IACR Transactions on Symmetric Cryptology
ISSN 2519-173X, Vol. 2020, No. 2, pp. 1–39. DOI:10.13154/tosc.v2020.i2.1-39

On the Composition of Single-Keyed Tweakable Even-Mansour for Achieving BBB Security

Avik Chakraborti[1,2], Mridul Nandi[1], Suprita Talnikar[1] and Kan Yasuda[2]

[1] Indian Statistical Institute, Kolkata, India,
avikchkrbrti@gmail.com,mridul.nandi@gmail.com,suprita45@gmail.com
[2] NTT Secure Platform Laboratories, Tokyo, Japan, yasuda.kan@lab.ntt.co.jp

Abstract. Observing the growing popularity of random permutation (RP)-based designs (e.g, Sponge), Bart Mennink in CRYPTO 2019 has initiated an interesting research in the direction of RP-based pseudorandom functions (PRFs). Both are claimed to achieve beyond-the-birthday-bound (BBB) security of $2n/3$ bits (n being the input block size in bits) but require two instances of RPs and can handle only one-block inputs. In this work, we extend research in this direction by providing two new BBB-secure constructions by composing the tweakable Even-Mansour appropriately. Our first construction requires only one instance of an RP and requires only one key. Our second construction extends the first to a nonce-based Message Authentication Code (MAC) using a universal hash to deal with multi-block inputs. We show that the hash key can be derived from the original key when the underlying hash is the Poly hash. We provide matching attacks for both constructions to demonstrate the tightness of the proven security bounds.

Keywords: PDMMAC · Davis-Meyer · PRF · MAC · permutation · beyond the birthday bound security

1 Introduction

There is significant research on the design of PRFs from PRPs and vice versa. The most relevant work based on PRP-from-PRF has been the Luby-Rackoff construction [LR88]. However, this direction is not very popular as PRPs are easier to build than PRFs and several cryptographic designs desire to be instantiated with PRFs. In fact, the research community has found it a better proposition to go the other way around - constructing PRFs from PRPs. Numerous works have been done in this area. The main reason behind this is that a PRP can be more easily designed from a PRF than a PRF from a PRP.

PRP-Based PRFs

The most obvious way a PRF can be constructed is to consider a PRP P (popular choice is an n-bit block cipher with uniformly sampled key e_K for some integer n) itself as a PRF. However, this leads to an $n/2$-bit secure PRF. This result comes from the fact that $2^{n/2}$ evaluations of the PRF will lead to a collision with significant probability while the collision probability in case of a PRP will be zero. This is also termed the PRP-PRF switching [BKR00, BR06, CN08, HWKS98]. In light of the recent research in lightweight cryptography, this bound may not be acceptable to designers. The value of n is generally chosen to be small because the state size of the PRF directly depends on n and lightweight designs aim to optimize it. For example, several lightweight block ciphers [BSS+13, BCG+12, BJK+16, BKL+07, BPP+17] that are proposed with a 64-bit state (i.e, $n = 64$) achieve only 32-bit security and can be broken with practical query

Received: 2020-03-02, Accepted: 2020-05-01, Published: 2020-07-24

complexity. This idea has resulted in several attempts to design a PRF from a PRP with more than $n/2$-bit security. They are popularly known as *Beyond-the-Birthday-bound* (BBB)-secure PRFs.

A first attempt to construct such a BBB-secure PRF denoted by $\mathsf{Trunc}_m(e_K(x))$, was proposed by Hall et al. [HWKS98], where $m < n$ (note that a block cipher is a popular candidate for a PRP), and its security was bounded by $2^{n-m/2}$ queries [BI99, GG16]. Later in [BKR98], Bellare et al. proposed n-bit security [BI99, DHT17, Luc00, Pat10] of $e_{K_1}(x) \oplus e_{K_2}(x)$ where K_1 and K_2 are independently sampled. Seurin et al. proposed a $2^{2n/3}$ query-secure PRF, which they called EDM [CS18],

$$e_{K_2}(e_{K_1}(x) \oplus x).$$

The security of this construction has been improved by Mennink [MN17] using Patarin's mirror theory [NPV17, Pat05, Pat10, Pat16]. Note that all constructions are deterministic (no use of nonce) and are instantiated with block ciphers with inputs considered to be of fixed length. However, there are a number of BBB-secure constructions that deal with arbitrary length inputs.

Generally, the technique is to incorporate a nonce and a keyed hash. The nonce is processed with a deterministic PRF and the output is properly integrated with the hashed value of the arbitrary length message. Note that, except a few, most of the PRFs do not allow nonce misuse. The WC-MAC [CW79, WC81] (Wegman-Carter MAC) is one of such constructions where the nonce is processed with a PRP-based PRF and a universal hash processes the message. Next, both the outputs are added and passed through another instance of PRP to generate the output. This design is vulnerable to nonce misuse but secure up to only birthday bound under nonce respect. Later, Cogliati and Seurin updated the WC MAC and designed the EWC-MAC [CS18] (Encrypted Wegman-Carter): $e_{K_2}(f_{K_1}(x) \oplus \mathcal{H}_{K_h}(x))$ (f is a deterministic PRF, $\mathcal{H}$ is a key universal hash and K_1, K_2 and K_h are unform and independent), which is birthday bound secure under both nonce misuse and respect scenario (can be proved using the PRP-PRF switching lemma). The most important question that arises is "*How can a BBB secure PRF be designed?*" The first prominent design in this area is the EWCDM construction:

$$e_{K_2}((e_{K_1}(N) \oplus N) \oplus \mathcal{H}_{K_h}(x))$$

by Cogliati et al. [CS18], where the PRF is instantiated by Davis-Meyer and is used in the EWC mode. This design achieves BBB security of $2n/3$-bits (though n-bit security was conjectured and proved by Mennink et al. [MN17] using mirror theory) under nonce respect and birthday bound under nonce misuse. However, this construction is not minimal in structure as it uses two independent instances of keys K_1 and K_2. Datta et al., in [DDNY18a, DDNY18b] proposed DWCDM which is a BBB secure construction (under nonce respect), and uses only one instance of the PRP where e_{K_2} is replaced by $e^{-1}_{K_1}$. In the security proof, the authors extended mirror theory and provide a concrete proof of security up to $2^{2n/3}$queries under nonce respect and birthday bound complexity under nonce misuse. Nevertheless, the bound is not tight as there does not exist any attack below 2^n queries. In fact the design is conjectured to have n-bit security.

Permutation-Based Designs

With the advent of public permutation-based designs and the efficiencies of permutations in the forward direction, several inverse-free hash and authenticated encryption schemes have been proposed. The most prominent of such designs are the Sponge designs introduced in SHA3 through the Keccak hash [BDPA15], this research direction later being extended by popular designs like PHOTON [GPP11]. Several AEAD designs like

Keyed Sponge [ADMA15, BDPA11b, BDPA11a, MRV15], SPONGENT [BKL+11], ASCON [DEMS16], Beetle [CDNY18] have later been proposed. Permutation-based designs generally provide lower security bounds and it can be highly interesting to design RP-based PRF with BBB security on the permutation size. Mennink et al., [CLM19a] recently in CRYPTO 2019, studied permutation-based PRFs and proposed two BBB secure constructions denoted as SOEM and SOKAC. However, both designs are not minimal in structure and cannot handle arbitrary-length data. Both use two independent instances of random permutations and at least one randomly sampled key. They are deterministic and do not handle nonce. In this paper, we explore this direction of research and address the following relevant questions: *Can we design minimally structured PRF? (i.e, with one instance of random permutation) Does there exist a nonce-based MAC constructed using an RP which is again minimal in structure and can handle arbitrary-length data?* We found the answer to be "yes", and we mainly propose two BBB secure deterministic and nonce based designs using only one instance of a random permutation and one uniformly sampled construction key. We list our contributions below.

1.1 Motivation

The initial motivation for our construction arises from the fact that there are no single key, single permutation-based MACs with BBB security. No similar BBB secure permutation-based (or even nonce-based) MAC construction currently exists other than SoEM22 [CLM19a], which is also based on two permutations. In fact, [CLM19a] also provides birthday bound attacks for the 2-permutations-1-keyed (π_1, π_2, K) and 1-permutation-2-keyed (π, K_1, K_2) constructions, thus leaving no scope for improvement in SoEM. It is therefore clear that a sequential construction is required for a minimization; SoKAC [CLM19a] is the only existing sequential construction, a birthday bound attack to which is present in [Nan20].

Two variants of SoKAC, namely SoKAC1 and SoKAC21 seem to have the following inconsistencies:

1. The authors claim a birthday bound security of SoKAC1 in Proposition 5 of [CLM19a], whose proof claims a distinguishing attack that does not seem to work. Hence, a corrected attack is required for SoKAC1.
2. SoKAC21 is claimed to achieve a tight $2n/3$-bit security in Proposition 6 of [CLM19a], accompanied by an attack with a query complexity of $\mathcal{O}\left(2^{2n/3}\right)$. This security is proved flawed in [Nan20], which shows a birthday bound attack on SoKAC21.

The main reason behind the above inconsistencies is the fixing of the input to the second permutation π_2 (or π) by the output of the first permutation π_1 (or π). Thus, although the final tag is a sum of the outputs of π_1, π_2 and a secret key, the fixing of the permutation input prevents construction of a transcript-inducing graph and subsequent use of Mirror theory.
This implies that the current form of SoKAC may not be a convincing construction to build upon. Our construction takes a different direction from SoKAC, and is inspired by DWCDM [DDNY18a, DDNY18b] - the output of only one permutation is involved in the tag generation and the sum of permutations occurs between the two permutation instances, allowing a query fixing the input and output of the construction (not the permutations) to be clearly described by an inducing graph, which was not the case in SoKAC. Thus, Mirror theory in its present form can be directly applied to our construction.

1.2 Our Contributions

We address the problem of designing a generic BBB secure MAC based on RP with the minimal structure. The term "minimal" refers to the number of instances of the

Table 1: Comparison of existing PRFs. #Keys and #Primitives denote number of key and primitive instances.

Construction	#Key Instances	#Primitive Instances	MAC Security in n-bits (tightness)	Nonce Based	Multi-Block Inputs
Based on permutations					
PDMMAC [**This work**]	1	1	$2n/3$ (tight)		
PDM*MAC [**This work**]	$1+1$ (hash key)	1	$2n/3$ (tight)	✓	✓
1K-PDM*MAC [**This work**]	1	1	$2n/3$ (tight)	✓	✓
SoEM1 [CLM19a]	2	1	- (birthday attack)		
SoEM21 [CLM19a]	1	2	- (birthday attack)		
SoEM22 [CLM19a]	2	2	$2n/3$ (tight)		
SoKAC1 [CLM19a]	2	1	- (birthday attack)		
SoKAC21 [CLM19a]	1	2	- (birthday attack) [Nan20]		
Based on Block Ciphers					
EDM [CS18]	2	2	$2n/3$ (not tight)		
EWCDM [CS18]	$2+1$ (hash key)	2	$2n/3$ (not tight)	✓	✓
DWCDM [DDNY18a, DDNY18b]	$1+1$ (hash key)	1	$2n/3$ (not tight)	✓	✓
1K-DWCDM [DDNY18a, DDNY18b]	1	1	$2n/3$ (not tight)	✓	✓

internal mathematical components (similar to DWCDM - Decrypted Wegman-Carter with Davies-Meyer, which minimizes the number of block cipher instances). Our proposal only uses one key and two calls of the same permutation (one forward and one inverse). The key is used to generate three sub-keys that are injected in between the two permutation calls. Precisely:

- We propose a deterministic MAC denoted by PDMMAC (***P**ermutation based **D**avis-**M**eyer*) using one permutation and one key instance. We prove its PRF (which also upper bounds the MAC security) security up to $2^{2n/3}$ queries under the random permutation model. We provide a proof using the coefficients-H technique. The bound has been proven to be tight with a matching attack with query complexity $2^{2n/3}$.

- The previous result sparks curiosity about the achievability of $2n/3$-bit security by a minimal construction that can process arbitrary length inputs. We propose a nonce-based MAC denoted by PDM*MAC using an additional keyed hash. We provide a BBB secure nonce-based MAC security proof of $2^{2n/3}$ query complexity under the nonce-respect scenario. We show the tightness of the proven security bound by demonstrating a matching attack.

- We propose a one keyed instance of PDM*MAC denoted by 1K-PDM*MAC by instantiating the hash key K_h as $K_h = \pi(K)$, where π is the underlying RP. In addition, the underlying nonce is chosen to be non zero and the hash function is chosen as Poly hash. This instance achieves the same security bound as PDM*MAC.

Table 1 describes the structures of several well known constructions in terms of the primitives and other design properties.

2 Preliminaries

The set of all n-bit binary strings is denoted by $\{0,1\}^n$, for an integer $n \in \mathbb{N}$. We denote the empty string by λ. $\{0,1\}^*$ set of all strings such that $\{0,1\}^* = \{0,1\}^+ \cup \{\lambda\}$. For $x, y \in \{0,1\}^*$, $x\|y$ is used to denote the concatenation of x and y. $\langle i \rangle_m$ denotes the binary representation of an integer i in m bits. We use $\mathsf{Func}(\mathsf{D}, \mathsf{R})$ to denote the set of all functions from the set D to R and $\mathsf{Perm}(S)$ to denote the set of all permutations over the set S. Typically the choices of D, R and S are taken over binary strings. We also denote a set of consecutive integers $\{1, \ldots, r\}$ simply by $[r]$.

2.1 PRF Security in the Random Permutation Model

Consider a function $f : \mathcal{K} \times \mathcal{M} \to \mathcal{T}$, where $\mathcal{K}$, $\mathcal{M}$ and $\mathcal{T}$ are the key space, message space and the tag space respectively. We discuss the pseudorandom security of f under the random permutation model. We assume that f makes internal public-random-permutation calls to π and π^{-1} (f can make calls to multiple random permutations when all of them are independent and uniform on the set of message blocks $\mathsf{Perm}(\mathcal{B})$). For simplicity, we use f_K^π to denote f with uniform K and uniform π. The distinguisher $\mathcal{D}$ is given access to either (f_K^π, π, π^{-1}) for $K \xleftarrow{\$} \{0,1\}^k$ or (ψ, π, π^{-1}) where $\psi \xleftarrow{\$} \mathsf{Func}(\mathcal{K} \times \mathcal{M}, \mathcal{T})$ is a random oracle. The distinguishing probability of $\mathcal{D}$ is represented by $\mathsf{Adv}_f^{\mathsf{prf}}(\mathcal{D})$, such that

$$\mathsf{Adv}_f^{\mathsf{prf}}(\mathcal{D}) = |\mathsf{Pr}[\mathcal{D}^{(f_K^\pi, \pi, \pi^{-1})} = 1] - \mathsf{Pr}[\mathcal{D}^{(\psi, \pi, \pi^{-1})} = 1]|.$$

To be precise, we call f an ϵ-PRF against (q_m, p)-adversaries if $\mathrm{Adv}_F^{PRF}(D) \leq \epsilon$ for all distinguishers $\mathcal{D}$ making q_m queries to f_K^π and p offline queries to π.

2.2 MAC Security in the Random Permutation Model

Consider f and another function $\mathsf{Ver} : \mathcal{K} \times \mathcal{M} \times \mathcal{T} \to \{0,1\}$ (similar to f_K^π, we use the notation Ver_K^π) such that for $(M,\ T)$, if $f_K^\pi(M) = T$ then $\mathsf{Ver}_K^\pi(M,T) = 1$ (otherwise $\mathsf{Ver}_K^\pi(M,T) = 0$). Consider a (q_m, p, q_v) adversary $\mathcal{A}$ making q_m queries to f_K^π, p queries to π and q_v queries to Ver_K^π. We say that $\mathcal{A}$ *forges* if any of its queries (M,T) to Ver_K^π returns 1, such that M has not been queried to f_K^π before. The advantage of $\mathcal{A}$ against the MAC security of f is defined as

$$\mathsf{Adv}_f^{\mathsf{MAC}}(\mathcal{A}) = \mathsf{Pr}[K \xleftarrow{\$} \mathcal{K}, \pi \xleftarrow{\$} \mathsf{Perm}(\mathcal{B}) :\ \mathcal{A}\ \textit{forges}].$$

To be precise, we call f an ϵ-MAC against (q_m, p, q_v)-adversaries if $\mathsf{Adv}_f^{\mathsf{MAC}}(\mathcal{A}) \leq \epsilon$ for all adversaries $\mathcal{A}$ making q_m queries to f_K^π, p queries to π and q_v queries to Ver_K^π.

2.3 Nonce-Based MAC Security in the Random Permutation Model

Consider nonce based versions of f and Ver (takes an additional input $N \in \mathcal{N}$.) such that for an input $(N,\ M,\ T)$, $\mathsf{Ver}_K^\pi(N,M,T) = 1$ if $f_K^\pi(N,M) = T$ and 0 otherwise. Consider a (q_m, p, q_v) adversary $\mathcal{A}$ making q_m queries to f_K^π without repeating the nonce, p queries to π and q_v queries to Ver_K^π. We say that $\mathcal{A}$ *forges* if any of its queries (N,M,T) to Ver_K^π, such that (N,M) has not been queried to f_K^π, returns 1. The advantage of $\mathcal{A}$ against the MAC security of f is defined as

$$\mathsf{Adv}_f^{\mathsf{MAC}}(\mathcal{A}) = \mathsf{Pr}[K \xleftarrow{\$} \mathcal{K}, \pi \xleftarrow{\$} \mathsf{Perm}(\mathcal{B}) :\ \mathcal{A}\ \textit{forges}].$$

We call f an ϵ-MAC against (q_m, p, q_v)-adversaries if $\mathsf{Adv}_f^{\mathsf{MAC}}(\mathcal{A}) \leq \epsilon$ for all (q_m, p, q_v)-adversaries $\mathcal{A}$.

2.3.1 Upper Bound on $\mathsf{Adv}_f^{\mathsf{MAC}}$ (Page 5, [DJN17]):

To get an upper bound for $\mathsf{Adv}_f^{\mathsf{MAC}}$, we consider a random oracle $\psi \xleftarrow{\$} \mathsf{Func}(\mathcal{K} \times \mathcal{N} \times \mathcal{M}, \mathcal{T})$ and reject oracle $\mathsf{Rej} : \mathcal{N} \times \mathcal{M} \times \mathcal{T} \to \{0\}$. The advantage $\mathsf{Adv}_f^{\mathsf{MAC}}$ is upper bounded by

$$\max_{\mathcal{D}} \left| \mathsf{Pr}[\mathcal{D}^{(f_K^\pi, \mathsf{Ver}_K^\pi, \pi, \pi^{-1})} = 1] - \mathsf{Pr}[\mathcal{D}^{(\psi, \mathsf{Rej}, \pi, \pi^{-1})} = 1] \right|.$$

2.4 Keyed Hash

Regular Hash: A function $\mathcal{H} : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$ is said to be an ϵ-regular hash function if $\forall$ $d \in \mathcal{D}$ and $r \in \mathcal{R}$,

$$\Pr_{K_h \xleftarrow{\$} \mathcal{K}} [\mathcal{H}(K_h, d) = r] \leq \epsilon.$$

AXU Hash: A function $\mathcal{H} : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$ is said to be an ϵ-AXU hash function if for two distinct d and d' from $\mathcal{D}$ and $r \in \mathcal{R}$,

$$\Pr_{K_h \xleftarrow{\$} \mathcal{K}} [\mathcal{H}(K_h, d) \oplus \mathcal{H}(K_h, d') = r] \leq \epsilon.$$

3-Way Regular Hash: A function $\mathcal{H} : \mathcal{K} \times \mathcal{D} \to \mathcal{R}$ is said to be an ϵ-3-way regular hash function if for three distinct d, d' and d'' from $\mathcal{D}$ and for any non-zero r from $\mathcal{R}$,

$$\Pr_{K_h \xleftarrow{\$} \mathcal{K}} [\mathcal{H}(K_h, d) \oplus \mathcal{H}(K_h, d') \oplus \mathcal{H}(K_h, d'') = r] \leq \epsilon.$$

An example of 3-way regular hash is Poly hash (with the secret key K_h) where the padded message $x^* = x_1 \| \cdots \| x_\ell$ is processed as

$$\mathsf{Poly}_{K_h}(x^*) = x_\ell \cdot K_h \oplus x_{\ell-1} \cdot K_h^2 \oplus \cdots \oplus x_1 \cdot K_h^\ell.$$

2.5 Coefficients-H Technique

We outline the Coefficients-H technique developed by Patarin, which serves as a convenient tool for bounding the advantage (see [Pat91, Vau03]). We will use this technique (without giving a proof) to prove our main theorem. Consider two oracles $\mathbf{O}_0 = (\$, \perp)$ (the ideal oracle for the relaxed [1] game) and $\mathbf{O}_1$ (real, i.e. our construction in the same relaxed game). Let $\mathcal{T}$ denote the set of all possible transcripts an adversary can obtain (i.e. the set of all *attainable* transcripts in the ideal world). We let X_{re} be the random variable that takes values $\tau \in \mathcal{T}$ when the adversary interacts with the real world and X_{id} to be the random variable that takes values $\tau \in \mathcal{T}$ when it interacts with the ideal world. Without loss of generality, we assume that the adversary is deterministic and fixed. Then the sample space for X_{re} and X_{id} is uniquely determined by the underlying oracle. As we deal with stateless oracles, these probabilities are independent of the order of query responses in the transcript. Suppose we have a set of transcripts, $\mathcal{T}_{\mathrm{good}} \subseteq \mathcal{T}$, which we call *good* transcripts, and the following conditions hold:

1. In the game involving the ideal oracle $\mathbf{O}_0$ (and the fixed adversary), the probability of getting a transcript in $\mathcal{T}_{\mathrm{good}}$ is at least $1 - \epsilon_1$.
2. For any transcript $\tau \in \mathcal{T}_{\mathrm{good}}$, we have
   $\Pr[\mathsf{X}_{\mathrm{re}} = \tau] \geq (1 - \epsilon_2) \cdot \Pr[\mathsf{X}_{\mathrm{id}} = \tau]$.

Then $|\Pr[\mathcal{D}^{\mathbf{O}_0} = 1] - \Pr[\mathcal{D}^{\mathbf{O}_1} = 1]| \leq \epsilon_1 + \epsilon_2$. The proof can be found in (say) [Vau03].

2.6 Two Sum-Capture Lemmas

Let $T^\star$ be a multiset of size q. Denote by $\mu(T^\star)$, the maximum over all subsets A, B of $\{0,1\}^n$, both of size q, of the quantity

$$\mu(T^\star, A, B) = |\{(T_i, \tilde{v}_a, \tilde{z}_b) \in T^\star \times A \times B \mid T_i = \tilde{v}_a \oplus \tilde{z}_b\}|.$$

[1] the term *relax* denotes that in addition to the query input-output tuples, additional state values may be supplied to the adversary (after all the queries are made) as a part of the transcripts

Recall lemma 1 of [CS18]:
Let T^* be a multiset of $q \geq 1$ uniformly random and independently chosen elements of $\{0,1\}^n$. Then-

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

This lemma can be slightly altered by simply taking the sizes of the multiset T^* and the sets A, B to be q, p_1, p_2, respectively:

Lemma 1. *Let $T^\star = \{T_1, \ldots, T_q\}$ be the multiset of all the tags received through the $[q]$ queries to the construction. Denote by $\mu(T^\star)$, the maximum over all subsets A, B of $\{0,1\}^n$, of size p_1 and p_2 respectively, of the quantity*

$$\mu(T^\star, A, B) = \left|\{(T_i, \tilde{v}_a, \tilde{z}_b) \in T^\star \times A \times B \mid T_i = \tilde{v}_a \oplus \tilde{z}_b\}\right|.$$

$$\Pr\left[\mu(T^\star) \geq \frac{p_1 p_2 q}{2^n} + \sqrt{3np_1p_2q}\right] \leq \frac{2}{2^n}. \tag{1}$$

If the set A is replaced by a multiset A^*, then this result is further modified into the following lemma:

Lemma 2. *Let T^*, A^* be multisets of $\{0,1\}^n$ and $B \subseteq \{0,1\}^n$. Define-*

$$\mu(T^*, A^*, B) = \left|\{(t, a, b) \in T^* \times A^* \times B : t = a \oplus b\}\right| \text{ and}$$

$$\mu(T^*) = \max_{\substack{A^*, B \\ |T^*|=q_1, |A^*|=q_2, |B|=p}} \mu(T^*, A^*, B).$$

If T^*, A^* are multisets of respectively q_1, q_2 uniformly random and independently chosen elements of $\{0,1\}^n$ and B is a subset of $\{0,1\}^n$ of size p, then

$$\Pr\left[\mu(T^*) \geq \frac{q_1 q_2 p}{2^n} + \sqrt{\frac{3np(q_1+q_2)}{2^n}}\right] \leq \frac{2}{2^n}. \tag{2}$$

A proof is available in Supplementary Material A

3 Mirror Theory

Mirror Theory: Mirror theory is a tool for finding the number of solutions to affine systems of equalities and non-equalities. Mirror theory by Patarin [NPV17,Pat05,Pat10,Pat16] provides a lower bound on such a number for a finite set of affine bi-variate equations, which is such that its variables are sampled without replacement. The proof is verifiable up to a bound of $2n/3$ bits.

Equation-Inducing Graph: Consider an undirected graph $\mathcal{G}_{\text{eq}} = (\mathsf{V}_{\text{eq}}, \mathsf{E}_{\text{eq}}, \mathcal{L})$, where $\mathsf{V}_{\text{eq}} = \{X_1, \ldots, X_m\}$ and the edge-label function $\mathcal{L} : \mathsf{E}_{\text{eq}} \to \mathbb{F}_{2^n}$ assigns a label λ to each edge $e \in \mathsf{E}_{\text{eq}}$.
If each vertex X_i is assumed to represent a unique variable (also denoted X_i, for the sake of convenience), then such a graph $\mathcal{G}_{\text{eq}}$ can be considered to induce a system of equations defined by-

$$X_i \oplus X_j = \lambda_{i,j}, \text{ whenever } e_{i,j} := \{X_i, X_j\} \in \mathsf{E}_{\text{eq}} \text{ and } \mathcal{L}(e_{i,j}) = \lambda_{i,j}.$$

Observe that should any of the following cases occur, the graph $\mathcal{G}_{\text{eq}}$ might induce a system of equations which is either inconsistent or has redundant equations:

- **Existence of a cycle:** A cycle arises in $\mathcal{G}_{\text{eq}}$ if there exists a sequence of edges $\{X_{i_1}, X_{j_1}\}, \ldots, \{X_{i_r}, X_{j_r}\} \in \mathsf{E}_{\text{eq}}$ such that $X_{j_a} = X_{i_{a+1}}\ \forall\ a \in [r-1]$ and $X_{j_r} = X_{i_1}$. A loop i.e. $X_i = X_j$ for some edge $\{X_i, X_j\} \in \mathsf{E}$ is also considered a cycle.
- **Zero Path Label:** The path label of a path P of edges in E_{eq} is defined as $\mathcal{L}(P) = \sum_{e \in P} \mathcal{L}(e)$. Thus, a zero path-label arises when there exists a path P in $\mathcal{G}$ such that $\mathcal{L}(P) = 0$.

Extended Mirror Theory: Extended mirror theory gives a lower bound for the number of solutions to a combination of a system of bi-variate affine equations (as in Mirror Theory) and a system of bi-variate affine non-equations of the form $X_i \oplus Y_i \neq c$. [DNT19] contains a detailed treatment of such a combination of systems.

Equations-and-Non-Equations-Inducing Graph: Consider an undirected graph $\mathcal{G}_{\text{eq}} = (\mathsf{V}_{\text{eq}}, \mathsf{E}_{\text{eq}}, \mathcal{L}_{\text{eq}})$, where $\mathsf{V}_{\text{eq}} = \{X_1, \ldots, X_m\}$ and the edge-label function $\mathcal{L}_{\text{eq}} : \mathsf{E}_{\text{eq}} \to \mathbb{F}_{2^n}$ assigns a label λ to each edge $e \in \mathsf{E}_{\text{eq}}$.

If each vertex X_i is assumed to represent a unique variable (also denoted X_i, for the sake of convenience), then such a graph $\mathcal{G}_{\text{eq}}$ can be considered to induce a system of equations defined by-

$$X_i \oplus X_j = \lambda_{i,j}, \text{ whenever } \mathsf{E}_{i,j} := \{X_i, X_j\} \in \mathsf{E}_{\text{eq}} \text{ and } \mathcal{L}_{\text{eq}}(\mathsf{E}_{i,j}) = \lambda_{i,j}.$$

Now consider an undirected graph $\mathcal{G}_{\text{eq,neq}} = (\mathsf{V}, \mathsf{E}_{\text{eq}} \sqcup \mathsf{E}_{\text{neq}}, \mathcal{L})$, where $\mathsf{V}_{\text{eq}} = \{X_1, \ldots, X_m\} \subseteq V = \{X_1, \ldots, X_v\}$ and the edge-label function $\mathcal{L} : \mathsf{E}_{\text{eq}} \sqcup \mathsf{E}_{\text{neq}} \to \mathbb{F}_{2^n}$ assigns a label λ to each edge $e \in \mathsf{E}_{\text{eq}} \sqcup \mathsf{E}_{\text{neq}}$.

Again assuming each vertex X_i, $\mathcal{G}_{\text{eq,neq}}$ can be considered to induce a system of equations and a system of non-equations defined by-

$$\begin{aligned}
X_i \oplus X_j = \lambda_{i,j}, \quad &\text{whenever} \quad e_{i,j} := \{X_i, X_j\} \in \mathsf{E}_{\text{eq}} \text{ and} \\
&\qquad \mathcal{L}(e_{i,j}) = \lambda_{i,j},\ \forall\, X_i, X_j \in \mathsf{V}_{\text{eq}} \\
X'_i \oplus X'_j \neq \lambda'_{i,j}, \quad &\text{whenever} \quad e'_{i,j} := \{X'_i, X'_j\} \in \mathsf{E}_{\text{neq}} \text{ and} \\
&\qquad \mathcal{L}(e'_{i,j}) = \lambda'_{i,j},\ \forall\, X'_i, X'_j \in V.
\end{aligned}$$

Let $\mathcal{G}_{\text{eq,neq}} = (\mathsf{V}, \mathsf{E}_{\text{eq}} \sqcup \mathsf{E}_{\text{neq}}, \mathcal{L})$ be a graph that induces a system of affine bivariate equations and non-equations over α distinct variables. Suppose $\mathcal{G}_{\text{eq,neq}}$ has α vertices and $q'_m + q_v$ edges with $|\mathsf{E}_{\text{eq}}| = q'_m, |\mathsf{E}_{\text{neq}}| = q_v$. Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be all the components (i.e. maximal subgraphs where any two vertices are connected to each other by a path) of $\mathcal{G}_{\text{eq}} = (\mathsf{V}, \mathsf{E}_{\text{eq}}, \mathcal{L}\big|_{\mathsf{E}_{\text{eq}}})$, $\mathcal{C}_i$ of size w_i, and let $\sigma_i = (w_1 + \cdots + w_i)$. Denote by $\xi_{\max}$, the size of the component of $\mathcal{G}_{\text{eq}}$ with the maximum number of vertices. Using an *extended version* of mirror theory, we can provide a lower bound on the number of injective solutions when the maximum component size is $\xi_{\max}$. We now state the following lemma, which summarizes the result of Theorem 3 in [DNT19]

Lemma 3. *The total number of injective solutions chosen from a set $\mathcal{Z}$ of size $2^n - c$, for some $c \geq 0$, for the induced system of equations and non-equations $\mathcal{G}_{\text{eq,neq}}$ is at least:*

$$(2^n)_\alpha \left(1 - \sum_{i=1}^{k} \frac{6\sigma_{i-1}^2 \binom{\xi_i}{2}}{2^{2n}} - \frac{2(q_v + c\alpha)}{2^n}\right),$$

provided $\sigma_k \xi_{\max} \leq 2^n/4$, and assuming $\sigma_0 = 0$.

This lemma thus provides a bound for a solution from a subset of $\{0,1\}^n$. However, applying this lemma to our results (Thm. 1, 2 and 3) generates the term $\frac{p(p+q)}{2^n}$ for the non-equations, as c takes the value p, which is not a constant. We wish for a beyond-the-birthday bound on this number, which could possibly have been achieved by the results in [DDNY18a, DDNY18b] . In spite of this providing a stronger bound, there are two problems. First, non-equations are unaccounted, which could be easily included (by the same method as in the proof of Cor. 2). Second, a maximum component size of only 2 is allowed for the equations-inducing subgraph. A modification of this result is presented here (Cor. 1 and Cor. 2), which not only takes non-equations into account and allows for a maximum size of 3 for equation-components, but also provides an improved bound.

3.1 Extended Mirror Theory

Some Probability Results

Recall the following result from [DDNY18b]: *Let $S' \subseteq \{0,1\}^n$ be a subset of size $(2^n - s')$ and $U_n \leftarrow \{0,1\}^n$. Let $(V, W) \xleftarrow[wor]{\$} S'^{(2)}$ be a WOR sample of size 2 drawn from S'. Then,*

$$V \oplus W \succ_{\epsilon_1(s')} U_n \text{ over } \mathbb{F}^\star_{2^n} := \mathbb{F}_{2^n} \setminus \{0^n\}, \tag{3}$$

where $\epsilon_1(s')$ is a quantity with value at most $\frac{s'^2}{(2^n - s')^2}$. This result can be extended for three random variables as follows:

Lemma 4. *Let $S' \subseteq \{0,1\}^n$ be a subset of size $(2^n - p')$ and $U_n, V_n \leftarrow \{0,1\}^n$. Let $(P, Q, R) \xleftarrow[wor]{\$} S'$ be a WOR sample of size 3 drawn from $S'^{(3)}$. Then,*

$$(P \oplus Q, Q \oplus R) \succ_{\epsilon_2(p')} (U_n, V_n), \tag{4}$$

where $\epsilon_2(p')$ is a quantity with value at most $\frac{3 \cdot 2^n \cdot p'^2 - p'^3}{(2^n - p')^3}$.

The proof is similar to that of the previous result, and is detailed in Sect. B of the Appendix.

Results on Mirror Theory

Eqn.s 3 and 4 can be easily extended for systems of equations as follows-

Corollary 1. *Let $S' \subseteq \{0,1\}^n$ be a subset of size $(2^n - p')$ and*

$$(X_1, X_2, \ldots, X_t, Y_1, Y_2, \ldots, Y_t, Z_1, Z_2, \ldots Z_t) \xleftarrow[wor]{\$} S'$$

be a WOR sample of size $3t$ drawn from $S'^{(3)}$. Then for constants $\lambda_1, \lambda_2, \ldots, \lambda_{2t}$ in $\{0,1\}^n$,

$$\Pr\left[(X_1 \oplus Y_1 = \lambda_1) \wedge (X_2 \oplus Y_2 = \lambda_2) \wedge \ldots \wedge (X_t \oplus Y_t = \lambda_t)\right] \geq \frac{1}{2^n}\left(1 - \frac{t \cdot p'^2}{(2^n - p')^2}\right), \tag{5}$$

by eqn.(3), and

$$\Pr\left[\begin{pmatrix} X_1 \oplus Y_1 = \lambda_1, \\ Z_1 \oplus Y_1 = \lambda_2 \end{pmatrix} \wedge \begin{pmatrix} X_2 \oplus Y_2 = \lambda_3, \\ Z_2 \oplus Y_2 = \lambda_4 \end{pmatrix} \wedge \ldots \wedge \begin{pmatrix} X_t \oplus Y_t = \lambda_{2t-1}, \\ Z_t \oplus Y_t = \lambda_{2t} \end{pmatrix}\right] \geq \frac{1}{2^{2nt}}\left(1 - \frac{3t \cdot 2^n \cdot p'^2}{(2^n - p')^3}\right), \tag{6}$$

by eqn.(4) of lemma 4.

Proof. Observe that by eqn.(3),

$$\Pr\left[(X_1 \oplus Y_1 = \lambda_1) \wedge (X_2 \oplus Y_2 = \lambda_2) \wedge \ldots \wedge (X_t \oplus Y_t = \lambda_t)\right]$$

$$\begin{aligned}
&= \Pr\left[X_1 \oplus Y_1 = \lambda_1 \;\middle|\; \begin{smallmatrix} X_1, Y_1 \in S' \\ \text{are distinct} \end{smallmatrix}\right] \times \cdots \times \Pr\left[X_t \oplus Y_t = \lambda_t \;\middle|\; \begin{smallmatrix} X_t, Y_t \in S' \\ \setminus \{X_1, \ldots, X_{t-1}, Y_1, \ldots, Y_{t-1}\} \\ \text{are distinct} \end{smallmatrix}\right] \\
&\geq \frac{1}{2^n}\left(1 - \epsilon_1(p')\right) \times \frac{1}{2^n}\left(1 - \epsilon_1(p'-2)\right) \cdots \times \frac{1}{2^n}\left(1 - \epsilon_1(p'-(2t-2))\right) \\
&\geq \prod_{i=1}^{t} \frac{1}{2^n}\left(1 - \epsilon_1(p')\right) \geq \frac{1}{2^{nt}} \sum_{i=1}^{t} \frac{1}{2^n}\left(1 - \epsilon_1(p')\right) \geq \frac{1}{2^{nt}}\left(1 - \frac{tp'^2}{\left(2^n - p'\right)^2}\right).
\end{aligned}$$

Similarly, by eqn.(4),

$$\Pr\left[\left(\begin{smallmatrix} X_1 \oplus Y_1 = \lambda_1, \\ Z_1 \oplus Y_1 = \lambda_2 \end{smallmatrix}\right) \wedge \left(\begin{smallmatrix} X_2 \oplus Y_2 = \lambda_3, \\ Z_2 \oplus Y_2 = \lambda_4 \end{smallmatrix}\right) \wedge \ldots \wedge \left(\begin{smallmatrix} X_t \oplus Y_t = \lambda_{2t-1}, \\ Z_t \oplus Y_t = \lambda_{2t} \end{smallmatrix}\right)\right]$$

$$\begin{aligned}
&= \Pr\left[\begin{smallmatrix} X_1 \oplus Y_1 = \lambda_1, \\ Z_1 \oplus Y_1 = \lambda_2 \end{smallmatrix} \;\middle|\; \begin{smallmatrix} X_1, Y_1, Z_1 \in S' \\ \text{are distinct} \end{smallmatrix}\right] \\
&\times \Pr\left[\begin{smallmatrix} X_2 \oplus Y_2 = \lambda_3, \\ Z_2 \oplus Y_2 = \lambda_4 \end{smallmatrix} \;\middle|\; \begin{smallmatrix} X_2, Y_2, Z_2 \in S' \setminus \{X_1, Y_1, Z_1\} \\ \text{are distinct} \end{smallmatrix}\right] \\
&\vdots \\
&\times \Pr\left[\begin{smallmatrix} X_t \oplus Y_t = \lambda_{2t-1}, \\ Z_t \oplus Y_t = \lambda_{2t} \end{smallmatrix} \;\middle|\; \begin{smallmatrix} X_t, Y_t, Z_t \in S' \setminus \{X_1, \ldots, X_{t-1}, Y_1, \ldots, Y_{t-1}, Z_1, \ldots, Z_{t-1}\} \\ \text{are distinct} \end{smallmatrix}\right] \\
&\geq \frac{1}{2^{2n}}\left(1 - \epsilon_2(p')\right) \times \frac{1}{2^{2n}}\left(1 - \epsilon_2(p'-3)\right) \cdots \times \frac{1}{2^{2n}}\left(1 - \epsilon_2(p'-3(t-1))\right) \\
&\geq \prod_{i=1}^{t} \frac{1}{2^{2n}}\left(1 - \epsilon_2(p')\right) \geq \frac{1}{2^{2nt}}\left(1 - \sum_{i=1}^{t} \epsilon_2(p')\right) \geq \frac{1}{2^{2nt}}\left(1 - \frac{3t \cdot 2^n \cdot p'^2}{(2^n - p')^3}\right).
\end{aligned}$$

□

The following bound on probability of a valid solution for a combination of a system of equations and a system of non-equations can also be obtained from Eqn.s 3 and 4-

Corollary 2. *Let $\mathcal{G}_{\text{eq,neq}} = (\mathsf{V}, \mathsf{E}_{\text{eq}} \sqcup \mathsf{E}_{\text{neq}}, \mathcal{L})$ be an equations-and-non-equations-inducing graph such that the subgraph $\mathcal{G}_{\text{eq}}$ only has components of size 2 or 3. If $|\mathsf{V} \setminus \mathsf{V}_{\text{eq}}| = q_v$ and λ_i $(i \in [q_m])$ are edge-labels of the edges in E_{eq} in the same order as the components, then the probability of the induced systems of equations and non-equations attaining any solution from a set $S' \subseteq \{0,1\}^n$ of size $(2^n - p')$ for all the variables represented only by the vertices in V_{eq} is bounded by-*

$$\frac{1}{2^{nq_m}}\left(1 - \frac{1200q_m^3 + 312(p' + 3q_v)q_m^2 + 2(p' + 3q_v)^2 q_m}{2^{2n}}\right)\left(1 - \frac{q_v}{2^n}\right). \tag{7}$$

Proof. Suppose $\mathcal{G}_{\text{eq}}^{\tau}$ has exactly $q_m - t$ components with-

1. t components $(X_i, Y_i, Z_i)_{i=1}^{t}$ of size 3 and
2. $q_m - 2t$ components $(X_i, Y_i)_{i=t+1}^{q_m - 2t}$ of size 2.

. Let $w_{i,j}$ be the number of edges in E_{neq} that connect one vertex of the i^{th} component of $\mathcal{G}_{\text{eq}}^{\tau}$ to one vertex of its j^{th} component. Also let $w(v)$ be the number of edges in E_{neq} from

some vertex in $\mathsf{V} \setminus \mathsf{V}_{\text{eq}}$ incident on a vertex $v \in \mathsf{V}_{\text{eq}}$. The number of solutions for all the variables represented by vertices in V_{eq} can then be computed as-

$$
\begin{aligned}
&\Pr\left[\begin{pmatrix} X_1 \oplus Y_1 = \lambda_1, \\ Z_1 \oplus Y_1 = \lambda_2 \end{pmatrix}, \begin{pmatrix} X_1 \oplus Y_1 = \lambda_2, \\ Z_1 \oplus Y_1 = \lambda_4 \end{pmatrix}, \ldots, \begin{pmatrix} X_t \oplus Y_t = \lambda_{2t-1}, \\ Z_t \oplus Y_t = \lambda_{2t} \end{pmatrix}, \begin{matrix} (X_{t+1} \oplus Y_{t+1} = \lambda_{2t+1}), (X_{t+2} \oplus Y_{t+2} = \lambda_{t+2}), \\ \ldots, (X_{q_m - 2t} \oplus Y_{q_m - 2t} = \lambda_{q_m}) \end{matrix}\right] \\
&= \Pr\left[\begin{matrix} X_1 \oplus Y_1 = \lambda_1, \\ Z_1 \oplus Y_1 = \lambda_2 \end{matrix} \;\middle|\; \begin{matrix} X_1, Y_1, Z_1 \in S' \\ \text{are distinct} \end{matrix}\right] \\
&\times \Pr\left[\begin{matrix} X_2 \oplus Y_2 = \lambda_3, \\ Z_2 \oplus Y_2 = \lambda_4 \end{matrix} \;\middle|\; \begin{matrix} X_2, Y_2, Z_2 \in S' \setminus \{X_1, Y_1, Z_1\} \\ \text{are distinct} \end{matrix}\right] \\
&\vdots \\
&\times \Pr\left[\begin{matrix} X_t \oplus Y_t = \lambda_{2t-1}, \\ Z_t \oplus Y_t = \lambda_{2t_1} \end{matrix} \;\middle|\; \begin{matrix} X_t, Y_t, Z_t \in S' \setminus \{X_1, \ldots, X_{t-1}, Y_1, \ldots, Y_{t-1}, Z_1, \ldots, Z_{t-1}\} \\ \text{are distinct} \end{matrix}\right] \\
&\times \Pr\left[X_{t+1} \oplus Y_{t+1} = \lambda_{2t+1} \;\middle|\; \begin{matrix} X_{t+1}, Y_{t+1} \in S' \\ \setminus \{X_1, \ldots, X_t, Y_1, \ldots, Y_t, Z_1, \ldots, Z_t\} \\ \text{are distinct} \end{matrix}\right] \\
&\vdots \\
&\times \Pr\left[X_{q_m - t} \oplus Y_{q_m - t} = \lambda_{q_m} \;\middle|\; \begin{matrix} X_{q_m - t}, Y_{q_m - t} \in S' \\ \setminus \{X_1, \ldots, X_{q_m - t}, Y_1, \ldots, Y_{q_m - t}, Z_1, \ldots, Z_{q_m - t}\} \\ \text{are distinct} \end{matrix}\right].
\end{aligned}
$$

The vertices in $\mathcal{G}_{\text{eq,neq}}$ representing X_1, Y_1 and Z_1 can be chosen after removing one value from S' for each non-equation edge joining one of these vertices to some other vertex of $\mathcal{G}_{\text{eq,neq}}$. Thus, the choice for their values must be made from a set of size $p' + w(X_1) + w(Y_1) + w(Z_1)$. Next, X_2, Y_2 and Z_2 can be chosen only after all the previously assigned values, all values conflicting with any non-equation edges connecting X_2, Y_2 and Z_2 to some vertex in $\mathsf{V} \setminus \mathsf{V}_{\text{eq}}$ and all values conflicting with any non-equation edges joining some vertex of the first component (i.e. X_1, Y_1 or Z_1) with the second component (i.e. X_2, Y_2 or Z_2) are removed from the set S'. This leaves a set of size no less than $p'+3+w(X_2)+w(Y_2)+w(Z_2)+w_{1,2}$. Similar calculations for the remaining components give the following lower bound for $\Pr\left[\begin{pmatrix} X_1 \oplus Y_1 = \lambda_1, Z_1 \oplus Y_1 = \lambda_2, \ldots, \\ X_t \oplus Y_t = \lambda_{2t-1}, Z_t \oplus Y_t = \lambda_{2t} \end{pmatrix}, \begin{pmatrix} X_{t+1} \oplus Y_{t+1} = \lambda_{2t+1}, \ldots, \\ X_{q_m - 2t} \oplus Y_{q_m - 2t} = \lambda_{q_m} \end{pmatrix}\right]$:

$$
\begin{aligned}
& \frac{1}{2^{2n}} \left(1 - \varepsilon_2 \left(p' + w(X_1) + w(Y_1) + w(Z_1)\right)\right) \\
\times \quad & \frac{1}{2^{2n}} \left(1 - \varepsilon_2 \left(p' + 3 + w(X_2) + w(Y_2) + w(Z_2) + w_{1,2}\right)\right) \\
& \vdots \\
\times \quad & \frac{1}{2^{2n}} \left(1 - \varepsilon_2 \left(p' + 3(t-1) + w(X_t) + w(Y_t) + w(Z_t) + \sum_{j=1}^{t-1} w_{j,t}\right)\right) \\
\times \quad & \frac{1}{2^{n}} \left(1 - \varepsilon_1 \left(p' + 3t + w(X_{t+1}) + w(Y_{t+1}) + \sum_{j=1}^{t} w_{j,t+1}\right)\right) \\
& \vdots \\
\times \quad & \frac{1}{2^{n}} \left(1 - \varepsilon_1 \left(p' + 3t + 2(q_m - t - 1) + w(X_{q_m - t}) + w(Y_{q_m - t}) + \sum_{j=1}^{q_m - t - 1} w_{j,q_m - t}\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \frac{1}{2^{2n}}\left(1-\frac{24(p'+3q_v)^2}{2^{2n}}\right)\\
&\times \frac{1}{2^{2n}}\left(1-\frac{24(p'+3+3q_v+9)^2}{2^{2n}}\right)\\
&\vdots\\
&\times \frac{1}{2^{2n}}\left(1-\frac{24(p'+3(t-1)+3q_v+9(t-1))^2}{2^{2n}}\right)\\
&\times \frac{1}{2^{n}}\left(1-\frac{4(p'+3t+2q_v+6t)^2}{2^{2n}}\right)\\
&\vdots\\
&\times \frac{1}{2^{n}}\left(1-\frac{4(p'+3t+2(q_m-t-1)+2q_v+6t+4(q_m-t-1))^2}{2^{2n}}\right)\\
&\geq \frac{1}{2^{2nt}}\left(1-\frac{24}{2^{2n}}\sum_{i=0}^{t-1}(p'+12i+3q_v)^2\right)\\
&\times \frac{1}{2^{n(q_m-2t)}}\left(1-\frac{4}{2^{2n}}\sum_{i=t}^{q_m-t}(p'+7t+6i+2q_v)^2\right)\\
&\geq \frac{1}{2^{2nt}}\left(1-\frac{24}{2^{2n}}\left(48q_m^3+12(p'+3q_v)q_m^2+(p'+3q_v)^2q_m\right)\right)\\
&\times \frac{1}{2^{n(q_m-2t)}}\left(1-\frac{4}{2^{2n}}\left(12q_m^3+6(p'+2q_v)q_m^2+(p'+2q_v)^2q_m\right)\right), \quad \text{since } t\leq q_m.
\end{aligned}
$$

Next, observe that the only vertices in V that remain after this computation are those connected by edges in E_{neq}. The number of valid solutions for these vertices is minimum when they form a single component. Since there can be at most $2q'_v \leq 2q_v$ vertices in $\mathsf{V}\setminus\mathsf{V}_{\text{eq}}$, the lower bound for the probability of any combination of values represented by these $2q'_v$ vertices is:

$$
\begin{aligned}
&\Pr\Big[(X'_1\oplus X'_2\neq\lambda'_1)\wedge(X'_2\oplus X'_3\neq\lambda'_2)\wedge\\
&\ldots\wedge\left(X'_{2q'_v-1}\oplus X'_{2q'_v}\neq\lambda'_{2q'_v-1}\right)\Big]\\
=\;& 1-\Pr\Big[(X'_1\oplus X'_2=\lambda'_1)\vee(X'_2\oplus X'_3=\lambda'_2)\vee\\
&\ldots\vee\left(X'_{2q'_v-1}\oplus X'_{2q'_v}=\lambda'_{2q'_v-1}\right)\Big]\\
\geq\;& 1-\Big(\Pr\left[X'_1\oplus X'_2=\lambda'_1\right]+\Pr\left[X'_2\oplus X'_3=\lambda'_2\right]+\\
&\ldots+\Pr\left[X'_{2q'_v-1}\oplus X'_{2q'_v}=\lambda'_{2q'_v-1}\right]\Big)
\end{aligned}
$$

$$\begin{aligned} &\geq 1 - \sum_{a=1}^{2q'_v-1} \frac{1}{2^n}\left(1 - \frac{p'^2}{(2^n-p')^2}\right) \\ &= 1 - \frac{(2q'_v-1)(2^n-2p')}{(2^n-p')^2} \\ &\geq 1 - \frac{q'_v}{2^n}, \text{ since } 2p' \leq 2^n/2. \end{aligned}$$

Since $q'_v \leq q_v$, any solution to the combined systems of equations and non-equations must therefore have a probability of at least-

$$\frac{1}{2^{nq_m}}\left(1 - \frac{1200q_m^3 + 312(p'+3q_v)q_m^2 + 2(p'+3q_v)^2 q_m}{2^{2n}}\right)\left(1 - \frac{q_v}{2^n}\right).$$

□

4 Related Work

We describe some constructions relevant for our proposals. We have also identified an issue in the cryptanalysis of SoKAC proposed in CRYPTO-19 [CLM19b].

Tweakable Even-Mansour: Even and Mansour pioneered the design and analysis of random permutation-based blockciphers [EM97]. Let π be an ideal n-bit (public) permutation and $K_1, K_2 \in \{0,1\}^n$ be the secret keys. The Even-Mansour construction is defined as follows:

$$\mathsf{EM}_{K_1,K_2}[\pi](x) := \pi(x \oplus K_1) \oplus K_2, \ \forall x \in \{0,1\}^n.$$

When $K_1 = K_2$, we simply write $\mathsf{EM}_{K_1}[\pi]$. In order to incorporate a tweak t in the Even-Mansour construction, Cogliati et al. replace the round keys by some functions $f_i(K_i, t)$ and called it Tweakable Even-Mansour (TEM) construction. This is exactly the spirit of the TWEAKEY framework introduced by Jean et al. [JNP14]. In this paper we consider the following simple instantiation of TEM.

$$\mathsf{TEM}_K[\pi](x,t) := \pi(x \oplus (2^t \cdot K)) \oplus (2^t \cdot K), \ \forall x,t \in \{0,1\}^n.$$

Here, 2 denotes a primitive element in the binary field $\{0,1\}^n$. Other similar known approaches can be found in [LRW02, Rog04, CLS15, Men16] etc.

Davis-Meyer: For a permutation π (public or keyed), Davis Meyer construction is defined as $\mathsf{DM}[\pi](x) := \pi(x) \oplus x$. This method has been popularly adopted to design both hash and PRF from an ideal permutation or cipher. When the permutation π is a blockcipher e_K, we write $\mathsf{DM}_K[e](x) := e_K(x) \oplus x$.

4.1 Some Examples of Permutation-based PRFs

SoEM: Sum of Even-Mansour. It is a permutation based PRF that uses two instances of EM to simply add them up to output the sum. Precisely,

$$\mathsf{SoEM}_{K_1,K_2}[\pi_1,\pi_2](x) := \mathsf{EM}_{K_1}[\pi_1](x) \oplus \mathsf{EM}_{K_2}[\pi_2](x).$$

SOEM has three instances denoted by

- SoEM1 with $\pi_1 = \pi_2$ and K_1, K_2 are independent,

- SoEM21 with π_1, π_2 are independent with $K_1 = K_2$ and
- SoEM22 with π_1, π_2 are independent and K_1, K_2 are independent.

Security of SoEM: Both SoEM1 and SoEM21 achieves the birthday bound security and associated with matching birthday attacks in query complexity. SoEM22 achieves BBB security of $2n/3$-bits with a matching attack in query complexity. Below, we will briefly discuss about the birthday bound attack on SoEM with a single random permutation (i.e, SoEM1). Note that, we use $\mathcal{O}(f(n))$ to denote $c \cdot f(n)$ computations, where c is a small constant. From now on, we use this notation throughout our paper when needed.

Attack Idea: The attack exploits the parallel structure of SoEM as well the usage of the same permutation in both the branches. In other words, if the inputs to the two branches swap then the final outputs will collide. Such a structure of inputs (M, M') can be obtained using $\mathcal{O}(2^{n/2})$ queries by adjusting the left and the right half of the inputs. The condition on the choice of (M, M') is $M \oplus M' = K_1 \oplus K_2$. This condition can be easily detected as the output of the messages M and M' would be same (see Fig.1).

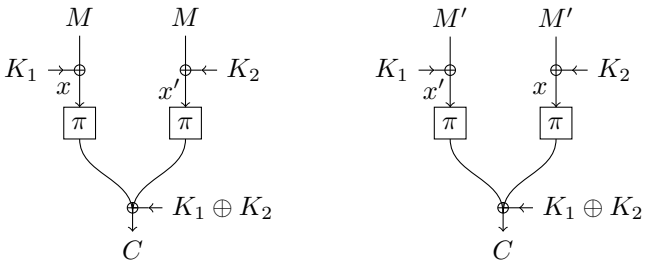

Figure 1: SoEM1 - Swapping x and x'.

SoKAC: These are mainly Even-Mansour followed by Davis-Meyer type of constructions. More precisely,

$$\mathsf{SoKAC1}_{K_1,K_2}[\pi_1](x) := \mathsf{DM}_{K_1}[e](\mathsf{EM}_{K_1,K_2}[\pi_1](x))$$
$$\mathsf{SoKAC21}(x) := \mathsf{DM}_{K}[e'](\mathsf{EM}_{K}[\pi_1](x))$$

where $e_K(x) = \pi_1(x) \oplus K$ and $e'_K(x) = \pi_2(x) \oplus K$.

Proposition 5 in [CLM19b] claims that the same birthday bound attack as on SoEM1 can be applied to SoKAC1. Also, Proposition 6 of the same paper claims that the same beyond birthday bound attack as on SoEM21 can be applied to SoKAC21. We observe that the attacks possibly do not work with the claimed complexities. The main reason is the serial structure of SoKAC, wherein a fresh input to the first permutation π_1 makes the internal state random. Hence, an extended attack on SoKAC is unknown to us. Recently, Nandi proposed a birthday bound attack on SoKAC21 in [Nan20], giving SoKAC21 a birthday bound security; a $2n/3$-bit security was claimed in Theorem 2 of [CLM19b]. Additionally, this paper presents an independent attack against SoKAC1 with $\mathcal{O}(2^{2n/3})$ query complexity in Fig. 3.

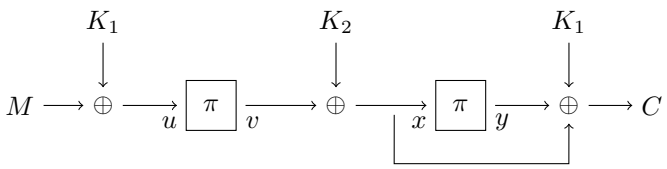

Figure 2: SoKAC1 - One permutation instance π $(= \pi_1 = \pi_2)$, two key instances K_1 and K_2.

$5 \cdot 2^{2n/3}$**-Query Attack on SoKAC1**

1. Queries $M_1, \ldots, M_q \leftarrow \{0,1\}^n$ to the authentication oracle $\mathbf{O}$ with $q = 2 \cdot 2^{2n/3}$ (say $M_i = \langle i \rangle_{2n/3} \| 0^{n/3}$ for $i < 2^{2n/3}$, $M_i = \langle i - 2^{2n/3} + 1 \rangle_{2n/3} \| 1 \| 0^{n/3-1}$ for $2^{2n/3} \leq i < 2 \cdot 2^{2n/3}$).

2. $\tilde{u}_1, \ldots, \tilde{u}_{p_1}$ with $p_1 = 2 \cdot 2^{2n/3}$ forward queries to the primitive π (say $\tilde{u}_a = 0^{n/3} \| \langle a \rangle_{2n/3}$ for $a < 2^{2n/3}$, $\tilde{u}_a = 0^{n/3-1} \| 1 \| \langle a - 2^{2n/3} + 1 \rangle_{2n/3}$ for $2^{2n/3} \leq a < 2 \cdot 2^{2n/3}$); receive responses $\tilde{v}_a = \pi(\tilde{u}_a)$, $a \in [p_1]$.

3. $\tilde{y}_1, \ldots, \tilde{y}_{p_2} \xleftarrow[\text{wor}]{\$} \{0,1\}^n$ with $p_2 = 2 \cdot 2^{2n/3}$ backward primitive queries to the primitive π; receive responses $\tilde{x}_b$, $b \in [p_2]$.

4. Set $\mathrm{Ext}_K := \{(i,a,b) \in [q] \times [p_1] \times [p_2] \mid (M_i \oplus \tilde{u}_a = K_1) \wedge (C_i \oplus \tilde{x}_b \oplus \tilde{y}_b = K_1)\}$ and set $\hat{\mathcal{K}} = \phi$.

5. For all $K \in \mathcal{K}$ with $|\mathrm{Ext}_K| \geq 2$, check whether:
   For all pairs of tuples $(i,a,b) \neq (i',a',b')$ in Ext_K, if $(\tilde{v}_a \oplus \tilde{x}_b \oplus \tilde{v}_{a'} \oplus \tilde{x}_{b'} = 0)$, then add K to $\hat{\mathcal{K}}$.

Figure 3: Interaction of the adversary with $(\mathbf{O}, \pi)$, where $\mathbf{O}$ is either the random oracle or the real construction oracle $\mathrm{SoKAC1}_K^\pi$ and the primitive π.

Analysis of the attack: Observe that for the values $q = p_1 = p_2 = 2 \cdot 2^{2n/3}$, the set Ext_K has size $\mathcal{O}(1)$ with high probability, for each value $K \in \mathcal{K}$. Furthermore, if $K^\star$ denotes the true key of the construction, then $\Pr[K^\star \in \hat{\mathcal{K}}] = \Pr[|\mathrm{Ext}_{K^\star}| \geq 2] \geq \frac{1}{4}$, and thus, the expected size, $\mathsf{E}\left[\left|\hat{\mathcal{K}}\right|\right]$, of the guess-key set $\hat{\mathcal{K}}$ is $\mathcal{O}(1)$.

4.2 Block cipher-based PRFs

EDM: Encrypted Davis Meyer. It encrypts the output of DM. More formally,

$$\mathsf{EDM}_{K_1,K_2}[e] := e_{K_2}(\mathsf{DM}_{K_1}[e](x)).$$

EDM is a $2n/3$-bit BBB secure PRF. The query complexity of the attack against EDM is $\mathcal{O}(2^n)$ query complexity. Hence, the security bound is not tight. Later, Mennink proposed the dual of EDM defined as $\mathsf{DM}_{K_2}[e](e_{K_1}(x))$. This design achieves the same security bound as EDM but the bound is not tight. It has even been proven to be n-bit secure using mirror theory. The proof is not verified and the attack complexity is again up to $\mathcal{O}(2^n)$ queries.
DDM: Decrypted Davis Meyer. DDM optimizes EDM in the number of block cipher instances. In other words DDM replaces the outer e_{K_2} by $e_{K_1}^{-1}$. Formally,

$$\mathsf{DDM}_K[e] := e_K^{-1}(\mathsf{DM}_K(x)).$$

The proven security bound of DDM is exactly the same as EDM. However, this bound is not known to be tight and is accompanied by an attack with $\mathcal{O}(2^n)$ queries.
EWCDM: All the constructions above can handle fixed length inputs. EWCDM [CS16] extends the input domain of EDM to handle multi-block inputs. It takes a nonce $N \in \mathcal{N}$

and an input $x \in \mathcal{M}$ (where $\mathcal{M}$ is the set of all multi-block inputs) to generate a tag $T \in \mathcal{T}$. $\mathsf{EWCDM}_{K_1,K_2,K_h}[e, \mathcal{H}]$ with N and x as the inputs is defined as

$$T = e_{K_2}(e_{K_1}(N) \oplus N \oplus \mathcal{H}_{K_h}(x)).$$

Here, $\mathcal{H}$ is ϵ_1-regular hash, ϵ_2-AXU hash and ϵ_3-3-way regular hash. For Poly hash, we have $\epsilon_1 = \epsilon_2 = \epsilon_3 = \frac{\ell}{2^n}$.

DWCDM: In CRYPTO 2018 [DDNY18b], Datta et al. proposed DWCDM which optimizes EWCDM the number of block cipher instances to one without any compromise in the security level.

It takes a nonce $N \in \mathcal{N}$ and an input $x \in \mathcal{M}$ ($\mathcal{M}$ is the set of all multi-block inputs) to generate a tag $T \in \mathcal{T}$. $\mathsf{DWCDM}_{K,K_h}[e, e^{-1}, \mathcal{H}]$ with N and x is defined as

$$T = e_K^{-1}(e_K(N) \oplus N \oplus \mathcal{H}_{K_h}(x)).$$

Here, the last $n/3$-bits of N are 0 and $\mathcal{H}$ is ϵ_1-regular hash, ϵ_2-AXU hash and ϵ_3-3-way regular hash. For Poly hash, we have $\epsilon_1 = \epsilon_2 = \epsilon_3 = \frac{\ell}{2^n}$.

5 PDMMAC and PDM*MAC Constructions

5.1 Specification and Security of PDMMAC

Specification of PDMMAC: Let $K \xleftarrow{\$} \{0,1\}^n$ and $\pi \xleftarrow{\$} \mathrm{Perm}(n)$. The PRF that we propose in this paper is a construction that takes a message $M \in \{0,1\}^n$ as an input and return n-bit tag $T := \mathsf{PDMMAC}_K^{\pi}(M)$. The construction PDMMAC is defined as

$$T = \pi^{-1}\left(\pi(K \oplus M) \oplus 3K \oplus M\right) \oplus 2K. \tag{8}$$

Design Rationale: Our design PDMMAC is motivated by DDM. Let

$$\mathsf{TEM}_K(t, M) = \pi(M \oplus 2^t \cdot K) \oplus 2^t \cdot K$$

be a specific instantiation of tweakable Even-Mansour construction. The construction PDMMAC can be equivalently described as (see Fig.4)

$$T = \mathsf{TEM}_K^{-1}\left(1, \mathsf{TEM}_K(0, M) \oplus M\right). \tag{9}$$

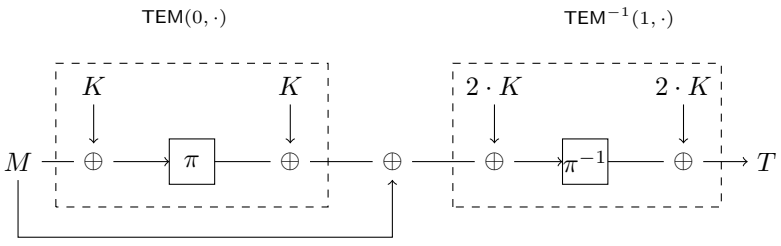

Figure 4: The construction PDMMAC

Security of PDMMAC: We prove that PDMMAC for one instance of uniform π and uniform key K is secure up to attack complexity $\mathcal{O}(2^{2n/3})$. We also propose an attack matching this bound.

Theorem 1. *Let $M \in \mathcal{M}$, and consider $\mathsf{PDMMAC}^{\pi}_{K}$ based on one permutation $\pi \xleftarrow{\$} \mathsf{Perm}(\{0,1\}^n)$ and one key $K \xleftarrow{\$} \{0,1\}^n$. For any distinguisher $\mathcal{D}$ making at most q construction queries at most p primitive queries to $\pi^{\pm}$, we have,*

$$\mathsf{Adv}^{\mathsf{prf}}_{\mathsf{PDMMAC}}(\mathcal{D}) \leq \frac{q^2 + 2q^3 + 3pq^2 + p^2q + 8q(p+q)^2}{2^{2n}} + \frac{6 + q + q\sqrt{3np} + \sqrt{6npq} + p\sqrt{3nq}}{2^n}.$$

The proof for this theorem can be found in Sect. 6. Note that the dominating term of advantage is $\sqrt{\frac{3n(pq^2+qp^2)}{2^{2n}}}$. So the construction is secure as long as $p, q \ll \frac{2^{2n/3}}{n^{1/3}}$.

A Matching Attack with $\mathcal{O}(2^{2n/3})$ Queries: We have a matching attack (up to the logarithmic factor). The attack is similar to that of PDM*MAC, and henceforth omitted. We include the attack for PDM*MAC instead of PDMMAC as it is more robust.

5.2 Specification and Security of PDM*MAC

Specification of PDM*MAC: The previous construction does not allow arbitrary-length messages. We now propose a construction similar to DWCDM, which uses a single ideal permutation $\pi \xleftarrow{\$} \mathrm{Perm}(n)$ and an n-bit key K. To process a message $M \in \{0,1\}^*$, a hash function $\mathcal{H}$ with a key K_h sampled independently of K is required, which is almost xor-universal, regular and 3-way regular. The construction PDM*MAC for an n-bit nonce N and a message $M \in \{0,1\}^*$, with $\mathcal{B} = \{0,1\}^n$ computes $T = \mathsf{PDM}^*\mathsf{MAC}^{\pi}_{K,K_h}(N, M)$ as follows:

$$T = \pi^{-1}\left(\pi(K \oplus N) \oplus 3K \oplus N \oplus \mathcal{H}_{K_h}(M)\right) \oplus 2K. \tag{10}$$

Instance of $\mathcal{H}$: PolyHash [MI11] is an example of a keyed hash which is $\frac{\ell}{2^n}$-regular, AXU and 3-way regular, where ℓ is the maximum number of n-bit blocks. The hash first uses an injective 10^* (one followed by zeros) padding to pad an input $M \in \{0,1\}^*$ to multiple of n-bits. Precisely, $M\|10^j = M_1\|M_2\|\ldots\|M_\ell$ where $j = n - |M| \mod n - 1$. The hash value is generated as

$$\mathsf{Poly}_{\mathcal{H}}(M) = M_\ell \cdot K_h \oplus M_{\ell-1} \cdot K_h^2 \oplus \cdots \oplus M_1 \cdot K_h^\ell.$$

Design Rationale: This construction is motivated by DWCDM. Like PDMMAC, the nonce and the hash of the message are XOR-ed between two permutation calls. Similar designs have been adapted for DWCDM from DDM. The construction PDM*MAC can be equivalently described as (see Fig.5)-

$$T = \mathsf{TEM}_K^{-1}\left(1, \mathsf{TEM}_K(0, N) \oplus N \oplus \mathcal{H}_{K_h}(M)\right). \tag{11}$$

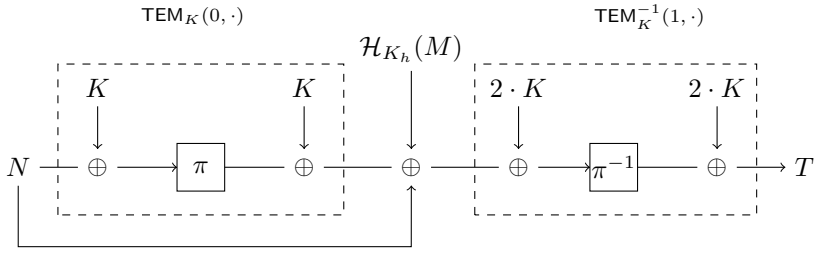

Figure 5: The construction PDM*MAC

SECURITY OF PDM*MAC: We prove the security of PDM*MAC up to an attack complexity of $\mathcal{O}(2^{2n/3})$ for one instance of uniform π and uniform key K. We also propose an attack matching this bound in Fig. 6.

A Matching Attack on PDM*MAC with $\mathcal{O}(2^{2n/3})$ Queries

1. Queries $(N_1, M), \ldots, (N_q, M)$ with $q = 2 \cdot 2^{2n/3}$ to authentication oracle $\mathbf{O}$ (say $N_i = \langle i \rangle_{2n/3} \| 0^{n/3}$ for $i < 2^{2n/3}$, $N_i = \langle i - 2^{2n/3} + 1 \rangle_{2n/3} \| 1 \| 0^{n/3-1}$ for $2^{2n/3} \leq i < 2 \cdot 2^{2n/3}$); receive responses $T_i = \mathbf{O}(M_i)$, $i \in [q]$.

2. $\tilde{u}_1, \ldots, \tilde{u}_{p_1}$ forward queries to the primitive π with $p_1 = 2 \cdot 2^{2n/3}$ (say $\tilde{u}_a = 0^{n/3} \| \langle a \rangle_{2n/3}$ for $a < 2^{2n/3}$, $\tilde{u}_a = 0^{n/3-1} \| 1 \| \langle a - 2^{2n/3} + 1 \rangle_{2n/3}$ for $2^{2n/3} \leq a < 2 \cdot 2^{2n/3}$); receive responses $\tilde{v}_a = \pi(\tilde{u}_a)$, $a \in [p_1]$.

3. $\tilde{y}_1, \ldots, \tilde{y}_{p_2} \xleftarrow[\text{wor}]{\$} \{0,1\}^n$ backward queries to the primitive π with $p_2 = 2 \cdot 2^{2n/3}$; receive responses $\tilde{x}_b$, $b \in [p_2]$.

4. Set $\text{Ext}_K := \{(i, a, b) \in [q] \times [p_1] \times [p_2] \mid (N_i \oplus \tilde{u}_a = K) \wedge (T_i \oplus \tilde{x}_b = 2K)\}$ and set $\hat{\mathcal{K}} = \phi$.

5. For all $K \in \mathcal{K}$ with $|\text{Ext}_K| \geq 2$, carry out the following check:
   For all pairs of tuples $(i, a, b) \neq (i', a', b')$ in Ext_K, if
   $$(N_i \oplus \tilde{v}_a \oplus \tilde{y}_b \oplus N_{i'} \oplus \tilde{v}_{a'} \oplus \tilde{y}_{b'} = 0)\,,$$
   then add K to $\hat{\mathcal{K}}$.

Figure 6: Interaction of the adversary with $(\mathbf{O}, \pi)$, where $\mathbf{O}$ is either the random oracle or the real construction oracle $\mathsf{PDM}^*\mathsf{MAC}^{\pi}_K$ and the primitive π.

Analysis of the attack: Observe that since $I_K := \{(i, a) \mid N_i \oplus \tilde{u}_a = K\}$ has size $\mathcal{O}(2^{n/3})$ for each value $K \in \mathcal{K}$, and for the values $q = p_1 = p_2 = 2 \cdot 2^{2n/3}$, the set Ext_K has size $\mathcal{O}(1)$ with high probability. Furthermore, if $K^\star$ denotes the true key of the construction, then $\Pr[K^\star \in \hat{\mathcal{K}}] = \Pr[|\text{Ext}_{K^\star}| \geq 2] \geq \frac{1}{4}$, and thus, the expected size, $\mathsf{E}\left[\left|\hat{\mathcal{K}}\right|\right]$, of the guess-key set $\hat{\mathcal{K}}$ is $\mathcal{O}(1)$.

Theorem 2. *Let $n \in \mathcal{N}$, and consider $\mathsf{PDM}^*\mathsf{MAC}^{\pi}_{K,K_h}$ based on one permutation $\pi \xleftarrow{\$} \mathsf{Perm}(\{0,1\}^n)$, one key $K \xleftarrow{\$} \{0,1\}^n$ and one hash key $K_h \xleftarrow{\$} \{0,1\}^n$. For any distinguisher $\mathcal{D}$ making at most q_m construction queries, at most p primitive queries to $\pi^{\pm}$ and at most q_v queries to the verification oracle, we have,*

$$\mathsf{Adv}^{\mathsf{MAC}}_{\mathsf{PDM}^*\mathsf{MAC}}(\mathcal{D}) \leq q_v\epsilon + \frac{q_m^2(1 + 1202q_m + 3p + 312(p + q_m + 3q_v)) + p^2(q_m + q_v)}{2^{2n}} + \frac{2(p + q_m + 3q_v)^2 q_m}{2^{2n}} + \frac{6 + 2q_m^2\epsilon + q_m + \sqrt{6npq_m} + q_m\sqrt{3np} + p\sqrt{3nq_m} + 3q_m^2 q_v\epsilon + q_v}{2^n}.$$

The proof for this theorem can be found in Sect. 7. If we assume $\epsilon \approx 2^{-n}$, the dominating term of advantage is $\sqrt{\frac{3n(pq_m^2 + q_m p^2)}{2^{2n}}}$. So the construction is secure as long as $p, q \ll \frac{2^{2n/3}}{n^{1/3}}$.

5.3 Single Keyed Version of PDM*MAC: 1K-PDM*MAC

The PDM*MAC construction calls one permutation, one key K associated with the permutation and one independent hash key K_h. We extend the specification of PDM*MAC to a single keyed version denoted by 1K-PDM*MAC. We use the technique of instantiating the hash key K_h by $K_h = \pi(K)$. We also assume that $N \neq 0$ and $\mathcal{H}$ is Poly hash. However, this technique is similar to that used in DWCDM (where $K_h = E_K(0)$). We prove that 1K-PDM*MAC for one instance of uniform π and uniform key K is secure up to attack complexity $\mathcal{O}(2^{2n/3})$.

Theorem 3. *Let $n \in \mathcal{N}$, and consider 1K-PDM*MAC$^{\pi}_{K}$ based on one permutation $\pi \xleftarrow{\$} \mathsf{Perm}(\{0,1\}^n)$, one key $K \xleftarrow{\$} \{0,1\}^n$. For any distinguisher $\mathcal{D}$ making at most q_m construction queries, at most p primitive queries to $\pi^{\pm}$ and at most q_v queries to the verification oracle, we have,*

$$\mathsf{Adv}^{\mathsf{MAC}}_{1K-PDM^*MAC}(\mathcal{D}) \leq q_v\epsilon + \frac{q_m^2(1+1202q_m+3p+312(p+q_m+3q_v))+p^2(q_m+q_v)}{2^{2n}} + \frac{2(p+q_m+3q_v)^2q_m}{2^{2n}} + \frac{6+q_m^2\epsilon(2+3q_v)+3q_m+2p+\sqrt{6npq_m}+q_m\sqrt{3np}+p\sqrt{3nq_m}+q_v}{2^n}.$$

The proof for this theorem can be found in Sect. 8.

6 Proof of Theorem 1

We use Coefficient-H technique [Pat91, Vau03] (described in Sect. 2.5) to prove the theorem. The details are given below.

Game Description

We denote by q, the number of queries that $\mathcal{D}$ makes to one of the construction oracles $\mathsf{PDMMAC}^{\pi}_{K}$ or φ, the queries being summarized by the transcript $\tau_q = \{(M_1, T_1), \ldots, (M_q, T_q)\}$. $\mathcal{D}$ also makes p queries to the primitive π, which are summarized by $\tau_p = \{(\tilde{u}_1, \tilde{v}_1), \ldots, (\tilde{u}_p, \tilde{v}_p)\}$. It may be assumed without loss of generality that both τ_q and τ_p have distinct elements.

After $\mathcal{D}$ has interacted with the oracles but before it has output its decision, the key K is also revealed to it. In the real world, this is the key used in the construction, while in the ideal world, it is a dummy value drawn uniformly at random from $\{0,1\}^n$. The full transcript of the interaction is denoted by $\tau = (\tau_q, \tau_p, K)$. The set of all attainable transcripts is denoted by $\mathcal{T}$, and we partition $\mathcal{T}$ as $\mathcal{T}_{\text{good}} \sqcup \mathcal{T}_{\text{bad}}$, as described shortly. We let X_{re} be the random variable that takes values $\tau \in \mathcal{T}$ when $\mathcal{D}$ interacts with the real world and X_{id} to be the random variable that takes values $\tau \in \mathcal{T}$ when $\mathcal{D}$ interacts with the ideal world.

Transcript Equations Induced by the Distinguishing Game

This distinguishing game results in a system of equations obtained through the queries to the construction and primitive oracles. These are of the form-

Construction equations:

$$\begin{aligned} \pi(M_1 \oplus K) \oplus \pi(T_1 \oplus 2K) &= 3K \oplus M_1 \\ &\vdots \\ \pi(M_q \oplus K) \oplus \pi(T_q \oplus 2K) &= 3K \oplus M_q \end{aligned}$$

Queries to primitive π:

$$\begin{aligned} \pi(\tilde{u}_1) &= \tilde{v}_1 \\ &\vdots \\ \pi(\tilde{u}_p) &= \tilde{v}_p \end{aligned}$$

Furthermore, these equations can be expressed graphically as described in the *supplementary material*.

6.1 Bad Events

A transcript $\tau = (\tau_q, \tau_p, K)$ is said to be in $\mathcal{T}_{\text{bad}}$ and is called a **bad transcript** if and only if at least one of the following is satisfied-

Collision amongst two construction queries-

B1. There exist $i \neq j \in [q]$ such that $(T_i \oplus M_j = 3K) \wedge (T_j \oplus M_i = 3K)$.

Collision within one construction query-

B2. There exists $i \in [q]$ such that $T_i \oplus M_i = 3K$.

Collision amongst three construction queries-

B3. There exist $i, j, k \in [q]$ such that $T_i \oplus M_j = T_j \oplus M_k = 3K$.

B4. There exist $i, j, k \in [q]$ such that $T_i = T_j = T_k$.

B5. There exist $i, j, k \in [q]$ such that $T_i = T_j = M_k \oplus 3K$.

Collision amongst two construction queries and one primitive query-

B6. There exist $i \neq j \in [q], k \in [p]$ such that $(M_i \oplus T_j = 3K) \wedge (2K \oplus T_i = \tilde{u}_k)$.

B7. There exist $i \neq j \in [q], k \in [p]$ such that $(M_i \oplus T_j = 3K) \wedge (K \oplus M_j = \tilde{u}_k)$.

Collision amongst one construction queries and two primitive queries-

B8. There exist $i \in [q], j, k \in [p]$ such that $(K \oplus M_i = \tilde{u}_k) \wedge (2K \oplus T_i = \tilde{u}_j)$.

Any transcript $\tau \in \mathcal{T}_{\text{good}} = \mathcal{T} \setminus \mathcal{T}_{\text{bad}}$ is said to be a **good transcript**. A figurative and graphical description of the bad events is given in Fig.7. In addition, a circled vertex in any graph describing a bad event denotes a collision with a primitive query.

6.1.1 Probability of Bad Transcripts

$$\text{Now, } \Pr[\tau \in \mathcal{T}_{\text{bad}}] \leq \sum_{i=1}^{8} \Pr[\text{B}i].$$

Probability of events B1, B2, B4 and B5. Consider event B1. Since there are q construction queries (with randomness only in T_i and T_j, but not in M_i and M_j), $\Pr[\text{B1}] \leq \frac{q^2}{2^{2n}}$, $\Pr[\text{B2}] \leq \frac{q}{2^n}$, $\Pr[\text{B4}] \leq \frac{q^3}{2^{2n}}$ and $\Pr[\text{B5}] \leq \frac{q^3}{2^{2n}}$.

Probability of event B3. Let A_3 be any constant value. Define $\Omega_3 = \{(j, i, k) | T_j \oplus M_j = T_i \oplus M_k\}$. Then-

$$\begin{aligned}
\Pr[\text{B3}] &= \Pr\left[(T_j \oplus M_j = T_i \oplus M_k) \wedge (3K = T_j \oplus M_k)\right] \\
&\leq \Pr\left[(3K = T_j \oplus M_k) \wedge (|\Omega_3| \geq A_3)\right] + \Pr\left[(3K = T_j \oplus M_k) \wedge (|\Omega_3| \leq A_3)\right] \\
&\leq \Pr[|\Omega_3| \geq A_3] \cdot \frac{1}{2^n} + A_3 \cdot \frac{1}{2^n}.
\end{aligned}$$

If $A_3 = \frac{pq^2}{2^n} + \sqrt{\frac{6npq}{2^n}}$, then by Lemma 2 of Sect. 2.6,

$$\Pr[\text{B7}] \leq \frac{pq^2}{2^{2n}} + \frac{\sqrt{6npq}}{2^n} + \frac{2}{2^n}.$$

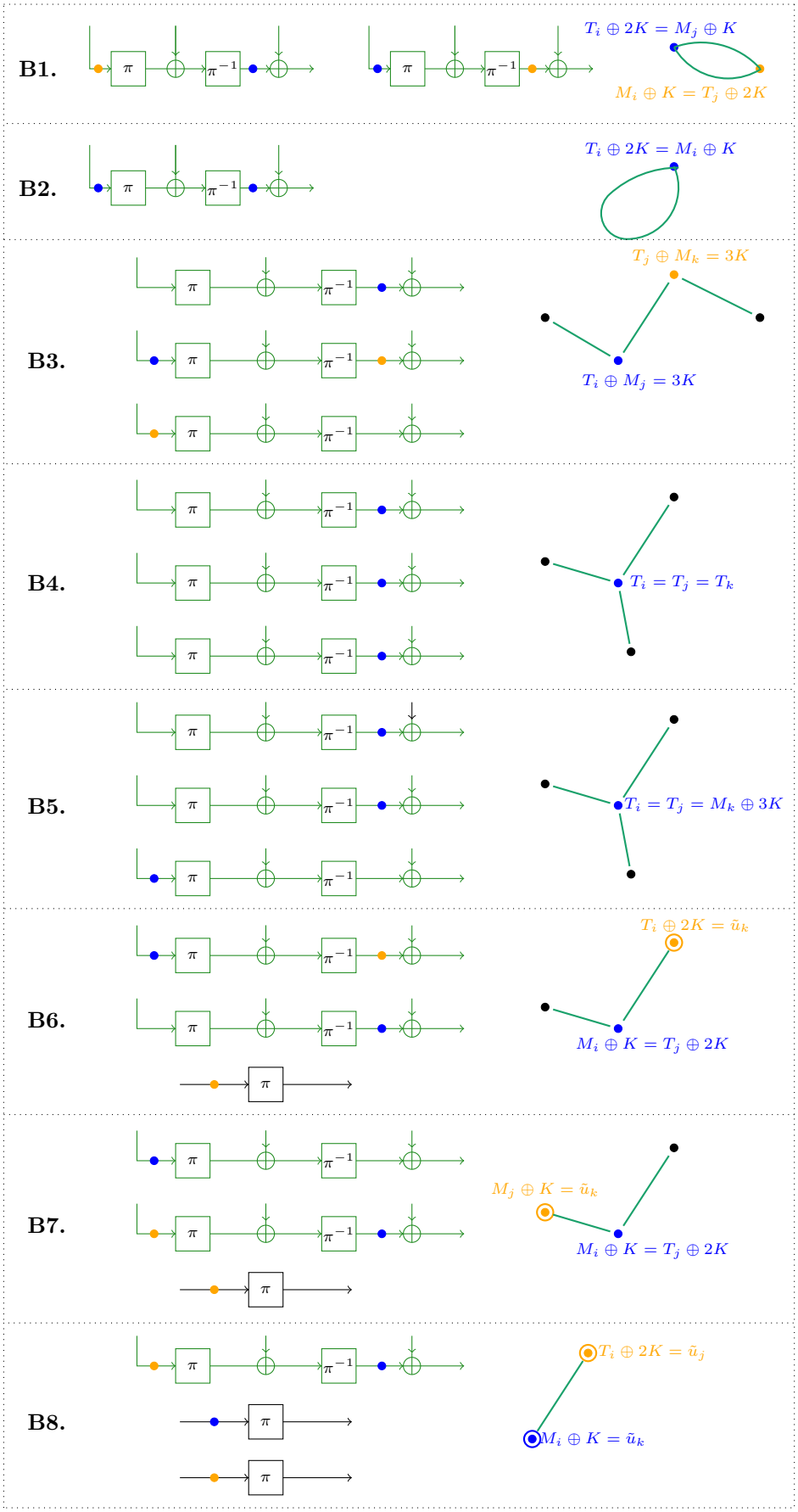

Figure 7: Collisions amongst construction equations and/or primitive queries - figurative and graphical representations of the bad events.

Probability of event B6. Since there are q construction queries and p queries to the primitive, $\Pr[\text{B6}] \leq \frac{pq^2}{2^{2n}}$.

Probability of event B7. Let A_7 be any constant value. Define $\Omega_7 = \{(j,i,k) | T_j \oplus 3M_j = M_i \oplus 3\tilde{u}_k\}$. Then-

$$\begin{aligned}\Pr[\text{B7}] &= \Pr\left[(T_j \oplus 3M_j = M_i \oplus 3\tilde{u}_k) \wedge (3K = T_j \oplus M_i)\right] \\ &\leq \Pr\left[(3K = T_j \oplus M_i) \wedge (|\Omega_7| \geq A_7)\right] + \Pr\left[(3K = T_j \oplus M_i) \wedge (|\Omega_7| \leq A_7)\right] \\ &\leq \Pr[|\Omega_7| \geq A_7] \cdot \frac{1}{2^n} + A_7 \cdot \frac{1}{2^n}.\end{aligned}$$

If $A_7 = \frac{pq^2}{2^n} + q\sqrt{3np}$, then by Lemma 1 of Sect. 2.6,

$$\Pr[\text{B7}] \leq \frac{pq^2}{2^{2n}} + \frac{q\sqrt{3np}}{2^n} + \frac{2}{2^n}.$$

Probability of event B8. Let A_8 be any constant value. Define $\Omega_8 = \{(i,k,j) | 2M_i \oplus T_i = 2\tilde{u}_k \oplus \tilde{u}_j\}$. Then-

$$\begin{aligned}\Pr[\text{B8}] &= \Pr\left[(2M_i \oplus T_i = 2\tilde{u}_k \oplus \tilde{u}_j) \wedge (2K = T_i \oplus \tilde{u}_j)\right] \\ &\leq \Pr\left[(2K = T_i \oplus \tilde{u}_j) \wedge (|\Omega_8| \geq A_8)\right] + \Pr\left[(2K = T_i \oplus \tilde{u}_j) \wedge (|\Omega_8| \leq A_8)\right] \\ &\leq \Pr[|\Omega_8| \geq A_8] \cdot \frac{1}{2^n} + A_8 \cdot \frac{1}{2^n}.\end{aligned}$$

If $A_8 = \frac{p^2q}{2^n} + p\sqrt{3nq}$, then by Lemma 1 of Sect. 2.6,

$$\Pr[\text{B8}] \leq \frac{p^2q}{2^{2n}} + \frac{p\sqrt{3nq}}{2^n} + \frac{2}{2^n}.$$

Thus,

$$\Pr[\tau \in \mathcal{T}_{\text{bad}}] \leq \frac{q^2 + 2q^3 + 3pq^2 + p^2q}{2^{2n}} + \frac{6 + q + q\sqrt{3np} + \sqrt{6npq} + p\sqrt{3nq}}{2^n}.$$

6.2 Good Transcripts

Observe that any good transcript $\tau \in \mathcal{T}_{\text{good}}$ must necessarily be induced by a graph $\mathcal{G}^{\tau}_{\text{eq}}$, which satisfies the following conditions:

- There is no cycle in $\mathcal{G}^{\tau}_{\text{eq}} = (\mathsf{V}_{\text{eq}}, \mathsf{E}_{\text{eq}}, \mathcal{L}_{\text{eq}})$.
- There is no path P in $\mathcal{G}^{\tau}_{\text{eq}}$ such that $\mathcal{L}_{\text{eq}}(P) := \sum_{e \in P} \mathcal{L}(e) = 0$.

Also, it may perhaps contain some circled vertices (denoting collisions with some permutation queries). In fact, every component of $\mathcal{G}^{\tau}_{\text{eq}}$ has size at most 3, due to the restrictions of bad events **B3**, **B4** and **B5**. Furthermore, no component of $\mathcal{G}^{\tau}_{\text{eq}}$ of size 3 has a circled vertex due to **B6** and **B7**, and components of size 2 may have at most one circled vertex due to **B8**. We first modify the good transcripts so as to make certain that none of the vertices of $\mathcal{G}^{\tau}_{\text{eq}}$ are circled, as follows:

- If there exists $i \in [q]$ and $k \in [p]$ such that $K \oplus M_i = \tilde{u}_k$, then remove (M_i, T_i) from τ_q and add $(2K \oplus T_i, 3K \oplus M_i \oplus \tilde{v}_k)$ to τ_p.

- If there exists $i \in [q]$ and $j \in [p]$ such that $2K \oplus T_i = \tilde{u}_j$, then remove (M_i, T_i) from τ_q and add $(K \oplus M_i, 3K \oplus M_i \oplus \tilde{v}_j)$ to τ_p.

Denote the new transcript of primitive queries by F, so that $|F| = p' = p+s$ and $q' = q-s$. Let $S' = \{0,1\}^n \setminus \{\tilde{v}_k \,|\, (\tilde{u}_k, \tilde{v}_k) \in F\}$. Denoting $Q = T \oplus 2K$ and $P = M \oplus K$, assume that for a modified good transcript τ, there are t_1 construction equations of the form

$$\begin{aligned} \pi(P_1) \oplus \pi(Q) &= \lambda_1 \\ \pi(P_2) \oplus \pi(Q) &= \lambda_2, \end{aligned}$$

t_2 construction equations of the form

$$\begin{aligned} \pi(P) \oplus \pi(Q_1) &= \lambda_1 \\ \pi(Q_1) \oplus \pi(Q_2) &= \lambda_2, \end{aligned}$$

and $q' - t_1 - t_2$ construction equations of the form $\pi(P) \oplus \pi(Q) = \lambda$.
Let p_{re} be the probability of a modified transcript τ satisfying the system of equations $\pi(M_i \oplus K) \oplus \pi(T_i \oplus 2K) = 3K \oplus M_i$, $i \in [q']$.

6.2.1 Good Transcript Analysis

The probabilities that X_{re} and X_{id} attain a particular value τ can be computed as

$$\begin{aligned} \Pr[\mathsf{X}_{\mathrm{id}} = \tau] &= \frac{1}{2^{nq}} \cdot \frac{1}{(2^n)_p} \cdot \frac{1}{2^n} \text{ and} \\ \Pr[\mathsf{X}_{\mathrm{re}} = \tau] &= p_{\mathrm{re}} \cdot \frac{1}{(2^n)_{p'}} \cdot \frac{1}{2^n}, \end{aligned}$$

where p_{re} can be computed using equations (3) and (4) as follows.

Probability that construction equations are satisfied.

Cases I and II. ($\pi(\mathbf{P_{2i-1}}) \oplus \pi(\mathbf{Q_i}) = \lambda_{\mathbf{2i-1}}, \pi(\mathbf{P_{2i}}) \oplus \pi(\mathbf{Q_i}) = \lambda_{\mathbf{2i}}$ **or** $\pi(\mathbf{P_{2t_1+j}}) \oplus \pi(\mathbf{Q_{t_1+2j}}) = \lambda_{\mathbf{2t_1+2j-1}}, \pi(\mathbf{Q_{t_1+2j-1}}) \oplus \pi(\mathbf{Q_{t_1+2j}}) = \lambda_{\mathbf{2t_1+2j}}$.)
By eqn.(6),

$$\begin{aligned} & \Pr\left[\begin{smallmatrix} \pi(P_1)\oplus\pi(Q_1)=\lambda_1, \pi(P_2)\oplus\pi(Q_1)=\lambda_2, \ldots, \\ \pi(P_{2t_1+t_2})\oplus\pi(Q_{t_1+2t_2-1})=\lambda_{2t_1+2t_2-1}, \pi(Q_{t_1+2t_2-1})\oplus\pi(Q_{t_1+2t_2})=\lambda_{2t_1+2t_2} \end{smallmatrix}\right] \\ \geq\ & \frac{1}{2^{2n(t_1+t_2)}} \left(1 - \frac{3 \cdot q' \cdot 2^n \cdot p'^2}{(2^n - p')^3}\right), \text{ since } t_1 + t_2 \leq q'. \end{aligned}$$

Case III. ($\pi(\mathbf{P_{2t_1+t_2+l}}) \oplus \pi(\mathbf{Q_{t_1+2t_2+l}}) = \lambda_{\mathbf{2t_1+2t_2+l}}$.)
By eqn.(5),

$$\begin{aligned} & \Pr\left[\begin{smallmatrix} \pi(P_{2t_1+t_2+1})\oplus\pi(Q_{t_1+2t_2+1})=\lambda_{2t_1+2t_2+1}, \ldots, \\ \pi(P_{q'-t_2})\oplus\pi(Q_{q'-t_1})=\lambda_{q'} \end{smallmatrix}\right] \\ \geq\ & \frac{1}{2^{n(q'-2t_1-2t_2)}} \left(1 - \frac{q' \cdot p'^2}{(2^n - p')^2}\right), \text{ since } q' - 2t_1 - 2t_2 \leq q'. \end{aligned}$$

$$\begin{aligned}
\text{Thus, } p_{\text{re}} &\geq \frac{1}{2^{nq'}}\left(1-\frac{3\cdot q'\cdot 2^n\cdot p'^2}{(2^n-p')^3}\right)\left(1-\frac{q'\cdot p'^2}{(2^n-p')^2}\right)\\
&\geq \frac{1}{2^{nq'}}\left(1-\frac{6\cdot 2^n\cdot q(p+q)^2}{2^{2n}}\right)\left(1-\frac{2\cdot 2^n\cdot q(p+q)^2}{2^{2n}}\right)\\
&\quad \left(\text{since } q\geq q',\ \frac{2}{2^n}\geq p'\geq p \text{ and } (p+q)\geq p'\right)\\
&\geq \frac{1}{2^{nq'}}\left(1-\frac{8\cdot 2^n\cdot q(p+q)^2}{2^{2n}}\right).
\end{aligned}$$

$$\begin{aligned}
\text{Thus, } \frac{\Pr\left[\mathsf{X}_{\text{re}}\right]}{\Pr\left[\mathsf{X}_{\text{id}}\right]} &\geq \frac{2^{nq}}{2^{nq'}}\cdot\frac{(2^n)_p}{(2^n)_{p'}}\cdot\left(1-\frac{8q(p+q)^2}{2^{2n}}\right)\geq\left(1-\frac{8q(p+q)^2}{2^{2n}}\right),\\
\text{i.e. } \frac{\Pr\left[\mathsf{X}_{\text{re}}\right]}{\Pr\left[\mathsf{X}_{\text{id}}\right]} &\geq (1-\epsilon_{\text{good}}),\ \text{where } \epsilon_{\text{good}}=\frac{8q(p+q)^2}{2^{2n}}.
\end{aligned}$$

7 Proof of Theorem 2

We use Coefficient-H technique [Pat91,Vau03] (described in Sect. 2.5) to prove the theorem.

Forging Game

An upper bound for the nonce-based MAC advantage can be computed by adapting the distinguishing game in Sect. 2.3 (the game is described in Page 5, [DJN17]) as follows. $\mathcal{D}$ makes q_m queries to one of the construction (authentication, or Auth) oracles $\mathsf{PDM}^*\mathsf{MAC}^{\pi}_{K,K_h}$ or φ, the queries being summarized by the authentication transcript

$$\tau_0^m = \{(N_1, M_1, T_1), \ldots, (N_{q_m}, M_{q_m}, T_{q_m})\},$$

and by q_v, the number of verification queries that $\mathcal{D}$ makes to one of the construction (verification, or Ver) oracles $\mathrm{Ver}^{\pi}_{K,K_h}$ or $\perp$, the queries being summarized by the verification transcript $\tau_0^v = \{(N'_1, M'_1, T'_1, b_1), \ldots, (N'_{q_v}, M'_{q_v}, T'_{q_v}, b_{q_v})\}$, where $\forall a, b_a \in \{0,1\}$ are the output values of the verification oracle (in the real world, the oracle checks if $\mathsf{Auth}(N'_a, M'_a) = T'_i$, and returns 1 or 0 according to whether the equality holds or not, respectively, while in the ideal world, $b_a = 0$ for all a). $\mathcal{D}$ also makes p queries to the primitive π, which are summarized by $\tau_p = \{(\tilde{u}_1, \tilde{v}_1), \ldots, (\tilde{u}_p, \tilde{v}_p)\}$. It may be assumed without loss of generality that each of τ_0^m, τ_0^v, τ has distinct elements.

After $\mathcal{D}$ has interacted with the oracles but before it has output its decision, the keys K and K_h are also revealed to it. In the real world, these are the keys used in the construction, while in the ideal world, they are dummy values drawn uniformly at random from $\{0,1\}^n$. The full transcript of the interaction is denoted by $\tau = (\tau_0^m, \tau_0^v, \tau_p, K, K_h)$. The set of all attainable transcripts is denoted by $\mathcal{T}$, and we partition $\mathcal{T}$ as $\mathcal{T}_{\text{good}} \sqcup \mathcal{T}_{\text{bad}}$, as described shortly.

Transcript Equations Induced by the Forging Game: The system of equations has a similar form, and is extended by a system of non-equations, as given below-

Authentication equations:

$$\begin{aligned}
\pi(N_1 \oplus K) \oplus \pi(T_1 \oplus 2K) &= 3K \oplus N_1 \oplus H_1\\
&\vdots\\
\pi(N_{q_m} \oplus K) \oplus \pi(T_{q_m} \oplus 2K) &= 3K \oplus N_{q_m} \oplus H_{q_m}
\end{aligned}$$

Verification non-equations:

$$\begin{aligned} \pi(N'_1 \oplus K) \oplus \pi(T'_1 \oplus 2K) &\neq 3K \oplus N'_1 \oplus H'_1 \\ &\vdots \\ \pi(N'_{q_v} \oplus K) \oplus \pi(T'_{q_v} \oplus 2K) &\neq 3K \oplus N'_{q_v} \oplus H'_{q_v} \end{aligned}$$

Queries to primitive π:

$$\begin{aligned} \pi(\tilde{u}_1) &= \tilde{v_1} \\ &\vdots \\ \pi(\tilde{u}_p) &= \tilde{v_p}, \end{aligned}$$

where $H_i = \mathcal{H}_{K_h}(M_i), \forall i \in [q_m]$ and $H'_j = \mathcal{H}'_{K_h}(M'_j), \forall j \in [q_v]$.

7.1 Bad Events

A transcript $\tau = (\tau_0^m, \tau_0^v, \tau_p, K, K_h)$ is said to be in $\mathcal{T}_{\text{bad}}$ and is called a **bad transcript** if and only if there exists a tuple $(N_i, M_i, T_i) \in \tau_0^m$, $(N'_a, M'_a, T'_a) \in \tau_0^v$ and $(\tilde{u}_j, \tilde{v}_j), (\tilde{x}_k, \tilde{y}_k) \in \tau_p$ such that at least one of the following is satisfied-

Collision amongst two authentication queries-

B1. There exist $i \neq j \in [q_m]$ such that $(T_i = T_j) \wedge (N_i \oplus H_i = N_j \oplus H_j)$.

B2. There exist $i \neq j \in [q_m]$ such that $(T_i \oplus N_j = 3K) \wedge (N_i \oplus H_i = N_j \oplus H_j)$.

B3. There exist $i \neq j \in [q_m]$ such that $(T_i \oplus N_j = 3K) \wedge (T_j \oplus N_i = 3K)$.

Collision within one authentication query-

B4. There exists $i \in [q_m]$ such that $T_i \oplus N_i = 3K$.

Collision amongst three authentication queries-

B5. There exist $i, j, k \in [q_m]$ such that $T_i \oplus N_j = T_j \oplus N_k = 3K$.

B6. There exist $i, j, k \in [q_m]$ such that $T_i = T_j = T_k$.

B7. There exist $i, j, k \in [q_m]$ such that $T_i = T_j = N_k \oplus 3K$.

Collision amongst two authentication queries and one primitive query-

B8. There exist $i \neq j \in [q_m], k \in [p]$ such that $(N_i \oplus T_j = 3K) \wedge (2K \oplus T_i = \tilde{u}_k)$.

B9. There exist $i \neq j \in [q_m], k \in [p]$ such that $(N_i \oplus T_j = 3K) \wedge (K \oplus N_j = \tilde{u}_k)$.

Collision amongst one authentication query and two primitive queries-

B10. There exist $i \in [q_m], j, k \in [p]$ such that $(K \oplus N_i = \tilde{u}_k) \wedge (2K \oplus T_i = \tilde{u}_j)$.

Collision amongst one verification query and two primitive queries-

B11. There exist $a \in [q_v], j, k \in [p]$ such that $(K \oplus N'_a = \tilde{u}_k) \wedge (2K \oplus T'_a = \tilde{u}_j)$.

Collision amongst one authentication and one verification query-

B12. There exist $i \in [q_m], a \in [q_v]$ such that $(N_i = N'_a) \wedge (H_i = H'_a) \wedge (T_i = T'_a)$.

Collision amongst two authentication queries and one verification query, with an extra condition-

B13. There exist $i, j \in [q_m], a \in [q_v]$ such that $(H_i \oplus H_j \oplus H'_a = N_i \oplus N_j \oplus N'_a \oplus 2K)$ and $(N'_a = N_i) \wedge (T_i \oplus N_j = 3K) \wedge (T_j = T'_a)$.

B14. There exist $i, j \in [q_m], a \in [q_v]$ such that $(H_i \oplus H_j \oplus H'_a = N_i \oplus N_j \oplus N'_a \oplus 2K)$ and $(T'_a \oplus N_i = 3K) \wedge (T_i \oplus N_j = 3K) \wedge (T_j \oplus N'_a = 3K)$.

B15. There exist $i, j \in [q_m], a \in [q_v]$ such that $(H_i \oplus H_j \oplus H'_a = N_i \oplus N_j \oplus N'_a \oplus 2K)$ and $(N'_a = N_i) \wedge (T_i = T_j) \wedge (T'_a \oplus N_j = 3K)$.

Any transcript $\tau \in \mathcal{T}_{\text{good}} = \mathcal{T} \setminus \mathcal{T}_{\text{bad}}$ is said to be a **good transcript**. A figurative and graphical description of the bad events is given in the *supplementary material*. In these figures, a circled vertex in any graph describing a bad event denotes a collision with a primitive query.

7.1.1 Probability of Bad Transcripts

$$\text{Now, } \Pr[\tau \in \mathcal{T}_{\text{bad}}] \leq \sum_{i=1}^{15} \Pr[\text{B}i].$$

Probability of events B1, B2, B3, B4, B6 and B7. Consider event B1. Since there are q_m authentication queries (with randomness only in T_i and T_j, but not in N_i and N_j) and since $\mathcal{H}$ is an ϵ-differential hash function, $\Pr[\text{B1}] \leq \frac{q_m^2 \epsilon}{2^n}$. Similarly, $\Pr[\text{B2}] \leq \frac{q_m^2 \epsilon}{2^n}$, $\Pr[\text{B3}] \leq \frac{q_m^2}{2^{2n}}$, $\Pr[\text{B4}] \leq \frac{q_m}{2^n}$, $\Pr[\text{B6}] \leq \frac{q_m^3}{2^{2n}}$ and $\Pr[\text{B7}] \leq \frac{q_m^3}{2^{2n}}$.

Probability of event B5. Let A_5 be any constant value. Define $\Omega_5 = \{(j, i, k) | T_j \oplus N_j = T_i \oplus N_k\}$. Then-

$$\begin{aligned}
\Pr[\text{B5}] &= \Pr\left[(T_j \oplus N_j = T_i \oplus N_k) \wedge (3K = T_j \oplus N_i)\right] \\
&\leq \Pr\left[(3K = T_j \oplus N_i) \wedge (|\Omega_5| \geq A_5)\right] + \Pr\left[(3K = T_j \oplus N_i) \wedge (|\Omega_5| \leq A_5)\right] \\
&\leq \Pr[|\Omega_5| \geq A_5] \cdot \frac{1}{2^n} + A_5 \cdot \frac{1}{2^n}.
\end{aligned}$$

If $A_5 = \frac{pq_m^2}{2^n} + \sqrt{\frac{6npq_m}{2^n}}$, then by Lemma 2 of Sect. 2.6,

$$\Pr[\text{B5}] \leq \frac{pq_m^2}{2^{2n}} + \frac{\sqrt{6npq_m}}{2^n} + \frac{2}{2^n}.$$

Probability of event B8. Since there are q_m authentication queries and p queries to the primitive, $\Pr[\text{B8}] \leq \frac{pq_m^2}{2^{2n}}$.

Probability of event B9. Let A_9 be any constant value. Define $\Omega_9 = \{(j, i, k) | T_j \oplus N_j = T_i \oplus N_k\}$. Then-

$$\begin{aligned}
\Pr[\text{B9}] &= \Pr\left[(T_j \oplus 3N_j = N_i \oplus 3\tilde{u}_k) \wedge (3K = T_j \oplus N_i)\right] \\
&\leq \Pr\left[(3K = T_j \oplus N_i) \wedge (|\Omega_9| \geq A_9)\right] + \Pr\left[(3K = T_j \oplus N_i) \wedge (|\Omega_9| \leq A_9)\right] \\
&\leq \Pr[|\Omega_9| \geq A_9] \cdot \frac{1}{2^n} + A_9 \cdot \frac{1}{2^n}.
\end{aligned}$$

If $A_9 = \frac{pq_m^2}{2^n} + q_m\sqrt{3np}$, then by Lemma 1 of Sect. 2.6,

$$\Pr[\text{B9}] \leq \frac{pq_m^2}{2^{2n}} + \frac{q_m\sqrt{3np}}{2^n} + \frac{2}{2^n}.$$

Probability of event B10. Let A_{10} be any constant value. Define $\Omega_{10} = \{(i,k,j)|2N_i \oplus T_i = 2\tilde{u}_k \oplus \tilde{u}_j\}$. Then-

$$\begin{aligned}\Pr[\text{B10}] &= \Pr\left[(2N_i \oplus T_i = 2\tilde{u}_k \oplus \tilde{u}_j) \wedge (2K = T_i \oplus \tilde{u}_j)\right] \\ &\leq \Pr\left[(2K = T_i \oplus \tilde{u}_j) \wedge (|\Omega_{10}| \geq A_{10})\right] + \Pr\left[(2K = T_i \oplus \tilde{u}_j) \wedge (|\Omega_{10}| \leq A_{10})\right] \\ &\leq \Pr[|\Omega_{10}| \geq A_{10}] \cdot \frac{1}{2^n} + A_{10} \cdot \frac{1}{2^n}.\end{aligned}$$

If $A_{10} = \frac{p^2 q_m}{2^n} + p\sqrt{3nq_m}$, then by Lemma 1 of Sect. 2.6,

$$\Pr[\text{B10}] \leq \frac{p^2 q_m}{2^{2n}} + \frac{p\sqrt{3nq_m}}{2^n} + \frac{2}{2^n}.$$

Probability of event B11. Since there are q_v verification queries and p queries to the primitive, $\Pr[\text{B11}] \leq \frac{p^2 q_v}{2^{2n}}$.

Probability of event B12. Since there are q_v verification queries and H is an ϵ-differential hash function, $\Pr[\text{B12}] \leq q_v\epsilon$.

Probability of events B13, B14 and B15. For all three events, $H_i \oplus H_j \oplus H'_a = (N_i \oplus N_j \oplus N'_a) \oplus 3K$. Since there are q_m authentication queries and q_v verification queries and assuming $\mathcal{H}$ is an ϵ-3-way-regular hash function, Pr[B13], Pr[B14] and Pr[B15] are all at most $\frac{q_m^2 q_v \epsilon}{2^n}$.

Thus,

$$\begin{aligned}\Pr[\tau \in \mathcal{T}_{\text{bad}}] \leq \frac{q_m^2 + 2q_m^3 + 3pq_m^2 + p^2 q_m + p^2 q_v}{2^{2n}} \\ + \frac{2q_m^2\epsilon + q_m + q_m\sqrt{3np} + \sqrt{6npq_m} + p\sqrt{3nq_m} + 6 + 3q_m^2 q_v \epsilon}{2^n} + q_v\epsilon.\end{aligned}$$

7.2 Good Transcripts

Observe that any good transcript $\tau \in \mathcal{T}_{\text{good}}$ must necessarily be induced by a graph $\mathcal{G}^{\tau}_{\text{eq,neq}}$, which satisfies the following conditions:

- There is no cycle of equation-inducing edges in $\mathcal{G}^{\tau}_{\text{eq}} = (\mathsf{V}_{\text{eq}}, \mathsf{E}_{\text{eq}}, \mathcal{L}\big|_{\mathsf{E}_{\text{eq}}})$.
- There is no path P in $\mathcal{G}^{\tau}_{\text{eq}}$ such that $\mathcal{L}(P) := \sum_{e \in P} \mathcal{L}(e) = 0$.
- For all the cycles C in $\mathcal{G}^{\tau}_{\text{eq,neq}}$ whose edge set consists of all but one equation edges $e \in \mathsf{E}_{\text{eq}}$ and exactly one non-equation edge $e' \in \mathsf{E}_{\text{neq}}$, $\mathcal{L}(C) \neq 0$.

It may perhaps contain some circled vertices (denoting collisions with some permutation queries). It shall be assumed that the edges in $\mathsf{E}^{\tau}_{\text{eq}}$ are continuous edges, colored green, and edges in $\mathsf{E}^{\tau}_{\text{neq}}$ are dotted edges, colored red. In fact, every component of $\mathcal{G}^{\tau}_{\text{eq}}$ has size at most 3, due to the restrictions of bad events **B5**, **B6** and **B7**. Furthermore, no component of $\mathcal{G}^{\tau}_{\text{eq}}$ of size 3 has a circled vertex due to **B8** and **B9**, and components of size 2 of $\mathcal{G}^{\tau}_{\text{eq}}$ as well as $\mathcal{G}^{\tau}_{\text{eq,neq}}$ may have at most one circled vertex due to **B10** and **B11**. Finally, the restrictions by bad events **B13**, **B14** and **B15** ensure that $\mathcal{G}^{\tau}_{\text{eq,neq}}$ satisfies the condition $\mathcal{L}(C) \neq 0$ for a cycle containig exactly one non-equation edge.

We first modify the good transcripts in such a way that no vertices remain circled:

- If there exists $i \in [q_m]$ and $k \in [p]$ such that $K \oplus N_i = \tilde{u}_k$, then remove (N_i, M_i, T_i) from τ_0^m and add $(2K \oplus T_i, 3K \oplus N_i \oplus H_i \oplus \tilde{v}_k)$ to τ_p.
- If there exists $i \in [q_m]$ and $j \in [p]$ such that $2K \oplus T_i = \tilde{u}_j$, then remove (N_i, M_i, T_i) from τ_0^m and add $(K \oplus N_i, 3K \oplus N_i \oplus H_i \oplus \tilde{v}_j)$ to τ_p.

Denote the new set of primitive transcripts by F, so that $|F| = p' = p + s$ and $q'_m = q_m - s$. Let $S' \subseteq \{0,1\}^n$ such that $S' = \{0,1\}^n \setminus \{\tilde{v}_k \,|\, (\tilde{u}_k, \tilde{v}_k) \in F\}$.
Let p_{re} be the probability of a modified transcript τ satisfying the system of equations $\pi(N_i \oplus K) \oplus \pi(T_i \oplus 2K) = 3K \oplus N_i \oplus \mathcal{H}_{K_h}(M_i)$, $i \in [q']$.

7.2.1 Good Transcript Analysis

The probabilities of X_{re} and X_{id} attaining a particular value τ can be computed as follows-

$$\begin{aligned}\Pr[\mathsf{X}_{\text{id}} = \tau] &= \frac{1}{2^{nq_m}} \cdot 1 \cdot \frac{1}{(2^n)_p} \cdot \left(\frac{1}{2^n}\right)^2 \text{ and} \\ \Pr[\mathsf{X}_{\text{re}} = \tau] &= p_{\text{re}} \cdot \frac{1}{(2^n)_{p'}} \cdot \left(\frac{1}{2^n}\right)^2.\end{aligned}$$

Probability that authentication equations and verification non-equations are satisfied.

By Corroloary 2,

$$\Pr\left[\begin{array}{c}\Big(\pi(P_1)\oplus\pi(Q_1)=\lambda_1, \pi(P_2)\oplus\pi(Q_1)=\lambda_2, \ldots, \\ \pi(P_{2t_1+t_2})\oplus\pi(Q_{t_1+2t_2-1})=\lambda_{2t_1+2t_2-1}, \pi(Q_{t_1+2t_2-1})\oplus\pi(Q_{t_1+2t_2})=\lambda_{2t_1+2t_2} \\ \pi(P_{2t_1+t_2+1})\oplus\pi(Q_{t_1+2t_2+1})=\lambda_{2t_1+2t_2+1}, \ldots, \\ \pi(P_{q'_m-t_2})\oplus\pi(Q_{q'_m-t_1})=\lambda_{q'_m}\Big) \wedge \\ \Big(\big(\pi(X'_1)\oplus\pi(X'_2)\neq\lambda'_1\big)\wedge\big(\pi(X'_2)\oplus\pi(X'_3)\neq\lambda'_2\big)\wedge\ldots\wedge\big(\pi(X'_{2q'_v-1})\oplus\pi(X'_{2q'_v})\neq\lambda'_{2q'_v-1}\big)\Big)\end{array}\right]$$

$$\leq \frac{1}{2^{nq'_m}}\Big(1 - \frac{1200{q'_m}^3 + 312(p'+3q_v){q'_m}^2 + 2(p'+3q_v)^2 q'_m}{2^{2n}}\Big)\Big(1 - \frac{q_v}{2^n}\Big).$$

Therefore, p_{re} must be at least $\frac{1}{2^{nq_m}}\Big(1 - \frac{1200q_m^3 + 312(p'+3q_v)q_m^2 + 2(p'+3q_v)^2 q_m}{2^{2n}}\Big)\Big(1 - \frac{q_v}{2^n}\Big)$, so that-

$$\begin{aligned}\frac{\Pr[\mathsf{X}_{\text{re}}]}{\Pr[\mathsf{X}_{\text{id}}]} &\geq \frac{2^{nq_m}}{2^{nq'_m}} \cdot \frac{(2^n)_p}{(2^n)_{p'}} \Big(1 - \frac{q_v}{2^n}\Big)\Big(1 - \frac{1200{q'_m}^3 + 312(p'+3q_v){q'_m}^2 + 2(p'+3q_v)^2 q'_m}{2^{2n}}\Big) \\ &\geq \Big(1 - \frac{q_v}{2^n}\Big) \cdot \left(1 - \frac{1200q_m^3 + 312(p+q_m+3q_v)q_m^2 + 2(p+q_m+3q_v)^2 q_m}{2^{2n}}\right), \\ &\quad \text{since } q'_m \leq q_m, p' \leq p + q_m \\ &\geq (1 - \epsilon_{\text{good}}), \text{ where} \\ \epsilon_{\text{good}} &= \frac{q_v}{2^n} + \frac{1200q_m^3 + 312(p+q_m+3q_v)q_m^2 + 2(p+q_m+3q_v)^2 q_m}{2^{2n}}.\end{aligned}$$

8 Proof of Theorem 3

The proof is similar to that of PDM*MAC, except for some extra bad cases. We add the following cases after **B14**. The cases are as follows.

B16. There exists $i \in [q_m]$ such that $T_i = 3K$.

B17. There exists $i \in [q_m]$ such that $\pi(N_i \oplus K) \oplus \mathcal{H}_{K_h}(M_i) \oplus N_i \oplus 3K = K_h$.

B18. There exists $k \in [p]$ such that $\tilde{u}_k = K$.

B19. There exists $k \in [p]$ such that $\tilde{y}_k = K_h$.

Probability of B16. There are q_m authentication queries. Hence, $\Pr[\text{B16}] \leq \frac{q_m}{2^n}$.

Probability of B17. In this case, N_i and M_i are fixed. Thus, $\Pr[\text{B17}] = \Pr[\pi(N_i \oplus K) \oplus 3K = \mathcal{H}_{K_h}(M_i) \oplus N_i \oplus K_h]$. As K and K_h are independently sampled in the ideal world, we obtain $\Pr[\text{B17}] \leq \frac{q_m}{2^n}$, by conditioning H.

Probability of B18 and B19. Since there are p queries to the primitive, $\Pr[\text{B18}], \Pr[\text{B19}] \leq \frac{p}{2^n}$.

Good Transcript Analysis

The good transcript analysis is exactly the same except in this case $\Pr[\mathsf{X}_{\text{re}} = \tau] = p_{\text{re}} \cdot \frac{1}{(2^n)_{p'}} \cdot \left(\frac{1}{2^n}\right)$ (as only the construction key K needs to be sampled, the last term in the expression is $\frac{1}{2^n}$ instead $(\frac{1}{2^n})^2$). However, this does not change the lower bound of $\frac{\Pr[\mathsf{X}_{\text{re}}]}{\Pr[\mathsf{X}_{\text{id}}]}$.

9 Open Problems

Our designs are minimal in structure in the number of permutation and key instances. However, PDMMAC makes two calls to one permutation π, one forward call to π and another inverse call to π^{-1}. We already know that PRFs with one permutation call can not provide more than birthday bound security and hence we need at least two calls to the permutation. Thus, the question

Can we design a BBB secure PRF with one permutation with two forward calls?

remains unanswered and the design of such a construction can be interesting to the community. A possible approach to proceed with this problem is to prove the $2n/3$-bit BBB security of SoKAC1. This design has been mentioned to be at most $n/2$-bit secure [CLM19a] accompanied by a birthday bound attack. However, the attack is possibly wrong and SoKAC1 may provide $2n/3$-bit BBB security.

Acknowledgments

The authors would like to thank Dr. Damian Vizer for his insightful comments and suggestions in preparing the final draft. We would also like to thank all the anonymous reviewers for their valuable comments. Avik Chakraborti, Mridul Nandi and Suprita Talnikar are supported by the project "Study and Analysis of IoT Security" under Government of India at R.C.Bose Centre for Cryptology and Security, Indian Statistical Institute, Kolkata.

References

[ADMA15] Elena Andreeva, Joan Daemen, Bart Mennink, and Gilles Van Assche. Security of keyed sponge constructions using a modular proof approach. In *FSE 2015*, pages 364–384, 2015.

[BCG+12] Julia Borghoff, Anne Canteaut, Tim Güneysu, Elif Bilge Kavun, Miroslav Knezevic, Lars R. Knudsen, Gregor Leander, Ventzislav Nikov, Christof Paar, Christian Rechberger, Peter Rombouts, Søren S. Thomsen, and Tolga Yalçin. PRINCE - A Low-Latency Block Cipher for Pervasive Computing Applications - Extended Abstract. In *ASIACRYPT 2012*, pages 208–225, 2012.

[BDPA11a] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. Duplexing the sponge: single-pass authenticated encryption and other applications. *IACR Cryptology ePrint Archive*, 2011:499, 2011.

[BDPA11b] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. On the Security of the Keyed Sponge Construction. In *Symmetric Key Encryption Workshop*, 2011.

[BDPA15] Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche. Keccak. *IACR Cryptology ePrint Archive*, 2015:389, 2015.

[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*, 1999:24, 1999.

[BJK+16] Christof Beierle, Jérémy Jean, Stefan Kölbl, Gregor Leander, Amir Moradi, Thomas Peyrin, Yu Sasaki, Pascal Sasdrich, and Siang Meng Sim. The SKINNY family of block ciphers and its low-latency variant MANTIS. In *CRYPTO 2016, Part II*, pages 123–153, 2016.

[BKL+07] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J. B. Robshaw, Yannick Seurin, and C. Vikkelsoe. PRESENT: An Ultra-Lightweight Block Cipher. In *CHES 2007*, pages 450–466, 2007.

[BKL+11] Andrey Bogdanov, Miroslav Knezevic, Gregor Leander, Deniz Toz, Kerem Varici, and Ingrid Verbauwhede. spongent: A lightweight hash function. In *CHES 2011*, pages 312–325, 2011.

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

[BKR00] Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of the cipher block chaining message authentication code. *J. Comput. Syst. Sci.*, 61(3):362–399, 2000.

[BPP+17] Subhadeep Banik, Sumit Kumar Pandey, Thomas Peyrin, Yu Sasaki, Siang Meng Sim, and Yosuke Todo. GIFT: A small present - towards reaching the limit of lightweight encryption. In *CHES 2017*, pages 321–345, 2017.

[BR06] Mihir Bellare and Phillip Rogaway. The security of triple encryption and a framework for code-based game-playing proofs. In *EUROCRYPT 2006*, pages 409–426, 2006.

[BSS+13] Ray Beaulieu, Douglas Shors, Jason Smith, Stefan Treatman-Clark, Bryan Weeks, and Louis Wingers. The simon and speck families of lightweight block ciphers. *IACR Cryptology ePrint Archive*, 2013:404, 2013.

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

[CLM19a] Yu Long Chen, Eran Lambooij, and Bart Mennink. How to build pseudorandom functions from public random permutations. In *CRYPTO 2019, Part I*, pages 266–293, 2019.

[CLM19b] Yu Long Chen, Eran Lambooij, and Bart Mennink. How to build pseudorandom functions from public random permutations. *IACR Cryptology ePrint Archive*, 2019:554, 2019.

[CLS15] Benoît Cogliati, Rodolphe Lampe, and Yannick Seurin. Tweaking even-mansour ciphers. In *CRYPTO 2015, Part I*, pages 189–208, 2015.

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

[CS16] Benoît Cogliati and Yannick Seurin. EWCDM: an efficient, beyond-birthday secure, nonce-misuse resistant MAC. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I*, pages 121–149, 2016.

[CS18] Benoît Cogliati and Yannick Seurin. Analysis of the single-permutation encrypted davies–meyer construction. *Designs, Codes and Cryptography*, 86, 03 2018.

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

[DDNY18a] Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Kan Yasuda. Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based mac. Cryptology ePrint Archive, Report 2018/500, 2018. https://eprint.iacr.org/2018/500.

[DDNY18b] Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Kan Yasuda. Encrypt or decrypt? to make a single-key beyond birthday secure nonce-based MAC. In *CRYPTO 2018, Part I*, pages 631–661, 2018.

[DEMS16] Christoph Dobraunig, Maria Eichlseder, Florian Mendel, and Martin Schläffer. Ascon v1.2. Submission to CAESAR, 2016. https://competitions.cr.yp.to/round3/asconv12.pdf.

[DHT17] Wei Dai, Viet Tung Hoang, and Stefano Tessaro. Information-theoretic indistinguishability via the chi-squared method. In *CRYPTO 2017, Part III*, pages 497–523, 2017.

[DJN17] Avijit Dutta, Ashwin Jha, and Mridul Nandi. Tight security analysis of ehtm MAC. *IACR Cryptology ePrint Archive*, 2017:837, 2017.

[DNT19] Avijit Dutta, Mridul Nandi, and Suprita Talnikar. Beyond birthday bound secure MAC in faulty nonce model. In *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part I*, pages 437–466, 2019.

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

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

[GPP11] Jian Guo, Thomas Peyrin, and Axel Poschmann. The PHOTON family of lightweight hash functions. In *CRYPTO 2011*, pages 222–239, 2011.

[HWKS98] Chris Hall, David A. Wagner, John Kelsey, and Bruce Schneier. Building prfs from prps. In *CRYPTO '98*, pages 370–389, 1998.

[JNP14] Jérémy Jean, Ivica Nikolic, and Thomas Peyrin. Tweaks and keys for block ciphers: The TWEAKEY framework. In *ASIACRYPT 2014*, pages 274–288, 2014.

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

[LRW02] Moses Liskov, Ronald L. Rivest, and David A. Wagner. Tweakable block ciphers. In *CRYPTO 2002*, pages 31–46, 2002.

[Luc00] Stefan Lucks. The sum of prps is a secure PRF. In *EUROCRYPT 2000*, pages 470–484, 2000.

[Men16] Bart Mennink. XPX: generalized tweakable even-mansour with improved security guarantees. In *CRYPTO 2016, Part I*, pages 64–94, 2016.

[MI11] Kazuhiko Minematsu and Tetsu Iwata. Building blockcipher from tweakable blockcipher: Extending FSE 2009 proposal. In *IMACC 2011*, pages 391–412, 2011.

[MN17] Bart Mennink and Samuel Neves. Encrypted davies-meyer and its dual: Towards optimal security using mirror theory. In *CRYPTO 2017, Part III*, pages 556–583, 2017.

[MRV15] Bart Mennink, Reza Reyhanitabar, and Damian Vizár. Security of full-state keyed sponge and duplex: Applications to authenticated encryption. In *ASIACRYPT 2015, Part II*, pages 465–489, 2015.

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing - Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[Nan20] Mridul Nandi. Mind the composition: Birthday bound attacks on ewcdmd and sokac21. *IACR Cryptology ePrint Archive*, 2020:236, 2020.

[NPV17] Valérie Nachef, Jacques Patarin, and Emmanuel Volte. *Feistel Ciphers - Security Proofs and Cryptanalysis*. Springer, 2017.

[Pat91] J. Patarin. Etude des Générateurs de Permutations Basés sur le Schéma du D.E.S. Phd Thèsis de Doctorat de l'Université de Paris 6, 1991.

[Pat05] Jacques Patarin. On linear systems of equations with distinct variables and small block size. In *ICISC 2005, Revised Selected Papers*, pages 299–321, 2005.

[Pat10] Jacques Patarin. Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. *IACR Cryptology ePrint Archive*, 2010:287, 2010.

[Pat16] Jacques Patarin. Mirror theory and cryptography. *IACR Cryptology ePrint Archive*, 2016:702, 2016.

[Rog04] Phillip Rogaway. Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In *ASIACRYPT 2004*, pages 16–31, 2004.

[Vau03] Serge Vaudenay. Decorrelation: A Theory for Block Cipher Security. *J. Cryptology*, 16(4):249–286, 2003.

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

Supplementary Material

A Proof of Sum-Capture Lemma 2

Here, we provide a proof for Lemma 2. For a basic review of the definitions and notations used in this proof, please refer [CS18].

Lemma 2. Let T^*, A^* be multisets of $\{0,1\}^n$ and $B \subseteq \{0,1\}^n$. Define-

$$\mu(T^*, A^*, B) = |\{(t,a,b) \in T^* \times A^* \times B : t = a \oplus b\}| \text{ and}$$

$$\mu(T^*) = \max_{\substack{A^*,B \\ |T^*|=q_1,|A^*|=q_2,|B|=p}} \mu(T^*, A^*, B).$$

If T^*, A^* are multisets of respectively q_1, q_2 uniformly random and independently chosen elements of $\{0,1\}^n$ and B is a subset of $\{0,1\}^n$ of size p, Then

$$\Pr\left[\mu(T^*) \geq \frac{q_1 q_2 p}{2^n} + \sqrt{\frac{3np(q_1+q_2)}{2^n}}\right] \leq \frac{2}{2^n}.$$

Proof.

$$\begin{aligned}
\mu(T^*, A^*, B) &= \sum_{t,a \in \{0,1\}^n} \delta_{T^*}(t)\delta_{A^*}(a)\mathbb{1}_B(b) \\
&= \sum_{t \in \{0,1\}^n} \delta_{T^*}(t)\left(\delta_{A^*} \star \mathbb{1}_B\right)(t) \\
&= 2^n \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{T^*}}(\alpha)\left(\widehat{\delta_{A^*} \star \mathbb{1}_B}\right)(\alpha) \\
&= 2^{2n} \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{T^*}}(\alpha)\widehat{\delta_{A^*}}(\alpha)\widehat{\mathbb{1}_B}(\alpha) \\
&= 2^{2n}\widehat{\delta_{T^*}}(0)\widehat{\delta_{A^*}}(0)\widehat{\mathbb{1}_B}(0) + 2^{2n}\sum_{\alpha \neq 0} \widehat{\delta_{T^*}}(\alpha)\widehat{\delta_{A^*}}(\alpha)\widehat{\mathbb{1}_B}(\alpha), \\
&\text{where} \quad \widehat{\delta_{T^*}}(0) = \frac{|T^*|}{2^n},\ \widehat{\delta_{A^*}}(0) = \frac{|A^*|}{2^n},\ \widehat{\mathbb{1}_B}(0) = \frac{|B|}{2^n} \text{ imply} \\
\mu(T^*, A^*, B) &= \frac{q_1 q_2 p}{2^n} + 2^{2n}\sum_{\alpha \neq 0} \widehat{\delta_{T^*}}(\alpha)\widehat{\delta_{A^*}}(\alpha)\widehat{\mathbb{1}_B}(\alpha) \\
&\leq \frac{q_1 q_2 p}{2^n} + 2^{2n}\sum_{\alpha \neq 0} \left|\widehat{\delta_{T^*}}(\alpha)\right|\left|\widehat{\delta_{A^*}}(\alpha)\right|\left|\widehat{\mathbb{1}_B}(\alpha)\right| \\
&\leq \frac{q_1 q_2 p}{2^n} + \Phi(T^*)\Phi(A^*)\sum_{\alpha \neq 0}\left|\widehat{\mathbb{1}_B}(\alpha)\right|, \\
&\text{where} \quad \Phi(T^*) = \max_{\alpha \neq 0}\left\{2^n\left|\widehat{\delta_{T^*}}(\alpha)\right|\right\} \\
&\text{and} \quad \Phi(A^*) = \max_{\alpha \neq 0}\left\{2^n\left|\widehat{\delta_{A^*}}(\alpha)\right|\right\}.
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{\alpha} \in \{0,1\}^n \left|\widehat{\mathbb{1}_B}(\alpha)\right|^2 &\geq \left(\sum_{\alpha} \neq 0 \left|\widehat{\mathbb{1}_B}(\alpha)\right|\right)^2 - 2 \cdot \sum_{0 \leq \alpha < \beta < 2^n} \left|\widehat{\mathbb{1}_B}(\alpha)\right| \cdot \left|\widehat{\mathbb{1}_B}(\beta)\right| \\
\implies \sum_{\alpha \neq 0}\left|\widehat{\mathbb{1}_B}(\alpha)\right| &\leq \sqrt{\frac{|B|}{2^n} + 2\sum_{0 \leq \alpha < \beta < 2^n}\left|\widehat{\mathbb{1}_B}(\alpha)\right| \cdot \left|\widehat{\mathbb{1}_B}(\beta)\right|} \leq \sqrt{\frac{|B|}{2^n}}.
\end{aligned}$$

Therefore, $\mu(T^*, A^*, B) \leq \frac{q_1 q_2 p}{2^n} + \Phi(T^*)\Phi(A^*) \cdot \sqrt{\frac{p}{2^n}}$. Since this holds for any $A^*, B \subseteq \{0,1\}^n$, it follows that

$$\frac{q_1 q_2 p}{2^n} + \sqrt{\frac{p}{2^n}} \cdot C \leq \mu(T^*) \leq \frac{q_1 q_2 p}{2^n} + \Phi(T^*)\Phi(A^*) \cdot \sqrt{\frac{p}{2^n}}$$

for some appropriate value of C, which implies $\Pr\left[\mu(T^*) \geq \frac{q_1 q_2 p}{2^n} + \sqrt{\frac{p}{2^n}} \cdot C\right] \leq \Pr\left[\Phi(T^*)\Phi(A^*) \geq C\right]$. Denote $T^* = \{t_1, \ldots, t_{q_1}\}$ and $A^* = \{a_1, \ldots, a_{q_2}\}$ using arbitrary orders. Then-

$$\begin{aligned}
\Phi(T^*) &= \max_{\alpha \neq 0} \left\{ 2^n \cdot \left| \widehat{\delta_{T^*}}(\alpha) \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \delta_{T^*}(x) \cdot (-1)^{\alpha \cdot x} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \sum_{i=1}^{q_1} \mathbb{1}_{\{t_i\}}(x) \cdot (-1)^{\alpha \cdot x} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{i=1}^{q_1} (-1)^{\alpha \cdot t_i} \right| \right\}.
\end{aligned}$$

Similarly, $\Phi(A^*) = \max\limits_{\alpha \neq 0} \left\{ \left| \sum\limits_{j=1}^{q_2} (-1)^{\alpha \cdot a_j} \right| \right\}$.

$$\begin{aligned}
\therefore \Phi(T^*)\Phi(A^*) &= \max_{\alpha \neq 0} \left\{ \left| \sum_{i=1}^{q_1} (-1)^{\alpha \cdot t_i} \right| \right\} \cdot \max_{\alpha \neq 0} \left\{ \left| \sum_{j=1}^{q_2} (-1)^{\alpha \cdot a_j} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{i=1}^{q_1} (-1)^{\alpha \cdot t_i} \cdot \sum_{j=1}^{q_2} (-1)^{\alpha \cdot a_j} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{(i,j) \in [q_1] \times [q_2]} (-1)^{\alpha \cdot (t_i + a_j)} \right| \right\}
\end{aligned}$$

For $\alpha \neq 0$, denoting $A^{(\alpha)}_{(i,j)} = (-1)^{\alpha \cdot (t_i + a_j)}$ and $A^{(\alpha)} = \sum\limits_{(i,j) \in [q_1] \times [q_2]} (-1)^{\alpha \cdot (t_i + a_j)}$, one obtains $\Phi(T^*)\Phi(A^*) = \max\limits_{\alpha \neq 0} \left\{ \left| A^{(\alpha)} \right| \right\}$. The random variable $A^{(\alpha)}$ is the sum of $q_1 + q_2$ independent random variables $A^{(\alpha)}_{(i,j)}$ such that $\Pr\left[A^{(\alpha)}_{(i,j)} = 1\right] = \Pr\left[A^{(\alpha)}_{(i,j)} = -1\right] = \frac{1}{2}$. Therefore, by the Chernoff bound given in Corollary 4.8 of [MU05], for any $a > 0$, $\Pr\left[\left|A^{(\alpha)}\right| \geq a\right] \leq 2e^{-a^2/2(q_1+q_2)}$.
Let $C \geq \sqrt{3n(q_1+q_2)}$. Then $\Pr\left[\left|A^{(\alpha)}\right| \geq C\right] \leq 2e^{-C^2/2(q_1+q_2)}$

$$\begin{aligned}
\implies \Pr\left[\mu(T^*) \geq \frac{q_1 q_2 p}{2^n} + \sqrt{\frac{p}{2^n}} \cdot C\right] &\leq \Pr\left[\Phi(T^*)\Phi(A^*) \geq C\right] \\
&= \Pr\left[\max_{\alpha \neq 0} \left\{ \left| A^{(\alpha)} \right| \right\} \geq C\right] \\
&\leq \sum_{\alpha \neq 0} \Pr\left[\left| A^{(\alpha)} \right| \geq C\right] \\
&\leq 2e^{-C^2/2(q_1+q_2)} \leq \frac{2}{2^n}, \\
&\text{since} \quad e^{3/4} \geq 2.
\end{aligned}$$

□

B Proof of Lemma 4

Consider a set S' of size $2^n - p'$, and three random variables $P, Q, R \xleftarrow[\text{wor}]{\$} S'$. Fix $\lambda_1, \lambda_2 \in \mathbb{F}_{2^n}$. For $i \in \{1, 2, 3\}$, let

$$A_i = \{(a_1, a_2, a_3) | a_1 \oplus a_2 = \lambda_1, a_2 \oplus a_3 = \lambda_2, a_i \notin S'\},$$

so that $|A_i| \leq p'$. Thus,

$$\{(p, q, r) \in S'^{(3)} | p \oplus q = \lambda_1, q \oplus r = \lambda_2\} = \\ \{(p, p \oplus \lambda_1, p \oplus \lambda_1 \oplus \lambda_2) | p \in \{0,1\}^n\} \setminus (A_1 \cup A_2 \cup A_3),$$

which is a set of size no less than $2^n - 3p'$. Hence,

$$\begin{aligned} \Pr\left[\begin{smallmatrix} P \oplus Q = \lambda_1, \\ Q \oplus R = \lambda_2 \end{smallmatrix}\right] &= \frac{2^n - |A_1 \cup A_2 \cup A_3|}{(2^n - p')(2^n - p' - 1)(2^n - p' - 2)} \\ &\geq \frac{2^n - 3p'}{(2^n - p')(2^n - p')(2^n - p')} \\ &= \frac{1}{2^{2n}} \left(1 - \frac{3 \cdot 2^n \cdot p'^2 - p'^3}{(2^n - p')^3}\right). \end{aligned}$$

C Figures Describing Bad Events for PDM*MAC

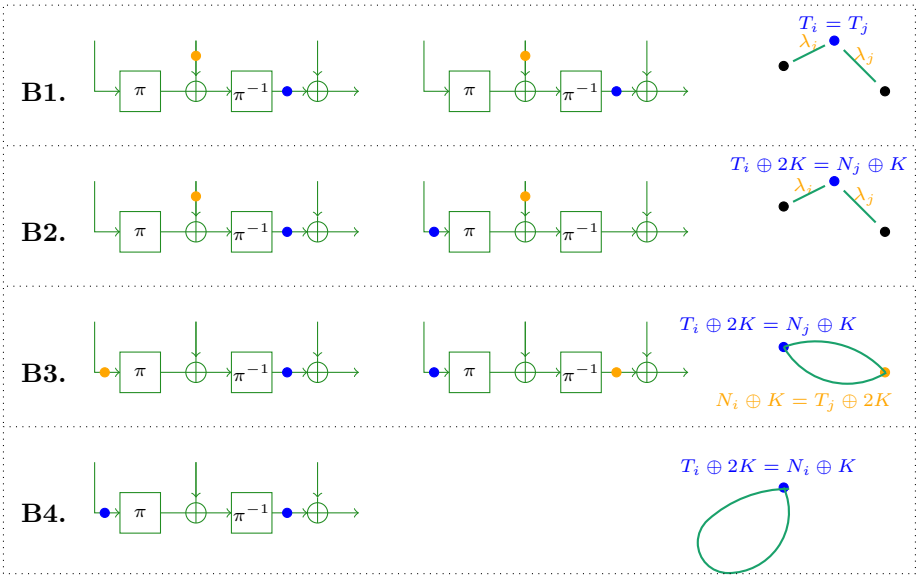

Figure 8: Collisions amongst one or two authentication queries - Figurative and graphical representations of the bad events.

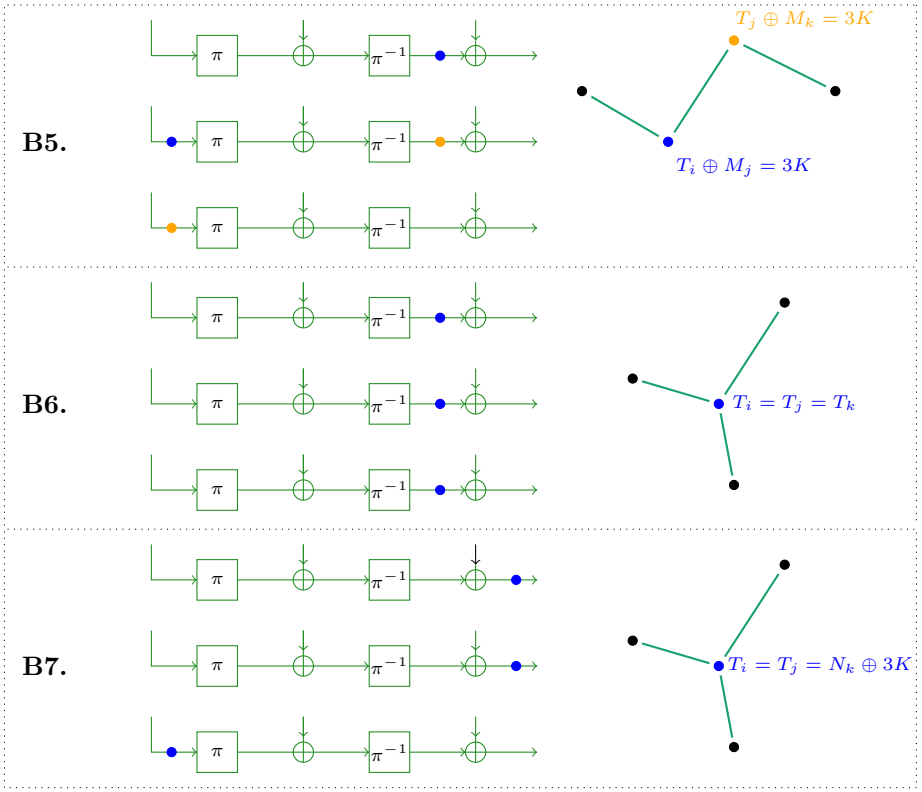

Figure 9: Collisions amongst three authentication queries - Figurative and graphical representations of the bad events.

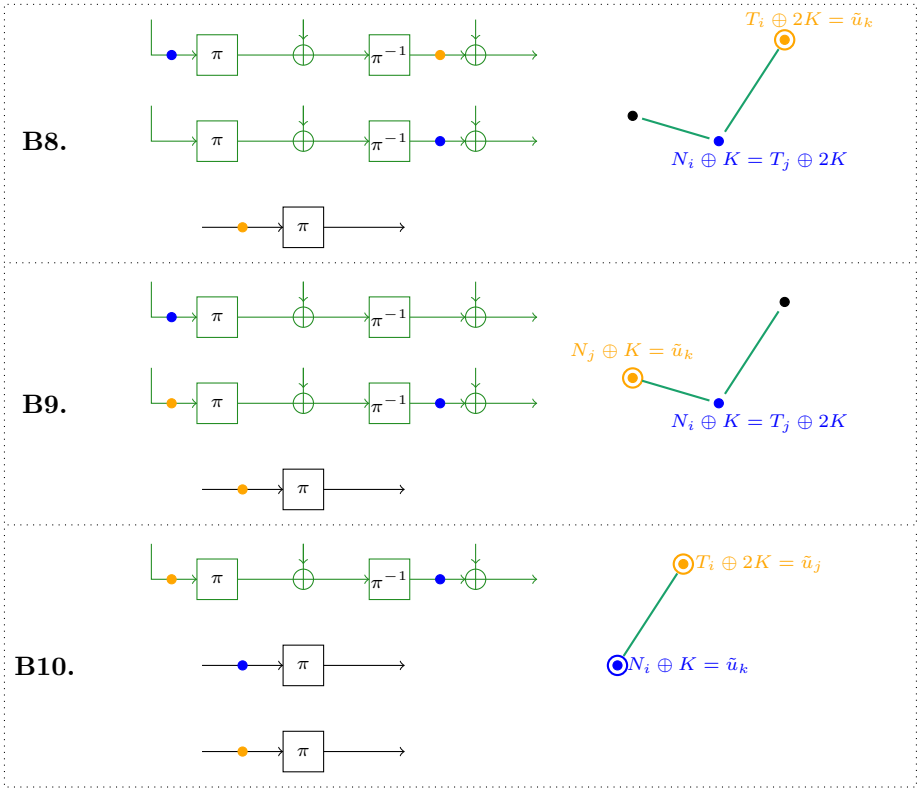

Figure 10: Collisions amongst authentication queries and primitive queries - Figurative and graphical representations of the bad events.

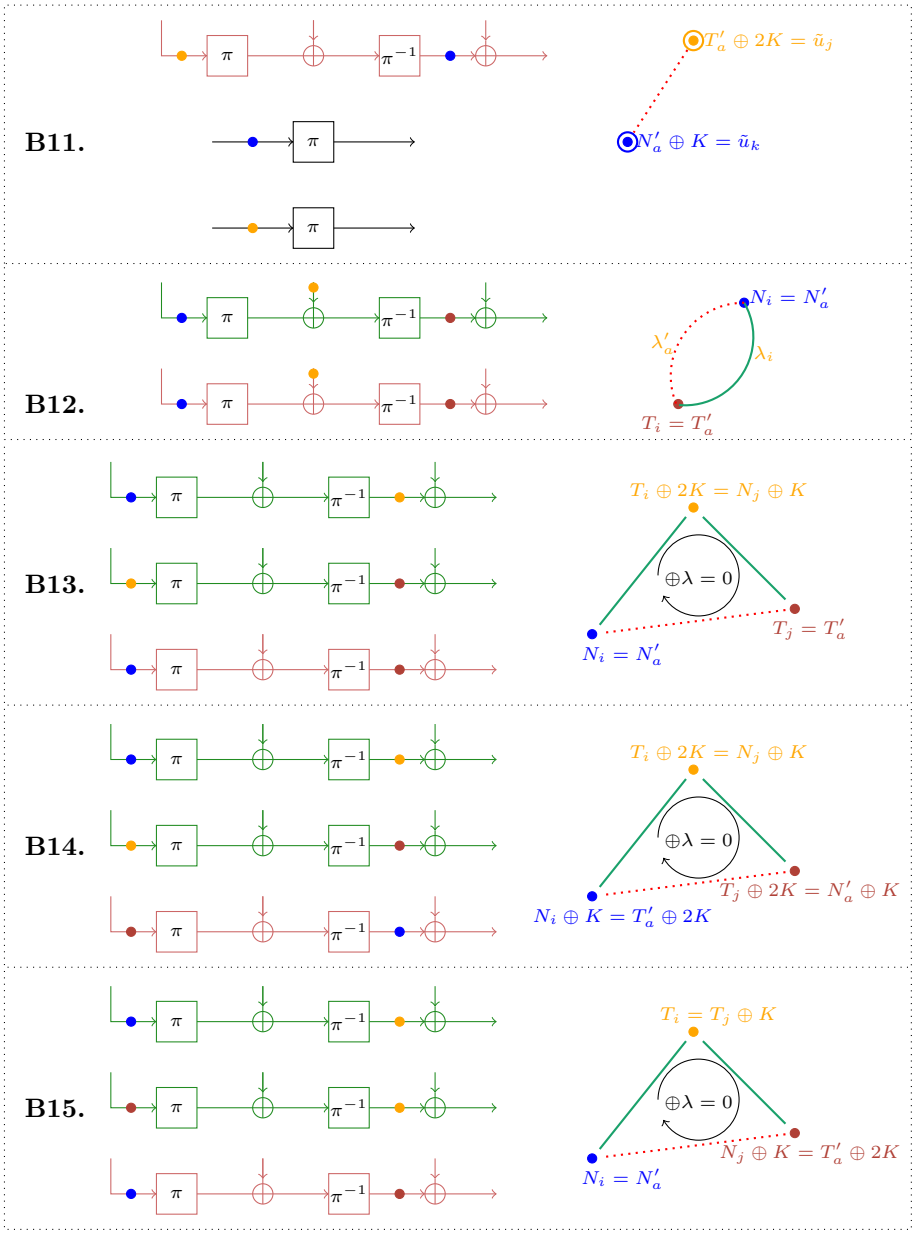

Figure 11: Collisions of verification queries with authentication and/or primitive queries - Figurative and graphical representations of the bad events.