Noname manuscript No.
(will be inserted by the editor)

Analysis of the Single-Permutation Encrypted Davies-Meyer Construction

Benoît Cogliati · Yannick Seurin

Received: date / Accepted: date

Abstract We consider the so-called *Encrypted Davies-Meyer* (EDM) construction, which turns a permutation P on $\{0,1\}^n$ into a function from $\{0,1\}^n$ to $\{0,1\}^n$ defined as $P(P(x) \oplus x)$. A similar construction using two independent permutations, namely $P'(P(x) \oplus x)$, was previously analyzed by Cogliati and Seurin (CRYPTO 2016) who showed that when P and P' are secret and random, then any black-box adversary needs at least roughly $2^{2n/3}$ queries to distinguish the construction from a uniformly random function from $\{0,1\}^n$ to $\{0,1\}^n$. In this paper, we focus on the single-permutation variant of the construction. Our main result is that the PRF-security of the single-permutation EDM construction is also (at least) roughly $2^{2n/3}$, in the sense that any black-box adversary needs at least this number of queries to distinguish the construction from a uniformly random function. This yields the first PRP-to-PRF conversion method which uses a single permutation, does not shrink the original domain nor range of the permutation, and provides security beyond the birthday bound.

Keywords block cipher · pseudorandomness · PRP-to-PRF conversion · beyond-birthday-bound security

Mathematics Subject Classification (2000) 94A60

This is the final version submitted by the authors to Designs, Codes and Cryptography. The final publication is available at http://link.springer.com/article/10.1007/s10623-018-0470-9.

B. Cogliati
University of Versailles, France
E-mail: benoitcogliati@hotmail.fr

Y. Seurin
ANSSI, Paris, France
E-mail: yannick.seurin@m4x.org

1 Introduction

PRP-to-PRF Conversion. Block ciphers are ubiquitous in cryptography. A block cipher takes a key k from some key space $\mathcal{K}$ and a plaintext x from some domain $\mathcal{X}$ and returns a ciphertext $y \in \mathcal{X}$. For each key k, the mapping $E_k : x \mapsto E(k, x)$ is an (efficiently invertible given the key k) permutation of $\mathcal{X}$. The standard security notion for a block cipher is that it should be a pseudorandom permutation (PRP) [GGM86], which means that no adversary having black-box access to some permutation of $\mathcal{X}$ and with limited resources (oracle queries and computation time) should be able to distinguish with noticeable advantage whether it is interacting with the block cipher under a random key k or with a uniformly random permutation of $\mathcal{X}$.

Even though invertibility might seem like a basic requirement when using a block cipher for encryption, this intuition turns out wrong for many *modes of operation*. Consider for example encryption using counter mode [BDJR97] based on a block cipher with domain $\mathcal{X} = \{0,1\}^n$: the message to be encrypted is split into n-bit message blocks m_i which are encrypted as $y_i = m_i \oplus E_k(c_i)$, where c_i is some non-repeating counter. It is well known that this is only secure as long as at most $2^{n/2}$ message blocks are encrypted under the same key. After that point, the ciphertexts can be distinguished from random by the adversary (indeed, for uniformly random y_i's, the adversary expects to see collisions among values $y_i \oplus m_i$, whereas this cannot happen for outputs of a real encryption oracle since $y_i \oplus m_i = E_k(c_i)$, where the c_i's are non-repeating). On the other hand, this problem vanishes if instead of using a PRP, one uses a pseudorandom *function* (PRF). As a PRP, a PRF takes as input a key $k \in \mathcal{K}$ and a plaintext $x \in \mathcal{X}$, and returns a "ciphertext" y in some range $\mathcal{Y}$ (which might be in general different from domain $\mathcal{X}$). The security requirement is now that no adversary with oracle access to some function from $\mathcal{X}$ to $\mathcal{Y}$ should be able to distinguish whether it interacts with $F_k = F(k, \cdot)$ for some random key k or with a uniformly random function from $\mathcal{X}$ to $\mathcal{Y}$ (as opposed to a random permutation of $\mathcal{X}$ for a PRP). It is easy to see that using a PRF in counter mode (i.e., encrypting message blocks as $y_i = m_i \oplus F_k(c_i)$) yields a security-preserving encryption mode, in the sense that the advantage of any adversary against the encryption mode is upper bounded by the advantage against the PRF itself (the mode itself does not incur any security loss, unlike when using a PRP).

Another example is the Wegman-Carter MAC construction [WC81], which relies on a PRF F and an almost-xor universal (AXU) hash function H to construct a nonce-based message authentication code defined as

$$\mathsf{WC}[F, H](\nu, m) = F_k(\nu) \oplus H_{k'}(m),$$

where ν is the nonce (a value which should never repeat) and m is the message to be authenticated. The Wegman-Carter construction enjoys a very strong security bound when used with a good PRF and a good AXU hash function (for n-bit tags, the forgery probability can be close to $q_f/2^n$, where q_f is the number of forgery attempts of the adversary, plus a term related to the PRF-security

of F). However, if F is replaced by a block cipher, the provable security bound drops to birthday bound [Sho96, Ber05].

These two examples show that invertibility can become a liability in many constructions based on block ciphers. Unfortunately, cryptographers have focused on designing good block ciphers, and efficient and secure PRFs are not readily available. Hence, a natural question is whether it is possible to turn a PRP E into a PRF $F[E]$ as efficiently as possible and in a security-preserving way, meaning that the PRF-advantage of any adversary A against $F[E]$ should be close to the PRP-advantage of a related adversary A' with similar resources against E, without any extra security loss. Note that any PRP E *is* a secure PRF, albeit secure only up to the so-called birthday bound, i.e., roughly $2^{n/2}$ queries (even when E is secure as a PRP in face of much more than $2^{n/2}$ queries). Indeed, at this number of queries, the adversary expects to see (with good probability) collisions in the outputs of a random function, whereas this cannot happen when it interacts with a random permutation. This result is often called the PRP-PRF switching lemma [BR06]. Hence, any PRP-to-PRF conversion method must at the bare minimum overcome the birthday bound in order to be of any value.

The converse problem, namely building a PRP from a PRF, has been solved almost 30 years ago in a celebrated paper by Luby and Rackoff [LR88] using the 3-round Feistel construction (if one wants a PRP secure against adversary making two-sided queries to the black-box permutation, this requires four rounds [LR88, Pat90]). For this reason, the PRP-to-PRF conversion problem is sometimes called "Luby-Rackoff backwards" [BKR98].

Previous Work. A significant number of constructions have been suggested to solve the PRP-to-PRF conversion problem. Perhaps the simplest one is truncation: namely, one drops m bits of the output and simply uses the, say, $n - m$ leftmost bits of the output of E_k as the output of $F[E]_k$. This has been analyzed by Hall *et al.* [HWKS98], who showed that this is secure up to roughly $\min\{2^{(n+m)/2}, 2^{2(n-m)/3}\}$ adversarial queries (this bound was subsequently improved by Bellare and Impagliazzo [BI99], and recently by Gilboa *et al.* [GGM18]). In the same paper [HWKS98], Hall *et al.* also studied an inefficient yet security-preserving construction based on ordering the outputs $E_k(1\|x), \ldots, E_k(d\|x)$. Note that these constructions do not preserve the range (nor the domain for the latter) of the original permutation.

Another option suggested by Bellare, Krovetz, and Rogaway [BKR98] is to use data-dependent re-keying. In the simple case where the block cipher's key space $\mathcal{K}$ is equal to its message space $\mathcal{X}$, this construction is defined as $F(k, x) = E(E(k, x), x)$. This construction enjoys a good security bound (in particular, beyond birthday), albeit only in the ideal cipher model for E (in the standard model, security drops to birthday bound).

Another simple method is what we call the XOR construction, which simply consists in summing the output of $r \geq 2$ independent encryptions of the input, namely, assuming the domain of E is $\{0,1\}^n$,

$$F_{(k_1,\ldots,k_r)}(x) = E_{k_1}(x) \oplus E_{k_2}(x) \oplus \cdots \oplus E_{k_r}(x),$$

where $k_1, \ldots, k_r$ are independent keys. This construction was first analyzed by Lucks [Luc00] who showed that it is secure up to roughly $2^{rn/(r+1)}$ adversarial queries. At about the same time, Bellare and Impagliazzo [BI99] independently proved that for $r = 2$, the advantage is upper bounded by $O(n)(q/2^n)^{3/2}$ (in other words, security is ensured up to roughly $2^n/n^{2/3}$ queries). Patarin [Pat08a, Pat13] proved in two different ways that the construction for $r = 2$ is already "optimally" secure, i.e., secure up to approximately 2^n adversarial queries. A slightly worse bound of $2^{(2r+1)n/(2r+2)}$ queries for the general case (yet with a sharper threshold) was also proved by Cogliati *et al.* [CLP14]. The XOR construction can be slightly tweaked to work with a single key with negligible security loss by defining what Lucks calls the "TWIN" construction [Luc00], namely

$$F_k(x) = E_k(0\|x) \oplus E_k(1\|x) \oplus \cdots \oplus E_k(r-1\|x).$$

However this slightly shrinks the domain of the resulting PRF by $\lceil \log_2(r) \rceil$ bits. Recently, Dai *et al.* [DHT17] gave a tight yet much simpler security proof for the XOR and the TWIN constructions with $r = 2$ based on a new technique called the chi-squared method (see also [BN18]).

The Encrypted Davies-Meyer Construction. A natural idea that comes to mind to turn a PRP into a PRF is to define $F_k(x) = E_k(x) \oplus x$. When F is seen as a $2n$-bit to n-bit compression function, this is called the Davies-Meyer (DM) construction. We will use this terminology here as well, even though our focus is on pseudorandomness and not hashing. Although the DM construction breaks the bijectivity of E_k, it is easy to see that this construction is not more secure as a PRF than E is. Indeed, the adversary can simply compute $F_k(x) \oplus x = E_k(x)$, and hence apply the standard collision attack.

Recently, Cogliati and Seurin [CS16] proposed a new construction called *Encrypted Davies-Meyer* (EDM for short), which simply consists in encrypting the output of the DM construction with an independent key, namely

$$F_{k,k'}(x) = E_{k'}(E_k(x) \oplus x). \tag{1}$$

Note that this thwarts the collision attack as the adversary is now unable to compute $E_k(x)$ from the outputs of $F_{k,k'}$. And indeed, Cogliati and Seurin actually did prove that this construction is secure up to $2^{2n/3}$ adversarial queries (and conjectured that it might actually be secure up to close to 2^n queries). In fact, Cogliati and Seurin were primarily interested in proving the security of a related MAC construction called *Encrypted Wegman-Carter with Davies-Meyer* (EWCDM), which uses an additional AXU hash function H and computes a tag from a message m and a nonce ν as

$$\mathsf{EWCDM}[E, H]_{k,k',k''}(\nu, m) = E_{k'}\big(E_k(\nu) \oplus \nu \oplus H_{k''}(m)\big). \tag{2}$$

At the heart of the security proof of this construction as a MAC is a proof that the construction defined in Equation (1) is a secure PRF.

Our Contribution. In this work, we consider the single-key variant of the construction of Equation (1), which turns a block cipher E with key space $\mathcal{K}$ and domain $\{0,1\}^n$ into a keyed function $\mathsf{EDM}[E]$ from $\{0,1\}^n$ to $\{0,1\}^n$ with the same key space as E defined as

$$\mathsf{EDM}[E]_k(x) = E_k(E_k(x) \oplus x). \tag{3}$$

We prove that the security of this construction is very similar to the one of the two-key version, namely security is ensured up to at least roughly $2^{2n/3}$ adversarial queries.[1]

Observe that since the construction of Equation (3) does not use any form of data-dependent re-keying, a standard hybrid argument allows to replace E_k by a uniformly random permutation P (at the cost of a term related to the PRP-security of E) and to focus on the simpler construction defined, overloading notation $\mathsf{EDM}[.]$, as

$$\mathsf{EDM}[P](x) = P(P(x) \oplus x). \tag{4}$$

The problem is now purely information-theoretic, as we now have to upper bound the advantage of any (even computationally unbounded) adversary in distinguishing $\mathsf{EDM}[P]$ with a random permutation from a uniformly random function within q queries. Our proof uses the H-coefficients technique [Pat08b, CS14]. We remark that the PRF-security of the single-permutation EDM construction is somehow related to the so-called *iterated random permutation problem* recently considered by Minaud and Seurin [MS15], which asks how many queries are required to distinguish the square $P(P(\cdot))$ of a random permutation (or more generally the r-th iterate of a random permutation) from a uniformly random permutation (the answer proven in [MS15] being $\Theta(2^n)$ queries, for any fixed number r of iterations). However, since the EDM construction breaks the bijectivity of the underlying permutation, the simple game-based technique from [MS15], which replaces the random permutation by a random *cyclic* permutation, does not seem to apply here. Instead, we rely on a more basic (yet also more cumbersome) counting technique that was used in the work of Chen *et al.* [CLL+14] about the single-permutation 2-round Even-Mansour cipher.

We observe that the single-permutation EDM construction is the first PRP-to-PRF conversion method achieving all the following properties at the same time:

(i) it uses a single permutation (i.e., a single key for the underlying PRP);
(ii) it does not use data-dependent re-keying;
(iii) it does not shrink the domain nor the range of the original permutation;
(iv) it is provably secure beyond the birthday bound.

[1] Actually the security bound for the single-key version is slightly worse than for the two-key version since it has a term $O(nq/2^{2n/3})$.

Related Work. For the XOR construction, a stronger security property than pseudorandomness, namely *indifferentiability* from a random function (which informally means that the construction behaves as a random function even for an adversary having oracle access to the underlying permutations on which the construction is based), was also investigated in [MPN10, MP15].

Open Problems. We conjecture that our security bound is not tight. Actually, we are not aware of any attack requiring $o(2^n)$ queries, so that the single-permutation EDM construction might well be optimally secure. Recently, at CRYPTO 2017, two independent papers gave security bounds for the EDM construction with two independent permutations that improve on Cogliati and Seurin's bound [CS16]: Dai *et al.* [DHT17] gave a $3n/4$-bit security proof using a new technique called the chi-squared method, and Mennink and Neves [MN17] gave an optimal n-bit security proof based on Patarin's Mirror Theory [Pat10]. Mennink and Neves gave compelling evidence that their technique is unlikely to be applicable in the single-permutation case. We are not aware of any such argument for the chi-squared method though.

Another important open problem is to extend our result to the single-key version of the EWCDM construction of Equation (2), where the same key k is used for both calls to the block cipher (the hashing key k'' remaining independent from k). As explained in [CS16], it seems difficult to build on the PRF-security of the EDM construction to prove in a black-box way the MAC-security of the EWCDM construction. For now, we have been unable to extend the current (already cumbersome) counting used for the proof of the single-permutation EDM construction to the more complicated case of single-key EWCDM.

Organization. We start with basic notation and definitions in Section 2. In Section 3, we prove a technical "sum-capture" result which will be useful for our main theorem. Finally, in Section 4, we prove our main theorem about the security of the single-permutation EDM construction.

2 Preliminaries

Basic Notation. For non-empty sets $\mathcal{X}$ and $\mathcal{Y}$, the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathsf{Func}(\mathcal{X}, \mathcal{Y})$, and the set of all permutations of $\mathcal{X}$ is denoted $\mathsf{Perm}(\mathcal{X})$. The set of binary strings of length n is denoted $\{0,1\}^n$. The set of all functions from $\{0,1\}^n$ to $\{0,1\}^n$ is simply denoted $\mathsf{Func}(n)$, and the set of all permutations of $\{0,1\}^n$ is simply denoted $\mathsf{Perm}(n)$. For integers $1 \leq b \leq a$, we will write $(a)_b = a(a-1)\cdots(a-b+1)$ and $(a)_0 = 1$ by convention. Given a non-empty set $\mathcal{X}$, we denote $x \leftarrow_{\$} \mathcal{X}$ the draw of an element x from $\mathcal{X}$ uniformly at random. Note that the probability that a random permutation $P \leftarrow_{\$} \mathsf{Perm}(n)$ satisfies q equalities $P(x_i) = y_i$ for distinct x_i's and distinct y_i's is exactly $1/(2^n)_q$.

PRFs AND PRPs. A keyed function with key space $\mathcal{K}$, domain $\mathcal{X}$, and range $\mathcal{Y}$ is a function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$. We denote $F_k(x)$ for $F(k, x)$. A (q, t)-adversary against F is an algorithm A with oracle access to a function from $\mathcal{X}$ to $\mathcal{Y}$, making at most q oracle queries, running in time at most t, and outputting a single bit. The advantage of A in breaking the PRF-security of F is defined as

$$\mathbf{Adv}_F^{\mathrm{PRF}}(\mathsf{A}) = \left|\Pr\left[k \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{F_k} = 1\right] - \Pr\left[R \leftarrow_{\$} \mathsf{Func}(\mathcal{X}, \mathcal{Y}) : \mathsf{A}^{R} = 1\right]\right|.$$

A block cipher with key space $\mathcal{K}$ and domain $\mathcal{X}$ is a mapping $E : \mathcal{K} \times \mathcal{X} \to \mathcal{X}$ such that for any key $k \in \mathcal{K}$, $x \mapsto E(k, x)$ is a permutation of $\mathcal{X}$. We denote $E_k(x)$ for $E(k, x)$. A (q, t)-adversary against E is an algorithm A with oracle access to a permutation of $\mathcal{X}$, making at most q oracle queries, running in time at most t, and outputting a single bit. The advantage of A in breaking the PRP-security of E is defined as

$$\mathbf{Adv}_E^{\mathrm{PRP}}(\mathsf{A}) = \left|\Pr\left[k \leftarrow_{\$} \mathcal{K} : \mathsf{A}^{E_k} = 1\right] - \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(\mathcal{X}) : \mathsf{A}^{P} = 1\right]\right|.$$

3 Yet Another Sum-Capture Lemma

In the section we prove a technical result that will be needed in the proof of our main theorem. It is a "sum-capture" type result, meaning it upper bounds the quantity

$$\max_{\substack{A,B \\ |A|=|B|=|Y|}} \left|\{(y, a, b) \in Y \times A \times B : y = a \oplus b\}\right|$$

for a random subset Y of $\{0,1\}^n$ (or more generally of an abelian group). This kind of result typically considers a set Y drawn at random *without* replacement [Bab89, Ste13]. The lemma below considers the case where Y is sampled at random *with* replacement (i.e., Y is a multiset), which is what we will need later.

For any multiset Y^* with elements in $\{0,1\}^n$ and any two subsets A and B of $\{0,1\}^n$, let

$$\mu(Y^*, A, B) = \left|\{(y, a, b) \in Y^* \times A \times B \ : \ y = a \oplus b\}\right|$$

and let

$$\mu(Y^*) = \max_{\substack{A,B \\ |A|=|B|=|Y^*|}} \mu(Y^*, A, B).$$

Lemma 1 *Let Y^* be a multiset of $q \geq 1$ uniformly random and independently chosen elements of $\{0,1\}^n$. Then*

$$\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] \leq \frac{2}{2^n}.$$

Proof. We recall some useful results on Fourier analysis on $\mathbb{Z}_2^n$ in Appendix A. Let A and B be any two subsets of size q of $\{0,1\}^n$. For any subset $S \subset \{0,1\}^n$, we denote $\mathbb{1}_S : \{0,1\}^n \to \{0,1\}$ the characteristic function of S. Note that some values may be repeated several times in Y^*. We denote $\delta_{Y^*} : \{0,1\}^n \to \mathbb{N}$ the function that counts the multiplicity of a value in Y^*. Then one has

$$\begin{aligned}
\mu(Y^*, A, B) &= \sum_{y,a \in \{0,1\}^n} \delta_{Y^*}(y) \mathbb{1}_A(a) \mathbb{1}_B(y \oplus a) \\
&= \sum_{y \in \{0,1\}^n} \delta_{Y^*}(y) (\mathbb{1}_A * \mathbb{1}_B)(y) \\
&= 2^n \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{Y^*}}(\alpha) \widehat{(\mathbb{1}_A * \mathbb{1}_B)}(\alpha) \\
&= 2^{2n} \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{Y^*}}(\alpha) \widehat{\mathbb{1}_A}(\alpha) \widehat{\mathbb{1}_B}(\alpha) \\
&= 2^{2n} \widehat{\delta_{Y^*}}(0) \widehat{\mathbb{1}_A}(0) \widehat{\mathbb{1}_B}(0) + 2^{2n} \sum_{\alpha \neq 0} \widehat{\delta_{Y^*}}(\alpha) \widehat{\mathbb{1}_A}(\alpha) \widehat{\mathbb{1}_B}(\alpha).
\end{aligned}$$

Note that, for any subset S of $\{0,1\}^n$ one has $\widehat{\mathbb{1}_S}(0) = \frac{|S|}{2^n}$ and

$$\begin{aligned}
\widehat{\delta_{Y^*}}(0) &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x) (-1)^{0 \cdot x} \\
&= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x) \\
&= \frac{q}{2^n}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mu(Y^*, A, B) &= \frac{q^3}{2^n} + 2^{2n} \sum_{\alpha \neq 0} \widehat{\delta_{Y^*}}(\alpha) \widehat{\mathbb{1}_A}(\alpha) \widehat{\mathbb{1}_B}(\alpha) \\
&\leq \frac{q^3}{2^n} + 2^{2n} \sum_{\alpha \neq 0} |\widehat{\delta_{Y^*}}(\alpha)| \cdot |\widehat{\mathbb{1}_A}(\alpha)| \cdot |\widehat{\mathbb{1}_B}(\alpha)| \\
&\leq \frac{q^3}{2^n} + 2^n \Phi(Y^*) \sum_{\alpha \neq 0} |\widehat{\mathbb{1}_A}(\alpha)| \cdot |\widehat{\mathbb{1}_B}(\alpha)|,
\end{aligned}$$

where

$$\Phi(Y^*) = \max_{\alpha \neq 0} \left\{ 2^n |\widehat{\delta_{Y^*}}(\alpha)| \right\}.$$

By Cauchy-Schwarz, and using the fact that, for any subset $S \subseteq \{0,1\}^n$,

$$\sum_{\alpha \in \{0,1\}^n} |\widehat{\mathbb{1}_S}(\alpha)|^2 = \frac{|S|}{2^n},$$

we get

$$\begin{aligned}\mu(Y^*, A, B) &\leq \frac{q^3}{2^n} + 2^n \Phi(Y^*) \sqrt{\sum_{\alpha \in \{0,1\}^n} |\widehat{\mathbb{1}_A}(\alpha)|^2} \cdot \sqrt{\sum_{\alpha \in \{0,1\}^n} |\widehat{\mathbb{1}_B}(\alpha)|^2} \\ &\leq \frac{q^3}{2^n} + \Phi(Y^*) \sqrt{|A| \cdot |B|} \\ &\leq \frac{q^3}{2^n} + q \cdot \Phi(Y^*).\end{aligned}$$

Since this holds for any subsets A and B, it follows that

$$\mu(Y^*) \leq \frac{q^3}{2^n} + q \cdot \Phi(Y^*),$$

which implies

$$\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] \leq \Pr\left[\Phi(Y^*) \geq \sqrt{3nq}\right].$$

Denote $Y^* = \{y_1, \ldots, y_q\}$ using an arbitrary order. Then one has

$$\begin{aligned}\Phi(Y^*) &= \max_{\alpha \neq 0} \left\{ 2^n |\widehat{\delta_{Y^*}}(\alpha)| \right\} \\ &= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x) (-1)^{\alpha \cdot x} \right| \right\} \\ &= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \sum_{i=1}^{q} \mathbb{1}_{\{y_i\}}(x) (-1)^{\alpha \cdot x} \right| \right\} \\ &= \max_{\alpha \neq 0} \left\{ \left| \sum_{i=1}^{q} (-1)^{\alpha \cdot y_i} \right| \right\}.\end{aligned}$$

For $\alpha \neq 0$, let us denote $A_i^{(\alpha)} = (-1)^{\alpha \cdot y_i}$ and $A^{(\alpha)} = \sum_{i=1}^{q} A_i^{(\alpha)}$. Then $\Phi(Y^*) = \max_{\alpha \neq 0} \{|A^{(\alpha)}|\}$. The random variable $A^{(\alpha)}$ is the sum of q independent random variables $A_i^{(\alpha)}$ such that $\Pr\left[A_i^{(\alpha)} = 1\right] = \Pr\left[A_i^{(\alpha)} = -1\right] = \frac{1}{2}$. The Chernoff bound tailored to this special case [MU05, Corollary 4.8] gives us, for any $a > 0$,

$$\Pr\left[|A^{(\alpha)}| \geq a\right] \leq 2e^{\frac{-a^2}{2q}}.$$

If we take $a = \sqrt{3nq} > 0$, then

$$\Pr\left[|A^{(\alpha)}| \geq \sqrt{3nq}\right] \leq 2e^{\frac{-a^2}{2q}} = 2e^{-3n/2} \leq \frac{2}{2^{2n}}$$

since $e^{3/4} \geq 2$. Finally, one has

$$\begin{aligned}\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] &\leq \Pr\left[\max_{\alpha\neq 0}\{|A^{(\alpha)}|\} \geq \sqrt{3nq}\right] \\ &\leq \sum_{\alpha\neq 0} \Pr\left[|A^{(\alpha)}| \geq \sqrt{3nq}\right] \\ &\leq \frac{2}{2^n}.\end{aligned}$$

□

4 The Encrypted Davies-Meyer Construction With a Single Permutation

4.1 Statement of the Result and Overview of the Proof

In this section, we consider the Encrypted Davies-Meyer construction

$$\mathsf{EDM}[P](x) = P(P(x) \oplus x),$$

where P is a random permutation of $\{0,1\}^n$. Cogliati and Seurin [CS16] previously considered a similar construction with two independent permutations, namely $P'(P(x) \oplus x)$, and proved that it was a secure PRF up to roughly $2^{2n/3}$ adversarial queries. Here, we prove that the single-permutation variant is also secure up to roughly $2^{2n/3}$ adversarial queries. More precisely, one has the following theorem.

Theorem 1 *Assume that $n \geq 9$ and $q \leq 2^n/8$. Let A be an adversary with oracle access to a function from $\{0,1\}^n$ to $\{0,1\}^n$, making at most q oracle queries, and returning a single bit. Then its advantage $\mathbf{Adv}(\mathsf{A})$ in distinguishing the EDM construction from a uniformly random function, defined as*

$$\left|\Pr\left[P \leftarrow_{\$} \mathsf{Perm}(n) : \mathsf{A}^{\mathsf{EDM}[P]} = 1\right] - \Pr\left[R \leftarrow_{\$} \mathsf{Func}(n) : \mathsf{A}^{R} = 1\right]\right|,$$

satisfies:

$$\mathbf{Adv}(\mathsf{A}) \leq \frac{36q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}} + \frac{2}{2^n}.$$

As a straightforward corollary, we obtain the following for the corresponding keyed construction.

Corollary 1 *Let E be a block cipher with key space $\mathcal{K}$ and domain $\{0,1\}^n$. Let $\mathsf{EDM}[E]$ be the keyed function defined by Equation (3). Assume that $n \geq 9$ and $q \leq 2^n/8$. Then for any (q,t)-adversary against the PRF-security of $\mathsf{EDM}[E]$, there exists a $(2q,t')$-adversary against the PRP-security of E with $t' = t + O(q)$, such that*

$$\mathbf{Adv}^{\mathrm{PRF}}_{\mathsf{EDM}[E]}(\mathsf{A}) \leq \mathbf{Adv}^{\mathrm{PRP}}_{E}(\mathsf{A}') + \frac{36q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}} + \frac{2}{2^n}.$$

The remaining of the paper is devoted to the proof of Theorem 1. First of all, remark that the result is trivial if $q > 2^{2n/3}$. Hence we are going to assume that $q \leq 2^{2n/3}$, which implies that

$$\frac{q^3}{2^{2n}} \leq \frac{q^2}{2^{4n/3}} \leq \frac{q^{3/2}}{2^n} \leq \frac{q}{2^{2n/3}}. \tag{5}$$

The proof uses the H-coefficients technique [Pat08b, CS14]: the real world corresponds to $\mathsf{EDM}[P]$, while the ideal world corresponds to R. Fix an adversary A making q oracle queries, and consider the transcript τ of the queries x of the adversary and corresponding answers y: more precisely, τ contains a pair $(x, y) \in (\{0,1\}^n)^2$ *iff* the adversary made an oracle query x that was answered with y (as usual for stateless oracles, the order of the queries is unimportant for the reasoning). In the following, we sometimes refer to a pair $(x, y) \in \tau$ simply as a query. We denote

$$X = \{x : (x, y) \in \tau\},$$
$$Y = \{y : (x, y) \in \tau\}.$$

Note that $|X| = q$, assuming *wlog* that the adversary never repeats a query. On the other hand, there might be collisions among oracle answers. We say that a query $(x, y) \in \tau$ is *colliding* if there exists a distinct query $(x', y') \in \tau$ such that $y = y'$, otherwise we say it is *non-colliding*. We denote Y^* the multiset of all oracle answers (i.e., oracle answers counted with multiplicity).

We say that a transcript is attainable if there exists a function $R \in \mathsf{Func}(n)$ such that A interacting with R results in transcript τ. We denote Θ the set of attainable transcripts. We also denote T_{re}, resp. T_{id}, the probability distribution of the transcript τ induced by the real world, resp. the ideal world.

The main lemma of the H-coefficients technique is the following one (see e.g. [CS14] or [CLL+14] for the proof).

Lemma 2 *Fix an adversary* A. *Let* $\Theta = \Theta_{\mathrm{good}} \sqcup \Theta_{\mathrm{bad}}$ *be a partition of the set of attainable transcripts. Assume that there exists* ε_1 *such that for any* $\tau \in \Theta_{\mathrm{good}}$, *one has*[2]

$$\frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} \geq 1 - \varepsilon_1,$$

and that there exists ε_2 *such that* $\Pr[T_{\mathrm{id}} \in \Theta_{\mathrm{bad}}] \leq \varepsilon_2$. *Then* $\mathbf{Adv}(\mathsf{A}) \leq \varepsilon_1 + \varepsilon_2$.

4.2 Definition and Probability of Bad Transcripts

In all the following, we let

$$M = \frac{q}{2^{n/3}}. \tag{6}$$

Note that, assuming $n \geq 9$, one has

$$q - 3M \geq \frac{q}{2}. \tag{7}$$

[2] Recall that for an attainable transcript, one has $\Pr[T_{\mathrm{id}} = \tau] > 0$.

We define bad transcripts as follows.

Definition 1 We say that an attainable transcript τ is bad if one of the following conditions is fulfilled:

(i) the number of colliding queries is larger than M;
(ii) then number $\alpha(\tau)$ of triples $(y, x, x') \in Y^* \times X \times X$ such that $y = x \oplus x'$ is larger than $q^3/2^n + q\sqrt{3nq}$.

Otherwise, we say that τ is good. We denote Θ_{bad}, resp. Θ_{good}, the set of bad, resp. good transcripts.

We start by upper bounding the probability to get a bad transcript in the ideal world.

Lemma 3 *One has*

$$\Pr\left[T_{\text{id}} \in \Theta_{\text{bad}}\right] \leq \frac{q}{2^{2n/3}} + \frac{2}{2^n}.$$

Proof. We first consider condition (i). Since in the ideal world the oracle answers y are uniformly random and independent, the expected number of colliding queries is lower than $q^2/2^n$. If we denote C the random variable defined as the number of colliding queries, then, by Markov's inequality,

$$\Pr\left[C \geq M\right] \leq \frac{q^2}{M2^n} = \frac{q}{2^{2n/3}}.$$

We now consider condition (ii). Note that one has $\alpha(\tau) = \mu(Y^*, X, X) \leq \mu(Y^*)$ (see Section 3 for the definition of μ). Since in the ideal world, Y^* is a multiset of uniformly random and independent values, using Lemma 1, one has

$$\Pr\left[\alpha(\tau) \geq q^3/2^n + q\sqrt{3nq}\right] \leq \Pr\left[\mu(Y^*) \geq q^3/2^n + q\sqrt{3nq}\right] \leq \frac{2}{2^n}.$$

The result follows by the union bound. □

4.3 Analysis of Good Transcripts

From now on, we fix a good transcript τ. We need to lower bound the probability to obtain τ in the real world. As usual, this probability is exactly the probability that the real oracle be "compatible" with the transcript (see e.g. [CS14]), i.e.,

$$\Pr\left[T_{\text{re}} = \tau\right] = \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(n) : \forall (x, y) \in \tau, P(P(x) \oplus x) = y\right].$$

In all the following, we let

$$r = |Y|$$

be the number of distinct oracle answers appearing in the transcript and

$$s = |\{(x, y) \in \tau : \forall (x', y') \in \tau \setminus \{(x, y)\}, y' \neq y\}|$$

be the number of non-colliding queries. Recall that $\alpha(\tau)$ denotes the number of triples $(y, x, x') \in Y^* \times X \times X$ such that $y = x \oplus x'$. Since the transcript is good, one has

$$q \geq s \geq q - M \tag{8}$$

$$\alpha(\tau) \leq \frac{q^3}{2^n} + q\sqrt{3nq}. \tag{9}$$

As just explained, in order to lower bound the probability of obtaining τ in the real world, we need to lower bound the number of permutations P such that

$$\forall (x, y) \in \tau,\ P(P(x) \oplus x) = y. \tag{10}$$

What makes this counting hard is that these equalities are not "independent". E.g., if there exists two queries (x, y) and (x', y') in τ such that $P(x) \oplus x = x'$, then one must have $P(x') = y$. Similarly, if $P(x) = y'$, then one must have $P(x') \oplus x' = x$. One could count only permutations P such that for any query $(x, y) \in \tau$, $P(x) \oplus x \notin X \cup Y$, however this only leads to a birthday bound. Hence, to get a bound beyond the birthday bound, we will need a more precise counting. As we will see now, it will be sufficient to consider permutations P such that $P(x) \oplus x = x'$ for t pairs $((x, y), (x', y'))$ of distinct non-colliding queries, for t in some sufficiently large range. However, we must ensure that the choice of these t pairs does not create constraints incompatible with other queries in the transcript. To this end, we introduce the following definition.

Definition 2 An unordered set of t (ordered) pairs of distinct non-colliding queries

$$\Sigma = \{((x_1, y_1), (x'_1, y'_1)), \ldots, ((x_t, y_t), (x'_t, y'_t))\}$$

is said *good* if the following conditions are fulfilled:

(a) for all $i \in \{1, \ldots, t\}$, $y_i \oplus x'_i \notin X$;
(b) for all $i \in \{1, \ldots, t\}$, $x_i \oplus x'_i \notin Y$;
(c) all values $y_i \oplus x'_i$, $i \in \{1, \ldots, t\}$, are distinct;
(d) all values $x_i \oplus x'_i$, $i \in \{1, \ldots, t\}$, are distinct.

Then we have the following lemma, which shows that the number of good sets Σ is close to $(s)_{2t}/t!$, the total number of unordered sets of t pairs of non-colliding queries. (Recall that s denotes the number of non-colliding queries in τ.)

Lemma 4 *Fix an integer t such that $0 \leq t \leq M$. Then the number $N_\Sigma(t)$ of good sets Σ of t pairs of non-colliding queries is at least*

$$N_\Sigma(t) \geq \frac{(s)_{2t}}{t!} \left(1 - \frac{12q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}}\right).$$

Proof. First, observe that among the $s(s-1)$ possible pairs of non-colliding queries, at most $2\alpha(\tau)$ of them do *not* satisfy conditions (a) and (b). Indeed, by definition of a good transcript (more precisely, condition (ii)), there cannot be more than $\alpha(\tau)$ pairs $((x,y),(x',y'))$ such that $y \oplus x' \in X$, and there cannot be more than $\alpha(\tau)$ pairs $((x,y),(x',y'))$ such that $x \oplus x' \in Y$. Hence, we can lower bound $N_\Sigma(t)$ as follows:

- we can choose $((x_1,y_1),(x'_1,y'_1))$ among at least $s(s-1)-2\alpha(\tau)$ possibilities;
- once $((x_1,y_1),(x'_1,y'_1))$ is fixed, we can choose (x_2,y_2) freely from the remaining $(s-2)$ possibilities; then, (x'_2,y'_2) must be different from (x_1,y_1), (x'_1,y'_1), and (x_2,y_2), and must be such that $x'_2 \neq y_2 \oplus y_1 \oplus x'_1$ in order to satisfy (c), and such that $x'_2 \neq x_2 \oplus x_1 \oplus x'_1$ in order to satisfy (d), which removes at most two possibilities since all queries made by the distinguisher are distinct; hence, overall there are at least $(s-5)$ possibilities for (x'_2,y'_2); after removing the at most $2\alpha(\tau)$ pairs of queries not satisfying (a) and (b), there remains at least $(s-2)(s-5)-2\alpha(\tau)$ possibilities for the pair $((x_2,y_2),(x'_2,y'_2))$;
- assume $((x_1,y_1),(x'_1,y'_1)),\ldots,((x_{i-1},y_{i-1}),(x'_{i-1},y'_{i-1}))$ have been chosen; we can choose (x_i,y_i) freely from the $(s-2i+2)$ remaining possibilities; then, (x'_i,y'_i) must be different from (x_1,y_1), (x'_1,y'_1), $\ldots$, (x_{i-1},y_{i-1}),(x'_{i-1},y'_{i-1}), and (x_i,y_i); moreover, it must be such that $x'_i \neq y_i \oplus y_j \oplus x'_j$ for all $j \in \{1,\ldots,i-1\}$ in order to satisfy (c), and such that $x'_i \neq x_i \oplus x_j \oplus x'_j$ for all $j \in \{1,\ldots,i-1\}$ in order to satisfy (d); overall, there are at least $(s-4i+3)$ possibilities for (x'_i,y'_i); after removing the at most $2\alpha(\tau)$ pairs not satisfying (a) and (b), there remains at least $(s-2i+2)(s-4i+3)-2\alpha(\tau)$ possibilities for the pair $((x_i,y_i),(x'_i,y'_i))$.

Since we consider an *unordered* set of t pairs, the number $N_\Sigma(t)$ of good sets Σ is at least

$$N_\Sigma(t) \geq \frac{1}{t!}\prod_{i=0}^{t-1}\big((s-2i)(s-4i-1)-2\alpha(\tau)\big).$$

Then

$$\begin{aligned}
N_\Sigma(t) &\geq \frac{(s)_{2t}}{t!}\prod_{i=0}^{t-1}\frac{(s-2i)(s-4i-1)-2\alpha(\tau)}{(s-2i)(s-2i-1)} \\
&\geq \frac{(s)_{2t}}{t!}\prod_{i=0}^{t-1}\left(1-\frac{2si-4i^2+2\alpha(\tau)}{(s-2i)(s-2i-1)}\right) \\
&\geq \frac{(s)_{2t}}{t!}\left(1-\sum_{i=0}^{t-1}\frac{2si-4i^2+2\alpha(\tau)}{(s-2i)(s-2i-1)}\right) \\
&\geq \frac{(s)_{2t}}{t!}\left(1-\sum_{i=0}^{t-1}\frac{2si+2\alpha(\tau)}{(s-2M)^2}\right) \\
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{sM^2+2\alpha(\tau)M}{(s-2M)^2}\right)
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{qM^2+2\alpha(\tau)M}{(q-3M)^2}\right) && \text{by Equation (8)}\\
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{4qM^2+8\alpha(\tau)M}{q^2}\right) && \text{by Equation (7)}\\
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{4q}{2^{2n/3}}-\frac{8\alpha(\tau)}{q2^{n/3}}\right) && \text{by Equation (6)}\\
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{4q}{2^{2n/3}}-\frac{8q^2}{2^{4n/3}}-\frac{8\sqrt{3nq}}{2^{n/3}}\right) && \text{by Equation (9)}\\
&\geq \frac{(s)_{2t}}{t!}\left(1-\frac{12q}{2^{2n/3}}-\frac{8\sqrt{3nq}}{2^{n/3}}\right) && \text{by Equation (5)},
\end{aligned}$$

as claimed. □

From now on, we fix some integer t such that $0 \leq t \leq M$ and some good set of t pairs of non-colliding queries

$$\Sigma = \{((x_1, y_1), (x'_1, y'_1)), \ldots, ((x_t, y_t), (x'_t, y'_t))\}.$$

We will lower bound the number of permutations P satisfying Equation (10) such that for any $i \in \{1, \ldots, t\}$, $P(x_i) \oplus x_i = x'_i$. Note that such a permutation satisfies Equation (10) for the $2t$ queries appearing in Σ *iff*

$$\forall i \in \{1, \ldots, t\}, \begin{cases} P(x_i) = x_i \oplus x'_i \\ P(x'_i) = y_i \\ P(y_i \oplus x'_i) = y'_i. \end{cases} \tag{11}$$

This set of $3t$ equalities is "satisfiable" in the sense that all inputs, resp. outputs, are distinct by conditions (a) and (c), resp. (b) and (d), characterizing a good set Σ (and also since all x_i's are distinct by assumption, and all y_i's are distinct for non-colliding queries). In the following, we denote

$$\begin{aligned}
X' &= X \cup \{y_i \oplus x'_i : i \in \{1, \ldots, t\}\},\\
Y' &= Y \cup \{x_i \oplus x'_i : i \in \{1, \ldots, t\}\}.
\end{aligned}$$

Note that $|X'| = q + t$ and $|Y'| = r + t$.

It remains to consider the $q - 2t$ queries $(u, v) \in \tau$ not appearing in Σ. Let

$$\begin{aligned}
q' &= q - 2t\\
r' &= r - 2t\\
s' &= s - 2t
\end{aligned}$$

be respectively the number of these queries, the number of distinct oracle answers appearing in these queries, and the number of such queries that are non-colliding. We group these remaining queries so that all queries with the same output are consecutive, and write them as

$$\begin{aligned}
\tau' = \big(&(u_{1,1}, v_1), \ldots, (u_{1,q_1}, v_1),\\
&\ldots,\\
&(u_{r',1}, v_{r'}), \ldots, (u_{r',q_{r'}}, v_{r'})\big),
\end{aligned}$$

where $v_1, \ldots, v_{r'}$ are distinct and $\sum_{i=1}^{r'} q_i = q'$. In order to ease computations later, we also assume that we ordered the queries so that non-colliding queries come first, i.e., $q_i = 1$ for $i \in \{1, \ldots, s'\}$ and $q_i > 1$ for $i \in \{s'+1, \ldots, r'\}$. Note that since the transcript is good, one has

$$r - s = r' - s' \leq \sum_{i=s'+1}^{r'} q_i \leq M = \frac{q}{2^{n/3}} \tag{12}$$

where the second inequality holds as otherwise condition (i) of a bad transcript would be fulfilled.

Our goal is now to lower bound the number of permutations P which, in addition to satisfying Equation (11), also satisfies

$$\forall (u, v) \in \tau', \ P(P(u) \oplus u) = v.$$

For this, we will consider all possible "intermediate" values $z_i = P^{-1}(v_i)$. Formally, we need the definition below.

Definition 3 A tuple of r' values $\mathbf{z} = (z_1, \ldots, z_{r'})$ is said *good* if all z_i's are distinct and outside X', and all values $z_i \oplus u_{i,j}$ for $i \in \{1, \ldots, r'\}$ and $j \in \{1, \ldots, q_i\}$ are distinct and outside Y'.

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

$$\begin{cases} \forall i \in \{1, \ldots, r'\}, \ \forall j \in \{1, \ldots, q_i\}, \ P(u_{i,j}) = z_i \oplus u_{i,j} \\ \forall i \in \{1, \ldots, r'\}, \ P(z_i) = v_i \end{cases} \tag{13}$$

is "satisfiable" and "compatible" with equalities of Equation (11) in the sense all inputs, resp. all outputs appearing in equalities of Equation (11) and Equation (13) are distinct by definition of a good set Σ and a good tuple $\mathbf{z}$. Moreover, a permutation P satisfying Equation (13) is such that $P(P(u) \oplus u) = v$ for all $(u, v) \in \tau'$.

We will now prove in the lemma below that the number of good tuples $\mathbf{z}$ is close to $(2^n - q - r - 2t)_{s'} (2^n)^{r'-s'}$. The rather complicated form of the bound will simplify computations later. Note that the term

$$\prod_{i=0}^{s'-1} \left(1 - \frac{i}{2^n - 3q - i}\right)$$

in the lower bound of this lemma is a "birthday" term since $s' \sim q$, however we will be able to cancel it with another term later.

Lemma 5 *Fix t and Σ as above. Then the number $N_{\mathbf{z}}(t)$ of good tuples $\mathbf{z}$ is at least*

$$N_{\mathbf{z}}(t) \geq (2^n - q - r - 2t)_{s'} (2^n)^{r'-s'} \left(1 - \frac{4q}{2^{2n/3}}\right) \prod_{i=0}^{s'-1} \left(1 - \frac{i}{2^n - 3q - i}\right).$$

Proof. Recall that $|X'| = q + t$ and $|Y'| = r + t$. We will lower bound $N_{\mathbf{z}}(t)$ as follows:

- z_1 must be such that $z_1 \notin X'$ and $z_1 \oplus u_{1,j} \notin Y'$, which leaves at least $2^n - q - t - q_1(r+t)$ possibilities for z_1;
- once z_1 is fixed, there are at least $2^n - q - t - 1 - q_2(r + t + q_1)$ possibilities for z_2, since z_2 must be different from z_1 and from $z_1 \oplus u_{1,j} \oplus u_{2,j'}$ for all $j \in \{1, \ldots, q_1\}$ and all $j' \in \{1, \ldots, q_2\}$; we also want $z_2 \notin X'$ and $z_2 \oplus u_{2,i} \notin Y'$ for all $i \in \{1, \ldots, q_2\}$;
- once z_1 and z_2 are fixed, there are at least $2^n - q - t - 2 - q_3(r + t + q_1 + q_2)$ possibilities for z_3, since z_3 must be different from z_1, z_2, $z_1 \oplus u_{1,j} \oplus u_{3,j'}$ for all $j \in \{1, \ldots, q_1\}$ and all $j' \in \{1, \ldots, q_3\}$, and from $z_2 \oplus u_{2,j} \oplus u_{3,j'}$ for all $j \in \{1, \ldots, q_2\}$ and all $j' \in \{1, \ldots, q_3\}$; we also want $z_3 \notin X'$ and $z_3 \oplus u_{3,i} \notin Y'$ for all $i \in \{1, \ldots, q_3\}$;
- etc.

Hence, the number of good tuples $\mathbf{z}$ is at least

$$N_{\mathbf{z}}(t) \geq \prod_{i=0}^{r'-1} \left(2^n - q - t - i - q_{i+1} \left(r + t + \sum_{j=1}^{i} q_j \right) \right).$$

We will further lower bound this quantity by separating terms corresponding to non-colliding queries ($0 \leq i \leq s' - 1$) for which $q_i = 1$ and colliding queries ($s' \leq i \leq r' - 1$). We have

$$\begin{aligned} N_{\mathbf{z}}(t) &\geq \prod_{i=0}^{s'-1} \left(2^n - q - t - i - q_{i+1} \left(r + t + \sum_{j=1}^{i} q_j \right) \right) \\ &\quad \times \prod_{i=s'}^{r'-1} \left(2^n - q - t - i - q_{i+1} \left(r + t + \sum_{j=1}^{i} q_j \right) \right) \\ &\geq \underbrace{\prod_{i=0}^{s'-1} (2^n - q - r - 2t - 2i)}_{N_{\mathbf{z},1}(t)} \underbrace{\prod_{i=s'}^{r'-1} (2^n - 2q - 2qq_{i+1})}_{N_{\mathbf{z},2}(t)}, \end{aligned}$$

where for $s' \leq i \leq r' - 1$ we used that

$$\begin{aligned} q + t + i &\leq q + t + r' - 1 = q + r - t - 1 \leq 2q \\ \text{and } r + t + \sum_{j=1}^{i} q_j &\leq r + t + q' = r + q - t \leq 2q. \end{aligned}$$

Moreover,

$$\begin{aligned}\frac{N_{\mathbf{z},1}(t)}{(2^n-q-r-2t)_{s'}} &= \prod_{i=0}^{s'-1}\left(\frac{2^n-q-r-2t-2i}{2^n-q-r-2t-i}\right)\\ &= \prod_{i=0}^{s'-1}\left(1-\frac{i}{2^n-q-r-2t-i}\right)\\ &\geq \prod_{i=0}^{s'-1}\left(1-\frac{i}{2^n-3q-i}\right),\end{aligned}$$

where for the last inequality we used that $r \leq q$ and $2t \leq 2M = 2q/2^{n/3} \leq q$, and

$$\begin{aligned}\frac{N_{\mathbf{z},2}(t)}{(2^n)^{r'-s'}} &= \prod_{i=s'}^{r'-1}\left(\frac{2^n-2q-2qq_{i+1}}{2^n}\right)\\ &\geq \prod_{i=s'}^{r'-1}\left(1-\frac{4qq_{i+1}}{2^n}\right)\\ &\geq 1-\frac{4q\sum_{i=s'}^{r'} q_{i+1}}{2^n}\\ &\geq 1-\frac{4q^{3/2}}{2^n} && \text{by Equation (12)}\\ &\geq 1-\frac{4q}{2^{2n/3}} && \text{by Equation (5)},\end{aligned}$$

which concludes the proof. □

We are now ready to wrap up the counting and prove the following result.

Lemma 6 *Assume $n \geq 9$ and $q \leq 2^n/8$. Then for any good transcript τ, one has*

$$\frac{\Pr\left[T_{\mathrm{re}}=\tau\right]}{\Pr\left[T_{\mathrm{id}}=\tau\right]} \geq 1-\frac{35q}{2^{2n/3}}-\frac{8\sqrt{3nq}}{2^{n/3}}.$$

Proof. For each integer t with $0 \leq t \leq M$, each possible choice of good set Σ of t pairs of non-colliding queries, and each possible choice of good tuple $\mathbf{z}$, the probability that a random permutation P satisfies equalities of Equation (11) and Equation (13) (which implies that it satisfies Equation (10)) is exactly

$$\frac{1}{(2^n)_{q+r'+t}}.$$

Indeed, there are exactly $3t$ equalities in Equation (11) and exactly $q'+r' = q-2t+r'$ equalities in Equation (13), hence $q+r'+t$ equalities in total to satisfy, and these equalities are "compatible" by the very definition of a good

set Σ and a good tuple $\mathbf{z}$. Summing over t, Σ, and $\mathbf{z}$, we obtain that the probability to get the transcript τ in the real world satisfies

$$\Pr\left[T_{\mathrm{re}} = \tau\right] \geq \sum_{0 \leq t \leq M} \frac{N_{\Sigma}(t) N_{\mathbf{z}}(t)}{(2^n)_{q+r'+t}}.$$

Since the probability to obtain τ in the ideal world is simply $1/(2^n)^q$, the ratio of probabilities is at least

$$\rho \stackrel{\mathrm{def}}{=} \frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} \geq \sum_{0 \leq t \leq M} \frac{(2^n)^q N_{\Sigma}(t) N_{\mathbf{z}}(t)}{(2^n)_{q+r'+t}}.$$

Injecting successively Lemma 4 and Lemma 5 in this inequality, one has

$$\begin{aligned}
\rho &\geq \left(1 - \frac{12q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}}\right) \sum_{0 \leq t \leq M} \frac{(s)_{2t} (2^n)^q N_{\mathbf{z}}(t)}{t!(2^n)_{q+r'+t}} \\
&\geq \left(1 - \frac{12q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}}\right) \left(1 - \frac{4q}{2^{2n/3}}\right) \\
&\quad \times \sum_{0 \leq t \leq M} \frac{(s)_{2t} (2^n)^q (2^n - q - r - 2t)_{s'} (2^n)^{r'-s'}}{t!(2^n)_{q+r'+t}} \prod_{i=0}^{s'-1} \left(1 - \frac{i}{2^n - 3q - i}\right).
\end{aligned}$$

Since

$$\begin{aligned}
(2^n)_{q+r'+t} &= (2^n)_q (2^n - q)_{r'-s'} (2^n - q - r' + s')_{s'+t} \\
&= (2^n)_q (2^n - q)_{r'-s'} (2^n - q - r + s)_{s'+t},
\end{aligned}$$

we get

$$\begin{aligned}
\rho \geq \left(1 - \frac{16q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}}\right) \underbrace{\frac{(2^n)^q}{(2^n)_q} \prod_{i=0}^{s'-1} \left(1 - \frac{i}{2^n - 3q - i}\right)}_{A} \underbrace{\frac{(2^n)^{r'-s'}}{(2^n - q)_{r'-s'}}}_{\geq 1} \\
\times \underbrace{\sum_{0 \leq t \leq M} \frac{(s)_{2t} (2^n - q - r - 2t)_{s'}}{t!(2^n - q - r + s)_{s'+t}}}_{B}. \quad (14)
\end{aligned}$$

We will now lower bound A and B. For A, we have

$$\begin{aligned}
A &= \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{s'-1} \left(1 - \frac{i}{2^n - 3q - i}\right) \\
&\geq \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \left(1 - \frac{i}{2^n - 3q - i}\right) \\
&= \prod_{i=0}^{q-1} \left(1 - \frac{3qi + i^2}{(2^n - i)(2^n - 3q - i)}\right)
\end{aligned}$$

$$\begin{aligned}
&\geq \prod_{i=0}^{q-1} \left(1 - \frac{4qi}{(2^n-q)(2^n-4q)}\right) \\
&\geq 1 - \frac{4q \cdot q^2/2}{(7 \cdot 2^n/8)(2^n/2)} && (q \leq 2^n/8) \\
&\geq 1 - \frac{5q}{2^{2n/3}} && \text{by Equation (5).} \qquad (15)
\end{aligned}$$

In order to lower bound B, we are going to appeal to a trick previously used in [CLL+14] and use the fact that the terms of this sum are close to the hypergeometric distribution. This is a discrete probability distribution which describes the probability of k successes in u draws without replacement, from a finite population of U elements that contains exactly K "good" elements and exactly $U-K$ "bad" ones. The probability that exactly k elements are drawn from the set of K "good" elements is thus

$$\mathbf{Hyp}_{U,K,u}(k) = \frac{\binom{K}{k}\binom{U-K}{u-k}}{\binom{U}{u}} = \frac{(u)_k (K)_k (U-K)_{u-k}}{k!(U)_u}.$$

The mean of the distribution $\mathbf{Hyp}_{U,K,u}$ is uK/U. Since

$$\mathbf{Hyp}_{2^n-q,s,s}(t) = \frac{(s)_t (s)_t (2^n-q-s)_{s-t}}{t!(2^n-q)_s} = \frac{(s)_t (s)_t (2^n-q-s)_{s'+t}}{t!(2^n-q)_{s'+2t}},$$

we can rewrite B as

$$\sum_{0 \leq t \leq M} \underbrace{\frac{(s)_{2t}}{(s)_t (s)_t}}_{C} \cdot \underbrace{\frac{(2^n-q)_{s'+2t}(2^n-q-r-2t)_{s'}}{(2^n-q-s)_{s'+t}(2^n-q-r+s)_{s'+t}}}_{D} \cdot \mathbf{Hyp}_{2^n-q,s,s}(t). \qquad (16)$$

We are now going to lower bound C and D independently from t and then lower bound B using Markov's inequality. First, for any t with $0 \leq t \leq M$,

$$\begin{aligned}
C = \frac{(s)_{2t}}{(s)_t (s)_t} \geq \frac{(s-2M)^{2t}}{s^{2t}} &\geq 1 - \frac{4tM}{s} \\
&\geq 1 - \frac{4M^2}{q-M} && \text{by Equation (8)} \\
&\geq 1 - \frac{8M^2}{q} && \text{by Equation (7)} \\
&= 1 - \frac{8q}{2^{2n/3}} && \text{by Equation (6).} \qquad (17)
\end{aligned}$$

For D, one has

$$\begin{aligned}
D &= \frac{(2^n-q)_{s'+t}(2^n-q-s'-t)_t(2^n-q-r-2t)_{s'}}{(2^n-q-s)_t(2^n-q-s-t)_{s'}(2^n-q-r+s)_{s'+t}} \\
&= \underbrace{\frac{(2^n-q)_{s'+t}}{(2^n-q-r+s)_{s'+t}}}_{\geq 1} \cdot \underbrace{\frac{(2^n-q-s+t)_t}{(2^n-q-s)_t}}_{\geq 1} \cdot \frac{(2^n-q-r-2t)_{s'}}{(2^n-q-s-t)_{s'}}
\end{aligned}$$

$$
\begin{aligned}
&\geq \prod_{i=0}^{s'-1} \frac{2^n - q - r - 2t - i}{2^n - q - s - t - i} \\
&= \prod_{i=0}^{s'-1} \left(1 - \frac{r - s + t}{2^n - q - s - t - i}\right) \\
&\geq 1 - \frac{s'(r - s + M)}{2^n - 3q} \\
&\geq 1 - \frac{2qM}{2^n - 3q} && \text{by Equation (12) and } s' \leq q \\
&\geq 1 - \frac{4q^2}{2^{4n/3}} && \text{by Equation (6) and } 3q \leq 2^n/2 \\
&\geq 1 - \frac{4q}{2^{2n/3}} && \text{by Equation (5).} \qquad (18)
\end{aligned}
$$

Since the mean of the hypergeometric distribution $\mathbf{Hyp}_{2^n-q,s,s}$ is $\frac{s^2}{2^n-q}$, we have

$$\sum_{t>M} \mathbf{Hyp}_{2^n-q,s,s}(t) \leq \frac{s^2}{M(2^n - q)} \leq \frac{2q^2}{M2^n} = \frac{2q}{2^{2n/3}}$$

using successively Markov's inequality, $q \leq 2^n/2$, and Equation (6). It follows that

$$\sum_{0 \leq t \leq M} \mathbf{Hyp}_{2^n-q,s,s}(t) \geq 1 - \frac{2q}{2^{2n/3}}.$$

Combining this with Equation (16), Equation (17), and Equation (18), we get

$$B \geq \left(1 - \frac{8q}{2^{2n/3}}\right)\left(1 - \frac{4q}{2^{2n/3}}\right) \sum_{0 \leq t \leq M} \mathbf{Hyp}_{N-q,s,s}(t) \geq 1 - \frac{14q}{2^{2n/3}}. \qquad (19)$$

Combining Equation (14), Equation (15), and Equation (19), we finally obtain

$$\rho \geq 1 - \frac{35q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}}, \qquad (20)$$

as claimed. □

4.4 Concluding the Proof

We are now ready to complete the proof of Theorem 1. Combining Lemma 2, Lemma 3, and Lemma 6, we obtain that the distinguishing advantage of any adversary is at most

$$\frac{36q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}} + \frac{2}{2^n}.$$

References

Bab89. László Babai. The Fourier Transform and Equations over Finite Abelian Groups: An introduction to the method of trigonometric sums. Lecture notes, December 1989. Available at http://people.cs.uchicago.edu/~laci/reu02/fourier.pdf.

BDJR97. Mihir Bellare, Anand Desai, E. Jokipii, and Phillip Rogaway. A Concrete Security Treatment of Symmetric Encryption. In *Symposium on Foundations of Computer Science - FOCS '97*, pages 394–403. IEEE Computer Society, 1997.

Ber05. Daniel J. Bernstein. Stronger Security Bounds for Wegman-Carter-Shoup Authenticators. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 164–180. Springer, 2005.

BI99. Mihir Bellare and Russell Impagliazzo. A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptology ePrint Archive, Report 1999/024, 1999. Available at http://eprint.iacr.org/1999/024.

BKR98. Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-Rackoff Backwards: Increasing Security by Making Block Ciphers Non-invertible. In Kaisa Nyberg, editor, *Advances in Cryptology - EUROCRYPT '98*, volume 1403 of *LNCS*, pages 266–280. Springer, 1998.

BN18. Srimanta Bhattacharya and Mridul Nandi. A note on the chi-square method: A tool for proving cryptographic security. *Cryptography and Communications*, 2018. https://doi.org/10.1007/s12095-017-0276-z.

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*, volume 4004 of *LNCS*, pages 409–426. Springer, 2006. Full version available at http://eprint.iacr.org/2004/331.

CLL+14. Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the Two-Round Even-Mansour Cipher. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology - CRYPTO 2014 (Proceedings, Part I)*, volume 8616 of *LNCS*, pages 39–56. Springer, 2014. Full version available at http://eprint.iacr.org/2014/443.

CLP14. Benoît Cogliati, Rodolphe Lampe, and Jacques Patarin. The Indistinguishability of the XOR of k Permutations. In Carlos Cid and Christian Rechberger, editors, *Fast Software Encryption - FSE 2014*, volume 8540 of *LNCS*, pages 285–302. Springer, 2014.

CS14. Shan Chen and John Steinberger. Tight Security Bounds for Key-Alternating Ciphers. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 327–350. Springer, 2014. Full version available at http://eprint.iacr.org/2013/222.

CS16. Benoît Cogliati and Yannick Seurin. EWCDM: An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 (Proceedings, Part I)*, volume 9814 of *LNCS*, pages 121–149. Springer, 2016.

DHT17. Wei Dai, Viet Tung Hoang, and Stefano Tessaro. Information-theoretic Indistinguishability via the Chi-squared Method. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology - CRYPTO 2017 (Proceedings, Part III)*, volume 10403 of *LNCS*, pages 497–523. Springer, 2017. Full version at http://eprint.iacr.org/2017/537.

GGM86. Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

GGM18. Shoni Gilboa, Shay Gueron, and Ben Morris. How Many Queries are Needed to Distinguish a Truncated Random Permutation from a Random Function? *J. Cryptology*, 31(1):162–171, 2018.

HWKS98. Chris Hall, David Wagner, John Kelsey, and Bruce Schneier. Building PRFs from PRPs. In Hugo Krawczyk, editor, *Advances in Cryptology - CRYPTO '98*, volume 1462 of *LNCS*, pages 370–389. Springer, 1998.

LR88. Michael Luby and Charles Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, 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*, volume 1807 of *LNCS*, pages 470–484. Springer, 2000.

MN17. Bart Mennink and Samuel Neves. Encrypted Davies-Meyer and Its Dual: Towards Optimal Security Using Mirror Theory. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology - CRYPTO 2017 (Proceedings, Part III)*, volume 10403 of *LNCS*, pages 556–583. Springer, 2017. Full version at http://eprint.iacr.org/2017/473.

MP15. Bart Mennink and Bart Preneel. On the XOR of Multiple Random Permutations. In Tal Malkin, Vladimir Kolesnikov, Allison Bishop Lewko, and Michalis Polychronakis, editors, *Applied Cryptography and Network Security - ACNS 2015*, volume 9092 of *LNCS*, pages 619–634. Springer, 2015.

MPN10. Avradip Mandal, Jacques Patarin, and Valérie Nachef. Indifferentiability beyond the Birthday Bound for the Xor of Two Public Random Permutations. In Guang Gong and Kishan Chand Gupta, editors, *Progress in Cryptology - INDOCRYPT 2010*, volume 6498 of *LNCS*, pages 69–81. Springer, 2010.

MS15. Brice Minaud and Yannick Seurin. The Iterated Random Permutation Problem with Applications to Cascade Encryption. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 (Proceedings, Part I)*, volume 9215 of *LNCS*, pages 351–367. Springer, 2015.

MU05. Michael Mitzenmacher and Eli Upfal. *Probability and computing - randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

Pat90. Jacques Patarin. Pseudorandom Permutations Based on the DES Scheme. In Gérard D. Cohen and Pascale Charpin, editors, *EUROCODE '90*, volume 514 of *LNCS*, pages 193–204. Springer, 1990.

Pat08a. Jacques Patarin. A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations. In Reihaneh Safavi-Naini, editor, *Information Theoretic Security - ICITS 2008*, volume 5155 of *LNCS*, pages 232–248. Springer, 2008. Full version available at http://eprint.iacr.org/2008/010.

Pat08b. Jacques Patarin. The "Coefficients H" Technique. In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *Selected Areas in Cryptography - SAC 2008*, volume 5381 of *LNCS*, pages 328–345. Springer, 2008.

Pat10. Jacques Patarin. Introduction to Mirror Theory: Analysis of Systems of Linear Equalities and Linear Non Equalities for Cryptography. IACR Cryptology ePrint Archive, Report 2010/287, 2010. Available at http://eprint.iacr.org/2010/287.

Pat13. Jacques Patarin. Security in $O(2^n)$ for the Xor of Two Random Permutations: Proof with the Standard H Technique. IACR Cryptology ePrint Archive, Report 2013/368, 2013. Available at http://eprint.iacr.org/2013/368.

Sho96. Victor Shoup. On Fast and Provably Secure Message Authentication Based on Universal Hashing. In Neal Koblitz, editor, *Advances in Cryptology - CRYPTO '96*, volume 1109 of *LNCS*, pages 313–328. Springer, 1996.

Ste13. John Steinberger. Counting solutions to additive equations in random sets. arXiv Report 1309.5582, 2013. Available at http://arxiv.org/abs/1309.5582.

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.

A Basics of Discrete Fourier Analysis

We recall some classical results on Fourier analysis over the abelian group $\mathbb{Z}_2^n$, taken from [CLL+14]. In the following, given a subset $S \subset \{0,1\}^n$, we denote $\mathbb{1}_S : \{0,1\}^n \to \{0,1\}$ the characteristic functions of S, namely $\mathbb{1}_S(x) = 1$ if $x \in S$ and $\mathbb{1}_S(x) = 0$ if $x \notin S$. Given two functions $f, g : \{0,1\}^n \to \mathbb{R}$, we denote

$$\langle f, g \rangle = \mathbb{E}[fg] = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x)$$

the inner product of f and g, and, for all $x \in \{0,1\}^n$, we denote

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

the convolution of f and g. Given $\alpha \in \{0,1\}^n$, we denote $\chi_\alpha : \{0,1\}^n \to \{\pm 1\}$ the *character* associated with α defined as

$$\chi_\alpha(x) = (-1)^{\alpha \cdot x}.$$

The all-one character χ_0 is called the *principal character*. All other characters $\chi \neq 1$ corresponding to $\alpha \neq 0$ are called *non-principal characters*. The set of all characters forms a group for the pointwise product operation $(\chi_\alpha \chi_\beta)(x) = \chi_\alpha(x)\chi_\beta(x)$ and one has $\chi_\alpha \chi_\beta = \chi_{\alpha \oplus \beta}$.

Given a function $f : \{0,1\}^n \to \mathbb{R}$ and $\alpha \in \{0,1\}^n$, the *Fourier coefficient* of f corresponding to α is

$$\widehat{f}(\alpha) \stackrel{\text{def}}{=} \langle f, \chi_\alpha \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)(-1)^{\alpha \cdot x}.$$

The coefficient corresponding to $\alpha = 0$ is called the *principal Fourier coefficient*, all the other ones are called *non-principal Fourier coefficients*. Note that for a set $S \subseteq \{0,1\}^n$ one has

$$\widehat{\mathbb{1}_S}(0) = \frac{|S|}{2^n},$$

namely the principal Fourier coefficient of $\mathbb{1}_S$ is equal to the relative size of the set. We will also use the following three classical results, holding for any functions $f, g, : \{0,1\}^n \to \mathbb{R}$, any $\alpha \in \{0,1\}^n$, and any $S \subseteq \{0,1\}^n$:

$$\sum_{x \in \{0,1\}^n} f(x)g(x) = 2^n \sum_{\alpha \in \{0,1\}^n} \widehat{f}(\alpha)\widehat{g}(\alpha) \tag{21}$$

$$\widehat{(f * g)}(\alpha) = 2^n \widehat{f}(\alpha)\widehat{g}(\alpha) \tag{22}$$

$$\sum_{\alpha \in \{0,1\}^n} |\widehat{\mathbb{1}_S}(\alpha)|^2 = \frac{|S|}{2^n}. \tag{23}$$