On the Number of Restricted Solutions to Constrained Systems and their Applications

Benoît Cogliati[1], Jordan Ethan[2], Ashwin Jha[3], Mridul Nandi[4], and Abishanka Saha[4]

[1]Thales DIS France SAS, Meudon, France
benoit.cogliati@gmail.com
[2]CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
jordan.ethan@cispa.de
[3]Ruhr-Universität Bochum, Bochum, Germany
letterstoashwin@gmail.com
[4]Indian Statistical Institute, Kolkata, India
{mridul.nandi,sahaa.1993}@gmail.com

Abstract. In this paper, we formulate a special class of systems of linear equations over finite fields that appears naturally in the provable security analysis of several MAC and PRF modes of operation. We derive lower bounds on the number of solutions for such systems adhering to some predefined restrictions, and apply these lower bounds to derive tight PRF security for several constructions. We show security up to $2^{3n/4}$ queries for the single-keyed variant of the Double-block Hash-then-Sum (`DBHtS`) construction, called `1k-DBHtS`, under appropriate assumptions on the underlying hash function. We show that the single-key variants of `PMAC+` and `LightMAC+`, called `1k-PMAC+` and `1k-LightMAC+` achieve the required hash function properties, and thus, achieve $3n/4$-bit security. Additionally, we show that the sum of r independent copies of the Even-Mansour cipher is a secure PRF up to $2^{\frac{r}{r+1}n}$ queries.

Keywords: `PMAC+`, `LightMAC+`, Sum of Even-Mansour, tight security

1 Introduction

For some $k \geq 2$, let $\Pi_1, \ldots, \Pi_k$ denote k mutually independent and uniform random permutations of $\{0,1\}^n$, and consider the function $\mathtt{F} : \{0,1\}^n \to \{0,1\}^n$ defined by the mapping

$$\mathtt{F}(x) := \Pi_1(x) \oplus \Pi_2(x) \oplus \ldots \oplus \Pi_k(x)$$

It is well-known [20,23] that $\mathtt{F}$ — the well-known sum of k permutations — is statistically indistinguishable from a length-preserving uniform random function, provided the permutations are secret and the number of queried points $q \leq 2^{n-1}$. Over the years, several proof techniques [4,32,23,17,21,15,20] have been used to prove this result, with various degrees of success. In particular, Patarin's mirror theory [36,37], has been the main tool to study the underlying combinatorial problem.

Suppose $k = 2$ and the adversary makes q queries to the oracle at hand. Let $Y_1^i := \Pi_1(x_i)$, $Y_2^i := \Pi_2(x_i)$, and λ_i denote the oracle output for any $1 \le i \le q$. A typical mirror theory based proof studies the system of equations $\{Y_1^i \oplus Y_2^i = \lambda_i\}$ and aims to count all solutions $(y_1^i, y_2^i : i \in [q])$, such that $y_b^i \ne y_b^j$ for all $i \ne j$. In a more general setting, one can study a system of bivariate equations endowed with a partition of the set of variables such that any two variables in the same partition must be assigned distinct values. We call this structure a *constrained system*. It is not difficult to see that for random outputs, the expected number of solutions is $(2^n)_q \times (2^n)_q / 2^{nq}$, where $(2^n)_q$. Dutta et al. and Cogliati and Patarin studied [21,15] the problem specific to the sum of permutations and showed a lower bound close to the expectation, while $q \le 2^n/24$ and the solution space is $\{0,1\}^{2n}$, and as a result a good bound on the advantage. Although this approach works when the permutations are secret, it does not apply directly when the adversary has oracle access to the permutations.

This is, for instance, the case with the sum of Even-Mansour or `SOEM` construction [12] defined by the mapping

$$\mathtt{F}(x) := \Pi_1(x \oplus K_1) \oplus \Pi_2(x \oplus K_2),$$

where (K_1, K_2) denotes the key. Since the adversary can now make primitive queries, certain solutions are forbidden for fresh permutation inputs for any construction query. More specifically, if $\mathcal{P}_1$ and $\mathcal{P}_2$ denote the set of primitive query outputs then, for any construction query with fresh inputs to Π_1 and Π_2, a valid solution must lie outside $(\mathcal{P}_1 \times \{0,1\}^n) \cup (\{0,1\}^n \times \mathcal{P}_2)$. As it turns out, existing mirror theory approaches cannot be extended directly in this general setting. In fact, the best lower bounds [19,12,29] show that the number of solutions are just $(1 - O(q^3/2^{2n}))$-close to the expectation, provided $q \le 2^{2n/3}$.

A similar situation also arises in the secret permutations regime. For instance, all single-keyed attempts at DbHtS-based MACs, like `1k-PMAC+` [19] and `n1kf9` [38] are shown to be secure up to $2^{2n/3}$ queries. The main bottleneck: a (possibly) suboptimal lower bound for the number of solutions for the underlying constrained system. Several previous works [19,14,29] mention this as the primary hurdle in improving the security bound for a class of single-keyed constructions. This motivates us to study the aforementioned combinatorial problem in its full generality.

1.1 Related Works

Single-keyed DbHtS. Most common message authentication code (MAC) and pseudorandom function (PRF) constructions are based on block ciphers. Prominent examples include `CBC-MAC` [5], `PMAC` [7], `OMAC` [26], `LightMAC` [33] etc. At a high level, these constructions come under the umbrella of UHF-then-PRF designs, where first a message is compressed to a short string by a universal[1] hash

[1] A keyed function is called universal, if it is collision-resistant under a uniformly chosen key.

function (UHF) and then a block cipher is applied on this string to generate the tag. As a consequence, a collision on the hash output implies a collision on the construction output, and vice-versa. In general, this gives a birthday bound distinguishing attack, which is a problem when the MAC is instantiated with small block ciphers, like PRESENT [9], LED [24], and GIFT -64 [3].

To go beyond birthday bound, one possible way to improve upon the UHF-then-PRF design is to double the hash output size, which then renders the aforementioned collision attack ineffective. One can obviously use a bigger block cipher to encrypt the hash output. A more prudent and popular approach is to apply the sum-of-permutations transformation, i.e., each block is encrypted separately and the resulting pair are XORed to generate the output. Datta et al. [18] formalized this, naming the design *diblock hash-then-sum* or DBHtS. They proved that several constructions, like PolyMAC [8,6,40], SUM-ECBC [41], PMAC+ [42], 3kf9 [43], LightMAC+ [35] etc. follow the DBHtS design paradigm and achieve $2n/3$-bit security. In [31], Leurent et al. presented a $3n/4$-bit attack against DBHtS schemes, and later, Kim et al. [30] proved $3n/4$-bit security of the above constructions, closing the gap. Independently, Jha and Nandi also proved [28] a $3n/4$-bit security for the general DBHtS construction with independent hash functions.

While the DBHtS paradigm is quite popular as it can be very efficient, provided the underlying hash is efficient and parallelizable, it requires multiple block cipher keys. Indeed, three keys are required: one each for the two block ciphers and an independent key for the hash function. On the other hand, having a single key variant of DBHtS is quite desirable from a practical point of view. Datta et al. proposed [19] a single-key variant for PMAC+, called the 1k-PMAC+, and show that this construction is $2n/3$-bit secure. In a similar vein, Shen and Sibleyras proposed a single-keyed variant of 3kf9, called n1kf9, and showed a similar security bound. However, there is no matching attack, and in fact, these constructions are believed to have a similar security bound as their original three-key counterparts.

Sum of Even-Mansour. All the constructions discussed so far are block-cipher based. In recent years, however, several new constructions have instead employed cryptographic permutations or to use the theory parlance, a *public*[2] *random permutation*. SipHash [2] and keyed sponge [1,34] are probably the first two such constructions achieving birthday bound security. In [12], Chen et al. first proposed to use a public random permutation to construct beyond-the-birthday bound PRFs. In particular, they proposed the *sum of Even-Mansour* or $\mathtt{SOEM}^2$ construction. The basic idea is to instantiate the block ciphers in the sum of permutations construction with public permutation based block cipher $\mathtt{EM}^{\Pi}(\mathsf{K}, m) = \Pi(\mathsf{K} \oplus m) \oplus \mathsf{K}$. Chen et al. showed that the sum of two independent Even-Mansour constructions, $\mathtt{SOEM}^2_{\Pi_1,\Pi_2}(\mathsf{K}_1, \mathsf{K}_2, m) = \mathtt{EM}^{\Pi_1}(\mathsf{K}_1, m) \oplus \mathtt{EM}^{\Pi_2}(\mathsf{K}_2, m)$ is a $2n/3$-bit secure PRF. In fact, they showed that the independence of the permutation and keys is a necessary condition as any weaker assumption would degrade the security to a birthday-bound. In [39], Sibleyras et al. demonstrated that while independence is

[2] The adversary is allowed black-box access to the permutation.

essential, the post-adding of keys is redundant. Indeed, a similar security bound is achievable with a more efficient design called the *keyed sum of permutations*, $\mathsf{KSoP}_{\Pi_1,\Pi_2}(\mathsf{K}_1, \mathsf{K}_2, m) = \Pi_1(\mathsf{K}_1 \oplus m) \oplus \Pi_2(\mathsf{K}_2 \oplus m)$. In this paper, we use $\texttt{SOEM}^2$ to refer to this slightly modified construction. Since the initial $\texttt{SOEM}^2$ proposal, several follow up works [10,22,11,39] came up with new variants relaxing certain conditions, although still within the $2n/3$-bit security range.

1.2 Our Contributions

Table 1. Summary of Results. ℓ, s, and n denote the the message length in n-bit blocks, the counter size in bits, and the block size of the primitive, respectively. The security bounds denote an upper bound on the number of queries (ignoring any length factors).

Construction	#Keys	#Calls	Security bound	Tightness
DBHtS [18]	3	2^+	$O(2^{3n/4})$ [30]	✓ [31]
PMAC+ [42]	3	$\ell + 2$	$O(2^{3n/4})$ [30]	✓ [31]
3kf9 [43]	3	$\ell + 2$	$O(2^{3n/4})$ [30]	✓ [31]
LightMAC+ [35]	3	$\ell\left(1 + \frac{s}{n-s}\right) + 2$	$O(2^{3n/4})$ [30]	✓ [31]
1k-PMAC+ [19]	1	$\ell + 2$	$O(2^{2n/3})$ [19]	–
n1kf9 [38]	1	$\ell + 2$	$O(2^{2n/3})$ [38]	–
$\texttt{SOEM}^2$ [12,39]	2	2	$\widetilde{O}(2^{2n/3})$ [12,39]	✓ [12,39]
1k-DBHtS	1	2^+	$O(2^{3n/4})$	✓
1k-PMAC+	1	$\ell\left(1 + \frac{2}{n-2}\right) + 2$	$O(2^{3n/4})$	✓
1k-LightMAC+	1	$\ell\left(1 + \frac{s+2}{n-s-2}\right) + 2$	$O(2^{3n/4})$	✓
$\texttt{SOEM}^r$	r	r	$\widetilde{O}(2^{rn/r+1})$	✓

Our Contributions are twofold:

1. In section 3, we formalize a general constrained system over any arbitrary finite field. This of course includes the omnipotent extensions of binary field. In section 4, we derive a lower bound on the number of solutions for a large class of constrained systems that encompasses all the known instances in literature.
2. As an application, we prove tight security bounds for several class of constructions:
   - *Tight bounds for single-keyed DbHtS:* In section 5, we define the single-key variant of DBHtS, called 1k-DBHtS, and prove that it achieves security up to $2^{3n/4}$ queries (and the total number of blocks) as long as the hash function satisfies some properties introduced in section 2.1. This solves

the open problem posed by Datta et al. in [19]. In section 6, we study two instances of `1k-DBHtS`, namely, the single-keyed variant of `1k-PMAC+` and `1k-LightMAC+`. In particular, we show that the corresponding hash functions, called `TPHash` and `TLightHash` achieve the desired properties, and thus, `1k-PMAC+` and `1k-LightMAC+` achieve security up to $2^{3n/4}$ queries. Our security bounds are tight in the number of queries as the Leurent-Nandi-Sibleyras attack [31] on `DBHtS` also applies to `1k-DBHtS`.

- *Tight Security of Sum of Even-Mansour:* In section 7, for $r \geq 2$, we define the sum of r Even-Mansour ciphers — an extension of the Sibleyras-Todo [39] variant of the sum of two Even-Mansour construction [12] by Chen et al. We show that this construction achieves security up to $2^{\frac{r}{r+1}n}$ queries, which can be shown to be tight by a simple key recovery argument [12,39]. This directly generalizes the previous results, both in terms of design and security.

Table 1 summarizes our results and gives a brief comparison with relevant existing results.

2 Preliminaries

For any prime power N, $\mathbb{F}_N$ denotes the finite field of order N. With a slight abuse of notation, we use $\oplus$ and $\cdot$ to denote the addition and multiplication operations in any finite field, and $\ominus$ to denote the subtraction operation. For $m, n \in \mathbb{N}^+$, $\mathbb{F}_N^m$ and $\mathbb{F}_N^{m \times n}$ denote the m-dimensional vector space and the set of all $(m \times n)$-matrices over $\mathbb{F}_N$, respectively. For any $\boldsymbol{v} \in \mathbb{F}_N^m$, $\mathtt{H}(\boldsymbol{v})$ denotes the number of non-zero coordinates in $\boldsymbol{v}$.

For any $n \in \mathbb{N}^+$, we identify $\mathbb{F}_{2^n}$ with $\{0,1\}^n$, the set of all n-bit strings. We write $\{0,1\}^* := \cup_{n=0}^{\infty}\{0,1\}^n$. For any $k \leq n \in \mathbb{N}^+$, $(n)_k := n(n-1)\ldots(n-k+1)$ denotes the falling factorial, and $(n)_0 = 1$ by convention.

Sum Capture: For some $k \geq 2$, let $\boldsymbol{\alpha} \in \mathbb{F}_N^k$ and $\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k \subseteq \mathbb{F}_N$ such that $\mathtt{H}(\boldsymbol{\alpha}) = k$. Define

$$\mathcal{SC}_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}) := \left\{ b \in \mathcal{B}_1 \times \ldots \times \mathcal{B}_k : \bigoplus_{i=1}^{k} \alpha_i \cdot b_i \in \mathcal{A} \right\}, \quad (1)$$

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}) := |\mathcal{SC}_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B})|, \quad (2)$$

where $\mathcal{B} = (\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_k})$ denotes an arbitrary ordering of the constituent sets. For $\alpha = (1,1,\ldots,1)$, $\mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}_{i_1}, \mathcal{B}_{i_2}, \ldots, \mathcal{B}_{i_k}) = \mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}_{j_1}, \mathcal{B}_{j_2}, \ldots, \mathcal{B}_{j_k})$ for any two permutations $(i_1\, i_2\, \ldots\, i_k)$ and $(j_1\, j_2\, \ldots\, j_k)$ of $[k]$. We drop the mask from notation whenever $\boldsymbol{\alpha} = (1,1,\ldots,1)$.

For any $k \geq 2$ and $p \geq 0$, we define

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, p) := \max_{\substack{\mathcal{B}_1,\ldots,\mathcal{B}_k \subseteq \mathbb{F}_N \\ |\mathcal{B}_i| \leq p}} \mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}), \quad (3)$$

The following lemma is a restatement of [27, Theorem 2], which in turn is a simple generalization of a similar result proved for the $k = 2$ case in [16].

Lemma 1. *For all but $(4/N)$-fraction of (multi)sets $\mathcal{A} \subseteq \mathbb{F}_N$ such that $|\mathcal{A}| = q$ and any $\boldsymbol{\alpha} \in \mathbb{F}_N^k$ with $H(\boldsymbol{\alpha}) \geq 2$, we have*

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, p) \leq \left(\frac{qp^k}{N} + 2p^{k-1}\sqrt{\ln(N)q} \right).$$

Some Useful Results: Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)$ denote a sequence of $\mathbb{F}_{2^n}$-valued random variables. Define

$$\Delta := \max_d |\{i \in [q] : \lambda_i = d\}|$$
$$\nabla := \max_{d \neq 0} |\{(i,j) : i \neq j \in [q],\ \lambda_i \oplus \lambda_j = d\}|$$

Proposition 1. *For any with replacement sampled $\boldsymbol{\lambda}$, $\mathbb{E}(\Delta) \leq 6n\lceil q/2^n \rceil$.*

Proof. First, assume $q \leq 2^n$. For any positive integer $k \leq q$, we have

$$\Pr(\Delta \geq k) \leq \sum_{x \in \mathbb{F}_{2^n}} \binom{q}{k} \frac{1}{2^{nk}} \leq 2^n \times \binom{q}{k} \frac{1}{2^{nk}} \leq 2^n \frac{q^k}{2^{nk} k!} \leq 2^n \left(\frac{qe}{k2^n} \right)^k,$$

where we use $k! \geq (k/e)^k$ for the last inequality. Then

$$\mathbb{E}(\Delta) \leq (k-1) + q \Pr(\Delta \geq k) \leq (k-1) + q2^n \left(\frac{qe}{k2^n} \right)^k \leq 2^{2n} \left(\frac{qe}{k2^n} \right)^k.$$

For $q \leq 2^n$, the result then follows by setting $k = 6n$. For $q > 2^n$, divide the q sample into $\lceil q/2^n \rceil$ blocks, each of size at most 2^n. The result then follows by exploiting the linearity of expectation. ◻

Proposition 2. *For any without replacement sampled $\boldsymbol{\lambda}$, $\mathbb{E}(\nabla) \leq 6n\lceil q^2/2^n \rceil$.*

Proof. First, assume $q^2 \leq 2^n$. For any positive integer $k \leq 2^{n-1}$, we have

$$\Pr(\nabla \geq k) \leq \sum_{x \in \mathbb{F}_{2^n}^*} \binom{q^2}{k} \frac{2}{2^{nk}} \leq 2^n \times \binom{q^2}{k} \frac{2}{2^{nk}} \leq 2^{n+1} \frac{q^{2k}}{2^{nk} k!} \leq 2^{n+1} \left(\frac{q^2 e}{k2^n} \right)^k,$$

The result now follows by arguing as in the previous Proposition. ◻

Let $\sim$ denote an equivalence relation on $[q]$ such that $i \sim j$ if $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_j$. Let $\mathcal{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_c\}$ denote the corresponding partition. Let $\mathtt{NI} \subseteq [c]$ denote the set of indices corresponding to non-singleton blocks. Let $\nu_i = |\mathcal{P}_i|$ for all $i \in [c]$, and $\nu_{\max} = \max\{\nu_i : i \in [c]\}$.

The following two propositions are a simple restatement of two results from [28].

Proposition 3 (Lemma 4.3 in [28]). *Suppose $\boldsymbol{\lambda}$ satisfies the condition, for distinct $i, j \in [q]$, $\Pr(\lambda_i = \lambda_j) \leq \epsilon$. Then, we have*

$$\mathbb{E}\left(\sum_{i \in \mathtt{NI}} \nu_i^2 \right) \leq 2q^2\epsilon.$$

Proposition 4 (Corollary 4.1 in [28]). *Suppose* $\boldsymbol{\lambda}$ *satisfies the condition, for distinct* $i, j \in [q]$, $\Pr(\lambda_i = \lambda_j) \leq \epsilon$. *Then,*

$$\Pr(\nu_{\max} \geq a) \leq \frac{2q^2\epsilon}{a^2}.$$

Proposition 5. *For any real-valued random variable* $\mathtt{X}$*, we have*

$$\mathbb{E}(|\mathtt{X} - \mathbb{E}(\mathtt{X})|) \leq \sqrt{\mathbb{V}(\mathtt{X})}.$$

Proof. We have

$$\begin{aligned}\mathbb{E}(|\mathtt{X} - \mathbb{E}(\mathtt{X})|) &= \sqrt{\mathbb{E}(|\mathtt{X} - \mathbb{E}(\mathtt{X})|)^2} \\ &\leq \sqrt{\mathbb{E}((\mathtt{X} - \mathbb{E}(\mathtt{X}))^2)} = \sqrt{\mathbb{V}(\mathtt{X})},\end{aligned}$$

where the inequality also follows from Jensen's inequality among others. □

2.1 Hash Functions

A $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$-keyed function H is the function family $\{H_K : \{0,1\}^* \to \mathcal{Y}\}_{K \in \mathcal{K}}$.

We often call H a *diblock hash function*, if we can write $\mathcal{Y}$ as $\mathcal{Z}^2$ for some $\mathcal{Z}$. For any diblock hash function H, we write $(H^1_K(m), H^2_K(m)) := (z_1, z_2)$, where $z_1, z_2 \in \mathcal{Z}$, whenever $H_K(m) = y = (z_1, z_2)$.

Permutation-based Hash Functions. A $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$-hash function is said to be permutation-based if $\mathcal{K} \subseteq \mathcal{P}(n)^r$ for some $r \in \mathbb{N}$. For any such hash function H, the *block function*, $\beta_H : \mathcal{P}(n) \times \{0,1\}^* \to \mathbb{N}$, is defined by the mapping:

$$(\pi^r, m) \mapsto \beta_{(\pi^r, m)},$$

where $\pi^r = (\pi_1, \ldots, \pi_r)$ and $\beta_{(\pi^r,m)}$ denotes the minimum number of invocations[3] of π needed to compute $H_\pi(m)$.

In this paper, we fix $r = 1$, and make the following two plausible assumptions on β_H:

1. β_H is functionally independent of the permutation, whence we drop the permutation from the parameters.
2. there exists a constant $c \in \mathbb{R}^+$ such that for any $m \in \{0,1\}^*$, $\beta_H(m) := c\lceil |m|/n \rceil$. We refer to such an H a *rate-*c^{-1} hash function.

Note that, 1 follows from 2. We state it explicitly for brevity.

We remark that the underlying hash functions in almost all the popular constructions, including `LightMAC`, `PMAC`, `LightMAC+`, `PMAC+`, `3kf9` etc. are rate-1, and `SUM-ECBC` is rate-2^{-1}. Thus, the above assumption is indeed plausible, and $c \leq 2$ in most applications.

[3] Note that, there exists a circuit for H such that on every input, H makes (possibly) a large but bounded number of black-box calls to π^r. Thus, $\beta_{\pi^r,m}$ is well-defined for any π^r and m.

Coverfree Hash Functions. For any $(\mathcal{K}, \{0,1\}^*, \mathcal{Y}^2)$-diblock hash function H, any $r \geq 3$, $s \geq 2$, and any $\boldsymbol{m} := (m_1, \ldots, m_q) \in (\{0,1\}^*)_q$, we define the following events

$\mathsf{COLL1}_H(\boldsymbol{m})$: $\exists^* i, j \in [q]$ such that $H^1_{\mathtt{K}}(m_i) = H^1_{\mathtt{K}}(m_j)$;

$\mathsf{COLL2}_H(\boldsymbol{m})$: $\exists^* i, j \in [q]$ such that $H^2_{\mathtt{K}}(m_i) = H^2_{\mathtt{K}}(m_j)$;

$\mathsf{AP1}^r_H(\boldsymbol{m})$: $\exists^* i_1, \ldots, i_r \in [q]$ such that $H^1_{\mathtt{K}}(m_{i_1}) = H^1_{\mathtt{K}}(m_{i_2}), H^2_{\mathtt{K}}(m_{i_2}) = H^2_{\mathtt{K}}(m_{i_3}), \ldots, H^1_{\mathtt{K}}(m_{i_{r-1}}) = H^1_{\mathtt{K}}(m_{i_r})$;

$\mathsf{AP2}^r_H(\boldsymbol{m})$: $\exists^* i_1, \ldots, i_r \in [q]$ such that $H^2_{\mathtt{K}}(m_{i_1}) = H^2_{\mathtt{K}}(m_{i_2}), H^1_{\mathtt{K}}(m_{i_2}) = H^1_{\mathtt{K}}(m_{i_3}), \ldots, H^2_{\mathtt{K}}(m_{i_{r-1}}) = H^2_{\mathtt{K}}(m_{i_r})$;

$\mathsf{MC1}^s_H(\boldsymbol{m})$: $\exists^* i_1, \ldots, i_s \in [q]$ such that $H^1_{\mathtt{K}}(m_{i_1}) = H^1_{\mathtt{K}}(m_{i_2}) = \cdots = H^1_{\mathtt{K}}(m_{i_s})$;

$\mathsf{MC2}^s_H(\boldsymbol{m})$: $\exists^* i_1, \ldots, i_s \in [q]$ such that $H^2_{\mathtt{K}}(m_{i_1}) = H^2_{\mathtt{K}}(m_{i_2}) = \cdots = H^2_{\mathtt{K}}(m_{i_s})$,

$\mathsf{COLL}_H(\boldsymbol{m})$: $\exists^* i, j \in [q]$ such that $H_{\mathtt{K}}(m_i) = H_{\mathtt{K}}(m_j)$.

where the randomness is induced by $\mathtt{K} \twoheadleftarrow \mathcal{K}$.

Definition 1. *For some* $\epsilon_1, \delta : \mathbb{N}^3 \to [0,1]$ *and* $\epsilon_2, \epsilon_3 : \mathbb{N}^4 \to [0,1]$*, a* $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$*-diblock hash function* H *is said to be an* $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$*-Coverfree Hash or CfH if and only if for any* $\rho = (q, \ell, \sigma) \in \mathbb{N}^3$*, any* $\boldsymbol{m} = (m_1, \ldots, m_q) \in (\{0,1\}^{n\ell})_q$ *containing at most* σ *blocks, any* $r \geq 3$*, and any* $s \geq 2$*, it satisfies*

$$\Pr(\mathsf{COLL1}_H(\boldsymbol{m})) \leq \epsilon_1(\rho), \quad \Pr(\mathsf{AP1}^r_H(\boldsymbol{m})) \leq \epsilon_2(\rho, r), \quad \Pr(\mathsf{MC1}^s_H(\boldsymbol{m})) \leq \epsilon_3(\rho, s),$$

$$\Pr(\mathsf{COLL2}_H(\boldsymbol{m})) \leq \epsilon_1(\rho), \quad \Pr(\mathsf{AP2}^r_H(\boldsymbol{m})) \leq \epsilon_2(\rho, r), \quad \Pr(\mathsf{MC2}^s_H(\boldsymbol{m})) \leq \epsilon_3(\rho, s),$$

and $\Pr(\mathsf{COLL}_H(\boldsymbol{m})) \leq \delta(\rho)$.

Double-block Hash-then-Sum. Let H be a $(\mathcal{K}, \{0,1\}^*, \{0,1\}^{2n})$-diblock hash function. The *DiBlock Hash-then-Sum* construction is a $(\mathcal{K} \times \mathcal{P}(n)^2, \{0,1\}^*, \{0,1\}^n)$-keyed function $\mathtt{DBHtS}_H$ defined by the mapping:

$$(K, \pi_1, \pi_2, m) \mapsto \pi_1(H^1_K(m)) \oplus \pi_2(H^2_K(m)) \tag{4}$$

Several beyond-the-birthday bound MAC constructions, including `SUM-ECBC` [41], `PMAC+` [42], `LightMAC+` [35] etc. follow this paradigm.

2.2 Security Definitions

In this paper, we assume that the distinguisher is non-trivial, i.e. it never makes a duplicate query, and it never makes a query for which the response is already known due to some previous query. Let $\mathbb{A}(q, \ell, \sigma, t)$ be the class of all non-trivial distinguishers limited to q oracle queries of each of length up to ℓ blocks and a total of σ blocks, and t computations. Any $\mathcal{A} \in \mathbb{A}(q, \ell, \sigma, t)$ is referred as a (q, ℓ, σ, t)-adversary.

In our analyses, especially security proofs, it will be convenient to work in the information-theoretic setting. Accordingly, we always skip the boilerplate hybrid steps and often assume that the adversary is computationally unbounded, i.e., $t = \infty$, and deterministic. A computational equivalent of all our security proofs can be easily obtained by a simple hybrid argument.

The advantage of any adversary $\mathcal{A}$ in distinguishing some oracle $\mathcal{O}_1$ from another oracle $\mathcal{O}_0$ is defined as

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) := \left|\Pr\left(\mathcal{A}^{\mathcal{O}_1} = 1\right) - \Pr\left(\mathcal{A}^{\mathcal{O}_0} = 1\right)\right|.$$

PRF Security: The PRF advantage of distinguisher $\mathcal{A}$ against a $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function $\mathtt{F}$ instantiated with a key $\mathtt{K} \twoheadleftarrow \mathcal{K}$ is defined as

$$\mathbf{Advt}_{\mathtt{F}}^{\mathrm{prf}}(\mathcal{A}) = \Delta_{\mathtt{F};\Gamma}(\mathcal{A}). \tag{5}$$

In this paper, we also consider the security model where the distinguisher is given oracle access to the internal primitives of the construction. More specifically, suppose $\mathtt{F}$ is constructed on top of k uniform random permutations $\Pi = (\Pi_1, \ldots, \Pi_k)$ of $\{0,1\}^n$, denoted $\mathtt{F}[\Pi]$. Then, the PRF advantage of $\mathcal{A}$ is defined as

$$\mathbf{Advt}_{\mathtt{F}[\Pi]}^{\mathrm{prf}}(\mathcal{A}) = \Delta_{(\mathtt{F}[\Pi],\Pi^{\pm});(\Gamma,\Pi^{\pm})}(\mathcal{A}), \tag{6}$$

where the superscript $\pm$ denotes a bidirectional access to Π.

2.3 The Expectation Method

Let $\mathcal{A}$ be a computationally unbounded and deterministic distinguisher that tries to distinguish between two oracles $\mathcal{O}_0$ and $\mathcal{O}_1$ via black box interaction with one of them. We denote the query-response tuple of $\mathcal{A}$'s interaction with its oracle by a transcript ω. This may also include any additional information the oracle chooses to reveal to the distinguisher at the end of the query-response phase of the game. We denote by Θ_{re} (res. Θ_{id}) the random transcript variable when $\mathcal{A}$ interacts with $\mathcal{O}_1$ (res. $\mathcal{O}_0$). The probability of realizing a given transcript ω in the security game with an oracle $\mathcal{O}$ is known as the *interpolation probability* of ω with respect to $\mathcal{O}$. Since $\mathcal{A}$ is deterministic, this probability depends only on the oracle $\mathcal{O}$ and the transcript ω. A transcript ω is said to be *attainable* if $\Pr(\Theta_{\mathrm{id}} = \omega) > 0$.

Lemma 2 (Fine-grained Expectation Method). *Let Ω be the set of all transcripts. For some $\varepsilon_{\mathrm{bad}} \geq 0$ and $\varepsilon_{\mathrm{ratio}} : \Omega \to \mathbb{R}$, suppose there is a set $\Omega_{\mathrm{bad}} \subseteq \Omega$ satisfying the following conditions:*
- $\Pr(\Theta_{\mathrm{id}} \in \Omega_{\mathrm{bad}}) \leq \varepsilon_{\mathrm{bad}}$,
- $\varepsilon_{\mathrm{ratio}}$ *is non-negative on* $\Omega_{\mathrm{good}} = \Omega \setminus \Omega_{\mathrm{good}}$,

- *for any $\omega \in \Omega_{\text{good}}$, ω is attainable and* $\dfrac{\Pr(\Theta_{\text{re}} = \omega)}{\Pr(\Theta_{\text{id}} = \omega)} \geq 1 - \varepsilon_{\text{ratio}}(\omega)$.

Then for any distinguisher $\mathcal{A}$ trying to distinguish between $\mathcal{O}_1$ and $\mathcal{O}_0$, we have the following bound on its distinguishing advantage:

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) \leq \varepsilon_{\text{bad}} + \mathbb{E}_{\Theta_{\text{id}}}(1_{\text{good}}\varepsilon_{\text{ratio}}),$$

where 1_{good} denotes the indicator variable corresponding to Ω_{good}.

The expectation method due to Hoang and Tessaro [25] is a simple corollary of the above result, when $\varepsilon_{\text{ratio}}$ is non-negative over the entire transcript space.

Corollary 1 (Expectation Method). *Suppose there is a non-negative function $\varepsilon_{\text{ratio}} : \Omega \to [0, \infty)$ satisfying the following conditions:*

- $\Pr(\Theta_{\text{id}} \in \Omega_{\text{bad}}) \leq \varepsilon_{\text{bad}}$;
- *For any $\omega \notin \Omega_{\text{bad}}$, ω is attainable and* $\dfrac{\Pr(\Theta_{\text{re}} = \omega)}{\Pr(\Theta_{\text{id}} = \omega)} \geq 1 - \varepsilon_{\text{ratio}}(\omega)$.

Then for any distinguisher $\mathcal{A}$ trying to distinguish between $\mathcal{O}_1$ and $\mathcal{O}_0$, we have the following bound on its distinguishing advantage:

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) \leq \varepsilon_{\text{bad}} + \mathbb{E}_{\Theta_{\text{id}}}(\varepsilon_{\text{ratio}}).$$

3 Constrained Systems

System of Linear Equations: Fix some $q, r \leq N$. Any system of q linear equations in r variables, $A\boldsymbol{x} = \boldsymbol{\lambda}$, over $\mathbb{F}_N$ can be compactly represented by the augmented matrix $A|\boldsymbol{\lambda}$, where $A \in \mathbb{F}_N^{q \times r}$ and $\boldsymbol{\lambda} \in \mathbb{F}_N^q$.

System-graph and Components: It would be often convenient to look at a graph-theoretic representation of the system $A|\boldsymbol{\lambda}$. Formally, to any system $A|\boldsymbol{\lambda}$, we associate an undirected, labeled, bipartite graph $G(A|\boldsymbol{\lambda}) = (\text{row}(A|\boldsymbol{\lambda}), \text{col}(A), \mathcal{E})$ where $\text{row}(A|\boldsymbol{\lambda})$ and $\text{col}(A)$ denote the two disjoint sets of vertices, and

$$\mathcal{E} = \{(\{A_{i\bullet}|\lambda_i, A_{\bullet j}\}, A_{ij}) \ : \ (i,j) \in [q] \times [r], A_{i,j} \neq 0^n\}$$

denotes the edge-set. Each edge $e = (\{A_{i\bullet}|\lambda_i, A_{\bullet j}\}, A_{ij}) \in \mathcal{E}$ is often written in a more illustrative notation as $A_{i\bullet}|\lambda_i \xrightarrow{A_{ij}} A_{\bullet j}$ or simply $i^- \text{——} j^|$ whenever convenient, where the superscripts $-$ and $|$ are used to differentiate row and column index, respectively. We call $G(A|\boldsymbol{\lambda})$ a *system-graph*.

In this context, we say that two rows $A_{i\bullet}|\lambda_i$ and $A_{i'\bullet}|\lambda_{i'}$ are *adjacent*, denoted $A_{i\bullet}|\lambda_i \sim A_{i'\bullet}|\lambda_{i'}$, if and only if there exists an $A_{\bullet j} \in \text{col}(A)$ such that $i^- \text{——} j^| \text{——} i'^-$.[4] The relation $\sim$ on $\text{row}(A|\boldsymbol{\lambda})$ is reflexive and symmetric, but not transitive.

We say that two rows $A_{i\bullet}$ and $A_{j\bullet}$ are *connected*, denoted $A_{i\bullet}|\lambda_i \rightsquigarrow A_{j\bullet}|\lambda_j$, if and only if they are connected in $G(A|\boldsymbol{\lambda})$. $\rightsquigarrow$ is an equivalence relation on

[4] Any two rows of a matrix are said to be disjoint, if they do not share a common column index with non-zero entry, and non-disjoint otherwise.

$\mathrm{row}(A|\boldsymbol{\lambda})$, effectively partitioning $\mathrm{row}(A|\boldsymbol{\lambda}) = A_1|\boldsymbol{\lambda}_1 \sqcup \cdots \sqcup A_c|\boldsymbol{\lambda}_c$. For each component $A_i|\boldsymbol{\lambda}_i$ of $A|\boldsymbol{\lambda}$, let $\overline{A}_i$ denote the *column-reduced form* of A_i, which is obtained by simply dropping all the zero columns from A_i. Then, it is easy to see that the induced subgraph $G[A_i|\boldsymbol{\lambda}_i, \mathrm{col}(\overline{A}_i)]$ is a component $G(A|\boldsymbol{\lambda})$, and a system-graph in its own right. As a consequence, with a slight abuse of notations, we also write $A_i|\boldsymbol{\lambda}_i$ to denote the $q_i \times (r+1)$ submatrix (also referred as a *component*) of $A|\boldsymbol{\lambda}$ corresponding to the equivalence class $A_i|\boldsymbol{\lambda}_i = \{A_{j_1\bullet}|\lambda_{j_1}, \ldots, A_{j_{q_i}\bullet}|\lambda_{j_{q_i}}\}$, i.e.

$$A_i|\boldsymbol{\lambda}_i = \begin{pmatrix} A_{j_1\bullet}|\lambda_{j_1} \\ \vdots \\ A_{j_{q_i}\bullet}|\lambda_{j_{q_i}} \end{pmatrix},$$

where $\sum_i q_i = q$. Let $r_i := |\mathrm{col}(\overline{A}_i)|$ and $\sum_i r_i = r$. For any $i \in [c]$, we say that $A_i|\boldsymbol{\lambda}_i$ is *isolated* if $q_i = 1$. By extension, $A|\boldsymbol{\lambda}$ is said to be isolated if $A_i|\boldsymbol{\lambda}_i$ is isolated for all $i \in [c]$.

Note that, both $\sim$ and $\rightsquigarrow$ are independent of $\boldsymbol{\lambda}$. Accordingly, we often view them as relations on $\mathrm{row}(A)$.

Definition 2 (Canonical Component Form). *Let $A_1|\boldsymbol{\lambda}_1 \sqcup \ldots \sqcup A_c|\boldsymbol{\lambda}_c$ be the partitioning of* $\mathrm{row}(A|\boldsymbol{\lambda})$ *with respect to $\rightsquigarrow$. The component form (CF) of $A|\boldsymbol{\lambda}$ with respect to an arbitrary ordering $(A_{i_1}|\boldsymbol{\lambda}_{i_1}, \ldots, A_{i_c}|\boldsymbol{\lambda}_{i_c})$ is defined as the block matrix*

$$\mathrm{CF}(A|\boldsymbol{\lambda}) := \begin{pmatrix} \overline{A}_{i_1} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\lambda}_{i_1} \\ \mathbf{0} & \overline{A}_{i_2} & \cdots & \mathbf{0} & \boldsymbol{\lambda}_{i_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \overline{A}_{i_c} & \boldsymbol{\lambda}_{i_c} \end{pmatrix}$$

$A|\boldsymbol{\lambda}$ can have several component forms. Unless stated otherwise, we always assume that the system $A|\boldsymbol{\lambda}$ is in some component form, for if not, it can be placed in CF by a swapping of rows and columns.

Definition 3 (Acyclic System). *Any system $A|\boldsymbol{\lambda}$ is said to be cyclic if and only if the corresponding system-graph $G(A|\boldsymbol{\lambda})$ is cyclic, and acyclic otherwise.*

The following proposition is a trivial consequence of the acyclic nature of the system-graph.

Proposition 6. *Any acyclic system has full row-rank.*

See Example 1 for a short explanation on the notations and definitions introduced thus far.

Example 1. Consider the following system of 6 equations in 15 variables over $\mathbb{F}_N$:

$$A|\boldsymbol{\lambda} = \left(\begin{array}{ccccccccccccccc|c} \alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 \\ 0 & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \\ \alpha_7 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_8 & \alpha_9 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \\ 0 & 0 & 0 & 0 & \alpha_{10} & 0 & \alpha_{11} & 0 & 0 & 0 & 0 & \alpha_{12} & 0 & 0 & 0 & \lambda_4 \\ 0 & 0 & \alpha_{13} & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{14} & \alpha_{15} & 0 & 0 & 0 & 0 & \lambda_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{16} & \alpha_{17} & \alpha_{18} & \lambda_6 \end{array}\right)$$

for non-zero $\alpha_1, \ldots, \alpha_{18} \in \mathbb{F}_N$. The corresponding system-graph is illustrated in Figure 1.

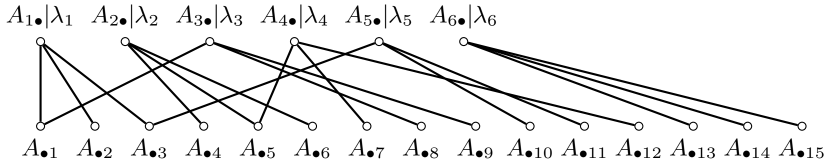

Fig. 1. The system-graph corresponding to the system in Example 1. The edge labels are omitted for readability.

Here,

- $A_{3\bullet}|\lambda_3 \sim A_{1\bullet}|\lambda_1 \sim A_{5\bullet}|\lambda_5$ giving $A_1|\boldsymbol{\lambda}_1 = \{A_{1\bullet}|\lambda_1, A_{3\bullet}|\lambda_3, A_{5\bullet}|\lambda_5\}$,
- $A_{2\bullet}|\lambda_2 \sim A_{4\bullet}|\lambda_4$ giving $A_2|\boldsymbol{\lambda}_2 = \{A_{2\bullet}|\lambda_2, A_{4\bullet}|\lambda_4\}$, and
- $A_{6\bullet}|\lambda_6 \sim A_{6\bullet}|\lambda_6$ giving $A_3|\boldsymbol{\lambda}_3 = \{A_{6\bullet}|\lambda_6\}$,

resulting in the following component form:

$$\left(\begin{array}{ccc|c} \overline{A}_1 & \mathbf{0} & \mathbf{0} & \boldsymbol{\lambda}_1 \\ \mathbf{0} & \overline{A}_2 & \mathbf{0} & \boldsymbol{\lambda}_2 \\ \mathbf{0} & \mathbf{0} & \overline{A}_3 & \boldsymbol{\lambda}_3 \end{array}\right) = \left(\begin{array}{ccccccccccccccc|c} \alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 \\ \alpha_7 & 0 & 0 & \alpha_8 & \alpha_9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \\ 0 & 0 & \alpha_{13} & 0 & 0 & \alpha_{14} & \alpha_{15} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_5 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{10} & 0 & \alpha_{11} & \alpha_{12} & 0 & 0 & 0 & \lambda_4 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{16} & \alpha_{17} & \alpha_{18} & \lambda_6 \end{array}\right)$$

The resulting system $\mathrm{CF}(A|\boldsymbol{\lambda})$ is acyclic and same as $A|\boldsymbol{\lambda}$ up to a relabeling of variables and constants. Furthermore, one of the components $A_3|\boldsymbol{\lambda}_3$ is isolated, although the overall system itself is non-isolated.

Solutions to a System of Equations: Let $\eta(A|\boldsymbol{\lambda})$ denote the number of solutions to the system $A|\boldsymbol{\lambda}$. Throughout we assume that the system is consistent, i.e., $\mathrm{rank}(A|\boldsymbol{\lambda}) = \mathrm{rank}(A)$, otherwise $\eta(A|\boldsymbol{\lambda}) = 0$.

The component form of a system gives a very simple product rule for the number of solutions:

$$\eta(A|\boldsymbol{\lambda}) = \prod_{i=1}^{c} \eta(\overline{A}_i|\boldsymbol{\lambda}_i), \tag{7}$$

which stems from the simple observation that any two components are completely disjoint, i.e., involve distinct variables.

Definition 4 (Constrained System). *For any positive integers q, r, t such that $q, t < r$, a (q, r, t)-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}; \mathsf{P})$ over $\mathbb{F}_N$ is the system $A|\boldsymbol{\lambda}$ of q equations in r variables, over $\mathbb{F}_N$, endowed with an equivalence relation P on $\mathrm{col}(A)$ resulting in the partition $\mathrm{col}(A) = \mathsf{P}_1 \sqcup \ldots \sqcup \mathsf{P}_t$.*

The dimension and rank of $\mathbb{S}$, denoted $\dim(\mathbb{S})$ and $\mathrm{rank}(\mathbb{S})$, are simply the dimension and rank of A, respectively.

For what follows, we fix a (q, r, t)-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$, where $A|\boldsymbol{\lambda}$ is in a component form. Whenever convenient, we drop P from the notation.

Since $\mathbb{S}$ is effectively a system of equations, all the notations and notions are analogously extended unless stated otherwise, except for a minor change in the definition of the system-graph $G(\mathbb{S})$ associated with $\mathbb{S}$ which is now endowed with an implicit coloring of the vertices $\mathrm{col}(A)$ that has a one to one correspondence with P. More precisely, for any $i \in [t]$, any two columns $A_{\bullet j}, A_{\bullet j'} \in \mathsf{P}_i$ share the same implicit color.

The ordered sequence $(\mathbb{S}_1 \prec \cdots \prec \mathbb{S}_c)$ denotes the component form of $\mathbb{S}$, denoted $\mathrm{CF}(\mathbb{S})$, where each $\mathbb{S}_i$ is the (q_i, r_i, t_i)-constrained system $(\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P}^{(i)})$, with $\mathsf{P}^{(i)} \subseteq \mathsf{P}$ being the equivalence relation on the set $\mathrm{col}(\overline{A}_i) \subseteq [r]$, that partitions $\mathrm{col}(\overline{A}_i)$ into t_i subsets $\mathsf{P}^{(i)}_1, \ldots, \mathsf{P}^{(i)}_{t_i}$.

$\mathbb{S}$ is said to be:

- a *clique* iff for all $j, j' \in \mathrm{col}(A)$, $(j, j') \in \mathsf{P}$.
- a *partite* iff for all $A_{i\bullet} \in \mathrm{row}(A)$, and for all $j, j' \in \mathrm{col}(\overline{A}_{i\bullet})$, $(j, j') \notin \mathsf{P}$.

Since $\mathsf{P}^{(i)} \subseteq \mathsf{P}$, for brevity we continue to use P instead of $\mathsf{P}^{(i)}$ for all i. Wlog we also assume that $\mathbb{S}$ is in component form or simply CF.

See Example 2 for an explanation on the notations and definitions related to constrained systems.

Example 2. Recall Example 1, and endow the system $A|\boldsymbol{\lambda}$ with an implicit equivalence relation P (as evident from the updated system-graph illustrated in Figure 2), resulting in the partition $\mathrm{col}(A) = \mathsf{P}_1 \sqcup \mathsf{P}_2 \sqcup \mathsf{P}_3$, where $\mathsf{P}_i = \{j \in [15] : j \equiv i \pmod 3\}$ for all $i \in [3]$.

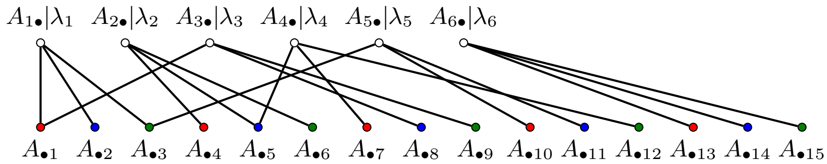

Fig. 2. The system-graph corresponding to the constrained system in Example 2. Yet again the edge labels are omitted for readability.

For the $(6, 15, 3)$-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$, we have

- $\dim(\mathbb{S}) = 6 \times 15$, $\mathrm{rank}(\mathbb{S}) = 6$,
- $\mathrm{CF}(\mathbb{S}) = (\mathbb{S}_1 \prec \mathbb{S}_2 \prec \mathbb{S}_3)$, where $\mathbb{S}_i = (\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P})$,
- $\mathbb{S}_3$ is isolated, but $\mathbb{S}$ is not, and
- $\mathbb{S}$ is acyclic and partite.

4 Solutions to a Constrained System

Definition 5 (Solution to a Constrained System). *For a family of sets $\mathcal{R} = \{\mathcal{R}_i \subseteq \mathbb{F}_N\}_{i \in [t]}$, any $\boldsymbol{y} = (y_1, \ldots, y_r) \in \mathbb{F}_N^r$ is said to be an $\overline{\mathcal{R}}$-solution for $\mathbb{S}$ if and only if the following conditions are satisfied:*

1. *$\boldsymbol{y}$ satisfies the system $A|\boldsymbol{\lambda}$,*
2. *for any $i \in [t]$, and any $j \in \mathsf{P}_i$, $y_j \notin \mathcal{R}_i$,*
3. *for any $i \in [t]$, and any $j \neq j' \in \mathsf{P}_i$, $y_j \neq y_{j'}$.*

In words, all elements in $\mathcal{R}_1, \ldots, \mathcal{R}_t$ are forbidden. In this context, $\mathcal{R}_i$ are referred as *forbidden sets*. Furthermore any two distinct P-related variables[5] must have distinct values.

Let $(\mathbb{S}|\mathcal{R})$ denote the $\overline{\mathcal{R}}$-solution space of $\mathbb{S}$ and $\eta(\mathbb{S}|\mathcal{R}) := |(\mathbb{S}|\mathcal{R})|$, the number of $\overline{\mathcal{R}}$-solutions of $\mathbb{S}$. The central problem that we study in this work is to find a good lower bound on $\eta(\mathbb{S}|\mathcal{R})$ under some assumptions on A, $\boldsymbol{\lambda}$ and $\mathcal{R}$.

Fix a (q, r, t)-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}; \mathsf{P})$ and a family of sets $\mathcal{R} = \{\mathcal{R}_i\}_{i \in [t]}$. Fix a component form $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$ for $\mathbb{S}$. For any $(i, j) \in [c] \times [t]$, let $r_i^{(j)} := \left|\mathrm{col}(\overline{A}_i) \cap \mathsf{P}_j\right|$, and define $r^{(j)} = \sum_{i=1}^{c} r_i^{(j)}$.
We will often use the following basic property which is easily verifiable using Definition 5.

Fact 1 *For any (q, r, t)-system $\mathbb{S}$ and any $\mathcal{R}' = (\mathcal{R}'_1, \ldots, \mathcal{R}'_t)$, where $\mathcal{R}'_i \subseteq \mathcal{R}_i$ for all $i \in [t]$, $\eta(\mathbb{S}|\mathcal{R}) \leq \eta(\mathbb{S}|\mathcal{R}')$.*

Without loss of generality, we assume $|\mathcal{R}_i| = s_i \leq s$ for some $s < N$, or else, $(\mathbb{S}|\mathcal{R}) = \varnothing$. Then, under the assumption that $\boldsymbol{\lambda}$ is uniform at random, one would expect that the number of $\overline{\mathcal{R}}$-solutions for $\mathbb{S}$ is approximately

$$\mathbb{E}\left(\mathbb{S}|\mathcal{R}\right) := \frac{\prod_{j=1}^{t} (N - s_i)_{r^{(j)}}}{N^q} \tag{8}$$

Of course, the assumption and the expression are both quite speculative at a first glance. However, as we show later, $\eta(\mathbb{S}|\mathcal{R})$ is very close to $\mathbb{E}\left(\mathbb{S}|\mathcal{R}\right)$ for a large class of constrained systems. Indeed, for certain binary matrices A and $\mathcal{R} = \varnothing$ case, Cogliati et al. prove [13] exactly this result. We aim to prove it in a more general setting where $\mathcal{R}$ may not be empty.

While tackling the problem in its full generality is an interesting and technically challenging endeavor, it might not captivate the general cryptography community. Instead, we focus on a specific class of constrained systems that includes, among other things, known instances in symmetric cryptography, particularly those discussed in this paper.

Definition 6 (Weight). *The weight of any $A \in \mathbb{F}_N^{q \times r}$ is defined as*

$$\mathtt{H}(A) := \min\{\mathtt{H}(\boldsymbol{v}) : \boldsymbol{v} \in \mathrm{rowsp}^+(A)\},$$

where $\mathrm{rowsp}^+(A) := \{a_1 A_{1\bullet} \oplus \cdots \oplus a_q A_{q\bullet} : \forall (a_1, \ldots, a_q) \neq \mathbf{0}\}$ *and* $\mathtt{H}(\boldsymbol{v})$ *denotes the number of non-zero coordinates in v.*

[5] The equivalence relation P on $\mathrm{col}(A)$ can be equivalently defined over the set of variables $\{x_1, \ldots, x_r\}$.

We have the following fact that relates the weight of a matrix (and its components) with its row rank.

Proposition 7. *Suppose* $A \in \mathbb{F}_N^{q\times r}$ *has* $\mathtt{H}(A) = k > 0$*. Then,*
(1) *A has full row rank.*
(2) *for every $r' \geq r - k + 1$ and $1 \leq i_1 < \cdots < i_{r'} \leq r$, the matrix $A' = (A_{\bullet i_1}|\cdots|A_{\bullet i_{r'}})$ has full row rank, where $A_{\bullet i}$ denotes the i-th column of A viewed as a q-dimensional vector.*
(3) *$r - k + 1 \geq q$.*

Proof. (1) follows from the definition. For (2), suppose to the contrary that A' does not have full rank. Then, we must have $\mathbf{0} \in \text{rowsp}^+(A')$. Specifically, one can find $(a_1, \ldots, a_q) \neq \mathbf{0} \in \mathbb{F}_2^q$, such that $a_1 A'_{1\bullet} \oplus \cdots \oplus a_q A'_{q\bullet} = \mathbf{0}$. Then, $\boldsymbol{v} = a_1 A_{1\bullet} \oplus \cdots \oplus a_q A_{q\bullet} \in \text{rowsp}^+(A)$, and $\mathtt{H}(\boldsymbol{v}) \leq r - r' \leq k - 1$. Thus, $\mathtt{H}(A) < k$, which is a contradiction. Finally, (3) follows from (2). ◻

Looking ahead momentarily the higher the weight of a system, the closer our bound to $\mathbb{E}(\mathbb{S}|\mathcal{R})$, and point (2) and (3) of Proposition 7 play a crucial role towards establishing this fact. The following definition and subsequent results provide an easy-to-check condition for determining the weight of a matrix.

Definition 7 (Regularity). *Any* $A \in \mathbb{F}_N^{q\times r}$ *is said to be k-regular if and only if* $\mathtt{H}(A_{i\bullet}) = k$*, for all* $i \in [q]$*.*

Note that, the above definition can be equivalently formulated as $\text{row}(A|\boldsymbol{\lambda})$ is regular[6] in $G(A|\boldsymbol{\lambda})$. The following propositions show that acyclic and highly regular systems have high weight.

Proposition 8. *For any $k \geq 2$, any k-regular and acyclic* $A \in \mathbb{F}_N^{q\times r}$ *has* $\mathtt{H}(A) = k$*.*

Proof. The result is trivial for $q = 1$. Assume for contradiction that $\mathtt{H}(A) < k$ for some $q \geq 2$. Then, for some $2 \leq l \leq q$, there exists a sequence of rows $A_{i_1\bullet}, \ldots, A_{i_l\bullet}$ and a sequence of non-zero field elements $a_1, \ldots, a_l$, such that $\boldsymbol{v} = a_1 A_{i_1\bullet} \oplus \ldots \oplus a_l A_{i_l\bullet}$ has $H(\boldsymbol{v}) < k$. Since, A is acyclic, one can always find two distinct rows $A_{i_a\bullet}$ and $A_{i_b\bullet}$ such that there exists at most one $A_{i_c\bullet} \in \{A_{i_1\bullet}, \ldots, A_{i_l\bullet}\} \setminus A_{i_a\bullet}$ and one $A_{i_d\bullet} \in \{A_{i_1\bullet}, \ldots, A_{i_l\bullet}\} \setminus A_{i_b\bullet}$ such that $A_{i_a\bullet} \sim A_{i_c\bullet}$ and $A_{i_b\bullet} \sim A_{i_d\bullet}$, respectively. For if not, then due to the finiteness of l, the matrix

$$\begin{pmatrix} A_{i_1\bullet} \\ \vdots \\ A_{i_l\bullet} \end{pmatrix}$$

is cyclic which contradicts the acyclic nature of A. Then, using the k-regularity of A, at least $k - 1 \geq 1$ non-zero columns in each of $A_{i_a\bullet}$ and $A_{i_b\bullet}$ have a single non-zero entry. Therefore, these columns contribute non-zero coordinates to $\boldsymbol{v}$. Thus, $H(\boldsymbol{v}) \geq 2k - 2$ which is at least k for $k \geq 2$. ◻

[6] A vertex set is said to be regular if all the constituent vertices have the same degree.

Proposition 9. *For any $q \geq 2$ and any $k \geq 3$, let $A \in \mathbb{F}_N^{q\times r}$ be acyclic and k-regular. Then, for any $1 \leq i_1 < \ldots < i_k \leq r$, the matrix $A' = A \setminus \{A_{\bullet i_1}, \ldots, A_{\bullet i_k}\}$ has:*

$$\mathrm{rank}(A') = \begin{cases} q-1 & \text{if } \{i_1,\ldots,i_k\} = \mathrm{col}(\overline{A}_{j\bullet}) \text{ for some } j \in [q], \\ q & \text{otherwise.} \end{cases}$$

Proof. First consider the case: $\{i_1,\ldots,i_k\} = \mathrm{col}(\overline{A}_{j\bullet})$ for some $j \in [q]$, i.e., all the non-zero columns of $A_{j\bullet}$ are deleted, and hence $A_{j\bullet}$ can be dropped without affecting the rank of A'. Thus, $\mathrm{rank}(A') \leq q-1$. Furthermore, since the system is acyclic and A is k-regular, A' must be acyclic and at least $(k-1)$-regular. Then, using Proposition 8, we have $\mathtt{H}(A') \geq k-1 \geq 2$, and thus using Proposition 7, $\mathrm{rank}(A') = q-1$.

Now suppose $\{i_1,\ldots,i_k\} \neq \mathrm{col}(\overline{A}_{j\bullet})$ for all $j \in [q]$. Thus, A' has q non-zero rows. Assume towards a contradiction that $\mathrm{rank}(A') < q$. Then one can find a sequence of distinct rows $A'_{j_1\bullet}, A'_{j_2\bullet}, \ldots, A'_{j_l\bullet} \in \mathrm{row}(A')$ and a sequence of non-zero coefficients $a_1, a_2, \ldots, a_l$ such that $\boldsymbol{v} = a_1 A'_{j_1\bullet} \oplus \ldots \oplus a_l A'_{j_l\bullet} = \mathbf{0}$. Let

$$A'' = \begin{pmatrix} A_{j_1\bullet} \\ A_{j_2\bullet} \\ \vdots \\ A_{j_l\bullet} \end{pmatrix}$$

We claim that the number of columns in A'' with a single non-zero entry in each of these columns is at least $2k-2$. Indeed, in the worst case, all the rows are connected to each other. So after a relabeling of rows one can find a sequence $A_{j'_1\bullet} \sim A_{j'_2\bullet} \sim \ldots \sim A_{j'_{l'}\bullet}$ for some $l' \leq l$. Since A'' is acyclic and k-regular, $A_{j'_1\bullet}$ and $A_{j'_{l'}\bullet}$ contribute at least $k-1$ columns each with a single non-zero entry. Now, even if one deletes k columns from A'', there are still at least $k-2 \geq 1$ columns that contribute non-zero entries in any linear combination, including $\boldsymbol{v} = a_1 A'_{j_1\bullet} \oplus \ldots \oplus a_l A'_{j_l\bullet}$. Therefore, $v \neq \mathbf{0}$, contradicting $\mathrm{rank}(A') < q$. □

Column-Uniform System: Any k-regular system $\mathbb{S} = (A|\boldsymbol{\lambda}; \mathsf{P})$ is said to be *column-uniform system*, if there exists a unique sequence (or set) of non-zero coefficients $\alpha_{\mathbb{S}} = (\alpha_1, \ldots, \alpha_k : \alpha_i)$ such that:

1. for each column j, there exists a unique $\alpha_{(j)} := \alpha_l$ for some $l \in [k]$, such that for all row i of A the following condition holds:

$$A_{ij} = \begin{cases} \alpha_{(j)} & \text{if } A_{ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

2. for all $i \in [q]$, the i-th equation in $\mathbb{S}$ is of the form $\alpha_1 x_{j_1} \oplus \cdots \oplus \alpha_k x_{j_k} = \lambda_i$, where $\{j_1, \ldots, j_k\} = \mathrm{col}(\overline{A}_{i\bullet})$.
3. for $k \geq 2$, if $j_1, j_2 \in \mathsf{P}_l$ then $\alpha_{(j_1)} = \alpha_{(j_2)}$.

While the above definition looks quite artificial, it is remarkably satisfied by most of the known instances of constrained systems in symmetric-key literature, including, for instance, the sum of k permutations and `DBHtS`.

In this paper, *we focus on lower bounding $\eta(\mathbb{S}\,|\,\mathcal{R})$ for column-uniform, acyclic and k-regular* (or k-CAR) *system* $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$.

Additional Notations and Conventions: Without loss of generality assume a component form $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$, such that all the isolated components appear before the non-isolated ones. Let $\mathtt{NI}(\mathbb{S})$ denote the set of indices of all the non-isolated components, $\xi_{\mathbb{S}} := \max\{r_i : i \in [c]\}$, $\Delta_{\mathbb{S}} := \max_d |\{i \in [q] : \lambda_i = d\}|$, and for any $i \in [c]$, let:

- $\mathbb{S}_{\le i}$ denote the system $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_i)$,
- $\boldsymbol{y}_{\le i}$ denote the solution of the sub-system $\mathbb{S}_{\le i}$,
- $\mathcal{P}$ and $\mathcal{F}$ define families of set indexed by $j \in [t]$ such that

$$\mathcal{P}_j(\boldsymbol{y}_{\le i}) := \{y_k \in \boldsymbol{y}_{\le i} : k \in \mathsf{P}_j\} \quad \text{and} \quad \mathcal{F}_j(\boldsymbol{y}_{\le i}) := \mathcal{R}_j \sqcup \mathcal{P}_j(\boldsymbol{y}_{\le i}).$$

Let $|\mathcal{P}_j(\boldsymbol{y}_{\le i})| := r_{\le i}^{(j)}$ and $|\mathcal{F}_j(\boldsymbol{y}_{\le i})| = f_{\le i}^{(j)} := s_j + r_{\le i}^{(j)}$.
Extending the notation for $i = 0$, let $\boldsymbol{y}_{\le 0}$ denote any empty sequence, and thus, $\mathcal{P}_j(\boldsymbol{y}_{\le 0}) = \varnothing$ and $\mathcal{F}_j(\boldsymbol{y}_{\le 0}) = \mathcal{R}_j$. In addition, for the sake of convenience, we also assume that $0 \in \mathcal{R}_j$ for all $j \in [t]$. Note that, $r_{\le i}^{(j)}$ and hence $f_{\le i}^{(j)}$ are independent of the actual elements in $\mathcal{P}_j(\boldsymbol{y}_{\le i})$ and $\mathcal{F}_j(\boldsymbol{y}_{\le i})$, respectively. In particular, we have $r_{\le i}^{(j)} \le q$, as each equation can have at most one variable in P_j, and thus, $f_{\le i}^{(j)} \le s_j + q \le s + q$.

4.1 The Case of CAR Partite System

For any t-CAR and partite (t-CARP) (q, r, t)-system $\mathbb{S}$, we have the obvious[7] bijective map $\alpha_j \longmapsto \mathsf{P}_j$. With this in mind, we define three families of sets $\hat{\mathcal{R}}$, $\hat{\mathcal{P}}$ and $\hat{\mathcal{F}}$ indexed by $j \in [t]$ such that

$$\begin{aligned} \hat{\mathcal{R}}_j &:= \alpha_j \cdot \mathcal{R}_j \\ \hat{\mathcal{P}}_j(\boldsymbol{y}_{\le i}) &:= \{\alpha_j \cdot y_k \in \boldsymbol{y}_{\le i} : k \in \mathsf{P}_j\} \\ \hat{\mathcal{F}}_j(\boldsymbol{y}_{\le i}) &:= \hat{\mathcal{R}}_j \sqcup \hat{\mathcal{P}}_j(\boldsymbol{y}_{\le i}). \end{aligned}$$

It is obvious that $|\hat{\mathcal{R}}_j| = s_j$, $|\hat{\mathcal{P}}_j(\boldsymbol{y}_{\le i})| = r_{\le i}^{(j)}$ and $|\hat{\mathcal{F}}_j(\boldsymbol{y}_{\le i})| = f_{\le i}^{(j)}$.

Theorem 1 (Partite Bound). *Let $t \ge 2$, and $\mathcal{R}$ be a family of sets. For any (q, r, t)-constrained system $\mathbb{S}$ which is t-CARP and satisfies $\xi_{\mathbb{S}}(s + q) \le N/2$, we have $\eta(\mathbb{S}\,|\,\mathcal{R}) \ge (1 - \varepsilon)\,\mathbb{E}\,(\mathbb{S}\,|\,\mathcal{R})$, where*

$$\varepsilon \le \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}, \mathcal{R})}{N^{t-1}} + \frac{2q\Delta_{\mathbb{S}}}{N^{t-1}} + \frac{6q(s+q)^t}{N^t} + \sum_{i \in \mathtt{NI}(\mathbb{S})} \left(\frac{2r_i^t(s+q)^t}{N^t} + \varepsilon_{\text{odd}}(q, r, s, t) \right),$$

where

$$\varepsilon_{\text{odd}}(q, r, s, t) = \begin{cases} \frac{2q_i(s+q)^{t-1}}{N^{t-1}} & \textit{for odd } t, \\ 0 & \textit{for even } t. \end{cases}$$

[7] This is made obvious by a (possible) relabeling of coefficient indices.

A proof of this result is derived in two stages. First, in Lemma 4, we derive an initial bound that would be useful when the local[8] error terms can be shown to be sufficiently small in expectation for a random constrained system. We then go on to derive a bound on the global error term which completes the proof of the aforementioned theorem.

Consider the i-th component $\mathbb{S}_i = (\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P})$. Since $\mathbb{S}$ is in CF, $\mathrm{col}(\overline{A}_i) = \{r_{\le(i-1)}+1,\ldots,r_{\le(i-1)}+t\}$, where $r_{\le(i-1)} = r_1+\ldots+r_{i-1}$. For brevity, we ignore the $r_{\le(i-1)}$ shift in indexing.

Now, towards a proof of Theorem 1, observe that

$$\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R}) = \sum_{\boldsymbol{y}_{\le(i-1)}} \eta(\mathbb{S}_i\,|\,\mathcal{F}(\boldsymbol{y}_{\le(i-1)})), \tag{9}$$

For a fixed $\boldsymbol{y}_{\le(i-1)}$, the set of $\overline{\mathcal{R}}$-solutions to $\mathbb{S}_i$ is given by

$$(\mathbb{S}_i\,|\,\mathcal{F}) \coloneqq \{y = (y_1,\ldots,y_{r_i}) \in \overline{\mathcal{F}}_{(1)} \times \ldots \times \overline{\mathcal{F}}_{(r_i)} : \overline{A}_i y = \boldsymbol{\lambda}_i\},$$

where, for all $j \in [r_i]$, $\mathcal{F}_{(j)} \coloneqq \mathcal{F}_k(\boldsymbol{y}^{\le i-1})$ for a unique $k \in [t]$. Let $f_{(j)} = |\mathcal{F}_{(j)}|$, and thus $f_{(j)} = f^{(k)}_{\le(i-1)}$ for a unique $k \in [t]$. Let $\mathcal{A}_\varnothing \coloneqq \{y \in \mathbb{F}_N^{r_i} \;:\; \overline{A}_i y = \boldsymbol{\lambda}_i\}$. Moreover, for each $j \in [r_i]$, we define

$$\mathcal{A}_{\{j\}} \coloneqq \mathcal{A}_\varnothing \bigcap (\mathbb{F}_N^{j-1} \times \mathcal{F}_{(j)} \times \mathbb{F}_N^{t-j}).$$

Then, we have

$$(\mathbb{S}_i\,|\,\mathcal{F}) = \mathcal{A}_\varnothing \setminus \left(\bigcup_{j\in\mathrm{col}(\overline{A}_i)} \mathcal{A}_{\{j\}} \right).$$

For any non-empty $\mathcal{J} \subseteq \mathrm{col}(\overline{A}_i)$, let $\mathcal{A}_\mathcal{J} \coloneqq \cap_{j\in\mathcal{J}} \mathcal{A}_{\{j\}}$. Using the principal of inclusion-exclusion, we have

$$\begin{aligned} \eta(\mathbb{S}_i\,|\,\mathcal{F}) &= |\mathcal{A}_\varnothing| - \left| \left(\bigcup_{j\in[r_i]} \mathcal{A}_{\{j\}} \right) \right| \\ &= \sum_{\mathcal{J}\subseteq[r_i]} (-1)^{|\mathcal{J}|} |\mathcal{A}_\mathcal{J}| \end{aligned} \tag{10}$$

Now, $|\mathcal{A}_\varnothing| = N^{r_i-q_i}$ follows from elementary linear algebra; In fact, by virtue of $\mathbb{S}$ being an acyclic and t-regular system, Proposition 8 and 7 allows for an analogous argument to prevail for any $\mathcal{A}_\mathcal{J}$ with $|\mathcal{J}| \le t-1$. In particular, for any $\mathcal{J} = \{l_1,\ldots,l_{|\mathcal{J}|}\}$, and any $y_\mathcal{J} = (y_{l_1},\ldots,y_{l_{|\mathcal{J}|}}) \in \mathcal{F}_{(l_1)} \times \ldots \times \mathcal{F}_{(l_{|\mathcal{J}|})}$, we obtain an equation in exactly $r_i - |\mathcal{J}| \ge r_i - t + 1 \ge q_i$ variables, which has exactly $N^{r_i-|\mathcal{J}|-q_i}$ solutions. There are exactly $f_{(\mathcal{J})} = f_{(l_1)} \cdots f_{(l_{|\mathcal{J}|})}$ such $y_\mathcal{J}$. Thus, we have $|\mathcal{A}_\mathcal{J}| = f_{(\mathcal{J})} \cdot N^{t-|\mathcal{J}|-q_i}$ for all $\mathcal{J} \subset [t]$.

[8] The adjective “local” here corresponds to individual components.

Crude Bound: Digressing a little, from (10) and the above discussion, we have

$$N^{r_i-q_i} - r_i(s+q)N^{r_i-q_i-1} \leq \eta(\mathbb{S}_i \,|\, \mathcal{F}) \leq N^{r_i-q_i}$$

for any acyclic system $\mathbb{S}$ where we use the fact $f_{(j)} \leq (s+q)$ for all $j \in [r_i]$. This along with (9) gives the following crude bound.

Fact 2 *For any acyclic (q,r,t)-system and any $\mathcal{R} = (\mathcal{R}_1,\ldots,\mathcal{R}_t)$, we have*

$$N^{r_i-q_i-1}\left(N - r_i(s+q)\right)\eta(\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R}) \leq \eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \leq N^{r_i-q_i}\eta(\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R}) \qquad (11)$$

Now coming back to (10) for a proof of Theorem 1, we study the right hand side separately for isolated and non-isolated components, starting with an isolated component.

Lemma 3. *Suppose $\mathbb{S}_i$ is isolated. Then, for any $y_{\leq(i-1)} \in (\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R})$, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \geq \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(1 + (-1)^t\frac{2}{N^{t-1}}\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\leq(i-1)}}{N}\right)\right),$$

where $f^{([t])}_{\leq(i-1)} = f^{(1)}_{\leq(i-1)}\cdot\ldots\cdot f^{(t)}_{\leq(i-1)}$.

Proof. Since $\mathbb{S}_i$ is t-regular, partite and isolated, $r_i = t$ and $q_i = 1$. Then, recall from (10) and the subsequent discussion

$$\begin{aligned}
\eta(\mathbb{S}_i \,|\, \mathcal{F}) &= \sum_{\mathcal{J}\subseteq[t]}(-1)^{|\mathcal{J}|}|\mathcal{A}_{\mathcal{J}}| \\
&= \sum_{\mathcal{J}\subset[t]}(-1)^{|\mathcal{J}|}f_{(\mathcal{J})}N^{t-|\mathcal{J}|-1} + (-1)^t\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) \\
&= \frac{1}{N}\left(\sum_{\mathcal{J}\subset[t]}(-1)^{|\mathcal{J}|}f_{(\mathcal{J})}N^{t-|\mathcal{J}|} + f_{([t])} - f_{([t])} + (-1)^t N\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F})\right) \\
&= \frac{1}{N}\left(\prod_{j=1}^{t}(N - f_{(j)}) + (-1)^t N\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f_{([t])}}{N}\right)\right) \\
&= \frac{1}{N}\left(\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)}) + (-1)^t N\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\leq(i-1)}}{N}\right)\right) \\
&\geq \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(1 + (-1)^t\frac{N}{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\leq(i-1)}}{N}\right)\right) \\
&\geq \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(1 + (-1)^t\frac{2}{N^{t-1}}\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\leq(i-1)}}{N}\right)\right), \qquad (12)
\end{aligned}$$

where the second equality is due to (2), the fifth equality is from a simple relabeling, and the last inequality follows from the fact that $f^{(j)}_{\leq(i-1)} \leq (s+q)$ and $t(s+q) \leq \xi_{\mathbb{S}}(s+q) \leq N/2$. ◻

Now, on to a lower bound on $\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R})$ for isolated $\mathbb{S}_i$.

Lemma 4. *Suppose $\mathbb{S}_i$ is isolated. Then, we have*

$$\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R}) \ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N}\left(1 - \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R})}{N^{t-1}} - \frac{2\Delta_{\mathbb{S}}}{N^{t-1}} - \frac{6(s+q)^t}{N^t}\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}).$$

Proof. From (9) and Lemma 3, we have

$$\begin{aligned}
\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R}) &= \sum_{\boldsymbol{y}_{\le(i-1)}} \eta(\mathbb{S}_i\,|\,\mathcal{F}(\boldsymbol{y}_{\le(i-1)}))\\
&\ge \sum_{\boldsymbol{y}_{\le(i-1)}} \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N}\left(1 + (-1)^t\frac{2}{N^{t-1}}\left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\le(i-1)}}{N}\right)\right)\\
&\ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N}\left(\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \frac{2f^{([t])}_{\le(i-1)}}{N^t}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \frac{2}{N^{t-1}}\sum_{\boldsymbol{y}_{\le(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F})\right)\\
&\ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N}\left(\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \frac{2f^{([t])}_{\le(i-1)}}{N^t}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \frac{2}{N^{t-1}}\sum_{\boldsymbol{y}_{\le(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F})\right)
\end{aligned} \tag{13}$$

Claim. We claim

$$\sum_{\boldsymbol{y}_{\le(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) \le \left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R}) + \Delta_{\mathbb{S}} + \frac{2(s+q)^t}{N}\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R})$$

Proof. We have

$$\begin{aligned}
\sum_{\boldsymbol{y}_{\le(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) &= \sum_{\boldsymbol{y}_{\le(i-1)}}\sum_{\mathcal{I}\subseteq[t]}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_{\mathcal{I}},\hat{\mathcal{R}}_{[t]\setminus\mathcal{I}})\\
&= \sum_{\mathcal{I}\subseteq[t]}\sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_{\mathcal{I}},\hat{\mathcal{R}}_{[t]\setminus\mathcal{I}})
\end{aligned}$$

where $\hat{\mathcal{P}}_{\mathcal{I}} = \hat{\mathcal{P}}_{j_1}\times\ldots\times\hat{\mathcal{P}}_{j_m}$ and $\hat{\mathcal{R}}_{[t]\setminus\mathcal{I}} = \hat{\mathcal{R}}_{k_1}\times\ldots\times\hat{\mathcal{R}}_{k_{m'}}$ for every $\mathcal{I} = \{j_1,\ldots,j_m\}$ and $[t]\setminus\mathcal{I} = \{k_1,\ldots,k_{m'}\}$. For brevity we simply write $\mathcal{I} = [m]$. Consider the following two cases:

- Case A: $\mathcal{I} = \emptyset$. In this case the definition straightaway gives

$$\sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_{[t]}) = \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R})\times\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}).$$

  We remark that for $i = 1$ this is the only possible case.

- Case B: $\mathcal{I} \ne \emptyset \subseteq [t]$. Fix some $(a_{t-m+1},\ldots,a_t) \in \hat{\mathcal{R}}_{[t]\setminus\mathcal{I}}$ and define $a_\oplus := a_{t-m+1}\oplus\ldots\oplus a_t$, with $a_\oplus = 0$ whenever $\mathcal{I} = [t]$. Fix some $(y_{l_1},\ldots,y_{l_m}) \in \hat{\mathcal{P}}_1\times\ldots\times\hat{\mathcal{P}}_m$. Then, we have

$$\sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,y_{l_1},\ldots,y_{l_m},a_{t-m+1},\ldots,a_t) = \sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i\ominus a_\oplus,y_{l_1},\ldots,y_{l_m}) \tag{14}$$

Thus, we want to count the number of solutions for $\mathbb{S}_{\le(i-1)}$ that additionally satisfies the equation $\alpha_{l_1} \cdot \boldsymbol{x}_{l_1} \oplus \ldots \oplus \alpha_{l_m} \cdot \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \ominus a_\oplus$.
Let $\mathbb{S}'_{\le(i-1)} = \mathbb{S}_{\le(i-1)} \cup \{\alpha_{l_1} \cdot \boldsymbol{x}_{l_1} \oplus \ldots \oplus \alpha_{l_m} \cdot \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \ominus a_\oplus\}$ be the constrained system $\mathbb{S}_{\le(i-1)}$ extended with the additional equation $\alpha_{l_1} \cdot \boldsymbol{x}_{l_1} \oplus \ldots \oplus \alpha_{l_m} \cdot \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \ominus a_\oplus$. Then, by definition, we have

$$\sum_{\boldsymbol{y}_{\le(i-1)}} \mu(\boldsymbol{\lambda}_i \ominus a_\oplus, y_{l_1}, \ldots, y_{l_m}) = \eta(\mathbb{S}'_{\le(i-1)} \mid \mathcal{R}).$$

Let $A'_{\le(i-1)}$ denote the corresponding coefficient matrix. We can have two cases based on the rank of $A'_{\le(i-1)}$:

- Case B1: $A'_{\le(i-1)}$ has full row rank. Suppose $l_m \in \mathrm{col}_{\overline{A}_j}$ for some $j \le (i-1)$ and let $\mathbb{S}_{\le(i-1)\setminus j}$ denote the constrained system that excludes $\mathbb{S}_j$. Then, using the fact that $A'_{\le(i-1)}$ is full rank, we have

$$\eta(\mathbb{S}'_{\le(i-1)} \mid \mathcal{R}) \le N^{t-2} \times \eta(\mathbb{S}_{\le(i-1)\setminus j} \mid \mathcal{R}),$$

  and further, using the crude bound (11), we have $\eta(\mathbb{S}_{\le(i-1)} \mid \mathcal{R}) \ge (N^{t-1} - t(s+q)N^{t-2}) \times \eta(\mathbb{S}_{\le(i-1)\setminus j} \mid \mathcal{R})$ holds as $\mathbb{S}_{\le(i-1)}$ is acyclic and t-regular. Thus,

$$\eta(\mathbb{S}'_{\le(i-1)} \mid \mathcal{R}) \le \frac{2}{N} \eta(\mathbb{S}_{\le(i-1)} \mid \mathcal{R}),$$

  where we use the fact that $t(s+q) \le N/2$. There are at most $\binom{t}{m}$ choices for $\mathcal{I}$ and for each such choice there are at most $q^m s^{t-m}$ choices for $(l_1, \ldots, l_m, a_{t-m+1}, \ldots, a_t)$, which finally gives

$$\sum_{\mathcal{I} \subseteq [t]} \sum_{\boldsymbol{y}_{\le(i-1)}} \mu(\boldsymbol{\lambda}_i, \hat{\mathcal{P}}_{\mathcal{I}}, \hat{\mathcal{R}}_{[t]\setminus \mathcal{I}}) \le \frac{2(s+q)^t}{N} \eta(\mathbb{S}_{\le(i-1)} \mid \mathcal{R}).$$

- Case B2: $A'_{\le(i-1)}$ does not have full row rank. This case is only possible if the additional equation is defined by the equations in $\mathbb{S}_{\le(i-1)}$. Since $\mathbb{S}_{\le(i-1)}$ is isolated, this case is only possible if the additional equation is redundant, i.e., $\mathcal{I} = [t]$, $\{l_1, \ldots, l_t\} = \mathrm{col}(\overline{A}_j)$ for some $j \le (i-1)$, and $\lambda_j = \lambda_i$. Since there is only one choice for $\mathcal{I}$, and at most $\Delta_{\mathbb{S}}$ choices for j, the number of solutions in this case is bounded by $\Delta_{\mathbb{S}} \eta(\mathbb{S}_{\le(i-1)} \mid \mathcal{R})$.

The claim then follows by combining the bounds in all cases, and the lemma follows by substituting the claimed bound in (13). □

Now on to non-isolated components.

Lemma 5. *Suppose $\mathbb{S}_i$ is non-isolated. Then, we have*

$$\eta(\mathbb{S}_i \mid \mathcal{F}) \ge \frac{\prod_{j=1}^{t} (N - f^{(j)}_{\le(i-1)})^{r_i^{(j)}}}{N^{q_i}} \left(1 - \frac{2 r_i^t (s+q)^t}{N^t} - \varepsilon_{\mathrm{odd}}(q,r,s,t)\right),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s,t) = \begin{cases} \frac{2q_i(s+q)^{t-1}}{N^{t-1}} & \text{for odd } t, \\ 0 & \text{for even } t. \end{cases}$$

Proof. Recall from (10) that

$$\eta(\mathbb{S}_i \mid \mathcal{F}) = \sum_{\mathcal{J} \subseteq [r_i]} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}|.$$

First consider the even t case, where using Bonferroni's inequality, we have

$$\begin{aligned}
\eta(\mathbb{S}_i \mid \mathcal{F}) &\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t-1}} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t-1}} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{r_i - |\mathcal{J}| - q_i} \\
&\geq \frac{1}{N^{q_i}} \left(\sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t}} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{r_i - |\mathcal{J}|} - \sum_{\substack{\mathcal{J}' \subseteq [r_i] \\ |\mathcal{J}'| = t}} f_{(\mathcal{J}')} N^{r_i - t} \right) \\
&\geq \frac{1}{N^{q_i}} \left(\prod_{j=1}^{r_i} (N - f_{(j)}) - r_i^t (s+q)^t N^{r_i - t} \right) \\
&\geq \frac{\prod_{j=1}^{r_i} (N - f_{(j)})}{N^{q_i}} \left(1 - \frac{2 r_i^t (s+q)^t}{N^t} \right),
\end{aligned} \tag{15}$$

where the last inequality follows from the fact that $f_{(j)} \leq (s+q)$ for any j and $r_i(s+q) \leq \xi_{\mathbb{S}}(s+q) \leq N/2$.

As for the odd t case, using Bonferroni's inequality, we have

$$\begin{aligned}
\eta(\Psi_i \mid \mathcal{F}) &\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t}} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| < t}} (-1)^{|\mathcal{J}|} f_{\mathcal{J}} N^{r_i - |\mathcal{J}| - q_i} - \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \frac{1}{N^{q_i}} \left(\sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| < t}} (-1)^{|\mathcal{J}|} f_{\mathcal{J}} N^{r_i - |\mathcal{J}|} - N^{q_i} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right) \\
&\geq \frac{1}{N^{q_i}} \left(\prod_{j=1}^{r_i} (N - f_{(j)}) - N^{q_i} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right) \\
&\geq \frac{\prod_{j=1}^{r_i} (N - f_{(j)})}{N^{q_i}} \left(1 - \frac{2}{N^{r_i - q_i}} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right)
\end{aligned} \tag{16}$$

Claim. We claim

$$\sum_{\substack{\mathcal{J}\subseteq[r_i]\\|\mathcal{J}|=t}} |\mathcal{A}_{\mathcal{J}}| \le q_i(s+q)^{t-1}N^{r_i-t-q_i+1} + r_i^t(s+q)^t N^{r_i-t-q_i}.$$

Proof. Let $\mathcal{J} = \{l_1,\dots,l_t\}$ and suppose $\mathbb{S}_i'$ denote the updated system after the removal of these t columns from $\mathbb{S}_i$. Using Proposition 9, we have two cases:

- Case A: $\mathcal{J} = \mathrm{col}(\overline{A}_{j\bullet})$ for some $A_{j\bullet} \in \mathrm{row}(A_i)$. From Proposition 9 we know that $\mathrm{rank}(\mathbb{S}_i') = q_i - 1$. Thus, we have

$$\sum_{\substack{\mathcal{J}=\mathrm{col}(\overline{A}_{j\bullet})\\A_{j\bullet}\in\mathrm{row}(A_i)}} |\mathcal{A}_{\mathcal{J}}| \le q_i(s+q)^{t-1}N^{r_i-t-q_i+1}.$$

- Case B: $\mathcal{J} \neq \mathrm{col}(\overline{A}_{j\bullet})$ for all $A_{j\bullet} \in \mathrm{row}(A_i)$. From Proposition 9 we know that $\mathrm{rank}(\mathbb{S}_i') = q_i$. Thus, we have

$$\sum_{\mathcal{J}\neq\mathrm{col}(\overline{A}_{j\bullet})} |\mathcal{A}_{\mathcal{J}}| \le r_i^t(s+q)^t N^{r_i-t-q_i}.$$

This proves the claim. □

The result follows by substituting the claimed bound in (16) by realizing that

$$\prod_{j=1}^{r_i}(N - f_{(j)}) = \prod_{k=1}^{t}(N - f_{\le(i-1)}^{k})^{r_i^{(k)}}$$ □

Since the bound in Lemma 5 is independent of $\boldsymbol{y}_{\le(i-1)}$, we have the following corollary.

Corollary 2. *Suppose $\mathbb{S}_i$ is non-isolated. Then, we have*

$$\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R}) \ge \frac{\prod_{j=1}^{t}(N - f_{\le(i-1)}^{(j)})^{r_i^{(j)}}}{N^{q_i}}\left(1 - \frac{2r_i^t(s+q)^t}{N^t} - \varepsilon_{\mathrm{odd}}(q_i,r_i,s,t)\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s,t) = \begin{cases} \frac{2q_i(s+q)^{t-1}}{N^{t-1}} & \text{for odd } t,\\ 0 & \text{for even } t.\end{cases}$$

Theorem 1 now follows from the appropriate recursive application of Lemma 4 and Corollary 2 for all i from c down to 1, carefully accumulating the bound for non-isolated components.

A Better Bound for 2-Regular Systems. For the special case of 2-regular systems, we provide a finer version of Lemma 4, and hence, a finer version of Theorem 1. Let

$$\nabla_{\hat{\mathcal{R}}} := \max_{d\neq 0}\left|\{(x_1,x_2)\in(\hat{\mathcal{R}}_1\times\hat{\mathcal{R}}_2) : x_1\oplus x_2 = d\}\right|.$$

Lemma 6. *Suppose* $\mathbb{S}_i$ *is* 2*-CARP. Then, we have*

$$\eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \geq \frac{\prod_{j=1}^{2}(N - f^{(j)}_{\leq (i-1)})}{N} (1-\epsilon)\, \eta(\mathbb{S}_{\leq (i-1)} \,|\, \mathcal{R}),$$

where

$$\epsilon \leq \frac{2}{N} \left| \mu_{\alpha_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R}) - \frac{|\mathcal{R}_1 \times \mathcal{R}_2|}{N} \right| + \frac{12(s+q)(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}})}{N^2} + \frac{24(s+q)^3}{N^3}$$

A proof of this lemma is provided in Appendix A.

Corollary 3. *Suppose* $\mathbb{S}$ *is* 2*-CARP. Then, we have* $\eta(\mathbb{S} \,|\, \mathcal{R}) \geq (1-\varepsilon)\, \mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R})$, *where*

$$\varepsilon \leq \frac{2}{N} \sum_{i \notin \mathtt{NI}(\mathbb{S})} \frac{2}{N} \left| \mu_{\alpha_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R}) - \frac{|\mathcal{R}_1 \times \mathcal{R}_2|}{N} \right| + \frac{12q(s+q)(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}})}{N^2} + \frac{24q(s+q)^3}{N^3} + \frac{2(s+q)^2}{N^2} \sum_{j \in \mathtt{NI}(\mathbb{S})} r_j^2 .$$

4.2 The Case of CAR Clique System

Towards a variation of Theorem 1, suppose $\mathbb{S}$ is column-uniform, acyclic, k-regular (k-CARC) for some $k \geq 2$, and clique. Thus, $t = 1$ in this case.

A system $\mathbb{S}$ is said to be *trivial* if and only if there exists $\boldsymbol{v} \in \mathrm{rowsp}^+(A)$ such that

$$H(\boldsymbol{v}) = 2 \quad \text{and} \quad (\boldsymbol{v}|0) \in \mathrm{rowsp}^+(A|\boldsymbol{\lambda}),$$

and *non-trivial* otherwise. For all trivial systems, $\eta(\mathbb{S} \,|\, \mathcal{R}) = 0$ can be trivially 0 for certain fields, such as the characteristic 2 fields with binary matrices. Accordingly, we assume that the system is non-trivial. Beyond this obvious limitation, the case of clique systems is quite similar to the partite case.

Indeed we reuse the same notations and arguments to a large extent. First, we redefine

$$\mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R}) \coloneqq \frac{(N-s)_r}{N^q}$$

Next, suppose $\overline{\mathbb{S}}$ denote an arbitrary partite version of $\mathbb{S}$. Set $\mathcal{R}_1 = \cdots = \mathcal{R}_k$, $s_1 = \ldots = s_k$, and reuse the definitions of $\mathcal{A}_\varnothing$ and $\mathcal{A}_{\{j\}}$ for any $j \in \mathrm{col}(\overline{A}_i)$. Furthermore, for each $j_1 \neq j_2 \in \mathrm{col}(\overline{A}_i)$, let

$$\mathsf{EQ}_{j_1,j_2} \coloneqq \{y = (y_1, \ldots, y_{r_i}) \in \mathbb{F}_N^{r_i} \;:\; \overline{A}_i y = \boldsymbol{\lambda}_i \wedge y_{j_1} = y_{j_2}\}.$$

Then, for any $i \in [q]$, we have

$$(\mathbb{S}_i \,|\, \mathcal{F}) = \mathcal{A}_\varnothing \setminus \left(\left(\bigcup_{j=1}^{r_i} \mathcal{A}_{\{j\}} \right) \cup \left(\bigcup_{j_1 < j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2} \right) \right),$$

More importantly,

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) = |\mathcal{A}_\varnothing| - \left| \bigcup_{j=1}^{r_i} \mathcal{A}_{\{j\}} \right| - \left| \bigcup_{j_1 < j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2} \right|$$

$$= \eta(\overline{\mathbb{S}}_i \,|\, \mathcal{F}) - \left| \bigcup_{j_1 < j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2} \right|$$
$$\geq \eta(\overline{\mathbb{S}}_i \,|\, \mathcal{F}) - \binom{r_i}{2} N^{r_i - 1 - q_i}$$

where the inequality follows from the fact that $|\mathsf{EQ}_{j_1,j_2}| \leq N^{r_i-1-q_i}$ as $\mathtt{H}(A) \geq k \geq 2$. This gives the following clique counterparts for the results derived in the partite case.

Lemma 7. *Suppose $\mathbb{S}_i$ is isolated and non-trivial. Then, for any $y_{\leq(i-1)} \in (\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R})$, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \geq \frac{(N - f_{\leq(i-1)})^k}{N} \left(1 + (-1)^k \frac{2}{N^{k-1}} \left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f^k_{\leq(i-1)}}{N} \right) - \frac{k^2}{N} \right).$$

Lemma 8. *Suppose $\mathbb{S}_i$ is isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \geq \frac{(N - f_{\leq(i-1)})^k}{N} \left(1 - \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R})}{N^{k-1}} - \frac{2\Delta_{\mathbb{S}}}{N^{k-1}} - \frac{6(s+kq)^k}{N^k} - \frac{k^2}{N} \right).$$

Lemma 9. *Suppose $\mathbb{S}_i$ is non-isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \geq \frac{(N - f_{\leq(i-1)})^{r_i}}{N^{q_i}} \left(1 - \frac{2r_i^k(s+kq)^k}{N^k} - \varepsilon_{\mathrm{odd}}(q,r,s) - \frac{r_i^2}{N} \right),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s) = \begin{cases} \frac{2q_i(s+kq)^{k-1}}{N^{k-1}} & \text{for odd } k, \\ 0 & \text{for even } k. \end{cases}$$

Corollary 4. *Suppose $\mathbb{S}_i$ is non-isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \geq \frac{(N - f_{\leq(i-1)})^{r_i}}{N^{q_i}} \left(1 - \frac{2r_i^k(s+kq)^k}{N^k} - \varepsilon_{\mathrm{odd}}(q,r,s) - \frac{r_i^2}{N} \right) \eta(\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R}),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s) = \begin{cases} \frac{2q_i(s+kq)^{k-1}}{N^{k-1}} & \text{for odd } k, \\ 0 & \text{for even } k. \end{cases}$$

Theorem 2 (Clique Bound). *Let $k \geq 2$ and $\mathcal{R}$ be a family of sets. For any $(q,r,1)$-constrained system $\mathbb{S}$ which is non-trivial, k-CARC and which satisfies $\xi_{\mathbb{S}}(q+s) \leq N/2$, we have $\eta(\mathbb{S} \,|\, \mathcal{R}) \geq (1-\varepsilon)\, \mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R})$, where*

$$\varepsilon \leq \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}, \mathcal{R})}{N^{k-1}} + \frac{2q\Delta_{\mathbb{S}}}{N^{k-1}} + \frac{6q(s+kq)^k}{N^k} + \frac{2qk^2}{N} + \sum_{i \in \mathtt{NI}(\mathbb{S})} \left(\frac{2r_i^k(s+kq)^k}{N^k} + \frac{q_i(s+kq)^{k-1}}{N^{k-1}} + \frac{r_i^2}{N} \right).$$

A Better Bound for 2-Regular Systems. As in the case of partite systems, we provide a refined bound for 2-regular systems. Let

$$\nabla_{\hat{\mathcal{R}}} := \max_{d \neq 0} \left| \{x_1, x_2 \in \hat{\mathcal{R}} : x_1 \oplus x_2 = d\} \right|.$$

Lemma 10. *Suppose $\mathbb{S}_i$ is 2-CARC. Then, we have*

$$\eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \geq \frac{\prod_{j=1}^{2} (N - f^{(j)}_{\leq(i-1)})}{N} (1 - \epsilon)\, \eta(\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R}),$$

where

$$\epsilon \leq \frac{2}{N} \left| \mu_{\alpha_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R}) - \frac{|\mathcal{R}|^2}{N} \right| + \frac{12(s+2q)(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}})}{N^2} + \frac{24(s+2q)^3}{N^3} + \frac{4}{N}$$

Corollary 5. *Suppose $\mathbb{S}$ is 2-CARC. Then, we have $\eta(\mathbb{S} \,|\, \mathcal{R}) \geq (1 - \varepsilon)\, \mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R})$, where*

$$\varepsilon \leq \frac{4q}{N} + \frac{2}{N} \sum_{i \notin \mathtt{NI}(\mathbb{S})} \left| \mu_{\alpha_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R}) - \frac{|\mathcal{R}|^2}{N} \right| + \frac{12q(s+2q)(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}})}{N^2} + \frac{24q(s+2q)^3}{N^3} + \left(\frac{2(s+2q)^2}{N^2} + \frac{1}{N} \right) \sum_{j \in \mathtt{NI}(\mathbb{S})} r_j^2.$$

5 Single-keyed Double-block Hash-then-Sum

Let π be a permutation of $\{0,1\}^n$. We define three injective functions $\pi_0, \pi_1, \pi_2 : \{0,1\}^{n-2} \to \{0,1\}^n$ as follows:

$$\pi_0(\cdot) := \pi(00\|\cdot) \qquad \pi_1(\cdot) := \pi(01\|\cdot) \qquad \pi_2(\cdot) := \pi(10\|\cdot)$$

For $0 \leq j \leq 2$, we define $\mathcal{I}_j(n) := \{\pi_j : \pi \in \mathcal{P}(n)\}$.

Definition 8 (Single-keyed Permutation-based DBHtS). *For some permutation π of $\{0,1\}^n$ and a permutation-based rate-c^{-1} diblock hash function $H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^{n-2} \times \{0,1\}^{n-2}$, we define the single-keyed* DBHtS, *denoted* $\mathtt{1k\text{-}DBHtS}_{\pi,H}$ *construction by the mapping:*

$$m \mapsto \pi_1(H_{\pi_0}(m)) \oplus \pi_2(H_{\pi_0}(m)). \tag{17}$$

The construction is illustrated in Fig. 3.

We drop the parameters π and H whenever they are clear from the context. We reemphasize here that the π_0,π_1,π_2 are all domain-separated versions of the same permutation π.
The following theorem shows that 1k-DBHtS is a secure PRF up to $\overline{\sigma} \leq 2^{3n/4}$, assuming appropriate bounds for the hash function. We remark that it might be possible to improve some of the constants slightly. However, the bound is asymptotically tight in terms of number of queries due to an attack by Leurent et al. [31] on the general DBHtS construction.

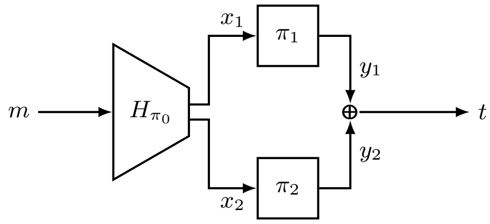

Fig. 3. The $\texttt{1k-DBHtS}_{\pi,H}$ construction.

Theorem 3. *Let $c, q, \ell, \sigma \geq 0$ satisfying $q, \ell \leq \sigma$ and $\overline{\sigma} = c\sigma + 2q \leq 2^{n-3}$. Suppose $H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^{2n-4}$ is a ratec^{-1} $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CFH. Then, for $\rho = (q, \ell, \sigma)$ and $\rho' = (2, \ell, 2\ell)$, the PRF advantage of any ρ-distinguisher $\mathcal{A}$ against $\texttt{1k-DBHtS}_{\Pi,H}$ satisfies*

$$\mathbf{Advt}^{\mathrm{prf}}_{\texttt{1k-DBHtS}_{\Pi,H}}(\mathcal{A}) \leq \epsilon_1 + \epsilon_2,$$

where

$$\epsilon_1 := 2\epsilon_2(\rho, 4) + \delta(\rho) + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3\left(\rho, 2^n/4\overline{\sigma}\right).$$

$$\epsilon_2 := \frac{16q^2\overline{\sigma}^2\epsilon_1(\rho')}{2^{2n}} + \frac{8q^2\epsilon_1(\rho')}{2^n} + \frac{q\overline{\sigma}}{2^{\frac{3n-4}{2}}} + \frac{120q\overline{\sigma}^3}{2^{3n}} + \frac{6q}{2^n}.$$

Proof. Without loss of generality assume that $\mathcal{A}$ is deterministic. Let
- $\mathsf{M}^i := (\mathsf{M}^i_1, \ldots, \mathsf{M}^i_{\ell_i})$, denote the i-th query of the distinguisher, containing $\ell_i \leq \ell$ blocks.
- T^i, denote the i-th response of the oracle.

In addition, the oracle releases additional information to the distinguisher, once the distinguisher is done querying the oracle, but before it outputs its decision bit.

In the real world, the oracle releases:
- $\mathsf{X}^i := (\mathsf{X}^i_1, \mathsf{X}^i_2) = H_{\Pi_0}(\mathsf{M}^i)$, the $(2n-2)$-bit internal hash output, or *finalization input* corresponding to the i-th query.
- $\mathsf{Y}^i := (\mathsf{Y}^i_1, \mathsf{Y}^i_2) = (\Pi_1(\mathsf{X}^i_1), \Pi_2(\mathsf{X}^i_2))$, the $2n$-bit *finalization output* corresponding to the i-th query.
- $\mathcal{R}$, the set of all image points sampled during the computation of $H_{\Pi_0}(\mathsf{M}^i)$ for all $i \in [q]$. Since H is a rate-c^{-1} hash function, $|\mathcal{R}| = c\sigma$ for M.

Thus, the full real world transcript can be described as

$$\Theta_{\mathrm{re}} := ((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i, \mathsf{Y}^i : i \in [q]), \mathcal{R}).$$

In the ideal world, the oracle first samples a dummy random permutation Π', and then computes $\mathsf{X}^i := H_{\Pi'_0}(\mathsf{M}^i)$ for all $i \in [q]$. In other words, X^i is generated faithfully for all $i \in [q]$. Note that, $\mathcal{R}$ can be derived here as well, as the ideal oracle is faithfully generating the hash outputs.

Sampling Y in the ideal world: The sampling mechanism for Y^i is on the other hand a bit more sophisticated. The goal is to sample Y^i's in such a way that

$$(\mathsf{X}_1^i = \mathsf{X}_1^j \iff \mathsf{Y}_1^i = \mathsf{Y}_1^j), \qquad (\mathsf{X}_2^i = \mathsf{X}_2^j \iff \mathsf{Y}_2^i = \mathsf{Y}_2^j),$$

is satisfied for all $i \neq j \in [q]$. We refer to this predicate as the *permutation compatibility* or PC *condition*.

For any $i \in [q]$, let $(i)_1 := \min\{j < i \ : \ \mathsf{X}_1^i = \mathsf{X}_1^j\}$ and $(i)_2 := \min\{j < i \ : \ \mathsf{X}_2^i = \mathsf{X}_2^j\}$. Let $r = |\{(i)_1, (i)_2 \ : \ i \in [q]\}|$. Consider the 2-regular and binary, $(q, r, 1)$-constrained system $\mathbb{S} := \{\mathsf{Y}_1^{(i)_1} \oplus \mathsf{Y}_2^{(i)_2} = \mathsf{T}^i : i \in [q]\}$.

Any $\overline{\mathcal{R}}$-solution for $\mathbb{S}$ satisfies the PC condition, apart from fully defining Y. As long as the system is acyclic and non-trivial, we can use the results developed in the previous section. Keeping this in mind, we now define some bad predicates on the partial transcript $((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i : i \in [q]), \mathcal{R})$:

$$\begin{aligned}
&\mathsf{A}_1 : \exists^* i, j, k, l \in [q], && \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^j = \mathsf{X}_2^k \wedge \mathsf{X}_1^k = \mathsf{X}_1^l.\\
&\mathsf{A}_2 : \exists^* i, j \in [q], && \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{T}^i \oplus \mathsf{T}^j = 0^n.\\
&\mathsf{A}_3 : \exists^* k \geq 2^{n-2}/(c\sigma + 2q), i_1, \ldots, i_k \in [q], && \mathsf{X}_1^{i_1} = \mathsf{X}_1^{i_2} = \ldots = \mathsf{X}_1^{i_k}.\\
&\mathsf{B}_1 : \exists^* i, j, k, l \in [q], && \mathsf{X}_2^i = \mathsf{X}_2^j \wedge \mathsf{X}_1^j = \mathsf{X}_1^k \wedge \mathsf{X}_2^k = \mathsf{X}_2^l.\\
&\mathsf{B}_2 : \exists^* i, j \in [q], && \mathsf{X}_2^i = \mathsf{X}_2^j \wedge \mathsf{T}^i \oplus \mathsf{T}^j = 0^n.\\
&\mathsf{B}_3 : \exists^* k \geq 2^{n-2}/(c\sigma + 2q), i_1, \ldots, i_k \in [q], && \mathsf{X}_2^{i_1} = \mathsf{X}_2^{i_2} = \ldots = \mathsf{X}_2^{i_k}.\\
&\mathsf{C} : \exists^* i \in [q], && \mathsf{T}^i = 0^n.\\
&\mathsf{D} : \exists^* i, j \in [q], && \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^i = \mathsf{X}_2^j.\\
&\mathsf{E} : \exists^* i, j, k \in [q], && \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^j = \mathsf{X}_2^k \wedge \mathsf{T}^i \oplus \mathsf{T}^j \oplus \mathsf{T}^k = 0^n.
\end{aligned}$$

Define $\mathsf{Cyclic} := \mathsf{A}_1 \vee \mathsf{B}_1 \vee \mathsf{D}$, $\mathsf{Trivial} := \mathsf{A}_2 \vee \mathsf{B}_2 \vee \mathsf{C} \wedge \mathsf{E}$, and $\mathsf{Giant} := \mathsf{A}_3 \vee \mathsf{B}_3$. It is not difficult to see that as long as Cyclic, $\mathsf{Trivial}$, and Giant are false, $\mathbb{S}$ is acyclic and non-trivial, and satisfies $\chi_{\mathbb{S}}(c\sigma + 2q) \leq 2^{n-1}$ for $(c\sigma + 2q) < 2^{3n/4}$. For notational convenience, let $s = c\sigma$.
Sampling Y*:* Onwards we describe the sampling of Y conditioned on the fact that $\neg(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant})$ holds. Let $\mathrm{CF}(\mathbb{S}) = (\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$ such that all the isolated components appear before the non-isolated ones. Let $\mathtt{NI}(\mathbb{S})$ denote the set of all non-isoltaed components. We define $\mathsf{Y} \twoheadleftarrow (\mathbb{S} | \mathcal{R})$. This concludes the sampling in the ideal world, and finally the ideal world transcript is given by

$$\Theta_{\mathrm{id}} := ((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i, \mathsf{Y}^i : i \in [q]), \mathcal{R}).$$

where the PC condition is satisfied as long as $\neg(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant})$ holds; otherwise the transcript is defined arbitrarily.

(Bad) Transcript Definition and Analysis: The set of transcripts Ω is the set of all tuples $\omega = ((m^i, t^i, x^i, y^i : i \in [q]), \mathcal{R})$, where $m^i \in \{0,1\}^*$, $t^i \in \{0,1\}^n$, $x^i \in \{0,1\}^{2n-2}$, $y^i \in \{0,1\}^{2n}$ and $\mathcal{R} \subseteq (\{0,1\}^n)^{c\sigma}$, where $\sigma = \sum_{i=1}^{q} \lceil |m^i|/n \rceil$.

A transcript ω is said to be *bad*, i.e., $\omega \in \Omega_{\rm bad}$ if and only if it satisfies $\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant}$, and *good* otherwise.

Lemma 11. *Suppose H is an $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-coverfree hash function. Then*

$$\Pr(\Theta_{\rm id} \in \Omega_{\rm bad}) \le 2\epsilon_2(\rho, 4) + \delta(\rho) + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3\left(\rho, \frac{2^{n-2}}{c\sigma + 2q}\right).$$

Proof. Let $s' = 2^{n-2}/(c\sigma + 2q)$. We have

$$\begin{aligned}
\Pr(\Theta_{\rm id} \in \Omega_{\rm bad}) &= \Pr(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant}) \\
&\le \Pr(\mathsf{Cyclic}) + \Pr(\mathsf{Trivial}) + \Pr(\mathsf{Giant}) \\
&\le \Pr(\mathsf{A}_1) + \Pr(\mathsf{B}_1) + \Pr(\mathsf{D}) + \Pr(\mathsf{A}_2) + \Pr(\mathsf{B}_2) + \Pr(\mathsf{C}) + \Pr(\mathsf{E}) + \Pr(\mathsf{A}_3) + \Pr(\mathsf{B}_3) \\
&\le \Pr\left(\mathsf{AP1}^4_H(\mathsf{M})\right) + \Pr\left(\mathsf{AP2}^4_H(\mathsf{M})\right) + \Pr(\mathsf{COLL}_H(\mathsf{M})) + \frac{\Pr(\mathsf{COLL1}_H(\mathsf{M}))}{2^n} \\
&\quad + \frac{\Pr(\mathsf{COLL2}_H(\mathsf{M}))}{2^n} + \frac{q}{2^n} + \frac{\Pr\left(\mathsf{AP1}^3_H(\mathsf{M})\right)}{2^n} + \Pr\left(\mathsf{MC1}^{s'}_H(\mathsf{M})\right) + \Pr\left(\mathsf{MC2}^{s'}_H(\mathsf{M})\right) \\
&\le 2\epsilon_2(\rho, 4) + \delta + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3(\rho, s'),
\end{aligned}$$

where the the first three (in)equalities follow from the definition and a trivial application of union bound, the fourth inequality just maps the bad predicates to corresponding coverfree hash events, and finally the fifth inequality follows from the coverfree bound of H. □

Good Transcript Analysis: Fix a good transcript $\omega \in \Omega_{\rm good}$. We will recycle notations from the sampling phase.

In the real world, Π is sampled exactly $s+r$ times ($|\mathcal{R}| = s$ and $|\{(i)_1, (i)_2 : i \in [q]\}| = r$). Thus, we have

$$\Pr(\Theta_{\rm re} = \omega) = \frac{1}{(2^n)_{s+r}} \tag{18}$$

In the ideal world, first T is sampled uniformly from a set of size 2^{nq}, followed by $\mathcal{R}$ which is sampled faithfully via Π. Finally, Y is sampled uniformly from $(\mathbb{S}|\mathcal{R})$. Using Corollary 5, we have

$$\begin{aligned}
\Pr(\Theta_{\rm id} = \omega) &= \frac{1}{2^{nq}} \times \frac{1}{(2^n)_s} \times \frac{1}{\eta(\mathbb{S}|\mathcal{R})} \\
&= \frac{1}{2^{nq}} \times \frac{1}{(2^n)_s} \times \frac{1}{(1-\varepsilon)\mathbb{E}(\mathbb{S}|\mathcal{R})}
\end{aligned} \tag{19}$$

where

$$\varepsilon \le \frac{4q}{2^n} + \frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\left|\mu(\mathsf{T}^{(i)}, \mathcal{R}) - \frac{|\mathcal{R}|^2}{2^n}\right| + \frac{12q(s+2q)(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}})}{2^{2n}} + \frac{24q(s+2q)^3}{2^{3n}} + \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right)\sum_{j \in \mathtt{NI}(\mathbb{S})} r_j^2.$$

Then, on dividing (18) by (19), we have

$$\frac{\Pr(\Theta_{\rm re} = \omega)}{\Pr(\Theta_{\rm id} = \omega)} \geq (1-\varepsilon)\,. \tag{20}$$

To apply Corollary 1, we have to compute the following expectations

$$\mu := \mathbb{E}\left(\frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\left|\mu(\mathsf{T}^{(i)},\mathcal{R}) - \frac{|\mathcal{R}|^2}{N}\right|\right)$$
$$\nu := \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right)\mathbb{E}\left(\sum_{j \in \mathtt{NI}(\mathbb{S})} r_j^2\right)$$
$$\delta := \mathbb{E}(\varDelta_{\mathbb{S}})$$
$$\gamma := \mathbb{E}(\nabla_{\mathcal{R}})$$

Using Proposition 1 and 2 and $s + 2q \leq 2^n$, we have $\delta \leq 6n$ and $\gamma \leq 7nq^2/2^n$.

Let $\sim_1$ (res. $\sim_2$) be equivalence relations on $[q]$, such that $i \sim_1 j$ (res. $i \sim_2 j$) if and only if $\mathsf{X}_1^i = \mathsf{X}_1^j$ (res. $\mathsf{X}_2^i = \mathsf{X}_2^j$). Let $\mathcal{C}_1^1, \ldots, \mathcal{C}_{t_1}^1$ and $\mathcal{C}_1^2, \ldots, \mathcal{C}_{t_2}^2$ denote the non-singleton equivalence classes of $[q]$ with respect to $\sim_1$ and $\sim_2$, respectively. For $i \in [t_1]$ and $j \in [t_2]$, let $\mathtt{mc}_i^{(1)} = |\mathcal{C}_i^1|$ and $\mathtt{mc}_j^{(2)} = |\mathcal{C}_j^2|$.

$$\begin{aligned}
\nu &= \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right)\mathbb{E}\left(\sum_{i' \in \mathtt{NI}(\mathbb{S})} r_{i'}^2\right) \\
&\leq \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right) \times 2\left(\sum_{j=1}^{t_1}\mathbb{E}\left(\mathtt{mc}_j^{(1)}\right) + \sum_{j'=1}^{t_2}\mathbb{E}\left(\mathtt{mc}_{j'}^{(2)}\right)\right) \\
&\leq \frac{16q^2(s+2q)^2\epsilon_1(2,\ell,2\ell)}{2^{2n}} + \frac{8q^2\epsilon_1(2,\ell,2\ell)}{2^n}.
\end{aligned} \tag{21}$$

Using Proposition 5, we have

$$\begin{aligned}
\mu &= \mathbb{E}\left(\frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\left|\mu(\mathsf{T}^{(i)},\mathcal{R}) - \frac{|\mathcal{R}|^2}{N}\right|\right) \\
&= \frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\mathbb{E}\left(\left|\mu(\mathsf{T}^{(i)},\mathcal{R}) - \frac{|\mathcal{R}|^2}{N}\right|\right) \\
&\leq \frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right)} + \frac{2}{2^n}\sum_{i \notin \mathtt{NI}(\mathbb{S})}\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) - \frac{s^2}{N}\right|,
\end{aligned} \tag{22}$$

We claim:

$$\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) - \frac{s^2}{2^n}\right| \leq \frac{3s}{2^n} \tag{23}$$

$$\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right)} \leq \frac{\sqrt{2}s}{2^{n/2}} + \frac{4s^2}{2^{3n/2}} \tag{24}$$

A proof of this claim is given in Appendix B. Theorem 3 then follows from Lemma 11 and (21)-(24). □

6 Instantiations of Cover-free Hash functions

For a diblock hash function $H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$ we can construct the truncated diblock hash $\mathtt{T}H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^{n-2} \times \{0,1\}^{n-2}$ as $\mathtt{T}H(x) \coloneqq (\mathtt{Trunc}(H_1(x)), \mathtt{Trunc}(H_2(x)))$, where $\mathtt{Trunc} : \{0,1\}^n \to \{0,1\}^{n-2}$ truncates the first two bits of its n-bit input.

Now let us define the functions $\mathtt{PHash} : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$ and $\mathtt{LightHash} : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$, as follows:

$\mathtt{PHash}_{\Pi_0}$	$\mathtt{LightHash}_{\Pi_0}$
Input: $m = m[1]\|\cdots\|m[k] \in (\{0,1\}^{n-2})^k$	Input: $m = m[1]\|\cdots\|m[k] \in (\{0,1\}^{n-s})^k$
$\Delta_0 \leftarrow \mathtt{Trunc}(\Pi_0(0))$	for $i \in [k]$,
$\Delta_1 \leftarrow \mathtt{Trunc}(\Pi_0(1))$	$Z[i] \leftarrow \Pi_0(\langle i \rangle_{s-2} \| m[i])$
for $i \in [k]$,	$x[1] \leftarrow Z[1] \oplus Z[2] \oplus \cdots \oplus Z[k]$
$W[i] \leftarrow m[i] \oplus 2^i \cdot \Delta_0 \oplus 2^{2i} \cdot \Delta_1$	$x[2] \leftarrow 2^{k-1} \cdot Z[1] \oplus 2^{k-2} \cdot Z[2] \cdots \oplus Z[k]$
$Z[i] \leftarrow \Pi_0(W[i])$	return $x \coloneqq (x[1]\|x[2])$
$x[1] \leftarrow Z[1] \oplus Z[2] \cdots \oplus Z[k]$	
$x[2] \leftarrow Z[1] \oplus 2 \cdot Z[2] \cdots \oplus 2^{k-1} \cdot Z[k]$	
return $x \coloneqq (x[1]\|x[2])$	

Two instances of CfHs will be the truncated versions of the above hash functions, $\mathtt{TPHash}$ and $\mathtt{TLightHash}$, respectively.

6.1 Affine bad events.

For a diblock hash function H, any $\boldsymbol{x} = (x_1, \ldots, x_q) \in (\mathcal{X})_q$, and $c, c_1, c_2, c_3 \in \{0,1\}^2$, we define:

$$\mathsf{COLL}_H^{c_1,c_2}(\boldsymbol{x}) : \ \exists^* i,j \in [q] \text{ such that } H_K(x_i) \oplus H_K(x_j) = (c_1 \| 0^{n-2}, c_2 \| 0^{n-2})$$

$$\mathsf{COLL1}_H^{c}(\boldsymbol{x}) : \ \exists^* i,j \in [q] \text{ such that } H_K^1(x_i) \oplus H_K^1(x_j) = c \| 0^{n-2}.$$

$$\mathsf{COLL2}_H^{c}(\boldsymbol{x}) : \ \exists^* i,j \in [q] \text{ such that } H_K^2(x_i) \oplus H_K^2(x_j) = c \| 0^{n-2}.$$

$$\mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x}) : \ \exists^* i,j,k,l \in [q] \text{ such that}$$

$$H_K^1(x_i) \oplus H_K^1(x_j) = c_1 \| 0^{n-2} \wedge H_K^2(x_j) \oplus H_K^2(x_k) = c_2 \| 0^{n-2}$$
$$\wedge\, H_K^1(x_k) \oplus H_K^1(x_l) = c_3 \| 0^{n-2}.$$

$$\mathsf{AP2}_H^{c_1,c_2,c_3}(\boldsymbol{x}) : \ \exists^* i,j,k,l \in [q] \text{ such that}$$

$$H_K^2(x_i) \oplus H_K^2(x_j) = c_1 \| 0^{n-2} \wedge H_K^1(x_j) \oplus H_K^1(x_k) = c_2 \| 0^{n-2}$$
$$\wedge\, H_K^2(x_k) \oplus H_K^2(x_l) = c_3 \| 0^{n-2}.$$

$\mathsf{AP2}_H^{c_1,c_2}(\boldsymbol{x})$: $\exists^* \, i,j,k \in [q]$ such that

$$H_K^2(x_i) \oplus H_K^2(x_j) = c_1 \| 0^{n-2} \wedge H_K^1(x_j) \oplus H_K^1(x_k) = c_2 \| 0^{n-2}$$

One can readily check that

$$\begin{aligned} \mathsf{COLL1}_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{c \in \{0,1\}^2} \mathsf{COLL1}_H^c(\boldsymbol{x}) & \mathsf{COLL2}_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{c \in \{0,1\}^2} \mathsf{COLL2}_H^c(\boldsymbol{x}) \\ \mathsf{AP1}_{\mathtt{T}H}^4(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2,c_3) \\ \in (\{0,1\}^2)^3}} \mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x}) & \mathsf{AP2}_{\mathtt{T}H}^4(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2,c_3) \\ \in (\{0,1\}^2)^3}} \mathsf{AP2}_H^{c_1,c_2,c_3}(\boldsymbol{x}) \\ \mathsf{COLL}_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2) \\ \in (\{0,1\}^2)^2}} \mathsf{COLL}_H^{c_1,c_2}(\boldsymbol{x}) & \mathsf{AP1}_{\mathtt{T}H}^3(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2) \\ \in (\{0,1\}^2)^3}} \mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x}) \end{aligned} \tag{25}$$

6.2 `TPHash`

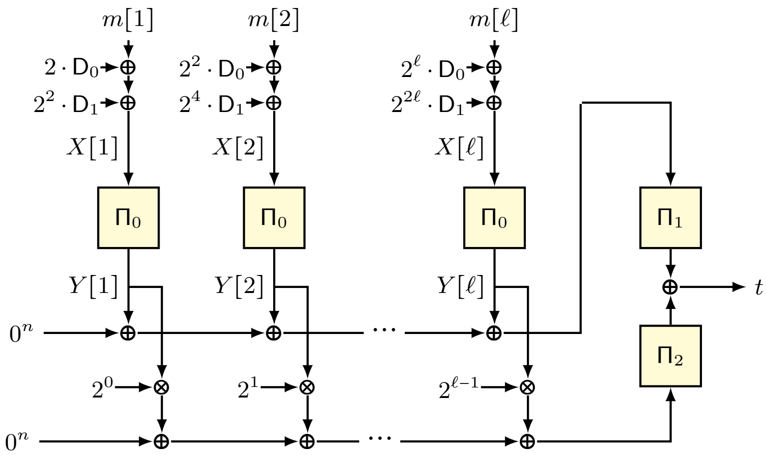

Fig. 4. `1k-PMAC+`

Our bad event analysis heavily depends on the one presented in [30]. We tailor their bounds according to our needs while highlighting the main aspects of similarity and departure between their results and ours.

Similar to the `PMAC+` analysis in [30] we define analogous auxiliary events as follows: Let the i-th message be $m^i = m^i[1] \| \cdots \| m^i[\ell_i] \in (\{0,1\}^{n-2})^{\ell_i}$, $i \in [q]$. For $i \neq j \in [q]$, let $\ell = \min\{\ell_i, \ell_j\}$ and $\ell' = \max\{\ell_i, \ell_j\}$, then we can define the index set for which $m^i[k] \neq m^j[k]$ as

$$I_{ij} := \{k \in [\ell] : m^i[k] \neq m^j[k]\} \sqcup [\ell + 1..\ell']$$

We define the following random variables: $\mathsf{D}_0 := \mathtt{Trunc}(\Pi_0(0))$, $\mathsf{D}_1 := \mathtt{Trunc}(\Pi_0(1))$, and $\mathsf{W}^i = \mathsf{W}^i[1]\|\cdots\|\mathsf{W}^i[\ell_i]$, where $\mathsf{W}^i[k] = m^i[k] \oplus 2^k \cdot \mathsf{D}_0 \oplus 2^{2k} \cdot \mathsf{D}_1$. We further define the random index sets where W^i and W^j collide as follows:

$$\mathtt{I}_{\mathsf{col}} = \{(i,j) \in ([q])_2 : \exists^* k, k' \text{ such that } \mathsf{W}^i[k] = \mathsf{W}^j[k']\}$$
$$\mathsf{J}_{\mathsf{col}} = \{(i,j) \in ([q])_2 : \min\{I_{ij}\} \le \ell_i \text{ and } \exists k \text{ such that } \mathsf{W}^i[\min\{I_{ij}\}] = \mathsf{W}^j[k]\}$$

Then the auxiliary events are:
Aux_1 **:** $\mathsf{D}_0 = 0 \lor \mathsf{D}_1 = 0$
Aux_2 **:** $\exists i \in [q], \exists^* k, k'$ such that $\mathsf{W}^i[k] = \mathsf{W}^i[k']$.
Aux_3 **:** $\exists i \in [q], k \in [\ell_i]$ such that $\mathsf{W}^i[k] \in \{0, 1, \Pi_0^{-1}(0)\}$.
Aux_4 **:** $|\mathtt{I}_{\mathsf{col}}| > q'$, where $q' = q/2^{n/4}$.
Aux_5 **:** $|\mathsf{J}_{\mathsf{col}}| > \sigma'$ where $\sigma' = \ell q$.
and let $\mathsf{Aux} = \bigvee_{i \in [5]} \mathsf{Aux}_i$.

Lemma 12. *For $\boldsymbol{m} = (m^i : i \in [q])$ and $c, c_1, c_2, c_3 \in \{0,1\}^2$,*

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{PHash_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) \le \frac{4\ell q^2}{2^{2n}}$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{PHash_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) \le \frac{2\sigma'^2}{2^{2n}} + \frac{4q'}{2^n} + \frac{2}{2^n} + \frac{2\sqrt{2}q^2}{2^{3n/2}} + \frac{8q'q^2}{2^{2n}} + \frac{96q^2}{2^{2n}} + \frac{8q^4}{2^{3n}}$$

Proof Sketch: First we note that, the following pairs of events, the left defined in [30] and the right defined in this paper, are equivalent in the single-key scenario:

$$\mathsf{Bad}_1 \equiv \mathsf{COLL}^{0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}), \qquad \mathsf{Bad}_2 \equiv \mathsf{AP1}^{0,0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$$

Analogous to Eq. (10) and (11) of [30], we can write, for any $c \in \{0,1\}^2$,

$$\mathtt{PHash}^1_{\Pi_0}(m^i) \oplus \mathtt{PHash}^1_{\Pi_0}(m^j) = c\|0^{n-2} \iff A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = c\|0^{n-2}$$
$$\mathtt{PHash}^2_{\Pi_0}(m^i) \oplus \mathtt{PHash}^2_{\Pi_0}(m^j) = c\|0^{n-2} \iff B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = c\|0^{n-2}$$

where, for $(i,j) \in ([q])_2$, $\{\mathsf{W}[1], \ldots, \mathsf{W}[t]\} := \{\mathsf{W}^i[1], \ldots, \mathsf{W}^i[\ell_i]\} \cup \{\mathsf{W}^j[1], \ldots, \mathsf{W}^j[\ell_j]\}$, and for $k \in [t]$, $\mathsf{Z}[k] := \Pi_0(\mathsf{W}[k])$.

Thus, borrowing from the analysis of [30], each of the events in the statement of this lemma can be written as an event that a system of equations $\boldsymbol{AZ} = \boldsymbol{c}$ holds, where $\boldsymbol{Z}$ is a vector with k-th component $\mathsf{Z}[k]$, and $\boldsymbol{c}$ depends on the indices c, c_1, c_2, c_3 of the corresponding event. If $\boldsymbol{c} \notin \mathcal{C}(\boldsymbol{A})$, then this system of equations will hold with 0 probability. If $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{A})$ then the probability that this system of equations holds, depends on the rank of $\boldsymbol{A}$ and not on $\boldsymbol{c}$. So we have that

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) \le \Pr\left(\mathsf{COLL}^{0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) = \Pr\left(\mathsf{Bad}_1 \land \neg\mathsf{Aux}\right)$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) \le \Pr\left(\mathsf{AP1}^{0,0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \land \neg\mathsf{Aux}\right) = \Pr\left(\mathsf{Bad}_2 \land \neg\mathsf{Aux}\right)$$

Thus we can use the bounds on the corresponding bad events from [30] to get our result. ◻

The probability analysis of the events $\mathsf{AP2}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{AP1}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ are similar to the analysis of the events $\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{COLL}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$, respectively.

Lemma 13. *For $\ell \le 2^{n-2}$, $m \ne m' \in (\{0,1\}^{n-2})^{\le \ell}$, and $c \in \{0,1\}^2$, we have*

$$\Pr\left(\mathit{PHash}^1_{\Pi_0}(m) \oplus \mathit{PHash}^1_{\Pi_0}(m') = c\|0^{n-2}\right) \le \frac{26\ell}{2^n}$$
$$\Pr\left(\mathit{PHash}^2_{\Pi_0}(m) \oplus \mathit{PHash}^2_{\Pi_0}(m') = c\|0^{n-2}\right) \le \frac{26\ell}{2^n}$$

Proof. Let $m \in (\{0,1\}^{n-2})^{\ell}$ and $m' \in (\{0,1\}^{n-2})^{\ell'}$. Note that the claim is trivial $\ell = 1$ and we ignore this case.

Let i be the maximum block-index where m and m' are distinct, precisely,

$$i = \begin{cases} \ell, & \text{if } \ell > \ell' \\ \max\{j \le \ell : m[j] \ne m'[j]\}, & \text{if } \ell = \ell' \end{cases}$$

Consider the random variables:

$$\begin{aligned} &\mathsf{D}_0 = \mathsf{trunc}(\Pi(0)), && \mathsf{D}_1 = \mathsf{trunc}(\Pi(1)), \\ &\mathsf{W}[i] = m[i] \oplus 2^i \cdot \mathsf{D}_0 \oplus 2^{2i} \cdot \mathsf{D}_1, && \mathsf{Z}[i] = \Pi_0(\mathsf{W}[i]), \qquad i \in [\ell] \\ &\mathsf{W}'[i] = m'[i] \oplus 2^i \cdot \mathsf{D}_0 \oplus 2^{2i} \cdot \mathsf{D}_1, && \mathsf{Z}'[i] := \Pi_0(\mathsf{W}'[i]), \qquad i \in [\ell'] \end{aligned}$$

Let us define the following events:

$$\begin{aligned} &\mathsf{E}_1 : \mathsf{D}_0 = 0 \\ &\mathsf{E}_2 : \bigvee_{j\in[\ell]} (\mathsf{W}[j] = 0 \vee \mathsf{W}[j] = 1) \vee \bigvee_{j\in[\ell']} (\mathsf{W}'[j] = 0 \vee \mathsf{W}'[j] = 1) \\ &\mathsf{E}_3 : \bigvee_{\substack{j\in[\ell] \\ j\ne i}} (\mathsf{W}[i] = \mathsf{W}[j]) \vee \bigvee_{j\in[\ell']} (\mathsf{W}[i] = \mathsf{W}'[j]) \end{aligned}$$

Note that $\Pr(c \cdot \mathtt{Trunc}(\Pi(a)) = b) = 4/2^n$ for any $a \in \{0,1\}^n$ and $b, c \in \{0,1\}^{n-2}$ with $c \ne 0$. Hence, for any $a_1, \ldots, a_r \in \{0,1\}^n$ and $b, c_1, \ldots, c_r \in \{0,1\}^{n-2}$ with $c_r \ne 0$, we have

$$\begin{aligned} &\Pr(c_1 \cdot \mathtt{Trunc}(\Pi(a_1)) \oplus \cdots \oplus c_r \cdot \mathtt{Trunc}(\Pi(a_r)) = b) \\ &= \sum_{\substack{b'_1,\ldots,b'_{r_1} \\ \in\{0,1\}^n \\ \text{all distinct}}} \Pr(\mathtt{Trunc}(\Pi(a_r)) = b') \Pr(\Pi(a_i) = b'_i \ \forall i \in [r-1]) \\ &\le \frac{4}{2^n - r + 1} \end{aligned}$$

where $b_i = \mathsf{trunc}(b'_i)$ and $b' = c_r^{-1} \cdot (b \oplus c_1 \cdot b_1 \oplus \cdots \oplus c_{r-1} \cdot b_{r-1})$. Similarly for any $a_1, \ldots, a_r \in \{0,1\}^n$ and $b, c_1, \ldots, c_r \in \{0,1\}^{n-2}$ with at least one $c_i \ne 0$, we have

$$\Pr(c_1 \cdot \Pi(a_1) \oplus \cdots \oplus c_r \cdot \Pi(a_r) = b) \le \frac{1}{2^n - r + 1}. \qquad (26)$$

This implies $\Pr(\mathsf{E}_1) = \Pr(\mathsf{trunc}(\Pi(0)) = 0) = 4/2^n$, $\Pr(\mathsf{E}_2 \mid \mathsf{E}_1^c) \le 4\ell \cdot 4/2^n$, and $\Pr(\mathsf{E}_3 \mid \mathsf{E}_1^c \wedge \mathsf{E}_2^c) \le (2\ell - 1) \cdot 4/2^n$.

Now the event $\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c \| 0^{n-2}$, is equivalent to $\mathsf{Z}[1] \oplus \cdots \oplus \mathsf{Z}[\ell] \oplus \mathsf{Z}'[1] \oplus \cdots \oplus \mathsf{Z}'[\ell'] = c\|0^{n-2}$. Of course, if any two Z-random variables are identically equal then they cancel out. However, conditional on $\mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c$ we have $\mathsf{Z}[i] \ne \mathsf{Z}[j], \mathsf{Z}'[j']$ for any $j \in [m] \setminus \{i\}, j' \in [m']$ and $\mathsf{Z}[i] \ne 0, \Pi(0), \Pi(1)$. Hence from Eq. (26), we have

$$\Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2} \,\middle|\, \mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c\right) \le \frac{1}{2^n - (m-1) - m' - 2} \le \frac{1}{2^n - 2\ell} \le 2/2^n$$

assuming $\ell \le 2^{n-2}$.

Since for any two events A and B, we have $\Pr(\mathsf{A}) = \Pr(\mathsf{A} \wedge \mathsf{B}) + \Pr(\mathsf{A} \wedge \mathsf{B}^c)$ and $\Pr(\mathsf{A} \wedge \mathsf{B}) \le \Pr(\mathsf{A})$ and $\Pr(\mathsf{A} \wedge \mathsf{B}) \le \Pr(\mathsf{A} \mid \mathsf{B})$, we have

$$\begin{aligned}
&\Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2}\right) \\
&\le \Pr(\mathsf{E}_1) + \Pr(\mathsf{E}_2 \mid \mathsf{E}_1^c) + \Pr(\mathsf{E}_3 \mid \mathsf{E}_1^c \wedge \mathsf{E}_2^c) \\
&\quad + \Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2} \,\middle|\, \mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c\right) \\
&\le \frac{4}{2^n} + \frac{16\ell}{2^n} + \frac{8\ell - 4}{2^n} + \frac{2}{2^n} \le \frac{26\ell}{2^n}
\end{aligned}$$

Same argument shows that $\Pr\left(\mathtt{PHash}^2_{\Pi_0}(m) \oplus \mathtt{PHash}^2_{\Pi_0}(m') = c\|0^{n-2}\right) \le 26\ell/2^n$.

Corollary 6.

$$\Pr\left(\mathsf{COLL1}^c_{PHash_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{13\ell q^2}{2^n} \qquad \Pr\left(\mathsf{COLL2}^c_{PHash_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{13\ell q^2}{2^n}$$

$$\Pr\left(\mathsf{MC1}^s_{TPHash_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{52\ell q^2}{s^2 \cdot 2^n} \qquad \Pr\left(\mathsf{MC2}^s_{TPHash_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{52\ell q^2}{s^2 \cdot 2^n}$$

The Corollary 6 follows from Lemma 13 and Proposition 4.

Finally, we bound the auxilliary events

Lemma 14. *We have*

$$\Pr(\mathsf{Aux}_1 \vee \mathsf{Aux}_3) \le \frac{3\ell q}{2^n - 2} + \frac{2}{2^n} \qquad \Pr(\mathsf{Aux}_2) \le \frac{\ell^2 q}{2^{n+1}}$$

$$\Pr(\mathsf{Aux}_4) \le \frac{\ell^2 q^2}{q' \cdot 2^n} \qquad \Pr(\mathsf{Aux}_5) \le \frac{\ell q^2}{\sigma' \cdot 2^n}$$

Combining these bounds we have

$$\Pr(\mathsf{Aux}) \le \frac{(\ell^2 + 8\ell)q}{2^{n+1}} + \frac{\ell^2 q^2}{q' \cdot 2^n} + \frac{\ell q^2}{\sigma' \cdot 2^n}$$

Combining Eq. (25), Lemma 12, Corollary 6 and Lemma 14 we have the following result:

Lemma 15. *$\mathtt{TPHash}_{\Pi_0}$ is a $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CfH where*

$$\epsilon_1(\rho) = \frac{26\ell q^2}{2^n}, \qquad \epsilon_2(\rho, 3) = \frac{16\ell q^2}{2^{2n}}, \qquad \epsilon_3(\rho, s) = \frac{52\ell q^2}{s^2 \cdot 2^n}, \qquad \delta(\rho) = \frac{16\ell q^2}{2^{2n}}$$

$$\epsilon_2(\rho, 4) = 8 \cdot \left(\frac{2\sigma'^2}{2^{2n}} + \frac{4q'}{2^n} + \frac{2}{2^n} + \frac{2\sqrt{2}q^2}{2^{3n/2}} + \frac{8q'q^2}{2^{2n}} + \frac{96q^2}{2^{2n}} + \frac{8q^4}{2^{3n}} \right)$$

6.3 `TLightHash`.

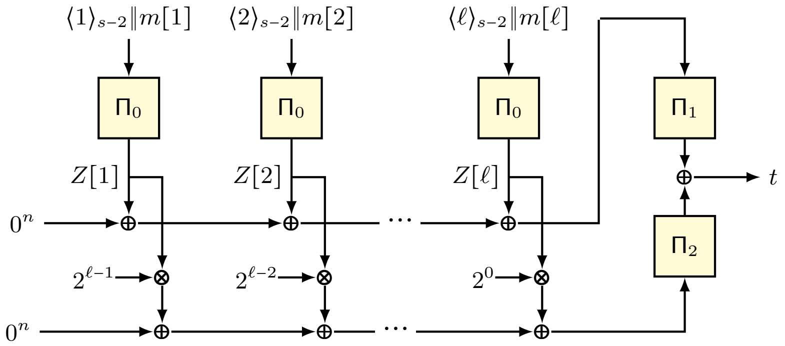

Fig. 5. `1k-LightMAC+`

As before, we let the i-th message be $m^i = m^i[1] \| \cdots \| m^i[\ell_i] \in (\{0,1\}^{n-s})^{\ell_i}$, $i \in [q]$. Note that, $m^i[k] \neq m^j[k] \iff \mathsf{Z}^i[k] \neq \mathsf{Z}^j[k]$ for any $k \in [\max\{\ell_i, \ell_j\}]$, where $\mathsf{Z}^i[k] := \Pi_0(\langle k \rangle_{s-2} \| m^i[k])$. Moreover, $\mathsf{Z}^i[k] \neq \mathsf{Z}^j[k']$ for any $k \neq k'$, $i, j \in [q]$. Let us fix $(i,j) \in ([q])_2$, define $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\} := \{\mathsf{Z}^i[k] : k \in [\ell_i]\} \cup \{\mathsf{Z}^j[k] : k \in [\ell_j]\}$ and partition $[t] := \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{i\bar{j}}$, where

$$\mathtt{I}_{\bar{i}j} := \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^i[k'] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j\}]\}$$
$$\mathtt{I}_{\bar{i}\bar{j}} := \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^i[k'] = \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j\}]\}$$
$$\mathtt{I}_{i\bar{j}} := \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^j[k'] \neq \mathsf{Z}^i[k'], k' \in [\max\{\ell_i, \ell_j\}]\}$$

Then we have

$$\begin{aligned} \mathtt{LightHash}^1_{\Pi_0}(m^i) \oplus \mathtt{LightHash}^1_{\Pi_0}(m^j) &= c \| 0^{n-2} \\ &\iff A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = c \| 0^{n-2} \\ \mathtt{LightHash}^2_{\Pi_0}(m^i) \oplus \mathtt{LightHash}^2_{\Pi_0}(m^j) &= c \| 0^{n-2} \\ &\iff B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = c \| 0^{n-2} \end{aligned}$$

where

- $A_k = 1$ for $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, $A_k = 0$, otherwise.

- $B_k = 2^\beta$ for some β, if $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, otherwise $B_k = 2^\beta \oplus 2^\gamma$ for some β, γ.

Due to this similarity with PHash, the argument given in Lemma 12 also holds here, giving us

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{COLL}^{0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{AP1}^{0,0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$
$$\Pr\left(\mathsf{AP2}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{AP2}^{0,0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$

Lemma 16. *Assume $\ell \leq 2^n/4$. Then in the ideal world,*

$$\Pr\left(\mathsf{COLL}^{0,0}_{LightHash_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{2q^2}{2^{2n}}$$

Proof. We fix $(i,j) \in ([q])_2$ as above, thus fixing $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\}$ and partitioning $[t] = \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{i\bar{j}}$. We can make the following observations about the index sets:

- $\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} \neq \varnothing$ since otherwise m^i and m^j will not be distinct.
- $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \geq 2$ because otherwise $\mathtt{LightHash}^1_{\Pi_0}(m^i) \neq \mathtt{LightHash}^1_{\Pi_0}(m^j)$.

If we consider the system of linear equations representing the events $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)$ and $\mathtt{LightHash}^2_{\Pi_0}(m^i) = \mathtt{LightHash}^2_{\Pi_0}(m^j)$, respectively:

$$A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = 0^n$$
$$B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = 0^n$$

then the above observations about the index sets imply that there are two distinct indices $k, k' \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$ such that $A_k = A_{k'} = 1$ and $B_k = 2^\beta$, $B_{k'} = 2^\gamma$ for distinct β and γ. This implies that the above system of linear equations has rank 2, and hence will be satisfied with probability $(2^n)_{t-2}/(2^n)_t = 1/(2^n - t + 2)(2^n - t + 1) \leq (2^n - 2\ell + 2)(2^n - 2\ell + 1) \leq 4/2^{2n}$ for $\ell \leq 2^n/4$. Since there are $q(q-1)/2$ tuples $(i,j) \in ([q])_2$, we have our result.

Lemma 17. *Assume that $\ell \leq 2^n/8$. Then in the ideal world, one has,*

$$\Pr\left(\mathsf{AP1}^{0,0,0}_{LightHash_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{q^4}{3 \cdot 2^{3n}} + \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n} + \frac{96q^2}{2^{2n}}$$

Proof. Let us fix $(i,j,r,s) \in ([q])_4$. We want to find the probability of the event

$$\begin{aligned}\mathsf{E}(i,j,r,s) : \left(\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)\right)\\ \wedge \left(\mathtt{LightHash}^2_{\Pi_0}(m^j) = \mathtt{LightHash}^2_{\Pi_0}(m^r)\right)\\ \wedge \left(\mathtt{LightHash}^1_{\Pi_0}(m^r) = \mathtt{LightHash}^1_{\Pi_0}(m^s)\right)\end{aligned}$$

Let $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\} = \{\mathsf{Z}^i[k] : k \in [\ell_i]\} \cup \{\mathsf{Z}^j[k] : k \in [\ell_j]\} \cup \{\mathsf{Z}^r[k] : k \in [\ell_r]\} \cup \{\mathsf{Z}^s[k] : k \in [\ell_s]\}$. Also let us partition $[t]$ in three ways as $[t] = \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{ij} = \mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}} \sqcup \mathtt{I}_{\bar{j}\bar{r}} \sqcup \mathtt{I}_{jr} = \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}} \sqcup \mathtt{I}_{\bar{r}\bar{s}} \sqcup \mathtt{I}_{rs}$ where

$$\mathtt{I}_{\bar{i}j} := \{k : \mathsf{Z}[k] = \mathsf{Z}^i[k'] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$

$$\mathtt{I}_{i\bar{j}} := \{k : \mathsf{Z}[k] = \mathsf{Z}^j[k'] \neq \mathsf{Z}^i[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$
$$\mathtt{I}_{\bar{i}\bar{j}} := \{k : \mathsf{Z}[k] = \mathsf{Z}^i[k'] = \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$
$$\mathtt{I}_{ij} := \{k : \mathsf{Z}[k] \neq \mathsf{Z}^i[k'], \mathsf{Z}[k] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$

and the rest of the index sets are defined analogously.

Then the above event can be represented by the following system of equations

$$A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = 0^n$$
$$B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = 0^n$$
$$C_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus C_t \cdot \mathsf{Z}[t] = 0^n$$

where

- $A_k = 1$ if $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, and $A_k = 0$ otherwise.
- $B_k = 2^\beta$ for some β if $k \in \mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}}$, $B_k = 2^\beta \oplus 2^\gamma$ for some β, γ if $k \in \mathtt{I}_{\bar{j}\bar{r}}$, and $B_k = 0$ otherwise.
- $C_k = 1$ if $k \in \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}$, and $C_k = 0$ otherwise.

As observed in the proof of Lemma 16, $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \geq 2$ and $|\mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}| \geq 2$. Let us call the coefficient matrix of the above system of equations $M^{(i,j,r,s)}$, its first row as $A^{(i,j,r,s)}$, second row as $B^{(i,j,r,s)}$ and third row as $C^{(i,j,r,s)}$. Let us write $([q])_4$ as union of three index sets, $([q])_4 = \mathtt{J}_1 \sqcup \mathtt{J}_2 \sqcup \mathtt{J}_3$, where $\mathtt{J}_i$ are defined as follows:

$$\mathtt{J}_1 := \{(i,j,r,s) : \mathrm{rank}(M^{(i,j,r,s)}) = 3\}$$
$$\mathtt{J}_2 := \{(i,j,r,s) : A^{(i,j,r,s)} = C^{(i,j,r,s)}\}\}$$
$$\mathtt{J}_2 := \{(i,j,r,s) : B^{(i,j,r,s)} = c_1 A^{(i,j,r,s)} \oplus c_2 C^{(i,j,r,s)} \text{ for } c_1, c_2 \neq 0\}$$

For $(i,j,r,s) \in \mathtt{J}_1$, the probability of the Z-variables satisfying the system of equations is $(2^n)_{t-3}/(2^n)_t \leq 8/2^{3n}$ for $\ell \leq 2^n/8$, since $t \leq 4\ell$. Thus we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_1} \mathsf{E}(i,j,r,s)\right] \leq \frac{q^4}{3 \cdot 2^{3n}} \tag{27}$$

Now let us define the equivalence relation over $([q])_2$ as $(i,j) \sim (r,s)$ if $\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} = \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}$. If $(i,j) \sim (r,s)$, then $A^{(i,j,r,s)} = C^{(i,j,r,s)}$, which implies $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j) \iff \mathtt{LightHash}^1_{\Pi_0}(m^r) = \mathtt{LightHash}^1_{\Pi_0}(m^s)$. Suppose the above relations partitions $([q])_2$ into c equivalence classes $([q])_2 = C_1 \sqcup \cdots \sqcup C_c$. For $a = 1, \ldots, c$, consider the events E_a that $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)$ for every $(i,j) \in C_a$. Thus from Eq. (26) we have that

$$\Pr[\mathsf{E}_a] \leq \frac{1}{2^n - 2\ell + 1}$$

since $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \leq 2\ell$ for all $(i,j) \in C_a$. Now we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_2} \mathsf{E}(i,j,r,s)\right] = \Pr\left[\bigvee_{a\in[c]} \bigvee_{(i,j),(r,s)\in C_a} \mathsf{E}(i,j,r,s)\right]$$

$$\leq \sum_{a=1}^{c} \Pr\left[\bigvee_{(i,j),(r,s)\in C_a} \mathsf{E}(i,j,r,s)\right]$$
$$= \sum_{a=1}^{c} \Pr[\mathsf{E}_a]\cdot \Pr\left(\bigvee_{(i,j),(r,s)\in C_a} \mathtt{LightHash}^2_{\Pi_0}(m^j) = \mathtt{LightHash}^2_{\Pi_0}(m^r) \,\middle|\, \mathsf{E}_a\right)$$
$$\leq \sum_{a=1}^{c} \frac{1}{2^n-2\ell+1}\cdot \min\left\{\frac{|C_a|^2}{2(2^n-2\ell+1)},1\right\}$$

where the last inequality follows from Eq. (26) and the facts that $A^{(i,j,r,s)}$ and $B^{(i,j,r,s)}$ are linearly independent, and that $|\mathtt{I}_{\bar{j}r}\sqcup\mathtt{I}_{j\bar{r}}\sqcup\mathtt{I}_{\bar{j}\bar{r}}|\leq 2\ell$ for all $(j,r)\in C_a$. Note that $1/(2^n-2\ell+1)\leq 2/2^n$ for $\ell\leq 2^n/8$. Subject to the condition that $\sum_{a=1}^{c}|C_a| = \binom{q}{2}$, the sum $\sum_{a=1}^{c}\min\{|C_a|^2/(2(2^n-2\ell+1)),1\}$ is maximized when $c = \lfloor\binom{q}{2}/2^{n/2}\rfloor+1$, $|C_a| = 2^{n/2}$ for $a\in[c-1]$ and $|C_c| = \binom{q}{2}-(c-1)2^{n/2}$, in which case we have

$$\sum_{c=1}^{a}\frac{2}{2^n}\cdot\min\left\{\frac{|C_a|^2}{2^n},1\right\}\leq\frac{q^2}{2\cdot 2^{3n/2}}+\frac{2}{2^n}.$$

Thus we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_1}\mathsf{E}(i,j,r,s)\right]\leq\frac{q^2}{2\cdot 2^{3n/2}}+\frac{2}{2^n} \tag{28}$$

Finally we consider $(i,j,r,s)\in\mathtt{J}_3$. In this case $B^{(i,j,r,s)} = c_1A^{(i,j,r,s)} + c_2C^{(i,j,r,s)}$. This linear dependence implies the following:

- $c_1 = 2^\beta$ and $c_2 = 2^\gamma$ for some β,γ.
- $(\mathtt{I}_{\bar{i}j}\sqcup\mathtt{I}_{i\bar{j}})\,\triangle\,(\mathtt{I}_{\bar{r}s}\sqcup\mathtt{I}_{r\bar{s}}) = \mathtt{I}_{\bar{j}r}\sqcup\mathtt{I}_{j\bar{r}}$.[9] Also $B_k, k\in\mathtt{I}_{\bar{j}r}$ are all distinct, and similarly, $B_k, k\in\mathtt{I}_{j\bar{r}}$ are all distinct
- $(\mathtt{I}_{\bar{i}j}\sqcup\mathtt{I}_{i\bar{j}})\cap(\mathtt{I}_{\bar{r}s}\sqcup\mathtt{I}_{r\bar{s}}) = \mathtt{I}_{\bar{j}\bar{r}}$. From the definition of the index sets, this reduces to $\mathtt{I}_{i\bar{j}}\cap\mathtt{I}_{\bar{r}s} = \mathtt{I}_{\bar{j}\bar{r}}$. If for $k\in\mathtt{I}_{\bar{j}\bar{r}}$, $\mathsf{Z}[k] = \mathsf{Z}^j[k'] = \mathsf{Z}^r[k']$, then $B_k = 2^{\ell_j-k'}+2^{\ell_r-k'}$. Since $2^a+2^b = 2^c+2^d$ implies either $(a,b)=(c,d)$ or $(a,b)=(d,c)$, and since in this case for every $k\in\mathtt{I}_{\bar{j}\bar{r}}$, $B_k = 2^\beta+2^\gamma$, we have $|\mathtt{I}_{\bar{j}\bar{r}}| = 1$.

Thus the following assumptions made in proof of Lemma 4 of [30] holds true:

- $B^{(i,j,r,s)}$ does not contain the same entry more than twice.
- $B^{(i,j,r,s)}$ contains at least two different non-zero entries.
- Each of $A^{(i,j,r,s)}$ and $C^{(i,j,r,s)}$ contains at least three ones.

The rest of the analysis is exactly the one presented in the proof of Lemma 4 of [30], except the ignorable fact that the coefficient of $\mathsf{Z}^j[k']$ is $2^{\ell_j-k'}$ (instead of $2^{k'}$ as in the [30]), which however makes no changes in the argument presented. Thus following the proof of Lemma 4 of [30], we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_3}(i,j,r,s)\right]\leq\frac{24q^2}{(2^n-4\ell+1)(2^n-4\ell+2)}\leq\frac{96q^2}{2^{2n}} \tag{29}$$

for $\ell\leq 2^n/8$.

Combining Eqs. (27), (28) and (29) we have our result.

[9] For two sets A,B, we denote their symmetric difference as $A\triangle B := (A\setminus B)\cup(B\setminus A)$

The probability analysis of the events $\mathsf{AP2}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{AP1}^{c_1,c_2}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ are similar to the analysis of the events $\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{COLL}^{c_1,c_2}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})$, respectively, and we get the same probability bounds.

The exact same arguments given to prove Lemma 13 can be used to prove the following statement, keeping in mind that we do not need to consider the events E_1 and E_2, described in the proof of Lemma 13, for `LightHash`:

Lemma 18. *For $\ell \le 2^{n-2}$, $m \ne m' \in (\{0,1\}^{n-2})^{\le \ell}$, and $c \in \{0,1\}^2$, we have*

$$\Pr\left(\mathtt{LightHash}^1_{\Pi_0}(m) \oplus \mathtt{LightHash}^1_{\Pi_0}(m') = c\|0^{n-2}\right) \le \frac{8\ell}{2^n}$$
$$\Pr\left(\mathtt{LightHash}^2_{\Pi_0}(m) \oplus \mathtt{LightHash}^2_{\Pi_0}(m') = c\|0^{n-2}\right) \le \frac{8\ell}{2^n}$$

Corollary 7.

$$\Pr\left(\mathsf{COLL1}^c_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{4\ell q^2}{2^n} \qquad \Pr\left(\mathsf{COLL2}^c_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{4\ell q^2}{2^n}$$
$$\Pr\left(\mathsf{MC1}_{\mathtt{TLightHash}_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{16\ell q^2}{s^2 \cdot 2^n} \qquad \Pr\left(\mathsf{MC2}_{\mathtt{TLightHash}_{\Pi_0}}(\boldsymbol{m})\right) \le \frac{16\ell q^2}{s^2 \cdot 2^n}$$

Thus we get our desired result:

Lemma 19. *$\mathtt{TLightHash}_{\Pi_0}$ is a $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CfH, where*

$$\epsilon_1(\rho) = \frac{8\ell q^2}{2^n}, \qquad \epsilon_2(\rho,3) = \frac{8\ell q^2}{2^{2n}}, \qquad \epsilon_3(\rho,s) = \frac{16\ell q^2}{s^2 \cdot 2^n}, \qquad \delta(\rho) = \frac{8q^2}{2^{2n}}$$
$$\epsilon_2(\rho,4) = 8 \cdot \left(\frac{q^4}{3 \cdot 2^{3n}} + \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n} + \frac{96q^2}{2^{2n}}\right)$$

6.4 Instantiating `1k-DBHtS`

For any $\pi \in \mathcal{P}(n)$, we define

$$\mathtt{1k\text{-}PMAC+}_\pi := \mathtt{1k\text{-}DBHtS}_{\pi,\mathtt{TPHash}} \qquad \mathtt{1k\text{-}LightMAC+}_\pi := \mathtt{1k\text{-}DBHtS}_{\pi,\mathtt{TLightHash}}$$

Using Lemma 15 and 19 in Theorem 3 we have:

Corollary 8. *Let $c, q, \ell, \sigma \ge 0$ satisfying $q, \ell \le \sigma$ and $\overline{\sigma} = \sigma + 2q \le 2^{n-3}$. Then, for $\rho = (q, \ell, \sigma)$ and $\rho' = (2, \ell, 2\ell)$, the PRF advantage of any ρ-distinguisher $\mathcal{A}$ against `1k-PMAC+`$_\Pi$ satisfies*

$$\mathbf{Advt}^{\mathrm{prf}}_{\mathtt{1k\text{-}PMAC+}_\Pi}(\mathcal{A}) \le \frac{39q}{2^n} + \frac{64q}{2^{5n/4}} + \frac{16\ell q^2}{2^{3n}} + \frac{q^2(32\ell^2 + 260\ell + 1536)}{2^{2n}} + \frac{q\overline{\sigma}^3 + 2080\ell q^2\overline{\sigma}^2}{2^{3n}} + \frac{128q^3}{2^{9n/4}} + \frac{96q\overline{\sigma}}{2^{3n/2}}$$

Corollary 9. *Let $c, q, \ell, \sigma \ge 0$ satisfying $q, \ell \le \sigma$ and $\overline{\sigma} = \sigma + 2q \le 2^{n-3}$. Then, for $\rho = (q, \ell, \sigma)$ and $\rho' = (2, \ell, 2\ell)$, the PRF advantage of any ρ-distinguisher $\mathcal{A}$ against `1k-LightMAC+`$_\Pi$ satisfies*

$$\mathbf{Advt}^{\mathrm{prf}}_{\mathtt{1k\text{-}LightMAC+}_\Pi}(\mathcal{A}) \le \frac{39q}{2^n} + \frac{1560q^2}{2^{2n}} + \frac{80q^2\ell}{2^{2n}} + \frac{126q\overline{\sigma}^3}{2^{3n}} + \frac{160\ell q^2\overline{\sigma}^2}{2^{3n}} + \frac{16q\overline{\sigma}}{2^{3n/2}}.$$

7 PRF Security of Sum of r Even-Mansour

For any $r \geq 2$, let $(\pi_1, \ldots, \pi_r) \twoheadleftarrow \mathcal{P}(n)^r$ be a tuple of r permutations of $\{0,1\}^n$ and let $(K_1, \ldots, K_r) \in (\{0,1\}^n)^r$ be a r-tuple of n-bit strings.

One-round Even-Mansour construction is a keyed permutation of $\{0,1\}^n$ defined by the mapping

$$x \longmapsto \pi_1(x \oplus K_1) \oplus K_1,$$

where K_1 denotes the key.

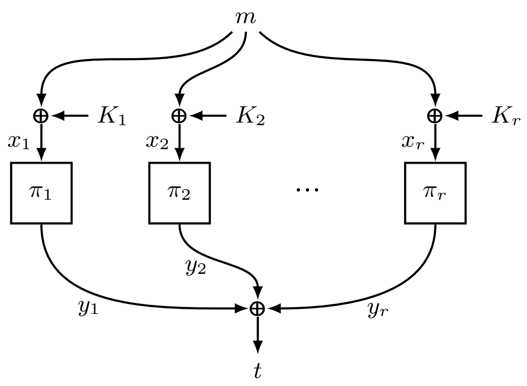

Fig. 6. The π-$\mathtt{SOEM}^r$ construction instantiated with key $K = (K_1, \ldots, K_r)$.

The r-sum of Even-Mansour construction, π-$\mathtt{SOEM}^r$ is a length-preserving keyed function of $\{0,1\}^n$ defined by the mapping

$$m \longmapsto \bigoplus_{i=1}^{r} \pi_i(m \oplus K_i),$$

where $K = (K_1, \ldots, K_r)$ denotes the key. See Figure 6 for a pictorial illustration. Notice that we skipped the post-permutation key masking. This is motivated by a similar simplification [39] by Sibleyras and Todo who studied the $r = 2$ case. Thus, we study the same problem for any arbitrary $r \geq 2$.

Theorem 4. *Fix some $r \geq 2$, $q + p \leq 2^{\frac{r}{r+1}n - \log_2(n)}$, and $\Pi = (\Pi_1, \ldots, \Pi_r) \twoheadleftarrow \mathcal{P}(n)^r$. For any (q,p)-distinguisher $\mathcal{A}$ we have*

$$\mathbf{Advt}^{\mathrm{prf}}_{\Pi\text{-}\mathtt{SOEM}^r}(\mathcal{A}) \leq \frac{4}{2^n} + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{10\sqrt{nq}(2p+2q)^{r-1}}{2^{n(r-1)}} + \frac{10q(2p+2q)^r}{2^{nr}}.$$

Proof. For the purpose of this proof let $\mathsf{F}_\mathsf{K}(\cdot) = \Pi\text{-}\mathtt{SOEM}^r_\mathsf{K}(\cdot)$, and let $\Gamma \twoheadleftarrow \{0,1\}^n$. $\mathcal{A}$'s goal is to distinguish between the *real* oracle $(\mathsf{F}_\mathsf{K}, \Pi^\pm)$ and the *ideal* oracle $(\Gamma, \Pi^\pm)$, where F_K and Γ are referred as the construction oracle and $\Pi^\pm$ is referred as the primitive oracle.
Fix a (q,p)-distinguisher $\mathcal{A}$. Let

- $(\mathsf{M}^i, \mathsf{T}^i)$ denote the i-th query-response tuple corresponding to the construction oracle. Let $\mathsf{M} := \{\mathsf{M}^i : i \in [q]\}$ and $\mathsf{T} := \{\mathsf{T}^i : i \in [q]\}$.
- $(\mathsf{U}_j^i, \mathsf{V}_j^i)$ denote the i-th query-response tuple corresponding to the permutation Π_j. Unless stated otherwise, we assume that all these queries are in the forward direction. Let $\mathsf{U}_j := \{\mathsf{U}_j^i : i \in [p]\}$, $\mathsf{V}_j := \{\mathsf{V}_j^i : i \in [p]\}$, $\mathsf{U} := (\mathsf{U}_1, \ldots, \mathsf{U}_r)$, and $\mathsf{V} := (\mathsf{V}_1, \ldots, \mathsf{V}_r)$.
- $(\mathsf{X}_j^i, \mathsf{Y}_j^i)$ denote the input-output tuple to the j-th permutation, for all $j \in [r]$, within the i-th construction query in the real world, i.e., $\mathsf{X}_j^i = \mathsf{M}^i \oplus \mathsf{K}_j$. Let $\mathsf{X}^i := (\mathsf{X}_j^i : j \in [r])$ and $\mathsf{Y}^i := (\mathsf{Y}_j^i : j \in [r])$. Let $\mathsf{X} := \{\mathsf{X}^i : i \in [q]\}$ and $\mathsf{Y} := \{\mathsf{Y}^i : i \in [q]\}$.

We study a modified game where the real oracle releases (X, Y) to $\mathcal{A}$ once the query-response phase is over, but before $\mathcal{A}$ outputs. This obviously does not decrease $\mathcal{A}$'s advantage.

Ideal World Transcript Extension: Naturally, in the ideal world, the sampling is extended to generate this additional information. We have

$$\mathcal{SC}(\mathsf{T}, \mathsf{V}) = \left\{(\mathsf{T}^i, \mathsf{V}_1^{j_1}, \mathsf{V}_2^{j_2}, \ldots, \mathsf{V}_r^{j_r}) \in \mathsf{T} \times \mathsf{V} : \bigoplus_{k=1}^{r} \mathsf{V}_k^{j_k} = \mathsf{T}^i\right\}$$
$$\mu(\mathsf{T}, \mathsf{V}) = |\mathcal{SC}(\mathsf{T}, \mathsf{V})|$$

Further due to the increasing nature of $\mu(\mathsf{T}, \cdot)$, $\mu(\mathsf{T}, \mathsf{V}) \le \mu(\mathsf{T}, p+q)$. We define the predicate

$$\mathsf{LSC}(\mathsf{T}, p+q) : \left(\mu(\mathsf{T}, p+q) > \frac{q(p+q)^r}{2^n} + 2(p+q)^{r-1}\sqrt{nq}\right)$$

The subsequent two-step sampling mechanism for (X, Y) in the ideal world is defined under the condition that $\neg\mathsf{LSC}(\mathsf{T}, p+q)$ holds:

1. In the first step, a dummy key tuple is sampled uniformly at random, i.e., $\mathsf{K} \twoheadleftarrow (\{0,1\}^n)^r$, which determines $\mathsf{X}_j^i := \mathsf{M}^i \oplus \mathsf{K}_j$. Consider the following predicates:

$$\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}) : \exists\, i \in [q],\, j_1, \ldots, j_r \in [p] \text{ such that } \left(\forall k \in [r],\, \mathsf{X}_k^i = \mathsf{U}_k^{j_k}\right)$$
$$\mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}) : \exists\, (i, j_1, j_2, \ldots, j_r) \in \mathcal{SC}(\mathsf{T}, \mathsf{V}),\, k \in [r], \text{ such that } (\mathsf{X}_k^i \ne \mathsf{U}_k^{j_k}) \text{ and } \left(\forall k' \ne k,\, \mathsf{X}_{k'}^i = \mathsf{U}_{k'}^{j_{k'}}\right)$$

   Going forward we assume that $\neg(\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}) \vee \mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}))$ holds. For each $i \in [q]$:
   (a) if there exists $j \in [p]$ and $k \in [r]$, such that $\mathsf{X}_k^i = \mathsf{U}_k^j$, then define $\mathsf{Y}_k^i := \mathsf{V}_k^j$;
   (b) let $\mathcal{I}_i = \{j \in [r] : \mathsf{X}_j^i \notin \mathsf{U}_j\}$ to be the set of permutation indices with fresh input for the i-th construction query.
   (c) let $\sim$ be a relation on $[q]$ defined as: $i_1 \sim i_2 \iff \mathcal{I}_{i_1} = \mathcal{I}_{i_2}$. Clearly, $\sim$ is an equivalence relation. Let $\mathcal{Q}_{(0)}^{(1)} \sqcup \ldots \mathcal{Q}_{(0)}^{(r)} \sqcup \mathcal{Q}_{(1)} \sqcup \ldots \sqcup \mathcal{Q}_{(c)}$ denote the corresponding partitioning of $[q]$, where $\mathcal{Q}_{(0)}^{(j)} = \{i \in [q] : \mathcal{I}_i = \{j\}\}$. Let

$|\mathcal{Q}_{(0)}^{(j)}| = q_0^{(j)}$, $q_0 := \sum_{j\in[r]} q_0^{(j)}$ and $|\mathcal{Q}_{(i)}| = q_i$. Then $q_0 + \sum_{i\in[c]} q_i = q$. Also, note that, $c \le \sum_{j=2}^{r-1} \binom{r}{j} \le 2^r$.

(d) for all $j \in [r]$ and $i \in \mathcal{Q}_{(0)}^{(j)}$, define $\mathsf{Y}_j^i := \oplus_{l\in[r]\setminus j} \mathsf{Y}_l^i \oplus \mathsf{T}^i$ and

$$\mathsf{Y}^{(0)} = \{\mathsf{Y}_j^i \oplus_{l\in[r]\setminus j} \mathsf{Y}_l^i \oplus \mathsf{T}^i \;:\; j \in [r], i \in \mathcal{Q}_{(0)}^{(j)}\}.$$

This concludes the first step. We encourage the readers to verify that at the end of this step Y_j^i is undefined for exactly the indices in $\mathcal{I}_i$ and $|\mathcal{I}_i| \ge 2$. Furthermore, due to $\neg(\mathsf{KG}(\mathsf{MU},\mathsf{K}) \vee \mathsf{SC}(\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K}))$, the partially defined (X,Y) is permutation-consistent.

Constrained system formulation: For each $i \in [c]$, let $\mathcal{J}_{(i)} = \{j_1,\ldots,j_{t_i}\}$ denote the set of permutation indices with fresh input for the i-th equivalence class $\mathcal{Q}_{(i)}$. Let $r_i = q_i t_i$.

Then, for each $i \in [c]$, we obtain a (q_i, r_i, t_i)-constrained system $\mathbb{S}^{(i)}$:

$$\mathbb{S}^{(i)} = \left\{ \bigoplus_{k\in\mathcal{J}_{(i)}} \mathsf{Y}_k^j = \mathsf{T}^j \bigoplus_{k'\in[r]\setminus\mathcal{J}_{(i)}} \mathsf{Y}_{k'}^j \right\}_{j\in\mathcal{Q}_{(i)}}$$

which is binary, acyclic, partite, isolate and t_i-regular.

2. In the second step, we sample a solution for each of the c constrained systems. First fix any arbitrary ordering of $\mathbb{S}^{(1)},\ldots,\mathbb{S}^{(c)}$. Now, for the i-th system:
   - let $\mathcal{R}_{\le(i-1)}^{(j)} = \mathsf{V}_j \cup \{\mathsf{Y}_j^k \;:\; k \in \mathcal{Q}_{(0)}^{(j)}\} \cup \{\mathsf{Y}_j^k \;:\; k \in \mathcal{Q}_{(1)} \sqcup \ldots \sqcup \mathcal{Q}_{(i-1)}\}$, for all $j \in [r]$, and let $|\mathcal{R}_{\le(i-1)}^{(j)}| = r_{\le(i-1)}^{(j)} \le (p+q)$,
   - let $\mathcal{R}_{\le(i-1)} = (\mathcal{R}_{\le(i-1)}^{(j)} \;:\; j \in [r])$ and $\widehat{\mathcal{R}}_{\le(i-1)} = (\mathcal{R}_{\le(i-1)}^{(j)} \;:\; j \in \mathcal{J}_{(i)})$,
   - let $\mathsf{T}^{(i)} = (\mathsf{T}^k \;:\; k \in \mathcal{Q}_{(i)})$ and $\widehat{\mathsf{T}}^{(i)} = (\mathsf{T}^k \oplus_{j\in[r]\setminus\mathcal{J}_{(i)}} \mathsf{Y}_j^k \;:\; k \in \mathcal{Q}_{(i)})$. Then, $|\mathsf{T}^{(i)}|, |\widehat{\mathsf{T}}^{(i)}| \le q_i$.
   - let $\mathsf{Y}^{(i)} = \{\mathsf{Y}_j^k \;:\; k \in \mathcal{Q}_{(i)}, j \in \mathcal{J}_{(i)}\}$. Then, $|\mathsf{Y}^{(i)}| = r_i$.

   We sample $\mathsf{Y}^{(i)} \twoheadleftarrow (\mathbb{S}^{(i)} \mid \widehat{\mathcal{R}}_{\le(i-1)})$, where using Theorem 1, we have

$$\eta(\mathbb{S}^{(i)} \mid \widehat{\mathcal{R}}_{\le(i-1)}) \ge \frac{\prod_{j\in\mathcal{J}_{(i)}} (2^n - r_{\le(i-1)}^{(j)})_{q_i}}{2^{nq_i}} \left(1 - \varepsilon^{(i)}\right) \tag{30}$$

$$\varepsilon^{(i)} \le \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\le(i-1)})}{2^{n(t_i-1)}} + \frac{2q_i \Delta_{\mathbb{S}^{(i)}}}{2^{n(t_i-1)}} + \frac{6q_i(p+q)^{t_i}}{2^{nt_i}} \tag{31}$$

   Since the solution for each system is sampled in a consistent manner given a consistent solution for the previous system, the cumulative sampling is also permutation-compatible. This completes the second step.

At this stage the full transcript in the ideal world, i.e., $\Theta_{\mathrm{id}} = (\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K},\mathsf{Y})$ is fully determined.

Some Notations on Transcripts: For any wo $\in \{\mathrm{re}, \mathrm{id}\}$, and $\Theta_{\mathrm{wo}} = (\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K},\mathsf{Y})$, let:

- $\Theta_{\mathrm{wo}}^{\mathsf{key}}$ denote the restriction of Θ_{wo} to the key K,

- $\Theta_{\mathrm{wo}}^{\mathsf{con}}$ denote the restriction of Θ_{wo} to the construction query-response tuple (M, T),
- $\Theta_{\mathrm{wo}}^{\mathsf{prim}}$ denote the restriction of Θ_{wo} to the key (U, V),
- $\Theta_{\mathrm{wo}}^{\mathsf{int}}$ denote the restriction of Θ_{wo} to the construction-specific primitive query-response (X, Y).

Bad Transcript Definition and Analysis: A transcript $\omega = (M, T, U, V, K, Y) \in \Omega$ is said to be *bad* if and only if $\mathsf{LSC}(T, p+q) \vee \mathsf{KG}(M, U, K) \vee \mathsf{SC}(M, T, U, V, K)$ holds.

Lemma 20.

$$\Pr(\Theta_{\mathrm{id}} \in \Omega_{\mathrm{bad}}) \leq \frac{4}{2^n} + \frac{2\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}} + \frac{2q(p+q)^r}{2^{nr}}$$

Proof. We have

$$\begin{aligned}
\Pr(\Theta_{\mathrm{id}} \in \Omega_{\mathrm{bad}}) &= \Pr(\mathsf{LSC}(\mathsf{T}, p+q) \vee \mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}) \vee \mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K})) \\
&\leq \Pr(\mathsf{LSC}(\mathsf{T}, p+q)) + \Pr(\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K})) + \Pr(\mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}) \mid \neg\mathsf{LSC}(\mathsf{T}, p+q)) \\
&\leq \frac{4}{2^n} + \frac{qp^r}{2^{nr}} + \frac{q(p+q)^r}{2^{nr}} + \frac{2(p+q)^{r-1}\sqrt{nq}}{2^{n(r-1)}},
\end{aligned}$$

where the first term on the right hand side corresponds to $\Pr(\mathsf{LSC}(\mathsf{T}, p+q))$ and follows from Lemma 1, the second term corresponds to $\Pr(\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}))$ and follows from the uniformity of K. The last two terms correspond to $\Pr(\mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}) \mid \neg\mathsf{LSC}(\mathsf{T}, p+q))$. To argue this, first notice that given $\neg\mathsf{LSC}(\mathsf{T}, p+q)$, we have

$$\mu(\mathsf{T}, \mathsf{V}) \leq \frac{q(p+q)^r}{2^n} + 2(p+q)^{r-1}\sqrt{nq}.$$

For each choice of $k \in [r]$, the predicate $\forall k' \neq k$, $\mathsf{X}_{k'}^i = \mathsf{U}_{k'}^{j_{k'}}$ is satisfied with at most $2^{-n(r-1)}$ probability. Now, we get the desired terms using union bound. □

Good Transcript Analysis: Let $\omega = (M, T, U, V, K, Y)$ be a good transcript. Since the transcript is good, $\neg(\mathsf{LSC}(T, p+q) \vee \mathsf{KG}(M, U, K) \vee \mathsf{SC}(M, T, U, V, K))$ holds.

Before moving forward, recall the notations introduced while discussing the sampling in the ideal world. We assume analogous notations for any arbitrary transcript.

We also ignore the probability computation of obvious events, such as: the message tuple being realized.

Real World: In the real world, we have

$$\begin{aligned}
\Pr(\Theta_{\mathrm{re}} = \omega) &= \Pr\left(\Theta_{\mathrm{re}}^{\mathrm{key}} = K, \Theta_{\mathrm{re}}^{\mathrm{prim}} = (U, V), \Theta_{\mathrm{re}}^{\mathrm{int}} = (X, Y), \Theta_{\mathrm{re}}^{\mathrm{con}} = (M, T)\right) \\
&= \Pr\left(\Theta_{\mathrm{re}}^{\mathrm{key}} = K\right) \times \Pr\left(\Theta_{\mathrm{re}}^{\mathrm{prim}} = (U, V)\right) \times \Pr\left(\Theta_{\mathrm{re}}^{\mathrm{int}} = (X, Y) \,\middle|\, \Theta_{\mathrm{re}}^{\mathrm{key}}, \Theta_{\mathrm{re}}^{\mathrm{prim}}\right)
\end{aligned}$$

$$= \frac{1}{2^{nr}} \times \frac{1}{(2^n)_p^r} \times \Pr\left(\Theta_{\text{re}}^{\text{int}} = (X, Y) \,\middle|\, \Theta_{\text{re}}^{\text{key}}, \Theta_{\text{re}}^{\text{prim}}\right),$$

where the first term on the right hand side follows from the uniformity of K, the second term follows from the uniformity of $\Pi = (\Pi_1, \dots, \Pi_r)$.

As for the last term, consider the partition imposed by $\sim$ in an arbitrary order, and also the associated notations introduced earlier. Then, conditioned on $(\Theta_{\text{re}}^{\text{key}}, \Theta_{\text{re}}^{\text{prim}})$, we have

$$\Pr\left(\Theta_{\text{re}}^{\text{int}} = (X, Y) \,\middle|\, \Theta_{\text{re}}^{\text{key}}, \Theta_{\text{re}}^{\text{prim}}\right) = \prod_{j=1}^{r} \frac{1}{(2^n - p)_{q_0^{(j)}}} \times \prod_{\substack{i \in [c] \\ j' \in \mathcal{J}_{(i)}}} \frac{1}{(2^n - r_{\le(i-1)}^{(j')})_{q_i}}.$$

Indeed, the first product term corresponds to the query indices with exactly one fresh primitive input, i.e. the ones in $\mathcal{Q}_{(0)}^{(j)}$ for some $j \in [r]$, and the second product correspond to the query indices with at least two fresh primitive inputs, computed using a simple application of chain rule over the partitions $\mathcal{Q}_{(1)}, \dots, \mathcal{Q}_{(c)}$. By combining everything, we have

$$\Pr(\Theta_{\text{re}} = \omega) = \frac{1}{2^{nr}} \times \frac{1}{(2^n)_p^r} \times \prod_{j=1}^{r} \frac{1}{(2^n - p)_{q_0^{(j)}}} \times \prod_{\substack{i \in [c] \\ j' \in \mathcal{J}_{(i)}}} \frac{1}{(2^n - r_{\le(i-1)}^{(j')})_{q_i}}, \tag{32}$$

Ideal World: In the ideal world, we have

$$\begin{aligned}
\Pr(\Theta_{\text{id}} = \omega) &= \Pr\left(\Theta_{\text{id}}^{\text{key}} = K, \Theta_{\text{id}}^{\text{prim}} = (U, V), \Theta_{\text{id}}^{\text{int}} = (X, Y), \Theta_{\text{id}}^{\text{con}} = (M, T)\right) \\
&= \Pr\left(\Theta_{\text{id}}^{\text{key}} = K\right) \times \Pr\left(\Theta_{\text{id}}^{\text{con}} = (M, T)\right) \times \Pr\left(\Theta_{\text{id}}^{\text{prim}} = (U, V)\right) \\
&\qquad \times \Pr\left(\Theta_{\text{id}}^{\text{int}} = (X, Y) \,\middle|\, \Theta_{\text{id}}^{\text{key}}, \Theta_{\text{id}}^{\text{prim}}, \Theta_{\text{id}}^{\text{con}}\right) \\
&= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \Pr\left(\Theta_{\text{id}}^{\text{int}} = (X, Y) \,\middle|\, \Theta_{\text{id}}^{\text{key}}, \Theta_{\text{id}}^{\text{prim}}, \Theta_{\text{id}}^{\text{con}}\right) \\
&= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{i \in [c]} \Pr\left(\mathsf{Y}^{(i)} = Y^{(i)} \,\middle|\, \widehat{\mathcal{R}}_{\le(i-1)}\right) \\
&= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{i \in [c]} \frac{1}{\eta(\mathbb{S}^{(i)} \,|\, \widehat{\mathcal{R}}_{\le(i-1)})}
\end{aligned}$$

where the first three terms are obvious. The fourth term corresponds to the indices in $\mathcal{Q}_{(i)}$ for all $i \in [c]$. Further, using (30), we have

$$\begin{aligned}
\Pr(\Theta_{\text{id}} = \omega) &\ge \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{\substack{i \in [c] \\ j' \in [r]}} \frac{2^{nq_i}}{\left(1 - \varepsilon^{(i)}\right) (2^n - r_{\le(i-1)}^{(j')})_{q_i}} \\
&= \frac{1}{2^{nr}} \times \frac{1}{2^{nq_0}} \times \frac{1}{(2^n)_p^r} \times \prod_{\substack{i \in [c] \\ j' \in [r]}} \frac{1}{\left(1 - \varepsilon^{(i)}\right) (2^n - r_{\le(i-1)}^{(j')})_{q_i}},
\end{aligned} \tag{33}$$

where the equality follows from the fact that $q = q_0 \sum_{i\in[c]} q_i$.

The Ratio: On dividing (32) by (33), we have

$$\frac{\Pr(\Theta_{\mathrm{re}} = \omega)}{\Pr(\Theta_{\mathrm{id}} = \omega)} \geq \prod_{i\in[c]} \left(1 - \varepsilon^{(i)}\right) \tag{34}$$

$$\geq 1 - \sum_{i\in[c]} \varepsilon^{(i)}$$

$$\geq 1 - \underbrace{\sum_{i\in[c]} \left(\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(t_i-1)}} + \frac{2q_i \varDelta_{\mathbb{S}^{(i)}}}{2^{n(t_i-1)}} + \frac{6q_i(p+q)^{t_i}}{2^{nt_i}} \right)}_{\varepsilon_{\mathrm{ratio}}(\omega)}. \tag{35}$$

Now, we have

$$\mathbb{E}\left(1_{\mathrm{good}}\varepsilon_{\mathrm{ratio}}\right) = \sum_{i\in[c]} \mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}}) \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(t_i-1)}}\right) + \sum_{i\in[c]} \frac{2\mathbb{E}(q_i)\,\mathbb{E}(\varDelta_{\mathbb{S}^{(i)}})}{2^{n(t_i-1)}} + \sum_{i\in[c]} \frac{6\mathbb{E}(q_i)(p+q)^{t_i}}{2^{nt_i}} \tag{36}$$

$$\leq \sum_{i\in[c]} \mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}}) \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(t_i-1)}}\right) + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{6q(2(p+q))^r}{2^{nr}} \tag{37}$$

where the first equality follows from linearity of expectation and the fact that $\mathbb{E}(\chi\mathtt{R}) \leq \mathbb{E}(\mathtt{R})$ for any non-negative random variable $\mathtt{R}$ and indicator random variable χ. The second/third term in the second inequality follows from $\mathbb{E}(q_i) \leq qp^{r-t_i}/2^{n(r-t_i)} \leq q(p+q)^{r-t_i}/2^{n(r-t_i)}$, $t_i \geq 2$, $c \leq 2^r$. Additionally, due to the uniformity of T and $q < 2^n$, $\mathbb{E}(\varDelta_{\mathbb{S}^{(i)}}) \leq 4n$. Now, for the first term, when $t_i = r$, we have

$$\mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}}) \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(r-1)}}\right) \leq \frac{2\mu(\mathsf{T}, \mathsf{V})}{2^{n(r-1)}}$$

$$\leq \frac{2\mu^r(\mathsf{T}, p+q)}{2^{n(r-1)}}$$

$$\leq \frac{2q(p+q)^r}{2^{nr}} + \frac{4\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}}, \tag{38}$$

where the last inequality follows from $1_{\mathrm{good}}(\Theta_{\mathrm{id}}) = 1$. For, $t_i < r$, let $\mathcal{J}_{(i)} = \{j_1, \ldots, j_{t_i}\}$, $[r] \setminus \mathcal{J}_{(i)} = \{j'_1, \ldots, j'_{r-t_i}\}$, and

$$\mathcal{KSC}_{(i)} := \left\{ (\mathsf{T}^{i'}, \mathsf{V}^{k_1}_{j'_1}, \ldots, \mathsf{V}^{k_{r-t_i}}_{j'_{r-t_i}}, \mathsf{Z}_{\mathcal{J}_{(i)}}) \in \mathcal{SC}(\mathsf{T}, \mathsf{V}_{[r]\setminus\mathcal{J}_{(i)}}, \mathcal{R}^{(\mathcal{J}_{(i)})}_{\leq(i-1)}) \;:\; \mathsf{X}^{i'}_{j'_l} = \mathsf{U}^{k_l}_{j'_l} \right\}$$

Then, $|\mathcal{KSC}_{(i)}| = \mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})$, and thus

$$\mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}}) \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(t_i-1)}}\right) \leq \frac{2}{2^{n(t_i-1)}} \mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}}) |\mathcal{KSC}_{(i)}|\right)$$

$$\leq \frac{2}{2^{n(t_i-1)}} \times \frac{\mu^r(\mathsf{T}, p+q)}{2^{n(r-t_i)}}$$
$$\leq \frac{2q(p+q)^r}{2^{nr}} + \frac{4\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}} \tag{39}$$

where the second inequality follows from the uniformity of K, and the last inequality follows from $1_{\text{good}}(\Theta_{\text{id}}) = 1$. Using (38) and (39) in (37), we have

$$\mathbb{E}\left(1_{\text{good}}\varepsilon_{\text{ratio}}\right) \leq \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{8\sqrt{nq}(2(p+q))^{r-1}}{2^{n(r-1)}} + \frac{8q(2(p+q))^r}{2^{nr}} \tag{40}$$

Finally, using the fine-grained variant of the Expectation method (see Lemma 2) along with Lemma 20 and (40), we have

$$\mathbf{Advt}^{\text{prf}}_{\Pi\text{-}\mathtt{SOEM}^r}(\mathcal{A}) \leq \frac{4}{2^n} + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{10\sqrt{nq}(2p+2q)^{r-1}}{2^{n(r-1)}} + \frac{10q(2p+2q)^r}{2^{nr}},$$

which completes the proof. □

References

1. Andreeva, E., Daemen, J., Mennink, B., Assche, G.V.: Security of keyed sponge constructions using a modular proof approach. In: Leander, G. (ed.) Fast Software Encryption - FSE 2015, Revised Selected Papers. Lecture Notes in Computer Science, vol. 9054, pp. 364–384. Springer (2015). https://doi.org/10.1007/978-3-662-48116-5_18
2. Aumasson, J., Bernstein, D.J.: Siphash: A fast short-input PRF. In: Galbraith, S.D., Nandi, M. (eds.) Progress in Cryptology - INDOCRYPT 2012. Proceedings. Lecture Notes in Computer Science, vol. 7668, pp. 489–508. Springer (2012). https://doi.org/10.1007/978-3-642-34931-7_28
3. Banik, S., Pandey, S.K., Peyrin, T., Sasaki, Y., Sim, S.M., Todo, Y.: GIFT: A small present - towards reaching the limit of lightweight encryption. In: Fischer, W., Homma, N. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2017, Proceedings. Lecture Notes in Computer Science, vol. 10529, pp. 321–345. Springer (2017). https://doi.org/10.1007/978-3-319-66787-4_16
4. Bellare, M., Impagliazzo, R.: A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptol. ePrint Arch. p. 24 (1999)
5. Bellare, M., Kilian, J., Rogaway, P.: The security of the cipher block chaining message authentication code. J. Comput. Syst. Sci. **61**(3), 362–399 (2000). https://doi.org/10.1006/JCSS.1999.1694
6. Bierbrauer, J., Johansson, T., Kabatianskii, G., Smeets, B.J.M.: On families of hash functions via geometric codes and concatenation. In: Stinson, D.R. (ed.) Advances in Cryptology - CRYPTO 1993, Proceedings. Lecture Notes in Computer Science, vol. 773, pp. 331–342. Springer (1993). https://doi.org/10.1007/3-540-48329-2_28
7. Black, J., Rogaway, P.: A block-cipher mode of operation for parallelizable message authentication. In: Knudsen, L.R. (ed.) Advances in Cryptology - EUROCRYPT 2002, Proceedings. Lecture Notes in Computer Science, vol. 2332, pp. 384–397. Springer (2002). https://doi.org/10.1007/3-540-46035-7_25

8. den Boer, B.: A simple and key-economical unconditional authentication scheme. J. Comput. Secur. **2**, 65–72 (1993)
9. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: an ultra-lightweight block cipher. In: Paillier, P., Verbauwhede, I. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2007, Proceedings. Lecture Notes in Computer Science, vol. 4727, pp. 450–466. Springer (2007). https://doi.org/10.1007/978-3-540-74735-2_31
10. Chakraborti, A., Nandi, M., Talnikar, S., Yasuda, K.: On the composition of single-keyed tweakable even-mansour for achieving BBB security. IACR Trans. Symmetric Cryptol. **2020**(2), 1–39 (2020). https://doi.org/10.13154/TOSC.V2020.I2.1-39
11. Chen, Y.L., Dutta, A., Nandi, M.: Multi-user BBB security of public permutations based MAC. Cryptogr. Commun. **14**(5), 1145–1177 (2022). https://doi.org/10.1007/S12095-022-00571-W
12. Chen, Y.L., Lambooij, E., Mennink, B.: How to build pseudorandom functions from public random permutations. In: Boldyreva, A., Micciancio, D. (eds.) Advances in Cryptology - CRYPTO 2019, Proceedings, Part I. Lecture Notes in Computer Science, vol. 11692, pp. 266–293. Springer (2019). https://doi.org/10.1007/978-3-030-26948-7_10
13. Cogliati, B., Dutta, A., Nandi, M., Patarin, J., Saha, A.: Proof of mirror theory for a wide range of $\xi _{\max }$. In: Hazay, C., Stam, M. (eds.) Advances in Cryptology - EUROCRYPT 2023, Proceedings, Part IV. Lecture Notes in Computer Science, vol. 14007, pp. 470–501. Springer (2023). https://doi.org/10.1007/978-3-031-30634-1_16
14. Cogliati, B., Jha, A., Nandi, M.: How to build optimally secure prfs using block ciphers. In: Moriai, S., Wang, H. (eds.) Advances in Cryptology - ASIACRYPT 2020, Proceedings, Part I. Lecture Notes in Computer Science, vol. 12491, pp. 754–784. Springer (2020). https://doi.org/10.1007/978-3-030-64837-4_25
15. Cogliati, B., Patarin, J.: Mirror theory: A simple proof of the pi+pj theorem with xi_max=2. IACR Cryptol. ePrint Arch. p. 734 (2020), https://eprint.iacr.org/2020/734
16. Cogliati, B., Seurin, Y.: Analysis of the single-permutation encrypted davies-meyer construction. Des. Codes Cryptogr. **86**(12), 2703–2723 (2018). https://doi.org/10.1007/S10623-018-0470-9
17. Dai, W., Hoang, V.T., Tessaro, S.: Information-theoretic indistinguishability via the chi-squared method. In: Katz, J., Shacham, H. (eds.) Advances in Cryptology - CRYPTO 2017, Proceedings, Part III. Lecture Notes in Computer Science, vol. 10403, pp. 497–523. Springer (2017). https://doi.org/10.1007/978-3-319-63697-9_17
18. Datta, N., Dutta, A., Nandi, M., Paul, G.: Double-block hash-then-sum: A paradigm for constructing BBB secure PRF. IACR Trans. Symmetric Cryptol. **2018**(3), 36–92 (2018). https://doi.org/10.13154/TOSC.V2018.I3.36-92
19. Datta, N., Dutta, A., Nandi, M., Paul, G., Zhang, L.: Single key variant of pmac_plus. IACR Trans. Symmetric Cryptol. **2017**(4), 268–305 (2017). https://doi.org/10.13154/TOSC.V2017.I4.268-305
20. Dinur, I.: Tight indistinguishability bounds for the XOR of independent random permutations by fourier analysis. In: Joye, M., Leander, G. (eds.) Advances in Cryptology - EUROCRYPT 2024, Proceedings, Part I. Lecture Notes in Computer Science, vol. 14651, pp. 33–62. Springer (2024). https://doi.org/10.1007/978-3-031-58716-0_2

21. Dutta, A., Nandi, M., Saha, A.: Proof of mirror theory for $\xi_{\max} = 2$. IEEE Trans. Inf. Theory **68**(9), 6218–6232 (2022). https://doi.org/10.1109/TIT.2022.3171178
22. Dutta, A., Nandi, M., Talnikar, S.: Permutation based EDM: an inverse free BBB secure PRF. IACR Trans. Symmetric Cryptol. **2021**(2), 31–70 (2021). https://doi.org/10.46586/TOSC.V2021.I2.31-70
23. Eberhard, S.: More on additive triples of bijections. CoRR **abs/1704.02407** (2017), http://arxiv.org/abs/1704.02407
24. Guo, J., Peyrin, T., Poschmann, A., Robshaw, M.J.B.: The LED block cipher. In: Preneel, B., Takagi, T. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2011. Proceedings. Lecture Notes in Computer Science, vol. 6917, pp. 326–341. Springer (2011). https://doi.org/10.1007/978-3-642-23951-9_22
25. Hoang, V.T., Tessaro, S.: Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In: Robshaw, M., Katz, J. (eds.) Advances in Cryptology - CRYPTO 2016, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9814, pp. 3–32. Springer (2016). https://doi.org/10.1007/978-3-662-53018-4_1
26. Iwata, T., Kurosawa, K.: OMAC: one-key CBC MAC. In: Johansson, T. (ed.) Fast Software Encryption - FSE 2003, Revised Papers. Lecture Notes in Computer Science, vol. 2887, pp. 129–153. Springer (2003). https://doi.org/10.1007/978-3-540-39887-5_11
27. Jha, A.: A note on the generalized sum-capture problem for rings. Online: https://ashwin-jha.github.io/papers/gsc.pdf (2024)
28. Jha, A., Nandi, M.: Tight security of cascaded LRW2. J. Cryptol. **33**(3), 1272–1317 (2020). https://doi.org/10.1007/S00145-020-09347-Y, https://doi.org/10.1007/s00145-020-09347-y
29. Jha, A., Nandi, M.: A survey on applications of h-technique: Revisiting security analysis of PRP and PRF. Entropy **24**(4), 462 (2022). https://doi.org/10.3390/E24040462
30. Kim, S., Lee, B., Lee, J.: Tight security bounds for double-block hash-then-sum macs. In: Advances in Cryptology - EUROCRYPT 2020, Proceedings, Part I. pp. 435–465 (2020). https://doi.org/10.1007/978-3-030-45721-1_16
31. Leurent, G., Nandi, M., Sibleyras, F.: Generic attacks against beyond-birthday-bound macs. In: Shacham, H., Boldyreva, A. (eds.) Advances in Cryptology - CRYPTO 2018. Proceedings, Part I. Lecture Notes in Computer Science, vol. 10991, pp. 306–336. Springer (2018). https://doi.org/10.1007/978-3-319-96884-1_11
32. Lucks, S.: The sum of prps is a secure PRF. In: Preneel, B. (ed.) Advances in Cryptology - EUROCRYPT 2000, Proceeding. Lecture Notes in Computer Science, vol. 1807, pp. 470–484. Springer (2000). https://doi.org/10.1007/3-540-45539-6_34
33. Luykx, A., Preneel, B., Tischhauser, E., Yasuda, K.: A MAC mode for lightweight block ciphers. In: Peyrin, T. (ed.) Fast Software Encryption - FSE 2016, Revised Selected Papers. Lecture Notes in Computer Science, vol. 9783, pp. 43–59. Springer (2016). https://doi.org/10.1007/978-3-662-52993-5_3
34. Mennink, B., Reyhanitabar, R., Vizár, D.: Security of full-state keyed sponge and duplex: Applications to authenticated encryption. In: Iwata, T., Cheon, J.H. (eds.) Advances in Cryptology - ASIACRYPT 2015, Proceedings, Part II. Lecture Notes in Computer Science, vol. 9453, pp. 465–489. Springer (2015). https://doi.org/10.1007/978-3-662-48800-3_19

35. Naito, Y.: Blockcipher-based macs: Beyond the birthday bound without message length. In: Advances in Cryptology - ASIACRYPT 2017, Proceedings, Part III. pp. 446–470 (2017). https://doi.org/10.1007/978-3-319-70700-6_16
36. Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. IACR Cryptol. ePrint Arch. p. 287 (2010), http://eprint.iacr.org/2010/287
37. Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. **28**(4), 321–338 (2017). https://doi.org/10.1007/S00200-017-0326-Y
38. Shen, Y., Sibleyras, F.: Key-reduced variants of 3kf9 with beyond-birthday-bound security. In: Agrawal, S., Lin, D. (eds.) Advances in Cryptology - ASIACRYPT 2022, Proceedings, Part II. Lecture Notes in Computer Science, vol. 13792, pp. 525–554. Springer (2022). https://doi.org/10.1007/978-3-031-22966-4_18
39. Sibleyras, F., Todo, Y.: Keyed sum of permutations: A simpler rp-based PRF. In: Rosulek, M. (ed.) Topics in Cryptology - CT-RSA 2023, Proceedings. Lecture Notes in Computer Science, vol. 13871, pp. 573–593. Springer (2023). https://doi.org/10.1007/978-3-031-30872-7_22
40. Taylor, R.: An integrity check value algorithm for stream ciphers. In: Stinson, D.R. (ed.) Advances in Cryptology - CRYPTO 1993. Proceedings. Lecture Notes in Computer Science, vol. 773, pp. 40–48. Springer (1993). https://doi.org/10.1007/3-540-48329-2_4
41. Yasuda, K.: The sum of CBC macs is a secure PRF. In: Pieprzyk, J. (ed.) Topics in Cryptology - CT-RSA 2010. Proceedings. Lecture Notes in Computer Science, vol. 5985, pp. 366–381. Springer (2010). https://doi.org/10.1007/978-3-642-11925-5_25
42. Yasuda, K.: A new variant of PMAC: beyond the birthday bound. In: Rogaway, P. (ed.) Advances in Cryptology - CRYPTO 2011. Proceedings. Lecture Notes in Computer Science, vol. 6841, pp. 596–609. Springer (2011). https://doi.org/10.1007/978-3-642-22792-9_34
43. Zhang, L., Wu, W., Sui, H., Wang, P.: 3kf9: Enhancing 3gpp-mac beyond the birthday bound. In: Wang, X., Sako, K. (eds.) Advances in Cryptology - ASIACRYPT 2012. Proceedings. Lecture Notes in Computer Science, vol. 7658, pp. 296–312. Springer (2012). https://doi.org/10.1007/978-3-642-34961-4_19

A Proof of Lemma 6

From (9), Lemma 3 and $t = 2$, we have

$$\begin{aligned}
\eta(\mathbb{S}_{\le i} \,|\, \mathcal{R}) &= \sum_{\boldsymbol{y}_{\le(i-1)}} \eta(\mathbb{S}_i \,|\, \mathcal{F}(\boldsymbol{y}_{\le(i-1)})) \\
&\ge \sum_{\boldsymbol{y}_{\le(i-1)}} \frac{\prod_{j=1}^{2}(N - f_{\le(i-1)}^{(j)})}{N} \left(1 + \frac{2}{N}\left(\mu(\boldsymbol{\lambda}_i, \hat{\mathcal{F}}) - \frac{f_{\le(i-1)}^{([2])}}{N}\right)\right) \\
&\ge \sum_{\boldsymbol{y}_{\le(i-1)}} \frac{\prod_{j=1}^{2}(N - f_{\le(i-1)}^{(j)})}{N} \Bigg(1 + \frac{2}{N}\Bigg(\mu(\boldsymbol{\lambda}_i, \hat{\mathcal{R}}_1, \hat{\mathcal{R}}_2) - \frac{s_1 s_2}{N} + \mu(\boldsymbol{\lambda}_i, \hat{\mathcal{R}}_1, \hat{\mathcal{P}}_2) - \frac{s_1(i-1)}{N} + \\
&\qquad \mu(\boldsymbol{\lambda}_i, \hat{\mathcal{P}}_1, \hat{\mathcal{R}}_2) - \frac{(i-1)s_2}{N} + \mu(\boldsymbol{\lambda}_i, \hat{\mathcal{P}}_1, \hat{\mathcal{P}}_2) - \frac{(i-1)^2}{N}\Bigg)\Bigg)
\end{aligned}$$

$$\geq \frac{\prod_{j=1}^{2}(N - f^{(j)}_{\leq(i-1)})}{N}\Bigg(\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{2}{N}\left|\mu_{\alpha_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R}_1,\mathcal{R}_2) - \frac{s_1 s_2}{N}\right|\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$$
$$- \frac{2}{N}\underbrace{\left(\frac{s_1(i-1)}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_1,\hat{\mathcal{P}}_2)\right)}_{\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_1,\hat{\mathcal{P}}_2)}$$
$$- \frac{2}{N}\underbrace{\left(\frac{(i-1)s_2}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{R}}_2)\right)}_{\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{R}}_2)}$$
$$- \frac{2}{N}\underbrace{\left(\frac{(i-1)^2}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{P}}_2)\right)}_{\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{P}}_2)}\Bigg) \tag{41}$$

Claim. We claim

$$\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_1,\hat{\mathcal{P}}_2) \leq \left(\frac{2(s+q)}{N}(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}}) + \frac{4(s+q)q^2}{N^2}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$$
$$\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{R}}_2) \leq \left(\frac{2(s+q)}{N}(\Delta_{\mathbb{S}} + \nabla_{\hat{\mathcal{R}}}) + \frac{4(s+q)q^2}{N^2}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$$
$$\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{P}}_2) \leq \left(\frac{2q\Delta_{\mathbb{S}}}{\mathbb{S}} + \frac{4q^3}{N^2}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$$

Proof. Let $\boldsymbol{x}_k^1, \boldsymbol{x}_k^2$ denote the variables corresponding to the k-th equations. Analogously, for $k \leq (i-1)$, we write y_k^1, y_k^2 to denote the solution assigned to the variables corresponding to the k-th equation. Let

$$\mathcal{M}_1 \coloneqq \left\{(x,j) \in \mathcal{R}_1 \times [i-1] : \boldsymbol{\lambda}_j \neq \boldsymbol{\lambda}_i,\, \boldsymbol{\lambda}_i \ominus \alpha_1 \cdot x \notin \hat{\mathcal{R}}_2,\, \boldsymbol{\lambda}_j \ominus \boldsymbol{\lambda}_i \oplus \alpha_1 \cdot x \notin \hat{\mathcal{R}}_1\right\}$$
$$\mathcal{M}_2 \coloneqq \left\{(j,x) \in [i-1] \times \mathcal{R}_2 : \boldsymbol{\lambda}_j \neq \boldsymbol{\lambda}_i,\, \boldsymbol{\lambda}_i \ominus \alpha_2 \cdot x \notin \hat{\mathcal{R}}_1,\, \boldsymbol{\lambda}_j \ominus \boldsymbol{\lambda}_i \oplus \alpha_2 \cdot x \notin \hat{\mathcal{R}}_2\right\}$$
$$\mathcal{M}_3 \coloneqq \{(j,j') \in [i-1]^2 : \boldsymbol{\lambda}_j, \boldsymbol{\lambda}_{j'} \neq \boldsymbol{\lambda}_i\}$$

Then, it is easy to see that

$$(i-\Delta_{\mathbb{S}})(s_1 - 2\nabla_{\hat{\mathcal{R}}}) \leq |\mathcal{M}_1| \leq i s_1$$
$$(i-\Delta_{\mathbb{S}})(s_2 - 2\nabla_{\hat{\mathcal{R}}}) \leq |\mathcal{M}_2| \leq i s_2$$
$$(i-\Delta_{\mathbb{S}})(i-\Delta_{\mathbb{S}} - 1) \leq |\mathcal{M}_3| \leq i^2$$

We prove the first and third inequality. As for the second inequality, it can be argued in a similar fashion as the first one.
First consider

$$\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_1,\hat{\mathcal{P}}_2) \leq \frac{s_1(i-1)}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_1,\hat{\mathcal{P}}_2)$$

$$\leq \frac{s_1(i-1)}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\leq(i-1)}}\sum_{(x,j)\in\mathcal{M}_1}\mu(\boldsymbol{\lambda}_i, \alpha_1\cdot x, \alpha_2\cdot y_j^2)$$
$$= \frac{s_1(i-1)}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \sum_{(x,j)\in\mathcal{M}_1}\sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i, \alpha_1\cdot x, \alpha_2\cdot y_j^2)$$

Thus, we have to count the number of solutions $\boldsymbol{y}_{\leq(i-1)}$ that additionally satisfies $\alpha_1\cdot x\oplus\alpha_2\cdot y_j^2 = \boldsymbol{\lambda}_i$ for all valid (x,j). Let $\mathbb{S}_{\leq(i-1)\setminus\{j\}}$ denote the system that excludes $\mathbb{S}_j$. Then, the second summand on the right hand side can be rewritten as

$$\sum_{(x,j)\in\mathcal{M}_1}\sum_{\boldsymbol{y}_{\leq(i-1)}}\mu(\boldsymbol{\lambda}_i, \alpha_1\cdot x, \alpha_2\cdot y_j^2) = \sum_{(x,j)\in\mathcal{M}_1}\eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}'(x,j)),$$

where $\mathcal{R}'(x,j) := (\mathcal{R}'_1(x,j), \mathcal{R}'_2(x,j))$, $\mathcal{R}'_1(x,j) := \mathcal{R}_1\cup\left\{\frac{\boldsymbol{\lambda}_j\ominus\boldsymbol{\lambda}_i\oplus\alpha_1\cdot x}{\alpha_1}\right\}$, and $\mathcal{R}'_2(x,j) := \mathcal{R}_2\cup\left\{\frac{\boldsymbol{\lambda}_i\ominus\alpha_1\cdot x}{\alpha_2}\right\}$. Using Fact 1, it is obvious that

$$\eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}'(x,j)) \leq \eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}),$$

since $(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}'(x,j)) \subseteq (\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R})$.

Now, a solution in $(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R})$ is not in $(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}'(x,j))$ if and only if there exists a $k\in[i-1]\setminus\{j\}$ such that $y_k^1 = \frac{\boldsymbol{\lambda}_j\ominus\boldsymbol{\lambda}_i\oplus\alpha_1\cdot x}{\alpha_1}$ or $y_k^2 = \frac{\boldsymbol{\lambda}_i\ominus\alpha_1\cdot x}{\alpha_2}$. Using Fact 2, such solutions are at most $2\eta(\mathbb{S}_{\leq(i-1)\setminus\{j,k\}}\,|\,\mathcal{R})$ for each k, where $\mathbb{S}_{\leq(i-1)\setminus\{j,k\}}$ denote the system excluding $\mathbb{S}_j$ and $\mathbb{S}_k$. Thus, we have

$$\begin{aligned}\eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}'(x,j)) &\geq \eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}) - 2\sum_{k\in[i-1]\setminus\{j\}}\eta(\mathbb{S}_{\leq(i-1)\setminus\{j,k\}}\,|\,\mathcal{R})\\ &\geq \eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R}) - 2\sum_{k\in[i-1]\setminus\{j\}}\frac{\eta(\mathbb{S}_{\leq(i-1)\setminus\{j\}}\,|\,\mathcal{R})}{N-2s-2(i-1)}\\ &\geq \frac{\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})}{N}\left(1-\frac{2i}{N-2s-2(i-1)}\right),\end{aligned}$$

where the second and third inequalities follow from Fact 2, and $s_1+s_2\leq 2s$. Finally, using $2(s+i-1)\leq 2^{n-1}$, we have

$$\begin{aligned}\varepsilon(\boldsymbol{\lambda}_i, \hat{\mathcal{R}}_1, \hat{\mathcal{P}}_2) &\leq \frac{s_1(i-1)}{N}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{|\mathcal{M}_1|}{N}\left(1-\frac{4i}{N}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})\\ &\leq \left(\frac{s+q}{N}(\varDelta_{\mathbb{S}}+2\nabla_{\hat{\mathcal{R}}}) + \frac{4(s+q)q^2}{N^2}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}),\end{aligned} \tag{42}$$

and similarly,

$$\varepsilon(\boldsymbol{\lambda}_i, \hat{\mathcal{P}}_1, \hat{\mathcal{R}}_2) \leq \left(\frac{s+q}{N}(\varDelta_{\mathbb{S}}+2\nabla_{\hat{\mathcal{R}}}) + \frac{4(s+q)q^2}{N^2}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}). \tag{43}$$

Now, consider

$$\begin{aligned}\varepsilon(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{P}}_2) &\le \frac{(i-1)^2}{N}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_1,\hat{\mathcal{P}}_2)\\ &\le \frac{(i-1)^2}{N}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \sum_{\boldsymbol{y}_{\le(i-1)}}\sum_{(j,j')\in\mathcal{M}_3}\mu(\boldsymbol{\lambda}_i,\alpha_1\cdot y_j^1,\alpha_2\cdot y_{j'}^2)\\ &= \frac{(i-1)^2}{N}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \sum_{(j,j')\in\mathcal{M}_3}\sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,\alpha_1\cdot y_j^1,\alpha_2\cdot y_{j'}^2)\end{aligned}$$

We want to count the number of solutions $\boldsymbol{y}_{\le(i-1)}$ that additionally satisfies $\alpha_1\cdot y_j^1\oplus\alpha_2\cdot y_{j'}^2=\boldsymbol{\lambda}_i$ for all valid (j,j'). Let $\mathbb{S}_{\le(i-1)\cup\{j-j'\}}$ denote the system $\mathbb{S}_{\le(i-1)}\cup\{\alpha_1\cdot\boldsymbol{x}_j^1\oplus\alpha_2\cdot\boldsymbol{x}_{j'}^2=\boldsymbol{\lambda}_i\}$. Then, we can rewrite the second summand on the right hand side as

$$\sum_{(j,j')\in\mathcal{M}_3}\sum_{\boldsymbol{y}_{\le(i-1)}}\mu(\boldsymbol{\lambda}_i,\alpha_1\cdot y_j^1,\alpha_2\cdot y_{j'}^2) = \sum_{(j,j')\in\mathcal{M}_3}\eta(\mathbb{S}_{\le(i-1)\cup\{j-j'\}}\,|\,\mathcal{R}).$$

Let $\mathbb{S}_{\le(i-1)\setminus\{j'\}}$ denote the system after excluding $\mathbb{S}_{j'}\cup\{\alpha_1\cdot\boldsymbol{x}_j^1\oplus\alpha_2\cdot\boldsymbol{x}_{j'}^2=\boldsymbol{\lambda}_i\}$. Then, it is obvious that

$$\eta(\mathbb{S}_{\le(i-1)\cup\{j-j'\}}\,|\,\mathcal{R}) \le \eta(\mathbb{S}_{\le(i-1)\setminus\{j'\}}\,|\,\mathcal{R})$$

Now, a solution in $(\mathbb{S}_{\le(i-1)\cup\{j-j'\}}\,|\,\mathcal{R})$ is not in $(\mathbb{S}_{\le(i-1)\setminus\{j'\}}\,|\,\mathcal{R})$ if and only if there exists a $k\in[i-1]\setminus\{j'\}$ such that $y_k^2=\frac{\boldsymbol{\lambda}_i\ominus\alpha_1\cdot y_j^1}{\alpha_2}$ or $y_k^1=\frac{\boldsymbol{\lambda}_{j'}\ominus\boldsymbol{\lambda}_i\oplus\alpha_1\cdot y_j^1}{\alpha_1}$. Such solutions are at most $2\eta(\mathbb{S}_{\le(i-1)\setminus\{j',k\}}\,|\,\mathcal{R})$ for each k, where $\mathbb{S}_{\le(i-1)\setminus\{j',k\}}$ denote the system excluding $\mathbb{S}_{j'}$ and $\mathbb{S}_k$. Thus, we have

$$\begin{aligned}\eta(\mathbb{S}_{\le(i-1)\cup\{j-j'\}}\,|\,\mathcal{R}) &\ge \eta(\mathbb{S}_{\le(i-1)\setminus\{j'\}}\,|\,\mathcal{R}) - 2\sum_{k\in[i-1]\setminus\{j'\}}\eta(\mathbb{S}_{\le(i-1)\setminus\{j',k\}}\,|\,\mathcal{R})\\ &\ge \eta(\mathbb{S}_{\le(i-1)\setminus\{j'\}}\,|\,\mathcal{R}) - 2\sum_{k\in[i-1]\setminus\{j'\}}\frac{\eta(\mathbb{S}_{\le(i-1)\setminus\{j'\}}\,|\,\mathcal{R})}{N-2s-2(i-1)}\\ &\ge \frac{\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R})}{N}\left(1-\frac{2i}{N-2s-2(i-1)}\right),\end{aligned}$$

where the second and third inequalities follow from Fact 2, and $s_1+s_2\le 2s$. Finally, we have

$$\begin{aligned}\varepsilon(\boldsymbol{\lambda}_i,\mathcal{P}_1,\mathcal{P}_2) &\le \frac{(i-1)^2}{N}\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) - \frac{|\mathcal{M}_3|}{N}\left(1-\frac{4i}{N}\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R})\\ &\le \left(\frac{2q\Delta_{\mathbb{S}}}{N}+\frac{4q^3}{N^2}\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}) \qquad (44)\end{aligned}$$

Lemma 6 now follows by combining (41) with (42)-(44). □

B Residual Calculations

We aim to show:

$$\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) - \frac{s^2}{2^n}\right| \le \frac{3s}{2^n} \tag{45}$$

$$\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right)} \le \frac{\sqrt{2}s}{2^{n/2}} + \frac{4s^2}{2^{3n/2}} \tag{46}$$

First consider $\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) - \frac{s^2}{2^n}\right|$. We need both lower and upper bounds on $\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right)$. Let $\mathcal{I} = \{i_1,\dots,i_s\}$ be an arbitrary indexing of $\mathcal{R}$.

For all $j,j' \in \mathcal{I}$, let $\mathbb{1}_{j,j'}$ denote the indicator random variable corresponding to the event $\mathtt{A}_j \oplus \mathtt{B}_{j'} = \mathsf{T}^{(i)}$, where $\mathtt{A}_j, \mathtt{B}_{j'} \in \mathcal{R}$. Then, we have

$$\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) = \sum_{j \ne j' \in \mathcal{I}} \Pr\left(\mathbb{1}_{j,j'}\right) = \frac{s(s-1)}{2^n-1}, \tag{47}$$

since for any pair of (j,j'), $\Pr\left(\mathbb{1}_{j,j'}\right) = 1/(2^n-1)$ and there are at most $s(s-1)$ such pairs. (45) now follows easily.

Now, consider the second claim. We have to compute the variance of $\mu(\mathsf{T}^{(i)},\mathcal{R})$. First, using the above formulation, we have

$$\begin{aligned}
\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) &= \mathbb{V}\left(\sum_{j,j' \in \mathcal{I}} \mathbb{1}_{j,j'}\right) \\
&= \sum_{j,j' \in \mathcal{I}} \mathbb{V}\left(\mathbb{1}_{j,j'}\right) + \sum_{\substack{j_1,j_2,j_3,j_4 \in \mathcal{I} \\ \{j_1,j_2\} \ne \{j_3,j_4\}}} \mathbb{V}\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) \\
&\le \sum_{j,j' \in \mathcal{I}} \mathbb{E}\left(\mathbb{1}_{j,j'}\right) + \sum_{\substack{j_1,j_2,j_3,j_4 \in \mathcal{I} \\ \{j_1,j_2\} \ne \{j_3,j_4\}}} \mathbb{V}\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) \\
&\le \mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{R})\right) + \sum_{\substack{j_1,j_2,j_3,j_4 \in \mathcal{I} \\ \{j_1,j_2\} \ne \{j_3,j_4\}}} \mathbb{V}\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right)
\end{aligned} \tag{48}$$

All that remains is to bound the covariances for every choice of $(j_1,j_2) \ne (j_3,j_4)$. First, we have

$$\mathbb{V}\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) = \Pr\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) - \Pr\left(\mathbb{1}_{j_1,j_2}\right)\Pr\left(\mathbb{1}_{j_3,j_4}\right)$$

It is easy to see that $\Pr\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) \le 1/(2^n-1)(2^n-3)$, and thus

$$\sum_{\substack{j_1,j_2,j_3,j_4 \in \mathcal{I} \\ \{j_1,j_2\} \ne \{j_3,j_4\}}} \mathbb{V}\left(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}\right) \le s^4\left(\frac{1}{(2^n-1)(2^n-3)} - \frac{1}{(2^n-1)^2}\right) \le \frac{16s^4}{2^{3n}}. \tag{49}$$

(46) now follows by taking square root on both sides of (48) after appropriate substitutions from (47) and (49).