How to Build Optimally Secure PRFs Using Block Ciphers

Benoît Cogliati[1], Ashwin Jha[2], and Mridul Nandi[2]

[1]CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
[2]Indian Statistical Institute, Kolkata, India
benoit.cogliati@cispa.saarland,{ashwin.jha1991,mridul.nandi}@gmail.com

Abstract. In EUROCRYPT '96, Aiello and Venkatesan proposed two candidates for $2n$-bit to $2n$-bit pseudorandom functions (PRFs), called Benes and modified Benes (or mBenes), based on n-bit to n-bit PRFs. While Benes is known to be secure up to 2^n queries (Patarin, AFRICACRYPT '08), the security of mBenes has only been proved up to $2^{n(1-\epsilon)}$ queries for all $\epsilon > 0$ by Patarin and Montreuil in ICISC '05. In this work, we show that the composition of a $2n$-bit hash function with mBenes is a secure variable input length (VIL) PRF up to 2^{n-2} queries (given appropriate hash function bounds). We extend our analysis with block ciphers as the underlying primitive and obtain two optimally secure VIL PRFs using block ciphers. The first of these candidates requires 6 calls to the block cipher. The second candidate requires just 4 calls to the block cipher, but here the proof is based on Patarin's mirror theory. Further, we instantiate the hash function with a PMAC+/LightMAC+ like hash, to get six candidates for deterministic message authentication codes with optimal security.

Keywords: PRF, MAC, Benes, modified Benes, PMAC+, LightMAC+

1 Introduction

Pseudorandom functions (PRF) over variable length inputs are keyed functions that take as input a bit string of arbitrary length and output a fixed length bit string that should be indistinguishable from uniformly random bits. This primitive is useful in practice as it can serve as a Message Authentication Code (MAC) in order to provide integrity and authenticity of messages. Moreover, when adequately combined with an encryption scheme (e.g. using the generic SIV structure [1]), it can also provide authenticated encryption. Unfortunately, barring a few examples like SURF [2], SipHash [3] and AES-PRF [4], building a concrete secure PRF from scratch has remained elusive.

Block Cipher-based PRF: Given the ubiquity of block ciphers (BC), building a provably secure PRF from block ciphers has been a widely studied problem in symmetric cryptography. As far as fixed input length (FIL) is concerned, the problem is essentially solved as several highly secure constructions already exist. For example, given two n-bit permutations Π_1 and Π_2, the following PRP-to-PRF constructions offer security up to (roughly) 2^n adversarial queries:

– the sum $x \mapsto \Pi_1(x) \oplus \Pi_2(x)$ of both permutations and its single-keyed variant the TWIN construction $x \mapsto \Pi_1(0||x) \oplus \Pi_1(1||x)$: after their introduction by Bellare et al. [5], their security has been the subject of a long line of research [5,6,7], culminating with [8,9] and [10] where optimal security has been proven;
– the Encrypted Davies-Meyer (EDM) construction $x \mapsto \Pi_2(\Pi_1(x) \oplus x)$ and its dual (EDMD) $x \mapsto \Pi_2(\Pi_1(x)) \oplus \Pi_1(x)$: EDM has been introduced in [11], and security up to roughly $2^n/n$ queries has been proven in [12], while EDMD has been designed and proven optimally secure in [12].

However, for the case of variable input length (VIL), very few constructions actually provide security beyond the birthday bound. The most notable exceptions are, the SUM-ECBC construction [13], the PMAC+ construction [14] and its single-key variant 1k-PMAC+ [15], 3kf9 [16] and LightMAC+ [17] since they offer beyond the birthday bound (but still suboptimal) security. Those modes of operations use the relatively new Double-block Hash-then-Sum or DbHtS paradigm [18], which applies n-bit block cipher calls to the two n-bit halves of a $2n$-bit hash function and then sums the encrypted output. Although the DbHtS paradigm is known to achieve very high security [19,20], it is not yet known whether it can achieve optimal security. A more traditional approach towards PRF construction is the classical Hash-then-PRF paradigm [21], that relies on an n-bit block cipher along with two other components:

– a hash function with $2n$-bit output; and
– a $2n$-bit to n-bit PRF.

Designing the latter primitive is deeply linked to the problem of domain extension for PRFs, which has also been the subject of a long line of research. Since we focus on the problem of designing an optimally secure construction from a block cipher, this restricts the set of possible finalization constructions to the Benes construction and its variants [22], and Feistel networks with at least four rounds [23][1]. Unfortunately, optimal security for Feistel networks when round functions are instantiated with PRPs still remains to be proven. Hence, using Feistel networks as a finalization function would require implementing the round PRFs as the xor of two permutations, thus increasing the number of block cipher calls to 8. As we will see, considering other structures will allow the design of more efficient schemes.

Benes and modified Benes: In [22], Aiello and Venkatesan introduced the Benes and modified Benes (or mBenes) constructions that build a $2n$-bit to n-bit PRF[2] from respectively 6 and 4 independent n-bit PRFs, where each underlying PRF is called once for each call to the construction. Patarin showed that Benes transformation is n-bit secure [24]. For mBenes, although Aiello and Venkatesan conjecture n-bit security, until now only a high level proof idea is shown [25,24]

[1] The actual Feistel networks are from $2n$-bit to $2n$-bit. In that case, 5 rounds are required for optimal security. Since we only require n-bit outputs, the final round can actually be dropped.

[2] The actual Benes and mBenes constructions are from $2n$-bit to $2n$-bit, requiring 8 and 6 calls respectively (see section 3 for details). For now, just n-bit output suffices.

for security up to (roughly) $2^{n(1-\epsilon)}$ queries for all $\epsilon > 0$. In order to use PRPs as the underlying primitive in Benes and mBenes while keeping optimal security, the most obvious solution would be to rely on an optimally secure PRP-to-PRF conversion method. However, this would increase the number of PRP calls of the construction to 12 for the Benes construction, and 8 for the mBenes construction. Current proof techniques unfortunately are not sufficient to prove optimal security for PRP-based Benes and mBenes constructions using a smaller number of permutation calls. Indeed, the current best result by Jha and Nandi shows that mBenes using 4 block ciphers is secure up to $2^{3n/4}$ queries [19].

1.1 Our Contributions

Table 1.1: Summary of beyond-the-birthday bound secure variable input length pseudorandom functions. Here ℓ denotes the length of the input message after padding.

Scheme	Primitive		Security	
	Type	No. of calls	Bound	Restriction
3kf9 [16]	PRP	$\ell+2$	$O\left(\frac{\ell^2 q^{4/3}}{2^n}\right)$ [20]	–
PMAC+ [14]	PRP	$\ell+2$	$O\left(\frac{q^{4/3}\ell^{2/3}+\ell^2 q}{2^n}\right)$ [20]	$\ell \ll 2^{n/2}$ [20]
1k-PMAC+ [15]	PRP	$\ell+2$	$O\left(\frac{q\sigma^2}{2^{2n}}\right)$	–
LightMAC+ [1] [17]	PRP	$2\ell+2$	$O\left(\frac{q^{4/3}}{2^n}\right)$ [20]	–
LightMac+2[1] [17]	PRP	$2\ell+2+t$	$O\left(\frac{q^{t+1}}{2^t}\right)$	$t \le 7;\ \ell = O\left(2^{n/2}\right)$
mPMAC+-f	PRF/PRP	$\ell+3$	$O\left(\frac{\sigma}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$
mPMAC+-p1	PRP	$\ell+5$	$O\left(\frac{\sigma}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$
mPMAC+-p2	PRP	$\ell+3$	$O\left(\frac{\sigma}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$
mLightMAC+-f [1]	PRF/PRP	$2\ell+3$	$O\left(\frac{q}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$
mLightMAC+-p1 [1]	PRP	$2\ell+5$	$O\left(\frac{q}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$
mLightMAC+-p2 [1]	PRP	$2\ell+3$	$O\left(\frac{q}{2^n}\right)$	$\ell = O\left(2^{n/2}\right)$

[1] In order to simplify the comparison, we focus on the case $m = n/2$ for LightMAC+-based constructions.

Our contribution is twofold. First, we introduce a novel construction dubbed HtmB for Hash-then-modified-Benes. This construction captures the design of a VIL-PRF based on a FIL primitive where the input is first hashed, then given as input to mBenes. This hashing step is what allows us to avoid the main difficulties that are encountered when one tries to prove optimal security for the mBenes construction. In more details, we introduce a new statistical property for hash functions with $2n$-bit outputs: Diblock Almost q-Collision-free Universality or DbACU_q (see section 2.2). We then show that the composition of a DbACU_q

hash function and the mBenes construction is n-bit secure (see section 4), and propose several extensions:

- HtmB-f: the standard HtmB construction based on 4 functions;
- HtmB-p1: the HtmB construction where two functions are replaced with permutations, and the remaining ones are replaced with the sum of two permutations
- HtmB-p2: the standard HtmB based on 4 permutations.

It is worth noting that the security proofs for the first two constructions are straightforward and rely on the same technique as Patarin's classical proofs for Benes [24]. The security proof for the last construction relies on the fundamental result of Mirror Theory [9, Theorem 6]. Note that DbACU_q can be easily achieved by concatenation of two independent almost universal (AU) hash functions. Moreover, we will show two instances where this property is also achieved for concatenation of dependent AU hash functions.

Second, we define two families of block cipher modes of operation dubbed mLightMAC+ and mPMAC+ (see section 5). Both are concrete instantiations of HtmB where the hashing algorithm is based respectively on the LightMAC+ and PMAC+ algorithms. In more details, both schemes are provably secure PRFs with n-bit output and have the following properties:

- mPMAC+ processes n bits of (padded) input per block cipher call during the hashing phase and is secure as long as the number of (padded) queried blocks is small in front of 2^n and no query is longer than $2^{n/2}$ blocks;
- for any fixed integer $m \in \{1, \ldots, n-1\}$, mLightMAC+ processes $n-m$ bits of input per block cipher call during the hashing phase and is secure as long as the number of adversarial queries is small in front of 2^{n3} and no query is longer than 2^m blocks.

Table 1.1 summarizes this information and compares our modes with the original LightMAC+ and PMAC+ constructions, while Fig. 1.1 highlights the changes between mPMAC+-p2, our mPMAC+ instantiation based on HtmB-p2, and the original PMAC+ construction.

In [26], Naito proposed a PMAC variant based on PMAC+ like masking and claimed length-independent bounds on the collision probability of the underlying hash layer. However, the proof is incorrect owing to a flaw identified in [27], and apparently it cannot be fixed within the proof setup developed in [26] (see [27] for further details). Consequently, in section 6.2, we first discuss this flaw and then derive a slightly worse bound which is still sufficient to prove optimal security of mPMAC+.

The key sizes in HtmB could be an issue in some memory-constrained environments. In section 7, we address this problem and present some variants of HtmB that require lesser key material. Finally, we conclude in section 8 with some open problems.

[3] Note that this is true regardless of the total length of all adversarial queries.

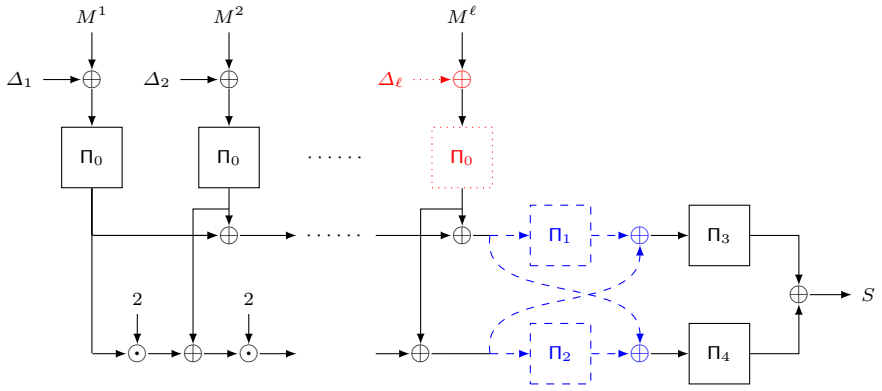

Fig. 1.1: Schematic of mPMAC+-p2, operating over a padded message of length ℓn bits. $\Pi_0, \ldots, \Pi_4$ are independent random permutations, and $\Delta_i = 2^i \odot \Pi_0(0) \oplus 2^{2i} \odot \Pi_0(1)$, where $\odot$ denotes the multiplication operator of $\mathrm{GF}(2^n)$. Components drawn in blue dashed lines represent the addition over the original PMAC+ construction. Components drawn in red dotted lines represent the deletion over the original PMAC+ construction. Note that the modified hash layer saves one block cipher call as compared to the one in PMAC+.

2 Preliminaries

Notational Setup: For $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, \ldots, n\}$, and $\{0,1\}^n$ denotes the set of bit strings of length n. Let $\mathsf{GF}(2^n)$ be the field of order 2^n. We identify bit string and finite field element of $\mathsf{GF}(2^n)$ by representing the string $a = a_{n-1} \ldots a_0 \in \{0,1\}^n$ as polynomial $a(x) = a_{n-1}x^{n-1} + \ldots + a_0 \in \mathsf{GF}(2^n)$ and vice versa. As usual, we define field addition $\oplus$ as polynomial addition, and multiplication $\odot$ as polynomial multiplication modulo the irreducible polynomial $f(x)$ used to represent $\mathsf{GF}(2^n)$. Therefore, we can view $\{0,1\}^n$ as the finite field $\mathsf{GF}(2^n)$ with $\oplus$ as field addition and $\odot$ as field multiplication. When the context is clear, we will denote by 2 the primitive element of $\mathsf{GF}(2^n)$. The set of all bit strings (including the empty string) is denoted $\{0,1\}^*$, and $|X|$ denotes the number of bits in $X \in \{0,1\}^*$. For any integer m, $\{0,1\}^{\leq m}$ denotes the set of all bit strings of bit length at most m. For $n \in \mathbb{N}$ and any two bit strings M and M', we denote by $M||M'$ the concatenation of M and M', and we define $\mathsf{pad}(M)$ as $M||10\cdots0$, such that $|\mathsf{pad}(M)|$ is the smallest multiple of n that is greater than $|M|$. For $i, m \in \mathbb{N}$ such that $i < 2^m$, we define $< i >_m$ as the m-bit little endian encoding of the integer i. For $n, r \in \mathbb{N}$, such that $0 \leq r \leq n$, we define the falling factorial $(n)_r := n!/(n-r)! = n(n-1)\cdots(n-r+1)$. The set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ is denoted $\mathcal{F}(\mathcal{X}, \mathcal{Y})$, and the set of all permutations of $\mathcal{X}$ is denoted $\mathcal{P}(\mathcal{X})$. We simply write $\mathcal{F}(a, b)$ and $\mathcal{P}(a)$, whenever $\mathcal{X} = \{0,1\}^a$ and $\mathcal{Y} = \{0,1\}^b$. For a finite set $\mathcal{X}$, $X \leftarrow\!\!\$\ \mathcal{X}$ denotes the uniform at random sampling of X from $\mathcal{X}$. For any property P of some random variable X, $\Pr[P[X]]$ denotes the probability that $P[X]$ is satisfied.

For $q \in \mathbb{N}$, X^q denotes the q-tuple $(X_1, X_2, \ldots, X_q)$. By an abuse of notation we also use X^q to denote the multiset $\{X_i : i \in [q]\}$. For $q \in \mathbb{N}$, for any set $\mathcal{X}$, $(\mathcal{X})_q$ denotes the set of all q-tuples with distinct elements from $\mathcal{X}$. For a pair of tuples X^q and Y^q, (X^q, Y^q) denotes the 2-ary q-tuple $((X_1, Y_1), \ldots, (X_q, Y_q))$. An n-ary q-tuple is defined analogously. For any tuple $X^q \in \mathcal{X}^q$, and for any function $f : \mathcal{X} \to \mathcal{Y}$, $f(X^q)$ denotes the tuple $(f(X_1), \ldots, f(X_q))$.

2.1 Keyed Functions and Block Ciphers

Keyed Function: A $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-*keyed function* F 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 write $F_K(X)$ for $F(K, X)$.

Block Cipher: A $(\mathcal{K}, \{0,1\}^n)$-*block cipher* E with key space $\mathcal{K}$ and block space $\{0,1\}^n$ is a $(\mathcal{K}, \{0,1\}^n, \{0,1\}^n)$-keyed function, such that for any key $K \in \mathcal{K}$, $X \mapsto E(K, X)$ is a permutation of $\{0,1\}^n$. We write $E_K(X)$ for $E(K, X)$.

Security Definitions: A (q, t)-distinguisher is an interactive algorithm with access to an oracle, that makes at most q oracle queries, runs in time at most t, and outputs a single bit. By convention, $t = \infty$ denotes computationally unbounded (information-theoretic) and deterministic distinguishers. In this paper, we assume that the distinguisher never makes a duplicate query.

Pseudorandom Function: The pseudorandom function or PRF advantage of any distinguisher $\mathscr{A}$ against a $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function F is defined as

$$\mathbf{Adv}^{\mathsf{prf}}_{F}(\mathscr{A}) = \mathbf{Adv}_{F;\mathsf{\Gamma}}(\mathscr{A}) := \left| \Pr_{\mathsf{K} \xleftarrow{\$} \mathcal{K}} \left[\mathscr{A}^{F_{\mathsf{K}}} = 1\right] - \Pr_{\mathsf{\Gamma} \xleftarrow{\$} \mathcal{F}(\mathcal{X},\mathcal{Y})} \left[\mathscr{A}^{\mathsf{\Gamma}} = 1\right] \right|. \quad (1)$$

Deterministic message authentication codes (or MAC) are keyed functions which provide both integrity and authenticity of data. It is a well-known fact [28] that a secure PRF is a good candidate of deterministic MAC.

Pseudorandom Permutation: The *pseudorandom permutation* or PRP advantage of any distinguisher $\mathscr{A}$ against a $(\mathcal{K}, \{0,1\}^n)$-block cipher E is defined as

$$\mathbf{Adv}^{\mathsf{prp}}_{E}(\mathscr{A}) = \mathbf{Adv}_{E;\mathsf{\Pi}}(\mathscr{A}) := \left| \Pr_{\mathsf{K} \xleftarrow{\$} \mathcal{K}} \left[\mathscr{A}^{E_{\mathsf{K}}} = 1\right] - \Pr_{\mathsf{\Pi} \xleftarrow{\$} \mathcal{P}(n)} \left[\mathscr{A}^{\mathsf{\Pi}} = 1\right] \right|. \quad (2)$$

Remark 2.1. All our results will be given in the information-theoretic setting, and their computational counterparts can be easily obtained via a boilerplate hybrid argument. In other words, instead of first starting with block ciphers (or PRFs), we will directly work with random permutations (or functions) as the underlying primitives.

Sum of Permutations: In 1998, two independent works [5,29] on building PRFs from PRPs proposed the Sum of Permutation (SoP) construction. For two independent random permutations $\mathsf{\Pi}_1, \mathsf{\Pi}_2 \xleftarrow{\$} \mathcal{P}(n)$, the SoP, denoted $\mathsf{\Pi}_1 \oplus \mathsf{\Pi}_2$, is defined as the mapping $X \mapsto \mathsf{\Pi}_1(X) \oplus \mathsf{\Pi}_2(X)$. After several attempts [6,7,9],

Dai et al. [10] finally showed that SoP is a secure PRF up to 2^n queries. In Proposition 2.1, we restate the well-known and celebrated result of [10]. A proof of Proposition 2.1 is available in [10].

Proposition 2.1. *For $n \geq 4$, $q \leq 2^{n-4}$, and all (q, ∞)-distinguisher $\mathscr{A}$ we have*

$$\mathbf{Adv}^{\mathsf{prf}}_{\Pi_1 \oplus \Pi_2}(\mathscr{A}) \leq \frac{q^{1.5}}{2^{1.5n}}.$$

2.2 Universal Hash Functions

We recall the usual definition of universal hash function. A $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function H is said to be *ϵ-almost universal (AU) hash function* if for any distinct $X, X' \in \mathcal{X}$, we have

$$\Pr_{\mathsf{K} \xleftarrow{\$} \mathcal{K}}[H_{\mathsf{K}}(X) = H_{\mathsf{K}}(X')] \leq \epsilon. \tag{3}$$

Let us fix a non-empty set $\mathcal{X} \subset \{0,1\}^*$. In this article, we are going to consider a slightly more general notion of universality. Namely, let H be a $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function that processes its inputs in n-bit blocks. H is said to be (q, σ, ϵ)-*Almost θ-Collision-free Universal* (or ACU_θ) if, for every $X^q \in (\mathcal{X})_q$ such that X^q contains at most σ blocks, one has $\Pr[C \geq \theta] \leq \epsilon$, where

$$C := |\{(i,j) \,:\, 1 \leq i < j \leq q,\, H_K(X_i) = H_K(X_j)\}|.$$

In the case of a (q, σ, ϵ)-ACU_1 hash function H, we simply say that H is (q, σ, ϵ)-AU. Note that if $q = 2$, we recover the standard AU notion. Moreover, the following proposition is a simple application of Markov's inequality.

Proposition 2.2. *For $q, \theta \in \mathbb{N}$ and $0 \leq \epsilon \leq 1$, let H be an ϵ-AU hash function. Then H is $(q, \infty, \frac{q^2\epsilon}{\theta})$-$ACU_\theta$.*

The proof of Proposition 2.2 follows from Markov's inequality and is thus skipped here.

We also define a new combined notion for the concatenation of two hash function. Namely, we say that a pair $H = (H_1, H_2)$ of two $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed hash functions H_1, H_2 is $(q, \sigma, \epsilon_2, \epsilon_1)$-*Diblock ACU_q* (or DbACU_q) if H is (q, σ, ϵ_2)-AU and H_1, H_2 are (q, σ, ϵ_1)-ACU_q. A simple example of DbACU_q hash function is the concatenation of two independent AU hash functions. In section 5, we present two other DbACU_q hash functions LightHash and PHash based respectively on the LightMAC+ and PMAC+ constructions.

The concatenation of two independent AU hash functions: Let H_1 and H_2 be two ϵ-AU hash functions with key space $\mathcal{K}$, message space $\mathcal{X}$ and range $\mathcal{Y}$. We define the concatenation $H = (H_1, H_2)$ of H_1 and H_2 as a $(\mathcal{K}^2, \mathcal{X}, \mathcal{Y}^2)$-keyed function defined as $H_{(K_1,K_2)}(X) = (H_{1,K_1}(X), H_{2,K_2}(X))$ for every $X \in \mathcal{X}$, $(K_1, K_2) \in \mathcal{K}^2$. The following result holds.

Proposition 2.3. *Let H_1, H_2 be two ϵ-AU hash functions keyed independently and $H = (H_1, H_2)$. For $q, \sigma \in \mathbb{N}$, H is $(q, \sigma, q^2\epsilon^2, q\epsilon)$-$DbACU_q$.*

A proof of Proposition 2.3 relies on the independence of both components and on Proposition 2.2.

2.3 Coefficient-H Technique

The coefficient-H technique by Patarin [30,31] is a tool to upper bound the distinguishing advantage of any deterministic and computationally unbounded distinguisher $\mathscr{A}$ in distinguishing the real oracle $\mathcal{R}$ from the ideal oracle $\mathcal{I}$. The collection of all queries and responses that $\mathscr{A}$ made and received to and from the oracle, is called the transcript of $\mathscr{A}$, denoted as τ.

Let $\mathbb{T}_{\mathsf{re}}$ and $\mathbb{T}_{\mathsf{id}}$ denote the transcript random variable induced by $\mathscr{A}$'s interaction with $\mathcal{R}$ and $\mathcal{I}$, respectively. Let $\mathcal{T}$ be the set of all transcripts. A transcript $\tau \in \mathcal{T}$ is said to be *attainable* if $\Pr[\mathbb{T}_{\mathsf{id}} = \tau] > 0$, i.e., it can be realized by $\mathscr{A}$'s interaction with $\mathcal{I}$. Following these notations, we state the main result of coefficient-H technique in Theorem 2.1. A proof of this theorem is available in [32,4], among others.

Theorem 2.1. *For $\epsilon_1, \epsilon_2 \geq 0$, suppose there is a set $\mathcal{T}_{\mathsf{bad}} \subseteq \mathcal{T}$, that we call the set of bad transcripts, such that the following conditions hold:*

- $\Pr[\mathbb{T}_{\mathsf{id}} \in \mathcal{T}_{\mathsf{bad}}] \leq \epsilon_1$*; and*
- *For any $\tau \notin \mathcal{T}_{\mathsf{bad}}$, τ is attainable and* $\dfrac{\Pr[\mathbb{T}_{\mathsf{re}} = \tau]}{\Pr[\mathbb{T}_{\mathsf{id}} = \tau]} \geq 1 - \epsilon_2$.

Then, for any computationally unbounded and deterministic distinguisher $\mathscr{A}$, we have

$$\mathbf{Adv}_{\mathcal{R};\mathcal{I}}(\mathscr{A}) \leq \epsilon_1 + \epsilon_2.$$

3 Benes and mBenes Transformations

Butterfly transformation: Given four functions $f_1, \ldots, f_4 \in \mathcal{F}(n,n)$, the Butterfly transformation (illustrated in Fig. 3.1) is a function from $\{0,1\}^{2n}$ to $\{0,1\}^{2n}$, which is defined as $\mathsf{Butterfly}[f_1, \ldots, f_4](L,R) := (X,Y)$, where

$$X := f_1(L) \oplus f_2(R) \text{ and } Y := f_3(L) \oplus f_4(R).$$

Benes transformation: Given eight functions $f_1, \ldots, f_8 \in \mathcal{F}(n,n)$, the Benes transformation (illustrated in Fig. 3.1) is a function from $\{0,1\}^{2n}$ to $\{0,1\}^{2n}$, which is defined as the composition of two Butterfly transformations, i.e. $\mathsf{Benes}[f_1, \ldots, f_8](L,R) := (S,T)$, where

$$\begin{aligned} S &:= f_5(f_1(L) \oplus f_2(R)) \oplus f_6(f_3(L) \oplus f_4(R)) = f_5(X) \oplus f_6(Y), \\ T &:= f_7(f_1(L) \oplus f_2(R)) \oplus f_8(f_3(L) \oplus f_4(R)) = f_7(X) \oplus f_8(Y). \end{aligned}$$

Modified Benes transformation: The modified Benes or mBenes transformation (illustrated in Fig. 3.1) is a simplification of the Benes transformation, where f_2 and f_3 are identity functions. So, we have $X = f_1(L) \oplus R$, $Y = f_4(R) \oplus L$, and $(S,T) = \mathsf{mBenes}[f_1, f_4, f_5, \ldots, f_8](L,R)$, such that $S = f_5(X) \oplus f_6(Y)$ and $T = f_7(X) \oplus f_8(Y)$.

For brevity we drop the parameters $f_1, \ldots, f_8$, whenever they are understood from the context.

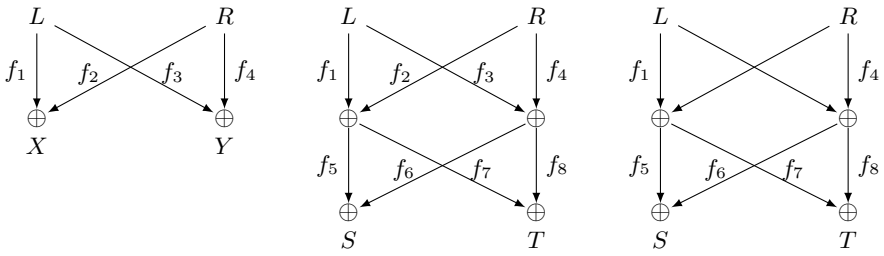

Fig. 3.1: Left to right: Butterfly, Benes and mBenes transformations. An edge (u, v) with label g denotes the mapping $v = g(u)$. Unlabelled edges are identity mapping.

3.1 Revisiting the Security Analysis of Benes and mBenes

Let (L^q, R^q) denote a q-tuple of inputs. Given $f_1, \ldots, f_4 \in \mathcal{F}(n, n)$, we can define (X^q, Y^q) by the definition of Benes or mBenes, as applicable.

Dependency Graph: To (L^q, R^q) and any $f_1, \ldots, f_4 \in \mathcal{F}(n, n)$, we associate the dependency graph $\mathcal{G}[L^q, R^q; f_{1,\ldots,4}] = ([q], \mathcal{E})$, over the set of all query indices $[q]$, where $\{i, j\} \in \mathcal{E}$ if and only if $X_i = X_j$ (the edge is colored red) or $Y_i = Y_j$ (the edge is colored blue). $\mathcal{G}[L^q, R^q; f_{1,\ldots,4}]$ may contain parallel edges, but their coloring will be different. Fig. 3.2 is a possible dependency graph for $q = 12$.

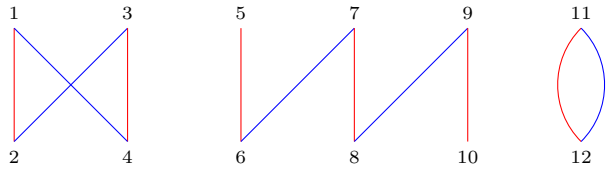

Fig. 3.2: A possible dependency graph for some 12-tuple of inputs.

Definition 3.1 (Alternating cycle). *An alternating cycle or circle of length $k \geq 2$, k even, is simply a cycle denoted by a sequence of $k + 1$ indices, $v^{k+1} = (v_1, \ldots, v_k, v_{k+1})$ such that*

- $v_{k+1} = v_1$,
- $\{v_i, v_{i+1}\} \in \mathcal{E}$ *for all* $i \in [k]$,
- $\{v_1, v_2\}$ *is colored red, and*
- $\{v_i, v_{i+1}\}$ *and* $\{v_{i+1}, v_{i+2}\}$ *do not share the same color, for all* $i \in [k-1]$.

Example 3.1. Any parallel edge is an example of alternating cycle. In Fig. 3.2, $(1, 2, 3, 4, 1)$ and $(11, 12, 11)$ are two possible alternating cycles.

Let $\mathtt{AC}[L^q, R^q; f_{1,\ldots,4}]$ denote the property that $\mathcal{G}[L^q, R^q; f_{1,\ldots,4}]$ contains an alternating cycle. We will drop the parameters $(L^q, R^q; f_{1,\ldots,4})$, whenever they are understood from the context.

For $f_1, \ldots, f_8 \xleftarrow{\$} \mathcal{F}(n, n)$, Aiello and Venkatesan [22] showed that PRF advantage of any distinguisher against Benes and mBenes is at most the probability

that AC is satisfied. Similar results were later also shown in [25,24]. Theorem 3.1 is a reformulation of [22, Lemma 2] (also [25, Theorem 5.2] and [24, Theorem 1]) in our notations.

Theorem 3.1. *For* $\Gamma_1 \ldots, \Gamma_8 \leftarrow_{\$} \mathcal{F}(n,n)$, $F \in \{\mathsf{Benes}, \mathsf{mBenes}\}$, *and any* (q,∞)*-distinguisher* $\mathscr{A}$, *we have*

$$\mathbf{Adv}^{\mathsf{prf}}_{F[\Gamma_1,\ldots,\Gamma_8]}(\mathscr{A}) \leq \mathrm{ACP}(q) := \max_{(L^q,R^q)} \Pr_{\Gamma_1,\ldots,\Gamma_4} [\mathtt{AC}[L^q,R^q;\Gamma_{1,\ldots,4}]].$$

A proof of Theorem 3.1 is available in [25] among others. For the sake of completeness, we reproduce it in the full version of this paper.

Aiello and Venkatesan [22] claimed that $\mathrm{ACP}(q) \leq q^2/2^{2n}$. Later, Patarin and Montreuil [25] showed that the initial analysis of $\mathrm{ACP}(q)$ by Aiello and Venkatesan was overly optimistic, and subsequently gave a non-tight estimate for Benes. The main idea of their analysis was to consider each equation in the alternating cycle, one-by-one, distinguishing whether the equation is dependent over the previous equations or not. If the i-th equation is independent then they freely choose the new index[4], i.e., $(i+1)$-th index in $q-i$ ways. However, when the equation is dependent, then there exist $j, j' < i$ such that $L_i = L_j$ and $R_i = R_{j'}$, hence we only have $i(i-1)$ ways to choose the $(i+1)$-th index. By continuing in this way and making some algebraic simplifications, they derive the upper bound

$$\mathrm{ACP}(q) \leq d(k)\frac{q^2}{2^{2n+1}} + \frac{q^4}{2^{4n+2}} + \frac{q^{k+1}}{2^{nk}},$$

for all $k \geq 1$, where $d(k) = 6.5 + \sum_{j=6}^{k} j^{2j} + k^{2k}$. So, for any k and sufficiently large n, we can claim security up to $q \leq \min\{2^{nk/k+1}, \sqrt{2^{2n}/d(k)}\}$. However, the bound becomes increasingly moot as we increase the value of k. Suppose we aim for security up to $2^{kn/k+1}$ queries. Then, for $k = 6$ we need $n > 112$, for $k = 7$ we need $n > 161$, and for $k = 9$ we need $n > 290$, where n denotes the output size of the underlying functions. Clearly, very high security (close to $0.9n$) is only possible for large output size ($n > 290$). In practice, with such a large output size, even a birthday bound security guarantee might suffice.

Patarin and Montreuil also claimed similar security bounds for mBenes [25]. However, they only gave a very high level and terse sketch of the proof. We refer the readers to [25] for details.

First Dependency and Tight Bound for Benes: Patarin [24] devised an elegant way to derive a more tighter estimate for $\mathrm{ACP}(q)$ in case of Benes.

Definition 3.2 (Alternating trail). *An alternating trail or line of length* $k \geq 2$ *is simply a trail denoted by a sequence of* $k+1$ *vertices,* $v^{k+1} = (v_1, \ldots, v_k, v_{k+1})$ *such that*

– $\{v_i, v_{i+1}\} \in \mathcal{E}$, *for all* $i \in [k]$.

[4] Each equation (except the last one) gives a new index.

– $\{v_i, v_{i+1}\}$ *and* $\{v_{i+1}, v_{i+2}\}$ *do not share the same color, for all* $i \in [k-1]$.

In addition, we say that v^{k+1} *is a red (res. blue) trail if* $\{v_1, v_2\}$ *is colored red (res. blue).*

Example 3.2. An alternating cycle is in fact a special type of alternating red trail with even length. In Fig. 3.2, $(1, 2, 3, 4, 1)$, $(5, 6, 7, 8, 9, 10)$, and $(11, 12, 11)$ are some of the possible alternating trails. Note that all these trails are red trails. On the other hand, $(2, 3, 4, 1, 2)$ is a blue trail.

Associated System of Equations: By definition, each edge in the dependency graph $\mathcal{G}$ corresponds to an equation. For example, say we have an edge $\{u, v\}$ with red color, then the associated equation is $X_u = X_v$. By extension, each connected component corresponds to a system of equations. In particular, any alternating trail (or cycle) v^{k+1} can be uniquely associated with a system of k equations. For example, suppose v^{k+1} is an alternating red trail of even length. Then, the associated system of equation is $X_{v_1} = X_{v_2}, \ldots, Y_{v_k} = Y_{v_{k+1}}$.

Example 3.3. In Fig. 3.2, we can have the following associated system of equations:

– For alternating cycle $(1, 2, 3, 4, 1)$: $X_1 = X_2,\ Y_2 = Y_3,\ X_3 = X_4,\ Y_4 = Y_1$.
– For alternating trail $(5, 6, 7, 8, 9, 10)$: $X_5 = X_6,\ Y_6 = Y_7,\ X_7 = X_8,\ Y_8 = Y_9,\ X_9 = X_{10}$.
– For parallel edge $(11, 12, 11)$: $X_{11} = X_{12},\ Y_{12} = Y_{11}$.

Definition 3.3 (First dependency [24]). *An alternating trail of length* $k \geq 2$ *is said to have first dependency if all the equations in the associated system of equations, except the last one are independent of others, and the last equation is a consequence of the previous equations.*

An alternating cycle of length $k \geq 2$ *is said to have first dependency if all the equations in the associated system of equations, except one are independent of others, and exactly one is a consequence of the other equations.*

Example 3.4. In Fig. 3.2, suppose $L_5 = L_9,\ L_6 = L_{10},\ R_5 = R_6,\ R_9 = R_{10}$. Then, $X_5 = X_6$ holds if $f_1(L_5) = f_1(L_6)$ (as $R_5 = R_6$). Similarly, $X_9 = X_{10}$ holds if $f_1(L_9) = f_1(L_{10})$ (as $R_9 = R_{10}$). But, $L_9 = L_5$ and $L_{10} = L_6$. Thus, $X_9 = X_{10}$ is a consequence of $X_5 = X_6$. Hence, $X_5 = X_6,\ Y_6 = Y_7,\ X_7 = X_8,\ Y_8 = Y_9,\ X_9 = X_{10}$ is an alternating trail of length 5 with first dependency.

Any alternating cycle of length k must have one of the following:

1. All the equations in the associated system of equations are independent.
2. The cycle has first dependency, i.e., all equations are independent except one.
3. The cycle contains an alternating trail of length $< k$ which has first dependency.

The first case is easy to bound as we have to choose k indices and we have k independent equations, which gives $O(q^k/2^{nk})$ bound. The second case is similar

to the last one, which is more general. Patarin argued that whenever an alternating trail has first dependency, then among the $k+1$ indices at least two are fixed once the other $k-1$ indices are chosen. Indices 6 and 9, for instance, are fixed once we choose indices 5 and 10 in Example 3.4. This observation immediately gives a bound of the form $O(q^{k-1}/2^{n(k-1)})$, since the first $k-1$ equations are independent. On combining the three cases, Patarin obtained the following bound on $\mathrm{ACP}(q)$ in case of Benes.

$$\mathrm{ACP}(q) \leq \frac{8590q^2}{2^{2n}} \tag{4}$$

Notice the large constant in the bound, which compels large n to get appreciable security in practice. The main component of this constant is an infinite sum $\sum_{k=3}^{\infty}\left(\frac{k^5}{2^{k-3}}\right)$. For large k, we observed that this sum can be approximated to 8588. In the same paper, Patarin also gave another improved bound [24, Theorem 9] using a more involved analysis which can be approximated to $26q^2/2^{2n} + 200076q^3/2^{4n}$ for large k.

First Dependency in mBenes: While the first dependency idea is quite useful for deriving tight security bound of Benes, Patarin noted that the same is not true in case of mBenes. In fact, a crucial argument—*among the* $k+1$ *indices* 2 *indices are fixed once we fix* $k-1$ *indices*—fails in case of mBenes. For example, suppose $X_1 = X_2,\ Y_2 = Y_3,\ X_3 = X_4$ is an alternating trail with first dependency, such that $L_1 = L_3$, $L_2 = L_4$, and $R_1 \oplus R_2 \oplus R_3 \oplus R_4 = 0$. It is clear to see that here only one index is fixed given the other three ($L_4 = L_2$ and $R_4 = R_1 \oplus R_2 \oplus R_3$). Consequently, Patarin speculates:

> *Therefore, a proof of security in* $O(2^n)$ *for the Modified Benes will be different, and probably more complex than our proof of security on* $O(2^n)$ *for the regular Benes.*

4 HtmB: Hash then modified Benes

Section 3 gives a clear indication that the exact security of mBenes is a difficult problem. The main difficulty in the analysis is a simple fact that the distinguisher has complete control over the inputs to mBenes. However, in practice PRFs are mostly required to work over arbitrary domains, which requires an additional preprocessing phase before the application of fixed input length PRF. This preprocessing is often done via a universal hash function—the so-called Hash-then-PRF paradigm [21]. This added layer of preprocessing somewhat curtails the distinguisher's ability to control the inputs to mBenes. Indeed, now we show that the composition of a universal hash function with mBenes leads to optimal security, with domain extension as byproduct.

Hash-then-modified-Benes: Let $\mathcal{M} \subseteq \{0,1\}^*$. Given a pair $H = (H_1, H_2)$ of two $(\mathcal{K}, \mathcal{M}, \{0,1\}^n)$-keyed hash functions ($H_1$ and H_2 may share the same key),

and $f_1, \ldots, f_4 \in \mathcal{F}(n,n)$, the Hash-then-modified-Benes or HtmB transformation is a function from $\mathcal{M}$ to $\{0,1\}^n$, which is defined as $\mathsf{HtmB}[H, f_1, \ldots, f_4](M) := S$, where

$$(L,R) := H_K(M) \quad X := f_1(L) \oplus R \quad Y := f_2(R) \oplus L \quad S = f_3(X) \oplus f_4(Y). \tag{5}$$

Remark 4.1. Note that, we reduced the output length of HtmB from $2n$ bits to n bits. This is mainly due to the fact that n bits of the output of a VIL PRF is sufficient to achieve 2^n query deterministic MAC security (a major inspiration for this work). In any case, another n-bit block can be easily generated by setting $T = f_5(X) \oplus f_6(Y)$ for some $f_5, f_6 \in \mathcal{F}(n,n)$.

We extend the dependency graph of section 3.1 to incorporate the hash function H. To any input $M^q \in (\mathcal{M})_q$, $K \in \mathcal{K}$, and $f_1, f_2 \in \mathcal{F}(n,n)$, we associate the dependency graph $\mathcal{G}[M^q; K, f_{1,2}] = ([q], \mathcal{E})$, where $\mathcal{E}$ is defined as before. Thus, $\mathcal{G}$ is again a bichromatic graph. We define $\mathtt{AC}[M^q; K, f_{1,2}]$, $\mathrm{ACP}(q)$, alternating trails, cycles, and the first dependency property analogously as in section 3.1. In the following subsections we present three security results on HtmB based on the choice of $f_1, \ldots, f_4$.

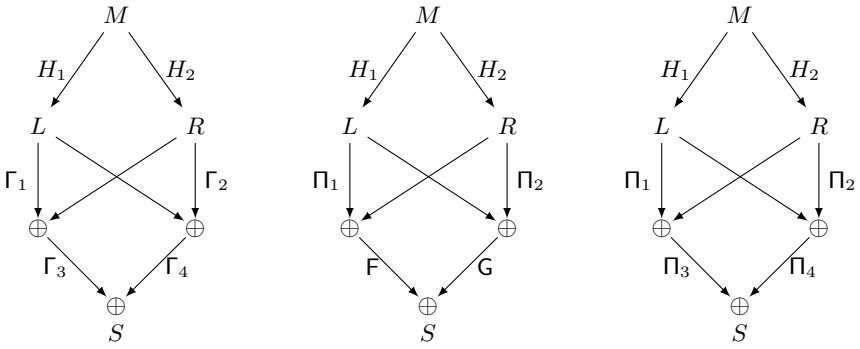

Fig. 4.1: The three instantiations of Hash-then-modified-Benes or HtmB transformation. $H = (H_1, H_2)$ is a DbACU_q hash function. From left to right: $\mathsf{HtmB\text{-}f}[H, \Gamma_1, \ldots, \Gamma_4] = \mathsf{HtmB}[H, \Gamma_1, \ldots, \Gamma_4]$ based on $\Gamma_1, \ldots, \Gamma_4 \leftarrow\!\!\$\, \mathcal{F}(n,n)$; $\mathsf{HtmB\text{-}p1}[H, \Pi_1, \ldots, \Pi_6] = \mathsf{HtmB}[H, \Pi_1, \Pi_2, \mathsf{F}, \mathsf{G}]$ based on $\Pi_1, \ldots, \Pi_6 \leftarrow\!\!\$\, \mathcal{P}(n)$, where $\mathsf{F}(X) = \Pi_3(X) \oplus \Pi_4(X)$ and $\mathsf{G}(Y) = \Pi_5(Y) \oplus \Pi_6(Y)$; and $\mathsf{HtmB\text{-}p2}[H, \Pi_1, \ldots, \Pi_4] = \mathsf{HtmB}[H, \Pi_1, \ldots, \Pi_4]$ based on $\Pi_1, \ldots, \Pi_4 \leftarrow\!\!\$\, \mathcal{P}(n)$. An edge (u, v) with label g denotes the mapping $v = g(u)$. Unlabelled edges are identity mapping.

4.1 HtmB-f: Random Function based Construction

Given $\Gamma_1, \ldots, \Gamma_4 \leftarrow\!\!\$\, \mathcal{F}(n,n)$, we obtain the hash-then-PRF instance where the PRF is instantiated with $\mathsf{mBenes}[\Gamma_1, \ldots, \Gamma_4]$ (truncated to first n-bit). Formally, we define $\mathsf{HtmB\text{-}f}[H, \Gamma_1, \ldots, \Gamma_4]$ (see Fig. 4.1) as $\mathsf{HtmB}[H, \Gamma_1, \ldots, \Gamma_4]$.

Recall that $\mathrm{ACP}(q)$ denotes the maximum probability of getting an alternating cycle in the dependency graph $\mathcal{G}$, where the probability is maximized over all choices of message tuple M^q. Lemma 4.1 gives a bound on $\mathrm{ACP}(q)$.

Lemma 4.1. *For $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq 2^{n-1}$, $(q, \sigma, \epsilon_2, \epsilon_1)$-DbACU$_q$ hash function H_K instantiated with $K \leftarrow\!\!\$\, \mathcal{K}$, and $\Gamma_1, \Gamma_2 \leftarrow\!\!\$\, \mathcal{F}(n, n)$, we have*

$$\mathrm{ACP}(q) \leq \frac{4q^2}{2^{2n}} + \frac{2q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Proof. Fix a q-tuple $M^q \in (\mathcal{M})_q$ that maximizes $\mathrm{ACP}(q)$. Recall that $(L^q, R^q) = H_K(M^q)$, $X_q = \Gamma_1(L^q) \oplus R^q$ and $Y^q = \Gamma_2(R^q) \oplus L^q$. We bound the probability of $\mathtt{AC}[M^q; K, \Gamma_{1,2}]$ conditioned on the following events:

- $\mathtt{Fresh}: \ \forall\, i, j \in [q],\ (L_i, R_i) \neq (L_j, R_j)$.
- $\mathtt{Lpairs}: \ |\{(i,j) \ : \ 1 \leq i < j \leq q, L_i = L_j\}| < q$.
- $\mathtt{Rpairs}: \ |\{(i,j) \ : \ 1 \leq i < j \leq q, R_i = R_j\}| < q$.

Let $\mathtt{Triv} = \neg(\mathtt{Fresh} \cap \mathtt{Lpairs} \cap \mathtt{Rpairs})$.

First, consider the probability of getting an alternating cycle of length 2 (parallel edge). Suppose the alternating cycle is $X_{i_1} = X_{i_2},\ Y_{i_1} = Y_{i_2}$, which can be rewritten as

$$\begin{aligned}\Gamma_1(L_{i_1}) \oplus R_{i_1} &= \Gamma_1(L_{i_2}) \oplus R_{i_2} \\ \Gamma_2(R_{i_1}) \oplus L_{i_1} &= \Gamma_2(R_{i_2}) \oplus L_{i_2}.\end{aligned}$$

Suppose $L_{i_1} = L_{i_2}$. Then, since $\mathtt{Fresh}$ holds, $R_{i_1} \neq R_{i_2}$, whence the first equation is not satisfied. Therefore, $L_{i_1} \neq L_{i_2}$. A similar argument implies $R_{i_1} \neq R_{i_2}$. Then, the system of equations must have full rank, i.e. rank 2. Using the randomness of Γ_1 and Γ_2, we get $q^2/2^{2n}$.

For even $k > 2$, let $X_{i_1} = X_{i_2},\ Y_{i_2} = Y_{i_3},\ \cdots,\ Y_{i_k} = Y_{i_1}$ be an alternating cycle of length k. Then, we can rewrite it as

$$\begin{aligned}\Gamma_1(L_{i_1}) \oplus R_{i_1} &= \Gamma_1(L_{i_2}) \oplus R_{i_2} \\ \Gamma_2(R_{i_2}) \oplus L_{i_2} &= \Gamma_2(R_{i_3}) \oplus L_{i_3} \\ &\vdots \\ \Gamma_2(R_{i_k}) \oplus L_{i_k} &= \Gamma_2(R_{i_1}) \oplus L_{i_1}.\end{aligned}$$

Now, we must have one of the following three cases:

1. *Independent cycle:* All k equations are independent, i.e., rank is k. Then, we can bound the probability to $q^k/2^{kn}$.
2. *Strict sub-trail with first dependency:* The cycle contains an alternating sub-trail of length $k' < k$, which has first dependency. Therefore, all the equations are independent except the last equation which is a consequence of previous equations. Without loss of generality, we assume that k' is odd. Then, we must have an associated system of equations

$$\begin{aligned}\Gamma_1(L_{i_1}) \oplus R_{i_1} &= \Gamma_1(L_{i_2}) \oplus R_{i_2} \\ \Gamma_2(R_{i_2}) \oplus L_{i_2} &= \Gamma_2(R_{i_3}) \oplus L_{i_3} \\ &\vdots \\ \Gamma_1(L_{i_{k'}}) \oplus R_{i_{k'}} &= \Gamma_1(L_{i_{k'+1}}) \oplus R_{i_{k'+1}}.\end{aligned}$$

Since the last equation is a consequence of previous equations, we must have some $i_j, i_{j'} < i_{k'}$, such that $L_{i_{k'}} = L_{i_j}$ and $L_{i_{k'+1}} = L_{i_{j'}}$. Using the fact that `Lpairs` holds, we can have at most q choices for $(i_{k'}, i_j)$ and at most q choices for $(i_{k'+1}, i_{j'})$. Similarly, we can use `Rpairs` when k' is even. The remaining $k'-3$ indices can be chosen in at most $q^{k'-3}$ ways. Finally, we bound the probability to at most $q^{k'-1}/2^{(k'-1)n}$ (as exactly $k'-1$ equations are independent).

3. *Circle has first dependency:* All the equations are independent except for the last one. This case can be handled in a similar manner as case 2. In fact, we get $q^{k-2}/2^{(k-1)n}$ which is a better bound as compared to case 2.

Combining the three cases we have

$$\begin{aligned}\Pr\left[\mathtt{AC}|\neg\mathtt{Triv}\right] &\leq \sum_{i=2}^{\infty}\frac{q^i}{2^{in}} + \sum_{j=4}^{\infty}\frac{q^{j-2}}{2^{(j-1)n}} + \sum_{k=3}^{\infty}\frac{q^{k-1}}{2^{(k-1)n}} \\ &\leq \frac{1}{1-\frac{q}{2^n}} \times \left(\frac{2q^2}{2^{2n}} + \frac{q^2}{2^{3n}}\right) \\ &\leq \frac{4q^2}{2^{2n}} + \frac{2q^2}{2^{3n}}, \end{aligned} \tag{6}$$

where the last inequality follows from $q \leq 2^{n-1}$. Finally, we have

$$\begin{aligned}\Pr\left[\mathtt{AC}\right] &\leq \Pr\left[\mathtt{AC}|\neg\mathtt{Triv}\right] + \Pr\left[\neg\mathtt{Fresh}\right] + \Pr\left[\neg\mathtt{Lpairs}\right] + \Pr\left[\neg\mathtt{Rpairs}\right] \\ &\leq \frac{4q^2}{2^{2n}} + \frac{2q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.\end{aligned}$$

At the last inequality, the third term on the right hand side follows from the (q, σ, ϵ_2)-AU property of H, and the fourth term follows from the (q, σ, ϵ_1)-ACU$_q$ property of H_1 and H_2. □

Remark 4.2. The utility of universal hash layer lies in the analysis of case 2 (and 3) in the proof. Specifically, we use the (q, σ, ϵ_1)-ACU$_q$ property of H_1 and H_2 to reduce the count of pairs with same L (or R) value from q^2 to q, which in turn helps us in reducing the overall choices for the $k'+1$ indices to $k'-1$.

Remark 4.3. This pair idea is not applicable to mBenes as the distinguisher has full control over the inputs (L_i, R_i). For instance, the distinguisher can fix a single L value across all q queries, so that we have exactly $q(q-1)$ pairs.

By now, it should be clear that Lemma 4.1 resolves the main hurdle in a proof of security up to $O(2^n)$ queries for HtmB-f. Theorem 4.1 quantifies the PRF security of HtmB-f.

Theorem 4.1. *For $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq 2^{n-1}$, $\Gamma_1 \ldots, \Gamma_4 \leftarrow\!\!\$\, \mathcal{F}(n,n)$, and $(q, \sigma, \epsilon_2, \epsilon_1)$-DbACU$_q$ hash function H_K instantiated with $K \leftarrow\!\!\$\, \mathcal{K}$, the PRF advantage of any (q, ∞)-distinguisher $\mathscr{A}$ against $\mathsf{HtmB\text{-}f}[H, \Gamma_1, \ldots, \Gamma_4]$ is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB\text{-}f}[H,\Gamma_1,\ldots,\Gamma_4]}(\mathscr{A}) \leq \frac{4q^2}{2^{2n}} + \frac{2q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Proof. A proof of this theorem can be derived using similar arguments as in case of Theorem 3.1 after substituting the bound of $\mathrm{ACP}(q)$ from Lemma 4.1.

4.2 HtmB-p1: Random Permutation based Construction

In this subsection, we aim to give a random permutation based instantiation of HtmB, called HtmB-p1. The obvious inspiration behind this is the wide availability of block ciphers which can be used to instantiate HtmB-p1.

A trivial way to achieve this is to replace the random functions with sum of independent random permutations. But this will cost 8 random permutation calls (2 calls for each f_i, $i \in [4]$). Instead, we observe that f_1 and f_2 can each be instantiated with single random permutation without any appreciable drop in security. This reduces the number of random permutation calls to 6.
Given $\Pi_1, \ldots, \Pi_6 \leftarrow_{\$} \mathcal{P}(n)$, we define the mappings, $\mathsf{F}, \mathsf{G} \in \mathcal{F}(n, n)$ as

$$\mathsf{F}(X) = \Pi_3(X) \oplus \Pi_4(X) \text{ and } \mathsf{G}(Y) = \Pi_5(Y) \oplus \Pi_6(Y),$$

and $\mathsf{HtmB\text{-}p1}[H, \Pi_1, \ldots, \Pi_6]$ (see Fig. 4.1) is defined as $\mathsf{HtmB}[H, \Pi_1, \Pi_2, \mathsf{F}, \mathsf{G}]$. Theorem 4.2 gives the PRF security of HtmB-p1.

Theorem 4.2. *For $n \geq 4$, $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq 2^{n-4}$, $\Pi_1, \ldots, \Pi_6 \leftarrow_{\$} \mathcal{P}(n)$, and $(q, \sigma, \epsilon_2, \epsilon_1)$-DbACU$_q$ hash function H_K instantiated with key $K \leftarrow_{\$} \mathcal{K}$, the PRF advantage of any (q, ∞)-distinguisher $\mathscr{A}$ against $\mathsf{HtmB\text{-}p1}[H, \Pi_1, \ldots, \Pi_6]$ is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB\text{-}p1}[H,\Pi_1,\ldots,\Pi_6]}(\mathscr{A}) \leq \frac{2q^{1.5}}{2^{1.5n}} + \frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Proof. Using hybrid argument, we replace F and G functions in the lower layer with independent random functions $\Gamma_3, \Gamma_4 \leftarrow_{\$} \mathcal{F}(n, n)$. This incurs a cost of $2q^{1.5}/2^{1.5n}$ (using Proposition 2.1). We denote the resulting construction by $\mathsf{HtmB}^{\star}$. Then we must have a (q, ∞)-distinguisher $\mathscr{B}$ against $\mathsf{HtmB}^{\star}$, such that

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB\text{-}p1}[H,\Pi_1,\ldots,\Pi_6]}(\mathscr{A}) \leq \mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB}^{\star}}(\mathscr{B}) + \frac{2q^{1.5}}{2^{1.5n}}. \tag{7}$$

Now, using a similar line of argument as used in Theorem 3.1, one can show that

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB}^{\star}}(\mathscr{B}) \leq \mathrm{ACP}(q). \tag{8}$$

Lemma 4.2 bounds $\mathrm{ACP}(q)$ to $\frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1$, which in combination with Eq. (7) and (8) gives the result. □

Lemma 4.2. *For $q \leq 2^{n-2}$, $K \leftarrow_{\$} \mathcal{K}$, and $\Pi_1, \Pi_2 \leftarrow_{\$} \mathcal{P}(n)$, we have*

$$\mathrm{ACP}(q) \leq \frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Proof. The proof idea is similar to the proof of Lemma 4.1 given in the previous subsection. So, we reuse the same set of notations and definitions.

Fix a q-tuple $M^q \in (\mathcal{M})_q$ that maximizes $\mathrm{ACP}(q)$. We bound the probability of $\mathtt{AC}[M^q; K, \Pi_{1,2}]$ conditioned on the following events:

– $\texttt{Fresh}: \ \forall\, i,j \in [q],\ (L_i, R_i) \neq (L_j, R_j)$.
– $\texttt{Lpairs}: \ |\{(i,j) \ : \ 1 \leq i < j \leq q, L_i = L_j\}| < q$.
– $\texttt{Rpairs}: \ |\{(i,j) \ : \ 1 \leq i < j \leq q, R_i = R_j\}| < q$.

The proof follows in exactly the same manner, except a minor change in the probability bound, due to a distributional change in the underlying randomness (random function to random permutation). It is easy to see that a system of k independent equations holds with probability less than $1/(2^n - k)^k$, when Π_1 and Π_2 are random permutations. We further simplify it to $2^k/2^{kn}$ using $k < q < 2^{n-1}$.

Using the above mentioned probability bound, along with the argumentation used in the proof of Lemma 4.1, we get

$$\begin{aligned}\Pr\left[\texttt{AC}|\texttt{Fresh} \cap \texttt{Lpairs} \cap \texttt{Rpairs}\right] &\leq \sum_{i=2}^{\infty} \frac{2^i q^i}{2^{in}} + \sum_{j=4}^{\infty} \frac{2^{j-1} q^{j-2}}{2^{(j-1)n}} + \sum_{k=3}^{\infty} \frac{2^{k-1} q^{k-1}}{2^{(k-1)n}} \\ &\leq \frac{1}{1 - \frac{q}{2^{n-1}}} \times \left(\frac{8q^2}{2^{2n}} + \frac{8q^2}{2^{3n}}\right) \\ &\leq \frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}}, \end{aligned} \tag{9}$$

where the last inequality follows from $q \leq 2^{n-2}$. Finally, we have

$$\mathrm{ACP}(q) \leq \frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

□

4.3 HtmB-p2: An Improvement over HtmB-p1

One can further reduce the number of permutation calls in HtmB-p1, if the generalized version of Mirror Theory [9,33,34] is correct. Specifically, we simply replace F and G in the definition of HtmB-p1 with the permutations Π_3 and Π_4 to get HtmB-p2. Formally, given $\Pi_1, \ldots, \Pi_4$ we define $\mathsf{HtmB\text{-}p2}[H, \Pi_1, \ldots, \Pi_4]$ (see Fig. 4.1) as $\mathsf{HtmB}[H, \Pi_1, \ldots, \Pi_4]$.

For any $M^q \in (\mathcal{M})_q$, $K \in \mathcal{K}$, and $\pi_1, \pi_2 \in \mathcal{P}(n)$, X^q, Y^q, and S^q are well-defined. In addition to $\texttt{AC}[M^q; K, \pi_{1,2}]$, we define two more properties on $\mathcal{G}[M^q; K, \pi_{1,2}]$:

– $\texttt{LC}[M^q; K, \pi_{1,2}](\xi)$: The largest component in $\mathcal{G}[M^q; K, \pi_{1,2}]$ contains at least $\xi + 1$ vertices.
– $\texttt{DG}[M^q; K, \pi_{1,2}]$: $\mathcal{G}[M^q; K, \pi_{1,2}]$ contains an alternating trail v^{k+1}, k odd, such that $\bigoplus_{j=1}^{k+1} S_{v_j} = 0$.

Patarin's Mirror Theory: Mirror theory [9,33,34] is a tool to obtain lower bound on the number of solutions of a system of equalities and non-equalities in finite groups. We restrict ourselves to the binary field $\mathrm{GF}(2^n)$ with $\oplus$ as the group operation. We use Mennink and Neves interpretation [12,35,19] of mirror theory, tailored to our needs and notational setup.

From X^q and Y^q, we define the mappings $\phi, \psi \in \mathcal{F}([q],[q])$ as $\phi(i) = \min\{j : X_j = X_i\}$ and $\psi(i) = \min\{k : Y_k = Y_i\}$. Let $\phi([q])$ and $\psi([q])$ denote the range of ϕ and ψ, respectively. Consider the set of equations $\mathcal{L} := \{U_{\phi(i)} \oplus V_{\psi(i)} = S_i : i \in [q]\}$, where U_j and V_k denote the unknowns for all $j \in \phi([q])$ and $k \in \psi([q])$. We define three properties on $\mathcal{L}$:

- *Circle-free*: $\mathcal{L}$ is called circle-free if $\mathtt{AC}[M^q; K, \pi_{1,2}]$ is false.
- *Non-degenerate*: $\mathcal{L}$ is called non-degenerate if $\mathtt{DG}[M^q; K, \pi_{1,2}]$ is false.
- *ξ-block-maximal:* $\mathcal{L}$ is called ξ-block-maximal if $\mathtt{LC}[M^q; K, \pi_{1,2}](\xi)$ is false.

Whenever $\mathcal{L}$ is circle-free, non-degenerate, and ξ-block-maximal, then we say that $\mathcal{L}$ is *mirror theory compatible till* ξ. The fundamental result of mirror theory [9, Theorem 6] is given in Theorem 4.3.

Theorem 4.3 (Theorem 3 in [12]). *Suppose $\mathcal{L}$, as defined above, is mirror theory compatible till ξ. Then, as long as $\xi^2 \cdot \max\{|\phi([q])|, |\psi([q])|\} \le 2^n/67$, the number of solutions for $\mathcal{L}$, such that $U_i \neq U_j$ for distinct $i, j \in \phi([q])$ and $V_k \neq V_\ell$ for distinct $k, \ell \in \psi([q])$, is at least*

$$\frac{(2^n)_{|\phi([q])|}(2^n)_{|\psi([q])|}}{2^{nq}}.$$

In [9], Patarin gave a very high level sketch of the proof. Later, in [34] Nachef, Patarin and Volte gave a proof that works till $q < 2^{n-3}$. In [12], Mennink and Neves gave a detailed exposition on mirror theory, and utilized the theory to get close to n-bit security bounds for EDM (and EWCDM [11], in nonce-respecting[5] setting). Jha and Nandi [19] developed a variant of mirror theory to derive tight security bounds for CLRW2 [36] and DbHtS. Independently, Kim et al. [20] used the theory to derive tight security bounds for several DbHtS MACs, including PMAC+ and LightMAC+. We use Theorem 4.3 in the security proof of HtmB-p2.

Theorem 4.4. *For $\epsilon_1, \epsilon_2, \sigma \ge 0$, $q \le \min\{2^{n-2}, 2^n/67n^2\}$, $\Pi_1, \ldots, \Pi_4 \leftarrow\!\!\$\, \mathcal{P}(n)$, and $(q, \sigma, \epsilon_2, \epsilon_1)$-DbACU$_q$ hash function H instantiated with key $K \leftarrow\!\!\$\, \mathcal{K}$, the PRF advantage of any (q, ∞)-distinguisher $\mathscr{A}$ against* HtmB-p2$[H, \Pi_1, \ldots, \Pi_4]$ *is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB\text{-}p2}[H,\Pi_1\ldots,\Pi_4]}(\mathscr{A}) \le \frac{16q^2}{2^{3n}} + \frac{36q^2}{2^{2n}} + \frac{4q}{2^n} + \epsilon_2 + 2\epsilon_1.$$

Proof approach: The idea is quite similar to the proof of HtmB-f. However, just avoiding AC in $\mathcal{G}$ is not enough. This is due to the switch from random functions to random permutations. For example, the system of equations $\Pi_3(X^q) \oplus \Pi_4(Y^q) = S^q$ should be non-degenerate. Otherwise, we might get a case where $\Pi_3(X_i) = \Pi_3(X_j)$ for $X_i \neq X_j$, which is clearly not possible. We show that the system is mirror theory compatible till n, except with very negligible probability as long as $q \le 2^{n-2}$. Then, we apply the fundamental result of mirror theory to get the proof of security using coefficient-H technique.

[5] Each query requires a distinct nonce input.

Proof. $\mathscr{A}$ tries to distinguish the real oracle $\mathcal{R} := (\mathsf{HtmB\text{-}p2}[H, \Pi_1, \ldots, \Pi_4])$ from the ideal oracle $\mathcal{I} := (\Gamma')$ for $\Gamma' \leftarrow_{\$} \mathcal{F}(\mathcal{M}, \{0,1\}^n)$. Let $[q]$ denote the set of all query indices, and (M^q, S^q) denote $\mathscr{A}$'s transcript, where M^q is the q-tuple of inputs and S^q is the q-tuple of outputs.

Consider a variant distinguishing game, where the oracle releases L^q, R^q, X^q, and Y^q, once the distinguisher has made all q queries. Note that this can only increase $\mathscr{A}$'s advantage, and not diminish it. In $\mathcal{R}$, this is quite straightforward, as L^q, R^q, X^q, and Y^q, are already computed during the query phase. The ideal oracle $\mathcal{I}$, samples dummy $K \leftarrow_{\$} \mathcal{K}$ and $\Pi_1, \Pi_2 \leftarrow_{\$} \mathcal{P}(n)$, and sets $(L^q, R^q) = H_K(M^q)$, $X^q = \Pi_1(L^q) \oplus R^q$ and $Y^q = \Pi_2(R^q) \oplus L^q$.

Bad Transcript: Let $\mathcal{T}$ denote the set of all transcripts. Let $\mathtt{Bad}$ denote the event that the system of equations $\mathcal{L} := \{U_{\phi(i)} \oplus V_{\psi(i)} = S_i : i \in [q]\}$ is not mirror theory compatible till n, and good otherwise. So $\mathtt{Bad}$ holds if at least one of $\mathtt{AC}$, $\mathtt{LC}(n)$, or $\mathtt{DG}$ is satisfied. We say that a transcript $(M^q, L^q, R^q, X^q, Y^q, S^q)$ is bad if $\mathtt{Bad}$ happens, and good otherwise. Let $\mathcal{T}_{\mathsf{bad}} \subset \mathcal{T}$ denote the set of all bad transcripts. Then, we have $\Pr[\mathbb{T}_{\mathsf{id}} \in \mathcal{T}_{\mathsf{bad}}] = \Pr[\mathtt{Bad}]$.
We bound the probability of $\mathtt{Bad}$ conditioned on the following events:

- $\mathtt{Fresh}$: $\forall\, i, j \in [q],\ (L_i, R_i) \neq (L_j, R_j)$.
- $\mathtt{Lpairs}$: $|\{(i,j) \ : \ 1 \leq i < j \leq q, L_i = L_j\}| < q$.
- $\mathtt{Rpairs}$: $|\{(i,j) \ : \ 1 \leq i < j \leq q, R_i = R_j\}| < q$.

Let $\mathtt{Triv} = \neg(\mathtt{Fresh} \cap \mathtt{Lpairs} \cap \mathtt{Rpairs})$. Then, we have

$$\begin{aligned}
\Pr[\mathtt{Bad}] &\leq \Pr[\mathtt{Bad}|\neg\mathtt{Triv}] + \Pr[\mathtt{Triv}] \\
&\overset{(*)}{\leq} \Pr[\mathtt{LC}(n)|\neg\mathtt{Triv}] + \Pr[\mathtt{DG}|\neg\mathtt{Triv}] + \Pr[\mathtt{AC}|\neg\mathtt{Triv}] + \Pr[\mathtt{Triv}] \\
&\overset{(**)}{\leq} \Pr[\mathtt{LC}(n)|\neg\mathtt{Triv}] + \Pr[\mathtt{DG}|\neg\mathtt{Triv}] + \frac{16q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1, \quad (10)
\end{aligned}$$

where inequality $(*)$ follows from the definition of $\mathtt{Bad}$, and inequality $(**)$ follows from Lemma 4.2. Lemma 4.3 bounds the probability of $\mathtt{LC}(n)$ and Lemma 4.4 bounds the probability of $\mathtt{DG}$ conditioned on $\neg\mathtt{Triv}$.

Good Transcript: Fix a good transcript $(M^q, L^q, R^q, X^q, Y^q, S^q)$. Since the ideal oracle faithfully (identical to the real oracle) simulates the computation of L^q, R^q, X^q, and Y^q, it is sufficient to concentrate on the ratio of the probabilities that (X^q, Y^q) maps to S^q in the real oracle and M^q maps to S^q in the ideal oracle.

$$\begin{aligned}
\frac{\Pr[\mathbb{T}_{\mathsf{re}} = (M^q, L^q, R^q, X^q, Y^q, S^q)]}{\Pr[\mathbb{T}_{\mathsf{id}} = (M^q, L^q, R^q, X^q, Y^q, S^q)]} &= \frac{\Pr[\Pi_3(X^q) \oplus \Pi_4(Y^q) = S^q]}{\Pr[\Gamma'(M^q) = S^q]} \\
&\overset{(*)}{=} 2^{nq} \times \frac{h_q}{(2^n)_{|\phi([q])|}(2^n)_{|\psi([q])|}} \\
&\overset{(**)}{\geq} 1. \quad (11)
\end{aligned}$$

where h_q denotes the number of solutions of the system of equations $\Pi_3(X^q) \oplus \Pi_4(Y^q) = S^q$, such that $\Pi_3(X_i) \neq \Pi_3(X_j)$ and $\Pi_4(Y_k) \neq \Pi_4(Y_\ell)$ for all $X_i \neq X_j$

and $Y_k \neq Y_\ell$. Further, each solution holds with exactly $1/(2^n)_{|\phi([q])|}(2^n)_{|\psi([q])|}$ probability, since Π_3 and Π_4 are invoked on exactly $|\phi([q])|$ and $|\psi([q])|$, respectively, distinct points. This justifies equality $(*)$. Let $U_{\phi(i)} = \Pi_3(X_i)$ and $V_{\psi(i)} = \Pi_4(Y_i)$ for all $i \in [q]$. Since the transcript is good, $\mathcal{L} := \{U_{\phi(i)} \oplus V_{\psi(i)} = S_i : i \in [q]\}$ is mirror theory compatible till n. Hence, using Theorem 4.3, we have

$$h_q \geq \frac{(2^n)_{|\phi([q])|}(2^n)_{|\psi([q])|}}{2^{nq}}. \tag{12}$$

This justifies the inequality $(**)$. The result follows from Eq. (10), Lemmata 4.3 and 4.4, and Theorem 2.1. □

Remark 4.4. In Eq. (11) we have substituted h_q with the lower bound claimed in the fundamental result of mirror theory (see Theorem 4.3). However, as reported in multiple works [10,37,35,19], a concrete proof of this result is still not available. Here, we discuss the impact of a weaker mirror theory result on Theorem 4.4. Suppose, in future we get a mirror theory proof that holds for some $\xi < n$ and the lower bound is

$$(1-\delta) \times \frac{(2^n)_{|\phi([q])|}(2^n)_{|\psi([q])|}}{2^{nq}},$$

for some $\delta > 0$. Here δ can be viewed as the degree of deviation from the perfect bound. Then, the bound in Theorem 4.4 is revised asymptotically to

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{HtmB\text{-}p2}[H,\Pi_1\ldots,\Pi_4]}(\mathscr{A}) = O\left(\frac{q^2}{2^{2n}}\right) + O\left(\frac{q^{\xi+1}}{2^{n\xi}}\right) + \delta + \epsilon_2 + 2\epsilon_1,$$

where the red colored terms are due to the degradation in mirror theory bound. Specifically, $O(q^{\xi+1}/2^{n\xi})$ arises in the bound of $\mathtt{LC}(\xi)|\neg\mathtt{Triv}$, and δ appears on the right hand side of Eq. (11) by substituting the weaker bound for h_q.

Lemma 4.3. *For $q \leq 2^{n-2}$, $K \leftarrow\!\!\$\, \mathcal{K}$, and $\Pi_1, \Pi_2 \leftarrow\!\!\$\, \mathcal{P}(n)$, we have*

$$\Pr\left[\mathtt{LC}(n)|\neg\mathtt{Triv}\right] \leq \frac{8q^2}{2^{2n}} + \frac{4q}{2^n}.$$

Lemma 4.4. *For $q \leq 2^{n-2}$, $K \leftarrow\!\!\$\, \mathcal{K}$, $\Pi_1, \Pi_2 \leftarrow\!\!\$\, \mathcal{P}(n)$, and $\Gamma' \leftarrow\!\!\$\, \mathcal{F}(\mathcal{M}, \{0,1\}^n)$, we have*

$$\Pr\left[\mathtt{DG}|\neg\mathtt{Triv}\right] \leq \frac{12q^2}{2^{2n}}.$$

Given the similarity of the proofs of Lemmata 4.3 and 4.4 with the proof of Lemma 4.2, they are deferred to the full version of this paper.

5 mLightMAC+ and mPMAC+

In this section, we define two families mLightMAC+ and mPMAC+ of deterministic MAC candidates based on block ciphers. Both families are constructed as the HtmB construction, where the DbACU_q hash functions (see section 2) are instantiated with the LightHash and PHash hash functions. In particular, our schemes have the following properties:

– they are secure VIL PRFs as long as the number of queried blocks are small in front of 2^n, where n denotes the block size;
– the calls to the underlying permutation can be computed in parallel.

5.1 mLightMAC+

In this section, we define the mLightMAC+ construction and prove its security. We are going to proceed in two steps: first, we define the LightHash family of permutation-based hash functions and upper bound the probability to get colliding outputs in Lemma 5.1, and then we use Theorems 4.1-4.4 to prove the actual security bound on mLightMAC+ in Corollary 5.1.

THE LightHash UNIVERSAL HASH FUNCTION: Given a permutation $\pi \in \mathcal{P}(n)$ and a positive integer $m \in [n-2]$, the LightHash universal hash function is a function from $\{0,1\}^{\leq (n-m)2^m-1}$ to $\{0,1\}^{2n}$ defined as follows. For all messages $M \in \{0,1\}^{\leq (n-m)2^m-1}$, we let $M' = \mathsf{pad}(M)$, $l = |M'|/(n-m)$ and $M' = M^1||\cdots||M^l$, where $|M^i| = n-m$ for all $i \in [l]$. The hash of the message M is defined as $\mathsf{LightHash}[\pi, m](M) = (\mathsf{LightHash}_1[\pi, m](M), \mathsf{LightHash}_2[\pi, m](M))$, where

$$\mathsf{LightHash}_1[\pi, m](M) = (\langle l \rangle_m || M^l) \oplus \bigoplus_{i=1}^{l-1} \pi\left(\langle i \rangle_m || M^i\right),$$

$$\mathsf{LightHash}_2[\pi, m](M) = (\langle l \rangle_m || M^l) \oplus \bigoplus_{i=1}^{l-1} 2^{l-i}\pi\left(< i >_m || M^i\right).$$

Note that LightHash requires 1 less block cipher call as compared to the hash layer in LightMAC+. The probability that two distinct messages generate colliding outputs in both components of LightHash can be upper bounded as follows.

Lemma 5.1. *Let $n \in \mathbb{N}$, $m \in [n-2]$. For any two distinct messages M_1, M_2 in $\{0,1\}^{\leq (n-m)2^m-1}$ and $\Pi \leftarrow\!\!\$\ \mathcal{P}(n)$, one has*

$$\Pr\left[\mathsf{LightHash}[\Pi, m](M_1) = \mathsf{LightHash}[\Pi, m](M_2)\right] \leq \frac{4}{2^{2n}},$$

$$\Pr\left[\mathsf{LightHash}_b[\Pi, m](M_1) = \mathsf{LightHash}_b[\Pi, m](M_2)\right] \leq \frac{2}{2^n},$$

for $b \in \{0,1\}$. In particular LightHash *is $(q, \infty, \frac{2q^2}{2^{2n}}, \frac{q}{2^n})$-DbACU$_q$.*

The proof of this Lemma can be found in section 6.1.

THE mLightMAC+ FAMILY OF PRFS: Given $\pi_0, \ldots, \pi_6 \in \mathcal{P}(n)$, $f_1, \ldots, f_4 \in \mathcal{F}(n,n)$ and an integer $m \in [n-2]$, the functions of the mLightMAC+ family are functions from $\{0,1\}^{\leq (n-m)2^m-1}$ to $\{0,1\}^n$ that are formally defined as

$$\mathsf{mLightMAC+\text{-}f}[\pi_0, f_1 \ldots, f_4, m] := \mathsf{HtmB\text{-}f}\left[\mathsf{LightHash}[\pi_0, m], f_1, \ldots, f_4\right],$$
$$\mathsf{mLightMAC+\text{-}p1}[\pi_0, \pi_1 \ldots, \pi_6, m] := \mathsf{HtmB\text{-}p1}\left[\mathsf{LightHash}[\pi_0, m], \pi_1, \ldots, \pi_6\right],$$
$$\mathsf{mLightMAC+\text{-}p2}[\pi_0, \pi_1 \ldots, \pi_4, m] := \mathsf{HtmB\text{-}p2}\left[\mathsf{LightHash}[\pi_0, m], \pi_1, \ldots, \pi_4\right].$$

Corollary 5.1 gives the PRF security of mLightMAC+.

Corollary 5.1. *For $q < 2^{n-4}$, $m \leq n-2$, and $\Pi_0, \ldots, \Pi_6 \leftarrow_\$ \mathcal{P}(n)$, $\Gamma_1, \ldots, \Gamma_4 \leftarrow_\$ \mathcal{F}(n,n)$, the PRF advantage of any (q,∞)-distinguisher $\mathscr{A}$ against* mLightMAC+ *is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mLightMAC+\text{-}f}[\Pi_0,\Gamma_1,\ldots,\Gamma_4,m]}(\mathscr{A}) \leq \frac{6q^2}{2^{2n}} + \frac{2q^2}{2^{3n}} + \frac{2q}{2^n},$$
$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mLightMAC+\text{-}p1}[\Pi_0,\ldots,\Pi_6,m]}(\mathscr{A}) \leq \frac{2q^{1.5}}{2^{1.5n}} + \frac{18q^2}{2^{2n}} + \frac{16q^2}{2^{3n}} + \frac{2q}{2^n},$$
$$\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mLightMAC+\text{-}p2}[\Pi_0,\ldots,\Pi_4,m]}(\mathscr{A}) \leq \frac{16q^2}{2^{3n}} + \frac{38q^2}{2^{2n}} + \frac{6q}{2^n}.$$

For the second and third inequalities, we also assume $n \geq 4$ and $q \leq 2^n/67n^2$, respectively.

Proof. This result is a direct combination of Lemma 5.1 and Theorems 4.1, 4.2 and 4.4. □

5.2 mPMAC+

As in the previous section, we define the mPMAC+ construction and prove its security. We first define the PHash family of permutation-based hash functions and upper bound the probability to get colliding outputs in Lemma 5.2, and then we use Theorems 4.1-4.4 to prove the actual security bound on mPMAC+ in Corollary 5.2.

The PHash Universal Hash Function: Given a permutation $\pi \in \mathcal{P}(n)$, the PHash universal hash function is a function from $\{0,1\}^{\leq n2^{n/2}-1}$ to $\{0,1\}^{2n}$ defined as follows. For all messages $M \in \{0,1\}^{\leq n2^{n/2}-1}$, we let $M' = \mathsf{pad}(M)$, $l = |M'|/n$ and $M' = M^1||\cdots||M^l$, where $|M^i| = n$ for all $i \in [l]$. The hash of the message M is then defined as $\mathsf{PHash}[\pi](M) = (\mathsf{PHash}_1[\pi](M), \mathsf{PHash}_2[\pi](M))$, where

$$\mathsf{PHash}_1[\pi](M) = M^l \oplus \bigoplus_{i=1}^{l-1} \pi\left(M^i \oplus 2^i\pi(0^n) \oplus 2^{2i}\pi(10^{n-1})\right),$$
$$\mathsf{PHash}_2[\pi](M) = M^l \oplus \bigoplus_{i=1}^{l} 2^{l-i}\pi\left(M^i \oplus 2^i\pi(0^n) \oplus 2^{2i}\pi(10^{n-1})\right).$$

Again note that PHash requires 1 less block cipher call as compared to the hash layer in PMAC+. One has the following result on the DbACU$_q$ bound of PHash.

Lemma 5.2. *Let $n \geq 6$. For $\Pi \leftarrow_\$ \mathcal{P}(n)$, $\sigma \in \mathbb{N}$, $\mathsf{PHash}[\Pi]$ is $(q,\sigma,\epsilon_2,\epsilon_1)$-DbACU$_q$ where*

$$\epsilon_2 \leq \frac{2\sigma^2 + 28q\sigma + 28q^2}{2^{2n}} + \frac{3q}{2^n - 2} + 3\frac{\sigma + q}{2^n - 1} \qquad \text{and} \qquad \epsilon_1 \leq \frac{4\sigma + 9q}{2^n}.$$

The proof of this Lemma can be found in section 6.2.

THE mPMAC+ FAMILY OF PRFS: Given $\pi_0, \ldots, \pi_6 \in \mathcal{P}(n)$ and $f_1, \ldots, f_4 \in \mathcal{F}(n,n)$, the functions of the mPMAC+ family are functions from $\{0,1\}^{n2^{n/2}-1}$ to $\{0,1\}^n$ that are formally defined as

$$\begin{aligned}
\mathsf{mPMAC+\text{-}f}[\pi_0, f_1, \ldots, f_4] &:= \mathsf{HtmB\text{-}f}\left[\mathsf{PHash}[\pi_0], f_1, \ldots, f_4\right], \\
\mathsf{mPMAC+\text{-}p1}[\pi_0, \pi_1, \ldots, \pi_6] &:= \mathsf{HtmB\text{-}p1}\left[\mathsf{PHash}[\pi_0], \pi_1, \ldots, \pi_6\right], \\
\mathsf{mPMAC+\text{-}p2}[\pi_0, \pi_1, \ldots, \pi_4] &:= \mathsf{HtmB\text{-}p2}\left[\mathsf{PHash}[\pi_0], \pi_1, \ldots, \pi_4\right]
\end{aligned}$$

Corollary 5.2 gives the PRF security of mPMAC+.

Corollary 5.2. *Let $n \geq 6$. For $q < 2^{n-4}$ and $\Pi_0, \ldots, \Pi_6 \leftarrow\!\!\$\, \mathcal{P}(n)$, and $\Gamma_1, \ldots, \Gamma_4 \leftarrow\!\!\$\, \mathcal{F}(n,n)$, the PRF advantage of any (q,∞)-distinguisher $\mathscr{A}$ against* mPMAC+ *is given by*

$$\begin{aligned}
\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mPMAC+\text{-}f}[\Pi_0,\Gamma_1,\ldots,\Gamma_4]}(\mathscr{A}) &\leq \frac{2q^2}{2^{3n}} + \frac{2\sigma^2 + 28q\sigma + 32q^2}{2^{2n}} + \frac{11\sigma + 15q}{2^n - 2}, \\
\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mPMAC+\text{-}p1}[\Pi_0,\ldots,\Pi_6]}(\mathscr{A}) &\leq \frac{2q^{1.5}}{2^{1.5n}} + \frac{16q^2}{2^{3n}} + \frac{2\sigma^2 + 28q\sigma + 44q^2}{2^{2n}} + \frac{11\sigma + 15q}{2^n - 2}, \\
\mathbf{Adv}^{\mathsf{prf}}_{\mathsf{mPMAC+\text{-}p2}[\Pi_0,\ldots,\Pi_4]}(\mathscr{A}) &\leq \frac{16q^2}{2^{3n}} + \frac{2\sigma^2 + 28q\sigma + 64q^2}{2^{2n}} + \frac{11\sigma + 19q}{2^n - 2},
\end{aligned}$$

where σ denotes an upper bound on the total number of n-bit blocks queried by $\mathscr{A}$. For the last inequality, we also assume $q \leq 2^n/67n^2$.

Proof. This result is a direct combination of Lemma 5.2 and Theorem 4.1, 4.2 and 4.4. □

6 Proofs Related to LightHash and PHash

6.1 Proof of Lemma 5.1

Let $q \in \mathbb{N}$, $m \in [n-2]$, $M^q \in \left(\{0,1\}^{(n-m)2^m-1}\right)_q$.

Let us now fix two distinct integers $i_1, i_2 \in [q]$, and let $M_1 = M_{i_1}$, $M_2 = M_{i_2}$.

The proof for the first inequality closely follows the proof of [17, Lemma 1] for the original LightHash construction, with slight changes to handle our variant. It is thus deferred to the full version of this paper for reasons of space.

We now consider the second inequality we have to prove, and denote by l_1 (resp. l_2) the length of $\mathsf{pad}(M_1)$ (resp. $\mathsf{pad}(M_2)$) in $(n-m)$-bit blocks. Note that $1 \leq l_1, l_2 \leq 2^m \leq 2^{n-2}$. Then the event

$$\mathsf{LightHash}_1[\Pi, m](M_1) = \mathsf{LightHash}_1[\Pi, m](M_2)$$

is equivalent to:

$$\left(\langle l_1 \rangle_m || M_1^{l_1}\right) \oplus \bigoplus_{i=1}^{l_1-1} \Pi\left(\langle i \rangle_m || M_1^i\right) = \left(\langle l_2 \rangle_m || M_2^{l_2}\right) \oplus \bigoplus_{i=1}^{l_2-1} \Pi\left(\langle i \rangle_m || M_2^i\right). \quad (13)$$

We consider two different cases: $l_1 \neq l_2$ and $l_1 = l_2$. Consider the first case. Let us assume that $1 \leq l_1 < l_2$. Thus, thanks to domain separation of the inputs and since at most $l_1 + l_2 \leq 2^{n-1}$ outputs appear in Eq. (13), fixing all the other outputs will provide a unique solution for $\Pi(\langle l_2 \rangle_m || M_2^{l_2})$. Hence, the probability that (13) is satisfied is at most $1/(2^n - l_1 - l_2 + 3)$. Now consider the second case. Since the adversary cannot repeat queries and our padding is injective, $\mathsf{pad}(M_1)$ and $\mathsf{pad}(M_2)$ must differ in at least one block. Let $i_0 \geq 1$ be the first such index. Then, even when eliminating the colliding outputs from Eq. (13), at least the outputs with index i_0 will remain. If $i_0 \leq l_1 - 1$, fixing all the other outputs will provide a unique solution for $\Pi(\langle i_0 \rangle_m || M_1^{i_0})$, and the probability that Eq. (13) is satisfied is also at most $1/(2^n - l_1 - l_2 + 3)$. Otherwise, if $i_0 = l_1$, Eq. (13) is reduced to $M_1^{l_1} = M_2^{l_2}$, which cannot hold by definition of i_0.

Overall, since $l_1 + l_2 \leq 2^{n-1}$, one has

$$\Pr\left[\mathsf{LightHash}_1[\Pi, m](M_1) = \mathsf{LightHash}_1[\Pi, m](M_2)\right] \leq \frac{2}{2^n}.$$

Similarly, one has

$$\Pr\left[\mathsf{LightHash}_2[\Pi, m](M_1) = \mathsf{LightHash}_2[\Pi, m](M_2)\right] \leq \frac{2}{2^n}.$$

We conclude the proof of the second part of Lemma 5.1 by summing over the $q(q-1)/2$ pairs of queries and using Markov's inequality.

6.2 Proof of Lemma 5.2

A FLAW IN [26]: The probability of observing a full collision in PHash has already been considered in [26]. However, Chakraborty et al. [27] identified a flaw in the argument. In more details, when considering what is referred to as Type-5 collisions, the author tries to upper bound the probability, over the random choice of two n-bit masks L_1 and L_2, that the following system is satisfied:

$$(2^{i_1} \oplus 2^{i_2})L_1 \oplus (2^{3i_1} \oplus 2^{3i_2})L_2 = X_1$$
$$(2^{i_3} \oplus 2^{i_4})L_1 \oplus (2^{3i_3} \oplus 2^{3i_4})L_2 = X_2$$

for some n-bit values X_1, X_2 and four integers i_1, i_2, i_3, i_4 such that at least three of them are distinct. It is then argued that either the system is of rank two, and has exactly one solution, or both equations are equal. In the second case, the author shows that $2^{i_1} \oplus 2^{i_2} = 2^{i_3} \oplus 2^{i_4}$ and $2^{3i_1} \oplus 2^{3i_2} = 2^{3i_3} \oplus 2^{3i_4}$ imply $i_1 = i_2 = i_3 = i_4$ which is impossible. However, it seems that another case is possible: the second equation can be a multiple of the first one. In that case, there exists a non-zero value α such that $\alpha(2^{i_1} \oplus 2^{i_2}) = 2^{i_3} \oplus 2^{i_4}$, $\alpha(2^{3i_1} \oplus 2^{3i_2}) = 2^{3i_3} \oplus 2^{3i_4}$ and $\alpha X_1 = X_2$, and the previous impossibility argument does not apply anymore. With a more complex analysis, it may still be possible to prove a bound that is independent from the length of the queries. Another approach could be to use a different masking, as demonstrated in [38,27], that avoids the above mentioned

case. In our work, we leave this question as an interesting open problem and we use a slightly worse bound that depends on the number of queried message blocks, but is still sufficient to provide optimal security.

Proof of Lemma 5.2 Let $n \geq 6$, $q \leq 2^n$ be two integers and let us fix a q-tuple of messages $M^q \in \left(\{0,1\}^{n2^{n/2}-1}\right)_q$ whose total block length is σ. We parse $\mathsf{pad}(M_i)$ as $M_i^1||\cdots||M_i^{l_i}$, where $i \in [q]$, $|M_i^j| = n$ for every $i \in [l_i]$, and $l_i \leq 2^{n/2}$. Note that, because of our padding, $\sum_{i=1}^{q} l_i \leq \sigma + q$. We are going to introduce several new random variables that depend on the uniformly random draw of Π:

- $L_1 = \Pi(0^n)$ and $L_2 = \Pi(10^{n-1})$;
- for all $i \in [q]$ and all $j \in [l_i - 1]$, $X_i^j = M_i^j \oplus 2^j L_1 \oplus 2^{2j} L_2$ and $Y_i^j = \Pi(X_i^j)$;
- for $i \in [q]$,

$$\Sigma_i = \mathsf{PHash}_1[\Pi](M_i) = M^{l_i} \oplus \bigoplus_{j=1}^{l_i-1} Y_i^j \text{ and}$$

$$\Theta_i = \mathsf{PHash}_2[\Pi](M_i) = M^{l_i} \oplus \bigoplus_{j=1}^{l_i-1} 2^{l_i-j} Y_i^j.$$

Let us fix two distinct integers i_1, i_2 in $[q]$, and assume w.l.o.g. that $l_{i_1} \geq l_{i_2}$. The first step of our proof is to upper bound the probability to create a collision in the output of PHash_1. More precisely, we want to upper bound the probability that $\Sigma_{i_1} = \Sigma_{i_2}$.

Claim 6.1 *One has*

$$\Pr\left[\Sigma_{i_1} = \Sigma_{i_2}\right] \leq 2\frac{l_{i_1} + l_{i_2} + 4}{2^n},$$

$$\Pr\left[\Theta_{i_1} = \Theta_{i_2}\right] \leq 2\frac{l_{i_1} + l_{i_2} + 4}{2^n}.$$

The proof of this claim is deferred to the full version of this paper for reasons of space.

Let C_1 (resp. C_2) be the number of Σ (resp. Θ) collisions. Summing over every pair of queries yields

$$\mathsf{Ex}\left[C_1\right] \leq \sum_{i_1 < i_2} 2\frac{l_{i_1} + l_{i_2} + 4}{2^n} \leq \frac{4q(\sigma + q) + 4q^2}{2^n} \leq \frac{4q\sigma + 9q^2}{2^n}.$$

Similarly, one has $\mathsf{Ex}\left[C_2\right] \leq \frac{4q\sigma+9q^2}{2^n}$. Using Markov's inequality ends the first part of the proof of this lemma.

Our goal is now to upper bound the probability of the following event (dubbed `Coll` in the following): there exist two distinct indices i_1 and i_2 such that

$$\mathsf{PHash}[\Pi](M_{i_1}) = \mathsf{PHash}[\Pi](M_{i_2}).$$

We are going to break this event into several different events that will be easier to handle:

– `Coll0`: there exist $i \in [q]$ and $j \in [l_i - 1]$ such that $X_i^j = 0^n$;
– `Coll1`: there exist $i \in [q]$ and $j \in [l_i - 1]$ such that $X_i^j = 10^{n-1}$;
– `3Coll`: there exist $i \in [q]$ and three pairwise distinct integers $j_1, j_2, j_3 \in [l_i - 1]$ such that $X_i^{j_1} = X_i^{j_2} = X_i^{j_3}$;
– `CleanColl`: this event corresponds to $\texttt{Coll} \wedge \neg\texttt{Coll0} \wedge \neg\texttt{Coll1} \wedge \neg\texttt{3Coll}$.

Clearly, one has

$$\Pr[\texttt{Coll}] \leq \Pr[\texttt{Coll0}] + \Pr[\texttt{Coll1}] + \Pr[\texttt{3Coll}] + \Pr[\texttt{CleanColl}]. \tag{14}$$

It is also easy to see that

$$\Pr[\texttt{Coll0}] \leq \frac{\sigma + q}{2^n - 1} \qquad \text{and} \qquad \Pr[\texttt{Coll1}] \leq \frac{\sigma + q}{2^n - 1}. \tag{15}$$

Let us now consider the event `3Coll`. Fix any $i \in [q]$ and any pairwise distinct $j_1, j_2, j_3 \in [l_i - 1]$. The system (S) of equations $X_i^{j_1} = X_i^{j_2} = X_i^{j_3}$ can be rewritten as

$$\begin{aligned}(2^{j_1} \oplus 2^{j_2})L_1 \oplus (2^{2j_1} \oplus 2^{2j_2})L_2 &= M_i^{j_1} \oplus M_i^{j_2} \\ (2^{j_1} \oplus 2^{j_3})L_1 \oplus (2^{2j_1} \oplus 2^{2j_3})L_2 &= M_i^{j_1} \oplus M_i^{j_3}\end{aligned}$$

Since j_1, j_2, j_3 are pairwise distinct and smaller than $2^n - 1$, the values $2^{j_1}, 2^{j_2}, 2^{j_3}$ are pairwise distinct and (S) is equivalent to

$$\begin{aligned}L_1 \oplus (2^{j_1} \oplus 2^{j_2})L_2 &= (M_i^{j_1} \oplus M_i^{j_2})/(2^{j_1} \oplus 2^{j_2}) \\ L_1 \oplus (2^{j_1} \oplus 2^{j_3})L_2 &= (M_i^{j_1} \oplus M_i^{j_3})/(2^{j_1} \oplus 2^{j_3}).\end{aligned}$$

Since $2^{j_2} \neq 2^{j_3}$, the system has a unique solution, and is verified with probability at most $1/2^n(2^n - 1)$.

Summing over every possible choice of i, j_1, j_2, j_3 yields

$$\Pr[\texttt{3Coll}] \leq \sum_{i=1}^{q} \frac{l_i^3}{2^n(2^n - 1)} \overset{(*)}{\leq} \sum_{i=1}^{q} \frac{l_i}{2^n - 1} \leq \frac{\sigma + q}{2^n - 1}, \tag{16}$$

where inequality $(*)$ comes from the fact that $l_i \leq 2^{n/2}$ for every $i \in [q]$.

We now have to handle the event `CleanColl`. We make the following claim.

Claim 6.2

$$\Pr[\texttt{CleanColl}] \leq \frac{2\sigma^2 + 28q\sigma + 28q^2}{2^{2n}} + \frac{3q}{2^n - 2}.$$

The proof this claim is deferred to the full version of this paper for reasons of space.

Combining Eqs (14), (15), (16) and Claim 6.2 yields

$$\Pr[\texttt{Coll}] \leq \frac{2\sigma^2 + 28q\sigma + 28q^2}{2^{2n}} + \frac{3q}{2^n - 2} + \frac{3(\sigma + q)}{2^n - 1},$$

which ends the proof.

7 Reducing the Number of Keys

HtmB-f, HtmB-p1, and HtmB-p2 need 4, 6, and 4 keys, respectively, apart from the hash key. This could be an issue in certain memory-restricted scenarios. In this section, we present some simple variants of these constructions that require less key material, albeit with a slight loss of security.
For any function $F \in \mathcal{F}$ and $b \in \{0,1\}^{<n}$, we define two mappings:

$$\widehat{F}^b := \lfloor F(b\|\cdot)\rfloor_{n-|b|} \qquad \widetilde{F}^b(X) := F(b\|\cdot),$$

where $\lfloor Y\rfloor_{n-d}$ denotes the $(n-d)$-least significant bits of Y for all $Y \in \{0,1\}^n$ and $d < n$. In the following discussion $\mathcal{M} \subseteq \{0,1\}^*$.

Single-key variant of HtmB-f: Given $\Gamma \leftarrow_\$ \mathcal{F}(n,n)$ and a pair $H = (H_1, H_2)$ of two $(\mathcal{K}, \mathcal{M}, \{0,1\}^{n-2})$-keyed hash functions, we define the single-key variant of HtmB-f, denoted 1k-HtmB-f, as:

$$\text{1k-HtmB-f}[H,\Gamma] := \text{HtmB-f}[H, \widehat{\Gamma}^{00}, \widehat{\Gamma}^{01}, \widetilde{\Gamma}^{10}, \widetilde{\Gamma}^{11}].$$

Theorem 7.1. *For $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq 2^{n-3}$, $\Gamma \leftarrow_\$ \mathcal{F}(n,n)$, and $(q,\sigma,\epsilon_2,\epsilon_1)$-DbACU$_q$ hash function H_K instantiated with $K \leftarrow_\$ \mathcal{K}$, the PRF advantage of any (q,∞)-distinguisher $\mathscr{A}$ against* 1k-HtmB-f$[H,\Gamma]$ *is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\text{1k-HtmB-f}[H,\Gamma]}(\mathscr{A}) \leq \frac{64q^2}{2^{2n}} + \frac{128q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Three-key variant of HtmB-p1: Given $\Pi_1, \Pi_2, \Pi_3 \leftarrow_\$ \mathcal{P}(n)$ and a pair $H = (H_1, H_2)$ of two $(\mathcal{K}, \mathcal{M}, \{0,1\}^{n-1})$-keyed hash functions, we define the three-key variant of HtmB-p1, denoted 3k-HtmB-p1, as:

$$\text{3k-HtmB-p1}[H,\Pi_1,\Pi_2,\Pi_3] := \text{HtmB-p1}[H, \widehat{\Pi}^0_1, \widehat{\Pi}^1_1, \widetilde{\Pi}^0_2, \widetilde{\Pi}^1_2, \widetilde{\Pi}^0_3, \widetilde{\Pi}^1_3].$$

Theorem 7.2. *For $n \geq 8$, $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq 2^{n-5}$, $\Pi_1, \Pi_2, \Pi_3 \leftarrow_\$ \mathcal{P}(n)$, and $(q,\sigma,\epsilon_2,\epsilon_1)$-DbACU$_q$ hash function H_K instantiated with key $K \leftarrow_\$ \mathcal{K}$, the PRF advantage of any (q,∞)-distinguisher $\mathscr{A}$ against* 3k-HtmB-p1$[H,\Pi_1,\Pi_2,\Pi_3]$ *is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\text{3k-HtmB-p1}[H,\Pi_1,\Pi_2,\Pi_3]}(\mathscr{A}) \leq \frac{2q}{2^n} + \frac{6q^{1.5}}{2^{1.5n}} + \frac{64q^2}{2^{2n}} + \frac{128q^2}{2^{3n}} + \epsilon_2 + 2\epsilon_1.$$

Two-key variant of HtmB-p2: Given $\Pi_1, \Pi_2 \leftarrow_\$ \mathcal{P}(n)$ and a pair $H = (H_1, H_2)$ of two $(\mathcal{K}, \mathcal{M}, \{0,1\}^{n-1})$-keyed hash functions, we define the two-key variant of HtmB-p2, denoted 2k-HtmB-p2, as:

$$\text{2k-HtmB-p2}[H,\Pi_1,\Pi_2] := \text{HtmB-p2}[H, \widehat{\Pi}^0_1, \widehat{\Pi}^1_1, \widetilde{\Pi}^0_2, \widetilde{\Pi}^1_2].$$

Theorem 7.3. *For $\epsilon_1, \epsilon_2, \sigma \geq 0$, $q \leq \min\{2^{n-3}, 2^n/67n^2\}$, $\Pi_1, \Pi_2 \leftarrow_\$ \mathcal{P}(n)$, and $(q,\sigma,\epsilon_2,\epsilon_1)$-DbACU$_q$ hash function H instantiated with key $K \leftarrow_\$ \mathcal{K}$, the PRF advantage of any (q,∞)-distinguisher $\mathscr{A}$ against* 2k-HtmB-p2$[H,\Pi_1,\Pi_2]$ *is given by*

$$\mathbf{Adv}^{\mathsf{prf}}_{\text{2k-HtmB-p2}[H,\Pi_1,\Pi_2]}(\mathscr{A}) \leq \frac{128q^2}{2^{3n}} + \frac{136q^2}{2^{2n}} + \frac{8q}{2^n} + \epsilon_2 + 2\epsilon_1.$$

The proofs of Theorem 7.1, 7.2, and 7.3 follow very similar strategies as used in the proofs of Theorem 4.1, 4.2, and 4.4, respectively. So, we skip formal proofs for economical reasons. For the sake of verification, we provide proof sketches in the full version of this paper.

8 Conclusion

In this paper, we proposed a novel method of constructing VIL PRFs, dubbed as the Hash-then-modified-Benes or HtmB transformation. Based on the type of internal primitive, we gave three instances of HtmB, viz. HtmB-f, HtmB-p1, and HtmB-p2. We showed that all three instances retain security for close to 2^n queries. We instantiate the three VIL PRFs using LightMAC+ and PMAC+ based hash functions, called LightHash and PHash, respectively. We explicitly derived relevant collision probability bounds for LightHash and PHash that, in combination with the bounds for HtmB instances, implies almost 2^n blocks security. Lastly, we proposed some reduced-key variants of HtmB-f, HtmB-p1, and HtmB-p2.

8.1 Further Discussion

On Single-key Variants for HtmB-p1 and HtmB-p2: There is a scope of further reducing the key size in case of HtmB-p1 and HtmB-p2 by using 2 and 1 extra bit(s), respectively, for domain separation. However, there is an obstacle in proving the security of resulting constructions. This obstacle stems from the fact that the permutation calls in the lower level are no longer independent of the permutation calls in the upper layer. As a result, the existing bounds on the sum of permutations [8,10] (in case of HtmB-p1) and mirror theory [9,33,34] (in case of HtmB-p2) are no longer applicable. It seems that we need a stronger result like sum of permutations under some added input/output restrictions. A partial positive result in this direction has been shown in [15], where the authors show similar result for queries up to $2^{2n/3}$. We leave it as an open problem to extend the result to close to 2^n queries under appropriate conditions.

On Hash Function Requirement: The reduced-key variants of HtmB need hash functions with unusual output sizes like $2n-2$ and $2n-4$ bits. However, one can easily generate such hash outputs by chopping appropriate bits of an ϵ-Almost XOR Universal (AXU) hash function, i.e. a hash function H_K such that for distinct inputs x, y and any difference δ, $\Pr_K\left[H_K(x) \oplus H_K(y) = \delta\right] \leq \epsilon$. Suppose we have a pair of n-bit hash functions $H = (H_1, H_2)$ that satisfies two properties:

- H_b are ϵ_1-AXU hash functions for $b \in [2]$, and
- H is an ϵ_2-AXU hash function.

Then, if we chop $d < n$ bits from each of H_1 and H_2, the resulting hash function can be shown to be $(q, \sigma, q^2 2^{2d}\epsilon_2, q2^d\epsilon_1)$-DbACU$_q$.

Unfortunately, LightHash and PHash of section 5 do not satisfy the AXU condition. Note that, we saved one block cipher call in LightHash and PHash

as compared to the hash layer in LightMAC+ and PMAC+, by absorbing the last data block directly. It would be interesting to see whether the original hash layer in LightMAC+ and PMAC+ can be used as appropriate replacements for LightHash and PHash, respectively, in the reduced-key variants.

References

1. Rogaway, P., Shrimpton, T.: A provable-security treatment of the key-wrap problem. In: Advances in Cryptology - EUROCRYPT '06. Proceedings. (2006) 373–390
2. Bernstein, D.J.: SURF: simple unpredictable random function (1997)
3. Aumasson, J., Bernstein, D.J.: Siphash: A fast short-input PRF. In: Progress in Cryptology - INDOCRYPT 2012. Proceedings. (2012) 489–508
4. Mennink, B., Neves, S.: Optimal prfs from blockcipher designs. IACR Trans. Symmetric Cryptol. **2017**(3) (2017) 228–252
5. Bellare, M., Krovetz, T., Rogaway, P.: Luby-rackoff backwards: Increasing security by making block ciphers non-invertible. In: Advances in Cryptology - EUROCRYPT '98. Proceeding. (1998) 266–280
6. Bellare, M., Impagliazzo, R.: A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptology ePrint Archive **1999** (1999) 24
7. Lucks, S.: The sum of prps is a secure PRF. In: Advances in Cryptology - EUROCRYPT '00. Proceeding. (2000) 470–484
8. Patarin, J.: A proof of security in o(2n) for the xor of two random permutations. In: Information Theoretic Security, Third International Conference, ICITS 2008. Proceedings. (2008) 232–248
9. Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. IACR Cryptology ePrint Archive **2010** (2010) 287
10. Dai, W., Hoang, V.T., Tessaro, S.: Information-theoretic indistinguishability via the chi-squared method. In: Advances in Cryptology - CRYPTO '17. Proceedings, Part III. (2017) 497–523
11. Cogliati, B., Seurin, Y.: EWCDM: an efficient, beyond-birthday secure, nonce-misuse resistant MAC. In: Advances in Cryptology - CRYPTO '16. Proceedings, Part I. (2016) 121–149
12. Mennink, B., Neves, S.: Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. In: Advances in Cryptology - CRYPTO '17. Proceedings, Part III. (2017) 556–583
13. Yasuda, K.: The sum of CBC macs is a secure PRF. In: Topics in Cryptology - CT-RSA 2010. Proceedings. (2010) 366–381
14. Yasuda, K.: A new variant of PMAC: beyond the birthday bound. In: Advances in Cryptology - CRYPTO '11. Proceedings. (2011) 596–609
15. Datta, N., Dutta, A., Nandi, M., Paul, G., Zhang, L.: Single key variant of pmac_plus. IACR Trans. Symmetric Cryptol. **2017**(4) (2017) 268–305
16. Zhang, L., Wu, W., Sui, H., Wang, P.: 3kf9: Enhancing 3gpp-mac beyond the birthday bound. In: Advances in Cryptology - ASIACRYPT '12. Proceedings. (2012) 296–312
17. Naito, Y.: Blockcipher-based macs: Beyond the birthday bound without message length. In: Advances in Cryptology - ASIACRYPT '17. Proceedings, Part III. (2017) 446–470

18. Datta, N., Dutta, A., Nandi, M., Paul, G.: Double-block hash-then-sum: A paradigm for constructing bbb secure prf. IACR Trans. Symmetric Cryptol. **2018**(3) (2018) 36–92
19. Jha, A., Nandi, M.: Tight security of cascaded LRW2. J. Cryptology **33**(3) (2020) 1272–1317
20. Kim, S., Lee, B., Lee, J.: Tight security bounds for double-block hash-then-sum macs. In: Advances in Cryptology - EUROCRYPT 2020, Proceedings, Part I. (2020) 435–465
21. Shoup, V.: Sequences of games: a tool for taming complexity in security proofs. IACR Cryptology ePrint Archive **2004** (2004) 332
22. Aiello, W., Venkatesan, R.: Foiling birthday attacks in length-doubling transformations - benes: A non-reversible alternative to feistel. In: Advances in Cryptology - EUROCRYPT '96. Proceeding. (1996) 307–320
23. Patarin, J.: Security of balanced and unbalanced feistel schemes with linear non equalities. IACR Cryptology ePrint Archive **2010** (2010) 293
24. Patarin, J.: A proof of security in $o(2^n)$ for the benes scheme. In: Progress in Cryptology - AFRICACRYPT '08. Proceedings. (2008) 209–220
25. Patarin, J., Montreuil, A.: Benes and butterfly schemes revisited. In: Information Security and Cryptology - ICISC '05. Revised Selected Papers. (2005) 92–116
26. Naito, Y.: The exact security of PMAC with two powering-up masks. IACR Trans. Symmetric Cryptol. **2019**(2) (2019) 125–145
27. Chakraborty, B., Chattopadhyay, S., Jha, A., Nandi, M.: On length independent security bounds for the PMAC family. IACR Cryptol. ePrint Arch. **2020** (2020) 656
28. Bellare, M., Goldreich, O., Mityagin, A.: The power of verification queries in message authentication and authenticated encryption. IACR Cryptology ePrint Archive **2004** (2004) 309
29. Hall, C., Wagner, D.A., Kelsey, J., Schneier, B.: Building prfs from prps. In: Advances in Cryptology - CRYPTO '98. Proceedings. (1998) 370–389
30. Patarin, J.: Etude des Générateurs de Permutations Pseudo-aléatoires Basés sur le Schéma du DES. PhD thesis, Université de Paris (1991)
31. Patarin, J.: The "coefficients H" technique. In: Selected Areas in Cryptography - SAC '08. Revised Selected Papers. (2008) 328–345
32. Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. In: Advances in Cryptology - EUROCRYPT '14. Proceedings. (2014) 327–350
33. Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. **28**(4) (2017) 321–338
34. Nachef, V., Patarin, J., Volte, E.: Feistel Ciphers - Security Proofs and Cryptanalysis. Springer (2017)
35. Mennink, B.: Towards tight security of cascaded LRW2. In: Theory of Cryptography - TCC '18. Proceedings, Part II. (2018) 192–222
36. Landecker, W., Shrimpton, T., Terashima, R.S.: Tweakable blockciphers with beyond birthday-bound security. In: Advances in Cryptology - CRYPTO '12. Proceedings. (2012) 14–30
37. Datta, N., Dutta, A., Nandi, M., Yasuda, K.: Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based MAC. In: Advances in Cryptology - CRYPTO '18. Proceedings, Part I. (2018) 631–661
38. Naito, Y.: The exact security of PMAC with three powering-up masks. IACR Cryptol. ePrint Arch. **2020** (2020) 731