Security of Permutation-based Compression Function $\mathsf{lp}231$

Jooyoung Lee[1]* and Daesung Kwon[2]

[1] Sejong University, Seoul, Korea, jlee05@ensec.re.kr
[2] The Attached Institute of Electronics and Telecommunications Research Institute, Daejeon, Korea, ds_kwon@ensec.re.kr

Abstract. In this paper, we study security of a certain class of permutation-based compression functions. Denoted $\mathsf{lp}231$ in [10], they are $2n$-to-n-bit compression functions using three calls to a single n-bit random permutation. We prove that $\mathsf{lp}231$ is asymptotically preimage resistant up to $2^{\frac{2n}{3}}/n$ query complexity and collision resistant up to $2^{\frac{n}{2}}/n^{1+\epsilon}$ query complexity for any $\epsilon > 0$. Based on a single permutation, $\mathsf{lp}231$ provides both efficiency and almost optimal collision security.

1 Introduction

A cryptographic hash function takes a message of arbitrary length, and returns a bit string of fixed length. The most common way of hashing variable length messages is to iterate a fixed-size compression function according to the Merkle-Damgård paradigm. The underlying compression function can either be constructed from scratch, or be built upon off-the-shelf cryptographic primitives such as blockciphers. For example, the Whirlpool hash function, adopted as ISO/IEC 10118-3 standard, is based on the Miyaguchi-Preneel construction using a modified version of AES [1]. Compression functions based on blockciphers have been widely studied [4–8, 14, 15]. Recently, researchers has begun to pay attention to building compression functions from fixed key blockciphers, where just a small number of constants are used as keys [2, 3, 9, 10, 12, 13]. Since each key of a blockcipher defines an independent random permutation in the ideal cipher model, such compression functions are often called *permutation-based*. Permutation-based compression functions have an obvious advantage over conventional blockcipher-based ones, since fixing the keys allows to save computational overload for key scheduling.

In earlier work, Black, Cochran and Shrimpton showed that any "highly-efficient" compression function using exactly one permutation call for each message block allows a query-efficient collision-finding attack [3]. Rogaway and Steinberger extended this result to a wide class of compression functions that map mn bits to rn bits using k calls to n-bit permutations [11]. Such compression functions, denoted $m \xrightarrow{k} r$, allow collision-finding attacks with $2^{n(1-(m-0.5r)/k)}$ query complexity, and preimage finding attacks with $2^{n(1-(m-r)/k)}$ query complexity.

In [10], the authors focused on the security of a special class of permutation-based compression functions, where the input to each permutation is given by a linear combination of the inputs to the compression function and the outputs of the previously called permutations. Such compression functions are called *linearly-dependent permutation-based*, and denoted by $\mathsf{LP}mkr$ if the compression function is based on independent random permutations, and by

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2013R1A1A2007488).

$\mathsf{lp}mkr$ if the compression function is based on a single random permutation. Taking into account the attacks presented in [11], they investigated the security of $\mathsf{LP}231$, $\mathsf{LP}241$, $\mathsf{LP}352$, $\mathsf{LP}362$ and their "lp variants". From a practical point of view, it is obvious that lp compression functions are more efficient compared to LP ones since an lp compression function uses its basing blockcipher with only one fixed key. However, [10] gives a concrete analysis only for $\mathsf{LP}231$. The analysis of the other compression functions rest on computer-aided approximation. Especially, the authors claim that analyzing $\mathsf{lp}231$ by hand would require about 30 times as much paper as $\mathsf{LP}231$.

In this paper, we give a concrete security analysis of $\mathsf{lp}231$ in terms of preimage resistance and collision resistance. Specifically, we prove preimage resistance up to $(2^{\frac{2n}{3}}/n)$ query complexity and collision resistance up to $(2^{\frac{n}{2}}/n^{1+\epsilon})$ query complexity for any $\epsilon > 0$. Our analysis is not only simpler than the authors of [10] estimated, but also elegant based on a recursive approach.

2 Preliminaries

General Notations For a positive integer n, let $[1,n] = \{1,2,\ldots,n\}$ and let $\mathit{\Pi}_n$ be the set of permutations on $\{0,1\}^n$. We let $\mathbb{F}_{2^n}$ denote a finite field of order 2^n. Throughout our work, we will identify $\mathbb{F}_{2^n}$ and $\{0,1\}^n$, assuming a fixed mapping between the two sets.

For positive integers s and t, let $\mathcal{M}_{\mathbb{F}_{2^n}}^{s\times t}$ denote the set of all $s \times t$ matrices over $\mathbb{F}_{2^n}$. For A, $B \in \mathcal{M}_{\mathbb{F}_{2^n}}^{2\times 1}$, $[A,B]$ is the 2×2 matrix obtained by the concatenation of A and B. The concatenation is similarly denoted for more than two matrices. For a 2×1 matrix

$$A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

let

$$A^* = \begin{bmatrix} a_2 \; a_1 \end{bmatrix}.$$

Note that $A^*A = 0$. For A, $B \in \mathcal{M}_{\mathbb{F}_{2^n}}^{2\times 1}$, A^*B is the determinant of $[A,B]$.

We write $u \overset{\$}{\leftarrow} U$ to denote uniform random sampling from the set U and assignment to u. For a multiset U, $\mathsf{mult}(U,u)$ is the multiplicity of u in U, and $\mathsf{mult}(U) = \max_{u\in U} \mathsf{mult}(U,u)$.

Linearly-dependent Permutation-based Compression Functions For positive integers m, k and r with $m > r$, let $\mathcal{M}_{\mathbb{F}_{2^n}}(m,k,r)$ be a set of $(k+r)\times(m+k)$ matrices $A = (a_{ij})$ over $\mathbb{F}_{2^n}$ such that

$$a_{ij} = 0 \text{ for } 1 \le i \le k \text{ and } j \ge m+i.$$

Then each matrix $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(m,k,r)$ defines a compression function lp_{mkr}^A with oracle access to a random permutation $\pi \in \mathit{\Pi}_n$ as follows.

$$\begin{aligned} \mathsf{lp}_{mkr}^A : (\{0,1\}^n)^m &\longrightarrow (\{0,1\}^n)^r \\ (v_1,\ldots,v_m) &\longmapsto (w_1,\ldots,w_r) \end{aligned}$$

where $(w_1,\ldots,w_r)$ is computed by the algorithm described in Figure 1(a). A function lp_{mkr}^A is called *linearly-dependent single-permutation-based*, and often simply denoted as $\mathsf{lp}mkr$ or lp^A. A compression function lp_{231}^A for $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$ is separately described in Figure 1(b).

```
Algorithm $\mathsf{lp}^A_{mkr}(v_1, \ldots, v_m)$
  for $i \leftarrow 1$ to $k$ do
      $x_i \leftarrow \sum_{j=1}^{m} a_{ij} v_j + \sum_{j=1}^{i-1} a_{i(m+j)} y_j$
      $y_i \leftarrow \pi(x_i)$
  for $i \leftarrow 1$ to $r$ do
      $w_i \leftarrow \sum_{j=1}^{m} a_{(k+i)j} v_j + \sum_{j=1}^{k} a_{(k+i)(m+j)} y_j$
  return $(w_1, \ldots, w_r)$
```

(a) lp^A_{mkr} for $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(m, k, r)$

```
Algorithm $\mathsf{lp}^A_{231}(v_1, v_2)$
  $x_1 \leftarrow a_{11} v_1 + a_{12} v_2$
  $y_1 \leftarrow \pi(x_1)$
  $x_2 \leftarrow a_{21} v_1 + a_{22} v_2 + a_{23} y_1$
  $y_2 \leftarrow \pi(x_2)$
  $x_3 \leftarrow a_{31} v_1 + a_{32} v_2 + a_{33} y_1 + a_{34} y_2$
  $y_3 \leftarrow \pi(x_3)$
  $w \leftarrow a_{41} v_1 + a_{42} v_2 + a_{43} y_1 + a_{44} y_2 + a_{45} y_3$
  return $w$
```

(b) lp^A_{231} for $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(2, 3, 1)$

Fig. 1. Compression function lp^A_{mkr}

Collision Resistance and Preimage Resistance For simplicity of notations, we will define collision resistance and preimage resistance focusing on linearly-dependent single-permutation-based compression functions, while these security notions can be extended in an obvious way to any hash function based on public ideal primitives.

Let lp^A_{mkr} be a compression function for $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(m, k, r)$. Given an information-theoretic adversary $\mathcal{A}$ with oracle access to π and π^{-1}, we execute the experiment $\mathbf{Exp}^{\mathsf{col}}_{\mathcal{A}}$ described in Figure 2(a) in order to quantify the collision resistance of lp^A_{mkr}. The experiment records the query-response pairs that the adversary $\mathcal{A}$ obtains into a *query history* $\mathcal{Q}$. A pair (x, y) is in the query history if $\mathcal{A}$ makes $\pi(x)$ and gets back y, or it makes $\pi^{-1}(y)$ and gets back x. Given a query history $\mathcal{Q}$, then $\mathsf{Map}_{\mathsf{lp}^A}(\mathcal{Q}) \subset (\{0,1\}^n)^m \times (\{0,1\}^n)^r$ is defined to be the set of pairs (v, w) such that there exist evaluations $(x_i, y_i) \in \mathcal{Q}$ satisfying the following equations.

$$x_i = \sum_{j=1}^{m} a_{ij} v_j + \sum_{j=1}^{i-1} a_{i(m+j)} y_j, \qquad i = 1, \ldots, k,$$

$$w_i = \sum_{j=1}^{m} a_{(k+i)j} v_j + \sum_{j=1}^{k} a_{(k+i)(m+j)} y_j, \qquad i = 1, \ldots, r, \tag{1}$$

where we write $v = (v_1, \ldots, v_m)$ and $w = (w_1, \ldots, w_r)$. Informally, $\mathsf{Map}_{\mathsf{lp}^A}(\mathcal{Q})$ is the set of the evaluations of lp^A_{mkr} that are determined by the query history $\mathcal{Q}$. Now the *collision-finding advantage* of $\mathcal{A}$ is defined to be

$$\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(\mathcal{A}) = \mathbf{Pr}\left[\mathbf{Exp}^{\mathsf{col}}_{\mathcal{A}} = 1\right].$$

The probability is taken over the random permutation π, and $\mathcal{A}$'s coins (if any). For $q > 0$, we define $\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(q)$ as the maximum of $\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(\mathcal{A})$ over all adversaries $\mathcal{A}$ making at most q queries.

The preimage resistance of lp^A_{mkr} is quantified similarly using the experiment $\mathbf{Exp}^{\mathsf{pre}}_{\mathcal{A}}$ described in Figure 2(b). The adversary $\mathcal{A}$ takes as input a random $w \in (\{0,1\}^n)^r$ before it begins making queries to π and π^{-1}. The *preimage-finding advantage* of $\mathcal{A}$ is defined to be

$$\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(\mathcal{A}) = \mathbf{Pr}\left[\mathbf{Exp}^{\mathsf{pre}}_{\mathcal{A}} = 1\right].$$

For $q > 0$, $\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(q)$ is the maximum of $\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(\mathcal{A})$ over all adversaries $\mathcal{A}$ making at most q queries.

```
Experiment $\mathbf{Exp}^{\mathsf{col}}_{\mathcal{A}}$
  $\pi \xleftarrow{\$} \mathit{\Pi}_n$
  $\mathcal{A}^{\pi,-\pi}$ updates $\mathcal{Q}$
  if $\exists\ v \neq v'$, $w$ s.t. $(v, w)$, $(v', w) \in \mathsf{Map}_{\mathsf{lp}^A}(\mathcal{Q})$ then
    output 1
  else
    output 0
```

(a) Quantification of collision resistance

```
Experiment $\mathbf{Exp}^{\mathsf{pre}}_{\mathcal{A}}$
  $\pi \xleftarrow{\$} \mathit{\Pi}_n$
  $w \xleftarrow{\$} I^r_n$
  $\mathcal{A}(w)^{\pi,-\pi}$ updates $\mathcal{Q}$
  if $\exists\ v$ s.t. $(v, w) \in \mathsf{Map}_{\mathsf{lp}^A}(\mathcal{Q})$ then
    output 1
  else
    output 0
```

(b) Quantification of preimage resistance

Fig. 2. Experiments for quantification of collision resistance and preimage resistance

3 Security of lp231

In this section, we will prove that a linearly-determined permutation-based compression function lp^A_{231} achieves good collision resistance and preimage resistance as long as the matrix $A \in \mathcal{M}_{\mathbb{F}_{2^n}}(m, k, r)$ satisfies a certain condition. Our strategy is to first define a certain "bad event" parameterized by a matrix, and then show that collision or preimage finding is hard without the occurrence of the bad event. The bad event is the union of some "auxiliary events", defined in the following subsection. Throughout this paper, we will write $N = 2^n$ and $N' = N - q$.

3.1 Auxiliary Events

In order to analyze the security of lp231, we need to define some auxiliary events. Suppose that an adversary $\mathcal{A}$ makes q adaptive queries to a random permutation π and its inverse π^{-1}, and records a query history

$$\mathcal{Q} = \{(x^j, y^j) \in \{0,1\}^n \times \{0,1\}^n : 1 \leq j \leq q\}$$

where (x^j, y^j) denotes the query-response pair obtained by the j-th query. Throughout this work, we will assume that x^j's are all distinct. This means that y^j's are also all distinct since π is a permutation. It would not affect $\mathcal{A}$'s collision finding advantage making any redundant query.

For $t \geq 1$ and, a_i, $b_i \in \mathbb{F}_{2^n}$ and A_i, $B_i \in \mathcal{M}^{2\times 1}_{\mathbb{F}_{2^n}}$, $i = 1, \ldots, t$, we define the following multisets.

$$U^t(a_1, b_1, \ldots, a_t, b_t) = \left\{ \sum_{i=1}^{t} \left(a_i x^{j_i} + b_i y^{j_i} \right) : j_1, \ldots, j_t \in [1, q] \right\},$$

$$U^t_{\neq}(a_1, b_1, \ldots, a_t, b_t) = \left\{ \sum_{i=1}^{t} \left(a_i x^{j_i} + b_i y^{j_i} \right) : j_1, \ldots, j_t \in [1, q] \text{ are all distinct} \right\},$$

$$V^t(A_1, B_1, \ldots, A_t, B_t) = \left\{ \sum_{i=1}^{t} \left(A_i x^{j_i} + B_i y^{j_i} \right) : j_1, \ldots, j_t \in [1, q] \right\}.$$

For a positive integer l, these multisets are associated with the following events.

$$\begin{aligned}\mathsf{E}^t(a_1,b_1,\ldots,a_t,b_t;l) &\Leftrightarrow \mathcal{A} \text{ sets } \mathsf{mult}(U^t(a_1,b_1,\ldots,a_t,b_t)) > l,\\ \mathsf{E}^t_{\neq}(a_1,b_1,\ldots,a_t,b_t;l) &\Leftrightarrow \mathcal{A} \text{ sets } \mathsf{mult}(U^t(a_1,b_1,\ldots,a_t,b_t)) > l,\\ \mathsf{F}^t(A_1,B_1,\ldots,A_t,B_t;l) &\Leftrightarrow \mathcal{A} \text{ sets } \mathsf{mult}(V^t(A_1,B_1,\ldots,A_t,B_t)) > l.\end{aligned}$$

We will often write $\mathsf{E}^t(l) = \mathsf{E}^t(a_1,b_1,\ldots,a_t,b_t;l)$, $\mathsf{E}^t_{\neq}(l) = \mathsf{E}^t_{\neq}(a_1,b_1,\ldots,a_t,b_t;l)$ and $\mathsf{F}^t(l) = \mathsf{F}^t(A_1,B_1,\ldots,A_t,B_t;l)$ for simplicity. The rest of this section is devoted to the estimation of the probability of these auxiliary events for small t's (say $t = 1,2,3$).

Theorem 1. *Let a_1, a_2, b_1, b_2 be nonzero elements in $\mathbb{F}_{2^n}$, and let*

$$\begin{aligned} f_1 = f_1(d_1) &= N\binom{q}{d_1+1}\left(\frac{1}{N'}\right)^{d_1+1},\\ f_2 = f_2(d_2) &= N\binom{q}{d_2+1}\left(\frac{2q}{N'}\right)^{d_2+1}\end{aligned}$$

for positive integers d_1 and d_2. Then the following hold.

1. $\mathbf{Pr}\left[\mathsf{E}^1(a_1,b_1;d_1)\right] \le f_1$.

2. $\mathbf{Pr}\left[\mathsf{E}^2_{\neq}(a_1,b_1,a_2,b_2;2d_1d_2)\right] \le f_2 + 2f_1$.

3. *If either $a_1 + a_2 \neq 0$ or $b_1 + b_2 \neq 0$, then*

$$\mathbf{Pr}\left[\mathsf{E}^2(a_1,b_1,a_2,b_2;2d_1d_2+d_1)\right] \le f_2 + 3f_1.$$

Proof. First we give a proof for the first inequality. Fix $c \in \mathbb{F}_{2^n}$. When $\mathcal{A}$ makes the j-th forward query $y = \pi(x^*)$, the probability that $a_1x^* + b_1y = c$, which is equivalent to $y = b_1^{-1}(c + a_1x^*)$, is not greater than $1/(2^n - (j-1))$. Similarly, when $\mathcal{A}$ makes the j-th backward query $x = \pi^{-1}(y^*)$, the probability that $a_1x + b_1y^* = c$ is not greater than $1/(2^n - (j-1))$. The event $\mathsf{E}^1(a_1,b_1;d_1)$ occurs when there exists a set $\{j_1,\ldots,j_{d_1+1}\} \subset [1,q]$ such that

$$a_1x^{j_1} + b_1y^{j_1} = \ldots = a_1x^{j_{d_1+1}} + b_1y^{j_{d_1+1}} = c$$

for some $c \in \mathbb{F}_{2^n}$. Since $1/(2^n - (j-1)) \le 1/N'$, it follows that

$$\mathbf{Pr}\left[\mathsf{E}^1(a_1,b_1;d_1)\right] \le N\binom{q}{d_1+1}\left(\frac{1}{N'}\right)^{d_1+1} = f_1.$$

For the proof of the second inequality, we define events

$$\begin{aligned}\mathsf{E}^2_j(c) &\Leftrightarrow \mathcal{A} \text{ sets } a_1x^{j_1} + b_1y^{j_1} + a_2x^{j_2} + b_2y^{j_2} = c \text{ with } j = \max\{j_1,j_2\},\\ \mathsf{E}^2(c;2d_1d_2) &\Leftrightarrow \mathcal{A} \text{ sets } \mathsf{mult}(U^2,c) > 2d_1d_2\end{aligned}$$

where we simply write $U^2 = U^2(a_1,b_1,a_2,b_2)$. Note that $\mathsf{E}^2_j(c)$ occurs when the j-th query increases the multiplicity of c in U^2 at least by one. In order to estimate $\mathbf{Pr}\left[\mathsf{E}^2(2d_1d_2)\right]$, we decompose $\mathsf{E}^2(2d_1d_2)$ as follows.

$$\mathsf{E}^2(2d_1d_2) = \bigvee_{c\in\mathbb{F}_{2^n}} \mathsf{E}^2(c;2d_1d_2) \Rightarrow \mathsf{E}_{ex} \vee \bigvee_{c\in\mathbb{F}_{2^n}} \left(\mathsf{E}^2(c;2d_1d_2) \wedge \neg\mathsf{E}_{ex}\right) \tag{2}$$

where

$$\mathsf{E}_{ex} = \mathsf{E}^1(a_1, b_1; d_1) \vee \mathsf{E}^1(a_2, b_2; d_1).$$

By the first inequality, it follows that

$$\mathbf{Pr}\left[\mathsf{E}_{ex}\right] \leq 2f_1. \tag{3}$$

We now analyze the event $\mathsf{E}^2(c; 2d_1d_2) \wedge \neg\mathsf{E}_{ex}$ for a fixed $c \in \mathbb{F}_{2^n}$. Suppose that a certain query, say the j-th query, completes equation

$$a_1x^j + b_1y^j + a_2x^{j'} + b_2y^{j'} = c$$

for some $j' < j$. For the fixed j-th query, the j'-th query should satisfy

$$a_2x^{j'} + b_2y^{j'} = c + a_1x^j + b_1y^j$$

and the number of such queries is at most d_1 without the occurrence of $\mathsf{E}^1(a_2, b_2; d_1)$. Taking into account the other position where the last query might contributes, we see that each query increases the multiplicity $\mathsf{mult}(U^2, c)$ at most by $2d_1$ without the occurrence of E_{ex}. This implies that the number of queries that increase $\mathsf{mult}(U^2, c)$ should be at least $d_2 + 1$. Therefore we have

$$\mathsf{E}^2(c; 2d_1d_2) \wedge \neg\mathsf{E}_{ex} \Rightarrow \bigvee_{\substack{J \subset [1,q] \\ |J| = d_2+1}} \left(\bigwedge_{j \in J} \left(\mathsf{E}^2_j(c) \wedge \neg\mathsf{E}_{ex}\right) \right). \tag{4}$$

In order to compute $\mathbf{Pr}\left[\bigwedge_{j\in J}\left(\mathsf{E}^2_j(c) \wedge \neg\mathsf{E}_{ex}\right)\right]$ for a fixed $J \subset [1, q]$ such that $|J| = d_2 + 1$, suppose that $\mathcal{A}$ makes the j^*-th query $\pi(x^*)$ for $j^* \in J$. Then we can upper bound the number of responses $y = \pi(x^*)$ that contribute the equation

$$a_1x^{j_1} + b_1y^{j_1} + a_2x^{j_2} + b_2y^{j_2} = c,$$

with $j^* = \max\{j_1, j_2\}$. If $j_1 = j^* > j_2$, then it should hold that

$$y = b_2^{-1}(c + a_1x^* + a_2x^{j_2} + b_2y^{j_2})$$

for some $j_2 < j^* \leq q$. So the number of possible responses for this case is at most q. The case $j_2 = j^* > j_1$ is analyzed similarly. Therefore the total number of possible responses $y = \pi(x^*)$ is at most $2q$. With an analogous argument for π^{-1}, we conclude that

$$\mathbf{Pr}\left[\bigwedge_{j\in J}\left(\mathsf{E}^2_j(c) \wedge \neg\mathsf{E}_{ex}\right)\right] \leq \left(\frac{2q}{N'}\right)^{d_2+1}. \tag{5}$$

Now by (2), (3), (4), and (5), we have

$$\mathbf{Pr}\left[\mathsf{E}^2\left(a_1, b_1, a_2, b_2; 2d_1d_2\right)\right] \leq N\binom{q}{d_2+1}\left(\frac{2q}{N'}\right)^{d_2+1} + 2f_1 = f_2 + 2f_1.$$

In order to prove the third inequality, note that

$$U^2\left(a_1, b_1, a_2, b_2\right) = U^2_{\neq}\left(a_1, b_1, a_2, b_2\right) \cup U^1\left(a_1 + a_2, b_1 + b_2\right).$$

If $a_1 + a_2 \neq 0$ and $b_1 + b_2 \neq 0$, then by the first inequality,

$$\mathbf{Pr}\left[\mathsf{E}^1\left(a_1 + a_2, b_1 + b_2; d_1\right)\right] \leq f_1.$$

This inequality also holds for the case that either $a_1 + a_2 \neq 0$ or $b_1 + b_2 \neq 0$: since π is a permutation, this special case implies

$$\mathbf{Pr}\left[\mathsf{E}^1\left(a_1 + a_2, b_1 + b_2; 1\right)\right] = 0.$$

Therefore by the first two inequalities,

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{E}^2(a_1, b_1, a_2, b_2; 2d_1d_2 + d_1)\right] &\leq \mathbf{Pr}\left[\mathsf{E}^2_{\neq}(a_1, b_1, a_2, b_2; 2d_1d_2)\right] + \mathbf{Pr}\left[\mathsf{E}^1\left(a_1 + a_2, b_1 + b_2; d_1\right)\right] \\ &\leq f_2 + 3f_1.\end{aligned}$$

□

Definition 1. *For $t \geq 1$, a matrix $M = [A_1, B_1, \ldots, A_t, B_t] \in \mathcal{M}^{2\times 2t}_{\mathbb{F}_{2^n}}$ is called* column-sum independent *if M satisfies the following conditions.*

1. $\left[\sum_{i\in I_1} A_i, \sum_{i\in I_2} A_i\right]$ *and* $\left[\sum_{i\in I_1} B_i, \sum_{i\in I_2} B_i\right]$ *are invertible for any pair of distinct nonempty subsets $I_1, I_2 \subset [1, t]$.*
2. $\left[\sum_{i\in I_1} A_i, \sum_{i\in I_2} B_i\right]$ *are invertible for any pair of (not necessarily distinct) nonempty subsets $I_1, I_2 \subset [1, t]$.*

Definition 1 will be used for compact statement of the following corollary. We point out some useful properties of column-sum independent matrices.

Property 1 *If $[A_1, B_1, \ldots, A_t, B_t]$ is column-sum independent, then $[A_{i_1}, B_{i_1}, \ldots, A_{i_s}, B_{i_s}]$ is also column-sum independent for every nonempty subset $\{i_1, \ldots, i_s\} \subset [1, t]$.*

Property 2 *If $[A_1, B_1, \ldots, A_t, B_t]$ is column-sum independent, then $\sum_{i\in I} A_i \neq 0$ and $\sum_{i\in I} B_i \neq 0$ for every nonempty subset $I \subset [1, t]$.*

Property 3 *Column-sum independence of M stipulates nonsingularity of*

$$2\left(2^t - 1\right)\left(2^t - 2\right) + \left(2^t - 1\right)^2$$

matrices determined by M.

Theorem 2. *Let $f_1 = f_1(d_1)$ and $f_2 = f_2(d_2)$ be defined as in Theorem 1, and let*

$$\begin{aligned} g_2 = g_2(d_1, d_2) &= \frac{d_2\left(4d_1q + 2d_1 + 1\right)q}{N'}, \\ g_3 = g_3(d_1, d_2, d_3) &= N^2\binom{q}{d_3 + 1}\left(\frac{6d_1d_2 + 6d_1 + 1}{N'}\right)^{d_3+1}\end{aligned}$$

for positive integers d_1, d_2 and d_3. If a matrix $[A_1, B_1, A_2, B_2, A_3, B_3] \in \mathcal{M}^{2\times 6}_{\mathbb{F}_{2^n}}$ is column-sum independent, then the following hold.

1. $\mathbf{Pr}\left[\mathsf{F}^1(A_1, B_1; 1)\right] = 0.$

2. $\mathbf{Pr}\left[\mathsf{F}^2(A_1, B_1, A_2, B_2; 1)\right] \leq g_2 + 6f_2 + 14f_1.$

3. $\mathbf{Pr}\left[\mathsf{F}^3(A_1, B_1, A_2, B_2, A_3, B_3; 7d_3)\right] \leq g_3 + 3g_2 + 24f_2 + 66f_1$.

Proof. The proof of the first equality is straightforward since $[A_1, B_1]$ is invertible. In order to prove the second inequality, we define the following events.

$$\mathsf{F}^2_{\mathsf{col}}(j) \Leftrightarrow \mathcal{A} \text{ sets } A_1x^{j_1} + B_1y^{j_1} + A_2x^{j_2} + B_2y^{j_2} = A_1x^{j_3} + B_1y^{j_3} + A_2x^{j_4} + B_2y^{j_4}$$
$$\text{where } j_3 < j_1 \leq j,\ j_4 < j_2 \leq j, \text{ and } j = \max\{j_1, j_2\}$$

and

$$\begin{aligned}\mathsf{E}_{ex} = \mathsf{E}^2&\left((B_1+B_2)^*A_1, (B_1+B_2)^*B_1, (B_1+B_2)^*A_2, (B_1+B_2)^*B_2; 2d_1d_2 + d_1\right)\\ \vee\, \mathsf{E}^2&\left((A_1+A_2)^*A_1, (A_1+A_2)^*B_1, (A_1+A_2)^*A_2, (A_1+A_2)^*B_2; 2d_1d_2 + d_1\right)\\ \vee\, \mathsf{E}^2_{\neq}&\left(B_1^*A_2, B_1^*B_2, B_1^*A_2, B_1^*B_2; 2d_1d_2\right)\\ \vee\, \mathsf{E}^2_{\neq}&\left(A_1^*A_2, A_1^*B_2, A_1^*A_2, A_1^*B_2; 2d_1d_2\right)\\ \vee\, \mathsf{E}^2_{\neq}&\left(B_2^*A_1, B_2^*B_1, B_2^*A_1, B_2^*B_1; 2d_1d_2\right)\\ \vee\, \mathsf{E}^2_{\neq}&\left(A_2^*A_1, A_2^*B_1, A_2^*A_1, A_2^*B_1; 2d_1d_2\right).\end{aligned}$$

Then it follows that

$$\mathsf{F}^2(A_1, B_1, A_2, B_2; 1) \Rightarrow \bigvee_{1 \leq j \leq q} \mathsf{F}^2_{\mathsf{col}}(j) \Rightarrow \mathsf{E}_{ex} \vee \bigvee_{1 \leq j \leq q} \left(\mathsf{F}^2_{\mathsf{col}}(j) \wedge \neg\mathsf{E}_{ex}\right) \tag{6}$$

and

$$\mathbf{Pr}\left[\mathsf{E}_{ex}\right] \leq 2\left(f_2 + 3f_1\right) + 4\left(f_2 + 2f_1\right) = 6f_2 + 14f_1 \tag{7}$$

by Theorem 1.

We now estimate the probability $\mathbf{Pr}\left[\mathsf{F}^2_{\mathsf{col}}(j) \wedge \neg\mathsf{E}_{ex}\right]$. Suppose that $\mathcal{A}$ makes the j-th query $\pi(x^*)$, and consider the following three cases where $y = \pi(x^*)$ contributes the equation

$$A_1x^{j_1} + B_1x^{j_1} + A_2x^{j_2} + B_2x^{j_2} = A_1x^{j_3} + B_1x^{j_3} + A_2x^{j_4} + B_2x^{j_4}. \tag{8}$$

Case 1: $j_1 = j_2 = j$. The equality (8) is reduced to

$$(A_1 + A_2)x^* + (B_1 + B_2)y = A_1x^{j_3} + B_1y^{j_3} + A_2x^{j_4} + B_2y^{j_4}. \tag{9}$$

Any response y satisfying (9) corresponds to a pair $(j_3, j_4) \in [1, j-1]^2$ such that

$$\begin{aligned}(B_1+B_2)^*A_1x^{j_3} + (B_1+B_2)^*B_1y^{j_3} + (B_1+B_2)^*A_2x^{j_4} + (B_1+B_2)^*B_2y^{j_4}&\\ = (B_1+B_2)^*(A_1+A_2)x^*.&\end{aligned} \tag{10}$$

The number of such pairs is at most $2d_1d_2 + d_1$ without the occurrence of E_{ex}.

Case 2: $j_1 = j$ and $j_2 \neq j$. The equality (8) is reduced to

$$A_1x^* + B_1y = A_2x^{j_2} + B_2y^{j_2} + A_1x^{j_3} + B_1y^{j_3} + A_2x^{j_4} + B_2y^{j_4}. \tag{11}$$

Any response y satisfying (11) corresponds to a triple $(j_2, j_3, j_4) \in [1, j-1]^3$ such that $j_2 \neq j_4$ and

$$B_1^*A_1x^{j_3} + (B_1^*A_2x^{j_2} + B_1^*B_2y^{j_2} + B_1^*A_2x^{j_4} + B_1^*B_2y^{j_4}) = B_1^*A_1x^*. \tag{12}$$

For each $j_3 \in [1, j-1]$, the number of pairs $(j_2, j_4) \in [1, j-1]^2$ satisfying (12) and $j_2 \neq j_4$ is at most $2d_1d_2$ without the occurrence of E_{ex}. Therefore the number of the triples satisfying (11) is at most $2d_1d_2q$ without the occurrence of E_{ex}.

Case 3: $j_2 = j$ and $j_1 \neq j$. The analysis of this case is essentially the same as Case 2.
To summarize, we conclude that

$$\mathbf{Pr}\left[\mathsf{F}^2_{\mathsf{col}}(j) \wedge \neg \mathsf{E}_{ex}\right] \leq \frac{d_2\left(4d_1q + 2d_1 + 1\right)}{N'}. \tag{13}$$

We can apply a similar argument for π^{-1}. So using a union bound, the second inequality is followed from (6), (7) and (13).

In order to prove the third inequality, we define events

$$\mathsf{F}^3_j(C) \Leftrightarrow \mathcal{A} \text{ sets } \sum_{i=1}^{3}\left(A_i x^{j_i} + B_i y^{j_i}\right) = C \text{ with } j = \max\{j_1, j_2, j_3\}, \tag{14}$$
$$\mathsf{F}^3(C; 7d_3) \Leftrightarrow \mathcal{A} \text{ sets } \mathsf{mult}(V^3, C) > 7d_3$$

for $C \in \mathcal{M}^{2\times 1}_{\mathbb{F}_{2^n}}$ and $j \in [1, q]$. Here we simply write $V^3 = V^3(A_1, B_1, A_2, B_2, A_3, B_3)$. The event $\mathsf{F}^3_j(C)$ occurs when the j-th query increases the multiplicity of C in V^3 at least by one. In order to estimate $\mathbf{Pr}\left[\mathsf{F}^3(7d_3)\right]$, we decompose $\mathsf{F}^3(7d_3)$ as follows.

$$\mathsf{F}^3(7d_3) = \bigvee_{C \in \mathcal{M}^{2\times 1}_{\mathbb{F}_{2^n}}} \mathsf{F}^3(C; 7d_3) \Rightarrow \mathsf{F}_{ex} \vee \bigvee_{C \in \mathcal{M}^{2\times 1}_{\mathbb{F}_{2^n}}} \left(\mathsf{F}^3(C; 7d_3) \wedge \neg \mathsf{F}_{ex}\right) \tag{15}$$

where

$$\begin{aligned}
\mathsf{F}_{ex} = {} & \mathsf{F}^1(A_1, B_1; 1) \vee \mathsf{F}^1(A_2, B_2; 1) \vee \mathsf{F}^1(A_3, B_3; 1) \\
& \vee \mathsf{F}^2(A_1, B_1, A_2, B_2; 1) \vee \mathsf{F}^2(A_2, B_2, A_3, B_3; 1) \vee \mathsf{F}^2(A_3, B_3, A_1, B_1; 1) \\
& \vee \mathsf{E}^1\left((A_2 + A_3)^* A_1, (A_2 + A_3)^* B_1; d_1\right) \vee \mathsf{E}^1\left((A_3 + A_1)^* A_2, (A_3 + A_1)^* B_2; d_1\right) \\
& \vee \mathsf{E}^1\left((A_1 + A_2)^* A_3, (A_1 + A_2)^* B_3; d_1\right) \vee \mathsf{E}^1\left((B_2 + B_3)^* A_1, (B_2 + B_3)^* B_1; d_1\right) \\
& \vee \mathsf{E}^1\left((B_3 + B_1)^* A_2, (B_3 + B_1)^* B_2; d_1\right) \vee \mathsf{E}^1\left((B_1 + B_2)^* A_3, (B_1 + B_2)^* B_3; d_1\right) \\
& \vee \mathsf{E}^2\left(A_3^* A_1, A_3^* B_1, A_3^* A_2, A_3^* B_2; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(A_1^* A_2, A_1^* B_2, A_1^* A_3, A_1^* B_3; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(A_2^* A_3, A_2^* B_3, A_2^* A_1, A_2^* B_1; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_3^* A_1, B_3^* B_1, B_3^* A_2, B_3^* B_2; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_1^* A_2, B_1^* B_2, B_1^* A_3, B_1^* B_3; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_2^* A_3, B_2^* B_3, B_2^* A_1, B_2^* B_1; 2d_1d_2 + d_1\right).
\end{aligned} \tag{16}$$

By Theorem 1 and the first two inequalities, it follows that

$$\mathbf{Pr}\left[\mathsf{F}_{ex}\right] \leq 3(g_2 + 6f_2 + 14f_1) + 6f_1 + 6(f_2 + 3f_1) = 3g_2 + 24f_2 + 66f_1. \tag{17}$$

We now analyze the event $\mathsf{F}^3(C; 7d_3) \wedge \neg \mathsf{F}_{ex}$ for a fixed $C \in \mathcal{M}^{2\times 1}_{\mathbb{F}_{2^n}}$. Since each query increases the multiplicity $\mathsf{mult}(V^3, C)$ at most by 7 without the occurrence of F_{ex} (according to the positions where the query contributes), the number of queries that increase $\mathsf{mult}(V^3, C)$ should be at least $d_3 + 1$. Therefore we obtain

$$\mathsf{F}^3(C; 7d_3) \wedge \neg \mathsf{F}_{ex} \Rightarrow \bigvee_{\substack{J \subset [1,q] \\ |J| = d_3 + 1}} \left(\bigwedge_{j \in J} \left(\mathsf{F}^3_j(C) \wedge \neg \mathsf{F}_{ex}\right)\right). \tag{18}$$

In order to compute $\mathbf{Pr}\left[\bigwedge_{j\in J}\left(\mathsf{F}_j^3(C)\wedge\neg\mathsf{F}_{ex}\right)\right]$ for a fixed $J\subset[1,q]$ such that $|J|=d_3+1$, suppose that $\mathcal{A}$ makes the j^*-th query $\pi(x^*)$ for $j^*\in J$. Then we can upper bound the number of responses $y=\pi(x^*)$ that contribute the equation

$$A_1x^{j_1}+B_1y^{j_1}+A_2x^{j_2}+B_2y^{j_2}+A_3x^{j_3}+B_3y^{j_3}=C \tag{19}$$

with $j^*=\max\{j_1,j_2,j_3\}$. For $s\in\{0,1,2\}$, consider the case where the j^*-th query contributes $3-s$ terms in equation (19). If $j_{s+1}=\cdots=j_3=j^*$, then the equation (19) is reduced to

$$\sum_{i=1}^{s}\left(A_ix^{j_i}+B_iy^{j_i}\right)+\bar{A}x^*+\bar{B}y=C \tag{20}$$

where $\bar{A}=\sum_{i=s+1}^{3}A_i\neq 0$ and $\bar{B}=\sum_{i=s+1}^{3}B_i\neq 0$ by the column-sum independence. By multiplying $\bar{B}^*$ on both sides of (20), we observe that each y satisfying (20) is associated with a solution $(j_1,\ldots,j_s)\in[1,j^*-1]^s$ to the following equation.

$$\sum_{i=1}^{s}\left(\bar{B}^*A_ix^{j_i}+\bar{B}^*B_iy^{j_i}\right)=\bar{B}^*C+\bar{B}^*\bar{A}x^*. \tag{21}$$

Without the occurrence of F_{ex}, the number of solutions $(j_1,\ldots,j_s)$ to (21) is at most 1 if $s=0$, d_1 if $s=1$ and $2d_1d_2+d_1$ if $s=2$. By symmetry in the positions of the j^*-th query and with an analogous argument for π^{-1}, we obtain

$$\mathbf{Pr}\left[\bigwedge_{j\in J}\left(\mathsf{F}_j^3(C)\wedge\neg\mathsf{F}_{ex}\right)\right]\leq\left(\frac{6d_1d_2+6d_1+1}{N'}\right)^{d_3+1}. \tag{22}$$

Using a union bound, the proof is complete from (15), (17), (18) and (22). □

3.2 Concrete Security Bounds for lp231

For $A\in\mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$, the system of equations (1) is rewritten as follows.

$$x_1=a_{11}v_1+a_{12}v_2 \tag{23}$$

$$x_2=a_{21}v_1+a_{22}v_2+a_{23}y_1 \tag{24}$$

$$x_3=a_{31}v_1+a_{32}v_2+a_{33}y_1+a_{34}y_2 \tag{25}$$

$$w=a_{41}v_1+a_{42}v_2+a_{43}y_1+a_{44}y_2+a_{45}y_3. \tag{26}$$

From equations (23) and (24), we obtain the following system of equations in variables v_1 and v_2.

$$\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}=\begin{bmatrix}x_1\\ x_2+a_{23}y_1\end{bmatrix}. \tag{27}$$

If $a_{11}a_{22}+a_{12}a_{21}\neq 0$, then we can solve the system of equations (27). By substituting its solution into equations (25) and (26), we obtain an equation of the following form.

$$A_1x_1+B_1y_1+A_2x_2+B_2y_2+A_3x_3+B_3y_3=Cw, \tag{28}$$

where A_i, B_i and C are matrices in $\mathcal{M}_{\mathbb{F}_{2^n}}^{2\times 1}$. We write $M(A)=[A_1,B_1,A_2,B_2,A_3,B_3]$ and $C(A)=C$, indicating these matrices are determined by the matrix A. Note that

$$A_3=\begin{bmatrix}1\\ 0\end{bmatrix}, B_3=\begin{bmatrix}0\\ a_{45}\end{bmatrix} \text{ and } C=\begin{bmatrix}0\\ 1\end{bmatrix}.$$

Preimage Resistance

Theorem 3. *Let $A = (a_{ij})$ be a matrix in $\mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$ such that $a_{11}a_{22} + a_{12}a_{21} \neq 0$. If $M(A)$ is column-sum independent, then for positive integers d_1 and d_2,*

$$\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(q) \leq \frac{(6d_1d_2 + 6d_1 + 1)q}{N'} + 6f_2 + 24f_1$$

where $f_1 = f_1(d_1)$ and $f_2 = f_2(d_2)$ are defined as in Theorem 1.

Proof. We will upper bound $\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(\mathcal{A})$ for an adversary $\mathcal{A}$ makes q queries to a random permutation π and its inverse π^{-1}. Let

$$\begin{aligned}
\mathsf{G}_{ex} = {} & \mathsf{E}^1\left((A_2 + A_3)^* A_1, (A_2 + A_3)^* B_1; d_1\right) \vee \mathsf{E}^1\left((A_3 + A_1)^* A_2, (A_3 + A_1)^* B_2; d_1\right) \\
& \vee \mathsf{E}^1\left((A_1 + A_2)^* A_3, (A_1 + A_2)^* B_3; d_1\right) \vee \mathsf{E}^1\left((B_2 + B_3)^* A_1, (B_2 + B_3)^* B_1; d_1\right) \\
& \vee \mathsf{E}^1\left((B_3 + B_1)^* A_2, (B_3 + B_1)^* B_2; d_1\right) \vee \mathsf{E}^1\left((B_1 + B_2)^* A_3, (B_1 + B_2)^* B_3; d_1\right) \\
& \vee \mathsf{E}^2\left(A_3^* A_1, A_3^* B_1, A_3^* A_2, A_3^* B_2; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(A_1^* A_2, A_1^* B_2, A_1^* A_3, A_1^* B_3; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(A_2^* A_3, A_2^* B_3, A_2^* A_1, A_2^* B_1; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_3^* A_1, B_3^* B_1, B_3^* A_2, B_3^* B_2; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_1^* A_2, B_1^* B_2, B_1^* A_3, B_1^* B_3; 2d_1d_2 + d_1\right) \\
& \vee \mathsf{E}^2\left(B_2^* A_3, B_2^* B_3, B_2^* A_1, B_2^* B_1; 2d_1d_2 + d_1\right).
\end{aligned}$$

Then by Theorem 1,

$$\mathbf{Pr}\left[\mathsf{G}_{ex}\right] \leq 6f_2 + 24f_1. \tag{29}$$

For $w \in \mathbb{F}_{2^n}$ and $j \in [1, q]$, we define an event

$$\begin{aligned}
\mathsf{P}_j(w) \Leftrightarrow \mathcal{A} \text{ sets } A_1x^{j_1} + B_1y^{j_1} + A_2x^{j_2} + B_2y^{j_2} + A_3x^{j_3} + B_3y^{j_3} = Cw \\
\text{with } j = \max\{j_1, j_2, j_3\}.
\end{aligned} \tag{30}$$

Since the occurrence of $\mathsf{P}_j(w)$ means that the j-th query determines a preimage of w, it follows that

$$\mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}(\mathcal{A}) \leq \max_{w \in \mathbb{F}_{2^n}} \mathbf{Pr}\left[\bigvee_{1 \leq j \leq q} \mathsf{P}_j(w)\right]. \tag{31}$$

For a fixed $w \in \mathbb{F}_{2^n}$, we use the following decomposition.

$$\mathbf{Pr}\left[\bigvee_{1 \leq j \leq q} \mathsf{P}_j(w)\right] \leq \mathbf{Pr}\left[\bigvee_{1 \leq j \leq q} \left(\mathsf{P}_j(w) \wedge \neg \mathsf{G}_{ex}\right)\right] + \mathbf{Pr}\left[\mathsf{G}_{ex}\right]. \tag{32}$$

Note that the event $\mathsf{P}_j(w)$ is identical with $\mathsf{F}^3_j(Cw)$ as defined in (14). So with the same analysis as $\mathsf{F}^3_j(Cw)$, the number of responses that determine a preimage of w is at most $6d_1d_2 + 6d_1 + 1$ for each query without the occurrence of G_{ex}. Therefore we obtain

$$\mathbf{Pr}\left[\bigvee_{1 \leq j \leq q} \left(\mathsf{P}_j(w) \wedge \neg \mathsf{G}_{ex}\right)\right] \leq \frac{(6d_1d_2 + 6d_1 + 1)q}{N'}. \tag{33}$$

The proof is complete from (29), (31), (32) and (33). □

Corollary 1. *Let $A = (a_{ij})$ be a matrix in $\mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$ such that $a_{11}a_{22} + a_{12}a_{21} \neq 0$. If $M(A)$ is column-sum independent, then*

$$\lim_{n\to\infty} \mathbf{Adv}^{\mathsf{pre}}_{\mathsf{lp}^A}\left(2^{\frac{2n}{3}}/n\right) = 0.$$

Proof. Let $q = 2^{\frac{2n}{3}}/n$, $d_1 = 2$ and $d_2 = 5 \cdot 2^{\frac{n}{3}} - 1$. Since

$$\frac{(6d_1d_2 + 6d_1 + 1)q}{N'} = \frac{(60 \cdot 2^{\frac{n}{3}} + 1)q}{N'} \leq \frac{122}{n}$$

$$f_1 = N\binom{q}{d_1+1}\left(\frac{1}{N'}\right)^{d_1+1} \leq N\left(\frac{2eq}{(d_1+1)N}\right)^{d_1+1} = N\left(\frac{2eq}{3N}\right)^3 = \frac{8e^3}{27n^3}$$

and

$$f_2 = N\binom{q}{d_2+1}\left(\frac{2q}{N'}\right)^{d_2+1} \leq N\left(\frac{2eq^2}{(d_2+1)N'}\right)^{d_2+1} \leq N\left(\frac{12e}{5n^2}\right)^{5\cdot 2^{\frac{n}{3}}}$$

using $N' \geq N/2$, it follows that

$$\lim_{n\to\infty} \frac{(6d_1d_2 + 6d_1 + 1)q}{N'} = \lim_{n\to\infty} f_2(d_1, d_2) = \lim_{n\to\infty} f_1(d_1) = 0.$$

The proof is complete from Theorem 3. □

Collision Resistance

Theorem 4. *Let $A = (a_{ij})$ be a matrix in $\mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$ such that $a_{11}a_{22} + a_{12}a_{21} \neq 0$, and let $M(A) = [A_1, B_1, A_2, B_2, A_3, B_3]$ and $C(A) = C$ satisfy the following conditions.*

1. *$M(A)$ is column-sum independent.*

2. *$[A_1, C]$, $[A_2, C]$, $[A_3, C]$, $[B_1, C]$ and $[B_2, C]$ are invertible.*

3. *The following 2×6 matrices are column-sum independent.*

$$\begin{aligned}
D^1 &= \left[[B_2, C]^{-1}[A_1, B_1], [B_1, C]^{-1}[A_2, B_2], [B_1, C]^{-1}[A_3, B_3]\right],\\
D^2 &= \left[[A_2, C]^{-1}[A_1, B_1], [A_1, C]^{-1}[A_2, B_2], [A_1, C]^{-1}[A_3, B_3]\right],\\
D^3 &= \left[[A_3, C]^{-1}[A_1, B_1], [A_3, C]^{-1}[A_2, B_2], [A_1, C]^{-1}[A_3, B_3]\right],\\
D^4 &= \left[[A_3, C]^{-1}[A_1, B_1], [A_3, C]^{-1}[A_2, B_2], [A_2, C]^{-1}[A_3, B_3]\right].
\end{aligned}$$

Then for positive integers d_1, d_2 and d_3,

$$\begin{aligned}
\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(q) \leq{}& Nq^2\left(\frac{6d_1d_2 + 6d_1 + 1}{N'}\right)^2 + \frac{3(2d_1d_2 + d_1)q + 3q\max\{7d_3q, 2d_1d_2 + d_1\}}{N'}\\
&+ 4g_3 + 15g_2 + 121f_2 + 333f_1,
\end{aligned}$$

where $f_1 = f_1(d_1)$, $f_2 = f_2(d_2)$, $g_2 = g_2(d_1, d_2)$ and $g_3 = g_3(d_1, d_2, d_3)$ are defined as in Theorem 1 and 2.

Proof. We will upper bound $\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(\mathcal{A})$ for an adversary $\mathcal{A}$ makes q queries to a random permutation π and its inverse π^{-1}. Let

$$\begin{aligned}\mathsf{H}_{ex} = \mathsf{F}_{ex} &\vee \mathsf{E}^2\left(\bar{B}^*A_1, \bar{B}^*B_1, \bar{B}^*A_2, \bar{B}^*B_2; 2d_1d_2 + d_1\right)\\ &\vee \mathsf{F}^3\left(D^1; 7d_3\right) \vee \mathsf{F}^3\left(D^2; 7d_3\right) \vee \mathsf{F}^3\left(D^3; 7d_3\right) \vee \mathsf{F}^3\left(D^4; 7d_3\right)\end{aligned}$$

where $\bar{B} = B_1 + B_2 + B_3$ and F_{ex} is defined as in (16). Then it follows that

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{H}_{ex}\right] &= 3g_2 + 24f_2 + 66f_1 + (f_2 + 3f_1) + 4(g_3 + 3g_2 + 24f_2 + 66f_1)\\ &= 4g_3 + 15g_2 + 121f_2 + 333f_1. \end{aligned}\tag{34}$$

For $j \in [1, q]$ and ρ^1, $\rho^2 \in \{0,1\}^3 \backslash \{(0,0,0)\}$, we define the following event.

$$\begin{aligned}\mathsf{C}^2_j(\rho^1, \rho^2) \Leftrightarrow &\mathcal{A} \text{ sets } A_1x^{j^1_1} + B_1y^{j^1_1} + A_2x^{j^1_2} + B_2y^{j^1_2} + A_3x^{j^1_3} + B_3y^{j^1_3} = Cw \text{ and}\\ &A_1x^{j^2_1} + B_1y^{j^2_1} + A_2x^{j^2_2} + B_2y^{j^2_2} + A_3x^{j^2_3} + B_3y^{j^2_3} = Cw,\\ &\text{ where } w \in \mathbb{F}_{2^n},\ (j^1_1, j^1_2, j^1_3) \neq (j^2_1, j^2_2, j^2_3), j = \max_{\substack{i=1,2,3\\ s=1,2}}\{j^s_i\} \text{ and}\\ &j^s_i = j \text{ if and only if } \rho^s_i = 1 \text{ for } i = 1,2,3 \text{ and } s = 1,2.\end{aligned}$$

Here ρ^1 and ρ^2 specify the positions where the j-th query contributes within the two different evaluations. Since ρ^1, $\rho^2 \neq (0,0,0)$, the occurrence of $\mathsf{C}^2_j(\rho^1, \rho^2)$ means that the *single* j-th query completes two colliding evaluations of lp^A_{231} at the same time. Let

$$\mathsf{C}^1 = \bigvee_{\substack{1 \le j_1 < j_2 \le q\\ w \in \mathbb{F}_{2^n}}} \left(\mathsf{P}_{j_1}(w) \wedge \mathsf{P}_{j_2}(w)\right)$$

where events $\mathsf{P}_{j_1}(w)$ and $\mathsf{P}_{j_2}(w)$ are defined as in (30) and let

$$\mathsf{C}^2 = \bigvee_{\substack{1 \le j \le q\\ (\rho^1, \rho^2) \in P}} \mathsf{C}^2_j(\rho^1, \rho^2)$$

where $P = P_1 \cup P_2 \cup P_3$ and

$$\begin{aligned}P_1 &= \left\{\left(\rho^1, \rho^2\right) : \exists\, i \in \{1,2,3\} \text{ such that } \rho^1_i = \rho^2_i = 1\right\},\\ P_2 &= \left\{\left((1,1,0),(0,0,1)\right), \left((1,0,1),(0,1,0)\right), \left((0,1,1),(1,0,0)\right)\right\},\\ P_3 &= \left\{\left((1,0,0),(0,1,0)\right), \left((1,0,0),(0,0,1)\right), \left((0,1,0),(0,0,1)\right)\right\}.\end{aligned}$$

Then by the symmetry of ρ^1 and ρ^2, it follows that

$$\mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}(\mathcal{A}) = \mathbf{Pr}\left[\mathsf{C}^1 \vee \mathsf{C}^2\right] \le \mathbf{Pr}\left[\mathsf{H}_{ex}\right] + \mathbf{Pr}\left[\mathsf{C}^1 \wedge \neg\mathsf{H}_{ex}\right] + \mathbf{Pr}\left[\mathsf{C}^2 \wedge \neg\mathsf{H}_{ex}\right]. \tag{35}$$

Estimation of $\mathbf{Pr}\left[\mathsf{C}^1 \wedge \neg\mathsf{H}_{ex}\right]$. If events P are defined as (30), then

$$\mathbf{Pr}\left[\mathsf{C}^1 \wedge \neg\mathsf{H}_{ex}\right] = \mathbf{Pr}\left[\bigvee_{\substack{1 \le j_1 < j_2 \le q\\ w \in \mathbb{F}_{2^n}}} \left(\mathsf{P}_{j_1}(w) \wedge \mathsf{P}_{j_2}(w) \wedge \neg\mathsf{H}_{ex}\right)\right].$$

With a similar argument as the analysis of preimage resistance, we obtain

$$\mathbf{Pr}\left[\mathsf{P}_{j_1^*}(w^*) \wedge \mathsf{P}_{j_2^*}(w^*) \wedge \neg\mathsf{H}_{ex}\right] \leq \left(\frac{6d_1d_2 + 6d_1 + 1}{N'}\right)^2$$

for fixed $w^* \in \mathbb{F}_{2^n}$ and $1 \leq j_1^* < j_2^* \leq q$. Therefore we have

$$\begin{aligned} \mathbf{Pr}\left[\mathsf{C}^1 \wedge \neg\mathsf{H}_{ex}\right] &= \mathbf{Pr}\left[\bigvee_{\substack{1\leq j_1<j_2\leq q\\ w\in\mathbb{F}_{2^n}}} \left(\mathsf{P}_{j_1}(w) \wedge \mathsf{P}_{j_2}(w) \wedge \neg\mathsf{H}_{ex}\right)\right] \\ &\leq Nq^2\left(\frac{6d_1d_2 + 6d_1 + 1}{N'}\right)^2. \end{aligned} \tag{36}$$

Estimation of $\mathbf{Pr}\left[\mathsf{C}^2 \wedge \neg\mathsf{H}_{ex}\right]$. In order to use

$$\mathbf{Pr}\left[\mathsf{C}^2 \wedge \neg\mathsf{H}_{ex}\right] \leq \sum_{\substack{1\leq j\leq q\\ (\rho^1,\rho^2)\in P}} \mathbf{Pr}\left[\mathsf{C}_j^2(\rho^1,\rho^2) \wedge \neg\mathsf{H}_{ex}\right] \tag{37}$$

we focus on the estimation of $\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1,\rho^2\right) \wedge \neg\mathsf{H}_{ex}\right]$ for each $(\rho^1,\rho^2) \in P$. We consider three cases as follows.

Case 1: We estimate the probability $\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1,\rho^2\right) \wedge \neg\mathsf{H}_{ex}\right]$ for $(\rho^1,\rho^2) \in P_1$. Suppose that $\rho^1 = \rho^2 = (1,0,0)$. If the event $\mathsf{C}_j^2((1,0,0),(1,0,0))$ occurs, then it holds that

$$A_1x^j + B_1y^j + A_2x^{j_2^1} + B_2y^{j_2^1} + A_3x^{j_3^1} + B_3y^{j_3^1} = Cw, \tag{38}$$

$$A_1x^j + B_1y^j + A_2x^{j_2^2} + B_2y^{j_2^2} + A_3x^{j_3^2} + B_3y^{j_3^2} = Cw \tag{39}$$

for some j_2^1, j_3^1, j_2^2, $j_3^2 < j$ and $w \in \mathbb{F}_{2^n}$. The equations (38) and (39) imply that

$$A_2x^{j_2^1} + B_2y^{j_2^1} + A_3x^{j_3^1} + B_3y^{j_3^1} = A_2x^{j_2^2} + B_2y^{j_2^2} + A_3x^{j_3^2} + B_3y^{j_3^2}.$$

Therefore it follows that

$$\mathsf{C}_j^2\left((1,0,0),(1,0,0)\right) \Rightarrow \mathsf{F}^2\left(A_2,B_2,A_3,B_3;1\right) \Rightarrow \mathsf{H}_{ex},$$

and hence,

$$\mathbf{Pr}\left[\mathsf{C}_j^2\left((1,0,0),(1,0,0)\right) \wedge \neg\mathsf{H}_{ex}\right] = 0.$$

The same argument applies to any event $\mathsf{C}_j^2\left(\rho^1,\rho^2\right)$ such that $\rho^1 \wedge \rho^2 \neq (0,0,0)$.

Case 2: We estimate the probability $\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1,\rho^2\right) \wedge \neg\mathsf{H}_{ex}\right]$ for $(\rho^1,\rho^2) \in P_2$. Say $\rho^1 = (1,1,0)$ and $\rho^2 = (0,0,1)$. Suppose that $\mathcal{A}$ makes the j-th query $\pi(x^*)$. Among $(N-(j-1))$ possible responses for $y = \pi(x^*)$, we need to upper bound the number of $y = \pi(x^*)$ satisfying

$$\left(A_1 + A_2\right)x^* + \left(B_1 + B_2\right)y + A_3x^{j_3^1} + B_3y^{j_3^1} = Cw, \tag{40}$$

$$A_1x^{j_1^2} + B_1y^{j_1^2} + A_2x^{j_2^2} + B_2y^{j_2^2} + A_3x^* + B_3y = Cw \tag{41}$$

for some j_3^1, j_1^2, $j_2^2 < j$ and $w \in \mathbb{F}_{2^n}$. Adding (40) and (41), we obtain

$$\bar{A}x^* + \bar{B}y + A_1 x^{j_1^2} + B_1 y^{j_1^2} + A_2 x^{j_2^2} + B_2 y^{j_2^2} + A_3 x^{j_3^1} + B_3 y^{j_3^1} = 0$$

which implies the following equation.

$$\bar{B}^* A_1 x^{j_1^2} + \bar{B}^* B_1 y^{j_1^2} + \bar{B}^* A_2 x^{j_2^2} + \bar{B}^* B_2 y^{j_2^2} + \left(\bar{B}^* A_3 x^{j_3^1} + \bar{B}^* B_3 y^{j_3^1}\right) = \bar{B}^* \bar{A} x^*. \qquad (42)$$

The number of solutions (j_3^1, j_1^2, j_2^2) to (42) is at most $(2d_1d_2 + d_1)q$ without the occurrence of H_{ex}. (We can first fix the j_3^1-th query, and then count the number of pairs (j_1^2, j_2^2) satisfying equation (42).) For events $\mathsf{C}_j^2\left((0,1,1),(1,0,0)\right)$ and $\mathsf{C}_j^2\left((1,0,1),(0,1,0)\right)$, we have the same equation as (42). Therefore we have

$$\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1, \rho^2\right) \wedge \neg \mathsf{H}_{ex}\right] \le \frac{(2d_1d_2 + d_1)q}{N'}.$$

Case 3: We estimate the probability $\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1, \rho^2\right) \wedge \neg \mathsf{H}_{ex}\right]$ for $(\rho^1, \rho^2) \in P_3$. Say $\rho^1 = (1,0,0)$ and $\rho^2 = (0,1,0)$. Suppose that $\mathcal{A}$ makes the j-th query $\pi(x^*)$. Among $(N - (j-1))$ possible responses for $y = \pi(x^*)$, we need to upper bound the number of $y = \pi(x^*)$ satisfying

$$\begin{aligned} A_1 x^* + B_1 y + A_2 x^{j_2^1} + B_2 y^{j_2^1} + A_3 x^{j_3^1} + B_3 y^{j_3^1} &= Cw, \\ A_1 x^{j_1^2} + B_1 y^{j_1^2} + A_2 x^* + B_2 y + A_3 x^{j_3^2} + B_3 y^{j_3^2} &= Cw, \end{aligned}$$

for some j_2^1, j_3^1, j_1^2, $j_3^2 < j$ and $w \in \mathbb{F}_{2^n}$. Removing variables y and w from this system of equations, we obtain the following equation.

$$\begin{aligned} &[B_1, C]^{-1} A_2 x^{j_2^1} + [B_1, C]^{-1} B_2 y^{j_2^1} + [B_1, C]^{-1} A_3 x^{j_3^1} + [B_1, C]^{-1} B_3 y^{j_3^1} \\ &\quad + [B_2, C]^{-1} A_1 x^{j_1^2} + [B_2, C]^{-1} B_1 y^{j_1^2} + \left([B_2, C]^{-1} A_3 x^{j_3^2} + [B_2, C]^{-1} B_3 y^{j_3^2}\right) \\ &\qquad = \left([B_1, C]^{-1} A_1 + [B_2, C]^{-1} A_2\right) x^*. \qquad (43) \end{aligned}$$

For each j_3^2, the number of solutions (j_2^1, j_3^1, j_1^2) to (43) is at most $7d_3$ without the occurrence of $\mathsf{F}^3\left(D^1; 7d_3\right)$. Therefore the number of solutions $(j_2^1, j_3^1, j_1^2, j_3^2)$ to (43) is at most $7d_3q$.

One special case is when ρ^1 or ρ^2 is $(0,0,1)$ and $\mathcal{A}$ makes a forward query $y = \pi(x^*)$. Since $[B_3, C]$ is not invertible, the above argument does not apply to this case. Say $\rho^1 = (0,0,1)$. Then the response $y = \pi(x^*)$ should satisfy

$$A_1 x^{j_1^1} + B_1 y^{j_1^1} + A_2 x^{j_2^1} + B_2 y^{j_2^1} + A_3 x^* + B_3 y = Cw, \qquad (44)$$

for some j_1^1, j_2^1 and $w \in \mathbb{F}_{2^n}$. Multiplying B_3^* on both sides of (44) and using $B_3^* C = 0$, we see that each y satisfying (44) is associated with a solution (j_1^1, j_2^1) to the following equation.

$$B_3^* A_1 x^{j_1^1} + B_3^* B_1 y^{j_1^1} + B_3^* A_2 x^{j_2^1} + B_3^* B_2 y^{j_2^1} = B_3^* A_3 x^*. \qquad (45)$$

The number of solutions (j_1, j_2) to (45) is at most $2d_1d_2 + d_1$ without the occurrence of $\mathsf{E}^2\left(B_3^* A_1, B_3^* B_1, B_3^* A_2, B_3^* B_2; 2d_1d_2 + d_1\right)$. Therefore we have

$$\mathbf{Pr}\left[\mathsf{C}_j^2\left(\rho^1, \rho^2\right) \wedge \neg \mathsf{H}_{ex}\right] \le \frac{\max\{7d_3q, 2d_1d_2 + d_1\}}{N'}.$$

To summarize the analysis for the three cases, we conclude that

$$\begin{aligned}\mathbf{Pr}\left[\mathsf{C}^2 \wedge \neg\mathsf{H}_{ex}\right] &\leq \sum_{\substack{1\leq j\leq q\\ (\rho^1,\rho^2)\in P}} \mathbf{Pr}\left[\mathsf{C}^2_j(\rho^1,\rho^2) \wedge \neg\mathsf{H}_{ex}\right] \\ &\leq \frac{3(2d_1d_2+d_1)q + 3q\max\{7d_3q, 2d_1d_2+d_1\}}{N'}. \qquad (46)\end{aligned}$$

Now the proof is complete from (34), (35), (36) and (46). □

Corollary 2. *Let $A = (a_{ij})$ be a matrix in $\mathcal{M}_{\mathbb{F}_{2^n}}(2,3,1)$ such that $a_{11}a_{22} + a_{12}a_{21} \neq 0$. If $M(A)$ and $C(A)$ satisfy the conditions described in Theorem 4, then*

$$\lim_{n\to\infty} \mathbf{Adv}^{\mathsf{col}}_{\mathsf{lp}^A}\left(2^{\frac{n}{2}}/n^{1+\epsilon}\right) = 0,$$

for any $\epsilon > 0$.

Proof. Let $q = 2^{\frac{n}{2}}/n^{1+\epsilon}$ and $(d_1, d_2, d_3) = (1, n, 3)$. Then it is easy to check that

$$\max\{7d_3q, 2d_1d_2 + d_1\} = 7d_3q$$

and

$$\begin{aligned} Nq^2\left(\frac{6d_1d_2+6d_1+1}{N'}\right)^2 &= O\left(\frac{1}{n^{2\epsilon}}\right), & \frac{3(2d_1d_2+d_1)q + 21d_3q^2}{N'} &= O\left(\frac{1}{n^{2+2\epsilon}}\right), \\ f_1(d_1) &= O\left(\frac{1}{n^{2+2\epsilon}}\right), & f_2(d_2) &= O\left(\frac{1}{n^{(2+2\epsilon)n}}\right), \\ g_2(d_1,d_2) &= O\left(\frac{1}{n^{1+2\epsilon}}\right), & g_3(d_1,d_2,d_3) &= O\left(\frac{1}{n^{4\epsilon}}\right). \end{aligned}$$

Since all the terms converge to 0 as n goes to infinity, and by Theorem 4, we obtain the corollary. □

Example 1. Let $n = 128$ and let $\mathbb{F}_{2^{128}} = \mathbb{F}[\zeta]/(\zeta^{128} + \zeta^7 + \zeta^2 + \zeta + 1)$ be a finite field of order 2^{128}, where $f(\zeta) = \zeta^{128} + \zeta^7 + \zeta^2 + \zeta + 1$ is an irreducible polynomial over $\mathbb{F}_2$. For simplicity of computation, assume that $a_{23} = 0$, $a_{45} = 1$, and

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then we have

$$M(A) = \begin{bmatrix} a_{31} & a_{33} & a_{32} & a_{34} & 1 & 0 \\ a_{41} & a_{43} & a_{42} & a_{44} & 0 & 1 \end{bmatrix} \quad \text{and} \quad C(A) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

If we set $a_{31} = a_{44} = \zeta$, $a_{33} = a_{42} = \zeta^3$, $a_{32} = a_{43} = \zeta^2 + \zeta$, and $a_{34} = a_{41} = 1$, then

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \zeta & \zeta^2+\zeta & \zeta^3 & 1 & 0 \\ 1 & \zeta^3 & \zeta^2+\zeta & \zeta & 1 \end{bmatrix}$$

and the corresponding matrices $M(A)$ and $C(A)$ satisfy all the conditions described in Theorem 4. With respect to the threshold distinguishing advantage 1/2, the resulting compression function lp^A is preimage resistant up to $2^{81.5}$ queries (with $(d_1, d_2) = (2, 2^{37.5})$) and collision resistant up to $2^{56.6}$ queries (with $(d_1, d_2, d_3) = (2, 8, 4)$).

References

1. P. Barreto and V. Rijmen. The Whirlpool hashing function. Primitve submitted to NESSIE, September 2000, revised on May 2003.
2. G. Bertoni, J. Daemen, M. Peeters and G. Van Assche. On the indifferentiability of the Sponge construction. Eurocrypt 2008, LNCS 4965, pp. 181–197, Springer-Verlag, 2008.
3. J. Black, M. Cochran and T. Shrimpton. On the impossibility of highly-efficient blockcipher-based hash functions. Eurocrypt 2005, LNCS 3494, pp. 526–541, Springer-Verlag, 2005.
4. J. Black, P. Rogaway and T. Shrimpton. Black-box analysis of the block-cipher-based hash-function construction from PGV. Crypto 2002, LNCS 2442, pp. 320–325, Springer-Verlag, 2002.
5. S. Hirose. Provably secure double-block-length hash functions in a black-box model. ICISC 2004, LNCS 3506, pp. 330–342, Springer-Verlag, 2005.
6. S. Hirose. Some plausible construction of double-block-length hash functions. FSE 2006, LNCS 4047, pp. 210–225, Springer-Verlag, 2006.
7. S. Matyas, S. Meyer and J. Oseas. Generating strong one-way functions with cryptographic algorithm. IBM Technical Disclosure Bulletin 27, 10a, pp. 5658–5659, 1985.
8. B. Preneel, R. Govaerts and J. Vandewalle. Hash functions based on block ciphers: A synthetic approach. Crypto 1993, LNCS 773, pp. 368–378, Springer-Verlag, 1994.
9. T. Ristenpart and T. Shrimpton. How to build a hash function from any collision-resistant function. Asiacrypt 2007, LNCS 4833, pp. 147–163, Springer-Verlag, 2007.
10. P. Rogaway and J. P. Steinberger. Constructing cryptographic hash functions from fixed-key blockciphers. Crypto 2008, LNCS 5157, pp. 433–450, Springer-Verlag, 2008.
11. P. Rogaway and J. P. Steinberger. Security/efficiency tradeoffs for permuation-based hashing. Eurocrypt 2008, LNCS 4965, pp. 220–236, Springer-Verlag, 2008.
12. T. Shrimpton and M. Stam. Building a collision-resistant function from non-compressing primitives. ICALP 2008, LNCS 5126, pp. 643–654, Springer-Verlag, 2008.
13. M. Stam. Beyond uniformity: Better security/efficiency tradeoffs for compression functions. Crypto 2008, LNCS 5157, pp. 397–412, Springer-Verlag, 2008.
14. J. P. Steinberger. The collision intractability of MDC-2 in the ideal-cipher model. Eurocrypt 2007, LNCS 4515, pp. 34–51, Springer-Verlag, 2008.
15. R. Winternitz. A secure one-way hash function built from DES. IEEE Symposium on Information Security and Privacy, pp. 88–90, IEEE Press, 1984.