Permutation Based EDM: An Inverse Free BBB Secure PRF

Avijit Dutta[1] and Mridul Nandi[2] and Suprita Talnikar[2]

Institute for Advancing Intelligence, TCG-CREST, Kolkata, India.
Indian Statistical Institute, Kolkata, India.
avirocks.dutta13@gmail.com,mridul.nandi@gmail.com,suprita45@gmail.com

Abstract. In CRYPTO 2019, Chen et al. have initiated an interesting research direction in designing PRF based on public permutations. They have proposed two beyond the birthday bound secure n-bit to n-bit PRF constructions, i.e., SoEM22 and SoKAC21, which are built on public permutations, where n is the size of the permutation. However, both of their constructions require two independent instances of public permutations. In FSE 2020, Chakraborti et al. have proposed a single public permutation based n-bit to n-bit beyond the birthday bound secure PRF, which they refer to as PDMMAC. Although the construction is minimal in the number of permutations, it requires the inverse call of its underlying permutation in their design. Coming up with a beyond the birthday bound secure public permutation based n-bit to n-bit PRF with a single permutation and two forward calls was left as an open problem in their paper. In this work, we propose pEDM, a single permutation based n-bit to n-bit PRF with two calls that do not require invertibility of the permutation. We have shown that our construction is secured against all adaptive information-theoretic distinguishers that make roughly up to $2^{2n/3}$ construction and primitive queries. Moreover, we have also shown a matching attack with similar query complexity that establishes the tightness of our security bound.

Keywords: Public Permutations · EDM · PDMMAC · Expectation Method.

1 Introduction

Luby and Rackoff [LR88], in their seminal work, have shown how to construct a keyed pseudorandom permutation (PRP) or, in other words, block cipher from secret keyed pseudorandom functions (PRF). Their work was a theoretical model for formally arguing the security of DES block cipher, which consists of r rounds of Feistel constructions invoking independent instances of keyed functions. However, it was soon realized the necessity of designing PRFs out of PRPs as primitives of cryptographic designs [BKR98]. Because we usually seek PRF security from a mode of operation and it is generally easier to design PRPs than PRFs. One of the biggest challenges in designing PRFs is to design a secure non-invertible round function that can be iterated multiple times to produce a secure PRF. However, iterating the non-invertible round function multiple times is hard to get right, as collision probabilities are amplified with each iteration [BDD+17, MN17b]. Nevertheless, Mennink and Neves [MN17b] designed a dedicated PRF called FastPRF from scratch, even though their design is based on grouping the round functions of a PRP. Moreover, there are plenty of cryptographic modes that do not require the invertibility of its underlying primitives [MV04, BR02, Iwa06, WC81, CDNY18, BCDM19, DEMS19, DDNP18, DDNY18, CS16, LPTY16]. Hence, in such cases, for realizing the PRF security of a mode of operation, it is a better and economical choice to use PRFs as the underlying primitive of the mode over PRPs, which are designed to be efficient in both forward

and inverse direction. In fact, as substantial evidence of our argument, counter mode of encryption generally offers a better security guarantee when instantiated with a PRF over a PRP because one can distinguish counter mode with PRP from the random encryption with $2^{n/2}$ queries, where n is the block size of the PRP. On the other hand, the counter mode with PRF behaves identically with the random encryption scheme modulo the PRF advantage of the keyed function.

Due to the classical result of PRF-PRP switching lemma [CN08, BKR94, BR06], a PRP E_k can be replaced with a PRF F_k until the number of invocations to the primitive exceeds $2^{n/2}$, where n is the block size of the permutation. Such a solution is adequate when the block size of the permutation is large, (e.g., AES 128). However, the solution may not be good enough when the block size is small (e.g., block size of 64 bits). This is paricularly relevant when one instantiate cryptographic schemes using lightweight block ciphers like PRESENT [BKL$^+$07], GIFT [BPP$^+$17] etc. The block size of such lightweight block ciphers is typically 64 bits. As a result, if one uses these block ciphers as PRF in cryptographic designs, it can ensure only 32 bits of security, which is not practical in today's world of computational power. As a remedy of this, exploring the cryptographic designs, which retains security even after invoking the primitive more than $2^{n/2}$ times, started to begin. Such designs are popularly known as **beyond birthday bound** (BBB) secure designs. In this direction, Hall et al. [HWKS98] have proposed a BBB secure PRF, called *Truncation* that takes an n-bit block cipher E_k and truncates the result to a bits. This construction was later proven to be secured upto $2^{n-a/2}$ queries [BI99, GG16]. Bellare et al. [BKR98] have proposed the Sum of Permutations (SoP) constructions which returns the xor of the outputs of two n-bit independent permutations.

$$\mathsf{SoP}^{\mathsf{P}_1,\mathsf{P}_2}(x) \overset{\Delta}{=} \mathsf{P}_1(x) \oplus \mathsf{P}_2(x).$$

This construction was proven to be secured upto $2^{2n/3}$ queries [Luc00] and recently it has been shown to be secured upto 2^n queries [DHT17]. Cogliati and Seurin [CS16] have proposed another candidate of beyond birthday bound secure PRF which they call *Encrpted Davis Meyer* (EDM) construction and they have shown that EDM achieves $2n/3$ bit security.

$$\mathsf{EDM}^{\mathsf{P}_1,\mathsf{P}_2}(x) \overset{\Delta}{=} \mathsf{P}_2(\mathsf{P}_1(x) \oplus x).$$

Later in [MN17a], Mennink and Neves showed an optimal security of the consruction. In the same paper, they also proposed a dual variant of EDM which he referred to as EDMD

$$\mathsf{EDMD}^{\mathsf{P}_1,\mathsf{P}_2}(x) \overset{\Delta}{=} \mathsf{P}_2(\mathsf{P}_1(x)) \oplus \mathsf{P}_1(x),$$

and showed its optimal PRF security. However, the proof of both the constructions are inherently based on a debated result of Mirror theory for general $\xi_{\max}$ [DNS20]. Guo et al. [GSWG19] have proposed SUMPIP, a contender of SoP construction

$$\mathsf{SUMPIP}^{\mathsf{P}}(x) \overset{\Delta}{=} \mathsf{P}(x) \oplus \mathsf{P}^{-1}(x).$$

In contrast to the single permutation variant of SoP which takes $n-1$ bit input, SUMPIP is the first single permutation based PRF that takes n-bit input and returns n-bit output. In the same paper, authors have also shown that single permutation variant of EDM and EDMD achieves $2n/3$-bit security. Concurrent to this, Cogliati and Seurin [CS18] have also shown $2n/3$ bit security for the single-keyed EDM construction. Very recently, Gunsing and Mennink [GM20] proposed a new approach to design a block cipher based PRF which they refer to as *Summation-Truncation Hybrid* (STH) technique. STH takes an $(n-1)$-bit input x, truncates the leftmost a bits of $\mathsf{E}(x\|0), \mathsf{E}(x\|1)$, and sums the discarded $n-a$ bits of $\mathsf{E}(x\|0)$ and $\mathsf{E}(x\|1)$ to produce an $(n+a)$-bit output. They showed the construction provides $2^{n-a/2}$ bits of security, where $n-a$ is the number of discarded bits.

1.1   Permutation Based Cryptography

All the above-discussed PRFs are built using block cipher as their underlying primitive and even stronger in most constructions this primitive is evaluated only in the forward direction. As block ciphers are designed to be efficient in both the forward and the inverse direction, block ciphers are thus over-engineered primitives for such purpose [CLM19a]. On the other extreme, cryptographic public permutations are particularly designed to be fast in the forward direction, but not necessarily in the inverse. Examples of such permutations include Keccak [BDPA13], Gimli [BKL$^+$17], SPONGENT [BKL$^+$13] etc. Moreover, in most of the cases evaluating an unkeyed public permutation is faster than evaluating a keyed block cipher, as the latter involves evaluating the underlying key scheduling algorithm each time the block cipher is invoked in the design. [1] Moreover, we do not need to store the round keys in permutation based designs, and designing permutation is usually simple over designing a block cipher. In this regard, we would like to quote the following statements of Bertoni et al. [BDP$^+$16]

" ... the inverse mapping of block ciphers imposes a separation of the processing of the n + k bits of the input. The key is processed in a key schedule and the data in the data path, and there can be no diffusion from the data path to the key schedule, which strongly limits the potential diffusion ... Such a restriction is not present in the design of cryptographic permutations as they do not make a distinction between the processing of key and data input as there is no specific key input."

With the advent of public permutation based designs and the efficiencies of evaluating it in the forward direction, numerous public permutation based inverse-free hash and authenticated encryption designs have been proposed [RBB03, CDNY18, NMSS19, DEMS19, BKL$^+$17, DEM$^+$20, CN19, DHP$^+$20]. The use of cryptographic permutation gained momentum during SHA-3 competition [RBB03]. Furthermore, the selection of the permutation based Keccak sponge function as the SHA-3 standard has given a high level of confidence in using this primitive in the community. Today, the permutation based sponge construction has become a successful and full-fledged alternative to the block cipher based modes. In fact, in the first round of the ongoing NIST lightweight competition [NIS18], 24 out of 57 submissions are based on cryptographic permutations, and out of 24, 16 permutation based proposals have qualified for the second round. These statistics depict the wide adoption of permutation based designs [CDNY18, BKL$^+$17, BCDM19, CN19, DHP$^+$20, DEMS19] in the community.

However, most of the permutation based cryptographic schemes generally provide lower security bound with respect to the permutation state size. For example, most of the sponge-based modes, in general, provides $c/2$ bits of security (exceptions are [CDNY18, DMA17]), where $c < b$ is the capacity part of the permutation, and b is its total state size. Nevertheless, the state size of a permutation is typically larger than the block size of a message (e.g., state size of KECCAK is 1600 bits), allowing the adequacy of the birthday bound in practice. However, the state size of lightweight permutations such as SPONGENT [BKL$^+$13] and PHOTON [GPP11] go as low as 88 and 100 bits, respectively. For these types of permutations, birthday bound solutions are inadequate. Thus, it can be highly interesting to design public permutation based cryptographic schemes that provide beyond the birthday bound security with respect to the permutation state size.

This line of research was initiated by Chen et al. in [CLM19b] where they proposed two fixed-input and fixed-output length beyond birthday bound secure PRFs based on public permutations - one is in the parallel mode and the other is in the sequential mode. (i)

---

[1]One might argue that caching the round keys of the block cipher eliminate the problem. But it requires more storage space than storing the master key of the block cipher, e.g., storing the round keys of AES-128 requires ten times more space than storing its master key.

For the parallel mode, they have shown that the sum of two independent instances of Even-Mansour [EM97] cipher, which they refer to as SoEM22,

$$\mathsf{SoEM22}^{\mathsf{P_1, P_2}}_{k_1, k_2}(x) \triangleq \mathsf{P}_1(x \oplus k_1) \oplus \mathsf{P}_2(x \oplus k_2) \oplus k_1 \oplus k_2,$$

provides a tight $2n/3$-bit security. This construction was later extended by Bhattacharya et al. [BDLN20], where they showed the beyond birthday bound security of the domain separated variant of SoEM22. They have also proved that one cannot reduce the number of keys of SoEM22 without degrading the security bound to the birthday limit. (ii) For the sequential mode, Chen et al. proposed SoKAC21,

$$\mathsf{SoKAC21}^{\mathsf{P_1, P_2}}_{k}(x) \triangleq \mathsf{P}_2(\mathsf{P}_1(x \oplus k) \oplus k) \oplus \mathsf{P}_1(x \oplus k) \oplus k,$$

which was proven to have a tight $2n/3$-bit security. However, later in [Nan20], Nandi exhibited a birthday bound attack on SoKAC21 and hence falsifying the security claim of the construction. In [CNTY20], Chakraborti et al. have proposed PDM MAC, a beyond birthday bound secure single permutation based fixed input and fixed output length PRF that opearates in sequential mode. The design of PDM MAC is motivated from the *Decrypted Davis-Meyer* (DDM) construction,

$$\mathsf{DDM}_k(x) \triangleq \mathsf{P}^{-1}(\mathsf{P}(x) \oplus x).$$

PDM MAC requires an n-bit key k and an n-bit public permutation P to generate the output as follows:

$$\mathsf{PDM}^{\mathsf{P}}_k(x) \triangleq \mathsf{P}^{-1}(\mathsf{P}(x \oplus k) \oplus (x \oplus 3k)) \oplus 2k.$$

They extended the construction towards designing a BBB secure single permutation and single keyed variant of nonce based MAC [2]. Although, minimally structured, PDM MAC and its related MAC constructions, i.e., PDM* MAC [CNTY20] and 1K-PDM* MAC [CNTY20] require the invertiblity of the permutation P (similar to the design of DWCDM [DDNY18]). However, inverse call in PDM MAC somewhat brings down one of the advantages of using cryptographic permutations in a mode, i.e, the efficiency of evaluating the permutation in forward direction. In fact, it was stated as an open problem [CNTY20] to design a BBB secure single permuation based PRF with two forward calls. Not only this, inverse-free designs become one of the important design aspects in today's cryptography as designs that rely solely on the forward call of the permutation makes a very low footprint in a combined implementation of the mode [BN16]. Therefore, until now we do not have any beyond birthday bound secure single permutation based fixed input and fixed output length PRF that opearates in sequential mode with two forward calls [3].

1.2   Our Contribution

In this paper, we propose pEDM, the first fixed-input and fixed-output length single permutation based beyond the birthday bound secure PRF that operates in a sequential mode without requiring the inverse call of the permutation. Our design is motivated by the EDM construction. In particular, pEDM with $2n$-bit keys and n-bit public permutation, takes an n-bit input and returns an n-bit output as follows:

$$\mathsf{pEDM}^{\mathsf{P}}_{k_1, k_2}(x) \triangleq \mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus (x \oplus k_1) \oplus k_2) \oplus k_1.$$

---

[2]Single permutation based nonce based MAC was also proposed in [DN20] that does not require invertibility of the permutation

[3]Chen et al. [CLM19b] shown a $n/2$-bit attack on SoKAC1 construction, $\mathsf{SoKAC1}^{\mathsf{P}}_{k_1, k_2}(x) = \mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus k_2) \oplus \mathsf{P}(x \oplus k_1) \oplus k_2 \oplus k_1$. However, Chakraborti et al. [CNTY20] claimed that the attack is possibly wrong and shown a $2n/3$-bit attack on it. They also conjectured that this attack bound is indeed tight.

We have shown that pEDM is secured against all adaptive information-theoretic distinguishers that make roughly up to $2^{2n/3}$ construction and primitive queries. We also show a matching attack of the same complexity and establish the tightness of the security bound. Note that we could directly realize a permutation based PRF by instantiating the block cipher of the single-keyed variant of the EDM construction with 2-round Even-Mansour cipher. But that leads to having 4 permutation calls in total with $6n$-bit keys. Compared to such a straightforward solution, our construction altogether saves 2 permutation calls and $4n$-bit keys. Although pEDM uses a single permutation call with no inverse functionality, the number of keys required is one more than the number of keys required in PDM MAC. Currently, we do not know whether our construction is prone to the birthday attack with a single key. However, we believe that it can be proven secure beyond the birthday bound with only an n-bit key. We show the PRF advantage of this construction through an extended distinguishing game and apply the expectation method to bound its distinguishing advantage. In table 1, we compare the structures of several public permutation based PRFs with single-block input, single-block output and multi-block input, multi-block output designs.

Table 1: Comparison table for permutation based PRFs. n denotes the state size of the permutation. Inv denotes whether the construction requires an inverse call of the permutation. $s \overset{\Delta}{=} n - \log(w+1)$, where $w \geq 1$, is the size of chunk in CENC based construction. The last three constructions require a keyed hash function with at most ℓ blocks input. The number of keys for those constructions includes the hash keys as well. All the constructions except CENCPP* and DS-CENCPP* requires two permutation calls. Although SoKAC1 has been shown to have a birthday bound attack and SoKAC21 is beyond the birthday bound secure [CLM19b], Chakraborti et al. [CNTY20] believed that the birthday bound attack on SoKAC1 is possibly wrong and shown an attack on it with $2^{2n/3}$ query complexity. Moreover, Nandi [Nan20] has shown a birthday bound attack on SoKAC21.

Constructions	(perm, keys)	Inv	(i/p, o/p)	Sec
SoEM1 [CLM19b]	$(1,2)$	x	(n,n)	$\Theta(n/2)$
SoEM21 [CLM19b]	$(2,1)$	x	(n,n)	$\Theta(n/2)$
SoEM22 [CLM19b]	$(2,2)$	x	(n,n)	$\Theta(2n/3)$
SoKAC1 [CLM19b] (†)	$(1,2)$	x	(n,n)	$\Omega(2n/3)$
SoKAC21 [CLM19b]	$(2,1)$	x	(n,n)	$\Theta(n/2)$
PDMMAC [CNTY20]	$(1,1)$	✓	(n,n)	$\Theta(2n/3)$
DS-SoEM [BDLN20]	$(1,2)$	x	$(n-1,n)$	$\Theta(2n/3)$
pEDM [This Paper]	$(1,2)$	x	(n,n)	$\Theta(2n/3)$
CENCPP* [BDLN20]	$(w+1,2)$	x	(n,wn)	$O(2n/3)$
DS-CENCPP* [BDLN20]	$(1,2)$	x	(s,wn)	$O(2n/3)$
nEHtM$_p$ [DN20]	$(1,2)$	x	$(n-1+\ell n,n)$	$\Theta(2n/3)$
PDM*MAC [CNTY20]	$(1,2)$	✓	$(n+\ell n,n)$	$\Theta(2n/3)$
1K-PDM*MAC [CNTY20]	$(1,1)$	✓	$(n+\ell n,n)$	$O(2n/3)$

We would like to mention here that DS-CENCPP* with $w = 1$ is a parallel construction with $n - 1$ bit input, requires field multiplication with a primitive element to derive $2n$-bit keys. However, our proposed construction is sequential with n bit input and does not require field multiplication to derive the keys. Although both of them comes with similar security bound (i.e., $2^{2n/3}$), but due to the sequential nature, pEDM requires less state size in hardware over the parallel construction DS-CENCPP*.

2   Preliminaries

BASIC NOTATIONS. For a set $\mathcal{X}$, $x \leftarrow_\$ \mathcal{X}$ denotes that x is sampled uniformly at random from $\mathcal{X}$ and is independent to all other random variables defined so far. We write $x \leftarrow y$ to denote that y is assigned in variable x. For any natural number q, $[q]$ denotes the set $\{1, \ldots, q\}$. We denote an empty set as $\emptyset$. We say two sets $\mathcal{X}$ and $\mathcal{Y}$ are disjoint if $\mathcal{X} \cap \mathcal{Y} = \emptyset$. We denote their union as $\mathcal{X} \sqcup \mathcal{Y}$ (which we refer to as *disjoint union*). Let $\mathfrak{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_s)$ be a finite collection of finite sets. We say $\mathfrak{X}$ is a *disjoint collection* if for each $j \neq j' \in [s]$, $\mathcal{X}_j$ and $\mathcal{X}_{j'}$ are disjoint. The size of $\mathfrak{X}$, denoted as $|\mathfrak{X}| = |\mathcal{X}_1| + \ldots + |\mathcal{X}_s|$. For a disjoint collection $\mathfrak{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_s, \mathcal{X}_{s+1})$, we write $\mathfrak{X} \setminus \mathcal{X}_{s+1}$ to denote the collection $(\mathcal{X}_1, \ldots, \mathcal{X}_s)$. For two disjoint collections $\mathfrak{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_s)$ and $\mathfrak{Y} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_{s'})$, we say $\mathfrak{X}$ is *inter disjoint* with $\mathfrak{Y}$ if for all $j \in [s], j' \in [s']$, $\mathcal{X}_j$ is disjoint with $\mathcal{Y}_{j'}$. If $\mathfrak{X}$ is inter disjoint with $\mathfrak{Y}$, then we denote their union as $\mathfrak{X} \sqcup \mathfrak{Y}$. Moreover, $|\mathfrak{X} \sqcup \mathfrak{Y}| = |\mathfrak{X}| + |\mathfrak{Y}|$. For a set $\mathcal{S}$ and for a finite disjoint collection of finite sets $\mathfrak{X} = (\mathcal{X}_1, \ldots, \mathcal{X}_s)$, we write $\mathcal{S} \setminus \mathfrak{X}$ to denote $\mathcal{S} \setminus (\mathcal{X}_1 \sqcup \ldots \sqcup \mathcal{X}_s)$. For a finite set $\mathcal{X} \subseteq \{0,1\}^n$ and for an arbitrary non-zero element $a \in \{0,1\}^n$, $\mathcal{X} \oplus a$ denotes the set $\{x \oplus a : x \in \mathcal{X}\}$.

For any natural number n, $\{0,1\}^n$ denotes the set of all binary strings of length n. We denote $|\{0,1\}^n|$ as $N = 2^n$ througout the paper. For integers $1 \leq b \leq a$, $(a)_b$ denotes $a(a-1)\ldots(a-b+1)$, where $(a)_0 = 1$ by convention. We denote the set of all n-bit permutations P as $\mathbb{P}(n)$. Let $\mathcal{Z}_1 = (z_1^1, \ldots, z_q^1)$ and $\mathcal{Z}_2 = (z_1^2, \ldots, z_q^2)$ be two finite tuples of length q such that for each $i \in [q]$, $z_i^1, z_i^2 \in \{0,1\}^n$. We say an n bit permutation $\mathsf{P} \in \mathbb{P}(n)$ maps $\mathcal{Z}_1$ to $\mathcal{Z}_2$, denoted as $\mathcal{Z}_1 \overset{\mathsf{P}}{\mapsto} \mathcal{Z}_2$, if for all $i \in [q]$, $\mathsf{P}(z_i^1) = z_i^2$. We say $\mathcal{Z}_1$ is *permutation compatible* to $\mathcal{Z}_2$ if there exists at least one $\mathsf{P} \in \mathbb{P}(n)$ such that $\mathcal{Z}_1 \overset{\mathsf{P}}{\mapsto} \mathcal{Z}_2$.

For a given tuple of ordered pairs $\mathcal{Q} = ((x_1, y_1), \ldots, (x_q, y_q))$, where the x_i's and the y_i's are pairwise distinct n-bit strings, we define the following two sets: $\mathsf{Dom}(\mathcal{Q}) = \{x_i \in \{0,1\}^n : (x_i, y_i) \in \mathcal{Q}\}$ and $\mathsf{Ran}(\mathcal{Q}) = \{y_i \in \{0,1\}^n : (x_i, y_i) \in \mathcal{Q}\}$. Clearly, $|\mathsf{Dom}(\mathcal{Q})| = |\mathsf{Ran}(\mathcal{Q})| = q$. We say that an n-bit permutation $\mathsf{P} \in \mathbb{P}(n)$ extends $\mathcal{Q}$, which we denote as $\mathsf{P} \mapsto \mathcal{Q}$, if for all $i \in [q], \mathsf{P}(x_i) = y_i$. We say that $\mathcal{Q}$ is *extendable* if there exists at least one $\mathsf{P} \in \mathbb{P}(n)$ such that $\mathsf{P} \mapsto \mathcal{Q}$.

We generalize this notion for more than one tuple of ordered pairs. Let $\widetilde{\mathcal{Q}} = (\mathcal{Q}_1, \ldots, \mathcal{Q}_s)$ such that for each $j \in [s]$, $\mathcal{Q}_j$ is defined as $\mathcal{Q}_j = ((x_1^j, y_1^j), \ldots, (x_{q_j}^j, y_{q_j}^j))$, where the x_i^j's and the y_i^j's are pairwise distinct n-bit strings. Now, for each $j \in [s]$, we define the following two sets: $\mathsf{Dom}(\mathcal{Q}_j) = \{x_i^j : (x_i^j, y_i^j) \in \mathcal{Q}_j\}$ and $\mathsf{Ran}(\mathcal{Q}_j) = \{y_i^j : (x_i^j, y_i^j) \in \mathcal{Q}_j\}$. Clearly, for each $j \in [s]$, $|\mathsf{Dom}(\mathcal{Q}_j)| = |\mathsf{Ran}(\mathcal{Q}_j)| = q_j$. Moreover, for all $j \neq j' \in [s]$, $\mathsf{Dom}(\mathcal{Q}_j)$ is disjoint with $\mathsf{Dom}(\mathcal{Q}_{j'})$ and $\mathsf{Ran}(\mathcal{Q}_j)$ is disjoint with $\mathsf{Ran}(\mathcal{Q}_{j'})$. Then $\mathfrak{X} = (\mathsf{Dom}(\mathcal{Q}_1), \ldots, \mathsf{Dom}(\mathcal{Q}_s))$ and $\mathfrak{Y} = (\mathsf{Ran}(\mathcal{Q}_1), \ldots, \mathsf{Ran}(\mathcal{Q}_s))$ becomes two disjoint collection of finite sets. We say that an n-bit permutation $\mathsf{P} \in \mathbb{P}(n)$ extends $\widetilde{\mathcal{Q}}$, which we denote as $\mathsf{P} \mapsto \widetilde{\mathcal{Q}}$, if for all $j \in [s], \mathsf{P} \mapsto \mathcal{Q}_j$. As an alternative notation of $\mathsf{P} \mapsto \widetilde{\mathcal{Q}}$, we also write $\mathfrak{X} \overset{\mathsf{P}}{\mapsto} \mathfrak{Y}$.

2.1   A Simple Result on Probability

In this section, we recall two simple probability results from [Dut20] that will be used while proving the security of the construction.

Proposition 1. *Let $\widetilde{Q} = (\mathcal{Q}_1, \ldots, \mathcal{Q}_{s+1})$ be an $s+1$ tuple of ordered pairs such that for $j \in [s+1]$, $\mathcal{Q}_j$ is defined as $\mathcal{Q}_j = ((x_1^j, y_1^j), \ldots, (x_{q_j}^j, y_{q_j}^j))$. Moreover, for each $j, j' \in [s+1]$, $\mathsf{Dom}(\mathcal{Q}_j) \cap \mathsf{Dom}(\mathcal{Q}_{j'}) = \emptyset$ and $\mathsf{Ran}(\mathcal{Q}_j) \cap \mathsf{Ran}(\mathcal{Q}_{j'}) = \emptyset$. Therefore, $\mathfrak{X} = (\mathsf{Dom}(\mathcal{Q}_1), \ldots, \mathsf{Dom}(\mathcal{Q}_{s+1}))$ and $\mathfrak{Y} = (\mathsf{Ran}(\mathcal{Q}_1), \ldots, \mathsf{Ran}(\mathcal{Q}_{s+1}))$ be two disjoint collection of finite sets such that for each $j \in [s+1]$, $|\mathsf{Dom}(\mathcal{Q}_j)| = |\mathsf{Ran}(\mathcal{Q}_j)| = q_j$. Then,*

we have

$$\Pr[\mathsf{P} \leftarrow_\$ \mathbb{P}(n) : \mathfrak{X} \setminus \mathsf{Dom}(\mathcal{Q}_{s+1}) \overset{\mathsf{P}}{\mapsto} \mathfrak{Y} \setminus \mathsf{Ran}(\mathcal{Q}_{s+1}) \mid \mathsf{P} \mapsto \mathcal{Q}_{s+1}] = \frac{1}{(N - q_{s+1})_{q_1 + \ldots + q_s}}.$$

By setting $s = 1$ in the above proposition gives the following simple corollary:

Corollary 1. *For two sets $\mathcal{Q}_1$ and $\mathcal{Q}_2$, where $\mathcal{Q}_1 = ((x_1^1, y_1^1), \ldots, (x_{q_1}^1, y_{q_1}^1))$ of cardinality q_1 and $\mathcal{Q}_2 = ((x_1^2, y_1^2), \ldots, (x_{q_2}^2, y_{q_2}^2))$ of cardinality q_2, such that $\mathsf{Dom}(\mathcal{Q}_1) \cap \mathsf{Dom}(\mathcal{Q}_2) = \emptyset$ and $\mathsf{Ran}(\mathcal{Q}_1) \cap \mathsf{Ran}(\mathcal{Q}_2) = \emptyset$. Then, we have*

$$\Pr[\mathsf{P} \leftarrow_\$ \mathbb{P}(n) : \mathsf{P} \mapsto \mathcal{Q}_1 \mid \mathsf{P} \mapsto \mathcal{Q}_2] = \frac{1}{(N - q_2)_{q_1}}.$$

2.2 Public Permutation Based Pseudorandom Functions

Let $\mathsf{F} : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ be a keyed function where $\mathcal{K}, \mathcal{X}$ and $\mathcal{Y}$ are the key space, input space and the output space respectively. We assume that F makes internal calls to the public random permutations $\mathbf{P} = (\mathsf{P}_1, \ldots, \mathsf{P}_d)$ for $d \geq 1$, where all of the d permutations are independent and uniformly sampled from $\mathbb{P}(n)$ for some $n \in \mathbb{N}$. Similarly, we write $\mathbf{P}^{-1} = (\mathsf{P}_1^{-1}, \ldots, \mathsf{P}_d^{-1})$ to denote the d tuple of inverse permutations. For simplicity, we write $\mathsf{F}_k^{\mathbf{P}}$ to denote F with uniform k and uniform $\mathbf{P}$.

A distinguisher D is given access to either of the oracle $\mathsf{F}_k^{\mathbf{P}}$ to denote F with uniform k and uniform $\mathbf{P}$ in the real world or a random function RF that maps elements from $\mathcal{X}$ to $\mathcal{Y}$ in the ideal world. Apart from making query to either of these two oracles, D can also make queries to the permutations $\mathbf{P}$ and $\mathbf{P}^{-1}$ in both of these worlds. Query of the former type, where the distinguisher is interacting with either $\mathsf{F}_k^{\mathbf{P}}$ or RF is called *construction query* and the query of the later type is called *primitive query*. A primitive query to the permutation is called *forward primitive query* and to the inverse of the permutation is called *inverse primitive query*. The prf advantage of D against F in the public permutation model is defined as

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{prf}}(\mathsf{D}) \triangleq \mid \Pr\left[\mathsf{D}^{(\mathsf{F}_k^{\mathbf{P}}, \mathbf{P}, \mathbf{P}^{-1})} \Rightarrow 1\right] - \Pr\left[\mathsf{D}^{(\mathsf{RF}, \mathbf{P}, \mathbf{P}^{-1})} \Rightarrow 1\right] \mid,$$

where (i) $\mathsf{D}^{\mathcal{O}} \Rightarrow 1$ denotes that the distinguisher D is given access to the oracle $\mathcal{O}$ to which it interacts with and after the interaction it outputs 1 and (ii) the above probability is defined over the randomness of $k \leftarrow_\$ \mathcal{K}$, $\mathsf{P}_1, \ldots, \mathsf{P}_d \leftarrow_\$ \mathbb{P}(n)$ and the randomness of the distinguisher (if any). We say D is a (q, p, t) distinguisher if D makes total q construction queries, p primitive queries and runs in at most t steps. We write

$$\mathbf{Adv}_{\mathsf{F}}^{\mathrm{prf}}(q, p, \mathsf{t}) \triangleq \max_{\mathsf{D}} \mathbf{Adv}_{\mathsf{F}}^{\mathrm{prf}}(\mathsf{D}),$$

where the maximum is taken over all (q, p, t)-distinguishers D. In this paper, we skip the time parameter of the distinguisher as we will assume throughout the paper that the distinguisher is computationally unbounded, and hence it is deterministic.

2.3 Sum Capture Lemma

In this section, we state a variant of the sum capture lemma [Bab02] used in [CLL+14]. Informally, the results states that when choosing a random subset $\mathcal{A}$ of $\mathrm{GF}(2^n)$ (or more generally any abelian group) of size q, the value

$$\mu(\mathcal{A}) \triangleq \max_{\mathcal{B}, \mathcal{C} \subseteq \mathrm{GF}(2^n)} |\{(a, b, c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C} : a = b \oplus c\}|,$$

is at most $q|\mathcal{B}||\mathcal{C}|/N$, except with negligible probabilty. Chen et al. [CLL+14] proved the result for a different setting where $\mathcal{A}$ arises from the interaction of an adversary with a random permutation P, namely $\mathcal{A} = x \oplus y : (x, y) \in \mathcal{Q}$, where $\mathcal{Q}$ is the transcript of the interaction between the adversary and the permutation. We employ the similar result in our setting which is stated as follows:

Lemma 1. *Let* RF *be a random function that maps elements from* $\{0,1\}^n$ *to* $\{0,1\}^n$. *Let* D *be some probabilititistic distinguisher that makes q adaptive queries to* RF. *Let* $\mathcal{Q} = ((x_1, y_1), \ldots, (x_q, y_q))$ *denotes the transcript of the interaction with* RF *to* D. *For any two subsets* $\mathcal{U}$ *and* $\mathcal{V}$ *of* $\{0,1\}^n$, *let*

$$\mu(\mathcal{Q}, \mathcal{U}, \mathcal{V}) = |\{((x,y), u, v) \in \mathcal{Q} \times \mathcal{U} \times \mathcal{V} : x \oplus u = y \oplus v\}|.$$

*Then assuming $9n \leq q \leq N/2$, we have*

$$\Pr_{\mathsf{RF},\omega} \left[\exists \mathcal{U}, \mathcal{V} \subseteq \{0,1\}^n : \mu(\mathcal{Q}, \mathcal{U}, \mathcal{V}) \geq \frac{q|\mathcal{U}||\mathcal{V}|}{N} + 3\sqrt{nq|\mathcal{U}||\mathcal{V}|} \right] \leq \frac{2}{N}, \tag{1}$$

where the probability is taken over the random choices of RF *and the random coins ω of* D.

As most part of its proof is similar to that of [CLL+14], we defer the proof of the above lemma in Supplementary Sect. 7.

3 pEDM: Permutation Based Encrypted Davis Meyer Construction

In this section, we propose pEDM, the first permutation based sequential beyond birthday bound secure pseudorandom function with two forward permutation calls. Our construction is permutation variant of the Encrypted Davis-Meyer (EDM) construction with $2n$ bit masking keys. pEDM takes an n-bit input x which is masked with an n-bit round key k_1 to generate the input of the first permutation call. Let this input be $x' = x \oplus k_1$. The permutation output $P(x')$ is then masked with $k_2 \oplus x'$ to generate the input for the second permutation call, which we denote as x'', where k_1 and k_2 are two independent n-bit round keys. Then the second permutation output $\mathsf{P}(x'')$ is masked with the round key k_1 to generate the final output y. Schematic diagram of the construction is shown in Fig. 3. In
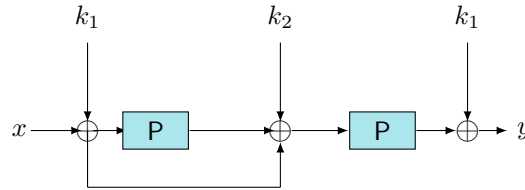
Figure 3.1: pEDM Construction with k_1 and k_2 indepedent keys and P is an n-bit permutation.

the following, we prove that pEDM is $2n/3$-bit secure in the public permutation model, where n is the state size of the permutation.

3.1 Security of pEDM

We show that pEDM is secure against all adversaries that make roughly $N^{2/3}$ construction and primitive queries in the random permutation model, where $N = 2^n$. In the following, we state the security result of pEDM, proof of which is deferred in Sect. 4.

Theorem 1. *Let* $P \leftarrow_\$ \mathbb{P}(n)$ *be an n-bit public random permutation and let* $k_1, k_2 \leftarrow_\$ \{0,1\}^n$ *be two independent n-bit keys. Then the PRF advantage for any* (q,p)-*distinguisher against the construction* $pEDM_{k_1,k_2}^P$ *that makes at most* q *construction queries and* p *primitive queries, is given by*

$$\mathbf{Adv}_{pEDM}^{prf}(q,p) \leq \frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{15q}{N^{2/3}} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{32pq^2}{N^2} + \frac{7qp^2}{N^2} + \frac{24q^3}{N^2}$$
$$+ \frac{2q^{3/2}}{N} + \frac{2p}{N^{4/3}} + \frac{3p\sqrt{nq}}{N} + \frac{2p\sqrt{q}}{N} + \frac{q\sqrt{p}}{N} + \frac{p^{3/2}}{N} + \frac{2}{N}.$$

Remarks 1.1. We would like to mention here that one can realize a construction by omitting the involvement of key k_1 in the feed-forward connection of pEDM. In other words, $P(P(x \oplus k_1) \oplus x \oplus k_2) \oplus k_1 = y$ is an another valid construction with similar level of security. Viewed in another way, our proposed construction can be viewed as a 2-round key-alternating cipher based on permutation based Davis-Meyer construction and a permutation, whereas the construction $P(P(x \oplus k_1) \oplus x \oplus k_2) \oplus k_1 = y$ can be equivalently viewed as Even-Mansour cipher based Davies-Meyer construction followed by an application of permutation. We believe that both are similar in performance and results to the similar security bound.

3.2   Matching Attack on pEDM

In this section, we show a matching key-recovery attack on pEDM with a total of $2^{2n/3+1}$ construction queries and $2^{2n/3+2}$ primitive queries. The idea of the attack is to collect a triplet of query indices $(i,j,k) \in [q] \times [p] \times [p]$ for each key k_1 in a set $\mathcal{S}_{k_1}$, where $q = 2^{2n/3+1}$ is the number of construction queries and $p = 2^{2n/3+1}$ is the number of primitive queries that distinguisher will make to the permutation P, such that $x_i \oplus k_1 = u_j^1, y_i \oplus k_1 = v_k^2$. We consider k_1 to be a potential candidate key if the number of triplets (i,j,k) in $\mathcal{S}_{k_1}$ is at least two such that

$$u_j^1 \oplus v_j^1 \oplus u_k^2 = u_{j'}^1 \oplus v_{j'}^1 \oplus u_{k'}^2$$

holds. We show that the true key belongs to the set of potential candidate keys with high probability and the size of the set of the candidate keys is not very large. We construct a deterministic adversary A that recovers the key of pEDM by making a total of $2^{2n/3+1}$ construction queries and $2^{2n/3+2}$ primitive queries as follows:

<u>NOTATION:</u> For a tuple $(x_1, x_2, \ldots, x_s)$ of length s, where each $x_i \in \{0,1\}^n$, we write $(x_1, x_2, \ldots, x_s) \xleftarrow{\text{wor}} \{0,1\}^n$ to denote that $x_i \leftarrow_\$ \{0,1\}^n \setminus \{x_1, \ldots, x_{i-1}\}$ for $i \geq 2$ with $x_1 \leftarrow_\$ \{0,1\}^n$. Similarly, $(x_1, x_2, \ldots, x_s) \xleftarrow{\text{wr}} \{0,1\}^n$ denotes that $x_i \leftarrow_\$ \{0,1\}^n$, independent to all $x_1, \ldots, x_{i-1}$ for $i \geq 2$ with $x_1 \leftarrow_\$ \{0,1\}^n$.

<u>ATTACK ALGORITHM:</u>

1. A chooses $2^{2n/3+1}$ construction queries $(x_1, \ldots, x_{2^{2n/3+1}}) \xleftarrow{\text{wor}} \{0,1\}^n$.

2. A chooses $2^{2n/3+2}$ forward primitive queries $(u_1, \ldots, u_{2^{2n/3+2}}) \xleftarrow{\text{wor}} \{0,1\}^n$.

3. A creates two lists: $\mathcal{U}_1$ that stores u_j for $j \in [2^{2n/3+1}]$ and another list $\mathcal{U}_2$ that stores u_j for $j \in [2^{2n/3+1}+1, 2^{2n/3+2}]$. We denote the elements of $\mathcal{U}_1$ as u_j^1, i.e., $u_j^1 \leftarrow u_j$ for $1 \leq j \leq 2^{2n/3+1}$ and denote the elements of $\mathcal{U}_2$ as u_k^2, i.e., $u_k^2 \leftarrow u_{k+2^{2n/3+1}}$ for $1 \leq k \leq 2^{2n/3+1}$.

4. Then A makes queries to the primitive P with u_j^1 for $j \in [2^{2n/3+1}]$ and obtains the responses $v_j^1 \leftarrow P(u_j^1)$. A makes another set of queries to the primitive P with u_k^2 for $k \in [2^{2n/3+1}+1, 2^{2n/3+2}]$ and obtains the corresponding responses $v_k^2 \leftarrow P(u_k^2)$.

5. For each k_1, A construct a set $\mathcal{S}_{k_1} = \{(i, j, k) \in [2^{2n/3}+1] \times [2^{2n/3}+1] \times [2^{2n/3}+1] : x_i \oplus u_j^1 = k_1 = y_i \oplus v_k^2\}$. We initialize $\mathcal{K}$ to the empty set $\emptyset$, which we call the set of *candidate keys.*

6. For all key $k_1 \in \{0,1\}^n$ with $|\mathcal{S}_{k_1}| \geq 2$, check if the following holds: for every pair $(i, j, k) \neq (i', j', k') \in \mathcal{S}_{k_1}$, one has

$$u_j^1 \oplus v_j \oplus u_k^2 \oplus u_{j'}^1 \oplus v_{j'}^1 \oplus u_{k'}^2 = 0. \tag{2}$$

If the above holds, then add $k_1 \in \mathcal{K}$.

Claim 1. *Let (k_1^*, k_2^*) be the true key, i.e., the pair of keys used in the construction. Then, we have*

$$\Pr[k_1^* \in \mathcal{K}] \geq 0.687 \tag{3}$$
$$\Pr[|\mathcal{K} \setminus \{k_1^*\}| \geq 128] \leq 0.5. \tag{4}$$

We defer the proof of the claim in the following section. However, the first equation of the claim says that the true key k_1^* belongs to the set of candidate keys with high probability, and the second equation says that the probability of the number of candidate keys is at least 128 is at most $1/2$. Before proceeding with the analysis of the attack, we recall the Chernoff-bound for the sum of independent Bernoulli trial as follows:

Lemma 2. *Let $X_1, X_2, \ldots, X_n$ be independent random variables following the bernoulli distribution such that X_i takes the value 1 with probability p_i and 0 with probability $(1-p_i)$. Let $X = X_1 + X_2 + \ldots + X_n$ and $\mu = \mathbf{E}[X]$. Then, for any $0 < \delta < 1$,*

$$\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}.$$

3.3   Analysis of the Key-Recovery Advantage

In this section, we prove Claim 1. In particular, we carry out the above two probability analysis of Claim 1 in the following two steps:

Step I: True key belongs to the set of candidate keys. According to step (6) of the algorithm, an element k_1 gets included in the set $\mathcal{K}$ if the following two conditions hold:

$$(a) \ |\mathcal{S}_{k_1}| \geq 2, \quad (b) \ u_j^1 \oplus v_j^1 \oplus u_k^2 = u_{j'}^1 \oplus v_{j'}^1 \oplus u_{k'}^2, \quad (i, j, k), (i', j', k') \in \mathcal{S}_{k_1},$$

where $\mathcal{S}_{k_1}$ is the set of all triplets $(i, j, k) \in ([2^{2n/3+1}])^3$, as defined in step (5) of the algorithm such that the following holds:

$$(\S) = \begin{cases} k_1 = x_i \oplus u_j^1 \\ k_1 = y_i \oplus v_k^2 \end{cases}$$

For the true key (k_1^*, k_2^*), Let z_i be the input variable of the second permutation call of the construction, i.e.,

$$z_i \stackrel{\Delta}{=} \mathsf{P}(x_i \oplus k_1^*) \oplus (x_i \oplus k_1^* \oplus k_2^*).$$

Note that all the x_i's are without replacement variables and so are $x_i \oplus k_1^*$. Moreover, the variables $\mathsf{P}(x_i \oplus k_1^*)$ are again sampled in without replacement manner and indepenent to variables $x_i \oplus k_1^*$. Therefore, each z_i is a sum of two independently sampled without replacement random variables and due to the result of the *sum of two independent permutations* [DNS20, DHT17], distribution of all z_i's are uniform.

Now note that, for the first part k_1^* of the true key pair (k_1^*, k_2^*), if it happens that for some $(i, j, k) \in ([2^{2n/3+1}])^3$, (§) holds, where k_1 in (§) is replaced by the true key k_1^*, then one can reveal the second part k_2^* of the true key pair (k_1^*, k_2^*) as $k_2^* = u_j^1 \oplus v_j^1 \oplus u_k^2$. As a result, for the true key k_1^* and for $(i, j, k), (i', j', k') \in \mathcal{S}_{k_1^*}$, the relation

$$u_j^1 \oplus v_j^1 \oplus u_k^2 = u_{j'}^1 \oplus v_{j'}^1 \oplus u_{k'}^2$$

gets automatically satisfied. Therefore, to bound Eqn. (3), it is enough to bound the probability that there exists at least two distinct tuples $(i, j, k), (i', j', k)$ exists such that

$$(\bigstar) = \begin{cases} k_1^* = x_i \oplus u_j^1 = y_i \oplus v_k^2 \\ k_1^* = x_{i'} \oplus u_{j'}^1 = y_{i'} \oplus v_{k'}^2. \end{cases}$$

Again, for the first part k_1^* of the true key pair (k_1^*, k_2^*), if it happens that for some $(i, j, k), (i', j', k') \in ([2^{2n/3+1}])^3$, the following equations are satisfied, namely

$$(\dagger) = \begin{cases} k_1^* = x_i \oplus u_j^1 \\ k_1^* = x_{i'} \oplus u_{j'}^1 \\ z_i = u_k^2 \\ z_{i'} = u_{k'}^2, \end{cases}$$

then it also satisfies $(\bigstar)$. As a result, it is enough to bound the probability that there exists at least two distinct tuples $(i, j, k), (i', j', k')$ such that $(\dagger)$ is satisfied. We bound the probability in two stages. In the first stage, we bound the number of i such that $z_i \in \mathcal{U}_2$ and we store such i in list $\mathcal{L}_1$. Let $\mathcal{L}_x$ be the set of all $x_i \oplus k_1^*$ such that $i \in \mathcal{L}_1$. In the second stage, we lower bound the probability that the number of j such that $u_j^1 \in \mathcal{L}_x$ is at least 2.

<u>STAGE-I:</u> Let $\mathbb{Z}_i$ be the indicator random variable that takes the value 1 if $z_i \in \mathcal{U}_2$. It is easy to see that $\mathbb{Z}_i$ are indepedent bernoulli random variables with success probability $2/2^{n/3}$. Let $Z = (\mathbb{Z}_1 + \ldots + \mathbb{Z}_{2^{2n/3+1}})$. Then $Z \sim \mathsf{Bin}(2^{2n/3+1}, 2/2^{n/3})$ and therefore, $\mathbf{E}[Z] = 4 \cdot 2^{n/3}$. By applying the Chernoff-bound as stated in Lemma (2) with $\delta = 1/2$, we have

$$\Pr[Z > 2^{n/3+1}] \geq 1 - \frac{1}{e^{2^{n/3-1}}}. \tag{5}$$

Therefore, Eqn. (5) says that the size of list $\mathcal{L}_1$ and in turn the size of list $\mathcal{L}_x$ is at least $2^{n/3} + 1$ holds with high probability.

<u>STAGE-II:</u> Let $\mathcal{L}_x$ be the list of all $x_i \oplus k_1^*$ such that $i \in \mathcal{L}_1$. Therefore, to bound the probability that there exists at least two distinct tuples $(i, j, k), (i', j', k)$ exists such that such that $(\bigstar)$ holds, we bound the following:

$$\Pr[|j : u_j^1 \in \mathcal{L}_x| \geq 2]. \tag{6}$$

We write Eqn. (6) as

$$(6) = 1 - \left(\underbrace{\Pr[u_j^1 \notin \mathcal{L}_x, \forall j \in [2^{2n/3+1}]]}_{(A)} + \underbrace{\sum_{j=1}^{2^{2n/3+1}} \Pr[u_j^1 \in \mathcal{L}_x \wedge u_k^1 \notin \mathcal{L}_x, \forall k \neq j]}_{(B)} \right). \tag{7}$$

<u>BOUNDING A:</u> To bound A, we would like to note here that $u_1^1, \ldots, u_{2^{2n/3+1}}^1$ are without replacement samples of $\{0, 1\}^n$. Moreover, $|\mathcal{L}_x| = 2^{n/3+1}$. By using a simple algebra, we have

$$(A) = \Pr[u_1^1, \ldots, u_{2^{2n/3}+1}^1 \notin \mathcal{L}_x] \leq \frac{(2^n - 2^{n/3+1})_{2^{2n/3+1}}}{(2^n)_{2^{2n/3+1}}} \leq \left(1 - \frac{2}{2^{2n/3}} \right)^{2 \cdot 2^{2n/3}} \leq \frac{1}{e^4}. \tag{8}$$

BOUNDING B: Bounding B is similar to that of A.

$$(B) = \sum_{j=1}^{2^{2n/3+1}} \Pr[u_j^1 \in \mathcal{L}_x] \cdot \Pr[u_k^1 \notin \mathcal{L}_x, \forall k \neq j] \leq \sum_{j=1}^{2^{2n/3+1}} \frac{2^{n/3+1}}{(2^n - 2^{2n/3+1} + 1)} \cdot \left(1 - \frac{2}{2^{2n/3}}\right)^{2 \cdot 2^{2n/3} - 1}$$

$$\leq 8\left(1 - \frac{2}{2^{2n/3}}\right)^{-1} \cdot \left(1 - \frac{2}{2^{2n/3}}\right)^{2 \cdot 2^{2n/3}} \leq \frac{16}{e^4}, \tag{9}$$

where the last inequality follows from $\frac{1}{(1-2/2^{2n/3})} \leq 2$ as $n \geq 3$. Therefore, from Eqn. (7), Eqn. (8), Eqn. (9) and by plug-in the value of $e \leq 3$, we have

$$\Pr[k_1^* \in \mathcal{K}] \geq 1 - \frac{17}{e^4}. \tag{10}$$

STEP-II: BOUNDING THE CARDINALITY OF $\mathcal{K} \setminus \{k_1^*\}$. To upper bound the probability that $|\mathcal{K} \setminus \{k_1^*\}| \geq 128$, we use the Markov's inequality. In particular, we have

$$\Pr[|\mathcal{K} \setminus \{k_1^*\}| \geq 128] \leq \frac{\mathbf{E}[|\mathcal{K} \setminus \{k_1^*\}|]}{128}. \tag{11}$$

Therefore, it is enough to upper bound the expected size of the set of candidate keys $\mathcal{K} \setminus \{k_1^*\}$. For each $k_1 \in \{0,1\}^n$, let $\mathbb{I}_{k_1}$ be the indicator random variable that takes the value 1 if there exists $(i, j, k), (i', j', k')$ such that the following holds:

$$(\S\S) = \begin{cases} u_j^1 \oplus v_j^1 \oplus u_k^2 = u_{j'}^1 \oplus v_{j'}^1 \oplus u_{k'}^2 \\ k_1 = x_i \oplus u_j^1 \\ k_1 = y_i \oplus v_k^2 \\ k_1 = x_{i'} \oplus u_{j'}^1 \\ k_1 = y_{i'} \oplus v_{k'}^2. \end{cases}$$

Otherwise, the indicator random variable $\mathbb{I}_{k_1}$ takes the value 0. It is easy to see using the linearity of expectation that

$$\sum_{k_1 \in \{0,1\}^n \setminus \{k_1^*\}} \mathbb{I}_{k_1} = |\mathcal{K} \setminus \{k_1^*\}| \Rightarrow \mathbf{E}[|\mathcal{K} \setminus \{k_1^*\}|] = \sum_{k_1 \in \{0,1\}^n \setminus \{k_1^*\}} \Pr[\mathbb{I}_{k_1} = 1]. \tag{12}$$

Therefore, it boils down to upper bound the probability that $\mathbb{I}_{k_1}$ takes the value 1. For a fixed choice of indices i, j, k and i', j', k', the above system of equations hold with probability at most 2^{-5n} as all the random variables are independent to each other. The number of choices of indices is at most $(2^{n/3+1})^6$. Therefore, we have

$$\Pr[\mathbb{I}_{k_1} = 1] \leq \frac{64}{2^n}. \tag{13}$$

From Eqn. (12) and Eqn. (13), we have the expected size of the set of candidate keys is at most 8. By plug-in this value into Eqn. (11), we have

$$\Pr[|\mathcal{K} \setminus \{k_1^*\}| \geq 128] \leq 1/2,$$

which concludes the proof of Claim 1.   □

Note that in the above attack, the distinguisher is information theoretically bounded. The run time of the attack [4] is more than 2^n. In particular, for each key k_1, the number of

---

[4]Note that the time complexity of the adversary here does not account for the number of times adversary makes offline primitive queries. The time complexity of the adversary solely means the time required to compute local operations.

steps required to populate set $\mathcal{S}_{k_1}$ is roughly 2^{2n}. Therefore, altogether step (5) of the algorithm takes at most 2^{4n} operations. For each key k_1 in step (6), algorithm takes at least one checking operation for Eqn. (2) in each set $\mathcal{S}_{k_1}$. Therefore, altogether step (5) of the algorithm takes 2^n operations. Therefore, the overall time complexity of the algorithm is roughly $O(2^{4n})$. Nevertheless, the number of construction queries is $2^{2n/3+1}$, and the total number of primitive queries is $2^{2n/3+2}$.

4 Proof of Theorem 1

Let us consider $\mathbf{k} = (k_1, k_2) \in \{0,1\}^{2n}$ be a pair of n-bit keys. We consider any information theoretic deterministic distinghisher D that interacts with the following oracles in either the real world or in the ideal world: in the real world it interacts with $(\mathsf{pEDM}^{\mathsf{P}}_{\mathbf{k}}, \mathsf{P})$ and in the ideal world it interacts with $(\mathsf{RF}, \mathsf{P})$, where RF is the random function over $\{0,1\}^n$ to $\{0,1\}^n$. We call the first oracle as *construction oracle* and the second one as *primitive oracle*. Query to the construction oracle is called the *construction query* and to that of the primitive oracle is called the *primitive query*. We summarize the construction queries in a transcript τ_c, where $\tau_c = \{(x_1, y_1), \ldots, (x_q, y_q)\}$ and the primitives queries in transcript $\tau_p = \{(u_1, v_1), \ldots, (u_p, v_p)\}$, where we assume that D makes total q construction and p primitive queries. For primitive queries, D can either make forward query u to its primitive P and receives response v or can make inverse query v to P^{-1} and receives response u. Since, we assume that D never makes pointless queries, none of the transcripts contain any duplicate elements.

We modify the experiment by releasing internal information to D after it has finished the interaction but has not output yet the decision bit. In the real world, we reveal the key $\mathbf{k}$ which is used in the construction and in the ideal world, we sample a pair of n-bit dummy keys $\mathbf{k} = (k_1, k_2)$ uniformly at random from the keyspace $\{0,1\}^n$ and reveal it to the distinguisher. In all the following, the complete transcript is $\tau = (\tau_c, \tau_p, \mathbf{k})$. Note that, the modified experiment only makes the distinguisher more powerful and hence the distinguishing advantage of D in this experiment is no way less than its distinguishing advantage in the former one. Let X_{re} denotes the random variable that takes a transcript τ realized in the real world. Similarly, X_{id} denotes the random variable that takes a transcript τ realized in the ideal world. The probability of realizing a transcript $\tau = (\tau_c, \tau_p, \mathbf{k})$ in the ideal (resp. real) world is called *ideal (resp. real) interpolation probability*. A transcript τ is said to be attainable with respect to D if its ideal interpolation probability is non-zero. Let Θ denotes the set of all attainable transcripts and $\phi : \Theta \to [0, \infty)$ be a non-negative function that maps any attainable transcripts to a non-negative real value. Following these notations, we state the main theorem of the Expectation Method [HT16] as follows:

Theorem 2 (Expectation Method). *Let* $\Theta = \mathsf{GoodT} \sqcup \mathsf{BadT}$ *be some partition of the set of attainable transcripts. Let* $\tau = (\tau_c, \tau_p, \mathbf{k}) \in \mathsf{GoodT}$ *be an arbitrary good transcript such that*

$$\frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} \triangleq \frac{\Pr[\mathsf{X}_{\mathrm{re}} = \tau]}{\Pr[\mathsf{X}_{\mathrm{id}} = \tau]} \geq 1 - \phi(\tau),$$

and there exists $\epsilon_{\mathrm{bad}} \geq 0$ *such that* $\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \leq \epsilon_{\mathrm{bad}}$. *Then,*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{pEDM}}(\mathsf{D}) \leq \mathbf{E}[\phi(\mathsf{X}_{\mathrm{id}})] + \epsilon_{\mathrm{bad}}. \tag{14}$$

Note that, the expectation method trivially boils down to the H-Coefficient technique if ϕ becomes a constant function such that for any attainable good transcripts τ, $\phi(\tau) = c$ for $0 \leq c \leq 1$. Having explained the Expectation Method in the view of our construction, we now state the following result.

Lemma 3. *Let $\tau = (\tau_c, \tau_p, \mathbf{k}) \in \Theta$ be an attainable transcript. Let $\mathsf{p}(\tau) \stackrel{\Delta}{=} \Pr[\mathsf{P} \leftarrow_\$ \mathbb{P}(n) : \mathsf{pEDM}^\mathsf{P}_\mathbf{k} \mapsto \tau_c \mid \mathsf{P} \mapsto \tau_p]$. Then, we have*

$$\rho(\tau) \stackrel{\Delta}{=} \frac{\mathsf{p}_{\mathrm{re}}(\tau)}{\mathsf{p}_{\mathrm{id}}(\tau)} = \mathsf{p}(\tau) \cdot N^q.$$

Recall that $\mathsf{pEDM}^\mathsf{P}_\mathbf{k} \mapsto \tau_c$ denotes $\mathsf{pEDM}^\mathsf{P}_\mathbf{k}(x_i) = y_i$ for all $(x_i, y_i) \in \tau_c$, i.e., for all $(x_i, y_i) \in \tau_c$, it must hold that $\mathsf{P}(\mathsf{P}(x_i \oplus k_1) \oplus x_i \oplus k_1 \oplus k_2) \oplus k_1 = y_i$, where $\mathbf{k} = (k_1, k_2)$. Proof of this lemma is trivial to follow as the ideal interpolation probability for a good transcript is $\frac{1}{(N)_p N^q}$, as the random function RF always outputs uniform random n-bit strings on each input query.

4.1   Definition and Probability of Bad Transcripts

In this section, we define and bound the probability of bad transcripts in the ideal world. For a transcript $\tau = (\tau_c, \tau_p, k_1, k_2)$, we define the following sets:

$$\mathcal{U} \stackrel{\Delta}{=} \{u \in \{0,1\}^n : (u,v) \in \tau_p\},$$

$$\mathcal{V} \stackrel{\Delta}{=} \{v \in \{0,1\}^n : (u,v) \in \tau_p\},$$

$$\alpha \stackrel{\Delta}{=} |\{(x,y) \in \tau_c : x \oplus k_1 \in \mathcal{U}\}|,$$

$$\beta \stackrel{\Delta}{=} |\{(x,y) \in \tau_c : y \oplus k_1 \in \mathcal{V}\}|,$$

$$\mathsf{C} \stackrel{\Delta}{=} |\{\{(x,y),(x',y')\} : (x,y),(x',y') \in \tau_c, y = y'\}|,$$

$$\sigma \stackrel{\Delta}{=} |\{\{(x,y),(x',y'),(x'',y'')\} : (x,y),(x',y'),(x'',y'') \in \tau_c, y \oplus x' \oplus k_2 = x'' \oplus k_1\}|,$$

$$\theta \stackrel{\Delta}{=} |\{\{(u,v),(u',v')\} : (u,v),(u',v') \in \tau_p, u \oplus v = u' \oplus v'\}|.$$

We say that a construction query $(x,y) \in \tau_c$ is *non-colliding* if $\forall (x',y') \in \tau_c, y \neq y'$. Now, we characterize the set of bad transcripts as follows. The main crux of identifying bad events is to identify the two-fold collisions, as depicted in Fig. 4.1

Definition 1. An attainable transcript $\tau = (\tau_c, \tau_p, \mathbf{k})$ is called a **bad** transcript if any one of the following holds:

1. *Inputs (resp. outputs) to the two consecutive permutation calls for a particular construction query are not fresh.*

   - B.1: $\exists (x,y) \in \tau_c, (u,v), (u',v') \in \tau_p$ such that $x \oplus k_1 = u, v \oplus u \oplus k_2 = u'$.
   - B.2: $\exists (x,y) \in \tau_c, (u,v), (u',v') \in \tau_p$ such that $y \oplus k_1 = v, u \oplus (x \oplus k_1) \oplus k_2 = v'$.
   - B.3: $\exists (x,y), (x',y') \in \tau_c, (u,v) \in \tau_p$ such that $x \oplus k_1 = u, v \oplus u \oplus k_2 = x' \oplus k_1$.
   - B.4: $\exists (x,y), (x',y') \in \tau_c, (u,v) \in \tau_p$ such that $y \oplus k_1 = v, u \oplus (x \oplus k_1) \oplus k_2 = y' \oplus k_1$.

2. *Both the input and output of a construction query are not fresh.*

   - B.5: $\exists (x,y) \in \tau_c, (u,v), (u',v') \in \tau_p$ such that $x \oplus k_1 = u, y \oplus k_1 = v'$.

3. *Inputs (resp. outputs) to the first (resp. second) permutation call of two construction queries collides with the input (resp. output) of two primitive queries, and the inputs (resp. outputs) to the second (resp. first) permutation call for those two construction queries collides.*

   - B.6: $\exists (x,y), (x',y') \in \tau_c, (u,v), (u',v') \in \tau_p$ such that $x \oplus k_1 = u, x' \oplus k_1 = u', u \oplus v = u' \oplus v'$.
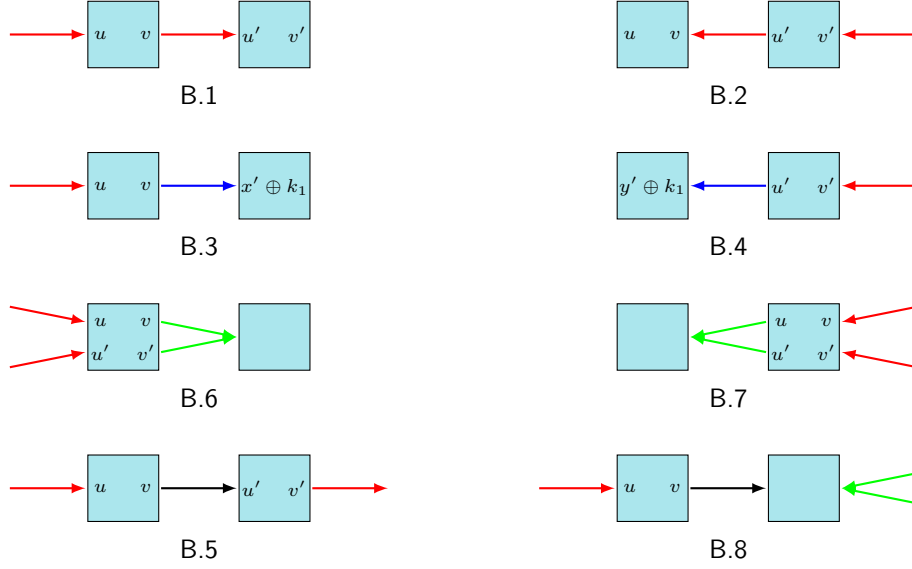
Figure 4.1: Different cases of two-fold collisions. Red edge denotes the input / output collides with prmitive input / output. Blue edge denotes that input collides with the input of some construction query or output of some construction query. Green edge denotes the collision among themselves.

- B.7: $\exists\, (x,y),(x',y') \in \tau_c, (u,v),(u',v') \in \tau_p$ such that $y \oplus k_1 = v, y' \oplus k_1 = v', u \oplus x = u' \oplus x'$.

4. *Additional Bad events.*

- B.8: $\exists\, (x,y),(x',y') \in \tau_c, (u,v) \in \tau_p$ such that $x \oplus k_1 = u, y = y'$.
- B.9: $\sigma \geq q^2/N^{1/3}$.
- B.10: $\mathsf{C} \geq q/N^{1/3}$.
- B.11: $\alpha \geq \sqrt{q}$.
- B.12: $\beta \geq \sqrt{q}$.
- B.13: $\theta \geq \sqrt{p}$.

Recall that $\mathsf{BadT} \subseteq \Theta$ be the set of all attainable bad transcripts and $\mathsf{GoodT} = \Theta \setminus \mathsf{BadT}$ be the set of all attainable good transcripts. We bound the probability of bad transcripts in the ideal world as follows.

Lemma 4. *Let $\tau = (\tau_c, \tau_p, \mathbf{k})$ be any attainable transcript. Let X_{id} and Θ_{b} be defined as above. Then*

$$\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \;\; \leq \;\; \frac{3qp^2}{N^2} + \frac{4pq^2}{N^2} + \frac{3p\sqrt{nq}}{N} + \frac{2p\sqrt{q}}{N} + \frac{q\sqrt{p}}{N} + \frac{p^{3/2}}{N} + \frac{2q}{N^{2/3}} + \frac{2}{N}.$$

Proof. Let $\tau = (\tau_c, \tau_p, k_1, k_2)$ be any attainable transcript. Recall that, in the ideal world k_1 and k_2 are sampled uniformly and independently from the keyspace. Using the union bound, we have

$$\Pr[\mathsf{X}_{\mathrm{id}} \in \mathsf{BadT}] \leq \Pr[\mathsf{B.7} \vee \mathsf{B.13}] + \sum_{\substack{1 \leq i \leq 13 \\ i \neq 7,13}} \Pr[\mathsf{B.i}]. \tag{15}$$

In the following, we bound the probabilities of all the bad events individually. The lemma will then follow by adding the individual bounds.

BOUNDING B.1. We consider the event B.1. For a fixed $(x, y) \in \tau_c$ and for a fixed $(u, v), (u', v') \in \tau_p$, the probability that

$$k_1 = x \oplus u, k_2 = u' \oplus v \oplus u$$

is N^{-2} due to the randomness of the key k_1 and k_2. By summing over all possible choices of $(x, y) \in \tau_c, (u, v), (u', v') \in \tau_p$, we have

$$\Pr[\mathsf{B.1}] \leq \frac{qp^2}{N^2}. \tag{16}$$

BOUNDING B.2. We consider the event B.2. For a fixed $(x, y) \in \tau_c$ and for a fixed $(u, v), (u', v') \in \tau_p$, the probability that

$$k_1 = y \oplus v, k_2 = v' \oplus u \oplus (x \oplus k_1)$$

is N^{-2} by using the randomness of k_1 and k_2. By summing over all possible choices of $(x, y) \in \tau_c, (u, v), (u', v') \in \tau_p$, we have

$$\Pr[\mathsf{B.2}] \leq \frac{qp^2}{N^2}. \tag{17}$$

BOUNDING B.3. We consider the event B.3. For a fixed $(x, y), (x', y') \in \tau_c$ and for a fixed $(u, v) \in \tau_p$, the probability that

$$k_1 = u \oplus x, k_2 = v \oplus (x' \oplus k_1) \oplus u$$

is N^{-2} by using the randomness of k_1 and k_2. By summing over all possible choices of $(x, y), (x', y') \in \tau_c, (u, v) \in \tau_p$, we have

$$\Pr[\mathsf{B.3}] \leq \frac{pq^2}{N^2}. \tag{18}$$

BOUNDING B.4. Using the similar reasoning as that of B.3, we have

$$\Pr[\mathsf{B.4}] \leq \frac{pq^2}{N^2}. \tag{19}$$

BOUNDING B.5. We consider the event B.5. To bound the event we consider the following set

$$\mathsf{BadK}_1 = \{k_1 \in \{0,1\}^n : \exists (x, y) \in \tau_c, (u, v), (u', v') \in \tau_p \text{ such that } k_1 = x \oplus u = y \oplus v'\}.$$

Note that $k \in \mathsf{BadK}_1 \Leftrightarrow \exists (x, y) \in \tau_c, (u, v), (u', v') \in \tau_p : k = x \oplus u = y \oplus v'$. Therefore, for any $\Delta > 0$, we have,

$$
\begin{aligned}
\Pr[\mathsf{B.5}] &= \Pr[k_1 \in \mathsf{BadK}_1] \\
&= \Pr[k_1 \in \mathsf{BadK}_1 \wedge |\mathsf{BadK}_1| \geq \Delta] + \Pr[k_1 \in \mathsf{BadK}_1 \wedge |\mathsf{BadK}_1| < \Delta] \\
&\leq \Pr[|\mathsf{BadK}_1| \geq \Delta]] + \frac{\Delta}{N}. \tag{20}
\end{aligned}
$$

Now, it is easy to see that

$$|\mathsf{BadK}_1| \leq \mathcal{Z} \stackrel{\Delta}{=} |\{((x, y), u, v') \in \tau_c \times \mathcal{U} \times \mathcal{V} : x \oplus u = y \oplus v'\}|.$$

Now, from Lemma 1, we have

$$\Pr[|\mathcal{Z}| \geq qp^2/N + 3p\sqrt{nq}] \leq 2/N.$$

Therefore, by setting $\Delta = qp^2/N + 3p\sqrt{nq}$ and by using Eqn. (20), we have

$$\Pr[\mathsf{B.5}] \leq \frac{qp^2}{N^2} + \frac{3p\sqrt{nq}}{N} + \frac{2}{N}. \tag{21}$$

<u>Bounding B.6 ∨ B.13.</u> We bound the event B.6 ∨ B.13. For this, we have the following:

$$\Pr[\mathsf{B.6} \vee \mathsf{B.13}] \leq \Pr[\mathsf{B.13}] + \Pr[\mathsf{B.6} \wedge \overline{\mathsf{B.13}}]. \tag{22}$$

To bound the probability of the event B.13, we define an indicator random variable $\mathbb{I}_{ij}$ which is set to 1 if and only if $(u_i, v_i), (u_j, v_j) \in \tau_p$ such that $u_i \oplus v_i = u_j \oplus v_j$. Therefore, we have

$$\theta = \sum_{i,j} \mathbb{I}_{ij}.$$

Now, for a fixed i, j, we have $\Pr[\mathbb{I}_{ij} = 1] = N^{-1}$. This is due to the fact that either both of $(u_i, v_i), (u_j, v_j)$ are backward queries in which u_i, u_j are random values or at least one of them is a forward query (w.l.og we assume (u_j, v_j) is a forward query) in which v_j is random. Hence, using the linearity of expectation, we have

$$\mathbf{E}[\theta] \quad = \quad \sum_{ij} \mathbf{E}[\mathbb{I}_{ij}] = \sum_{ij} \Pr[\mathbb{I}_{ij} = 1] \leq \frac{p^2}{N}. \tag{23}$$

Therefore, using Markov's inequality, we have

$$\Pr[\mathsf{B.13}] = \Pr[\theta \geq \sqrt{p}] \leq \frac{\mathbf{E}[\theta]}{\sqrt{p}} \overset{(1)}{\leq} \frac{p^{3/2}}{N}, \tag{24}$$

where (1) follows from Eqn. (23). Now, we bound the probability of B.6 ∧ $\overline{\mathsf{B.13}}$. To bound the event, for a fixed pair of $(x, y), (x', y') \in \tau_c$ and for a fixed pair of $(u, v), (u', v') \in \tau_p$ the probability that $x \oplus k_1 = u, x' \oplus k_1 = u', u \oplus v = u' \oplus v'$ is N^{-1} due to the randomness of k_1. Note that as $u \neq u'$, the probability of the event is well defined. <u>Since, we consider the probability of the conditional event B.6 conditioned on the event $\overline{\mathsf{B.13}}$, the number of such pairs of $(u, v), (u', v') \in \tau_p$ satisfies the the event is at most $\sqrt{p}$.</u> Moreover, the number of choices for $(x, y) \in \tau_c$ is q which makes the choice for $(x', y') \in \tau_c$ is at most 1, as choosing an (x, y) determines (x', y'), namely, $x' = u' \oplus u \oplus x$. Hence,

$$\Pr[\mathsf{B.6} \wedge \overline{\mathsf{B.13}}] \leq \frac{q\sqrt{p}}{N}. \tag{25}$$

By combining Eqn. (22), Eqn. (24) and Eqn. (25), we have

$$\Pr[\mathsf{B.6} \vee \mathsf{B.13}] \leq \frac{q\sqrt{p}}{N} + \frac{p^{3/2}}{N}. \tag{26}$$

<u>Bounding B.7.</u> To bound the event B.7 we need to bound the event $y = v \oplus k_1, y' = v' \oplus k_1, u \oplus x = u' \oplus x'$. For a fixed pair of $(x, y), (x', y') \in \tau_c$ and $(u, v), (u', v') \in \tau_p$, the above event holds with probability N^{-2} due to the independence of y and y'. Now, the number of choice for $(x, y), (x', y')$ is at most q^2 and the number of choice for (u, v) is at most p which makes the number of choice for (u', v') is at most 1. Hence, by varying over all possible choices of $(x, y), (x', y') \in \tau_c$ and $(u, v), (u', v') \in \tau_p$, we have

$$\Pr[\mathsf{B.7}] \leq \frac{pq^2}{N^2}. \tag{27}$$

<u>Bounding B.8.</u> To bound the event B.8, we fix $(x, y), (x', y') \in \tau_c$ and $(u, v) \in \tau_p$, the probability that

$$k_1 = u \oplus x, y = y'$$

is N^{-2}. By summing over all possible choices of $(x, y), (x', y') \in \tau_c, (u, v) \in \tau_p$, we have

$$\Pr[\mathsf{B.8}] \leq \frac{pq^2}{N^2}. \tag{28}$$

<u>Bounding B.9.</u> To bound the event B.9, we define an indicator random variable $\mathbb{I}_{ijk}$ which is set to 1 if and only if $(x_i, y_i), (x_j, y_j), (x_k, y_k) \in \tau_c$ such that $y_i \oplus x_j \oplus k_2 = x_k \oplus k_1$. Therefore, we have

$$\sigma = \sum_{i,j,k} \mathbb{I}_{ijk}.$$

Now, for a fixed i, j and k, we have $\Pr[\mathbb{I}_{ijk} = 1] = N^{-1}$ by using the randomness of k_1. Using the linearity of expectation, we have

$$\mathbf{E}[\sigma] \quad = \quad \sum_{ijk} \mathbf{E}[\mathbb{I}_{ijk}] = \sum_{ijk} \Pr[\mathbb{I}_{ijk} = 1] \leq \frac{q^3}{N}. \tag{29}$$

Therefore, using Markov's inequality, we have

$$\Pr[\mathsf{B.9}] = \Pr[\sigma \geq q^2/N^{1/3}] \leq \frac{\mathbf{E}[\sigma]}{q^2/N^{1/3}} \overset{(2)}{\leq} \frac{q}{N^{2/3}}, \tag{30}$$

where (2) follows from Eqn. (29).
<u>Bounding B.10.</u> To bound the event B.10, we define an indicator random variable $\mathbb{I}_{ij}$ which is set to 1 if and only if $(x_i, y_i), (x_j, y_j) \in \tau_c$ such that $y_i = y_j$. Therefore, we have

$$\mathsf{C} = \sum_{i,j} \mathbb{I}_{ij}.$$

Now, for a fixed i, j, we have $\Pr[\mathbb{I}_{ij} = 1] = N^{-1}$ by using the independence of y_i and y_j. Using the linearity of expectation, we have

$$\mathbf{E}[\mathsf{C}] \quad = \quad \sum_{ij} \mathbf{E}[\mathbb{I}_{ij}] = \sum_{ij} \Pr[\mathbb{I}_{ij} = 1] \leq \frac{q^2}{N}. \tag{31}$$

Therefore,

$$\Pr[\mathsf{B.10}] = \Pr[\mathsf{C} \geq q/N^{1/3}] \leq \frac{\mathbf{E}[\mathsf{C}]}{q/N^{1/3}} \overset{(3)}{\leq} \frac{q}{N^{2/3}}, \tag{32}$$

where (3) follows from Eqn. (31).
<u>Bounding B.11.</u> To bound the event B.11, we define an indicator random variable $\mathbb{I}_{ij}$ which is set to 1 if and only if $(x_i, y_i) \in \tau_c, (u_j, v_j) \in \tau_p$ such that $x_i \oplus k_1 = u_j$. Therefore, we have

$$\alpha = \sum_{i,j} \mathbb{I}_{ij}.$$

Now, for a fixed i, j, we have $\Pr[\mathbb{I}_{ij} = 1] = N^{-1}$ by using the randomness of k_1. Using the linearity of expectation, we have

$$\mathbf{E}[\alpha] \quad = \quad \sum_{ij} \mathbf{E}[\mathbb{I}_{ij}] = \sum_{ij} \Pr[\mathbb{I}_{ij} = 1] \leq \frac{qp}{N}. \tag{33}$$

Therefore, using Markov's inequality, we have

$$\Pr[\mathsf{B.11}] = \Pr[\alpha \geq \sqrt{q}] \leq \frac{\mathbf{E}[\alpha]}{\sqrt{q}} \overset{(4)}{\leq} \frac{p\sqrt{q}}{N}, \tag{34}$$

where (4) follows from Eqn. (33).

<u>Bounding B.12.</u> To bound the event B.12, we define an indicator random variable $\mathbb{I}_{ij}$ which is set to 1 if and only if $(x_i, y_i) \in \tau_c, (u_j, v_j) \in \tau_p$ such that $y_i \oplus k_1 = v_j$. Therefore, we have

$$\beta = \sum_{i,j} \mathbb{I}_{ij}.$$

Now, for a fixed i, j, we have $\Pr[\mathbb{I}_{ij} = 1] = N^{-1}$ by using the randomness of k_1. Using the linearity of expectation, we have

$$\mathbf{E}[\beta] \quad = \quad \sum_{ij} \mathbf{E}[\mathbb{I}_{ij}] = \sum_{ij} \Pr[\mathbb{I}_{ij} = 1] \leq \frac{qp}{N}. \tag{35}$$

Therefore, using Markov's inequality, we have

$$\Pr[\mathsf{B.12}] = \Pr[\beta \geq \sqrt{q}] \leq \frac{\mathbf{E}[\beta]}{\sqrt{q}} \overset{(5)}{\leq} \frac{p\sqrt{q}}{N}, \tag{36}$$

where (5) follows from Eqn. (35). Using the Eqn. (15)-Eqn. (36), the result follows. $\qquad \square$

4.2   Analysis of Good Transcripts

In this section, we state that for a good transcript $\tau = (\tau_c, \tau_p, \mathbf{k})$, realizing τ is almost as likely in the real world as in the ideal world. More formally,

Lemma 5 (Good Lemma). *Let* $\tau = (\tau_c, \tau_p, \mathbf{k}) \in \mathsf{GoodT}$ *be a good transcript. Let* X_{re} *and* X_{id} *be defined as above. For some positive integer* $0 \leq t \leq q/N^{1/3}$, *we have*

$$\frac{\Pr[\mathsf{X}_{\mathrm{re}} = \tau]}{\Pr[\mathsf{X}_{\mathrm{id}} = \tau]} \quad \geq \quad 1 - \left(\frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{13q}{N^{2/3}} + \frac{2pt}{q^2} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{28pq^2}{N^2} + \frac{4p^2q}{N^2} \right.$$
$$\left. + \frac{24q^3}{N^2} + \frac{2q^{3/2}}{N} \right).$$

Proof of this lemma is the most difficult part of the paper. Hence, we devote the following separate section for proving it. Therefore, by applying H-Coefficient technique (i.e., Theorem 2) with Lemma 4 and Lemma 5, the result follows. $\qquad \square$

5   Proof of Good Lemma

In this section, we prove that for a good transcript $\tau = (\tau_c, \tau_p, \mathbf{k})$, realizing it in the real world is as likely as realizing it in the ideal world. Note that, we have shown in Lemma 3 that to compute the ratio of real to ideal interpolation probability for a good transcript τ, one needs to compare

$$\mathsf{p}(\tau) \overset{\Delta}{=} \Pr[\mathsf{P} \leftarrow_{\$} \mathbb{P}(n) : \mathsf{pEDM}_{\mathbf{k}}^{\mathsf{P}} \mapsto \tau_c \mid \mathsf{P} \mapsto \tau_p]$$

with N^q. Therefore, it is enough to establish a lower bound of $\mathsf{p}(\tau)$.

5.1   Establishing Lower bound on $\mathsf{p}(\tau)$

First of all, for a good transcript $\tau = (\tau_c, \tau_p, \mathbf{k})$ recall that $\mathcal{U}$ is the set of all domain points of primitive queries and $\mathcal{V}$ is the set of all range points of it. Since, $\tau = (\tau_c, \tau_p, k_1, k_2)$ is a good transcript, we can partition the set of construction queries $\tau_c \in \tau$ into a finite number of disjoint groups as follows:

$$
\begin{aligned}
\mathcal{Q}_\mathcal{U} &\stackrel{\Delta}{=} \{(x,y) \in \tau_c : x \oplus k_1 \in \mathcal{U}\} \\
\mathcal{Q}_\mathcal{V} &\stackrel{\Delta}{=} \{(x,y) \in \tau_c : y \oplus k_1 \in \mathcal{V}\} \\
\mathcal{Q}_0 &\stackrel{\Delta}{=} \{(x,y) \in \tau_c : x \oplus k_1 \notin \mathcal{U}, y \oplus k_1 \notin \mathcal{V}\}
\end{aligned}
$$

Having defined the sets, we claim that the sets are disjoint and they exhaust the entire set of attainable good transcripts. By the definition of bad transcripts, we have $\mathcal{Q}_\mathcal{U} \cap \mathcal{Q}_\mathcal{V} = \emptyset$ and by definition we have $\mathcal{Q}_\mathcal{U} \cap \mathcal{Q}_0 = \emptyset, \mathcal{Q}_\mathcal{V} \cap \mathcal{Q}_0 = \emptyset$. Hence, we have the following result:

Proposition 2. *Let* $\tau = (\tau_c, \tau_p, k_1, k_2) \in \mathsf{GoodT}$ *be a good transcript. Then the sets* $(\mathcal{Q}_\mathcal{U}, \mathcal{Q}_\mathcal{V}, \mathcal{Q}_0)$ *are pairwise disjoint.*

Note that, since τ is a good transcript, we have, $\alpha = |\mathcal{Q}_\mathcal{U}| \leq \sqrt{q}$ and $\beta = |\mathcal{Q}_\mathcal{V}| \leq \sqrt{q}$. Let $\mathsf{E}_\mathcal{U}$ denote the event $\mathsf{pEDM}^\mathsf{P}_\mathbf{k} \mapsto \mathcal{Q}_\mathcal{U}$. Similarly, $\mathsf{E}_\mathcal{V}$ denote the event $\mathsf{pEDM}^\mathsf{P}_\mathbf{k} \mapsto \mathcal{Q}_\mathcal{V}$ and finally, E_0 denote the event $\mathsf{pEDM}^\mathsf{P}_\mathbf{k} \mapsto \mathcal{Q}_0$. Now, it is easy to see that

$$
\begin{aligned}
\mathsf{p}(\tau) &= \Pr[\mathsf{E}_\mathcal{U} \wedge \mathsf{E}_\mathcal{V} \wedge \mathsf{E}_0 \mid \mathsf{P} \mapsto \tau_p] \\
&= \underbrace{\Pr[\mathsf{E}_\mathcal{U} \wedge \mathsf{E}_\mathcal{V} \mid \mathsf{P} \mapsto \tau_p]}_{\mathsf{p}_1(\tau)} \cdot \underbrace{\Pr[\mathsf{E}_0 \mid \mathsf{E}_\mathcal{U} \wedge \mathsf{E}_\mathcal{V} \wedge \mathsf{P} \mapsto \tau_p]}_{\mathsf{p}_2(\tau)}
\end{aligned}
\tag{37}
$$

Thus, it is enough to establish a good lower bound on $\mathsf{p}_1(\tau)$ and $\mathsf{p}_2(\tau)$ for a good transcript τ.

5.2   Lower Bound of $\mathsf{p}_1(\tau)$

To lower bound $\mathsf{p}_1(\tau)$, we define the following sets:

$$
\begin{aligned}
\mathcal{S}_1 &\stackrel{\Delta}{=} \{x \oplus k_1 : (x,y) \in \mathcal{Q}_\mathcal{U}\}, \quad \mathcal{S}_2 \stackrel{\Delta}{=} \{x \oplus k_1 : (x,y) \in \mathcal{Q}_\mathcal{V}\} \\
\mathcal{D}_1 &\stackrel{\Delta}{=} \{y \oplus k_1 : (x,y) \in \mathcal{Q}_\mathcal{U}\}, \quad \mathcal{D}_2 \stackrel{\Delta}{=} \{y \oplus k_1 : (x,y) \in \mathcal{Q}_\mathcal{V}\}
\end{aligned}
$$

Note that, $\mathcal{S}_1 \subseteq \mathcal{U}$ and $|\mathcal{S}_1| = \alpha$. Similarly, $\mathcal{D}_2 \subseteq \mathcal{V}$ and $|\mathcal{D}_2| = \beta$. Conditioned on $\mathsf{P} \mapsto \tau_p$, P is fixed on exactly p input-output pairs. For each $(x,y) \in \mathcal{Q}_\mathcal{U}$, there is a unique $(u,v) \in \tau_p$ such that $x \oplus k_1 = u$, so that $\mathsf{P}(x \oplus k)$ is well defined, which is equal to v. Similarly, for each $(x,y) \in \mathcal{Q}_\mathcal{V}$, there is a unique $(u,v) \in \tau_p$ such that $y \oplus k_1 = v$, so that $\mathsf{P}^{-1}(y \oplus k_1)$ is well defined, which is equal to u. This leads us to define the following two additional sets:

$$
\begin{aligned}
\mathcal{X}_1 &\stackrel{\Delta}{=} \{\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2 : (x,y) \in \mathcal{Q}_\mathcal{U}\} \\
\mathcal{X}_2 &\stackrel{\Delta}{=} \{\mathsf{P}^{-1}(y \oplus k_1) \oplus x \oplus k_1 \oplus k_2 : (x,y) \in \mathcal{Q}_\mathcal{V}\}.
\end{aligned}
$$

In the following we state that every element of $\mathcal{D}_1$ is distinct and does not collide with any primitive query output. Similarly, every element of $\mathcal{S}_2$ is distinct and does not collide with any primitive query input.

Proposition 3. *Every element of $\mathcal{D}_1$ is distinct and does not collide with any primitive query output. Similarly, every element of $\mathcal{S}_2$ is distinct and does not collide with any primitive query input.*

Proof. The distinct property of $\mathcal{D}_1$ follows from $\neg$B.6. Moreover, if any elements of $\mathcal{D}_1$ collides with any primitive query output then it would satisfy condition B.2. This says that $\mathcal{D}_1 \cap \mathcal{V} = \emptyset \Rightarrow \mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$ and hence $|\mathcal{D}_1| = \alpha$.

By definition every element of $\mathcal{S}_2$ is unique and does not collide with any primitive query input (otherwise satifies condition B.2). Hence, $\mathcal{S}_2 \cap \mathcal{U} = \emptyset \Rightarrow \mathcal{S}_2 \cap \mathcal{S}_1 = \emptyset$ and hence $|\mathcal{S}_2| = \beta$. $\qquad\square$

However, the above result says that $|\mathcal{D}_1| = \alpha$ and $|\mathcal{S}_2| = \beta$. Now, we have the following proposition which states that every element of $\mathcal{X}_1$ and $\mathcal{X}_2$ are distinct and $\mathcal{X}_1$ is pairwise disjoint with $\mathcal{S}_1$ and $\mathcal{S}_2$. Similarly, every element of $\mathcal{X}_2$ is distinct and pairiwise disjoint with $\mathcal{D}_1$ and $\mathcal{D}_2$.

Proposition 4. *Every element of $\mathcal{X}_1$ is distinct and $\mathcal{X}_1 \cap \mathcal{S}_1 = \emptyset, \mathcal{X}_1 \cap \mathcal{S}_2 = \emptyset$. Moreover, every element of $\mathcal{X}_2$ is distinct and $\mathcal{X}_2 \cap \mathcal{D}_1 = \emptyset, \mathcal{X}_2 \cap \mathcal{D}_2 = \emptyset$.*

Proof. For the sake of contradiction, let us assume that $\mathsf{P}(x_i \oplus k_1) \oplus x_i \oplus k_1 \oplus k_2 = \mathsf{P}(x_j \oplus k_1) \oplus x_j \oplus k_1 \oplus k_2$ for some $(x_i, y_i),)(x_j, y_j) \in \mathcal{Q}_{\mathcal{U}}$. But this implies the condition B.7 to hold, which implies that τ is not a good transcript. Thus, every element of $\mathcal{X}_1$ is distinct. Moreover, none of the elements of $\mathcal{X}_1$ collides with any primitive query input, othwerwise it would satisfy condition B.1. This implies that $\mathcal{X}_1 \cap \mathcal{S}_1 = \emptyset$. Moreover, $\mathcal{X}_1 \cap \mathcal{S}_2 = \emptyset$ which follows due to $\neg$B.4. Thus, we have, $\mathcal{X}_1 \cap \mathcal{U} = \emptyset \Rightarrow \mathcal{X}_1 \cap \mathcal{S}_1 = \emptyset$ and $\mathcal{X}_1 \cap \mathcal{S}_2 = \emptyset$. Hence, $|\mathcal{X}_1| = \alpha$. For the second part of the proof, for the sake of contradiction, let us assume that $\mathsf{P}^{-1}(y_i \oplus k_1) \oplus (x_i \oplus k_1) \oplus k_2 = \mathsf{P}^{-1}(y_j \oplus k_1) \oplus (x_j \oplus k_1) \oplus k_2$ for some $(x_i, y_i), (x_j, y_j) \in \mathcal{Q}_{\mathcal{V}}$. But this implies the condition B.8 to hold, which implies that τ is not a good transcript. Thus, every element of $\mathcal{X}_2$ is distinct. Moreover, none of the elements of $\mathcal{X}_2$ collides with any primitive query output, otherwise it would satisfy condition B.3. This implies that $\mathcal{X}_2 \cap \mathcal{D}_2 = \emptyset$. Moreover, $\mathcal{X}_2 \cap \mathcal{D}_1 = \emptyset$ which follows due to $\neg$B.5. Thus, we have, $\mathcal{X}_2 \cap \mathcal{V} = \emptyset \Rightarrow \mathcal{X}_2 \cap \mathcal{D}_2 = \emptyset$ and $\mathcal{X}_2 \cap \mathcal{D}_1 = \emptyset$. Hence, $|\mathcal{X}_2| = \beta$. $\qquad\square$

Now, from Proposition 3 and Proposition 4, we have $|\mathcal{S}_1| = |\mathcal{X}_1| = |\mathcal{D}_1| = \alpha$ and $|\mathcal{S}_2| = |\mathcal{X}_2| = |\mathcal{D}_2| = \beta$. Also recall that $|\mathcal{U}| = |\mathcal{V}| = p$. Now, we consider the following two sequences:

$$
\begin{aligned}
\mathcal{Q}_1 &\triangleq \big((\mathsf{P}(x_i \oplus k_1) \oplus x_i \oplus k_1 \oplus k_2, y_i \oplus k_1) : (x_i, y_i) \in \mathcal{Q}_{\mathcal{U}}\big). \\
\mathcal{Q}_2 &\triangleq \big((x_i \oplus k_1, \mathsf{P}^{-1}(y_i \oplus k_1) \oplus x_i \oplus k_1 \oplus k_2) : (x_i, y_i) \in \mathcal{Q}_{\mathcal{V}}\big).
\end{aligned}
$$

From Proposition 3 and Proposition 4, it follows that the domain of $\mathcal{Q}_1$ is disjoint with the domain of $\mathcal{Q}_2$. Moreover, they are individually disjoint with $\mathcal{U}$. Similarly, the range of $\mathcal{Q}_1$ is disjoint with the range of $\mathcal{Q}_2$ and they are individually disjoint with $\mathcal{V}$. Therefore, we have $\mathfrak{X} = (\mathcal{U}, \mathcal{X}_1, \mathcal{S}_2)$ and $\mathfrak{Y} = (\mathcal{V}, \mathcal{D}_1, \mathcal{X}_2)$ are disjoint collections. Thus, from Proposition 2 one has,

$$
\mathsf{p}_1(\tau) \triangleq \Pr[\mathsf{P} \leftarrow_\$ \mathbb{P}(n) : \mathfrak{X} \setminus \mathcal{U} \xrightarrow{\mathsf{P}} \mathfrak{Y} \setminus \mathcal{V} \mid \mathsf{P} \mapsto \tau_p] = \frac{1}{(N-p)_{\alpha+\beta}}. \tag{38}
$$

5.3 Lower Bound on $\mathsf{p}_2(\tau)$

In the last section, we have seen that P has been fixed on $\alpha + \beta$ input-output (apart from p input-output primitive pairs). Moreover, the collection of input and output sets of P that have been explored in the last section is $\mathfrak{X} = (\mathcal{U}, \mathcal{X}_1, \mathcal{S}_2)$ and $\mathfrak{Y} = (\mathcal{V}, \mathcal{D}_1, \mathcal{X}_2)$. Recall that, we have defined the set

$$
\mathcal{Q}_0 = \{(x, y) \in \tau_c : x \oplus k_1 \notin \mathcal{U}, y \oplus k_1 \notin \mathcal{V}\}.
$$

For the sake of simplicity, we rename the elements of $\mathcal{Q}_0$ as $\mathcal{Q}_0 = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{q'}, y_{q'})\}$. It is easy to see that $|\mathcal{Q}_0| = q' = q - (\alpha + \beta)$. Let us define two sets $\mathcal{X} = \{x \in \{0, 1\}^n :$

$(x, y) \in (\tau_c \setminus \mathcal{Q}_\mathcal{U})\}$ and $\mathcal{Y} = \{y \in \{0,1\}^n : (x, y) \in \mathcal{Q}_0\}$, where $\mathcal{Y}$ is the set of the distinct number of responses. Let $r = |\mathcal{Y}|$ and let us denote

$$\mathcal{S} = \{(x, y) \in \mathcal{Q}_0 : \forall (x', y') \neq (x, y) \in \mathcal{Q}_0, y \neq y'\}$$

be the set of non-colliding queries of $\mathcal{Q}_0$ and $s' = |\mathcal{S}|$. Since, τ is a good transcript, $s' \geq q - M$ where $M = q/N^{1/3}$, otherwise B.10 would be satisfied. Now, we bound the probability that a permutation P realizes $\mathcal{Q}_0$, i.e., we need to lower bound the number of permutations P which are already fixed on $\alpha + \beta$ input-output pairs such that

$$\forall (x, y) \in \mathcal{Q}_0, \mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2) \oplus k_1 = y \tag{39}$$

holds. Note that the equations in Eqn. (39) are not independent as two permutations are identical. For example, if there exists two queries (x, y) and (x', y') in $\mathcal{Q}_0$ such that $\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2 = x' \oplus k_1$, then one must have $\mathsf{P}(x' \oplus k_1) = y \oplus k_1$. Similarly, if $\mathsf{P}(x \oplus k_1) = y' \oplus k_1$, then one must have $\mathsf{P}(x' \oplus k_1) \oplus x' \oplus k_1 \oplus k_2 = x \oplus k_1$. For simplicity, one could count only permutations P which are already fixed on $\alpha + \beta$ input-output pairs, such that for any query $(x, y) \in \mathcal{Q}_0$, $\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2 \notin \mathcal{X} \oplus k_1$, however this only leads to a birthday bound. Hence, to get a bound beyond the birthday, we need to allow collisions and a more precise counting. For doing this, we will be considering permutations P which are already fixed on $\alpha + \beta$ input-output pairs, such that $\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2 = x' \oplus k_1$ for t pairs of $((x, y), (x', y'))$ of distinct non-colliding queries, where t is some sufficiently large value. However, we must be careful in choosing the t-pairs of distinct non-colliding queries that do not create any incompatibility with other queries.

5.3.1　Counting Collisions

To this end, we define an index set $\mathcal{I} = \{i \in [q'] : (x_i, y_i) \in \mathcal{S}\}$ and $\mathcal{I}^{(2)}$ denotes the set of all ordered pairs of distinct elements of $\mathcal{I}$, i.e., $\mathcal{I}^{(2)} = \{(i, j) : i, j \in \mathcal{I}, i \neq j\}$.

Definition 2. For a fixed positive integer t, an unordered set of t ordered pair of indices

$$\mathcal{I}_t = \{(i_1, j_1), (i_2, j_2), \ldots, (i_t, j_t)\} \subseteq \mathcal{I}^{(2)},$$

is good if it satisfies the following conditions:

1. for $l \in [t], x_{j_l} \oplus x_{i_l}$ are distinct.

2. for $l \in [t], y_{i_l} \oplus x_{j_l}$ are distinct.

3. for $l \in [t], x_{j_l} \oplus x_{i_l} \oplus k_2 \notin \mathcal{V}$.

4. for $l \in [t], y_{i_l} \oplus x_{j_l} \oplus k_2 \notin \mathcal{U}$.

5. for $l \in [t], y_{i_l} \oplus k_2 \oplus x_{j_l} \notin \mathcal{X} \oplus k_1$.

6. for $l \in [t], x_{i_l} \oplus k_2 \oplus x_{j_l} \notin \mathcal{Y} \oplus k_1$.

7. for $l \in [t], x_{j_l} \oplus x_{i_l} \oplus k_2 \notin \mathcal{X}_2$.

8. for $l \in [t], y_{i_l} \oplus x_{j_l} \oplus k_2 \notin \mathcal{X}_1$.

Note that $\mathcal{I}_t$ is the set of t ordered pair of indices of non-colliding queries. We justify below why the above-listed conditions, if obeyed by the elements of $\mathcal{I}_t$, do not create any incompatibility with the other queries.

RATIONALE FOR THE CONDITIONS. We elaborate here the conditions stated in the Defn. 2 and justify why they ensure the compatibilities with other queries. We call an element in the set $\mathcal{I}_t$ a *dependency pair*. A dependency pair $(i_l, j_l) \in \mathcal{I}_t$ are dependent in one of the following two ways:

(a) $\mathsf{P}(x_{i_l} \oplus k_1) \oplus (x_{i_l} \oplus k_1) \oplus k_2 = x_{j_l} \oplus k_1$ or

(b) $\mathsf{P}(x_{i_l} \oplus k_1) = y_{j_l} \oplus k_1$.

for some permutations P which are already fixed on $\alpha + \beta$ input-output pairs. Such a dependency pair is said to be of *length 1*. Now, from (a) and (b), we have the following two equalities:

$$\widehat{(a)}\ \mathsf{P}(x_{i_l} \oplus k_1) = x_{i_l} \oplus x_{j_l} \oplus k_2, \quad \widehat{(b)}\ \mathsf{P}(x_{j_l} \oplus k_1) = x_{i_l} \oplus x_{j_l} \oplus k_2.$$

Both these equalities impose the distinctness of the permutation output, i.e., $x_{i_l} \oplus x_{j_l} \oplus k_2$ as its input, i.e, $x_{i_l} \oplus k_1$ or $x_{j_l} \oplus k_1$ are distinct which justifies condition (1) of Defn. 2. Similarly, $\widehat{(a)}$ and $\widehat{(b)}$ also imposes that the permutation outputs should not collide with any primitive output (i.e., the elements of $\mathcal{V}$), as their corresponding input does not collide with any primitive input (i.e., elements of $\mathcal{U}$), which justifies condition (3) of Defn. 2. Similarly, $x_{i_l} \oplus x_{j_l} \oplus k_2$ should not collide with any elements of $\mathcal{Y} \oplus k_1$, which justifies condition (6) of Defn. 2. Moreover, $x_{i_l} \oplus x_{j_l} \oplus k_2$ should not collide with any elements of $\mathcal{X}_2$, as $(x_{i_l}, y_{i_l}), (x_{j_l,y_{j_l}}) \notin \mathcal{Q}_{\mathcal{V}}$. This justifies conditions (7) of Defn. 2. Note that Eqn. (a) also imposes the following equality:

$$\widehat{(c)}\ \mathsf{P}(x_{j_l} \oplus k_1) = y_{i_l} \oplus k_1.$$

Now, we require that $\mathsf{P}(x_{j_l} \oplus k_1) \oplus (x_{j_l} \oplus k_1 \oplus k_2)$ or equivalently $y_{i_l} \oplus x_{j_l} \oplus k_2$ should be distinct, which justifies condition (2) of Defn. 2. Moreover, it should not collide with any other elements of $\mathcal{X} \oplus k_1$, otherwise that would extend the length of the dependency pair by 1. This phenomena justifies the condition (5) of the definition. Similarly $y_{i_l} \oplus x_{j_l} \oplus k_2$ should not collide with any primitive inputs, as $y_{j_l} \oplus k_1$ does not collide with any primitive output. This justifies condition (4) of the definition. We also require that $y_{i_l} \oplus x_{j_l} \oplus k_2$ should not collide with any elements of $\mathcal{X}_2$, otherwise $y_{j_l} \oplus k_1 \in \mathcal{D}_1$ which is not possible as $(x_{j_l}, y_{j_l}) \notin \mathcal{Q}_{\mathcal{U}}$. This justifies the condition (8) of Defn. 2.

Lemma 6. *Fix a positive integer t such that $0 \leq t \leq M$. Then the number of good sets $\mathcal{I}_t$ of t pairs of non-colliding queries is at least*

$$|\mathcal{I}_t| \geq \frac{(s')_{2t}}{t!} \left(1 - \frac{4q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2\sqrt{q}}{N^{1/3}}\right).$$

Proof. First, observe that among the $s'(s' - 1)$ possible pairs of non-colliding query indices (i_1, j_1), at most $(2\sigma + 2p + \alpha + \beta)$ of them do not satisfy conditions (3)-(8). Indeed, by definition of a good transcript (more precisely, condition (B.9)), there cannot be more than σ pairs $((x_{i_l}, y_{i_l}), (x_{j_l}, y_{j_l}))$ such that $y_{i_l} \oplus x_{j_l} \oplus k_2 \in \mathcal{X} \oplus k_1$ and there cannot be more than σ pairs $((x_{i_l}, y_{i_l}), (x_{j_l}, y_{j_l}))$ such that $x_{i_l} \oplus x_{j_l} \oplus k_2 \in \mathcal{Y} \oplus k_1$. Similarly, by condition (B.11), there cannot be more than α pairs such that $y_{i_l} \oplus x_{j_l} \oplus k_2 \in \mathcal{X}_1$. By condition (B.12), there cannot be more than β pairs such that $x_{j_l} \oplus x_{i_l} \oplus k_2 \in \mathcal{X}_2$. Hence, we can lower bound $\mathcal{I}_t$ as follows:

- we can choose (i_1, j_1) among at least $s'(s' - 1) - 2\sigma - 2p - \alpha - \beta$ possibilities

- once (i_1, j_1) is fixed, we can choose i_2 freely from the remaining $(s' - 2)$ possibilities; then, j_2 must be different from i_1, j_1 and i_2. Moreover, we also have $x_{j_2} \oplus x_{i_2} \neq x_{j_1} \oplus x_{i_1}$ and $y_{i_2} \oplus x_{j_2} \neq y_{i_1} \oplus x_{j_1}$. Hence, the choice for j_2 is $(s' - 5)$; after removing the at most $2\sigma + 2p + \alpha + \beta$ pairs of queries not satisfying (3)-(8), there remains at least $(s' - 2)(s' - 5) - 2\sigma - 2p - \alpha - \beta$ possibilities for the pair (i_2, j_2)

- assume $(i_1, j_1), (i_2, j_2), (i_{l-1}, j_{l-1})$ have been chosen, we can choose i_l freely from the $(s' - 2l + 2)$ remaining possibilities; then, j_l must be different from $i_1, j_1, \ldots, i_{l-1}, j_{l-1}, i_l$.

Moreover, it must be such that $x_{j_l} \oplus x_{i_l} \neq x_{j_d} \oplus x_{i_d}$ for $d \in [l-1]$ and $y_{i_l} \oplus x_{j_l} \neq y_{i_d} \oplus x_{j_d}$ for $d \in [l-1]$. Thus, overall there are at least $(s' - 4l + 3)$ possibilities for j_l; after removing the at most $2\sigma + 2p + \alpha + \beta$ pairs not satisfying (3)-(8), there remains at least $(s' - 2l + 2)(s' - 4l + 3) - 2\sigma - 2p - \alpha - \beta$ possibilities for the pair (i_l, j_l)

Since $\mathcal{I}_t$ is unordered set of t pairs, the number $|\mathcal{I}_t|$ of good sets is at least

$$|\mathcal{I}_t| \geq \frac{1}{t!} \prod_{l=0}^{t-1} \left((s' - 2l)(s' - 4l - 1) - 2\sigma - 2p - \alpha - \beta \right).$$

Then, we have

$$
\begin{aligned}
|\mathcal{I}_t| \quad &\geq \quad \frac{(s')_{2t}}{t!} \prod_{l=0}^{t-1} \frac{(s' - 2l)(s' - 4l - 1) - 2p - 2\sigma - \alpha - \beta}{(s' - 2l)(s' - 2l - 1)} \\
&\geq \quad \frac{(s')_{2t}}{t!} \prod_{l=0}^{t-1} \left(1 - \frac{2ls' - 4l^2 + 2p + 2\sigma + \alpha + \beta}{(s' - 2l)(s' - 2l - 1)} \right) \\
&\overset{(1)}{\geq} \quad \frac{(s')_{2t}}{t!} \prod_{l=0}^{t-1} \left(1 - \frac{2ls' + 2p + 2\sigma + \alpha + \beta}{(s' - 2M)^2} \right) \\
&\geq \quad \frac{(s')_{2t}}{t!} \left(1 - \sum_{l=0}^{t-1} \frac{2ls' + 2p + 2\sigma + \alpha + \beta}{(s' - 2M)^2} \right) \\
&\overset{(2)}{\geq} \quad \frac{(s')_{2t}}{t!} \left(1 - \frac{2s'M^2 + 2pt + 2\sigma M + \alpha M + \beta M}{q^2} \right) \\
&\overset{(3)}{\geq} \quad \frac{(s')_{2t}}{t!} \left(1 - \frac{2M^2}{q} - \frac{2pt}{q^2} - \frac{2\sigma M}{q^2} - \frac{\beta M}{q^2} - \frac{\alpha M}{q^2} \right) \\
&\overset{(4)}{\geq} \quad \frac{(s')_{2t}}{t!} \left(1 - \frac{2q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2q}{N^{2/3}} - \frac{2}{\sqrt{q}N^{1/3}} \right) \\
&\overset{(5)}{\geq} \quad \frac{(s')_{2t}}{t!} \left(1 - \frac{2q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2q}{N^{2/3}} - \frac{2\sqrt{q}}{N^{1/3}} \right) \\
&\geq \quad \frac{(s')_{2t}}{t!} \left(1 - \frac{4q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2\sqrt{q}}{N^{1/3}} \right).
\end{aligned}
$$

Note that, (1) follows as $l \leq t \leq M$. (2) follows as $t \leq M$ and $s' - 2M \leq q$. Moreover, (3) follows as $s' \leq q$. (4) follows as $M \leq q/N^{1/3}$, $\sigma \leq q^2/N^{1/3}$ and $\alpha, \beta \leq \sqrt{q}$. Moreover, (5) follows as $q \geq 1$ and hence the result follows. $\qquad\square$

From now on we fix a positive integer t such that $0 \leq t \leq M$ and a good set $\mathcal{I}_t = \{(i_1, j_1), \ldots, (i_t, j_t)\}$. Now, for a good set $\mathcal{I}_t$, we define the following set

$$\mathcal{Q}_{\mathcal{I}_t} \overset{\triangle}{=} \{(x_i, y_i) \in \mathcal{Q}_0 : (i, \star) \in \mathcal{I}_t \vee (\star, i) \in \mathcal{I}_t\}.$$

We are now interested to lower bound the number of permutations P that are already fixed on $\alpha + \beta$ input output pairs and satisfying the following

$$\mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2) \oplus k_1 = y, \forall(x, y) \in \mathcal{Q}_0 \tag{40}$$

such that for any $l \in [t]$, $\mathsf{P}(x_{i_l} \oplus k_1) \oplus x_{i_l} \oplus k_1 \oplus k_2 = x_{j_l} \oplus k_1$. Note that such a permutation P which is already fixed on $\alpha + \beta$ input output pairs and satisfying Eqn. (40) for the $2t$ queries appearing in $\mathcal{I}_t$ if and only if $\forall l \in [t]$ we have the following:

1. $\mathsf{P}(x_{i_l} \oplus k_1) = x_{j_l} \oplus x_{i_l} \oplus k_2$

2. $\mathsf{P}(x_{j_l} \oplus k_1) = y_{i_l} \oplus k_1$

3. $\mathsf{P}(y_{i_l} \oplus x_{j_l} \oplus k_2) = y_{j_l} \oplus k_1$

Note that this set of $3t$ equalities is input-output compatible as $\mathcal{I}_t$ is a good set. Now, it is easy to see that the collection of the following sets are pairwise disjoint: $\mathfrak{X}^+ = \{\mathcal{X} \oplus k_1, \mathcal{U}, \mathcal{X}_1, \{y_{i_l} \oplus x_{j_l} \oplus k_2 : l \in [t]\}\}$. Similarly, all the sets in the following collection are pairwise disjoint: $\mathfrak{Y}^+ = \{\mathcal{Y} \oplus k_1, \mathcal{V}, \mathcal{X}_2, \{x_{j_l} \oplus x_{i_l} \oplus k_2 : l \in [t]\}, \{x_{i_l} \oplus x_{j_l} \oplus k_2 : l \in [t]\}\}$. Thus, we define the following sets:

$$\begin{aligned} \mathcal{X}' &= \mathcal{X} \oplus k_1 \cup \mathcal{U} \cup \mathcal{X}_1 \cup \{y_{i_l} \oplus x_{j_l} \oplus k_2 : l \in [t]\} \\ \mathcal{Y}' &= \mathcal{Y} \oplus k_1 \cup \mathcal{V} \cup \mathcal{X}_2 \cup \{x_{j_l} \oplus x_{i_l} \oplus k_2 : l \in [t]\} \cup \{x_{i_l} \oplus x_{j_l} \oplus k_2 : l \in [t]\} \end{aligned}$$

It is easy to see that $|\mathcal{X}'| = q' + p + \alpha + t$ and $|\mathcal{Y}'| = r + 2t + p + \beta$, where recall that $r = |\mathcal{Y}|$. Now, it remains to consider the remaining $q' - 2t$ queries $(x, y) \in \mathcal{Q}_0$ such that $(x, y) \notin \mathcal{Q}_{\mathcal{I}_t}$. To this end, let $q'' = q' - 2t = q - \alpha - \beta - 2t$ be the number of remaining queries in $\mathcal{Q}_0 \setminus \mathcal{Q}_{\mathcal{I}_t}$, $s'' = s' - 2t$ be the number of non-colliding queries in $\mathcal{Q}_0 \setminus \mathcal{Q}_{\mathcal{I}_t}$ and $r' = r - 2t$ be the numbr of distinct oracle responses appearing in these queries.

By following the approach of [CS18], we regroup the elements of $\mathcal{Q}_0 \setminus \mathcal{Q}_{\mathcal{I}_t}$ such that all queries with the same output becomes consecutive. We write the queries as follows:

$$\tau' = \begin{cases} \Big((x_{11}, y_1), \ldots, (x_{1q_1}, y_1), \\ (x_{21}, y_2), \ldots, (x_{2q_2}, y_2), \\ \quad \vdots \quad \vdots \quad \vdots \quad \quad \vdots \\ (x_{r',1}, y_{r'}), \ldots, (x_{r',q_{r'}}, y_{r'})\Big), \end{cases}$$

where $y_1, \ldots, y_{r'}$ are distinct. Moreover, we also have $(q_1 + \ldots + q_{r'}) = q''$. For the ease of later computations, we assume that in these grouping all the non-colliding queries appear first followed by colliding quries, i.e., $q_i = 1$ for $i \in [s'']$ and $q_i > 1$ for $i \in \{s'' + 1, \ldots, r'\}$. Now, our goal is to lower bound the number of permutations $\mathsf{P} \in \mathbb{P}(n)$ which are already fixed on $\alpha + \beta$ input output pairs and in addition to satisfying above $3t$ equalities, also satisfies the following.

$$\forall (x, y) \in \tau', \mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2) \oplus k_1 = y. \tag{41}$$

For this, we sample all intermediate values $z = \mathsf{P}^{-1}(y \oplus k_1)$, which leads us to the second step of the proof.

5.3.2 Sampling Intermediate Values

Let us consider a sequence $\mathbf{z} = (z_1, z_2, \ldots, z_{r'})$ of r many n-bit values. We say $\mathbf{z}$ is *good* if it satisfies the following conditions:

1. each z_i's are distinct

2. for all $i \in [r'], z_i \notin \mathcal{X}'$

3. for all $i \in [r']$ and $j \in [q_i], z_i \oplus x_{i,j}$ are distinct

4. for all $i \in [r']$ and $j \in [q_i], z_i \oplus x_{i,j} \oplus k_1 \oplus k_2 \notin \mathcal{Y}'$

Note that for any good tuple $\mathbf{z} = (z_1, \ldots, z_{r'})$, the set of the following equalities

$$\begin{cases} \forall i \in [r'], \forall j \in [q_i], \mathsf{P}(x_{i,j} \oplus k_1) = z_i \oplus k_2 \oplus x_{i,j} \oplus k_1 \\ \forall i \in [r'], \mathsf{P}(z_i) = y_i \oplus k_1 \end{cases}$$

is compatible with previously defined all input output pairs. Moreover, a permutation P satisfying the equations is such that $\mathsf{P}(\mathsf{P}(x \oplus k_1) \oplus x \oplus k_1 \oplus k_2) = y \oplus k_1$ for all $(x, y) \in \tau'$. In the following, we count the number of good tuples $\mathbf{z}$.

Lemma 7. *Fix a positive integer t such that $0 \le t \le M$ and a good set $\mathcal{I}_t$. Then the number of good tuples $\mathbf{z} = (z_1, \ldots, z_{r'})$ is at least*

$$\mathcal{N}_{\mathbf{z}}(t) \ge (N-q-p-\alpha-\beta-3t-r)_{s''} \prod_{i=0}^{s''-1} \left(1 - \frac{p+i}{N-(5q+p)-i}\right)(N)^{r'-s''}\left(1 - \frac{6q^2}{N^{4/3}} - \frac{2pq}{N^{4/3}}\right).$$

Proof. To count the number of good tuples $\mathbf{z}$, the number of valid choices for z_1 is at least $N - (q + p + \alpha + t) - q_1(r + 2t + p + \beta)$ as $z_1 \notin \mathcal{X}'$ and $z_1 \oplus x_{1,j} \oplus k_1 \oplus k_2 \notin \mathcal{Y}'$ for $j \in [q_1]$, where recall that $|\mathcal{X}'| = q + p + \alpha + t$ and $|\mathcal{Y}'| = r + 2t + p + \beta$.

Once the value of z_1 is fixed, z_2 can be chosen in the following way:

- $z_2 \ne z_1$

- $z_2 \notin \mathcal{X}'$

- $z_2 \oplus k_2 \oplus x_{2,j} \oplus k_1 \notin \mathcal{Y}'$ for $j \in [q_2]$

- $z_2 \ne z_1 \oplus x_{1,j} \oplus x_{2,j'}$ for $j \in [q_1], j' \in [q_2]$.

Thus, the number of valid choices for z_2 is at least $N - 1 - (q + p + \alpha + t) - q_2(r + 2t + p + \beta + q_1)$. In general after choosing the values for $z_1, \ldots, z_{i-1}$, we choose the value for z_i. In that case, the number of valid choices for z_i is at least

$$N - (i-1) - (q + p + \alpha + t) - q_i\Big(r + 2t + p + \beta + \sum_{j=1}^{i-1} q_j\Big).$$

This is because z_i cannot be equal to $z_1, \ldots, z_{i-1}$ which accounts for $i - 1$ terms in the above equation. Moreover, $z_i \notin \mathcal{X}'$ introduces $(q + p + \alpha + t)$ in the equation. For $j \in [q_i], z_i \oplus k_2 \oplus x_{i,j} \oplus k_1 \notin \mathcal{Y}'$ which includes $q_i(r + 2t + p + \beta)$ term in the equation and $z_i \ne z_l \oplus x_{l,j} \oplus x_{i,j'}$ for $l \in [i-1], j \in [q_l], j' \in [q_i]$ which introduces the term $q_i(q_1 + \ldots + q_{i-1})$ in the above equation. Therefore, overall the number of good tuples $\mathbf{z}$ is at least

$$\mathcal{N}_{\mathbf{z}}(t) \ge \prod_{i=0}^{r'-1} \Big(N - (q + p + \alpha + t) - i - q_{i+1}\big(r + 2t + p + \beta + \sum_{j=1}^{i} q_j\big)\Big). \qquad (42)$$

For the ease of the computation, we split up Eqn. (63) into two parts: the first part is comprised of s'' many non-colliding queries and the next part is comprised of colliding

queries. Therefore, we have

$$
\begin{aligned}
\mathcal{N}_{\mathbf{z}}(t) \;\geq\; & \prod_{i=0}^{s''-1}\left(N-(q+p+\alpha+t)-i-r-2t-p-\beta-i\right) \\
& \prod_{i=s''}^{r'-1}\left(N-(q+p+\alpha+t)-i-q_{i+1}\big(r+2t+p+\beta+\sum_{j=1}^{i}q_j\big)\right) \\
\;\geq\; & \underbrace{\prod_{i=0}^{s''-1}\left(N-(q+2p+\alpha+\beta+3t)-2i-r\right)}_{\mathsf{A}} \\
& \underbrace{\prod_{i=s''}^{r'-1}\left(N-(q+p+\alpha+t)-i-q_{i+1}\big(r+2t+p+\beta+\sum_{j=1}^{i}q_j\big)\right)}_{\mathsf{B}}. \quad (43)
\end{aligned}
$$

COMPUTING A: To compute A, we have the following.

$$
\begin{aligned}
\mathsf{A} \;=\; & (N-q-p-\alpha-\beta-3t-r)_{s''}\cdot\prod_{i=0}^{s''-1}\frac{N-q-2p-\alpha-\beta-3t-2i-r}{N-q-p-\alpha-\beta-3t-r-i} \\
\;=\; & (N-q-p-\alpha-\beta-3t-r)_{s''}\cdot\prod_{i=0}^{s''-1}\left(1-\frac{p+i}{N-q-p-\alpha-\beta-3t-r-i}\right) \\
\;\overset{(1)}{\geq}\; & (N-q-p-\alpha-\beta-3t-r)_{s''}\cdot\prod_{i=0}^{s''-1}\left(1-\frac{p+i}{N-(5q+p)-i}\right), \quad (44)
\end{aligned}
$$

where (1) follows as $\alpha,\beta\leq q, r\leq q$ and $3t\leq q$. Now, we compute B as follows:

COMPUTING B: To compute B, we need the following inequalities: for $i\in\{s'',s''+1,\ldots,r'-2,r'-1\}$, we have

$$
q+p+\alpha+t+i\overset{(2)}{\leq}q+p+\alpha+t+r'-1\overset{(3)}{=}q+p+\alpha+r-t-1\overset{(4)}{\leq}3q+p\quad(45)
$$

$$
r+2t+p+\beta+\sum_{j=1}^{i}q_j\overset{(6)}{\leq}r+2t+p+\beta+q''\overset{(7)}{=}r+p+\beta+q'\overset{(8)}{\leq}3q+p.\quad(46)
$$

Note that, (2) follows as $i\leq r'_1$. (3) and (7) follows as $r'=r-2t$ and $q''=q'-2t$ respectively. (4) follows as $\alpha\leq q, r-t-1\leq q$. Moreover, (6) follows as $(q_1+\ldots q_{r'})=q''$

and (8) follows as $q' \leq q, \beta \leq q$ and $r \leq q$. Now, we have the following:

$$
\begin{aligned}
\mathsf{B} &= \prod_{i=s''}^{r'-1} \left(N - (q+p+\alpha+t) - i - q_{i+1}\Big(r + 2t + p + \beta + \sum_{j=1}^{i} q_j\Big) \right) \\
&\geq \prod_{i=s''}^{r'-1} \left(N - (3q+p) - q_{i+1}(3q+p) \right) \quad \text{(By applying Eqn. (45), Eqn. (46))} \\
&\geq (N)^{r'-s''} \prod_{i=s''}^{r'-1} \left(\frac{N - (3q+p) - q_{i+1}(3q+p)}{N} \right) \\
&\geq (N)^{r'-s''} \prod_{i=s''}^{r'-1} \left(1 - \frac{q_{i+1}(6q+2p)}{N} \right) \geq (N)^{r'-s''} \left(1 - \sum_{i=s''}^{r'} \frac{q_{i+1}(6q+2p)}{N} \right) \\
&\geq (N)^{r'-s''} \left(1 - \frac{(6q+2p)\sum\limits_{i=s''}^{r'} q_{i+1}}{N} \right) \overset{(9)}{\geq} (N)^{r'-s''} \left(1 - \frac{(6q+2p)M}{N} \right) \\
&\overset{(10)}{\geq} (N)^{r'-s''} \left(1 - \frac{6q^2}{N^{4/3}} - \frac{2pq}{N^{4/3}} \right),
\end{aligned}
\tag{47}
$$

where (9) holds as $\sum\limits_{i=s''}^{r'-1} q_{i+1} \leq M$ and (10) holds as $M \leq q/N^{1/3}$. Therefore, the result follows from Eqn. (43), Eqn. (44) and Eqn. (47). $\qquad\square$

5.3.3   Final Calculation

Now we come to the final part of the proof. We have seen from Sect. 5.3.1 that P has been fixed on $3t$ input-output pairs. Moreover, from Sect. 5.3.2, it is trivial to check that P has been additionally fixed on $q'' + r'$ input-output pairs. Thus, in total P is fixed on $3t + q'' + r' = q' + t + r'$ input-output pairs where the last equality arises due to $q'' = q' - 2t$. Also recall that P has already been fixed on $p + \alpha + \beta$ input-output pairs. Therefore, from Lemma 6 and Lemma 7, we have

$$
\begin{aligned}
\mathsf{p}_2(\tau) &\geq \sum_{0 \leq t \leq M} \frac{|\mathcal{I}_t| \cdot \mathcal{N}_{\mathbf{z}}(t)}{(N - p - \alpha - \beta)_{q'+t+r'}} \\
&\geq (N)^{r'-s''} \cdot \left(1 - \frac{6q^2}{N^{4/3}} - \frac{2pq}{N^{4/3}} \right) \sum_{0 \leq t \leq M} \left[\frac{(s')_{2t}}{t!} \cdot \left(1 - \frac{4q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2\sqrt{q}}{N^{1/3}} \right) \right. \\
&\quad \left. \cdot \prod_{i=0}^{s''-1} \left(1 - \frac{p+i}{N - (5q+p) - i} \right) \cdot \cdot \frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - p - \alpha - \beta)_{q'+t+r'}} \right].
\end{aligned}
\tag{48}
$$

Therefore, from Lemma 3, Eqn. (37), Eqn. (38) and Eqn. (48), we have

$$
\begin{aligned}
\rho(\tau) \;\geq\; & \frac{N^{q+r'-s''}}{(N-p)_{\alpha+\beta}} \cdot \left(1 - \frac{6q^2}{N^{4/3}} - \frac{2pq}{N^{4/3}}\right) \sum_{0\leq t\leq M}\left[\frac{(s')_{2t}}{t!}\cdot\left(1 - \frac{4q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2\sqrt{q}}{N^{1/3}}\right)\right.\\
& \left.\cdot\prod_{i=0}^{s''-1}\left(1 - \frac{p+i}{N-(5q+p)-i}\right)\cdot\cdot\frac{(N-q-p-\alpha-\beta-3t-r)_{s''}}{(N-p-\alpha-\beta)_{q'+t+r'}}\right]\\
\geq\; & \underbrace{\frac{N^{\alpha+\beta}}{(N-p)_{\alpha+\beta}}}_{\geq 1}\cdot\underbrace{\frac{(N)^{r'-s''}}{(N-p-\alpha-\beta-q')_{r'-s''}}}_{\geq 1}\cdot\underbrace{\frac{N^{q'}}{(N-p-\alpha-\beta)_{q'}}\cdot\prod_{i=0}^{s''-1}\left(1 - \frac{p+i}{N-(5q+p)-i}\right)}_{\text{D.1}}\\
& \cdot\left(1 - \frac{6q^2}{N^{4/3}} - \frac{2pq}{N^{4/3}}\right)\cdot\left(1 - \frac{4q}{N^{2/3}} - \frac{2pt}{q^2} - \frac{2\sqrt{q}}{N^{1/3}}\right)\\
& \cdot\underbrace{\left[\sum_{0\leq t\leq M}\frac{(s')_{2t}\cdot(N-q-p-\alpha-\beta-3t-r)_{s''}}{t!\cdot(n-p-\alpha-\beta-q'-r'+s'')_{s''+t}}\right]}_{\text{D.2}}.
\end{aligned}
\tag{49}
$$

COMPUTING D.1.

$$
\begin{aligned}
\text{D.1} \;=\;& \frac{N^{q'}}{(N-p-\alpha-\beta)_{q'}}\cdot\prod_{i=0}^{s''-1}\left(1 - \frac{p+i}{N-(5q+p)-i}\right)\\
\overset{(1)}{\geq}\;& \prod_{i=0}^{q'-1}\left(1 + \frac{p+i}{N-p-i}\right)\cdot\prod_{i=0}^{s''-1}\left(1 - \frac{p+i}{N-(5q+p)-i}\right)\\
\overset{(2)}{\geq}\;& \prod_{i=0}^{q'-1}\left[\left(1 + \frac{p+i}{N-p-i}\right)\cdot\left(1 - \frac{p+i}{N-i-5q-p}\right)\right]\\
\geq\;& \prod_{i=0}^{q'-1}\left[1 - \left(\frac{5q(p+i)+p^2+2pi+i^2}{(N-p-i)(N-p-i-5q)}\right)\right]\\
\overset{(3)}{\geq}\;& \prod_{i=0}^{q'-1}\left[1 - \frac{20q(p+i)}{N^2} - \frac{4p^2}{N^2} - \frac{8pi}{N^2} - \frac{4i^2}{N^2}\right]\\
\geq\;& \left(1 - \frac{20pq^2}{N^2} - \frac{20q^3}{N^2} - \frac{4p^2q}{N^2} - \frac{8pq^2}{N^2} - \frac{4q^3}{N^2}\right)\\
\geq\;& \left(1 - \frac{28pq^2}{N^2} - \frac{4p^2q}{N^2} - \frac{24q^3}{N^2}\right),
\end{aligned}
\tag{50}
$$

where (1) holds as $N-p-\alpha-\beta\leq N-p$. (2) holds as $s''\leq q'$ and (3) holds as $p+i\leq N/2, p+i+5q\leq N/2$.

COMPUTING D.2.

$$
\begin{aligned}
\text{D.2} \;=\;& \sum_{0\leq t\leq M}\frac{(s')_{2t}\cdot(N-q-p-\alpha-\beta-3t-r)_{s''}}{t!\cdot(N-p-\alpha-\beta-q'-r'+s'')_{s''+t}}\\
\overset{(1)}{\geq}\;& \sum_{0\leq t\leq M}\frac{(s')_{2t}}{\underbrace{(s')_t(s')_t}_{\text{E.1}}}\cdot\underbrace{\frac{(s')_t(s')_t}{t!}\cdot\frac{(N-q-p-\alpha-\beta-s')_{s'-t}}{(N-q-p-\alpha-\beta)_{s'}}}_{\text{E.2}}\\
& \cdot\underbrace{\frac{(N-q-p-\alpha-\beta)_{s'}}{(N-q-p-\alpha-\beta-s')_{s'-t}}\cdot\frac{(N-q-p-\alpha-\beta-3t-r)_{s''}}{(N-p-\alpha-\beta-q'-r+s')_{s''+t}}}_{\text{E.3}},
\end{aligned}
\tag{51}
$$

where (1) holds as $r' = r - 2t$ and $s'' = s' - 2t$. Now, we individually bound E.1 and E.3 as follows:

COMPUTING E.1.

$$\mathsf{E.1} = \frac{(s')_{2t}}{(s')_t(s')_t} \geq \frac{(s' - 2M)^{2t}}{(s')^{2t}} \geq 1 - \frac{4tM}{s'} \overset{(2)}{\geq} 1 - \frac{4M^2}{q - M} \overset{(3)}{\geq} 1 - \frac{8M^2}{q} \overset{(4)}{\geq} \left(1 - \frac{8q}{N^{2/3}}\right), \quad (52)$$

where (2) follows as $s' \geq q - M$, (3) follows as $q - 3M \leq q/2$ and finally (4) follows as $M = q/N^{1/3}$.

COMPUTING E.3.

$$
\begin{aligned}
\mathsf{E.3} \quad &= \quad \frac{(N - q - p - \alpha - \beta)_{s'}}{(N - q - p - \alpha - \beta - s')_{s'-t}} \cdot \frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - p - \alpha - \beta - q' - r + s')_{s''+t}} \\
&\overset{(5)}{=} \quad \frac{(N - q - p - \alpha - \beta)_{s''+2t}}{(N - q - p - \alpha - \beta - s')_{s''+t}} \cdot \frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - p - \alpha - \beta - q' - r + s')_{s''+t}} \\
&= \quad \frac{(N - q - p - \alpha - \beta)_{s''+t}}{(N - q - p - r + s')_{s''+t}} \cdot \frac{(N - q - p - \alpha - \beta - s'' - t)_t}{(N - q - p - \alpha - \beta - s')_t} \cdot \frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - q - p - \alpha - \beta - s' - t)_{s''}} \\
&\overset{(6)}{=} \quad \underbrace{\frac{(N - q - p - \alpha - \beta)_{s''+t}}{(N - q - p - r + s')_{s''+t}}}_{\mathsf{E.3.1}} \cdot \underbrace{\frac{(N - q - p - \alpha - \beta - s' + t)_t}{(N - q - p - \alpha - \beta - s')_t}}_{\geq 1} \cdot \underbrace{\frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - q - p - \alpha - \beta - s' - t)_{s''}}}_{\mathsf{E.3.2}},
\end{aligned}
$$

where (5) and (6) follows as $s'' = s' - 2t$. Now, we individually bound E.3.1 and E.3.2 as follows:

COMPUTING E.3.1.

$$
\begin{aligned}
\mathsf{E.3.1} \quad &= \quad \frac{(N - q - p - \alpha - \beta)_{s''+t}}{(N - q - p - r + s')_{s''+t}} = \prod_{i=0}^{s''+t-1} \frac{N - q - p - \alpha - \beta - i}{N - q - p - r + s' - i} \\
&= \quad \prod_{i=0}^{s''+t-1} \left(1 - \frac{\alpha + \beta - r + s'}{N - q - p - i - r + s'}\right) \\
&= \quad \prod_{i=0}^{s''+t-1} \left(1 - \frac{\alpha + \beta - (r - s')}{N - q - p - i - r + s'}\right) \\
&\overset{(7)}{\geq} \quad \prod_{i=0}^{s''+t-1} \left(1 - \frac{2(\alpha + \beta)}{N}\right) \\
&\overset{(8)}{\geq} \quad \prod_{i=0}^{s''+t-1} \left(1 - \frac{2\sqrt{q}}{N}\right) \overset{(9)}{\geq} \left(1 - \frac{2q^{3/2}}{N}\right), \quad (53)
\end{aligned}
$$

where (7) follows as $q + p + i + r - s' \leq N/2$, (8) follows as $\alpha, \beta \leq q$ and (9) follows as $s'' + t \leq q$.

COMPUTING E.3.2.

$$
\begin{aligned}
\mathsf{E.3.2} \quad &= \quad \frac{(N - q - p - \alpha - \beta - 3t - r)_{s''}}{(N - q - p - \alpha - \beta - s' - t)_{s''}} \\
&= \quad \prod_{i=0}^{s''-1} \left(1 - \frac{2t + r - s'}{N - q - p - \alpha - \beta - s' - t - i}\right) \overset{(10)}{\geq} \left(1 - \frac{2s''(2M + r - s')}{N}\right) \\
&\overset{(11)}{\geq} \quad \left(1 - \frac{6s''M}{N}\right) \overset{(12)}{\geq} \left(1 - \frac{6q^2}{N^{4/3}}\right), \quad (54)
\end{aligned}
$$

where (10) follows as $t \leq M$ and $q + p + \alpha + \beta + s' + t + i \leq N/2$. (11) follows as $r - s' \leq M$ and (12) follows as $s'' \leq q$ and $M = q/N^{1/3}$. Therefore, from Eqn. (53) and Eqn. (54) and by using the inequality $(1-a)(1-b) \geq (1-a-b)$ for $a, b \leq 1$, we have

$$\mathsf{E.3} \geq \left(1 - \frac{2q^{3/2}}{N} - \frac{6q^2}{N^{4/3}}\right). \tag{55}$$

Now by combining Eqn. (51), Eqn. (52) and Eqn. (55) and by using the inequality $(1-a)(1-b) \geq (1-a-b)$ for $a, b \leq 1$, we have

$$\mathsf{D.2} \geq \left(1 - \frac{8q}{N^{2/3}} - \frac{2q^{3/2}}{N} - \frac{6q^2}{N^{4/3}}\right) \cdot \sum_{0 \leq t \leq M} \mathsf{E.2}. \tag{56}$$

Note that, for a fixed t, $\mathsf{E.2} = \mathbf{Hyp}(t)_{N', s', s'}$, where $N' = N - q - p - \alpha - \beta$ with parameters N', s' and s'. It is a folklore result that the expectation of the hypergeometric distribution with parameters N', s', s' is s'^2/N'. Therefore, we have

$$\mathsf{D.2} \geq \left(1 - \frac{8q}{N^{2/3}} - \frac{2q^{3/2}}{N} - \frac{6q^2}{N^{4/3}}\right) \cdot \sum_{0 \leq t \leq M} \mathbf{Hyp}(t)_{N', s', s'}.$$

Now, using Markov's inequality, we have

$$\sum_{t > M} \mathbf{Hyp}(t)_{N', s', s'} \leq \frac{s'^2}{qN^{2/3}} \leq \frac{q}{N^{2/3}}, \tag{57}$$

where the first inequality appears due to the Markov's inequality and the second inequaluty follows as $s' \leq q$ and $M = q/N^{1/3}$. Therefore, using Eqn. (56) and Eqn. (57) and by using the inequality $(1-a)(1-b) \geq (1-a-b)$ for $a, b \leq 1$, we have

$$\mathsf{D.2} \geq \left(1 - \frac{9q}{N^{2/3}} - \frac{2q^{3/2}}{N} - \frac{6q^2}{N^{4/3}}\right). \tag{58}$$

Finally, by combining Eqn. (49), Eqn. (50) and Eqn. (58) and by using the inequality $(1-a)(1-b) \geq (1-a-b)$ for $a, b \leq 1$, we have

$$\rho(\tau) \geq 1 - \underbrace{\left(\frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{13q}{N^{2/3}} + \frac{2pt}{q^2} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{28pq^2}{N^2} + \frac{4p^2q}{N^2} + \frac{24q^3}{N^2} + \frac{2q^{3/2}}{N}\right)}_{\phi(\tau)}.$$

This completes the proof of Lemma 5. Now it only remains to compute the expectation of $\phi(\tau)$ as follows:

COMPUTING THE EXPECTATION. We now compute the expectation of $\phi(\tau)$ over the randomness of the permutation P as follows:

$$\begin{aligned}
\mathbf{E}_{\mathsf{P}}[\phi(\tau)] &= \left(\frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{13q}{N^{2/3}} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{28pq^2}{N^2} + \frac{4p^2q}{N^2} + \frac{24q^3}{N^2} + \frac{2q^{3/2}}{N}\right) + \mathbf{E}_{\mathsf{P}}\left[\frac{2pt}{q^2}\right] \\
&= \left(\frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{13q}{N^{2/3}} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{28pq^2}{N^2} + \frac{4p^2q}{N^2} + \frac{24q^3}{N^2} + \frac{2q^{3/2}}{N}\right) + \frac{2p}{q^2}\mathbf{E}_{\mathsf{P}}[t].
\end{aligned}$$

Now, it remains to compute the expectation of the random variable t over the randomness of the permutation P. Let t_i be the indicator random variable that takes the value 1 if $\mathsf{P}(x_i \oplus k_1) \oplus x_i \oplus k_2 \in \mathcal{X}$, for $1 \leq i \leq M$. Therefore, it is easy to see that

$$\Pr[t_i = 1] = \Pr[\mathsf{P}(x_i \oplus k_1) \oplus x_i \oplus k_2 \in \mathcal{X}] \leq \frac{q'}{N}.$$

Since, $t = t_1 + \ldots + t_M$, due to the linearity of expectation, we have

$$\mathbf{E}_{\mathsf{P}}[t] = \sum_{i=1}^{M} \mathbf{E}_{\mathsf{P}}[t_i] = \sum_{i=1}^{M} \Pr[t_i = 1] \leq \frac{q'M}{N} \leq \frac{q^2}{N^{4/3}}, \tag{59}$$

where the last inequality appears as $M = q/N^{1/3}$ and $q' \leq q$. Therefore, from Eqn. (59), we have

$$\mathbf{E}_{\mathsf{P}}[\phi(\tau)] \leq \left(\frac{12q^2}{N^{4/3}} + \frac{2pq}{N^{4/3}} + \frac{13q}{N^{2/3}} + \frac{2\sqrt{q}}{N^{1/3}} + \frac{28pq^2}{N^2} + \frac{4p^2q}{N^2} + \frac{24q^3}{N^2} + \frac{2q^{3/2}}{N} + \frac{2p}{N^{4/3}} \right). \tag{60}$$

The result of Theorem 1 follows from Theorem 2, Lemma 4 and Eqn. (60) which concludes the proof of the security result.

6 Conclusion and Future Works

This paper has proposed an inverse free single permutation based beyond the birthday bound secure PRF that requires $2n$ bit keys. One could also achieve the same goal using the single permutation based tweakable Even Mansour cipher [Dut20]. However, the solution comes at the cost of implementing the costly universal hash functions. Moreover, parallel modes like nEHtM$_p$, SoEM22 or DS-SoEM also achieves the beyond birthday bound PRF security, but again the former one requires to implement a universal hash function, SoEM22 requires two independent permutations and DS-SoEM takes $n - 1$ bit input. However, it would be interesting to study the sequential design of an inverse free single permutation based PRF with only n bit key. We believe that pEDM can be turned to a single permutation oriented beyond the birthday bound secure nonce based MAC by xoring an almost-xor universal hash function in between the two permutation calls (similar to the flavour of EWCDM [CS16]).

Acknowledgments

References

[Bab02]   László Babai. The fourier transform and equations over finite abelian groups (lec- ture notes, version 1.3). 2002.

[BCDM19] Tim Beyne, Yu Long Chen, Christoph Dobraunig, and Bart Mennink. Elephant. *NIST LWC*, 2019.

[BDD+17] Ritam Bhaumik, Nilanjan Datta, Avijit Dutta, Nicky Mouha, and Mridul Nandi. The iterated random function problem. In Tsuyoshi Takagi and Thomas Peyrin, editors, *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part II*, volume 10625 of *Lecture Notes in Computer Science*, pages 667–697. Springer, 2017.

[BDLN20] Arghya Bhattarcharjee, Avijit Dutta, Eik List, and Mridul Nandi. CENCPP - beyond-birthday-secure encryption from public permutations. *IACR Cryptol. ePrint Arch.*, 2020:602, 2020.

[BDP+16]   Guido Bertoni, Joan Daemen, Michaël Peeters, Gilles Van Assche, and
            Ronny Van Keer. Farfalle: parallel permutation-based cryptography. *IACR
            Cryptol. ePrint Arch.*, 2016:1188, 2016.

[BDPA13]   Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. Kec-
            cak. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual Interna-
            tional Conference on the Theory and Applications of Cryptographic Techniques,
            Athens, Greece, May 26-30, 2013. Proceedings*, pages 313–314, 2013.

[BI99]   M. Bellare and R. Impagliazzo. A tool for obtaining tighter security anal-
            yses of pseudorandom function based constructions, with applications to
            prp to prf conversion. Cryptology ePrint Archive, Report 1999/024, 1999.
            http://eprint.iacr.org/1999/024.

[BKL+07]   Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel
            Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe.
            PRESENT: an ultra-lightweight block cipher. In *CHES 2007, Proceedings*,
            pages 450–466, 2007.

[BKL+13]   Andrey Bogdanov, Miroslav Knezevic, Gregor Leander, Deniz Toz, Kerem
            Varici, and Ingrid Verbauwhede. SPONGENT: the design space of lightweight
            cryptographic hashing. *IEEE Trans. Computers*, 62(10):2041–2053, 2013.

[BKL+17]   Daniel J. Bernstein, Stefan Kölbl, Stefan Lucks, Pedro Maat Costa Massolino,
            Florian Mendel, Kashif Nawaz, Tobias Schneider, Peter Schwabe, François-
            Xavier Standaert, Yosuke Todo, and Benoît Viguier. Gimli : A cross-platform
            permutation. In *Cryptographic Hardware and Embedded Systems - CHES
            2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017,
            Proceedings*, pages 299–320, 2017.

[BKR94]   Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of cipher block
            chaining. In Yvo Desmedt, editor, *Advances in Cryptology - CRYPTO '94, 14th
            Annual International Cryptology Conference, Santa Barbara, California, USA,
            August 21-25, 1994, Proceedings*, volume 839 of *Lecture Notes in Computer
            Science*, pages 341–358. Springer, 1994.

[BKR98]   Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-rackoff backwards:
            Increasing security by making block ciphers non-invertible. In *EUROCRYPT
            '98, Proceeding.*, pages 266–280, 1998.

[BN16]   Ritam Bhaumik and Mridul Nandi. Olef: an inverse-free online cipher. an online
            SPRP with an optimal inverse-free construction. *IACR Trans. Symmetric
            Cryptol.*, 2016(2):30–51, 2016.

[BPP+17]   Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki,
            Siang Meng Sim, and Yosuke Todo. GIFT: A small present - towards reaching
            the limit of lightweight encryption. In Wieland Fischer and Naofumi Homma,
            editors, *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th
            International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*,
            volume 10529 of *Lecture Notes in Computer Science*, pages 321–345. Springer,
            2017.

[BR02]   John Black and Phillip Rogaway. A block-cipher mode of operation for
            parallelizable message authentication. In Lars R. Knudsen, editor, *Advances in
            Cryptology - EUROCRYPT 2002, International Conference on the Theory and
            Applications of Cryptographic Techniques, Amsterdam, The Netherlands, April*

28 - May 2, 2002, Proceedings, volume 2332 of *Lecture Notes in Computer Science*, pages 384–397. Springer, 2002.

[BR06]   Mihir Bellare and Phillip Rogaway. The security of triple encryption and a framework for code-based game-playing proofs. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, volume 4004 of *Lecture Notes in Computer Science*, pages 409–426. Springer, 2006.

[CDNY18]   Avik Chakraborti, Nilanjan Datta, Mridul Nandi, and Kan Yasuda. Beetle family of lightweight and secure authenticated encryption ciphers. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(2):218–241, 2018.

[CLL+14]   Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the two-round even-mansour cipher. In *Advances in Cryptology - CRYPTO 2014,*, pages 39–56, 2014.

[CLM19a]   Yu Long Chen, Eran Lambooij, and Bart Mennink. How to build pseudorandom functions from public random permutations. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, pages 266–293, 2019.

[CLM19b]   Yu Long Chen, Eran Lambooij, and Bart Mennink. How to build pseudorandom functions from public random permutations. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, pages 266–293, 2019.

[CN08]   Donghoon Chang and Mridul Nandi. A short proof of the PRP/PRF switching lemma. *IACR Cryptol. ePrint Arch.*, 2008:78, 2008.

[CN19]   Bishwajit Chakraborty and Mridul Nandi. Orange. *NIST LWC*, 2019.

[CNTY20]   Avik Chakraborti, Mridul Nandi, Suprita Talnikar, and Kan Yasuda. On the composition of single-keyed tweakable even-mansour for achieving BBB security. *IACR Trans. Symmetric Cryptol.*, 2020(2):1–39, 2020.

[CS16]   Benoît Cogliati and Yannick Seurin. EWCDM: an efficient, beyond-birthday secure, nonce-misuse resistant MAC. In *CRYPTO 2016, Proceedings, Part I*, pages 121–149, 2016.

[CS18]   Benoît Cogliati and Yannick Seurin. Analysis of the single-permutation encrypted davies-meyer construction. *Des. Codes Cryptogr.*, 86(12):2703–2723, 2018.

[DDNP18]   Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Goutam Paul. Double-block hash-then-sum: A paradigm for constructing bbb secure prf. *IACR Transactions on Symmetric Cryptology*, 2018(3):36–92, 2018.

[DDNY18]   Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Kan Yasuda. Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based MAC. In *Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part I*, pages 631–661, 2018.

[DEM+20]   Christoph Dobraunig, Maria Eichlseder, Stefan Mangard, Florian Mendel, Bart
           Mennink, Robert Primas, and Thomas Unterluggauer. Isap v2.0. *IACR Trans.
           Symmetric Cryptol.*, 2020(S1):390–416, 2020.

[DEMS19]   Christoph Dobraunig, Maria Eichlseder, Florian Mendel, and Martin Schläffer.
           Ascon v1.2. *NIST LWC*, 2019.

[DHP+20]   Joan Daemen, Seth Hoffert, Michaël Peeters, Gilles Van Assche, and Ronny Van
           Keer. Xoodyak, a lightweight cryptographic scheme. *IACR Trans. Symmetric
           Cryptol.*, 2020(S1):60–87, 2020.

[DHT17]   Wei Dai, Viet Tung Hoang, and Stefano Tessaro. Information-theoretic in-
           distinguishability via the chi-squared method. In *Advances in Cryptology -
           CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa
           Barbara, CA, USA, August 20-24, 2017, Proceedings, Part III*, pages 497–523,
           2017.

[DMA17]   Joan Daemen, Bart Mennink, and Gilles Van Assche. Full-state keyed duplex
           with built-in multi-user support. In Tsuyoshi Takagi and Thomas Peyrin,
           editors, *Advances in Cryptology - ASIACRYPT 2017 - 23rd International
           Conference on the Theory and Applications of Cryptology and Information
           Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part II*, volume
           10625 of *Lecture Notes in Computer Science*, pages 606–637. Springer, 2017.

[DN20]   Avijit Dutta and Mridul Nandi. BBB secure nonce based MAC using public
           permutations. In Abderrahmane Nitaj and Amr M. Youssef, editors, *Progress
           in Cryptology - AFRICACRYPT 2020 - 12th International Conference on
           Cryptology in Africa, Cairo, Egypt, July 20-22, 2020, Proceedings*, volume
           12174 of *Lecture Notes in Computer Science*, pages 172–191. Springer, 2020.

[DNS20]   Avijit Dutta, Mridul Nandi, and Abishanka Saha. Proof of mirror theory for
           $\xi_{\max}=2$. *IACR Cryptol. ePrint Arch.*, 2020:669, 2020.

[Dut20]   Avijit Dutta. Minimizing the two-round tweakable even-mansour cipher. *IACR
           Cryptol. ePrint Arch.*, 2020:1076, 2020.

[EM97]   Shimon Even and Yishay Mansour. A construction of a cipher from a single
           pseudorandom permutation. *J. Cryptology*, 10(3):151–162, 1997.

[GG16]   Shoni Gilboa and Shay Gueron. The advantage of truncated permutations.
           CoRR, abs/1610.02518, 2016.

[GM20]   Aldo Gunsing and Bart Mennink. The summation-truncation hybrid: Reusing
           discarded bits for free. In Daniele Micciancio and Thomas Ristenpart, editors,
           *Advances in Cryptology - CRYPTO 2020 - 40th Annual International Cryp-
           tology Conference, CRYPTO 2020, Santa Barbara, CA, USA, August 17-21,
           2020, Proceedings, Part I*, volume 12170 of *Lecture Notes in Computer Science*,
           pages 187–217. Springer, 2020.

[GPP11]   Jian Guo, Thomas Peyrin, and Axel Poschmann. The PHOTON family of
           lightweight hash functions. In *Advances in Cryptology - CRYPTO 2011 - 31st
           Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011.
           Proceedings*, pages 222–239, 2011.

[GSWG19]   Chun Guo, Yaobin Shen, Lei Wang, and Dawu Gu. Beyond-birthday secure
           domain-preserving prfs from a single permutation. *Des. Codes Cryptogr.*,
           87(6):1297–1322, 2019.

[HT16]   Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length
            extension: Exact bounds and multi-user security. In *Advances in Cryptology
            - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa
            Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 3–32, 2016.

[HWKS98]   Chris Hall, David A. Wagner, John Kelsey, and Bruce Schneier. Building prfs
            from prps. In *CRYPTO 1998, Proceedings*, pages 370–389, 1998.

[Iwa06]   Tetsu Iwata. New blockcipher modes of operation with beyond the birthday
            bound security. In *Fast Software Encryption, 13th International Workshop,
            FSE 2006, Graz, Austria, March 15-17, 2006, Revised Selected Papers*, pages
            310–327, 2006.

[LPTY16]   Atul Luykx, Bart Preneel, Elmar Tischhauser, and Kan Yasuda. A MAC mode
            for lightweight block ciphers. In *Fast Software Encryption - 23rd International
            Conference, FSE 2016, Bochum, Germany, March 20-23, 2016, Revised Selected
            Papers*, pages 43–59, 2016.

[LR88]   Michael Luby and Charles Rackoff. How to construct pseudorandom per-
            mutations from pseudorandom functions. *SIAM J. Comput.*, 17(2):373–386,
            1988.

[Luc00]   Stefan Lucks. The sum of prps is a secure PRF. In Bart Preneel, editor,
            *Advances in Cryptology - EUROCRYPT 2000, International Conference on the
            Theory and Application of Cryptographic Techniques, Bruges, Belgium, May
            14-18, 2000, Proceeding*, volume 1807 of *Lecture Notes in Computer Science*,
            pages 470–484. Springer, 2000.

[MN17a]   Bart Mennink and Samuel Neves. Encrypted davies-meyer and its dual:
            Towards optimal security using mirror theory. In *Advances in Cryptology -
            CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa
            Barbara, CA, USA, August 20-24, 2017, Proceedings, Part III*, pages 556–583,
            2017.

[MN17b]   Bart Mennink and Samuel Neves. Optimal prfs from blockcipher designs.
            IACR Trans. Symmetric Cryptol., 2017(3):228–252, 2017.

[MV04]   David A. McGrew and John Viega. The security and performance of the
            galois/counter mode (GCM) of operation. In *Progress in Cryptology - IN-
            DOCRYPT 2004, 5th International Conference on Cryptology in India, Chen-
            nai, India, December 20-22, 2004, Proceedings*, pages 343–355, 2004.

[Nan20]   Mridul Nandi. Mind the composition: Birthday bound attacks on EWCDMD
            and sokac21. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryp-
            tology - EUROCRYPT 2020 - 39th Annual International Conference on the
            Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May
            10-14, 2020, Proceedings, Part I*, volume 12105 of *Lecture Notes in Computer
            Science*, pages 203–220. Springer, 2020.

[NIS18]   NIST. Lightweight cryptography, 2018. Online: https://csrc.nist.gov/
            Projects/Lightweight-Cryptography. Accessed: August 01, 2019.

[NMSS19]   Yusuke Naito, Mitsuru Matsui, Takeshi Sugawara, and Daisuke Suzuki. SAEB:
            A lightweight blockcipher-based AEAD mode of operation. *IACR Cryptol.
            ePrint Arch.*, 2019:700, 2019.

[RBB03] Phillip Rogaway, Mihir Bellare, and John Black. Sha-3 standard. *ACM Transactions on Information and System Security (TISSEC)*, 6(3):365–403, 2003.

[WC81] Mark N. Wegman and Larry Carter. New hash functions and their use in authentication and set equality. *J. Comput. Syst. Sci.*, 22(3):265–279, 1981.

Supplementary Materials

7 Proof of Lemma 1

To prove the lemma, let us first recall some useful results on Fourier analysis on $\mathbb{Z}_2^n$.

NOTATION: For a given subset $\mathcal{S} \subset \{0,1\}^n$, we denote $\mathbb{I}_\mathcal{S} : \{0,1\}^n \to \{0,1\}$ to be the characteristic function of the set $\mathcal{S}$ such that $\mathbb{I}_\mathcal{S}(s) = 1$ if and only if $s \in \mathcal{S}$. Given two real-valued functions $f, g : \{0,1\}^n \to \mathbb{R}$, we denote the inner product of f and g as

$$\langle f, g \rangle = \mathbf{E}[fg] = \frac{1}{N} \sum_{x \in \{0,1\}^n} f(x)g(x).$$

The convolution of f and g is denoted as

$$(f \star g)(x) = \sum_{y \in \{0,1\}^n} f(y)g(x \oplus y), \forall x \in \{0,1\}^n.$$

For a given $\alpha \in \{0,1\}^n$, the character associated with α is denoted as $\chi_\alpha : \{0,1\}^n \to \{+1, -1\}$, defined as $\chi_\alpha(x) = (-1)^{\alpha \cdot x}$. χ_0 is called the principal character and all other $\chi \neq 1$ for $\alpha \neq 0$ are called non-principal character. Now, given a real-valued function $f : \{0,1\}^n \to \mathbb{R}$ and for $\alpha \in \{0,1\}^n$, we define the fourier coefficient of f corresponding to α is

$$\widehat{f}(\alpha) \stackrel{\Delta}{=} \langle f, \chi_\alpha \rangle = \frac{1}{N} \sum_{x \in \{0,1\}^n} f(x)(-1)^{\alpha \cdot x}.$$

The coefficient corresponding to $\alpha = 0$ is called the principal fourier coefficient and all other coefficients are called non-principal fourier coefficients. Note that the principal fourier coefficient for a characteristic function $\mathbb{I}_\mathcal{S}$ of a set $\mathcal{S}$ is

$$\widehat{\mathbb{I}}_\mathcal{S}(0) = \frac{|\mathcal{S}|}{N}.$$

Having defined the neccessary notations, we now recall three important results on fourier analysis holds for any functions $f, g : \{0,1\}^n \to \mathbb{R}$, any $\alpha \in \{0,1\}^n$ and any $\mathcal{S} \subseteq \{0,1\}^n$ as follows:

$$\sum_{x \in \{0,1\}^n} f(x)g(x) = N \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)\widehat{g}(\alpha), \tag{61}$$

$$\widehat{(f \star g)}(\alpha) = N\widehat{f}(\alpha)\widehat{g}(\alpha), \tag{62}$$

$$\sum_{\alpha \in \{0,1\}^n} |\widehat{\mathbb{I}}_\mathcal{S}(\alpha)|^2 = \frac{|\mathcal{S}|}{N}. \tag{63}$$

Now, we define the following two parameters associated to $\mathcal{Q}$:

$$\Phi_{\alpha,\beta}(\mathcal{Q}) \stackrel{\Delta}{=} N^2|\widehat{\mathbb{I}}_\mathcal{Q}(\alpha,\beta)| = |\sum_{(x,y) \in \mathcal{Q}} (-1)^{\alpha \cdot x \oplus \beta \cdot y}|,$$

$$\Phi(\mathcal{Q}) \stackrel{\Delta}{=} \max_{\alpha \neq 0, \beta \neq 0} \Phi_{\alpha,\beta}(\mathcal{Q}).$$

Now, to bound $\mu(\mathcal{Q}, \mathcal{U}, \mathcal{V})$, let us consider the following two sets: $\mathcal{N} = \mathcal{U} \times \mathcal{V} = \{(u, v) : u \in \mathcal{U}, v \in \mathcal{V}\}$ and $\mathcal{K} = \{(k, k) : k \in \{0, 1\}^n\}$. Note that, $((x, y), u, v) \in \mathcal{Q} \times \mathcal{U} \times \mathcal{V}$ if and only if $\exists k \in \{0, 1\}^n$ such that

$$(x, y) \oplus (u, v) = (k, k).$$

Therefore, we have

$$
\begin{aligned}
\mu(\mathcal{Q}, \mathcal{U}, \mathcal{V}) \quad &= \sum_{(x,y) \in (\{0,1\}^n)^2} \sum_{(u,v) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{Q}}(x, y) \mathbb{I}_{\mathcal{N}}(u, v) \mathbb{I}_{\mathcal{K}}(x \oplus u, y \oplus v) \\
&= \sum_{(x,y) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{Q}}(x, y) \sum_{(u,v) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{N}}(u, v) \mathbb{I}_{\mathcal{K}}(x \oplus u, y \oplus v) \\
&= \sum_{(x,y) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{Q}}(x, y)(\mathbb{I}_{\mathcal{N}} \star \mathbb{I}_{\mathcal{K}})(x, y) \\
&\overset{(1)}{=} N^2 \sum_{(\alpha,\beta) \in (\{0,1\}^n)^2} \widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \beta)(\widehat{\mathbb{I}_{\mathcal{N}} \star \mathbb{I}_{\mathcal{K}}})(\alpha, \beta) \\
&\overset{(2)}{=} N^4 \sum_{(\alpha,\beta) \in (\{0,1\}^n)^2} \widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta) \\
&\overset{(3)}{=} N^4 \frac{|\mathcal{Q}|}{N^2} \frac{|\mathcal{N}|}{N^2} \frac{|\mathcal{K}|}{N^2} + N^4 \sum_{(\alpha,\beta) \neq (0,0)} \widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta) \\
&\overset{(4)}{=} \frac{q|\mathcal{U}||\mathcal{V}|}{N} + N^4 \sum_{(\alpha,\beta) \neq (0,0)} \widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta) \widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta), \quad\quad (64)
\end{aligned}
$$

where (1) follows from Eqn. (61), (2) follows from Eqn. (62), (3) follows from separating the principal fourier coefficients from non-principal fourier coefficients and (4) follows from the cardinality of $\mathcal{Q}, \mathcal{N}$ and $\mathcal{K}$.

COMPUTING $\widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta)$. To compute $\widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta)$, we have the following:

$$
\begin{aligned}
\widehat{\mathbb{I}}_{\mathcal{N}}(\alpha, \beta) \quad &= \frac{1}{N^2} \sum_{(u,v) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{N}}(u, v)(-1)^{\alpha \cdot u \oplus \beta \cdot v} \\
&= \frac{1}{N^2} \sum_{(u,v) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{U}}(u) \mathbb{I}_{\mathcal{U}}(u)(-1)^{\alpha \cdot u \oplus \beta \cdot v} \\
&= \frac{1}{N^2} \left(\sum_{u \in (\{0,1\}^n)} \mathbb{I}_{\mathcal{U}}(u)(-1)^{\alpha \cdot u} \right) \left(\sum_{v \in (\{0,1\}^n)} \mathbb{I}_{\mathcal{V}}(v)(-1)^{\beta \cdot v} \right) \\
&= \widehat{\mathbb{I}}_{\mathcal{U}}(\alpha) \widehat{\mathbb{I}}_{\mathcal{V}}(\beta). \quad\quad (65)
\end{aligned}
$$

COMPUTING $\widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta)$. To compute $\widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta)$, we have the following:

$$
\begin{aligned}
\widehat{\mathbb{I}}_{\mathcal{K}}(\alpha, \beta) \quad &= \frac{1}{N^2} \sum_{(x,y) \in (\{0,1\}^n)^2} \mathbb{I}_{\mathcal{K}}(x, y)(-1)^{\alpha \cdot x \oplus \beta \cdot y} \\
&= \frac{1}{N^2} \sum_{y \in (\{0,1\}^n)} (-1)^{\alpha \cdot y \oplus \beta \cdot y}. \quad\quad (66)
\end{aligned}
$$

Note that, Eqn. (66) is evaluated to 0 if $\beta = \alpha$. Therefore, from Eqn. (64), Eqn. (65) and

Eqn. (66), we have

$$
\begin{aligned}
\mu(\mathcal{Q}, \mathcal{U}, \mathcal{V}) &= \frac{q|\mathcal{U}||\mathcal{V}|}{N} + N^3 \sum_{\alpha \neq 0} \widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \alpha) \widehat{\mathbb{I}}_{\mathcal{U}}(\alpha) \widehat{\mathbb{I}}_{\mathcal{V}}(\alpha) \\
&\leq \frac{q|\mathcal{U}||\mathcal{V}|}{N} + N^3 \sum_{\alpha \neq 0} |\widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \alpha)| \cdot |\widehat{\mathbb{I}}_{\mathcal{U}}(\alpha)| \cdot |\widehat{\mathbb{I}}_{\mathcal{V}}(\alpha)| \\
&\leq \frac{q|\mathcal{U}||\mathcal{V}|}{N} + N\Phi(\mathcal{Q}) \sum_{\alpha \neq 0} |\widehat{\mathbb{I}}_{\mathcal{U}}(\alpha)| \cdot |\widehat{\mathbb{I}}_{\mathcal{V}}(\alpha)|, \quad (67)
\end{aligned}
$$

where the last inequality follows from the definition of $\Phi(\mathcal{Q})$ that $|\widehat{\mathbb{I}}_{\mathcal{Q}}(\alpha, \alpha)| \leq \Phi(\mathcal{Q})/N^2$. Moreover, by using the Cauchy-Schwartz inequality we have,

$$
\sum_{\alpha \neq 0} |\widehat{\mathbb{I}}_{\mathcal{U}}(\alpha)| \cdot |\widehat{\mathbb{I}}_{\mathcal{V}}(\alpha)| \leq \sqrt{\sum_{\alpha \neq 0} |\widehat{\mathbb{I}}_{\mathcal{U}}(\alpha)|^2} \cdot \sqrt{\sum_{\alpha \neq 0} |\widehat{\mathbb{I}}_{\mathcal{V}}(\beta)|^2} \leq \frac{1}{N}\sqrt{|\mathcal{U}||\mathcal{V}|}. \quad (68)
$$

By plug-in the Eqn. (68) into Eqn. (66) we have,

$$
\mu(\mathcal{Q}, \mathcal{U}, \mathcal{V}) \leq \frac{q|\mathcal{U}||\mathcal{V}|}{N} + \Phi(\mathcal{Q})\sqrt{|\mathcal{U}||\mathcal{V}|}. \quad (69)
$$

From Lemma 5 of [CLL$^+$14] we know that if $A_1, \ldots, A_q$ is a sequence of random variables taking values in $\{+1, -1\}$ such that for all $i \in [q]$ and all $(a_1, \ldots, a_{i-1}) \in (\{+1, -1\})^{i-1}$ the following holds:

$$
\Pr[A_i = 1 \mid (A_1, \ldots, A_{i-1}) = (a_1, \ldots, a_{i-1})] \leq \frac{1}{2} + \epsilon,
$$

for some $\epsilon \in [0, 1/2]$. Then for any $\delta \in [0, 1]$, one has

$$
\Pr[\sum_{i=1}^{q} A_i \geq q(2\epsilon + \delta)] \leq e^{\frac{q\delta^2}{12}}.
$$

Now, using Lemma 5 of [CLL$^+$14] we claim the following:

Claim 1. *Assume $9n \leq q \leq N/2$, let D be some probabilitistic distinguisher that makes q adaptive queries to RF. Let $\mathcal{Q} = ((x_1, y_1), \ldots, (x_q, y_q))$ denotes the transcript of the interaction with RF to D. Then*

$$
\Pr_{\mathsf{RF}, \omega}[\Phi(\mathcal{Q}) \geq 3\sqrt{nq}] \leq \frac{2}{N},
$$

where the probability is defined over the randomness of RF and the random coin ω of the distinguisher D.

Proof. Proof of this claim is similar to that of Lemma 6 of [CLL$^+$14]. We define the random variables $A_i = (-1)^{\alpha \cdot x_i \oplus \beta \cdot y_i}$ where $(x_i, y_i) \in \mathcal{Q}$. Then we have, $|(A_1 + A_2 + \ldots + A_q)| = \Phi_{\alpha, \beta}(\mathcal{Q})$. Now, for the i-th query with input x_i, the output y_i is a uniform random variable over a set of size N. Moreover, once x_i is fixed, there are exactly $N/2$ y_i's such that $A_i = 1$ as $\beta \neq 0$. Therefore,

$$
\Pr[A_i = 1 \mid (A_1, \ldots, A_{i-1}) = (a_1, \ldots, a_{i-1})] = \frac{N/2}{N} = \frac{1}{2}.
$$

Hence, $\epsilon = 0$ and therefore,

$$
\Pr[\sum_{i=1}^{q} A_i \geq q\delta] \leq e^{-\frac{q\delta^2}{12}}.
$$

Using a similar reasoning and by setting $A_i' = -A_i$, we have

$$\Pr[\sum_{i=1}^{q} A_i \leq -q\delta] \leq e^{-\frac{q\delta^2}{12}}.$$

By combining these two equations, we have

$$\Pr[\Phi(\mathcal{Q}) \geq q\delta] \leq 2e^{-\frac{q\delta^2}{12}}.$$

The result follows by setting $\delta = \sqrt{12 \ln N / q}$. This makes $q \geq 9n$ which implies $\delta \leq 1$ and $\sqrt{12 \ln 2} \leq 3$.   □