$ is the random oracle R and $\mathtt{G}_2$ is the the simulator S^R.

The Value of ε. By proposition 1 and 2, for all $v \in \mathcal{V}_{\text{good}}$, we have a lower bound of $\Pr[v_{\mathtt{F}_1,\mathtt{F}_2} = v]$ and an upper bound of $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} = v]$. So, we can choose $\varepsilon = \frac{(3r+1)q_2 + rq_1}{2^s} + \frac{\sigma^2}{2^{n+1}}$. When $r = n - s$, $\varepsilon = \frac{(3(n-s)+1)q_2 + (n-s)q_1}{2^s} + \frac{\sigma^2}{2^{n+1}}$.

The Value of ε'. Now we compute ε' such that $\Pr[v_{\mathtt{G}_1,\mathtt{G}_2} \in \mathcal{V}_{\text{good}}] \geq 1 - \varepsilon'$, where $\mathtt{G}_1$ is the random oracle R and $\mathtt{G}_2$ is the simulator S^R. $v_{\mathtt{G}_1,\mathtt{G}_2} \in \mathcal{V}_{\text{good}}$ means that the view $v_{\mathtt{G}_1,\mathtt{G}_2}$ is r-good. Therefore, we have to prove that the upper bound of the probability that there is a r-multicollision among q uniformly and independently chosen $(n-s)$-bits is ε'. Let's compute this ε' as follows. Let us denote the $\mu(n-s, r, q)$ for the probability that there is a r-multicollision among q uniformly and independently chosen $(n-s)$-bits. Now it is easy to see that $\mu(n-s, r, q) \leq \frac{\binom{q}{r}}{2^{(n-s)(r-1)}}$. Now we choose $r = n-s$ and hence $\mu(n-s, r, q) \leq (q/2^{n-s-1})^r \leq q/2^{n-s-1}$ if $q \leq 2^{n-s-1}$. Since $\mathrm{chop}(S(\cdot))$ and $R(\cdot)$ uniformly and independently distributed over $\{0,1\}^{n-s}$, a $(n-s)$-good view is obtained by $\mathcal{A}^{S,R}$ with probability at least $1 - q/2^{n-s-1}$, where $q = q_1 + q_2$. Therefore, we can choose $\varepsilon' = q/2^{n-s-1}$ when $r = n - s$.

Now, by using proposition 1 and 2 and strong interpolation theorem we obtain our following main theorem of the section. Here, $\varepsilon^* = \varepsilon + \varepsilon'$.

Theorem 3. *The chopMD construction is $(t_{\mathcal{A}}, t_S, q, \sigma, \varepsilon^*)$-indifferentiable from a random oracle, in the random oracle model for the compression function, for any $t_{\mathcal{A}}$, with $t_S = \ell \cdot \mathrm{O}(q^2)$ and $\varepsilon^* = \frac{(3(n-s)+1)q_2+(n-s)q_1}{2^s} + \frac{q}{2^{n-s-1}} + \frac{\sigma^2}{2^{n+1}} = \mathrm{O}(\frac{nq}{2^s} + \frac{q}{2^{n-s}} + \frac{\sigma^2}{2^n})$, where $q = q_1 + q_2$.*

4 chopDBL hash functions and its security analysis

A r-multicollision for a hash function H is a r-set $\{X_1, \cdots, X_r\}$ such that $H(X_1) = \cdots = H(X_r)$. In [8] it is shown that the r-multicollision can be found in the classical MD hash function in roughly $2^{n/2}$ complexity. For a random oracle it needs [14] roughly $2^{n(r-1)/r}$ complexity. Moreover, Kelsey-Schneier [9] found a second preimage attack which needs roughly $2^{n/2}$ queries for classical MD hash function. But for a random oracle to have a second preimage attack we need at least 2^n queries. Thus MD hash function is not good in terms of multicollision and second-preimage attack. Lucks designed a wide pipe hash which is secure against these attacks.

We first define Lucks wide pipe design. In his design let $F : \{0,1\}^{w+b} \to \{0,1\}^w$ and $g : \{0,1\}^w \to \{0,1\}^n$ be two independently distributed random oracles. The wipe pipe hash [10] is defined as $g(\mathrm{MD}^F(M))$ for any padded message M. In [10], it was shown that

- the second preimage attack for the wide pipe hash needs $\min\{2^{w/2}, 2^n\}$ complexity.
- the k-multicollision for the wide pipe hash needs $\min\{2^{w/2}, 2^n\}$ complexity.

Here we show that the random oracle assumption of g is redundant. More precisely, we obtain almost similar bound when g is a simply chop function. Thus we define a chopDBL hash function as

$$\mathrm{chopDBL}^F(m_1, \cdots, m_\ell) = \mathrm{chop}_n(\mathrm{MD}^F(m_1, \cdots, m_\ell)).$$

One can compute $F : \{0,1\}^{2n+b} \to \{0,1\}^{2n}$ based on two independent random oracles $f_1, f_2 : \{0,1\}^{2n+b} \to \{0,1\}^n$ as $F(X) = f_1(X) \parallel f_2(X)$. As shown in the last section, we have an improved security analysis for chopMD. By using Theorem 3 we know that $\mathrm{chop}_n\mathrm{MD}^F$ is $2^n/(3n+1)$-indifferentiable secure.

Theorem 4. *The chopDBL construction is $(t_{\mathcal{A}}, t_S, q, \sigma, \varepsilon)$-indifferentiable from a random oracle, in the random oracle model for the compression function, for any $t_{\mathcal{A}}$, with $t_S = \ell \cdot O(q^2)$ and $\varepsilon = \mathrm{O}(\frac{nq}{2^n} + \frac{q}{2^n} + \frac{\sigma^2}{2^{2n}})$.*

The above theorem says that to have an indifferentiability attack we need at least $2^n/(3n+1)$ query complexity (the number of message blocks queried). Thus, if we can have second preimage attack of chopDBL with q query complexity then $q \geq 2^n/(3n+1)$. Otherwise we can distinguish chopDBL from a random oracle with less queries than $2^n/(3n+1)$. A similar argument shows that r-multicollision attack needs at least minimum of $2^{n(r-1)/r}$ and $2^n/(3n+1)$ queries. Thus in the random oracle model our new design of hash function is almost optimally secure (with respect to second preimage and multicollision).

5 Conclusion

In this paper, we present an improved security analysis for chopMD. This improved security analysis helps us how to get security beyond the birthday barrier. More precisely, we design an n-bit wide pipe hash function which has security level close to 2^n and hence we have beyond birthday barrier. The new design is much simpler and efficient. It would be interesting to see whether it preserves other properties more particularly, second preimage security.

Acknowledgement

Thank Anonymous referees for giving us valuable comments.

References

1. M. Bellare and P. Rogaway. Random Oracles Are Practical : A Paradigm for Designing Efficient Protocols. In *1st Conference on Computing and Communications Security*, ACM, pages 62–73, 1993.
2. M. Bellare and T. Ristenpart. Multi-Property-Preserving Hash Domain Extension and the EMD Transform. In *Asiacrypt'2006*, volume **4284** of *Lecture Notes in Computer Science*, pages 299–314, Springer-Verlag, 2006.
3. D. J. Bernstein. A short proof of the unpredictability of cipher block chaining. http://cr.yp.to/antiforgery/easycbc-20050109.pdf , 2005.
4. D. Chang, S. Lee, M. Nandi and M. Yung. Indifferentiable Security Analysis of Popular Hash Functions with Prefix-Free Padding. In *Asiacrypt'2006*, volume **4284** of *Lecture Notes in Computer Science*, pages 283–298, Springer-Verlag, 2006.
5. J. S. Coron, Y. Dodis, C. Malinaud and P. Puniya. Merkle-Damgard Revisited: How to Construct a Hash Function. In *Advances in Cryptology-Crypto'2005*, volume **3621** of *Lecture Notes in Computer Science*, pages 430–448, Springer-Verlag, 2005.
6. I. B. Damgard. A design principle for hash functions. In *Advances in Cryptology-Crypto'1989*, volume **435** of *Lecture Notes in Computer Science*, pages 416–427, Springer-Verlag, 1989.
7. S. Hirose, J. H. Park and A. Yun. A Simple Variant of the Merkle-Damgård Scheme with a Permutation. In *Asiacrypt'2007*, volume **4833** of *Lecture Notes in Computer Science*, pages 113–129, Springer-Verlag, 2007.
8. A. Joux. Multicollisions in iterated hash functions. Application to cascaded constructions. In *Advances in Cryptology-Crypto'2004*, volume **3152** of *Lecture Notes in Computer Science*, pages 306–316, Springer-Verlag, 2004.
9. J. Kelsey and B. Schneier. Second preimages on n-bit hash functions for much less than 2^n work. In *Advances in Cryptology-Eurocrypt'2005*, volume **3494** of *Lecture Notes in Computer Science*, pages 474–490, Springer-Verlag, 2005.
10. Stefan Lucks. A Failure-Friendly Design Principle for Hash Functions. In *Advances in Cryptology-Asiacrypt'2005*, volume **3788** of *Lecture Notes in Computer Science*, pages 474–494, Springer-Verlag, 2005.

11. U. Maurer, R. Renner and C. Holenstein. Indifferentiability, Impossibility Results on Reductions, and Applications to the Random Oracle Methodology. In *TCC'2004*, volume **2951** of *Lecture Notes in Computer Science*, pages 21–39, Springer-Verlag, 2004.

12. Ueli Maurer and Stefano Tessaro. Domain Extension of Public Random Functions: Beyond the Birthday Barrier. In *Advances in Cryptology-Crypto'2007*, volume **4622** of *Lecture Notes in Computer Science*, pages 187–204, Springer-Verlag, 2007.

13. R. C. Merkle. One way hash functions and DES. In *Advances in Cryptology-Crypto'1989*, volume **435** of *Lecture Notes in Computer Science*, pages 428–446, Springer-Verlag, 1990.

14. M. Nandi and D. R. Stinson. Multicollision Attacks on Some Generalized Sequential Hash Functions. Information Theory, IEEE Transactions on Volume **53**, Issue 2, Feb. 2007 pp 759 - 767.

15. C. Shannon. Communication theory of secrecy systems. Bell Systems Technical Journal, **28(4)**: pages 656–715, 1949.

Appendix.

Proof of Lemma 1. Let D be the set of all elements from $(\{0,1\}^b)^+$ whose MD^f values are determined from the relations $f(x_1,m_1)=z_1,\ \cdots,\ f(x_{q_2},m_{q_2}) = z_{q_2}$. Since v is irreducible, $M_i \notin D$ for all $1 \le i \le q_1$. Let P denotes the set of all nonempty prefixes of M_i's. More precisely,

$$P = \{M \in (\{0,1\}^b)^+ : M \text{ is a prefix of } M_i \text{ for some } 1 \le i \le q_1\}.$$

We enumerate the set $P \setminus (D \cup \{M_1, \cdots, M_{q_1}\}) := \{N_1, \cdots, N_{\sigma'}\}$. Note that, $|P| \le \sum_i ||M_i||$. Now, we have

$$\sigma = q_2 + \sum_i ||M_i|| \ \ge\ q_2 + |P| \ \ge\ q_2 + \sigma' + q_1 := \sigma'',$$

where σ is the total number of message blocks queried. Now we choose σ' distinct elements $z'_1, \cdots, z'_{\sigma'} \in \{0,1\}^n$ which are distinct from x_i's and IV. These values will be assigned as intermediate outputs of f during the computation of $\mathrm{MD}^f(M_i)$'s. We can choose such z'_i's in at least $\mathbf{P}(2^n - q_2 - 1, \sigma')$ ways. Now given any such choices of z'_i's we count the number of functions f such that

1. $f(x_1,m_1)=z_1,\cdots,f(x_{q_2},m_{q_2})=z_{q_2}$,
2. $\mathrm{MD}^f(M_1)=h_1,\cdots,\mathrm{MD}^f(M_{q_1})=h_{q_1}$ and
3. $\mathrm{MD}^f(N_1)=z'_1,\cdots,\mathrm{MD}^f(N_{\sigma'})=z'_{\sigma'}$.

Claim : relation 1,2,3 $\Leftrightarrow$ relation 1 and $f(a_1)=h_1,\cdots f(a_{q_1})=h_{q_1}, f(a'_1)=z'_1,\cdots,f(a'_{\sigma'})=z'_{\sigma'}$, where (x_i,m_i)'s, a_i's and a'_i's are all distinct. Moreover, the values of a_i's and a'_is are completely determined from the tuples v=$((x_1,m_1,z_1)$,

$\cdots$, $(x_{q_2}, m_{q_2}, z_{q_2})$, (M_1, h_1), $\cdots$, $(M_{q_1}, h_{q_1}))$ and $(z'_1, \cdots, z'_{\sigma'})$. More precisely, $a_i = (c_i, \text{last}(M_i))$ where

$$\begin{aligned} c_i &= z'_j \text{ if } \text{cut}(M_i) = N_j \\ &= \text{IV if } \text{cut}(M_i) = \lambda \\ &= h_j \text{ if } \text{cut}(M_i) = M_j \\ &= z_j \text{ if } \text{MD}^f(\text{cut}(M_i)) = z_j \text{ is determined from the relation 1} \end{aligned}$$

Similarly, $a'_i = (c'_i, \text{last}(N_i))$ where

$$\begin{aligned} c'_i &= z'_j \text{ if } \text{cut}(N_i) = N_j \\ &= \text{IV if } \text{cut}(N_i) = \lambda \\ &= h_j \text{ if } \text{cut}(N_i) = M_j \\ &= z_j \text{ if } \text{MD}^f(\text{cut}(N_i)) = z_j \text{ is determined from the relation 1} \end{aligned}$$

From the above discussion it is clear that the relations 1,2 and 3 equivalently correspond to the σ many distinct input-outputs of f. Thus the number of functions f satisfying 1,2 and 3 is exactly $2^{n(2^{n+b}-\sigma'')}$ where $\sigma'' = q_1 + q_2 + \sigma'$. By multiplying the number of choices of z'_is with $2^{n(2^{n+b}-\sigma'')}$, we obtain the number of functions satisfying 1 and 2 is at least

$$2^{n(2^{n+b}-\sigma'')} \times \mathbf{P}(2^n - q_2 - 1, \sigma') \geq \frac{|\mathcal{F}|}{2^{n(q_1+q_2)}} \times (1 - \frac{(\sigma' + q_2 + 1)^2}{2^{n+1}}) \geq \frac{|\mathcal{F}|}{2^{n(q_1+q_2)}} \times (1 - \frac{\sigma^2}{2^{n+1}}).$$

This follows from ineq-2 (stated in the beginning of the section). This proves the first part. The second part is trivial from the first part since $\mathbf{u}$ has uniform distribution on $\mathcal{F}$ and hence we need to divide the above quantity by $|\mathcal{F}|$. ■

Proof of Lemma 2. We denote ℓ_i as the number of pairs (x_k, m_k) such that $\text{chop}(x_k) = y^i$. More precisely, $\ell_i = |\{k : 1 \leq k \leq q_2, \text{chop}(x_k) = y^i\}|$. Since y^i's are distinct, $\ell_1 + \cdots + \ell_t \leq q_2$. Now we choose $w^i_j \in \{0,1\}^s$, $1 \leq j \leq r_i$, $1 \leq i \leq t$ such that

$$h^i_j = (w^i_j \parallel y^i)\text{'s are distinct and also distinct from } x_i\text{'s and IV.} \qquad \text{(A)}$$

The number ways we can choose w^i_j's satisfying the above condition (A) is at least

$$I_1 := (2^s - \ell_1 - 1)(2^s - \ell_1 - 2) \cdots (2^s - \ell_1 - r_1) \cdots (2^s - \ell_t - 1) \cdots (2^s - \ell_t - r_t).$$

We can choose w^1_1 in $2^s - \ell_1 - 1$ ways as there are ℓ_1 many x_k's with $\text{chop}(x_k) = y^1$ and chop(IV) can be equal to y^1. After choosing w^1_1 we can choose w^1_2 in $(2^s - \ell_1 - 2)$ ways and so on. Now, after choosing all $w^1_1, \cdots, w^1_{\ell_1}$ we can choose w^2_1 in $2^s - \ell_2 - 1$ ways since $y^2 \neq y^1$ and so on. Thus we have I_1 many w^i_j's with the condition (A). A straight forward simplification shows that $I_1 \geq 2^{sq_1}(1 - r(q_1 + q_2)/2^s)$ (we use the relations $\sum_i \ell_i \leq q_2$, $r_i \leq r$ and $\sum_i r_i = q_1$). Now for any fixed such choice of w^i_j's, the values h^i_j's are distinct from x_i's and IV. Thus, the tuple

$$v = ((x_1, m_1, z_1),\ \cdots,\ (x_{q_2}, m_{q_2}, z_{q_2}),\ (M^1_1, h^1_1),\ \cdots,\ (M^1_{r_1}, h^1_{r_1}),\ \cdots,\ (M^t_1, h^t_1),\ \cdots,\ (M^t_{r_t}, h^t_{r_t}))$$

is irreducible. Hence the number of functions $f \in \mathrm{Func}(n+b, n)$ such that

1. $f(x_1, m_1) = z_1, \cdots, f(x_{q_2}, m_{q_2}) = z_{q_2}$ and
2. $\mathrm{MD}^f(M^1_1) = h^1_1, \cdots, \mathrm{MD}^f(M^1_{r_1}) = h^1_{r_1}, \cdots, \mathrm{MD}^f(M^t_1) = h^t_1, \cdots, \mathrm{MD}^f(M^t_{r_t}) = h^t_{r_t}$

is at least $\frac{|\mathcal{F}|}{2^{n(q_1+q_2)}} \times (1 - \frac{\sigma^2}{2^{n+1}})$ (by using Lemma 1). So, the number of functions satisfying the relation in this lemma is at least

$$\frac{|\mathcal{F}|}{2^{n(q_1+q_2)}} \times (1 - \frac{\sigma^2}{2^{n+1}}) \times 2^{sq_1}(1 - \frac{r(q_1+q_2)}{2^s}) \geq |\mathcal{F}| \times \frac{1 - \Delta}{2^{nq_2+(n-s)q_1}},$$

where $\Delta = \frac{r(q_1+q_2)}{2^s} + \frac{\sigma^2}{2^{n+1}}$. The second part is followed from the first part. ■